Applied Mathematics and Computation 508 (2026) 129608

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Full Length Article

Generalizing the SVD of a matrix under nonstandard inner product and its applications to linear ill-posed problems

Haibo Li

School of Mathematics and Statistics, The University of Melbourne, VIC, 3010, Australia

ARTICLE INFO

Keywords:
Fredholm integral equation
Ill-posed problems
Weighted singular value decomposition
Subspace projection regularization
Weighted Golub-Kahan bidiagonalization
Weighted LSQR

ABSTRACT

The singular value decomposition (SVD) of a matrix is a powerful tool for many matrix computation problems. In this paper, we consider a generalization of the standard SVD to analyze and compute the regularized solution of linear ill-posed problems that arise from discretizing the first kind Fredholm integral equations. For the commonly used quadrature method for discretization, a regularizer of the form $\|x\|_M^2 := x^\top M x$ should be exploited, where M is symmetric positive definite. To handle this regularizer, we use the weighted SVD (WSVD) of a matrix under the M-inner product. Several important applications of the WSVD, such as low-rank approximation and solving the least squares problems with minimum $\|\cdot\|_M$-norm, are studied. We propose the weighted Golub-Kahan bidiagonalization (WGKB) to compute several dominant WSVD components and a corresponding weighted LSQR algorithm to iteratively solve the least squares problem. All the above tools and methods are used to analyze and solve linear ill-posed problems with the regularizer $\|x\|_M^2$. Several WGKB based iterative regularization and hybrid regularization methods are proposed to compute a good regularized solution, which can incorporate the prior information about x encoded by the regularizer $\|x\|_M^2$. Several numerical experiments are performed to illustrate the fruitfulness of our methods.

1. Introduction

The singular value decomposition (SVD) is a well-known matrix factorization tool for diagonalizing a matrix with arbitrary shape [42]. It generalizes the eigen-decomposition of a square normal matrix to any $m \times n$ matrix. Let $A \in \mathbb{R}^{m\times n}$, then there exist orthogonal matrices $\hat{U} = (\hat{u}_1, \ldots, \hat{u}_m) \in \mathbb{R}^{m\times m}$ and $\hat{V} = (\hat{v}_1, \ldots, \hat{v}_n) \in \mathbb{R}^{n\times n}$, such that

$$\hat{U}^\top A \hat{V} = \hat{\Sigma} \tag{1.1}$$

with

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_q \\ \mathbf{0} \end{pmatrix}, \ m \geq n \quad \text{or} \quad \hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_q & \mathbf{0} \end{pmatrix}, \ m < n, \tag{1.2}$$

where $q = \min\{m, n\}$ and $\hat{\Sigma}_q = \mathrm{diag}(\hat{\sigma}_1, \ldots, \hat{\sigma}_q) \in \mathbb{R}^{q\times q}$ is a diagonal matrix. The real values $\hat{\sigma}_1 \geq \cdots \geq \hat{\sigma}_q \geq 0$ are called singular values, and the corresponding vectors u_i and v_i are called right and left singular vectors, respectively.

E-mail address: haibo.li@unimelb.edu.au.

https://doi.org/10.1016/j.amc.2025.129608
Received 11 April 2025; Accepted 16 June 2025

Available online 26 June 2025

The applications of the SVD encompass a wide range of areas. The mathematical applications include determining the rank, range and null spaces of a matrix, computing the pseudoinverse of a matrix, determining a low-rank approximation of a matrix, solving linear least squares problems, solving discrete linear ill-posed problems, and many others [11,21,2,18]. We will review some of these applications in Section 2. Additionally, the SVD is highly useful in various areas of science and engineering, including signal processing, image processing, principal component analysis, control theory, recommender systems, and many others [41,38,22,24]. For a large-scale matrix A, a partial SVD can be computed by using the Golub-Kahan bidiagonalization (GKB), which generates two Krylov subspaces and projects A onto these two subspaces to get a small-scale bidiagonal matrix [16]. The bidiagonal structure of the projected matrix makes it convenient to develop efficient algorithms. For example, the dominant singular values and corresponding vectors of A can be well approximated by the SVD of the projected bidiagonal matrix [8]. It is also shown in [40] that a good low-rank approximation of A can be directly obtained from the GKB of A without directly computing any SVD. For large sparse least squares (LS) problem of the form $\min_{x\in\mathbb{R}^n}\|Ax-b\|_2$, one of the most commonly used LSQR solver is also based on GKB [35]. We remark that LSQR is mathematically equivalent to applying the conjugate gradient algorithm to the normal equation $A^\top A = A^\top b$.

In this paper, we focus on generalizing the SVD to analyze and compute the regularized solution of the first kind Fredholm equation [29]

$$g(s) = \int_{t_1}^{t_2} K(s,t)f(t)\mathrm{d}t + \sigma\dot{W}(s) \tag{1.3}$$

with $t\in[t_1,t_2]$ and $s\in[s_1,s_2]$, where $K(s,t)\in L^2\left([t_1,t_2]\times[s_1,s_2]\right)$ is a square-integrable kernel function, $\dot{W}(s)$ is the Gaussian white noise, more precisely, the generalized derivative of the standard Brownian motion $W(s)$ with $s\in[s_1,s_2]$ [26], σ is the scale of the noise, and $g(s)$ is the observation. The aim is to recover the unknown $f(t)$ from the noisy observation. To solve this problem, the first step is to discretize the above integral equation. Two commonly used discretization methods are the quadrature method and Galerkin method [28], and we focus on the first one in this paper. In the quadrature method, a quadrature rule with grid points $\{p_1,\dots,p_n\}$ and corresponding weights $\{w_1,\dots,w_n\}$ are chosen to approximate the integral as

$$\int_{t_1}^{t_2} K(s,t)f(t)\mathrm{d}t \approx \sum_{i=1}^{n} w_i K(s,p_i)f(p_i) \tag{1.4}$$

and the underlying unknown function we want to recover becomes the n-dimensional vector $x_{\text{true}} = (f(p_1),\dots,f(p_n))^\top$. The observations are chosen from m grid points in $[s_1,s_2]$ to get a m-dimensional noisy vector b. After discretization, the above integral equation (1.3) becomes

$$b = Ax + e, \tag{1.5}$$

where e is the noise vector in the discrete case, and b is a perturbed version of the unknown exact observation. We remark that for the discrete system, the noise comes from both the observational noise and the discretization error.

Since the integral operator T_K corresponding to the square-integrable kernel $K(s,t)$ is a linear compact operator, the singular values of T_K gradually decay to zero without a noticeable gap when T_K is not degenerate. As a result, after discretization, the matrix A becomes increasingly ill-conditioned as n increases, making the discrete linear system (1.5) ill-posed [21]. As a result, the naive solution to the LS problem $\min_{x\in\mathbb{R}^n}\|Ax-b\|_2$ will deviate very far from the true solution x_{true}. Tikhonov regularization is usually used to handle the ill-posedness property of this problem [43], where a specific regularization strategy depends both on the noise type and the prior information about the true solution. For example, for a Gaussian white noise e, the standard-form Tikhonov regularization has the form

$$\min_{x\in\mathbb{R}^n}\|Ax-b\|_2^2 + \lambda\|x\|_2^2, \tag{1.6}$$

where the 2-norm regularization term enforces some smoothness on the solution. However, we emphasize that the regularization term $\|x\|_2^2$ arises from the discretized version of $\|f\|_{L^2([t_1,t_2])}$; here $L^2([t_1,t_2])$ is the space of square-integrable functions defined on $[t_1,t_2]$ with Lebesgue measure. If all the weights in the quadrature rule for the integral in (1.4) have the same value, which corresponds to the midpoint rule for numerical integral, then the vector norm $\|x\|_2$ is a good approximation to $\|f\|_{L^2([t_1,t_2])}$. However, the midpoint rule has relatively lower accuracy, which results in a higher discretization error and amplifies the solution error due to the ill-posedness of the inverse problem. Using other quadrature rules with higher accuracy, such as the Simpson's rule or Gaussian quadrature, the weights will have different values, and a good approximation to $\|f\|_{L^2([t_1,t_2])}$ should be $\|x\|_M$ instead of $\|x\|_2$, where $\|x\|_M = (x^\top Mx)^{1/2}$ with $M = \mathrm{diag}(w_1,\dots,w_n)$. In this case, the standard-form Tikhonov regularization should be replaced by

$$\min_{x\in\mathbb{R}^n}\|Ax-b\|_2^2 + \lambda\|x\|_M^2. \tag{1.7}$$

Using the standard SVD of A, we can analyze and compute the ill-posed problem (1.5) with the regularizer $\|x\|_2^2$. For example, the naive solution to $\min_{x\in\mathbb{R}^n}\|Ax-b\|_2$ can be expressed by the SVD expansion form $x_{LS} = A^\dagger b = \sum_{i=1}^{n}\frac{\hat{u}_i^\top b}{\hat{\sigma}_i}\hat{v}_i$, where "$\dagger$" is the Moore-Penrose pseudoinverse [18, 5.5.2]. This motivates the so-called "truncated SVD" (TSVD) regularization method, where the regularized

solution is formed as $x_k = \sum_{i=1}^{k} \frac{\hat{u}_i^\top b}{\hat{\sigma}_i} \hat{v}_i$ by truncating the first k dominant SVD expansion from x_{LS} to discard those highly amplified noisy components [19]. Moreover, the Tikhonov regularized solution to (1.6) can be expressed as the filtered SVD expansion form $x_\lambda = \sum_{i=1}^{n} \frac{\hat{\sigma}_i^2}{\hat{\sigma}_i^2+\lambda} \frac{\hat{u}_i^\top b}{\hat{\sigma}_i} \hat{v}_i$, which tells us that the regularization parameter λ should be chosen such that the filter factors $f_i := \frac{\hat{\sigma}_i^2}{\hat{\sigma}_i^2+\lambda} \approx 1$ for small index i and $f_i \approx 0$ for large i to suppress noisy components. For large-scale ill-posed problems, the GKB process can be exploited to design iterative regularization algorithms [34,1]. The LSQR solver with early stopping rules is a standard algorithm to handle the regularizer $\|x\|_2^2$. Specifically, the regularization property of LSQR with early stopping rules has been analyzed using the SVD of A, where the k-th LSQR solution also has a filtered SVD expansion form [21].

To analyze the ill-posed problem (1.5) with the regularizer $\|x\|_M^2$, we use the generalization of the standard SVD of A under a nonstandard inner product. Specifically, the usual 2-inner product in $\mathbb{R}^n$ is replaced by the M-inner product $\langle z, z' \rangle := z^\top M z'$. The core idea is to treat the matrix $A \in \mathbb{R}^{m\times n}$ as a linear operator between the two finite-dimensional Hilbert spaces $\left(\mathbb{R}^n, \langle\cdot,\cdot\rangle_M\right)$ and $\left(\mathbb{R}^m, \langle\cdot,\cdot\rangle_2\right)$. The attempt to generalize SVD is not new, which can go back to Van Loan in 1976 [44], where he proposed the generalized SVD (GSVD) for a matrix pair. The GSVD is a powerful tool to analyze the general-form Tikhonov regularization term $\|Lx\|_2^2$, where $L \in \mathbb{R}^{p\times n}$ with $p \le n$ is a regularization matrix. In recent years, there have been several other generalized forms of the SVD, such as the weighted SVD of different forms proposed in [39,30,25]. Specifically, in [25], the authors used the weighted SVD to solve the discrete ill-posed problem arising from (1.3) when $f(t)$ and $g(s)$ are discretized on the same grid points with $[t_1, t_2] = [s_1, s_2]$ and $m = n$. In their paper, both the data fidelity term and regularization term use the weighted M-norm. However, we emphasize that for the discrete observation vector b with Gaussian white noise, the most appropriate form for the data fidelity term should be $\|Ax - b\|_2^2$.

In this paper, we consider a new generalization of the SVD, investigate its properties and propose numerical algorithms for its computation. Then we use this generalized SVD to investigate the solution of the linear least squares problems with minimum $\|x\|_M$ norm and propose an iterative algorithm for this problem. This iterative algorithm with proper early stopping rules can be used to handle the discrete ill-posed problem with regularization term $\|x\|_M^2$. In contrast to the work of [39,30], which focuses on the theoretical foundation of the WSVD, our results place greater emphasis on its applications and related computations, particularly in the context of ill-posed inverse problems.

The main contributions of this paper are listed as follows.

- For any symmetric positive definite matrix $M \in \mathbb{R}^{n\times n}$, we recall a generalization of the SVD of $A \in \mathbb{R}^{m\times n}$ called the weighted SVD (WSVD), where the right singular vectors constitute an M-orthonormal basis of $\left(\mathbb{R}^n, \langle\cdot,\cdot\rangle_M\right)$. We study several of its properties and applications, including a new form of low-rank approximation of A based on the WSVD and the WSVD form expression of the minimum $\|x\|_M$-norm solution to $\min_{x\in\mathbb{R}^n} \|Ax - b\|_2$.
- We propose a weighted Golub-Kahan bidiagonalization (WGKB) process, which can be used to compute several dominant WSVD components, namely the singular values and singular vectors corresponding to the WSVD. A WGKB-based iterative algorithm for solving the least squares problems is also proposed. It is a weighted form corresponding to the standard LSQR algorithms, thereby we name it weighted LSQR (WLSQR).
- Using the WSVD, we analyze the solution of the Tikhonov regularization problem (1.7). In order to utilize the information from the regularizer $\|x\|_M^2$ and avoid selecting λ in advance, we propose the subspace projection regularization (SPR) method. We show that the WGKB-based SPR method is just the WLSQR with early stopping rules, which can efficiently compute a satisfied regularized solution. To overcome the semi-convergence of the WGKB-based SPR method, we also propose a WGKB-based hybrid regularization method.

In the numerical experiments, we use both 1-dimensional and 2-dimensional first kind Fredholm equations to test our WGKB-based iterative regularization and hybrid regularization algorithms. With the Simpson's rule for discretizing the integral, we show that the proposed algorithms have better performance than the standard LSQR solver for regularizing the original problem. We remark that if we use the Galerkin method with a group of orthogonal basis functions for $f(t)$ to discretize (1.4), then we will also get a diagonal matrix M. However, if the basis functions are not orthogonal, then M can be a dense matrix, and in this case, the methods presented in this paper would face the challenge of computing M^{-1}.

The paper is organized as follows. In Section 2 we review some basic properties and applications of the SVD. In Section 3, we propose the weighted SVD of A with weight matrix M and investigate its properties. In Section 4, we propose the WGKB process to compute several dominant WSVD components and the WLSQR algorithms for solving least squares problems. All the above methods are used in Section 5 to analyze and develop both iterative regularization and hybrid regularization methods for linear ill-posed problems with regularizer $\|x\|_M^2$. In Section 6, we use several numerical examples to illustrate the effectiveness of the new methods. Finally, we conclude the paper in Section 7.

