

Scalable iterative data-adaptive RKHS regularization

Haibo Li¹, Jinchao Feng², and Fei Lu³

¹School of Mathematics and Statistics, The University of Melbourne, Victoria, Australia.

²Department of Mathematics, School of Sciences, Great Bay University, Dongguan, China.

³Department of Mathematics, Johns Hopkins University, Baltimore, USA.

Abstract

We present iDARR, a scalable iterative Data-Adaptive RKHS Regularization method, for solving ill-posed linear inverse problems. The method searches for solutions in subspaces where the true solution can be identified, with the data-adaptive RKHS penalizing the spaces of small singular values. At the core of the method is a new generalized Golub-Kahan bidiagonalization procedure that recursively constructs orthonormal bases for a sequence of RKHS-restricted Krylov subspaces. The method is scalable with a complexity of $O(kmn)$ for m -by- n matrices with k denoting the iteration numbers. Numerical tests on the Fredholm integral equation and 2D image deblurring show that it outperforms the widely used L^2 and l^2 norms, producing stable accurate solutions consistently converging when the noise level decays.

Keywords: iterative regularization; ill-posed inverse problem; reproducing kernel Hilbert space; Golub-Kahan bidiagonalization; deconvolution.

AMS subject classifications(MSC2020): 47A52, 65F22, 65J20

1 Introduction

This study considers large-scale ill-posed linear inverse problems with little prior information on the regularization norm. The goal is to reliably solve high-dimensional vectors $x \in \mathbb{R}^n$ from the equation

$$Ax + \mathbf{w} = b, \quad A \in \mathbb{R}^{m \times n}, \tag{1.1}$$

where A and b are data-dependent forward mapping and output, and \mathbf{w} denotes noise or measurement error. The problem is ill-posed in the sense that the least squares solution with minimal Euclidean norm, often solved by $x_{LS} = A^\dagger b$ or $x_{LS} = (A^\top A)^\dagger A^\top b$ with \dagger denoting pseudo-inverse, is sensitive to perturbations in b . Such an ill-posedness happens when the singular values of A decay to zero faster than the perturbation in b projected in the corresponding singular vectors.

Regularization is crucial to producing stable solutions for the ill-posed inverse problem. Broadly, it encompasses two integral components: a penalty term that defines the search domain and a hyperparameter that controls the strength of regularization. There are two primary approaches to implementing regularization: direct methods, which rely on matrix decomposition, e.g., the Tikhonov regularization [35], the truncated singular value decomposition (SVD) [11, 15]; and iterative methods, which use matrix-vector computations to scale for high-dimensional problems, see e.g., [7, 12, 41] for recent developments.

In our setting, we encounter two primary challenges: selecting an adaptive regularization norm and devising an iterative method to ensure scalability. The need for an adaptive norm arises from the variability of the forward map A across different applications and the often limited prior information about the regularity of x . Many existing regularization norms, such as the Euclidean norms used in Tikhonov methods in [15, 35] and the total variation norm in [33], lack this adaptability; for more examples, see the related work section below. Although a data-adaptive regularization norm has been proposed in [25, 26]

for nonparametric regression, it is implicitly defined and requires a spectral decomposition of the normal operator.

We introduce iDARR, an iterative Data-Adaptive Reproducing kernel Hilbert space Regularization method. This method resolves both challenges by iteratively solving the subspace-projected problem

$$x_k = \operatorname{argmin}_{x \in \mathcal{X}_k} \|x\|_{C_{rkhs}}, \quad \mathcal{X}_k = \{x : \min_{x \in \mathcal{S}_k} \|Ax - b\|_2\},$$

where $\|\cdot\|_{C_{rkhs}}$ is the implicitly defined semi-norm of a *data-adaptive RKHS* (DA-RKHS), and \mathcal{S}_k are subspaces of the DA-RKHS constructed by a *generalized Golub-Kahan bidiagonalization* (gGKB). By stopping the iteration early using the L-curve criterion, it produces a stable accurate solution without using any matrix decomposition.

The DA-RKHS is a space defined by the data and model, embodying the intrinsic nature of the inverse problem. Its closure is the data-dependent space in which the true solution can be recovered, particularly when A is deficient-ranked. Thus, when used for regularization, it confines the solution search in the right space and penalizes the small singular values, leading to stable solutions. We construct this DA-RKHS by reformulating eq. (1.1) as a weighted Fredholm integral equation of the first kind and examining the identifiability of the input signal, as detailed in Section 2.

Our key innovation is the gGKB. It constructs solution subspaces in the DA-RKHS without explicitly computing it. It is scalable with a cost of only $O(kmn)$, where k is the number of iterations. This cost is orders of magnitude much smaller than the cost of direct methods based on spectral decomposition of $A^\top A$, typically $O(n^3 + mn^2)$ operations.

The iDARR and gGKB have solid mathematical foundations. We prove that each subspace \mathcal{S}_k is restricted in the DA-RKHS, thereby named the *RKHS-restricted Krylov subspaces*. It is spanned by the orthonormal vectors produced by gGKB. Importantly, if not stopped early, the gGKB terminates when the RKHS-restricted Krylov subspace is fully explored, and the solution in each iteration is unique.

Systematic numerical tests employing the Fredholm integral equations demonstrate that iDARR surpasses traditional iterative methods employing l^2 and L^2 norms in the state-of-the-art **IR TOOLS** package [12]. Notably, iDARR delivers accurate estimators that consistently decay with the noise level. This superior performance is evident irrespective of whether the spectral decay is exponential or polynomial, or whether the true solution resides inside or outside the DA-RKHS. Furthermore, our application to image deblurring underscores both its scalability and accuracy.

Main contributions. Our main contribution lies in developing iDARR, a scalable iterative regularization method tailored for large-scale ill-posed inverse problems with little prior knowledge about the solution. The cornerstone of iDARR is the introduction of a new data-adaptive RKHS determined by the underlying model and the data. A key technical innovation is the gGKB, which efficiently constructs solution subspaces of the implicitly defined DA-RKHS.

1.1 Related work

Numerous regularization methods have been developed, and the literature on this topic is too vast to be surveyed here; we refer to [11, 12, 15] and references therein for an overview. In the following, we compare iDARR with the most closely related works.

Regularization norms Various regularization norms exist, such as Euclidean norms of Tikhonov in [15, 35], the total variation norm $\|x'\|_{L^1}$ of the Rudin–Osher–Fatemi method in [33], the L^1 norm $\|x\|_{L^1}$ of LASSO in [34], and the RKHS norm $\|x\|_R^2$ of an RKHS with a user-specified reproducing kernel R [3, 9, 37]. These norms, however, are often based on presumed properties of the solution and do not consider the specifics of each inverse problem. Our RKHS norm differs by adapting to the model and data: our RKHS has a reproducing kernel determined by the inverse problem, and its closure is the

space in which the solution can be identified, making it an apt choice for regularization in the absence of additional solution information.

Iterative regularization (IR) methods IR methods are scalable by accessing the matrix only via matrix-vector multiplications, producing a sequence of estimators until an early stopping, where the iteration number plays the role of the regularization parameter. IR has a rich and extensive history and continues to be a vibrant area of interest in contemporary studies [1, 12, 24, 30]. Different regularization terms lead to various methods. The LSQR algorithm [5, 30] with early stopping is standard for $\|x\|_2^2$ -regularization. It solves projected problems in Krylov subspaces before transforming back to the original domain. For $\|Lx\|_2^2$ with $L \in \mathbb{R}^{p \times n}$, the widely-used methods include joint bidiagonalization method [18, 19], generalized Krylov subspace method [21, 31], random SVD or generalized SVD method [38–40], modified truncated SVD method [2, 17], etc. For the general regularization norm in the form $x^T M x$ with a symmetric matrix M , the MLSQR in [1] treats positive definite M and the preconditioned GKB [24] handles positive semi-definite M 's. Our iDARR studies the case that M is unavailable but M^\dagger is ready to be used.