Throughout the paper, denote by $\mathcal{R}(C)$ and $\mathcal{N}(C)$ the range space and null space of a matrix C, by I_k the k-by-k identity matrix. We denote by $\mathbf{0}$ a zero vector or matrix with its order clear from the context.

2. Properties of the SVD and its applications

In this section, we review several important properties of the SVD and its applications to linear ill-posed problems. They motivate us to generalize the SVD to the nonstandard inner product case, which can be applied to handle the regularizer of the form (1.7).

Let us start from the connections between the SVD, Schatten p-norm [23, §7.4.7], and low-rank approximation of a matrix [18, §2.4].

Theorem 2.1 *(Schatten p-norm). Suppose the SVD of a matrix A is as in (1.1). For any integer $1 \le p \le \infty$, define*

$$|||A|||_p = \begin{cases} \left(\sum_{i=1}^{\min\{m,n\}} \hat{\sigma}_i^p\right)^{1/p}, & p \neq \infty \\ \hat{\sigma}_1, & p = \infty. \end{cases}$$

Then $|||\cdot|||_p$ is a matrix norm on $\mathbb{R}^{m\times n}$, called the Schatten p-norm.

The specific choice of p yields several commonly used matrix norms:

1. $p = 1$: gives the nuclear norm. It is commonly used in low-rank matrix completion algorithms.
2. $p = 2$: gives the Frobenius norm (often denoted by $\|\cdot\|_F$).
3. $p = \infty$: gives the spectral norm (often denoted by $\|\cdot\|_2$). Note that $|||A|||_\infty = \lim_{p\to\infty} |||A|||_p$.

All the Schatten norms are unitarily invariant, which means that $|||A|||_p = |||\bar{U}^\top A \bar{V}|||_p$ for any matrix A and all unitary matrices $\bar{U}$ and $\bar{V}$. In this paper, we focus on the spectral norm and use the popular notation $\|\cdot\|_2$ to denote it. From another definition of the spectral norm, we also have $\|A\|_2 = \max_{x\neq \mathbf{0}} \frac{\|Ax\|_2}{\|x\|_2} = \hat{\sigma}_1$.

One of the reasons that the SVD is so widely used is that it can be used to find the best low-rank approximation to a matrix. The following low-rank approximation property is often used such as in data compression, image compression and recommender systems.

Theorem 2.2 *(Eckhart-Young-Mirsky). Suppose $k < \operatorname{rank}(A) = r$ and let*

$$\hat{A}_k = \hat{U}_k \hat{\Sigma}_k \hat{V}_k^\top = \sum_{i=1}^{k} \hat{\sigma}_i \hat{u}_i \hat{v}_i^\top, \tag{2.1}$$

where $\hat{U}_k$ and $\hat{V}_k$ contain the first k columns of $\hat{U}$ and $\hat{V}$, and $\hat{\Sigma}_k$ is the top-left $k\times k$ submatrix of $\hat{\Sigma}$. Then

$$\min_{\operatorname{rank}(X)\le k} \|A - X\|_2 = \|A - \hat{A}_k\|_2 = \hat{\sigma}_{k+1}. \tag{2.2}$$

The Moore-Penrose pseudoinverse of a matrix is the most widely known generalization of the inverse matrix [18, 5.5.2]. Using the SVD of A, we can give its explicit form:

$$A^\dagger = \hat{V}\hat{\Sigma}^\dagger \hat{U}^\top, \quad \hat{\Sigma}^\dagger = \begin{pmatrix} \hat{\Sigma}_r^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{2.3}$$

One application of the Moore-Penrose pseudoinverse is to solve the rank-deficient least squares problems [18, §5.5.1]. In the rest part of the paper, for any matrix $C \in \mathbb{R}^{k\times l}$, vector $d \in \mathbb{R}^k$ and subspace $\mathcal{V} \subset \mathbb{R}^l$, we use $\{x \in \mathcal{V} : x \in \operatorname{argmin} \|Cx - d\|_2\}$ to denote the set of all $x \in \mathcal{V}$ satisfying $\|Cy - d\|_2 \ge \|Cx - d\|_2$ for any $y \in \mathcal{V}$.

Theorem 2.3. *Let $\operatorname{rank}(A) = r$ (can be smaller than $\min\{m,n\}$). Then the rank-deficient least squares problem with a minimum 2-norm*

$$\min_{x\in S} \|x\|_2, \quad S = \{x \in \mathbb{R}^n : x \in \operatorname{argmin} \|Ax - b\|_2\} \tag{2.4}$$

has a unique solution

$$x_{LS} = A^\dagger b = \sum_{i=1}^{r} \frac{\hat{u}_i^\top b}{\hat{\sigma}_i} \hat{v}_i. \tag{2.5}$$

For linear ill-posed problem with Tikhonov regularization (1.6), the regularized solution has a similar expression to (2.5) but with additional filter factors [21, §4.2]:

$$x_\lambda = \sum_{i=1}^{n} \frac{\hat{\sigma}_i^2}{\hat{\sigma}_i^2 + \lambda} \frac{\hat{u}_i^\top b}{\hat{\sigma}_i} \hat{v}_i, \tag{2.6}$$

where the regularization parameter λ should be chosen such that the filter factors $f_i := \frac{\hat{\sigma}_i^2}{\hat{\sigma}_i^2+\lambda} \approx 1$ for those small index i corresponding to dominant information about x_{true} and $f_i \approx 0$ for those large index i to filter out noisy components. Note that the summation index i runs from 1 to n, as the matrix A for the ill-posed problem does not have a well-defined numerical rank.

For small-scale matrices, the SVD can be efficiently computed by a variant of QR algorithm [16] or Jacobi rotation procedure [9,10]. For large-scale matrices, one commonly used SVD algorithm is based on the Golub-Kahan bidiagonalization (GKB), which applies a Lanczos-type iterative procedure to A to generate two Krylov subspaces and project A to be a small-scale bidiagonal matrix. Then the SVD of the reduced bidiagonal matrix is computed to approximate some dominant SVD components of A [13]. The GKB process is also a standard procedure used in LSQR for iteratively solving large-scale least square problems. At the k-th step of LSQR, it equivalently computes the solution

$$x_k = \min_{x\in S_k} \|Ax-b\|_2, \quad S_k = \mathcal{K}_k(A^\top A, A^\top b) := \operatorname{span}\{(A^\top A)^i, A^\top b\}_{i=0}^{k-1}.$$

The above LS problem can be transformed into a k-dimensional subproblem using the subspace projection procedure. For large-scale linear ill-posed problems, the above approach is very efficient and fruitful for handling the $\|x\|_2^2$ regularization term, where an early stopping rule should be used to avoid the semi-convergence behavior, meaning that the error $\|x_k - x_{\text{true}}\|_2$ first decreases and then increases as k runs from 1 to n. The hybrid regularization method is another type of iterative method that can stabilize the convergence behavior, which usually applies the Tikhonov regularization to the projected subproblem at each iteration; see e.g. [7,27,37,6].

If all the weights in the quadrature rule for the integral in (1.4) are equal, as in the case of the midpoint rule, the vector norm $\|x\|_2$ provides a reasonable approximation to $\|f\|_{L^2([t_1,t_2])}$. However, the midpoint rule tends to have lower accuracy, leading to higher discretization errors and an increased amplification of solution errors due to the ill-posed nature of the inverse problem. In contrast, using quadrature rules with higher accuracy, such as the Simpson's rule or Gaussian quadrature, results in different weight values. In this case, the more accurate approximation to $\|f\|_{L^2([t_1,t_2])}$ should be $\|x\|_M$ instead of $\|x\|_2$, where $M = \operatorname{diag}(w_1,\dots,w_n)$ represents the diagonal matrix of weights. In such a scenario, we should consider the Tikhonov regularization problem (1.6) and the corresponding iterative regularization methods. To analyze and solve this problem, in the following part, we generalize the SVD to the M-inner product case and propose corresponding iterative algorithms.

3. The weighted SVD with nonstandard inner-product

In this section, assume $A\in\mathbb{R}^{m\times n}$ be a matrix of rank $r<\min\{m,n\}$. Let $M\in\mathbb{R}^{n\times n}$ be a symmetric positive definite matrix. It can be either diagonal or non-diagonal. This matrix can introduce a new inner product in $\mathbb{R}^n$.

Definition 3.1. For any $A\in\mathbb{R}^{m\times n}$, define the linear operator $\mathcal{A}:(\mathbb{R}^n,\langle\cdot,\cdot\rangle_M)\to(\mathbb{R}^m,\langle\cdot,\cdot\rangle_2)$ as $\mathcal{A}: x\mapsto Ax$ for $x\in\mathbb{R}^n$ under the canonical bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, where $\langle\cdot,\cdot\rangle_2$ is the standard 2-inner product and $\langle x,x'\rangle_M := x^\top M x'$ is called the M-inner product.

The operator $\mathcal{A}$ is bounded since $(\mathbb{R}^n,\langle\cdot,\cdot\rangle_M)$ and $(\mathbb{R}^m,\langle\cdot,\cdot\rangle_2)$ are both finite dimensional Hilbert spaces. Thus, we can definite the operator norm of $\mathcal{A}$.

Definition 3.2. Define the M-weighted norm of A as

$$\|A\|_{M,2} := \|\mathcal{A}\| := \max_{x\neq\mathbf{0}} \frac{\|Ax\|_2}{\|x\|_M}. \tag{3.1}$$

Similar to the unitarily invariant property of $\|A\|_2$, we have the following property for $\|A\|_{M,2}$.

Proposition 3.1. *Let $\tilde U\in\mathbb{R}^{m\times m}$ and $\tilde V\in\mathbb{R}^{n\times n}$ are 2- and M-orthogonal matrices, respectively, i.e. $\tilde U^\top\tilde U = I_m$ and $\tilde V^\top M\tilde V = I_n$. Then we have*

$$\|\tilde U^\top A\tilde V\|_2 = \|A\|_{M,2} \tag{3.2}$$

Proof. Since $\tilde V^\top M\tilde V = I_n$ and M is positive definite, it follows that $\tilde V$ is invertible. Thus, we have

$$\begin{aligned}\|\tilde U^\top A\tilde V\|_2 &= \max_{x\neq\mathbf{0}} \frac{\|\tilde U^\top A\tilde V x\|_2}{\|x\|_2} = \max_{x\neq\mathbf{0}} \frac{\|\tilde U^\top A\tilde V x\|_2}{\|\tilde V x\|_M}\\ &= \max_{y\neq\mathbf{0}} \frac{\|\tilde U^\top A y\|_2}{\|y\|_M} = \max_{y\neq\mathbf{0}} \frac{\|Ay\|_2}{\|y\|_M} \quad (\text{let } y=\tilde V x)\\ &= \|A\|_{M,2},\end{aligned}$$

where we have used $\|\tilde V x\|_M^2 = x^\top\tilde V^\top M\tilde V x = x^\top x = \|x\|_2^2$. The proof is completed. □

The following result generates the SVD of a matrix A, which has a very similar form to (1.1).

Theorem 3.1. *Let $A\in\mathbb{R}^{m\times n}$, and $M\in\mathbb{R}^{n\times n}$ is symmetric positive definite. There exist 2-orthogonal matrix $U\in\mathbb{R}^{m\times m}$ and M-orthogonal matrix $V\in\mathbb{R}^{n\times n}$, and diagonal matrix $\Sigma_r = \operatorname{diag}(\sigma_1,\dots,\sigma_r)$ with $\sigma_1\geq\cdots\geq\sigma_r>0$, such that*

$$U^\top AV = \Sigma := \begin{pmatrix} \Sigma_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \Sigma \in \mathbb{R}^{m\times n}. \tag{3.3}$$

Proof. By the definition of $\|A\|_{M,2}$, there exist vectors $v_1 \in \mathbb{R}^n$ and $u_1 \in \mathbb{R}^m$ such that $\|v_1\|_M = \|u_1\|_2 = 1$ and $Av_1 = \sigma_1 u_1$ with $\sigma_1 = \|A\|_{M,2}$. Let $V_2 \in \mathbb{R}^{n\times(n-1)}$ and $U_2 \in \mathbb{R}^{m\times(m-1)}$ such that $V = (v_1, V_2)$ and $U = (u_1, U_2)$ are M- and 2-orthogonal, respectively. Then we get

$$U^\top AV = \begin{pmatrix} u_1^\top Av_1 & u_1^\top AV_2 \\ U_2^\top Av_1 & U_2^\top AV_2 \end{pmatrix} =: \begin{pmatrix} \sigma_1 & x^\top \\ \mathbf{0} & B \end{pmatrix} =: A_1,$$

where $x \in \mathbb{R}^{n-1}$ and $B \in \mathbb{R}^{(m-1)\times(n-1)}$. By Lemma 3.1 we have $\|A_1\|_2 = \|A\|_{M,2} = \sigma_1$. Let $\tilde{x} = (\sigma_1, x^\top)^\top$. It follows that

$$\sigma_1^2 \geq \frac{\|A_1\tilde{x}\|_2^2}{\|\tilde{x}\|_2^2} = \left\| \begin{pmatrix} \sigma_1^2 + x^\top x \\ Bx \end{pmatrix} \right\|_2^2 \Big/ \|\tilde{x}\|_2^2 \geq \|\tilde{x}\|_2^2 = \sigma_1^2 + x^\top x,$$

which leads to $x = \mathbf{0}$. Therefore, we have $U^\top AV = \begin{pmatrix} \sigma_1 & \mathbf{0}^\top \\ \mathbf{0} & B \end{pmatrix}$. Now (3.3) can be obtained by using mathematical induction. Since U and V are invertible, it follows that $\operatorname{rank}(A) = r = \operatorname{rank}(\Sigma_r)$. The proof is completed. □

The main difference between the two forms (1.1) and (3.3) is that the right vectors $\{v_i\}$ are M-orthonormal. We call (3.3) the *weighted SVD* (WSVD) of A with weight matrix M. For $M = I_n$, it is the same as the standard SVD. Similar to the standard SVD, the WSVD can be used to analyze and solve many problems with a nonstandard 2-norm. Specifically, it can be used to analyze and develop efficient algorithms for linear ill-posed problems with the $\|x\|_M^2$ regularization term.

Note that $V^\top MV = I_n$ implies that $VV^\top MV = V$. Multiplying V^{-1} from the right-hand side, we get $VV^\top = M^{-1}$. Therefore, from (3.3) we get

$$A = U \begin{pmatrix} \Sigma_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} V^\top M. \tag{3.4}$$

From (3.3) and (3.4) we have

$$Av_i = \sigma_i u_i, \tag{3.5}$$

$$A^\top u_i = \sigma_i M v_i. \tag{3.6}$$

Also, we have the WSVD expansion form of A: $A = \sum_{i=1}^r \sigma_i u_i v_i^\top M$. Besides, the range space and null space of A can be explicitly written as

$$\mathcal{R}(A) = \operatorname{span}\{u_1, \ldots, u_r\},$$

$$\mathcal{N}(A) = \operatorname{span}\{v_{r+1}, \ldots, v_n\},$$

where $\{v_{r+1}, \ldots, v_n\}$ is an M-orthonormal basis of $\mathcal{N}(A)$.

Using the WSVD, the Eckhart-Young-Mirsky theorem for low-rank approximation of a matrix under the $\|\cdot\|_{M,2}$ norm has the following form.

Theorem 3.2. *For any integer $1 \leq k < r$, we have*

$$\min_{X\leq k} \|A - X\|_{M,2} \geq \sigma_{k+1}, \tag{3.7}$$

where the minimum can be achieved if $X = A_k := \sum_{i=1}^k \sigma_i u_i v_i^\top M$.

Proof. First, if $X = A_k := \sum_{i=1}^k \sigma_i u_i v_i^\top M$, we have $\operatorname{rank}(X) = k$, and by Proposition 3.1 we have

$$\|A - X\|_{M,2} = \|U^\top(A - X)V\|_2 = \left\| U^\top U \begin{pmatrix} \mathbf{0} & & & \\ & \sigma_{k+1} & & \\ & & \ddots & \\ & & & \sigma_r \\ & & & \mathbf{0} \end{pmatrix} V^\top MV \right\|_2 = \sigma_{k+1}.$$

Thus, we only need to prove $\|A - X\|_{M,2} \geq \sigma_{k+1}$ for any $X \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(X) = k$. For such X we have $\dim(\mathcal{N}(X)) = n - k$, thereby we can find M-orthonormal vectors $\{w_1, \ldots, w_{n-k}\}$ such that $\mathcal{N}(X) = \operatorname{span}\{w_1, \ldots, w_{n-k}\}$. Notice that $\mathcal{N}(X) \cap \operatorname{span}\{v_1, \ldots, v_{k+1}\} \neq \{\mathbf{0}\}$ since the sum of dimensions of these two subspaces is $n + 1$. Let z be a nonzero vector in the intersection of the above two subspaces and $\|z\|_M = 1$. Using the WSVD of A, we get

$$Az = \sum_{i=1}^{r} \sigma_i u_i (v_i^\top M z) = \sum_{i=1}^{k+1} \sigma_i u_i (v_i^\top M z),$$

since z is M-orthogonal to $v_{k+2}, \dots, v_n$. It follows that

$$\|A - X\|_{M,2}^2 \geq \frac{\|(A-X)z\|_2^2}{\|z\|_M^2} = \|Az\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 (v_i^\top M z)^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} (v_i^\top M z)^2.$$

Since $\|z\|_M^2 = z^\top M z = z^\top M V V^\top M z = \|V^\top M z\|_2^2$, where we used $VV^\top = M^{-1}$, we have

$$\sum_{i=1}^{k+1} (v_i^\top M z)^2 = \sum_{i=1}^{n} (v_i^\top M z)^2 = \|V^\top M z\|_2^2 = \|z\|_M^2 = 1.$$

We finally obtain $\|A - X\|_{M,2} \geq \sigma_{k+1}$. □

For the rank-deficient least squares problem, we can write the solution set by using the WSVD, which is convenient to find the unique minimum $\|\cdot\|_M$ solution.

Theorem 3.3. *For the linear least squares problems* $\min_{x\in\mathbb{R}^n} \|Ax - b\|_2$*, the set of all solutions is*

$$\mathcal{X} = \sum_{i=1}^{r} \frac{u_i^\top b}{\sigma_i} v_i + \operatorname{span}\{v_{r+1}, \dots, v_n\}, \tag{3.8}$$

and the unique solution to

$$\min_{x\in\mathcal{X}} \|x\|_M, \quad \mathcal{X} = \{x \in \mathbb{R}^n : x \in \operatorname{argmin} \|Ax - b\|_2\} \tag{3.9}$$

is

$$x_* = \sum_{i=1}^{r} \frac{u_i^\top b}{\sigma_i} v_i \tag{3.10}$$

Proof. Write U and V as $U = (U_r, U_{r,\perp})$ and $V = (V_r, V_{r,\perp})$. Using the WSVD of A, we have

$$\begin{aligned} \|Ax - b\|_2^2 &= \left\| U \begin{pmatrix} \Sigma_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} V_r^\top \\ V_{r,\perp}^\top \end{pmatrix} Mx - b \right\|_2^2 = \left\| \begin{pmatrix} \Sigma_r V_r^\top M x \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} U_r^\top b \\ U_{r,\perp}^\top b \end{pmatrix} \right\|_2^2 \\ &= \|\Sigma_r V_r^\top M x - U_r^\top b\|_2^2 + \|U_{r,\perp}^\top b\|_2^2. \end{aligned}$$

Therefore, the minimizers of $\min_{x\in\mathbb{R}^n} \|Ax - b\|_2$ are the solutions to $\Sigma_r V_r^\top M x = U_r^\top b$, which is equivalent to $V_r^\top M x = \Sigma_r^{-1} U_r^\top b$. An obvious solution to the above equation is $x_* = V_r \Sigma_r^{-1} U_r^\top b = \sum_{i=1}^{r} \frac{u_i^\top b}{\sigma_i} v_i$. Since $\mathcal{N}(V_r^\top M) = \operatorname{span}\{v_{r+1}, \dots, v_n\}$, we have the expression of $\mathcal{X}$ as (3.8).

On the other hand, for any $x \in \mathcal{X}$ such that $x = x_* + \sum_{r+1}^{n} \gamma_i v_i$, since v_i are mutual M-orthogonal, we have

$$\|x\|_M^2 = \|x_*\|_M^2 + \sum_{i=r+1}^{n} \gamma_i^2 \geq \|x_*\|_M^2,$$

where the equality on the right holds if and only if $\gamma_{r+1} = \cdots = \gamma_n = 0$. Therefore (3.9) has the unique solution x_*. □

If we let $A^{\dagger_M} = V \Sigma^\dagger U^\top$, where $\Sigma^\dagger := \begin{pmatrix} \Sigma_r^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{n\times m}$. Then we can express x_* as $x_* = A^{\dagger_M} b$. This is a similar expression to the smallest 2-norm solution to $\min_{x\in\mathbb{R}^n} \|Ax - b\|_2$. But unfortunately, $A^{\dagger_M}$ is not a real pseudoinverse of a matrix since $(AA^{\dagger_M})^\top \neq AA^{\dagger_M}$. Thus, we do not discuss it anymore.

4. Iterative algorithm for the WSVD and applications

The WSVD of A is actually the singular value expansion of the linear compact operator $\mathcal{A} : (\mathbb{R}^n, \langle\cdot,\cdot\rangle_M) \to (\mathbb{R}^m, \langle\cdot,\cdot\rangle_2)$ that has a finite rank. This motivates us to apply the GKB process to $\mathcal{A}$ to approximate several dominant WSVD components of A; see [5] for the GKB process for compact linear operators.

Starting from a nonzero vector $b \in \mathbb{R}^m$,[1] the GKB process proceeds based on the following recursive relations:

$$\beta_1 p_1 = b, \tag{4.1}$$

$$\alpha_i q_i = \mathcal{A}^* p_i - \beta_i q_{i-1}, \tag{4.2}$$

$$\beta_{i+1} p_{i+1} = \mathcal{A} q_i - \alpha_i p_i, \tag{4.3}$$

where $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$. The iteration proceeds as $i = 1, 2, \dots$, and we set $q_0 := \mathbf{0}$. From the definition of $\mathcal{A}$ we have $\mathcal{A} q_i = A q_i$. In order to compute $\mathcal{A}^*$, we use the basic relation

$$\langle \mathcal{A}x, y \rangle_2 = \langle x, \mathcal{A}^* y \rangle_M$$

which is equivalent to $(Ax)^\top y = x^\top \mathcal{A}^* M y$ for any vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. It follows that $\mathcal{A}^* = M^{-1} A^T$. Therefore, we obtain the GKB process of $\mathcal{A}^*$, as summarized in Algorithm 1. We name it as the weighted GKB process with weight matrix M.

Algorithm 1 The k-step weighted GKB (WGKB).

Input: Matrix $A \in \mathbb{R}^{m\times n}$, positive definite $M \in \mathbb{R}^{n\times n}$, nonzero $b \in \mathbb{R}^m$
Output: $\{\alpha_i, \beta_i\}_{i=1}^{k+1}$, $\{p_i, q_i\}_{i=1}^{k+1}$
Let $\beta_1 = \|b\|_2$, $p_1 = b/\beta_1$
Compute $\bar{s} = A^\top p_1$, $s = M^{-1}\bar{s}$
Compute $\alpha_1 = (s^\top \bar{s})^{1/2}$, $q_1 = s/\alpha_1$
for $i = 1, 2, \dots, k$ **do**
  $r = A q_i - \alpha_i p_i$
  $\beta_{i+1} = \|r\|$, $p_{i+1} = r/\beta_{i+1}$
  $\bar{s} = A^\top p_{i+1} - \beta_{i+1} M q_i$, $s = M^{-1}\bar{s}$
  $\alpha_{i+1} = (s^\top \bar{s})^{1/2}$, $q_{i+1} = s/\alpha_{i+1}$
end for

Using the property of GKB for the compact operator $\mathcal{A}$, before the WGKB reaches the termination step, that is, $k_t := \max_i\{\alpha_i\beta_i > 0\}$, the k-step WGKB process generates two groups of vectors $\{p_1, \dots, p_{k+1}\}$ and $\{q_1, \dots, q_{k+1}\}$ that are 2- and M-orthornormal, respectively. If we let $P_{k+1} = (p_1, \dots, p_{k+1})$, $Q_k = (p_1, \dots, q_k)$ and

$$B_k = \begin{pmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \beta_3 & \ddots & \\ & & \ddots & \alpha_k \\ & & & \beta_{k+1} \end{pmatrix} \in \mathbb{R}^{(k+1)\times k}, \tag{4.4}$$

then we have

$$A Q_k = P_{k+1} B_k, \tag{4.5}$$

$$M^{-1} A^\top P_{k+1} = Q_k B_k^\top + \alpha_{k+1} q_k e_{k+1}^\top, \tag{4.6}$$

where e_{k+1} is the $(k+1)$-th column of I_{k+1}. Therefore, B_k is the projection of A onto subspaces $\mathrm{span}\{P_{k+1}\}$ and $\mathrm{span}\{Q_k\}$.

We can expect to approximate several dominant WSVD components of A by the SVD of B_k. Let the compact SVD of B_k be

$$B_k = Y_k \Theta_k H_k^\top, \quad \Theta_k = \mathrm{diag}\left(\theta_1^{(k)}, \dots, \theta_k^{(k)}\right), \quad \theta_i^{(k)} > \cdots > \theta_k^{(k)} > 0, \tag{4.7}$$

where $Y_k = \left(y_1^{(k)}, \dots, y_k^{(k)}\right) \in \mathbb{R}^{(k+1)\times k}$ and $H_k = \left(h_1^{(k)}, \dots, h_k^{(k)}\right) \in \mathbb{R}^{k\times k}$ are two orthornormal matrices. Then the approximation to the WSVD triplet $\left(\sigma_i, u_i, v_i\right)$ is $\left(\bar{\sigma}_i^{(k)}, \bar{u}_i^{(k)}, \bar{v}_i^{(k)}\right) := \left(\theta_i^{(k)}, P_{k+1} y_i^{(k)}, Q_k h_i^{(k)}\right)$.

Proposition 4.1. *The approximated WSVD triplet based on WGKB satisfies*

$$A\bar{v}_i^{(k)} - \bar{\sigma}_i^{(k)} \bar{u}_i^{(k)} = 0, \tag{4.8}$$

$$A^\top \bar{u}_i^{(k)} - \bar{\sigma}_i^{(k)} M \bar{v}_i^{(k)} = \alpha_{k+1} M q_{k+1} e_{k+1}^\top. \tag{4.9}$$

Proof. These two relations can be verified by directly using (4.5) and (4.6):

$$A\bar{v}_i^{(k)} - \bar{\sigma}_i^{(k)} \bar{u}_i^{(k)} = A Q_k h_i^{(k)} - \theta_i^{(k)} P_{k+1} y_i^{(k)} = P_{k+1}\left(B_k h_i^{(k)} - \theta_i^{(k)} y_i^{(k)}\right) = 0$$

[1] For the GKB process used in LSQR, it usually uses the right-hand side b as the starting vector. However, for using the GKB process to calculate the SVD or WSVD, it can use any nonzero vector in $\mathbb{R}^m$ as a starting vector.

and

$$\begin{aligned} A^\top \bar{u}_i^{(k)} - \bar{\sigma}_i^{(k)} M \bar{v}_i^{(k)} &= A^\top P_{k+1} y_i^{(k)} - \theta_i^{(k)} M Q_k h_i^{(k)} = M \left(Q_k B_k^\top + \alpha_{k+1} q_k e_{k+1}^\top \right) y_i^{(k)} - \theta_i^{(k)} M Q_k h_i^{(k)} \\ &= M Q_k (B_k^\top y_i^{(k)} - \theta_i^{(k)} h_i^{(k)}) + \alpha_{k+1} M q_k e_{k+1}^\top y_i^{(k)} \\ &= \alpha_{k+1} M q_k e_{k+1}^\top y_i^{(k)}. \end{aligned}$$

The proof is completed. □

Therefore, the triplet $\left(\bar{\sigma}_i^{(k)}, \bar{u}_i^{(k)}, \bar{v}_i^{(k)} \right)$ can be accepted as a satisfied WSVD triplet at the iteration that $\left| \alpha_{k+1} q_k e_{k+1}^\top y_i^{(k)} \right|$ is sufficiently small. This easily computed quantity can be used as a stopping criterion for iteratively computing WSVD triplets.