Golub-Kahan bidiagonalization (GKB) The GKB was first used to solve inverse problems in [29], which generates orthonormal bases for Krylov subspaces in $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_2)$ and $(\mathbb{R}^m, \langle \cdot, \cdot \rangle_2)$. This method extends to bounded linear compact operators between Hilbert spaces, with properties and convergence results detailed in [6]. Our gGKB extends the method to construct RKHS-restricted Krylov subspace in $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_{C_{rkhs}})$ and $(\mathbb{R}^m, \langle \cdot, \cdot \rangle_2)$, where C_{rkhs} is positive semidefinite, and in particular, only C_{rkhs}^\dagger is available.

The remainder of this paper is organized as follows: Section 2 introduces the adaptive RKHS with a characterization of its norm. Section 3 presents in detail the iDARR. Section 4 proves the desired properties of gGKB. In Section 5, we systematically examine the algorithm and demonstrate the robust convergence of the estimator when the noise becomes small. Finally, Section 6 concludes with a discussion on future developments.

2 A Data Adaptive RKHS for Regularization

This section introduces a data-adaptive RKHS (DA-RKHS) that adapts to the model and data in the inverse problem. The closure of this DA-RKHS is the function (or vector) space in which the true solution can be recovered, or equivalently, the inverse problem is well-defined in the sense that the loss function has a unique minimizer. Hence, when its norm is used for regularization, this DA-RKHS ensures that the minimization searches in the space where we can identify the solution.

To describe the DA-RKHS, we first present a unified notation that applies to both discrete and continuous time models using a weighted Fredholm integral equation of the first kind. Based on this notation, we write the normal operator as an integral operator emerging in a variational formulation of the inverse problem. The integral kernel is the reproducing kernel for the DA-RKHS. In other words, the normal operator defines the DA-RKHS. At last, we briefly review a DA-RKHS Tikhonov regularization algorithm, the DARTR algorithm.

2.1 Unified notation for discrete and continuous models

The linear equation eq. (1.1) can arise from discrete or continuous inverse problems. In either case, we can present the inverse problem using the prototype of the Fredholm integral equation of the first kind. We consider only the 1D case for simplicity, and the extension to high-dimensional cases is straightforward. Specifically, let $\mathcal{S}, \mathcal{T} \subset \mathbb{R}$ be two compact sets. We aim to recover the function $\phi : \mathcal{S} \rightarrow \mathbb{R}$ in the

Fredholm equation

$$y(t) = \int_{\mathcal{S}} K(t, s)\phi(s)\nu(ds) + \sigma\dot{W}(t) =: L_K\phi(t) + \sigma\dot{W}(t), \quad \forall t \in \mathcal{T} \quad (2.1)$$

from discrete noisy data

$$b = (y(t_1), \dots, y(t_m))^\top \in \mathbb{R}^m,$$

where we assume the observation index $\mathcal{T} = \{t_j\}_{j=1}^m$ to be $0 = t_0 < t_1 < \dots < t_m$ for simplicity. Here the measurement noise $\sigma\dot{W}(t)$ is the white noise scaled by σ ; that is, the noise at t_j has a Gaussian distribution $\mathcal{N}(0, \sigma^2(t_{j+1} - t_j))$ for each j . Such noise is integrable when the observation mesh refines, i.e., $\max_j(t_{j+1} - t_j)$ vanishes.

Here the finite measure ν can be either the Lebesgue measure with \mathcal{S} being an interval or an atom measure with \mathcal{S} having finitely many elements. Correspondingly, the Fredholm integral equation eq. (2.1) is either a continuous or a discrete model.

In either case, the goal is to solve for the function $\phi : \mathcal{S} \rightarrow \mathbb{R}$ in eq. (2.1). When seeking a solution in the form of $\phi = \sum_{i=1}^n x_i \phi_i$, where $\{\phi_i\}_{i=1}^n$ is a pre-selected set of basis functions, we obtain the linear equation eq. (1.1) with $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$ and the matrix A with entries

$$A(j, i) = \int_{\mathcal{S}} K(t_j, s)\phi_i(s)\nu(ds) = L_K\phi_i(t_j), \quad 1 \leq j \leq m, 1 \leq i \leq n. \quad (2.2)$$

In particular, when $\{\phi_i\}$ are piece-wise constants, we obtain A as follows.

- *Discrete model.* Let ν be an atom measure on $\mathcal{S} = \{s_i\}_{i=1}^n$, a set with n elements. Suppose that the basis functions are $\phi_i(s) = \mathbf{1}_{\{s_i\}}(s)$. Then, $\phi = x$ and the matrix A has entries $A(j, i) = K(t_j, s_i)\nu(s_i)$.
- *Continuous model.* Let ν be the Lebesgue measure on $\mathcal{S} = [0, 1]$, and $\phi_i(s) = \mathbf{1}_{[s_{i-1}, s_i]}(s)$ be piecewise constant functions on a partition of \mathcal{S} with $0 = s_0 < s_1 < \dots < s_n = 1$. Then, $\phi = \sum_{i=1}^n x_i \phi_i$ and the matrix A has entries $A(j, i) = K(t_j, s_i)(s_i - s_{i-1})$.

The default function spaces for ϕ and y above are $L_\nu^2(\mathcal{S})$ and $L_\mu^2(\mathcal{T})$. The loss function $\mathcal{E}(x) = \|Ax - b\|_2^2$ over $L_\nu^2(\mathcal{S})$ becomes

$$\mathcal{E}(\phi) := \|L_K\phi - y\|_{L_\mu^2(\mathcal{T})}^2 = \langle L_K\phi, L_K\phi \rangle_{L_\mu^2(\mathcal{T})} - 2\langle L_K\phi, y \rangle_{L_\mu^2(\mathcal{T})} + \|y\|_{L_\mu^2(\mathcal{T})}^2, \quad (2.3)$$

Eq.eq. (2.1) is a prototype of ill-posed inverse problems, dating back from Hadamard [14], and it remains a testbed for new regularization methods [15, 23, 28, 36].

The $L_\nu^2(\mathcal{S})$ norm is often a default choice for regularization. However, it has a major drawback: it does not take into account the operator L_K , particularly when L_K has zero eigenvalues, and it leads to unstable solutions that may blow up in the small noise limit [22]. To avoid such instability, particularly for iterative methods, we introduce a weighted function space and an RKHS that are adaptive to both the data and the model in the next sections.

2.2 The function space of identifiability

We first introduce a weighted function space $L_\rho^2(\mathcal{S})$, where the measure ρ is defined as

$$\frac{d\rho}{d\nu}(s) := \frac{1}{Z} \int_{\mathcal{T}} |K(t, s)|\mu(dt), \quad \forall s \in \mathcal{S}, \quad (2.4)$$

where $Z = \int_{\mathcal{S}} \int_{\mathcal{T}} |K(t, s)|\mu(dt)\nu(ds)$ is a normalizing constant. This measure quantifies the exploration of data to the unknown function through the integral kernel K at the output set \mathcal{T} , i.e., $\{K(t_j, \cdot)\}_{t_j \in \mathcal{T}}$,

hence it is referred to as an *exploration measure*. In particular, when (1.1) is a discrete model, the exploration measure is the normalized column sum of the absolute values of the matrix A .

The major advantage of the space $L_\rho^2(\mathcal{S})$ over the original space $L_\nu^2(\mathcal{S})$ is that it is adaptive to the specific setting of the inverse problem. In particular, this weighted space takes into account the structure of the integral kernel and the data points in \mathcal{T} . Thus, while the following introduction of RKHS can be carried out in both $L_\rho^2(\mathcal{S})$ and $L_\nu^2(\mathcal{S})$, we will focus only on $L_\rho^2(\mathcal{S})$.