To solve the large-scale least square problem (3.9), one method is to transform it to the standard one:

$$\min_{z \in \mathcal{Z}} \|z\|_2, \quad \mathcal{Z} = \{ z \in \mathbb{R}^n : z \in \operatorname{argmin} \| A L_M^{-1} z - b \|_2 \} \tag{4.10}$$

by the substitution $z = L_M x$, where L_M is the Cholesky factor of M, i.e. $M = L_M^\top L_M$, and then use the LSQR algorithm to solve it. However, this transformation needs to compute the Cholesky factorization of M in advance, which can be very costly for large-scale M. Noticing that the least square problem (3.9) can be obtained by the WSVD, that is $x_* = A^{\dagger_M} b$, we can expect to iteratively compute x_* based on the WGKB process of A with starting vector b. Note from (4.1) that $\beta P_{k+1} e_1 = b$, where e_1 is the first column of I_{k+1}. If the WGKB process does not terminate until the k-th step, i.e. $k < k_t$, then B_k has full column rank. In this case, we seek a solution to (3.9) in the subspace $\operatorname{span}\{Q_k\}$. By letting $x = Q_k y$ with $y \in \mathbb{R}^k$, we have

$$\min_{x \in \operatorname{span}\{Q_k\}} \|Ax - b\|_2 = \min_{y \in \mathbb{R}^n} \|AQ_k y - b\|_2 = \min_{y \in \mathbb{R}^n} \|P_{k+1} B_k y - \beta P_{k+1} e_1\|_2 = \min_{y \in \mathbb{R}^n} \|B_k y - \beta e_1\|_2$$

and $\|x\|_M = \|Q_k y\|_M = \|y\|_2$. Therefore, the problem (3.9) with $x \in \operatorname{span}\{Q_k\}$ becomes

$$\min_{y \in \mathcal{Y}_k} \|y\|_2, \quad \mathcal{Y}_k = \{ y \in \mathbb{R}^k : y \in \operatorname{argmin} \| B_k y - \beta e_1 \|_2 \} \tag{4.11}$$

This is a standard linear least squares problem with minimum 2-norm, which has the unique solution $y_k = B_k^\dagger \beta e_1$. Therefore, at the k-th iteration, we compute the iterative approximation to (3.9):

$$x_k = Q_k y_k, \quad y_k = B_k^\dagger \beta e_1. \tag{4.12}$$

The above procedure is similar to the LSQR algorithm for standard 2-norm least squares problems. Moreover, the bidiagonal structure of B_k allows us to design a recursive relation to update x_k step by step without explicitly computing $B_k^\dagger \beta e_1$ at each iteration; see [35, Section 4.1] for the similar recursive relation in LSQR. We summarized the iterative algorithm for iteratively solving (3.9) in Algorithm 2, which is named the weighted LSQR (WLSQR) algorithm.

Algorithm 2 Weighted LSQR (WLSQR).

Input: Matrix $A \in \mathbb{R}^{m \times n}$, positive definite $M \in \mathbb{R}^{n \times n}$, vector $b \in \mathbb{R}^m$
Output: Approximate solution to (3.9): x_k
Compute $\beta_1 p_1 = b$, $\alpha_1 q_1 = M^{-1} A^T q_1$,
Set $x_0 = \mathbf{0}$, $w_1 = q_1$, $\bar{\phi}_1 = \beta_1$, $\bar{\rho}_1 = \alpha_1$
for $i = 1, 2, \ldots$ until convergence, **do**
(Applying the WGKB process)
$\beta_{i+1} p_{i+1} = A q_i - \alpha_i p_i$
$\alpha_{i+1} q_{i+1} = M^{-1} A^\top p_{i+1} - \beta_{i+1} q_i$
(Applying the Givens QR factorization to B_k)
$\rho_i = (\bar{\rho}_i^2 + \beta_{i+1}^2)^{1/2}$
$c_i = \bar{\rho}_i / \rho_i$
$s_i = \beta_{i+1} / \rho_i$
$\theta_{i+1} = s_i \alpha_{i+1}$
$\bar{\rho}_{i+1} = -c_i \alpha_{i+1}$
$\phi_i = c_i \bar{\phi}_i$
$\bar{\phi}_{i+1} = s_i \bar{\phi}_i$
(Updating the solution)
$x_i = x_{i-1} + (\phi_i / \rho_i) w_i$
$w_{i+1} = v_{i+1} - (\theta_{i+1} / \rho_i) w_i$
end for

The following result shows that the WLSQR algorithm approaches the exact solution to (3.9) as the algorithm proceeds.

Theorem 4.1. *If the WGKB process terminates at step $k_t = \max_i \{\alpha_i \beta_i > 0\}$, then the iterative solution x_{k_t} is the exact solution to* (3.9).

Proof. By Theorem 3.3, a vector $x \in \mathbb{R}^n$ is the unique solution to (3.9) if and only if

$$Ax - b \perp \mathcal{R}(A), \quad x \perp_M \operatorname{span}\{v_{r+1}, \ldots, v_n\}.$$

Using the property of the GKB process of $\mathcal{A}$, the subspace $\operatorname{span}\{Q_k\}$ can be expressed as the Krylov subspace

$$\operatorname{span}\{Q_k\} = \mathcal{K}_k(\mathcal{A}^*\mathcal{A}, \mathcal{A}^*b) = \operatorname{span}\{(\mathcal{A}^*\mathcal{A})^i\mathcal{A}^*b\}_{i=0}^{k-1} = \operatorname{span}\{(M^{-1}A^\top A)^i M^{-1}A^\top b\}_{i=0}^{k-1}.$$

For any $k \le k_t$, since $x_k = Q_k y_k$, it follows

$$x_k \in \operatorname{span}\{(M^{-1}A^\top A)^i M^{-1}A^\top b\}_{i=0}^{k-1} \subseteq \mathcal{R}(M^{-1}A^\top) = M^{-1}\mathcal{N}(A)^\perp.$$

Using the WSVD of A, we have $\mathcal{N}(A) = \operatorname{span}\{v_{r+1}, \ldots, v_n\}$. For any $v \in \mathbb{R}^n$, it follows that

$$v \in M^{-1}\mathcal{N}(A)^\perp \Leftrightarrow Mv \in \mathcal{N}(A)^\perp \Leftrightarrow v^\top M v_i = 0,\ i = r+1, \ldots, n,$$

which leads to $M^{-1}\mathcal{N}(A)^\perp = \operatorname{span}\{v_1, \ldots, v_r\}$. Therefore, we get $x_k \in \operatorname{span}\{v_1, \ldots, v_r\}$ and thereby $x_k \perp_M \operatorname{span}\{v_{r+1}, \ldots, v_n\}$.

To prove $Ax_{k_t} - b \perp \mathcal{R}(A)$, we only need to show $A^\top(Ax_{k_t} - b) = \mathbf{0}$. By (4.2) and (4.3), we have

$$\begin{aligned}
A^\top(Ax_{k_t} - b) &= A^\top(AQ_{k_t}y_{k_t} - P_{k_t+1}\beta_1 e_1) \\
&= A^\top P_{k_t+1}(B_{k_t}y_{k_t} - \beta_1 e_1) \\
&= M(Q_{k_t}B_{k_t}^\top + \alpha_{k_t+1}q_{k_t+1}e_{k+1}^\top)(B_{k_t}y_{k_t} - \beta_1 e_1) \\
&= M\left[Q_{k_t}(B_{k_t}^\top B_{k_t}y_{k_t} - B_{k_t}^\top\beta_1 e_1) + \alpha_{k_t+1}\beta_{k_t+1}q_{k+1}e_{k_t}^\top y_{k_t}\right] \\
&= \alpha_{k_t+1}\beta_{k_t+1}Mq_{k_t+1}e_{k_t}^\top y_{k_t},
\end{aligned}$$

where we used $B_{k_t}^\top B_{k_t}y_{k_t} = B_{k_t}^\top\beta_1 e_1$ since y_{k_t} satisfies the normal equation of $\min_y \|B_{k_t}y - \beta_1 e_1\|_2$. Since WGKB terminates at k_t, which means that $\alpha_{k_t+1}\beta_{k_t+1} = 0$, we have $A^\top(Ax_{k_t} - b) = \mathbf{0}$. □

The following result shows that the WLSQR has the same effect as first transforming the original problem to (4.10) and then solving it, but it achieves this without the need for an explicit transformation.

Theorem 4.2. *Let the Cholesky factorization of M be $M = L_M^\top L_M$. Then the k-th iterative solution of WLSQR is $x_k = L_M^{-1}z_k$, where z_k is the k-th LSQR solution of* $\min_{z\in\mathbb{R}^n} \|AL_M^{-1}z - b\|_2$.

Proof. The k-th LSQR solution of $\min_{z\in\mathbb{R}^n} \|AL_M^{-1}z - b\|_2$ is the solution of the subspace constrained least squares problem

$$\min_{z\in\bar{S}_k} \|AL_M^{-1}z - b\|_2, \quad \bar{S}_k = \mathcal{K}_k\left((AL_M^{-1})^\top(AL_M^{-1}), (AL_M^{-1})^\top b\right),$$

where

$$\mathcal{K}_k\left((AL_M^{-1})^\top(AL_M^{-1}), (AL_M^{-1})^\top b\right) = \operatorname{span}\left\{(L_M^{-\top}A^\top AL_M^{-1})^i L_M^{-\top}A^\top b\right\}_{i=0}^{k-1} = L_M^{-\top}\operatorname{span}\left\{(A^\top AM^{-1})^i A^\top b\right\}_{i=0}^{k-1}.$$

Therefore, $x_k = L_M^{-1}z_k$ is the solution of the problem $\min_{x\in L_M^{-1}\bar{S}_k} \|Ax - b\|_2$. From the proof of Theorem 4.1, we have

$$L_M^{-1}\bar{S}_k = M^{-1}\operatorname{span}\left\{(A^\top AM^{-1})^i A^\top b\right\}_{i=0}^{k-1} = \operatorname{span}\{(M^{-1}A^\top A)^i M^{-1}A^\top b\}_{i=0}^{k-1} = \mathcal{K}_k(\mathcal{A}^*\mathcal{A}, \mathcal{A}^*b),$$

which is the k-th solution subspace $\mathcal{S}_k$ generated by WGKB. By writing any vector in $\mathcal{S}_k$ as $x = Q_k y$ with $y \in \mathbb{R}^k$, it is easy to verify that $\min_{x\in\mathcal{S}_k} \|Ax - b\|_2$ has the unique solution. It follows that x_k is the k-th WLSQR solution. □

5. Using the WSVD to analyze and solve linear ill-posed problems

For the Tikhonov regularization (1.7) with the $\|x\|_M^2$ regularization term, if we have the Cholesky factorization $M = L_M^\top L_M$, this problem can be transformed into the standard-form Tikhonov regularization problems $\min_{\bar{x}\in\mathbb{R}^n}\{\|AL_M^{-1}\bar{x} - b\|_2^2 + \lambda\|\bar{x}\|_2^2\}$ by letting $\bar{x} = L_M x$. Without the Cholesky factorization of M, we can write its solution explicitly using the WSVD of A.

Theorem 5.1. *The solution to the Tikhonov regularization* (1.7) *can be written as*

$$x_\lambda = \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2 + \lambda}\frac{u_i^\top b}{\sigma_i}v_i. \tag{5.1}$$

Proof. Since V is an M-orthogonal matrix, we can let $x = Vy$ for any $x \in \mathbb{R}^n$ where $y \in \mathbb{R}^n$. Using relations $AV = U\Sigma$ and $V^\top MV = I_n$ in the WSVD of A, (1.7) becomes

$$\min_{y\in\mathbb{R}^n} \{\|U\Sigma y - b\|_2^2 + \lambda\|y\|_2^2\}.$$

The normal equation of this problem is

$$\left[(U\Sigma)^\top(U\Sigma) + \lambda I_n\right] y = (U\Sigma)^\top b \;\Leftrightarrow\; (\Sigma^\top\Sigma + \lambda I_n)y = \Sigma^\top U^\top b,$$

which leads to the unique solution to (1.7) as (5.1).

$$x_\lambda = V(\Sigma^\top\Sigma + \lambda I_n)^{-1}\Sigma^\top U^\top b = \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2 + \lambda}\frac{u_i^\top b}{\sigma_i} v_i.$$

The proof is completed. □

The above expression of x_λ is similar to (2.6), where λ should be chosen properly to filter out the noisy components.

5.1. Iterative regularization by subspace projection

To avoid choosing a proper regularization parameter in advance for the Tikhonov regularization, we consider the *subspace projection regularization* (SPR) method following the idea in [12, §3.3], which can be formed as

$$\min_{x\in\bar{\mathcal{X}}_k} \|x\|_M, \quad \bar{\mathcal{X}}_k = \{x \in \mathcal{S}_k : x \in \operatorname{argmin}\|Ax - b\|_2\}. \tag{5.2}$$

Remark 5.1. The above SPR method is a generalization of the iterative regularization method corresponding to the $\|x\|_2^2$ regularization term. For example, the LSQR method with early stopping can be written as

$$\min_{x\in\bar{\mathcal{X}}_k} \|x\|_2, \quad \bar{\mathcal{X}}_k = \{x \in \mathcal{S}_k : x \in \operatorname{argmin}\|Ax - b\|_2\}, \quad \mathcal{S}_k = \mathcal{K}_k(A^\top A, A^\top b).$$

The success of the SPR method highly depends on the choice of solution subspaces $\mathcal{S}_k$, which should be elaborately constructed to incorporate the prior information about the desired solution encoded by the regularizer $\|x\|_M^2$. For the LSQR method, the solution subspaces $\mathcal{K}_k(A^\top A, A^\top b)$ can only deal with the $\|x\|_2^2$ regularization term. This motivates us to develop a new iterative process to construct solution subspaces to incorporate prior information encoded by $\|x\|_M^2$.

Remark 5.2. For any choice of a k-dimensional $\mathcal{S}_k$, there exists a unique solution to (5.2). To see it, let $x = W_k y$ with $y \in \mathbb{R}^k$ be any vector in $\mathcal{S}_k$, where $W_k \in \mathbb{R}^{n\times k}$ whose columns are M-orthonormal and span $\mathcal{S}_k$. Then the solution to (5.2) satisfies $x_k = W_k y_k$, where y_k is the solution to

$$\min_{y\in\mathcal{Y}_k} \|y\|_2, \quad \mathcal{Y}_k = \{y \in \mathbb{R}^k : y \in \operatorname{argmin}\|AW_k y - b\|_2\}.$$

This problem has a unique solution $y_k = (AW_k)^\dagger b$. Therefore, there exists a unique solution to (5.2).

Using the WSVD of A, if we choose the k-th solution subspace in (5.2) as $\mathcal{S}_k = \operatorname{span}\{v_1, \ldots, v_k\}$, then the solution to (5.2) is $x_k = V_k y_k$, where y_k is the solution to

$$\min_{y\in\mathcal{Y}_k} \|y\|_2, \quad \mathcal{Y}_k = \{y \in \mathbb{R}^k : y \in \operatorname{argmin}\|U_k\Sigma_k y - b\|_2\}.$$

Note that $U_k\Sigma_k$ has full column rank for $1 \le k \le r$. Therefore, $\mathcal{Y}_k$ has only one element, which is $y_k = \Sigma_k^{-1}U_k^\top b$, thereby

$$x_k = V_k\Sigma_k^{-1}U_k^\top b = \sum_{i=1}^{k} \frac{u_i^\top b}{\sigma_i} v_i.$$

Comparing this result with Theorem 3.3, we find that x_k can be obtained by truncating the first k components of x_*. Thus, we name this form of x_k as the truncated WSVD (TWSVD) solution. By truncating the above solution at a proper k, the TWSVD solution can capture the main information corresponding to the dominant right weighted singular vectors v_i while discarding the highly amplified noisy vectors corresponding to others.