Next, we consider the variational inverse problem over L_ρ^2 , and the goal is to find a minimizer of the loss function eq. (2.3) in L_ρ^2 . Since the loss function is quadratic, its Hessian is a symmetric positive linear operator, and it has a unique minimizer in the linear subspace where the Hessian is strictly positive. We assign a name to this subspace in the next definition.

Definition 2.1 (Function space of identifiability) *In a variational inverse problem of minimizing a quadratic loss function \mathcal{E} in L_ρ^2 , we call $\mathcal{L}_{\bar{G}} = \frac{1}{2}\nabla^2\mathcal{E}$ the normal operator, where $\nabla^2\mathcal{E}$ is the Hessian of \mathcal{E} , and we call $H = \text{Null}(\mathcal{L}_{\bar{G}})^\perp$ the function space of identifiability (FSOI).*

The next lemma specifies the FSOI for the loss function in eq. (2.3) (see [26] for its proof).

Lemma 2.2 *Assume $K \in C(\mathcal{T} \times \mathcal{S})$. For ρ in eq. (2.4), define $\bar{G} : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$ as*

$$\bar{G}(s, s') := \frac{G(s, s')}{\frac{d\rho}{d\nu}(s)\frac{d\rho}{d\nu}(s')}, \quad G(s, s') := \int_{\mathcal{T}} K(t, s)K(t, s')\mu(dt). \quad (2.5)$$

(a) *The normal operator for \mathcal{E} in eq. (2.3) over L_ρ^2 is $\mathcal{L}_{\bar{G}} : L_\rho^2 \rightarrow L_\rho^2$ defined by*

$$\mathcal{L}_{\bar{G}}\phi(s) := \int_{\mathcal{S}} \phi(s')\bar{G}(s, s')\rho(ds'), \quad (2.6)$$

and the loss function can be written as

$$\mathcal{E}(\phi) = \langle \mathcal{L}_{\bar{G}}\phi, \phi \rangle_{L_\rho^2} - 2\langle \phi^D, \phi \rangle_{L_\rho^2} + \text{const.}, \quad (2.7)$$

where ϕ^D comes from Riesz representation s.t. $\langle \phi^D, \phi \rangle_{L_\rho^2} = \langle L_K\phi, y \rangle_{L_\mu^2(\mathcal{T})}$ for any $\phi \in L_\rho^2$.

(b) *$\mathcal{L}_{\bar{G}}$ is compact, self-adjoint, and positive. Hence, its eigenvalues $\{\lambda_i\}_{i \geq 1}$ converge to zero and its orthonormal eigenfunctions $\{\psi_i\}_i$ form a complete basis of $L_\rho^2(\mathcal{S})$.*

(c) *The FSOI is $H := \overline{\text{span}\{\psi_i\}_{i:\lambda_i>0}} \subset L_\rho^2(\mathcal{S})$, and the unique minimizer of \mathcal{E} in H is $\hat{\phi} = \mathcal{L}_{\bar{G}}^{-1}\phi^D$, where $\mathcal{L}_{\bar{G}}^{-1}$ is the inversion of $\mathcal{L}_{\bar{G}} : H \rightarrow L_\rho^2$.*

Theorem 2.2 reveals the cause of the ill-posedness, and provides insights on regularization:

- The variational inverse problem is well-defined only in the FSOI H , which can be a proper subset of L_ρ^2 . Its ill-posedness in H depends on the smallest eigenvalue of the operator $\mathcal{L}_{\bar{G}}$ and the error in ϕ^D .
- When the data is noiseless, the loss function can only identify the H -projection of the true input function. When data is noisy, its minimizer $\mathcal{L}_{\bar{G}}^{-1}\phi^D$ is ill-defined in L_ρ^2 when $\phi^D \notin \mathcal{L}_{\bar{G}}(H)$.

As a result, when regularizing the ill-posed problem, an important task is to ensure the search takes place in the FSOI and to remove the noise-contaminated components making $\phi^D \notin \mathcal{L}_{\bar{G}}(H)$.

2.3 A Data-adaptive RKHS

Our *data-adaptive RKHS* is the RKHS with \bar{G} in eq. (2.5) as a reproducing kernel. Hence, it is adaptive to the integral kernel K and the data in the model. When its norm is used for regularization, it ensures that the search takes place in the FSOI because its L_ρ^2 closure is the FSOI; also, it penalizes the components in ϕ^D corresponding to the small singular values.

The next lemma characterizes this RKHS, and we refer to [26] for its proof.

Lemma 2.3 (Characterization of the adaptive RKHS) *Assume $K \in C(\mathcal{T} \times \mathcal{S})$. The RKHS H_G with \bar{G} as its reproducing kernel satisfies the following properties.*

- (a) $H_G := \mathcal{L}_{\bar{G}}^{-\frac{1}{2}}(L_\rho^2(\mathcal{S}))$ has inner product $\langle \phi, \phi \rangle_{H_G} = \langle \mathcal{L}_{\bar{G}}^{-\frac{1}{2}}\phi, \mathcal{L}_{\bar{G}}^{-\frac{1}{2}}\phi \rangle_{L_\rho^2(\mathcal{S})}$.
- (b) $\{\sqrt{\lambda_i}\psi_i\}_{\lambda_i > 0}$ is an orthonormal basis of H_G , where $\{(\lambda_i, \psi_i)\}_i$ are eigen-pairs of $\mathcal{L}_{\bar{G}}$.
- (c) For any $\phi = \sum_{i=1}^{\infty} c_i y_i \in H_G$, we have

$$\langle L_K \phi, L_K \phi \rangle_{L_\mu^2(\mathcal{T})} = \sum_{i=1}^{\infty} \lambda_i c_i^2, \quad \|\phi\|_{L_\rho^2}^2 = \sum_{i=1}^{\infty} c_i^2, \quad \|\phi\|_{H_G}^2 = \sum_{i=1}^{\infty} \lambda_i^{-1} c_i^2.$$

- (d) $H = \overline{H_G}$ with inclosure in $L_\rho^2(\mathcal{S})$, where $H = \overline{\text{span}\{y_i\}_{i:\lambda_i>0}}$ is the FSOI.

The next theorem shows the computation of the RKHS norm for the problem eq. (1.1) when it is written in the form eq. (2.1)-eq. (2.2). A key component is solving the eigenvalues of $\mathcal{L}_{\bar{G}}$ through a generalized eigenvalue problem.

Theorem 2.4 (Computation of RKHS norm) *Suppose that eq. (1.1) is equivalent to eq. (2.1) under basis functions $\{\phi_i\}_{i=1}^n$ with $n \leq \infty$ and eq. (2.2). Let B with entries $B(i,j) = \langle \phi_i, \phi_j \rangle_{L_\rho^2}$ be the non-singular basis matrix, where ρ is the measure defined in eq. (2.4). Then, the operator $\mathcal{L}_{\bar{G}}$ in eq. (2.6) has eigenvalues $(\lambda_1, \dots, \lambda_n)$ solved by the generalize eigenvalue problem:*

$$A^\top A V = B V \Lambda, \quad \text{s.t.,} \quad V^\top B V = I_n, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n), \quad (2.8)$$

and the eigenfunctions are $\{\psi_k = \sum_{j=1}^n V_{jk} \phi_j\}_k$. The RKHS norm of $\phi = \sum_{i=1}^n x_i \phi_i$ satisfies

$$\begin{aligned} \|\phi\|_{H_G}^2 &= \|x\|_{C_{rkh}s}^2 = x^\top C_{rkh}s x, \\ C_{rkh}s &= (V \Lambda V^\top)^\dagger = B(A^\top A)^\dagger B, \quad C_{rkh}s^\dagger = B^{-1}(A^\top A)B^{-1}. \end{aligned} \quad (2.9)$$

In particular, if $B = I_n$, we have $C_{rkh}s = (A^\top A)^\dagger$.

Proof. Denote $\Phi = (\phi_1, \dots, \phi_n)^\top$ and $\Psi = (\psi_1, \dots, \psi_n)^\top$. We first prove that the eigenvalues of $\mathcal{L}_{\bar{G}}$ are solved by eq. (2.8). We suppose $\{(\lambda_i, \psi_i)\}_{i=1}^n$ are the eigenvalues and eigen-functions of $\mathcal{L}_{\bar{G}}$ over L_ρ^2 with $\{\psi_i\}$ being an orthonormal basis of $L_\rho^2(\mathcal{S})$. Since $\mathcal{H} = \text{span}\{\phi_i\}_{i=1}^n \supseteq \mathcal{L}_{\bar{G}}(L_\rho^2)$, there exists $V \in \mathbb{R}^{n \times n}$ such that $\Psi = V^\top \Phi$, i.e., $\psi_k = \sum_{j=1}^n V_{jk} \phi_j$ for each $1 \leq k \leq n$. Then, the task is to verify that V and $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_n)$ satisfy $A^\top A V = B V \Lambda$ and $V^\top B V = I_n$.

The orthonormality of $\{\psi_i\}$ implies that

$$I_n = (\langle \psi_k, \psi_l \rangle_{L_\rho^2})_{1 \leq k, l \leq n} = \left(\langle \sum_{i=1}^n V_{ik} \phi_i, \sum_{j=1}^n V_{jl} \phi_j \rangle_{L_\rho^2} \right)_{1 \leq k, l \leq n} = V^\top B V.$$

Note that $[A^\top A](j, i) = \langle L_K \phi_i, L_K \phi_j \rangle_{L_\mu^2(\mathcal{T})} = \langle \phi_j, \mathcal{L}_{\bar{G}} \phi_i \rangle_{L_\rho^2}$ for $1 \leq i, j \leq n$. Then,

$$\langle \phi_j, \mathcal{L}_{\bar{G}} \psi_k \rangle_{L_\rho^2} = \sum_{i=1}^n [A^\top A](j, i) V_{ik}.$$

Meanwhile, the eigen-equation $\mathcal{L}_{\bar{G}}\psi_k = \lambda_k\psi_k$ implies that for each ϕ_j ,

$$\langle \phi_j, \mathcal{L}_{\bar{G}}\psi_k \rangle_{L^2_\rho} = \lambda_k \langle \phi_j, \psi_k \rangle_{L^2_\rho} = \lambda_k \langle \phi_j, \sum_{i=1}^n V_{ik}\phi_i \rangle_{L^2_\rho} = \lambda_k \sum_{i=1}^n B_{ji}V_{ik},$$

i.e., $(\langle \phi_j, \mathcal{L}_{\bar{G}}\psi_k \rangle_{L^2_\rho}) = B\Lambda$. Hence, these two equations imply that $A^\top AV = B\Lambda$.

Next, to compute the norm of $\phi = \sum_{i=1}^n x_i\phi_i \in H_G$, we write it as $\phi = x^\top \Phi = x^\top V^{-1}\Psi$. Then, its norm is

$$\|\phi\|_{H_G}^2 = \sum_{k=1}^n \lambda_k^{-1} (x^\top V^{-1})_k^2 = x^\top V^{-1}\Lambda^\dagger V^{-\top} x = x^\top (V\Lambda V^\top)^\dagger x.$$

Thus, $C_{rkhs} = (V\Lambda V^\top)^\dagger = B(A^\top A)^\dagger B$ and $C_{rkhs}^\dagger = V\Lambda V^\top = B^{-1}(A^\top A)B^{-1}$. ■

In particular, when eq. (1.1) is either a discrete model or a discretization of eq. (2.1) based on Riemann sum approximation of the integral, the exploration measure ρ is the normalized column sum of the absolute values of the matrix A , and $B = \text{diag}(\rho)$. See Section 5.1 for details.

2.4 DARTR: data-adaptive RKHS Tikhonov regularization

We review DARTR, a data-adaptive RKHS Tikhonov regularization algorithm introduced in [26].

Specifically, it solves the problem eq. (1.1) with regularization:

$$(\hat{x}_{\lambda_*}, \lambda_*) = \underset{x \in \mathbb{R}^n, \lambda \in \mathbb{R}^+}{\operatorname{argmin}} \mathcal{E}_\lambda(x), \quad \text{where } \mathcal{E}_\lambda(x) := \|Ax - b\|^2 + \lambda \|x\|_{C_{rkhs}}^2,$$

where the norm $\|\cdot\|_{C_{rkhs}}$ is the DA-RKHS norm introduced in Theorem 2.4. A direct solution minimizing $\mathcal{E}_\lambda(x)$ is to solve $(A^\top A + \lambda C_{rkhs})x_\lambda = A^\top b$. However, the computation of C_{rkhs} requires a pseudo-inverse that may cause numerical instability.

DARTR introduces a transformation matrix $C_* := V\Lambda^{1/2}$ to avoid using the pseudo-inverse. Note that $C_*^\top C_{rkhs} C_* = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} := \mathbf{I}_r$, where I_r is the identity matrix with rank r , the number of positive eigenvalues in Λ . Then, the linear equation $(A^\top A + \lambda C_{rkhs})x_\lambda = A^\top b$ is equivalent to

$$(C_* A^\top A C_* + \lambda \mathbf{I}_r) \tilde{x}_\lambda = C_* A^\top b \tag{2.10}$$

with $\tilde{x}_\lambda = C_*^{-1}x_\lambda$. Thus, DARTR computes \tilde{x}_λ in the above equation by least squares with minimal norm, and returns $x_\lambda = C_* \tilde{x}_\lambda$.

DARTR is a direct method based on matrix decomposition, and it takes $O(mn^2 + n^3)$ flops. Hence, it is computationally infeasible when n is large. The iterative method in the next section implements the RKHS regularization in a scalable fashion.

3 Iterative Regularization by DA-RKHS

This section introduces a subspace project method tailored to utilize the DA-RKHS for iterative regularization. As an iterative method, it achieves scalability by accessing the coefficient matrix only via matrix-vector multiplications, producing a sequence of estimators until reaching a desired solution. This section follows the notation conventions in Table 1.

3.1 Overview

Our regularization method is based on subspace projection in the DA-RKHS. It iteratively constructs a sequence of linear subspaces \mathcal{S}_k of the DA-RKHS $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_{C_{rkhs}})$, and recursively solves projected problems

$$x_k = \underset{x \in \mathcal{X}_k}{\operatorname{argmin}} \|x\|_{C_{rkhs}}, \quad \mathcal{X}_k = \{x : \min_{x \in \mathcal{S}_k} \|Ax - b\|_2\}. \tag{3.1}$$

Table 1: Table of notations.

A, B, C	matrix or array by capital letters
b, c, x, y, z, u, v	vector by regular letters
α, β, γ	scalar by Greek letters
$\mathcal{R}(A)$ and $\mathcal{N}(A)$	the range and null spaces of matrix A

This process yields a sequence of solutions $\{x_k\}$, each emerging from its corresponding subspace. We ensure the uniqueness of the solution within each iteration, as detailed in Theorem 4.7. The iteration proceeds until it meets an early stopping criterion, designed to exclude excessive noisy components and thereby achieve effective regularization. The spaces \mathcal{S}_k are called the *solution subspaces*, and the iteration number k plays the role of the regularization parameter.

Algorithm 1 outlines this procedure, which is a recursion of the following three parts.