From the above investigation, those dominant v_i play an important role in the regularized solution, since they contain the desirable information about x encoded by the regularizer $\|x\|_M^2$. As has been shown that WGKB can be used to approximate the WSVD triplets of A, this motivates us to design iterative regularization algorithms based on WGKB. This can be achieved by setting the k-th solution subspace as $\mathcal{S}_k = \operatorname{span}\{Q_k\}$. Following the same procedure for deriving WLSQR, the problem (5.2) becomes

$$\min_{y\in\mathcal{Y}_k}\|y\|_2,\quad \mathcal{Y}_k=\{y\in\mathbb{R}^k:\ y\in\operatorname{argmin}\|B_k y-\beta_1 e_1\|_2\},\tag{5.3}$$

which has the solution given by (4.12). Therefore, the SPR method with $S_k=\operatorname{span}\{Q_k\}$ is actually the WLSQR method. One important difference from solving the ordinary least squares problem is that here the iteration should be early stopped. This can be seen from Theorem 4.1 since the algorithm eventually converges to the naive solution to (3.9). This is the typical *semi-convergence* behavior for SPR methods: as the iteration proceeds, the iterative solution first gradually approximates to x_{true}, then the solution will deviate far from x_{true} and eventually converges to $A^{\dagger_M}b$ [12, §3.3]. This is because the solution subspace will contain more and more noisy components as it gradually expands. The iteration at which the corresponding solution has the smallest error is called the semi-convergence point. Note that the iteration number k in SPR plays a similar role as the regularization parameter in Tikhonov regularization. Here we adapt two criteria for choosing regularization parameters to estimate the semi-convergence point.

Early stopping rules

1. Discrepancy principle (DP). For the Gaussian white noise e, if an estimate of $\|e\|_2$ is known, one criterion for determining the early stopping iteration is the DP, which states that the discrepancy between the data and predicted output, which is $\|Ax_k-b\|_2$, should be of the order of $\|e\|_2$ [33]. Following the derivation of the procedure for updating x_k in [35], we have

   $$\|Ax_k-b\|_2=\|B_k y_k-\beta_1 e_1\|=\bar{\phi}_{k+1}.$$

   Note that $\|Ax_k-b\|_2$ decreases monotonically since x_k minimizes $\|Ax_k-b\|_2$ in the gradually expanding subspace S_k. Therefore, following DP, we should stop the iteration at the first k that satisfies

   $$\bar{\phi}_{k+1}\le\tau\|e\|_2<\bar{\phi}_k\tag{5.4}$$

   with $\tau>1$ slightly, such as $\tau=1.01$. Typically, the early stopping iteration determined by DP is slightly smaller than the semi-convergence point, thereby the corresponding solution is slightly over-smoothed.
2. L-curve (LC) criterion. The LC criterion is another early stopping rule that does not need an estimate on $\|e\|_2$ [20]. In this method, the log-log scale the curve

   $$\left(\log\|Ax_k-b\|_2,\log\|x_k\|_M\right)=\left(\log\bar{\phi}_{k+1},\log(\|x_k\|_M)\right),\tag{5.5}$$

   is plotted, which usually has a characteristic 'L' shape. The iteration corresponding to the corner of the L-curve, which is defined as the point of maximum curvature of the L-curve in a log-log plot, is usually a good early stopping iteration.
3. Generalized cross-validation (GCV). The GCV is a statistical approach for estimating the optimal regularization parameter especially for the Gaussian white noise case [17]. For the WLSQR method, the GCV function with respect to k should be

   $$\mathrm{GCV}(k)=\frac{\|Ax_k-b\|_2^2}{(\operatorname{trace}(I_m-AA_k^{\dagger}))^2},$$

   where $A_k^{\dagger}$ denotes the matrix that maps the right-hand side b to the k-th regularized solution x_k. By (5.3) and (4.1) we have

   $$x_k=Q_k y_k=Q_k B_k^{\dagger}\beta_1 e_1=Q_k B_k^{\dagger}P_{k+1}^{\top}b,$$

   which implies that $A_k^{\dagger}=Q_k B_k^{\dagger}P_{k+1}^{\top}$. Using (4.5) we get

   $$\operatorname{trace}(AA_k^{\dagger})=\operatorname{trace}(AQ_k B_k^{\dagger}P_{k+1}^{\top})=\operatorname{trace}(P_{k+1}B_k B_k^{\dagger}P_{k+1}^{\top})=\operatorname{trace}(B_k B_k^{\dagger})=k.$$

   Therefore, the final expression of the GCV function is

   $$\mathrm{GCV}(k)=\frac{\bar{\phi}_{k+1}^2}{(m-k)^2},\tag{5.6}$$

   and the minimizer of this function is used as early stopping iteration.

We remark that in using GCV or L-curve method, one must go a few iterations beyond the optimal k in order to verify the optimum. The whole process of the algorithm is summarized in Algorithm 3.

For the application of WLSQR to the first-kind Fredholm integral equations, the computational cost of obtaining the matrices A and b is relatively low when using the quadrature rule for numerical integral. The matrix M is simply a diagonal matrix with its diagonal elements being the weights of the quadrature rule, and A is the product of a matrix containing the values of the kernel function at the grid points and M. In our iterative regularization algorithm, handling M^{-1} is very efficient because M is diagonal.

5.2. Hybrid regularization based on WGKB

Although an early stopping criterion can be used to avoid the semi-convergence, the corresponding solution is still often over or under-regularized since the relative error is very sensitive near the semi-convergence point. The hybrid regularization method is

Algorithm 3 WLSQR with early stopping.

Input: Matrix $A \in \mathbb{R}^{m\times n}$, positive definite $M \in \mathbb{R}^{n\times n}$, vector $b \in \mathbb{R}^m$
Output: Final regularized solution corresponding to (5.2)
1: **for** $k = 1, 2, \ldots,$ **do**
2: Compute α_k, β_k, p_k, q_k by WGKB
3: Update x_k from x_{k-1}
4: Compute the $\|Ax_k - b\|_2$ and $\|x\|_M$
5: **if** Early stopping criterion is satisfied **then**
6: Estimate the semi-convergence point as k_1
7: Terminate the iteration to get x_{k_1}
8: **end if**
9: **end for**

another type of iterative method that can stabilize the convergence behavior, which usually applies Tikhonov regularization to the projected problem at each iteration; see e.g. [7,27,37,6].

At each iteration of the hybrid regularization, we seek a regularized solution in $\mathcal{S}_k = \text{span}\{Q_k\}$, which means we find $x = Q_k y$ with $y \in \mathbb{R}^k$ to the problem

$$\min_{x=Q_k y}\{\|Ax-b\|_2^2 + \lambda x^\top M x\} = \min_{y\in\mathbb{R}^k}\{\|B_k y - \beta_1 e_1\|_2^2 + \lambda_k \|y\|_2^2\},$$

where we have used $(Q_k y)^\top M(Q_k y) = y^\top (Q_k^\top M Q_k) y = \|y\|_2^2$. Note that λ_k should be determined at each iteration. Let λ^{opt} and λ_k^{opt} denote the optimal regularization parameters that minimize the solution error for the original problem and k-th subproblem, respectively. The main idea of the hybrid method is that as k gradually increases, the projected problem approximates the original problem. For a large enough k such that the solution subspace captures the main information of x_{true}, then λ_k^{opt} can well approximate λ^{opt} and thus the corresponding regularized solution will well approximate $x_{\lambda^{\text{opt}}}$ [27,37]. Once we have determined λ_k, the k-th hybrid regularized solution is

$$x_{k,\lambda_k} = Q_k y_{k,\lambda_k}, \quad y_{k,\lambda_k} = \underset{y\in\mathbb{R}^k}{\text{argmin}}\{\|B_k y - \beta_1 e_1\|_2^2 + \lambda_k \|y\|_2^2\}. \tag{5.7}$$

In order to determine λ_k^{opt} at each iteration, we adapt the weighted generalized cross-validation method (WGCV) first proposed in [7]. From (5.7) we have $y_{k,\lambda} = (B_k^\top B_k + \lambda I_k)^{-1} B_k^\top \beta_1 e_1 := B_{k,\lambda}^\dagger \beta_1 e_1$. At the k-th step, the WGCV method finds the minimizer of the following function with the weight parameter ω_k:

$$G(\omega_k, \lambda) = \frac{\|B_k y_{k,\lambda} - \beta_1 e_1\|_2^2}{\left(\text{trace}(I_{k+1} - \omega_k B_k B_{k,\lambda}^\dagger)\right)^2}, \tag{5.8}$$

and use this minimizer as λ_k. If $\omega_k = 1$ for all k, it is the standard GCV method. The weight parameter ω_k is initialized and automatically updated following the same strategy in [7]. Using the SVD of B_k, we can get the analytical expression of $G(\omega_k, \lambda)$ and find its minimizer using the MATLAB built-in function `fminbnd.m`. In the ideal situation, as k increases, λ_k will converge and $G(1, \lambda_k)$ will converge to a fixed value. We terminate the iteration at $k + s_1$ with k the first iteration satisfying

$$\left|\frac{G(1,\lambda_{i+1}) - G(1,\lambda_i)}{G(1,\lambda_1)}\right| < \texttt{tol1}, \quad i = k, \ldots, k+s_1, \tag{5.9}$$

where $s_1 + 1$ is the length of the window to avoid bumps. We set $s_1 = 4$ and $\texttt{tol1} = 10^{-6}$ by default.

If we have a good estimate of $\|e\|_2$, we can use the "secant update" (SU) method to update λ_k step by step quickly based on DP, which was first proposed for the Arnoldi-Tikhonov hybrid method [15]. A heuristic derivation of SU is as follows. At each iteration, define the function

$$\psi_k(\lambda) = \|Ax_k - b\|_2 = \|B_k y_{k,\lambda} - \beta_1 e_1\|_2, \tag{5.10}$$

and consider determining the proper k and λ_k simultaneously by solving the nonlinear equation $\psi_k(\lambda) = \tau\|e\|_2$ with a fixed $\tau > 1$ slightly. We remark that this equation has a solution only when k is sufficiently large. We use the following secant method to solve the above equation. Starting from an initial value λ_0, suppose we already have λ_{k-1} at the $(k-1)$-th iteration. Notice that $\psi_k(\lambda)$ monotonically increases with respect to λ [31]. It can be approximated by the linear function $f(\lambda) = \psi_k(0) + \frac{\psi_k(\lambda_{k-1}) - \psi_k(0)}{\lambda_{k-1}}\lambda$, which interpolates $\psi_k(\lambda)$ at 0 and λ_{k-1}. To update λ_k for the next step, we replace $\psi_k(\lambda)$ by $f(\lambda)$ and solve $f(\lambda) = \tau\|e\|_2$, which leads to $\lambda_k = \frac{\tau\|e\|_2 - \psi_k(0)}{\psi_k(\lambda_{k-1}) - \psi_k(0)}\lambda_{k-1}$. This update formula may suffer from instability for small k, since it holds that $\psi_k(0) > \tau\|e\|_2$. Therefore, we finally use

$$\lambda_k = \left|\frac{\tau\|e\|_2 - \psi_k(0)}{\psi_k(\lambda_{k-1}) - \psi_k(0)}\right|\lambda_{k-1} \tag{5.11}$$

as the practical formula to update λ_k. We set $\lambda_0 = 1.0$ by default. Numerically we find that (5.11) is very stable in the sense that when k is sufficiently large, both the regularization parameter λ_k and residual norm $\psi_k(\lambda_{k-1})$ tend to plateau. Since $\psi_k(0) = \|Ax_k - b\|_2 =$

$\bar{\phi}_{k+1}$, which is the residual norm of the k-th WLSQR solution, it can be updated efficiently by Algorithm 2 (without computing p_k and x_k). To terminate the iteration, we choose $k + s_2$ as the stopping iteration with k the first iteration satisfying

$$\psi_k(0) \le \tau \|e\|_2 \quad \text{and} \quad \left| \frac{\psi_{i+1}(\mu_i) - \psi_i(\mu_{i-1})}{\psi_i(\mu_{i-1})} \right| \le \texttt{tol2}, \quad i = k, \dots, k + s_2, \tag{5.12}$$

where $s_2 + 1$ is the length of the window to avoid bumps. We set $s_2 = 4$ and $\texttt{tol2} = 0.001$ by default.

To summarize, we show the pseudocode of the WGKB based hybrid regularization (WGKB_Hyb) algorithm using WGCV or SU in Algorithm 4.

Algorithm 4 WGKB based hybrid regularization (WGKB_Hyb).

Input: A, b, M, $\alpha > 0$, ω_1 or λ_0
for $k = 1, 2, \dots,$ **do**
  Compute p_k, q_k, α_k, β_k by `WGKB`
  if 'hybrid = WGCV' **then**
    Compute the SVD of B_k
    Determine λ_k by minimizing (5.8)
    Compute y_{k,λ_k} by solving (5.7)
    Update ω_k following [15]
  else if 'hybrid = SU' **then**
    Compute $y_{k,\lambda_{k-1}}$ by solving (5.7)
    Compute the residual norm $\psi_k(\lambda_{k-1})$ by (5.10)
    Computing $\bar{\phi}_{k+1}$ recursively by Algorithm 2
    Update λ_k by (5.11)
  end if
  Terminate iteration by (5.9) or (5.12) at k_2
end for
Compute $x_{k_2,\lambda_{k_2}} = Q_{k_2} y_{k_2,\lambda_{k_2}}$
Output: Final regularized solution $x_{k_2,\lambda_{k_2}}$

6. Numerical experiments

We consider the Fredholm integral equation of the first kind, with the goal of recovering the unknown function $f(t)$ from the noisy observation $g(s)$. While (1.3) is presented as a one-dimensional (1D) problem, our approach can also be used to solve d-dimensional problems with $d > 1$. First, we choose four 1D inverse problems to test the proposed method, then we use our method to solve a large-scale two-dimensional (2D) inverse problems.

6.1. One-dimensional inverse problems

Example 1 This example is chosen from [21] with the name shaw. It models a one-dimensional image restoration problem using the Fredholm integral equation (1.3), where the kernel K and solution f are given by

$$K(s,t) = (\cos s + \cos t)^2 \left(\frac{\sin u}{u} \right)^2, \quad u = \pi(\sin s + \sin t),$$
$$f(t) = 2\exp\left(-6(t-0.8)^2\right) + \exp\left(-2(t+0.5)^2\right),$$

where $t \in [-\pi/2, \pi/2]$ and $s \in [-\pi/2, \pi/2]$.