- P1 *Construct the solution subspaces.* We introduce a new generalized Golub-Kahan bidiagonalization (gGKB) to construct the solution subspaces in the DA-RKHS iteratively. The procedure is presented in Algorithm 2.
- P2 *Recursively update the solution to the projected problem.* We solve the least squares problem in the solution subspaces in eq. (3.1) efficiently by a new LSQR-type algorithm, Algorithm 3, updating $\|x_k\|_{C_{rkhs}}$ and the residual norm $\|Ax - b\|_2$ without even computing the residual.
- P3 *Regularize by early stopping.* We select the optimal k by either the discrepancy principle (DP) when we have an accurate estimate of $\|\mathbf{w}\|_2$ or the L-curve criterion otherwise.

Require: $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $B = \text{diag}(\rho)$, $x_0 = \mathbf{0}$, $\bar{x}_0 = \mathbf{0}$, where ρ is the exploration measure in eq. (2.4).

```

1: for  $k = 1, 2, \dots$ , do
2:   P1. Compute  $u_k, z_k, \bar{z}_k, \alpha_k, \beta_k$  by gGKB in Algorithm 2.
3:   P2. Update  $x_k, \bar{\gamma}_{k+1}, \|x_k\|_{C_{rkhs}}$ , etc. by Algorithm 3.
4:   P3: Stop at iteration  $k_*$  if Early stopping criterion is satisfied.            $\triangleright$  L-curve or DP

```

Ensure: Final solution x_{k_*}

Algorithm 1: iDARR: Iterative Data-Adaptive RKHS Regularization

In the next subsection, we present details for these key parts. Then, we analyze the computational complexity of the algorithm.

3.2 Algorithm details and derivations

This subsection presents detailed derivations for the three parts P1-P3 in Algorithm 1.

P1. Construct the solution subspaces. We construct the solution subspaces by elaborately incorporating the regularization term $\|\cdot\|_{C_{rkhs}}^2$ in the Golub-Kahan bidiagonalization (GKB) process. A key point is to use $C_{rkhs}^\dagger = B^{-1}A^\top AB^{-1}$ to avoid explicitly computing C_{rkhs} , which involves the costly spectral decomposition of the normal operator, see eq. (2.9) in Theorem 2.4.

Consider first the case where C_{rkhs} is positive definite. In this scenario, A has full column rank, and the C_{rkhs} -inner product Hilbert space $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_{C_{rkhs}})$ is a discrete representation of the RKHS H_G with the given basis functions. Note that the true solution is mapped to the noisy b by $A : (\mathbb{R}^n, \langle \cdot, \cdot \rangle_{C_{rkhs}}) \rightarrow (\mathbb{R}^m, \langle \cdot, \cdot \rangle_2)$. Let $A^* : (\mathbb{R}^m, \langle \cdot, \cdot \rangle_2) \rightarrow (\mathbb{R}^n, \langle \cdot, \cdot \rangle_{C_{rkhs}})$ be the adjoint of A , i.e. $\langle Ax, b \rangle_2 = \langle x, A^*b \rangle_{C_{rkhs}}$.

for any $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$. By definition, the matrix-form expression of A^* is

$$A^* = C_{rkhs}^{-1} A^\top, \quad (3.2)$$

since $\langle Ax, b \rangle_2 = x^\top A^\top b$ and $\langle x, A^* b \rangle_{C_{rkhs}} = x^\top C_{rkhs} A^* b$ for any x and b .

The Golub-Kahan bidiagonalization (GKB) process recursively constructs orthonormal bases for these two Hilbert spaces starting with the vector b as follows:

$$\beta_1 u_1 = b, \quad \alpha_1 z_1 = A^* u_1, \quad (3.3a)$$

$$\beta_{i+1} u_{i+1} = A z_i - \alpha_i u_i, \quad (3.3b)$$

$$\alpha_{i+1} z_{i+1} = A^* u_{i+1} - \beta_{i+1} z_i, \quad (3.3c)$$

where $u_i \in (\mathbb{R}^m, \langle \cdot, \cdot \rangle_2)$, $z_i \in (\mathbb{R}^n, \langle \cdot, \cdot \rangle_{C_{rkhs}})$ with $z_0 = \mathbf{0}$, and α_i , β_i are normalizing factor such that $\|u_i\|_2 = \|z_i\|_{C_{rkhs}} = 1$. The iteration starts with $u_1 = b/\beta_1$ with $\beta_1 = \|b\|_2$. Using A^* in eq. (3.2), we write eq. (3.3c) as

$$\alpha_{i+1} z_{i+1} = C_{rkhs}^{-1} A^\top u_{i+1} - \beta_{i+1} z_i \quad (3.4)$$

with $\alpha_{i+1} = \|C_{rkhs}^{-1} A^\top u_{i+1} - \beta_{i+1} z_i\|_{C_{rkhs}}$. To compute α_{i+1} without explicitly computing C_{rkhs} , define $\bar{z}_i = C_{rkhs} z_i$. Then we have

$$\alpha_{i+1} \bar{z}_{i+1} = A^\top u_{i+1} - \beta_{i+1} \bar{z}_i, \quad (3.5)$$

where $\bar{z}_0 := \mathbf{0}$. Let $p = A^\top u_{i+1} - \beta_{i+1} \bar{z}_i$. Then we obtain $\alpha_{i+1} = \|C_{rkhs}^{-1} p\|_{C_{rkhs}} = (p^\top C_{rkhs}^{-1} p)^{1/2}$, which uses $C_{rkhs}^{-1} = B^{-1} A^\top A B^{-1}$ without computing C_{rkhs} .

Next, consider that C_{rkhs} is positive semidefinite. The iterative process remains the same with C_{rkhs}^{-1} replaced by the pseudo-inverse C_{rkhs}^\dagger , because $C_{rkhs}^\dagger = B^{-1} A^\top A B^{-1}$ has the same form as C_{rkhs}^{-1} for the non-singular case. Specifically, the recursive relation eq. (3.4) becomes

$$\alpha_{i+1} z_{i+1} = C_{rkhs}^\dagger A^\top u_{i+1} - \beta_{i+1} z_i. \quad (3.6)$$

To compute α_{i+1} , we use the property that $z_i \in \mathcal{R}(C_{rkhs})$, which will be proved in Theorem 4.3. Note that $C_{rkhs}^\dagger C_{rkhs} = P_{\mathcal{N}(C_{rkhs})^\perp} = P_{\mathcal{R}(C_{rkhs})}$ since C_{rkhs} is symmetric, where $P_{\mathcal{X}}$ is the projection operator onto subspace \mathcal{X} . It follows that $C_{rkhs}^\dagger C_{rkhs} z_i = z_i$. Therefore, eq. (3.6) becomes

$$\alpha_{i+1} z_{i+1} = C_{rkhs}^\dagger (A^\top u_{i+1} - \beta_{i+1} C_{rkhs} z_i).$$

Letting $\bar{z}_i = C_{rkhs} z_i$ and $p = A^\top u_{i+1} - \beta_{i+1} \bar{z}_i$ again, we get $\alpha_{i+1} = \|C_{rkhs}^\dagger p\|_{C_{rkhs}} = (p^\top C_{rkhs}^\dagger p)^{1/2}$.

Thus, the two cases of C_{rkhs} lead to the same iterative process. We summarize the iterative process in Algorithm 2, and call it *generalized Golub-Kahan bidiagonalization* (gGKB).

Suppose the gGKB process terminates at step $k_t := \max_{k \geq 1} \{\alpha_k \beta_k > 0\}$. We show in Theorem 4.4–Theorem 4.5 that the output vectors $\{u_i\}_{i=1}^{k_t}$ and $\{z_i\}_{i=1}^{k_t}$ are orthonormal in $(\mathbb{R}^m, \langle \cdot, \cdot \rangle_2)$ and $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_{C_{rkhs}})$, respectively. In particular, they span two Krylov subspaces generated by $\{AC_{rkhs}^\dagger A^\top, b\}$ and $\{C_{rkhs}^\dagger A^\top A, C_{rkhs}^\dagger A^\top b\}$, respectively.