Example 2 This example is Phillips' famous test problem [36]. Define the function

$$\phi(x) = \begin{cases} 1 + \cos(\frac{\pi x}{3}), & |x| < 3 \\ 0, & |x| \ge 3. \end{cases}$$

Then the kernel K, the solution f and the exact observation are given by

$$K(s,t) = \phi(s-t),$$
$$f(t) = \phi(t),$$

where $t \in [-6, 6]$ and $s \in [-6, 6]$.

Example 3 This test problem is constructed by ourselves. Define the kernel function and true solution as

$$K(s,t) = e^{st},$$
$$f(t) = e^t \cos t$$

where $t \in [0, 1]$ and $s \in [0, 1]$.

Table 6.1
Properties of the four test examples.

Problem	Example 1	Example 2	Example 3	Example 4
$m \times n$	2500×2001	3000×2501	3500×3001	4000×3501
Condition number	5.89×10^{18}	2.14×10^{9}	5.98×10^{18}	1.27×10^{7}
Ill-posedness	severe	mild	severe	mild

Example 4 This test problem is constructed by ourselves. Define the kernel function and true solution as

$$K(s,t) = \begin{cases} s(1-t), & s < t \\ t(1-s), & s \geq t, \end{cases}$$
$$f(t) = t - 2t^2 + t^3,$$

where $t \in [0,1]$ and $s \in [0,1]$.

To discretize the Fredholm integral equation (1.3), we partition the interval $[t_1, t_2]$ into $2l$ uniform subintervals to get $n = 2l + 1$ grid points $t_1 = p_1 < p_2 < \cdots < p_{n-1} < p_n = t_2$. The whole integral is partitioned as

$$\int_{t_1}^{t_2} K(s,t) f(t) \mathrm{d}t = \sum_{i=1}^{l} \int_{p_{2i-1}}^{p_{2i+1}} K(s,t) f(t) \mathrm{d}t,$$

where each integral is approximated by Simpson's rule

$$\int_{p_{2i-1}}^{p_{2i+1}} K(s,t) f(t) \mathrm{d}t \approx \frac{p_{2i+1} - p_{2i-1}}{6} \left[K(s, p_{2i-1}) f(p_{2i-1}) + 4K(s, p_{2i}) f(p_{2i}) + K(s, p_{2i+1}) f(p_{2i+1}) \right].$$

Therefore, the whole integral is approximated as

$$\int_{t_1}^{t_2} K(s,t) f(t) \mathrm{d}t \approx \sum_{i=1}^{n} w_i K(s, p_i) f(p_i),$$

with weights

$$\frac{h}{3} \{1, 4, 2, 4, 2, 4, \ldots, 2, 4, 1\}, \quad h = (t_2 - t_1)/(n-1).$$

The observations are selected from m uniform points in $[s_1, s_2]$ to get an m-dimensional vector. The task is to recover the true vector $x_{\text{true}} = (f(p_1), \ldots, f(p_n))^{\top}$ from the noisy observation b constructed as follows:

$$b = Ax + e, \tag{6.1}$$

where $e \in \mathbb{R}^m$ is a discrete Gaussian white noise vector, and A is the discretized kernel:

$$A = \begin{pmatrix} K(s_1,t_1)w_1 & K(s_1,t_2)w_2 & \cdots & K(s_1,t_n)w_n \\ K(s_2,t_1)w_1 & K(s_2,t_2)w_2 & \cdots & K(s_2,t_n)w_n \\ \vdots & \vdots & \ddots & \vdots \\ K(s_m,t_1)w_1 & K(s_m,t_2)w_2 & \cdots & K(s_m,t_n)w_n \end{pmatrix} \in \mathbb{R}^{m \times n}. \tag{6.2}$$

The scale of the noise is controlled by the noise level $\varepsilon := \|e\|_2 / \|Ax_{\text{true}}\|_2$, which may vary for different test examples. The properties of the test examples are shown in Table 6.1. The discretized true solutions and noisy observations with $\varepsilon = 10^{-2}$ are shown in Fig. 6.1.

From the above discretization, the data fidelity term and regularization term for the discrete ill-posed linear system (6.1) should be $\|Ax - b\|_2^2$ and $\|x\|_M^2$, respectively, where M is the weight matrix

$$M = \frac{h}{3} \mathrm{diag}(1, 4, 2, 4, 2, 4, \ldots, 2, 4, 1). \tag{6.3}$$

We demonstrate the performance of WLSQR for regularizing the four linear ill-posed problems. The standard LSQR method is used as a comparison, where the convergence behaviors of the two methods are shown by plotting the variation of relative reconstruction error $\|x_k - x_{\text{true}}\|_2 / \|x_{\text{true}}\|_2$ with respect to k. To further show the effectiveness of WLSQR, we also use x_{true} to find the optimal Tikhonov regularization parameter λ_{opt} for (1.7) and the corresponding solution $x_{\lambda_{opt}}$, that is $\lambda_{opt} = \min_{\lambda > 0} \|x_\lambda - x_{\text{true}}\|_2$. We use this optimal solution as a baseline for comparing the two methods. All the experiments are performed using MATLAB R2023b.

The convergence behaviors of WLSQR and LSQR are shown in Fig. 6.2 using the relative error curves, where the noise levels are set as $\varepsilon = 10^{-3}, 10^{-2}, 10^{-1}$ for the four test examples. For each example, we find that both the two methods exhibit semi-convergence property, but the relative error of LSQR does not obviously decrease. In contrast, the relative error for WLSQR at the semi-convergence

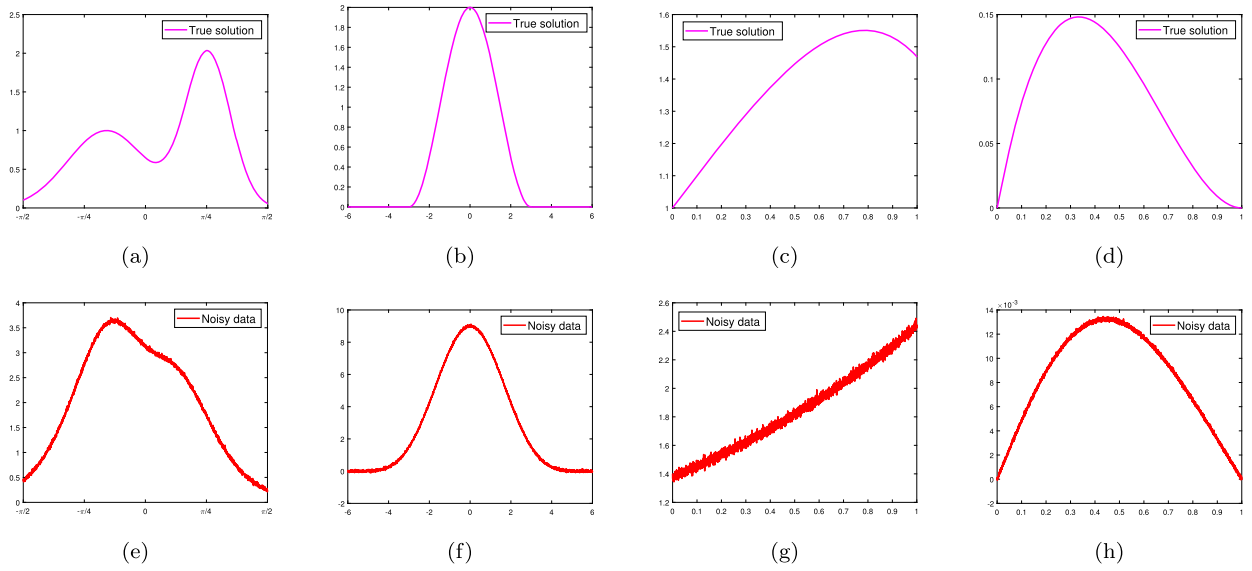

Fig. 6.1. The true solutions and corresponding noisy observations. The noise level for all four test examples is $\epsilon = 10^{-2}$. From left to right are: (a), (e) Example 1; (b), (f) Example 2; (c), (g) Example 3; (d), (h) Example 4.

Table 6.2
Relative errors of the final regularized solutions and corresponding early stopping iterations (in parentheses), where $\epsilon = 10^{-3}$.

Problem	Example 1	Example 2	Example 3	Example 4
Tikh-opt	0.0361	0.0060	0.0062	0.0038
WLSQR-opt	0.0331 (9)	0.0057 (11)	0.0037 (3)	0.0029 (7)
WLSQR-DP	0.0474 (7)	0.0089 (8)	0.0538 (2)	0.0066 (5)
WLSQR-LC	0.0451 (8)	0.0186 (14)	0.0037 (3)	0.0233 (12)
WLSQR-GCV	0.0474 (7)	0.0327 (15)	0.2599 (4)	0.0470 (15)
LSQR-opt	0.3178 (9)	0.3163 (11)	0.3166 (3)	0.3162 (7)
LSQR-DP	0.3194 (7)	0.3163 (8)	0.3206 (2)	0.3163 (5)
LSQR-LC	0.3191 (8)	0.3164 (14)	0.3166 (3)	0.3170 (12)
LSQR-GCV	0.3194 (7)	0.3177 (15)	0.4011 (4)	0.3194 (15)

point is much smaller, and it is usually a bit smaller than the best Tikhonov regularization solution. This confirms the regularization effect of WLSQR, which can incorporate the prior information encoded by $\|x\|_M^2$ into the solution subspaces.

Table 6.2 shows the performance of the WLSQR algorithm with early stopping rules by listing the estimated stopping iterations and corresponding relative errors. We can find that DP always under-estimates the optimal early stopping iteration for both WLSQR and LSQR, while LC and WGCV can either under-estimate or over-estimate the optimal early stopping iteration. Note that for both LSQR and WLSQR applied to Example 2 and Example 4, LC and GCV significantly over-estimate the optimal early stopping iteration, thereby leads to bad solutions with much larger errors. We remark that for WLSQR, the DP method for Example 3 get a regularized solution with a high error, but they are much more accurate than the corresponding LSQR solutions.

We depict the reconstructed solutions computed by WLSQR with DP as an early stopping rule in Fig. 6.3, where the optimal Tikhonov regularized solutions are used as a comparison. We remark that for any ill-posed inverse problems, DP always under-estimates the semi-convergence point and thereby leads to an over-smoothed solution; see e.g. [12, §4.3, 7.3]. The reconstructed solutions for Example 1, 2 and 4 by WLSQR are all of high quantity, very close to the true solutions. For Example 3, the DP solution by WLSQR is significantly over-smoothed, because the relative error is extremely sensitive near the semi-convergence point. Although we do not depict it, the optimal solution at the semi-convergence point of WLSQR for Example 3 is very close to the true solution. In practice, we can not use x_{true} to find the semi-convergence point, instead we should use a proper early stopping rule to find a solution with not-so-bad accuracy.

To overcome the instability of WLSQR caused by the semi-convergence, we apply the WGKB_Hyb algorithm for both the four examples, with the noise levels set to $\epsilon = 10^{-2}$. We denote the WGKB based hybrid methods with WGCV and SU by WGKB-WGCV and WGKB-SU, respectively. For WGKB-SU, we set $\tau = 1.001$ and $\lambda_0 = 1.0$. In Fig. 6.4 we show the relative errors of the two hybrid methods as k grows from 1 to 20. In Fig. 6.5 we display the reconstructed solutions obtained by WGKB-SU at the final iteration. It can be observed that the relative errors of both WGKB-WGCV and WGKB-SU eventually stabilize as k becomes sufficiently large, with values only slightly higher than the lowest relative errors of WLSQR. As shown in Fig. 6.5, the reconstructed solutions computed by WGKB-SU closely match the true solutions for all the four examples.

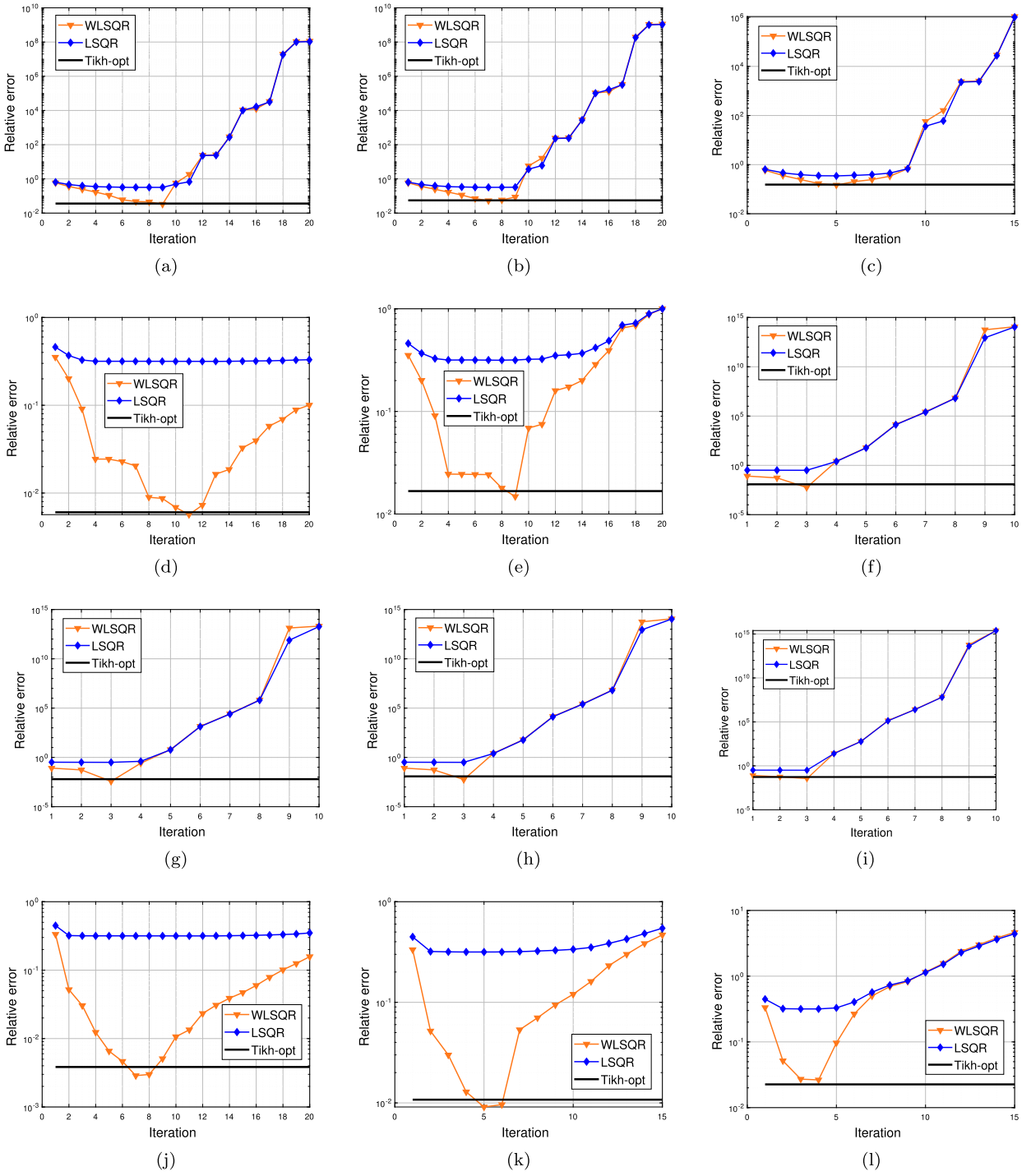

Fig. 6.2. Semi-convergence curves of LSQR and WLSQR for the four test examples. Figures from the top to bottom correspond to Example 1–Example 4; figures from left to right correspond to noise levels $\varepsilon = 10^{-3}, 10^{-2}, 10^{-1}$.

To further illustrate the regularization effect of WLSQR, we show how the relative error of the regularized solution decreases as the noise level approaches zero. We let the noise levels decreases as $\varepsilon = 3.2\times10^{-2}, 1.6\times10^{-2}, 8\times10^{-3}, 4\times10^{-3}, 2\times10^{-3}, 1\times10^{-3}$. The top four subfigures in Fig. 6.6 depict variations of relative error for the optimal regularized solutions computed by WLSQR, LSQR and Tikhonov regularization. We can find that for both WLSQR and Tikhonov regularization, the relative error of the regularized solution approaches zero as the noise decreases to zero. Additionally, the WLSQR solution typically exhibits better accuracy. But for LSQR, the relative error remains almost the same even if the noise level approaches zero, this is because LSQR does not make use of

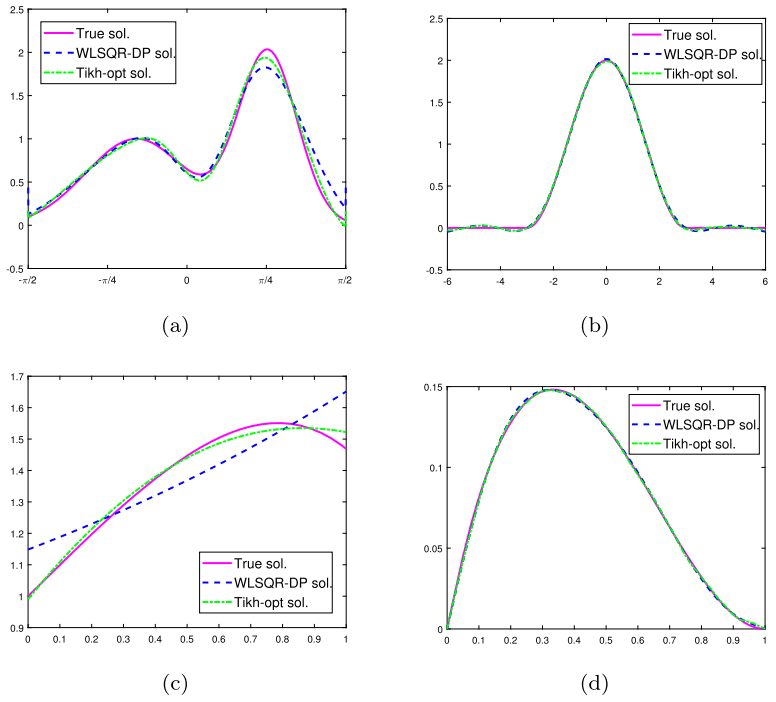

Fig. 6.3. The reconstructed solutions computed by the WLSQR algorithm with DP as the early stopping rule and the Tikhonov regularization (1.7) with optimal regularization parameter. The noise levels are $\epsilon = 10^{-2}$ for all four test examples: (a) Example 1; (b) Example 2; (c) Example 3; (d) Example 4.

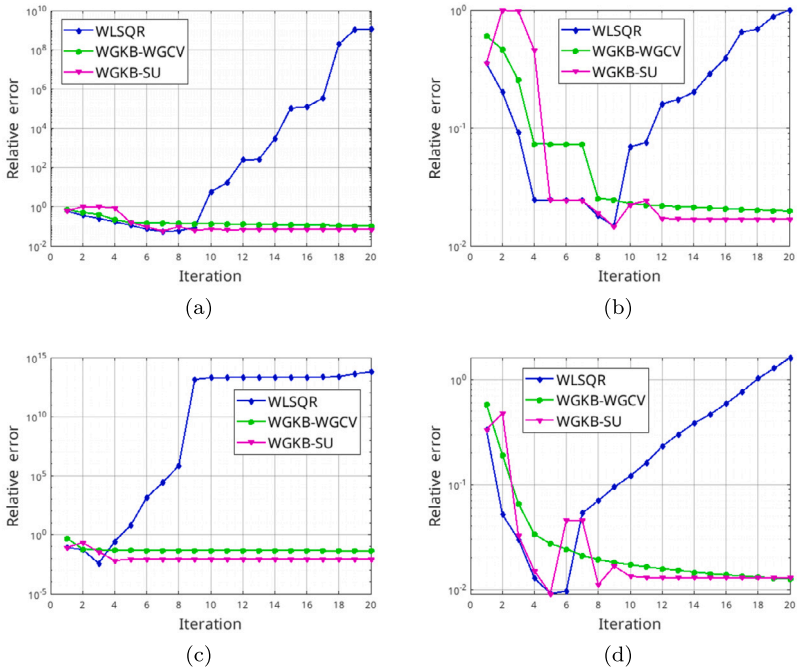

Fig. 6.4. Relative error curves for WLSQR, WGKB-WGCV and WGKB-SU. The noise levels are $\epsilon = 10^{-2}$ for all four test examples: (a) Example 1; (b) Example 2; (c) Example 3; (d) Example 4.

the prior information encoded by $\|x\|_M^2$. We note that for an effective regularization method, the corresponding relative error should converge to zero as the noise approaches zero; see e.g. [12, §3.1, 3.2]. The bottom four subfigures depict variations of relative error when using DP as early stopping rules. Similar to the above tests, DP does not work very well for Example 3, but it is very fruitful for the other three examples.

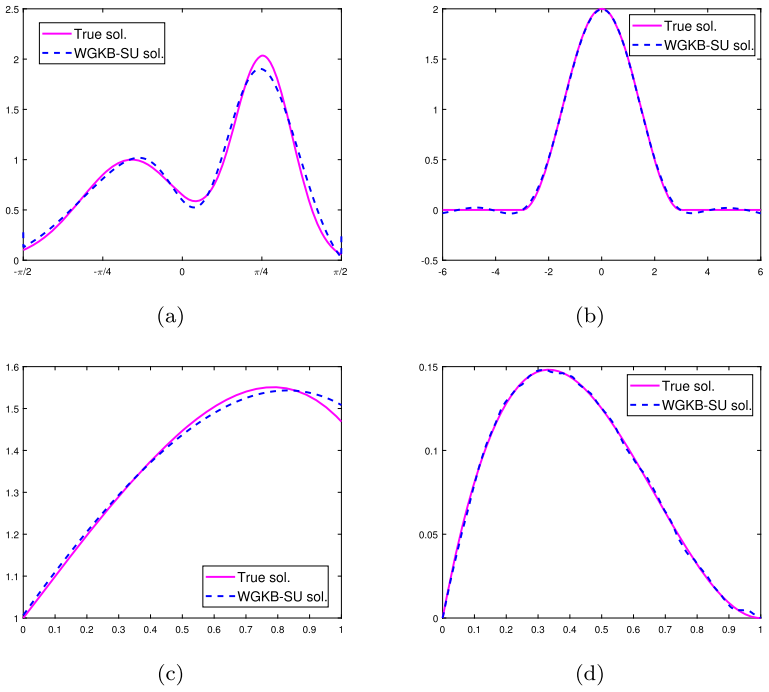

Fig. 6.5. The reconstructed solutions computed by WGKB-SU. The noise levels are $\epsilon = 10^{-2}$ for all four test examples: (a) Example 1; (b) Example 2; (c) Example 3; (d) Example 4.

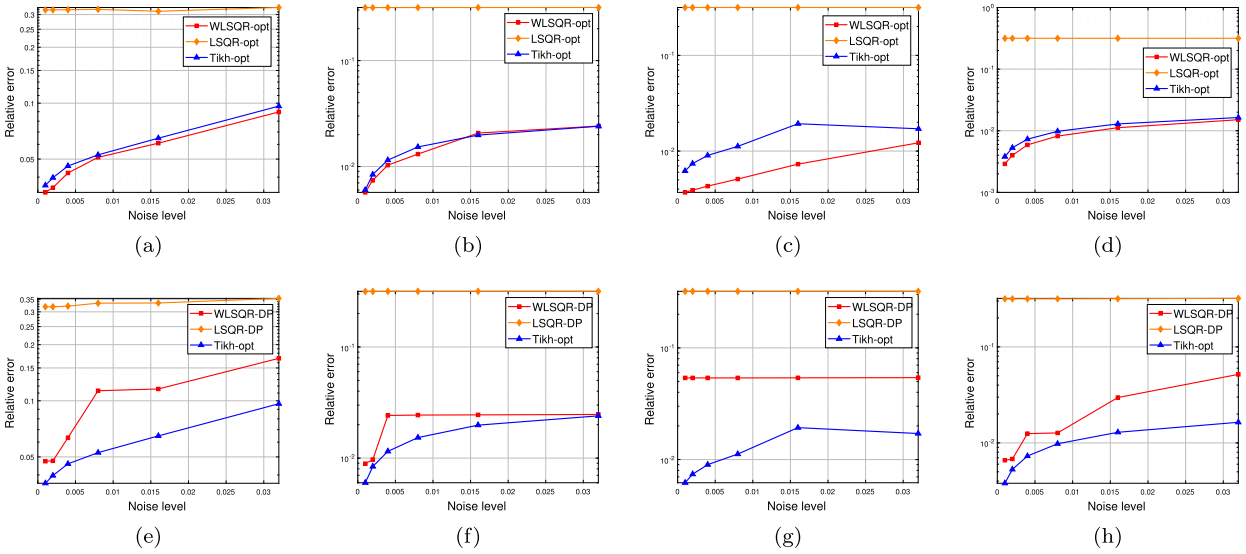

Fig. 6.6. Decay rate of the relative error of the regularized solution computed by WLSQR, LSQR, and Tikhonov regularization as the noise level decreases, with values of $\epsilon = 3.2 \times 10^{-2}, 1.6 \times 10^{-2}, 8 \times 10^{-3}, 4 \times 10^{-3}, 2 \times 10^{-3}, 1 \times 10^{-3}$. Figures on the top and bottom correspond to the relative errors at the early stopping iterations that are optimal and estimated by DP, respectively. Figures from left to right correspond to Example 1–Example 4.

6.2. A two-dimensional inverse problem: NMR relaxometry

We use both the WLSQR and WGKB_Hyb algorithms to solve a 2D inverse problem—nuclear magnetic resonance (NMR) relaxometry. The aim of NMR relaxometry is to reconstruct a distribution of relaxation times associated with the probed material, starting from a signal measured at given times; see [32] for its applications. Mathematically, the 2D NMR relaxometry is modeled using the following (noiseless) Fredholm integral equation of the first kind:

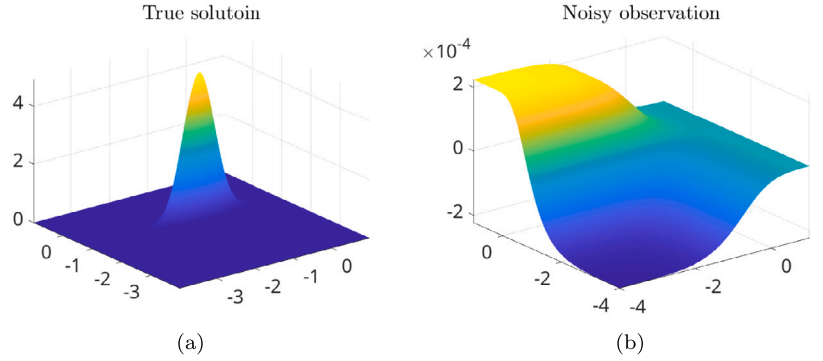

Fig. 6.7. Illustration of the NMR relaxometry problem: (a) the true solution as a function of $(\log_{10}(t_1), \log_{10}(t_2))$; (b) the noisy observation as a function of $(\log_{10}(\tau_1), \log_{10}(\tau_2))$, where the noise level is $\varepsilon = 10^{-3}$.

$$\int_0^{T_1}\int_0^{T_2} K(\tau_1,\tau_2,t_1,t_2) f(t_1,t_2)\mathrm{d}t_1\mathrm{d}t_2 = g(\tau_1,\tau_2), \tag{6.4}$$

where $g(\tau_1,\tau_2)$ is the noiseless signal as a function of experiment times (τ_1,τ_2) and $f(t_1,t_2)$ is the density distribution function.