In matrix form, the k -step gGKB process with $k < k_t$, starting with vector b , produces a 2-orthonormal matrix $U_{k+1} = (u_1, \dots, u_{k+1}) \in \mathbb{R}^{m \times (k+1)}$ (i.e., $U_{k+1}^\top U_{k+1} = I_{k+1}$), a C_{rkhs} -orthonormal matrix $Z_{k+1} = (z_1, \dots, z_{k+1}) \in \mathbb{R}^{n \times (k+1)}$, and a bidiagonal matrix

$$B_k = \begin{pmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \beta_3 & \ddots & \\ & & \ddots & \alpha_k \\ & & & \beta_{k+1} \end{pmatrix} \in \mathbb{R}^{(k+1) \times k} \quad (3.7)$$

Require: $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $B = \text{diag}(\rho)$

- 1: Initialize: let $\beta_1 = \|b\|_2$, $u_1 = b/\beta_1$, and compute $p = A^\top u_1$, $s = B^{-1}A^\top AB^{-1}p$.
- 2: Let $\alpha_1 = (s^\top p)^{1/2}$, $z_1 = s/\alpha_1$, $\bar{z}_1 = p/\alpha_1$.
- 3: **for** $i = 1, 2, \dots, k$, **do**
- 4: $r = Az_i - \alpha_i u_i$, $\beta_{i+1} = \|r\|_2$, $u_{i+1} = r/\beta_{i+1}$;
- 5: $p = A^\top u_{i+1} - \beta_{i+1} \bar{z}_i$, $s = B^{-1}A^\top AB^{-1}p$; $\triangleright C_{rkhs}^\dagger = B^{-1}A^\top AB^{-1}$
- 6: $\alpha_{i+1} = (s^\top p)^{1/2}$, $z_{i+1} = s/\alpha_{i+1}$, $\bar{z}_{i+1} = p/\alpha_{i+1}$. $\triangleright \bar{z}_i = C_{rkhs} z_i$

Ensure: $\{\alpha_i, \beta_i\}_{i=1}^{k+1}$, $\{u_i, z_i, \bar{z}_i\}_{i=1}^{k+1}$

Algorithm 2: Generalized Golub-Kahan bidiagonalization (gGKB)

such that the recursion in eq. (3.3a), eq. (3.3b) and eq. (3.6) can be written as

$$\beta_1 U_{k+1} e_1 = b, \quad (3.8a)$$

$$AZ_k = U_{k+1} B_k, \quad (3.8b)$$

$$C_{rkhs}^\dagger A^\top U_{k+1} = Z_k B_k^T + \alpha_{k+1} z_{k+1} e_{k+1}^\top, \quad (3.8c)$$

where e_1 and e_{k+1} are the first and $(k+1)$ -th columns of I_{k+1} . We emphasize that B_k is a bidiagonal matrix of full column rank, since $\alpha_i, \beta_i > 0$ for all $i \leq k+1$.

P2. Recursively update the solution to the projected problem. For each $k \leq k_t$, i.e., before the gGKB terminates, we solve eq. (3.1) in the subspace

$$\mathcal{S}_k := \text{span}\{z_1, \dots, z_k\} \quad (3.9)$$

and compute the RKHS norm of the solution.

We show first the uniqueness of the solution to eq. (3.1). From eq. (3.8b) we have $B_k = U_{k+1}^\top AZ_k$, which implies that B_k is a projection of A onto the two subspaces $\text{span}\{U_{k+1}\}$ and $\text{span}\{Z_k\}$. Since any vector x in \mathcal{S}_k can be written as $x = Z_k y$ with a $y \in \mathbb{R}^k$, we obtain from eq. (3.8a) and eq. (3.8b) that for any $x = Z_k y$,

$$\begin{aligned} \min_{x=Z_k y} \|Ax - b\|_2 &= \min_{y \in \mathbb{R}^k} \|AZ_k y - U_{k+1} \beta_1 e_1\|_2 \\ &= \min_{y \in \mathbb{R}^k} \|U_{k+1} B_k y - U_{k+1} \beta_1 e_1\|_2 = \min_{y \in \mathbb{R}^k} \|B_k y - \beta_1 e_1\|_2. \end{aligned} \quad (3.10)$$

Since $k \leq k_t$, B_k has full column rank, this k -dimensional least squares problem has a unique solution. Therefore, the unique solution to eq. (3.1) is

$$x_k = Z_k y_k, \quad y_k = \underset{y \in \mathbb{R}^k}{\text{argmin}} \|B_k y - \beta_1 e_1\|_2. \quad (3.11)$$

We note that the above uniqueness is independent of the specific construction of the basis vectors $\{z_k\}$. In general, as long as $\mathcal{S}_k \subseteq \mathcal{R}(C_{rkhs})$, the solution to eq. (3.1) is unique; see Theorem 4.7.

Importantly, we recursively compute x_k without explicitly solving y_k in eq. (3.11), avoiding the $O(k^3)$ computational cost. To this end, we adopt a procedure similar to the LSQR algorithm in [30] to iteratively update x_k from $x_0 = \mathbf{0}$. It starts from the following Givens QR factorization:

$$Q_k (B_k \quad \beta_1 e_1) = \begin{pmatrix} R_k & f_k \\ & \bar{\gamma}_{k+1} \end{pmatrix} := \begin{pmatrix} \rho_1 & \theta_2 & & & & \gamma_1 \\ & \rho_2 & \theta_3 & & & \gamma_2 \\ & & \ddots & \ddots & & \vdots \\ & & & & \rho_{k-1} & \theta_k & \gamma_{k-1} \\ \hline & & & & & \rho_k & \gamma_k \\ & & & & & & \bar{\gamma}_{k+1} \end{pmatrix},$$

where the orthogonal matrix Q_k is the product of a series of Givens rotation matrices, and R_k is a bidiagonal upper triangular matrix; see [13, §5.2.5]. We implement the Givens QR factorization using the procedure in [30], which recursively zeros out the subdiagonal elements β_i for each $2 \leq i \leq k+1$. Specifically, at the i -th step, a Givens rotation zeros out β_{i+1} by

$$\begin{pmatrix} c_i & s_i \\ s_i & -c_i \end{pmatrix} \begin{pmatrix} \bar{\rho}_i & 0 & \bar{\gamma}_i \\ \beta_{i+1} & \alpha_{i+1} & 0 \end{pmatrix} = \begin{pmatrix} \rho_i & \theta_{i+1} & \gamma_i \\ 0 & \bar{\rho}_{i+1} & \bar{\gamma}_{i+1} \end{pmatrix},$$

where the entries c_i and s_i of the Givens rotations satisfy $c_i^2 + s_i^2 = 1$, and the elements $\rho_i, \theta_{i+1}, \bar{\rho}_{i+1}, \gamma_i, \bar{\gamma}_{i+1}$ are recursively updated accordingly; see [30] for more details.

As a result of the QR factorization, we can write

$$\|B_k y - \beta_1 e_1\|_2^2 = \|Q_k (B_k - \beta_1 e_1) \begin{pmatrix} y \\ -1 \end{pmatrix}\|_2^2 = \|R_k y - f_k\|_2^2 + |\bar{\gamma}_{k+1}|^2. \quad (3.12)$$

Hence, the solution to $\min_{y \in \mathbb{R}^k} \|B_k y - \beta_1 e_1\|$ is $y_k = R_k^{-1} f_k$. Factorizing R_k as

$$R_k = D_k \bar{R}_k, \quad D_k := \begin{pmatrix} \rho_1 & & & \\ & \rho_2 & & \\ & & \ddots & \\ & & & \rho_k \end{pmatrix}, \quad \bar{R}_k := \begin{pmatrix} 1 & \theta_2/\rho_1 & & \\ & 1 & \theta_3/\rho_2 & \\ & & \ddots & \theta_k/\rho_{k-1} \\ & & & 1 \end{pmatrix}$$

and denoting $W_k = Z_k \bar{R}_k^{-1} = (w_1, \dots, w_k)$, we get

$$x_k = Z_k y_k = Z_k R_k^{-1} f_k = (Z_k \bar{R}_k^{-1})(D_k^{-1} f_k) = W_k (D_k^{-1} f_k) = \sum_{i=1}^k (\gamma_i / \rho_i) w_i.$$

Updating x_k recursively, and solving $W_k \bar{R}_k = Z_k$ by back substitution, we obtain:

$$x_i = x_{i-1} + (\gamma_i / \rho_i) w_i, \quad w_{i+1} = z_{i+1} - (\theta_{i+1} / \rho_i) w_i, \quad \forall 1 \leq i \leq k. \quad (3.13)$$

Lastly, we compute $\|x_i\|_{C_{rkh_s}}$ without an explicit C_{rkh_s} . We have from eq. (3.13):

$$C_{rkh_s} x_i = C_{rkh_s} x_{i-1} + (\gamma_i / \rho_i) C_{rkh_s} w_i, \quad C_{rkh_s} w_{i+1} = C_{rkh_s} z_{i+1} - (\theta_{i+1} / \rho_i) C_{rkh_s} w_i.$$

Letting $\bar{x}_i = C_{rkh_s} x_i$ and $\bar{w}_i = C_{rkh_s} w_i$, and recalling that $\bar{z}_i = C_{rkh_s} z_i$, we have

$$\|x_i\|_{C_{rkh_s}}^2 = \bar{x}_i^\top \bar{x}_i, \quad \bar{x}_i = \bar{x}_{i-1} + (\gamma_i / \rho_i) \bar{w}_i, \quad \bar{w}_{i+1} = \bar{z}_{i+1} - (\theta_{i+1} / \rho_i) \bar{w}_i. \quad (3.14)$$

- ```

1: Let $x_0 = \mathbf{0}$, $\bar{x}_0 = \mathbf{0}$, $w_1 = z_1$, $\bar{w}_1 = \bar{z}_1$, $\bar{\gamma}_1 = \beta_1$, $\bar{\rho}_1 = \alpha_1$
2: for $i = 1, 2, \dots, k$, do
3: $\rho_i = (\bar{\rho}_i^2 + \beta_{i+1}^2)^{1/2}$, $c_i = \bar{\rho}_i / \rho_i$, $s_i = \beta_{i+1} / \rho_i$
4: $\theta_{i+1} = s_i \alpha_{i+1}$, $\bar{\rho}_{i+1} = -c_i \alpha_{i+1}$, $\gamma_i = c_i \bar{\gamma}_i$, $\bar{\gamma}_{i+1} = s_i \bar{\gamma}_i$
5: $x_i = x_{i-1} + (\gamma_i / \rho_i) w_i$, $w_{i+1} = z_{i+1} - (\theta_{i+1} / \rho_i) w_i$
6: $\bar{x}_i = \bar{x}_{i-1} + (\gamma_i / \rho_i) \bar{w}_i$, $\bar{w}_{i+1} = \bar{z}_{i+1} - (\theta_{i+1} / \rho_i) \bar{w}_i$
7: $\|x_i\|_{C_{rkh_s}} = (\bar{x}_i^\top \bar{x}_i)^{1/2}$

```

Algorithm 3: Updating procedure

The whole updating procedure is described in Algorithm 3. This algorithm yields the residual norm  $\|Ax_k - b\|_2$  without explicitly computing the residual. In fact, by eq. (3.10) and eq. (3.12) we have

$$\bar{\gamma}_{k+1} = \|B_k y_k - \beta_1 e_1\|_2 = \|Ax_k - b\|_2. \quad (3.15)$$

Note that  $\bar{\gamma}_{k+1}$  decreases monotonically since  $x_k$  minimizes  $\|Ax - b\|_2$  in the gradually expanding subspace  $\text{span}\{Z_k\}$ .

Importantly, Algorithm 3 efficiently computes the solution  $x_k$  and  $\|x_k\|_{C_{rkh_s}}$ . At each step of updating  $x_i$  or  $w_{i+1}$ , the computation takes  $O(2n)$  flops. Similarly, for updating  $\bar{x}_i$  or  $\bar{w}_{i+1}$ , as well as for computing  $\|x_i\|_{C_{rkh_s}}$ , the number of flops are also  $O(2n)$ . Therefore, the dominant computational cost is  $O(10n)$ . In contrast, if  $y_k$  is solved explicitly at each step, it takes  $O(\sum_{i=1}^k i^3) \sim O(k^4)$  flops; together with the step of forming  $x_k = Z_k y_k$  that takes  $O(kn)$  flops, they lead to a total cost of  $O(kn + k^4)$  flops. Thus, the LSQR-type iteration in Algorithm 3 significantly reduces the number of flops from  $O(kn + k^4)$  to  $O(10n)$ .

**P3. Regularize by early stopping.** An early stopping strategy is imperative to prevent the solution subspace from becoming excessively large, which could otherwise compromise the regularization. This necessity is rooted in the phenomenon of *semi-convergence*: the iteration vector  $x_k$  initially approaches an optimal regularized solution but subsequently moves towards the unstable naive solution to  $\min_{x \in \mathcal{R}(C_{rkh_s})} \|Ax - b\|_2$ , as detailed in Theorem 4.6.

For early stopping, we adopt the L-curve criterion, as outlined in [12]. This method identifies the ideal early stopping iteration  $k_*$  at the corner of the curve represented by

$$(\log \|Ax_k - b\|_2, \|x_k\|_{C_{rkh_s}}) = (\log \bar{\gamma}_{k+1}, \log \|x_k\|_{C_{rkh_s}}). \quad (3.16)$$

Here  $\bar{\gamma}_{k+1}$  and  $\|x_k\|_{C_{rkh_s}}$  are computed with negligible cost in Algorithm 3. To construct the L-curve effectively, we set the gGKB to execute at least 10 iterations. Additionally, to enhance numerical stability, we stop the gGKB when either  $\alpha_i$  or  $\beta_i$  is near the machine precision, as inspired by Theorem 4.6.

It is noteworthy that the *discrepancy principle* (DP) presents a viable alternative when the measurement error  $\|\mathbf{w}\|_2$  in eq. (1.1) is available with a high degree of accuracy. The DP halts iterations at the earliest instance of  $k$  that satisfies  $\bar{\gamma}_{k+1} = \|Ax_k - b\|_2 \leq \tau \|\mathbf{w}\|_2$ , where  $\tau$  is chosen to be marginally greater than 1.

### 3.3 Computational Complexity

Suppose the algorithm takes  $k$  iterations and the basis matrix  $B$  is diagonal. Recall that  $A \in \mathbb{R}^{m \times n}$ ,  $B \in \mathbb{R}^{n \times n}$ ,  $u_i \in \mathbb{R}^m$  and  $z_i \in \mathbb{R}^n$ . The total computational cost of Algorithm 1 is about  $O(3mnk)$  when  $m \leq n/3$  or  $k < n/3$ ; and about  $O((m+k)n^2)$  when otherwise. The cost is dominated by the gGKB process since the cost of the update procedure in Algorithm 3 is only  $O(n)$  at each step.

The gGKB can be computed in two approaches. The first approach uses only matrix-vector multiplication. The main computations in each iteration of gGKB occur at the matrix-vector products  $p = A^\top u_i$  and  $s = B^{-1}A^\top AB^{-1}p = B^{-1}(A^\top(A(B^{-1}p)))$ , which take  $O(mn)$  and  $O(2mn)$  flops respectively. Thus, the total computational cost of gGKB is  $O(3mnk)$  flops. Another approach is using  $\mathbf{A} = A^\top A$  instead of  $A^\top$  and  $A$  to compute  $s$ . In this approach, the computation of  $\mathbf{A}$  from  $A$  takes  $O(mn^2)$  flops, and the matrix-vector multiplication  $\mathbf{A}v$  in each iteration takes about  $O(n^2)$  flops. Hence, the total cost of  $k$  iterations is  $O(mn^2 + kn^2)$ . The second approach is faster when  $mn^2 + n^2k < 3mnk$ , or equivalently,  $(3m - n)k > mn$ . That is, roughly speaking,  $m > n/3$  and  $k > mn/(3m - n) > n/3$ .