Following the setting in [14, §3.4], we use the separable kernel

$$K(\tau_1,\tau_2,t_1,t_2) = K_1(\tau_1,t_1)K_2(\tau_2,t_2) := (1-2e^{-\frac{\tau_1}{t_1}})e^{-\frac{\tau_2}{t_2}}$$

and the phantom 'organic' for the relaxation time distribution $f(t_1,t_2)$. This is a severely ill-posed problem. To discretize (6.4), we consider a change of variables by using the logarithmically equispaced nodes

$$t_1^1, t_2^1, \ldots, t_{n_1}^1 \quad \text{and} \quad t_1^2, t_2^2, \ldots, t_{n_2}^2$$

and use the Simpson's rule for the double integral in (6.4) and get sampled observations at the logarithmically equispaced points

$$\tau_1^1, \tau_2^1, \ldots, \tau_{m_1}^1 \quad \text{and} \quad \tau_1^2, \tau_2^2, \ldots, \tau_{m_2}^2.$$

This means that for $1 \le l_1 \le m_1, 1 \le l_2 \le m_2$, we have

$$g(\tau_{l_1}^1, \tau_{l_2}^2) = \sum_{k_1=1}^{n_1}\sum_{k_2=1}^{n_2} K_1(\tau_{l_1}^1, t_{k_1}^1)K_2(\tau_{l_2}^2, t_{k_2}^2) f(t_{k_1}^1, t_{k_2}^2) w(t_{k_1}^1, t_{k_2}^2),$$

where $w(t_{k_1}^1, t_{k_2}^2)$ is the weight value at the point $(t_{k_1}^1, t_{k_2}^2)$. Let $h_1 = \frac{T_1}{n_1-1}$, $h_2 = \frac{T_2}{n_2-1}$, $s_1 = \frac{h_1}{3}(1,4,2,4,2,4,\ldots,2,4,1)^\top \in \mathbb{R}^{n_1}$, and $s_2 = \frac{h_2}{3}(1,4,2,4,2,4,\ldots,2,4,1)^\top \in \mathbb{R}^{n_2}$. Then $w(t_{k_1}^1, t_{k_2}^2)$ is the (k_1,k_2)-element of the Simpson matrix $S = s_1 s_2^\top$; see e.g. [4, §4.8]. Therefore, we have the equation

$$A_1(X \circ S)A_2^\top = B,$$

where "$\circ$" is the Hadarmard product, and

$$\begin{cases} A_{l_1,k_1}^1 = K_1(\tau_{l_1}^1, t_{k_1}^1), & l_1 = 1,\ldots,m_1,\ k_1 = 1,\ldots,n_1 \\ A_{l_1,k_1}^1 = K_1(\tau_{l_2}^2, t_{k_2}^2), & l_2 = 1,\ldots,m_2,\ k_2 = 1,\ldots,n_2 \\ X_{k_1,k_2} = f(t_{k_1}^1, t_{k_2}^2), & k_1 = 1,\ldots,n_1,\ k_2 = 1,\ldots,n_2 \\ B_{l_1,l_2} = g(\tau_{l_1}^1, \tau_{l_2}^2), & l_1 = 1,\ldots,m_1,\ l_2 = 1,\ldots,m_2. \end{cases}$$

Using the relation $\mathrm{vec}(A_1(X \circ S)A_2^\top) = A_2 \otimes A_1 \cdot \mathrm{diag}(\mathrm{vec}(S)) \cdot \mathrm{vec}(X)$, we obtain the linear system $Ax = b$ with $A = A_2 \otimes A_1 \cdot \mathrm{diag}(\mathrm{vec}(S)) \in \mathbb{R}^{m_1 m_2 \times n_1 n_2}$ and $x = \mathrm{vec}(X) \in \mathbb{R}^{n_1 n_2}$. Then we add a Gaussian white noise to Ax with noise level $\varepsilon = 10^{-3}$ to get a noisy observation vector $b \in \mathbb{R}^{m_1 m_2}$. We set $T_1 = T_2 = 10$ and $n_1 = n_2 = 129$, while τ_1 and τ_2 vary from 10^{-4} to 10, and $m_1 = m_2 = 2n_1$. Based on this setting, we get the matrix $A \in \mathbb{R}^{66564\times 16641}$. The discretized true solution and noisy observation in logarithm scale are shown in Fig. 6.7.

Let $M = \mathrm{diag}(\mathrm{vec}(S))$. Then the approximation to the L^2-norm of f should be $\|x\|_M$, and we should consider the Tikhonov regularization $\min_x\{\|Ax-b\|_2^2 + \lambda\|x\|_M^2\}$ and the corresponding iterative regularization. In our experiment, we compare the regularization effect of the WGKB based algorithms with that of the LSQR algorithm with regularizer $\|x\|_2^2$.

To demonstrate the effective regularization of the WGKB-based algorithms, we present the relative error curves of the LSQR, WLSQR, WGKB-WGCV, and WGKB-SU algorithms together in Fig. 6.8. First, we find that all these WGKB based algorithms can

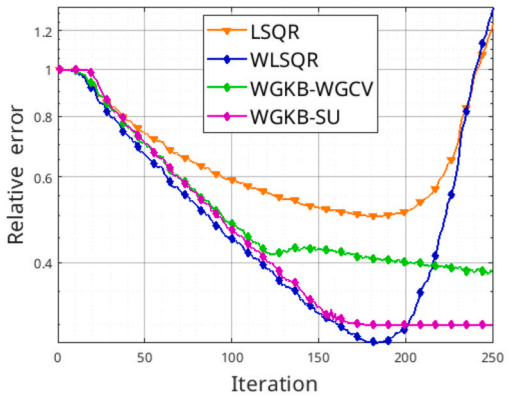

Fig. 6.8. Relative error curves for LSQR, WLSQR, WGKB-WGCV and WGKB-SU applied to the NMR relaxometry problem.

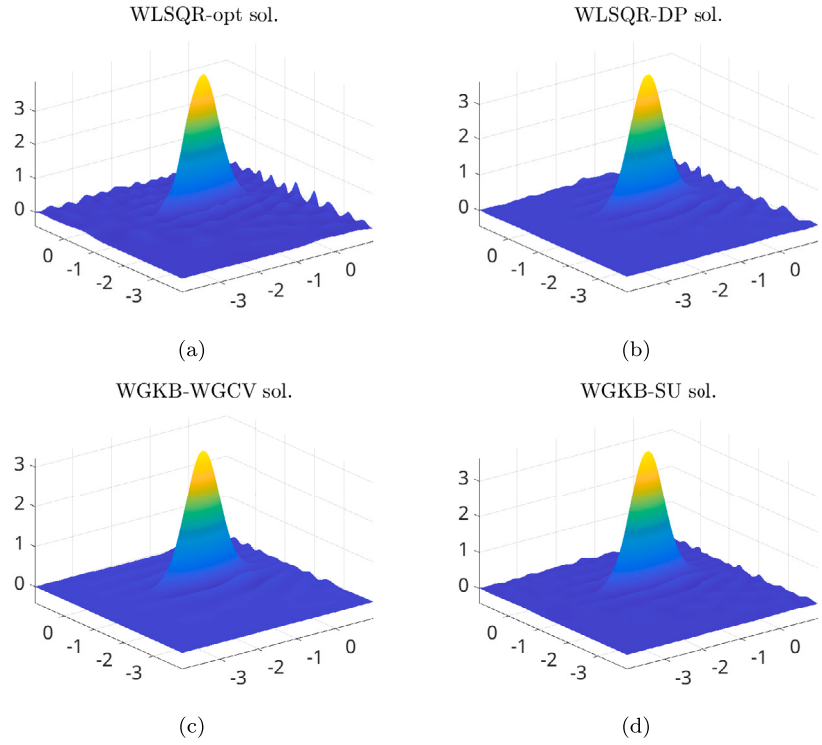

Fig. 6.9. The reconstructed solutions for the NMR relaxometry problem, computed by WLSQR, WGKB-WGCV and WGKB-SU, where we use $(\log_{10}(t_1), \log_{10}(t_2))$ for the time variables. The WLSQR-opt and WLSQR-DP solutions are obtained at $k = 184$ and $k = 153$, respectively.

achieve relative errors much smaller than the LSQR method. We observe the typical semi-convergence behavior of WLSQR, where the smallest relative error is achieved at the semi-convergence point $k = 184$. Using DP, we estimate the semi-convergence point to be $k = 153$. We also find that the semi-convergence behavior is avoided by the hybrid method, where the relative error for WGKB-SU is only slightly higher than the smallest relative error for WLSQR. The convergence of WGKB-WGCV is much more slower, and the relative error is larger than that of WGKB-SU even after 250 steps; this is a common potential flaw for hybrid methods [7,37]. The reconstructed solutions are depicted in Fig. 6.9. The WGKB-based methods can compute regularized solutions that do not deviate significantly from the true solution. We remark this example is just used for experiments, while a better regularization should include the incorporation of box constraints and smoothness priors on the solution [3].

To summarize, the aforementioned experiments for 1D and 2D inverse problems have confirmed that the WLSQR algorithm combined with a proper early stopping rule can obtain a good regularized solution for the regularizer $\|x\|_M^2$, and the WGKB based hybrid regularization method can overcome the semi-convergence behavior to get a more reliable solution. When the matrix M arising from the quadrature discretization is not an identity matrix, the WGKB based methods are better than the LSQR method, because it can approximate dominant WSVD components of A, which contains main information about the true solution.

7. Conclusions

To analyze and compute linear ill-posed problems with regularizer $\|x\|_M^2$, we have revisited and studied a generalization of the SVD under a nonstandard inner product, named the weighted SVD (WSVD). The WSVD shares several similar properties and applications as the standard SVD, such as the low-rank approximation property and solving the least squares problems. Meanwhile, it is very convenient to handle the matrix computation problems with $\|x\|_M$-norm. We have proposed a weighted Golub-Kahan bidiagonalization (WGKB) for computing several dominant WSVD components, and a WGKB-based algorithm, called the weighted LSQR (WLSQR) to solve iteratively least squares problems with minimum $\|x\|_M$-norm. Using the WSVD, we have analyzed the Tikhonov regularization of the linear ill-posed problem with regularizer $\|x\|_M^2$ and given the truncated WSVD solution. We have proposed the WGKB based subspace projection regularization method, which is equivalent to WLSQR with early stopping rules to efficiently compute the regularized solution. To avoid the semi-convergence of WLSQR, two WGKB based hybrid regularization algorithms are proposed. Several numerical experiments are performed to illustrate the fruitfulness of our methods.

Acknowledgements

The author thanks the two anonymous referees for their detailed reading of the manuscript and many invaluable comments.

Data availability

Data will be made available on request.

References

[1] Å. Björck, A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations, BIT Numer. Math. 28 (1988) 659–670.
[2] Å. Björck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996.
[3] V. Bortolotti, R. Brown, P. Fantazzini, G. Landi, F. Zama, Uniform penalty inversion of two-dimensional nmr relaxation data, Inverse Probl. 33 (2016) 015003.
[4] R.L. Burden, J.D. Faires, Numerical Analysis, Brooks Cole, 2010.
[5] N.A. Caruso, P. Novati, Convergence analysis of LSQR for compact operator equations, Linear Algebra Appl. 583 (2019) 146–164.
[6] J. Chung, S. Gazzola, Computational methods for large-scale inverse problems: a survey on hybrid projection methods, SIAM Rev. 66 (2024) 205–284.
[7] J. Chung, J.G. Nagy, D.P. O'leary, et al., A weighted GCV method for Lanczos hybrid regularization, Electron. Trans. Numer. Anal. 28 (2008).
[8] J. Demmel, W. Kahan, Accurate singular values of bidiagonal matrices, SIAM J. Sci. Stat. Comput. 11 (1990) 873–912.
[9] Z. Drmač, K. Veselić, New fast and accurate Jacobi SVD algorithm. I, SIAM J. Matrix Anal. Appl. 29 (2008) 1322–1342.
[10] Z. Drmač, K. Veselić, New fast and accurate Jacobi SVD algorithm. II, SIAM J. Matrix Anal. Appl. 29 (2008) 1343–1362.
[11] C. Eckart, G. Young, The approximation of one matrix by another of lower rank, Psychometrika 1 (1936) 211–218.
[12] H.W. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, Kluwer Academic Publishers, 2000.
[13] K.V. Fernando, B.N. Parlett, Accurate singular values and differential qd algorithms, Numer. Math. 67 (1994) 191–229.
[14] S. Gazzola, P.C. Hansen, J.G. Nagy, IR Tools: a MATLAB package of iterative regularization methods and large-scale test problems, Numer. Algorithms 81 (2019) 773–811.
[15] S. Gazzola, P. Novati, Automatic parameter setting for Arnoldi–Tikhonov methods, J. Comput. Appl. Math. 256 (2014) 180–195.
[16] G. Golub, W. Kahan, Calculating the singular values and pseudo-inverse of a matrix, J. Soc. Ind. Appl. Math., Ser. B Numer. Anal. 2 (1965) 205–224.
[17] G.H. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics 21 (1979) 215–223.
[18] G.H. Golub, C.F. Van Loan, Matrix Computations, 4th ed., The Johns Hopkins University Press, Baltimore, 2013.
[19] P.C. Hansen, The truncated svd as a method for regularization, BIT Numer. Math. 27 (1987) 534–553.
[20] P.C. Hansen, Analysis of discrete ill-posed problems by means of the L-curve, SIAM Rev. 34 (1992) 561–580.
[21] P.C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion, SIAM, Philadelphia, 1998.
[22] P.C. Hansen, J.G. Nagy, D.P. O'Leary, Deblurring Images: Matrices, Spectra and Filtering, SIAM, Philadelphia, 2006.
[23] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, 2012.
[24] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philos. Trans. R. Soc., Math. Phys. Eng. Sci. 374 (2016) 20150202.
[25] M. Jozi, S. Karimi, A weighted singular value decomposition for the discrete inverse problems, Numer. Linear Algebra Appl. 25 (2018) e2114.
[26] I. Karatzas, S. Shreve, Brownian Motion and Stochastic Calculus, vol. 113, Springer Science & Business Media, 2012.
[27] M.E. Kilmer, D.P. O'Leary, Choosing regularization parameters in iterative methods for ill-posed problems, SIAM J. Matrix Anal. Appl. 22 (2001) 1204–1221.
[28] A. Kirsch, et al., An Introduction to the Mathematical Theory of Inverse Problems, vol. 120, Springer, 2011.
[29] R. Kress, Linear Integral Equations, Applied Mathematical Sciences, Springer, New York, 2013.
[30] I. Kyrchei, Weighted singular value decomposition and determinantal representations of the quaternion weighted Moore–Penrose inverse, Appl. Math. Comput. 309 (2017) 1–16.
[31] B. Lewis, L. Reichel, Arnoldi–Tikhonov regularization methods, J. Comput. Appl. Math. 226 (2009) 92–102.
[32] J. Mitchell, T. Chandrasekera, L. Gladden, Numerical estimation of relaxation and diffusion distributions in two dimensions, Prog. Nucl. Magn. Reson. Spectrosc. 62 (2012) 34–50.
[33] V.A. Morozov, Regularization of incorrectly posed problems and the choice of regularization parameter, USSR Comput. Math. Math. Phys. 6 (1966) 242–251.
[34] D.P. O'Leary, J.A. Simmons, A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems, SIAM J. Sci. Stat. Comput. 2 (1981) 474–489.
[35] C.C. Paige, M.A. Saunders, LSQR: an algorithm for sparse linear equations and sparse least squares, ACM Trans. Math. Softw. 8 (1982) 43–71.
[36] D.L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, J. ACM 9 (1962) 84–97.
[37] R.A. Renaut, S. Vatankhah, V.E. Ardestani, Hybrid and iteratively reweighted regularization by unbiased predictive risk and weighted GCV for projected systems, SIAM J. Sci. Comput. 39 (2017) B221–B243.
[38] P. Resnick, H.R. Varian, Recommender systems, Commun. ACM 40 (1997) 56–58.
[39] I. Sergienko, E. Galba, Weighted singular value decomposition of matrices with singular weights based on weighted orthogonal transformations, Cybern. Syst. Anal. 51 (2015) 514–528.
[40] H.D. Simon, H. Zha, Low-rank matrix approximation using the Lanczos bidiagonalization process with applications, SIAM J. Sci. Comput. 21 (2000) 2257–2274.

[41] C.R. Smith III, Multivariable Process Control Using Singular Value Decomposition, The University of Tennessee, 1981.
[42] G.W. Stewart, On the early history of the singular value decomposition, SIAM Rev. 35 (1993) 551–566.
[43] A.N. Tikhonov, V.Y. Arsenin, Solutions of Ill-Posed Problems, 1977, Washington, DC.
[44] C.F. Van Loan, Generalizing the singular value decomposition, SIAM J. Numer. Anal. 13 (1976) 76–83.