In practice, the matrix-vector computation is preferred since the iteration number  $k$  is often small. The resulting iDARR algorithm takes about  $O(3mnk)$  flops.

## 4 Properties of gGKB

This section studies the properties of the gGKB in Algorithm 2, including the structure of the solution subspace, the orthogonality of the resulted vectors  $\{u_i\}_{i=1}^{k+1}$  and  $\{z_i\}_{i=1}^{k+1}$  and the number of iterations at termination, defined as:

$$\text{gGKB terminate step: } k_t := \max_{k \geq 1} \{\alpha_k \beta_k > 0\}. \quad (4.1)$$

Additionally, we show that the solution to eq. (3.1) is unique in each iteration.

Throughout this section, let  $r$  denote the rank of  $\Lambda$  and let  $V_r$  denote the first  $r$  columns of  $V$ , where  $\Lambda$  and  $V$  are matrices constituted by the generalized eigenvalues and eigenvectors of  $\{A, B\}$  in eq. (2.9) in Theorem 2.4. We have  $\text{rank}(A) = r$ . Recall that  $C_{rkhs} = (V\Lambda V^\top)^\dagger = B(A^\top A)^\dagger B$ . Note that the DA-RKHS is  $(\mathcal{R}(C_{rkhs}), \langle \cdot, \cdot \rangle_{C_{rkhs}})$ .

## 4.1 Properties of gGKB

We show first that the gGKB-produced vectors  $\{u_i\}_{i=1}^{k+1}$  and  $\{z_i\}_{i=1}^{k+1}$  are orthogonal in  $\mathbb{R}^n$  and in the DA-RKHS, and the solution subspaces of the gGKB are RKHS-restricted Krylov subspaces.

**Definition 4.1 (RKHS-restricted Krylov subspace)** Let  $A \in \mathbb{R}^{m \times n}$  and  $b \in \mathbb{R}^n$ , and let  $B \in \mathbb{R}^{n \times n}$  be a symmetric positive definite matrix. Let  $C_{rkhs} = B(A^\top A)^\dagger B$ , which defines an RKHS  $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_{C_{rkhs}})$ . The RKHS-restricted Krylov subspaces are

$$\mathcal{K}_{k+1}(C_{rkhs}^\dagger A^\top A, C_{rkhs}^\dagger A^\top b) = \text{span}\{(C_{rkhs}^\dagger A^\top A)^i C_{rkhs}^\dagger A^\top b\}_{i=0}^k, \quad k \geq 0. \quad (4.2)$$

The main result is the following.

**Theorem 4.2 (Properties of gGKB)** Recall  $k_t$  in eq. (4.1), and the gGKB generates vectors  $\{u_i\}_{i=1}^k$ ,  $\{z_i\}_{i=1}^k$  and  $\mathcal{S}_k = \text{span}\{z_i\}_{i=1}^k$ . They satisfy the following properties:

- (i)  $\{u_i\}_{i=1}^k$  and  $\{z_i\}_{i=1}^k$  are orthonormal in  $\mathbb{R}^n$  and in  $(\mathcal{R}(C_{rkhs}), \langle \cdot, \cdot \rangle_{C_{rkhs}})$ , respectively;
- (ii)  $\mathcal{S}_k = \mathcal{K}_k(C_{rkhs}^\dagger A^\top A, C_{rkhs}^\dagger A^\top b)$  for each  $k \leq k_t$ , and the termination iteration number is  $k_t = \dim(\mathcal{K}_\infty(C_{rkhs}^\dagger A^\top A, C_{rkhs}^\dagger A^\top b))$ .

**Proof.** Part (i) follows from Theorem 4.3, where we show that  $z_k \in \mathcal{R}(C_{rkhs})$ , and Theorem 4.4, where we show the orthogonality of these vectors.

For Part (ii),  $\mathcal{S}_k = \mathcal{K}_k(C_{rkhs}^\dagger A^\top A, C_{rkhs}^\dagger A^\top b)$  follows from that  $\{z_i\}_{i=1}^k$  form an orthonormal basis of the RKHS-restricted Krylov subspace. We prove  $k_t = \dim(\mathcal{K}_\infty(C_{rkhs}^\dagger A^\top A, C_{rkhs}^\dagger A^\top b))$  in Theorem 4.5. ■

**Proposition 4.3** For each  $z_i$  generated by gGKB in eq. (3.3), it holds that  $z_i \in \mathcal{R}(C_{rkhs})$ . Additionally, if  $q := C_{rkhs}^\dagger A^\top u_{i+1} - \beta_{i+1} z_i \neq \mathbf{0}$ , then  $\alpha_{i+1} = \|q\|_{C_{rkhs}} \neq 0$ .

**Proof.** We prove it by mathematical induction. For  $i = 1$ , we obtain from eq. (3.6) and Theorem 2.4 that  $\alpha_1 z_1 = C_{rkhs}^\dagger A^\top u_1 = V\Lambda V^\top A^\top u_1 \in \mathcal{R}(V_r) = \mathcal{R}(C_{rkhs})$ , where  $r$  is the rank of  $\Lambda$ . Suppose  $z_i \in \mathcal{R}(C_{rkhs})$  for  $i \geq 1$ . Using again eq. (3.6) and Theorem 2.4 we get

$$\alpha_{i+1} z_{i+1} = C_{rkhs}^\dagger A^\top u_{i+1} - \beta_{i+1} z_i = V\Lambda V^\top A^\top u_{i+1} - \beta_{i+1} z_i \in \mathcal{R}(C_{rkhs}).$$

Therefore,  $z_{i+1} \in \mathcal{R}(C_{rkhs})$ , and  $q \in \mathcal{R}(C_{rkhs})$ .

If  $\alpha_{i+1} = 0$ , then  $q \in \mathcal{N}(C_{rkhs}) = \mathcal{R}(C_{rkhs})^\perp$ . Therefore,  $q = \mathbf{0}$ . ■

Thus, even if  $C_{rkhs}$  is singular (positive semidefinite), the gGKB in Algorithm 2 does not terminate as long as the right-hand sides of eq. (3.3b) and eq. (3.6) are nonzero, since the iterative computation of  $\{\beta_{i+1}, u_{i+1}\}$  and  $\{\alpha_{i+1}, z_{i+1}\}$  can continue. Next, we show that these vectors are orthogonal.

**Proposition 4.4 (Orthogonality)** Suppose the  $k$ -step gGKB does not terminate, i.e.  $k < k_t$ , then  $\{u_i\}_{i=1}^{k+1}$  is a 2-orthonormal basis of the Krylov subspace

$$\mathcal{K}_{k+1}(AC_{rkhs}^\dagger A^\top, b) = \text{span}\{(AC_{rkhs}^\dagger A^\top)^i b\}_{i=0}^k, \quad (4.3)$$

and  $\{z_i\}_{i=1}^{k+1}$  is a  $C_{rkhs}$ -orthonormal basis for the RKHS-restricted Krylov subspace in eq. (4.2).

**Proof.** Note from Theorem 2.4 that  $C_{rkhs}^\dagger = V\Lambda V^\top$ . Let  $W_r = V_r\Lambda_r^{1/2}$ . Then  $W_r^\top C_{rkhs} W_r = I_r$ ,  $C_{rkhs}^\dagger = W_r W_r^\top$ , and  $\mathcal{R}(W_r) = \mathcal{R}(C_{rkhs})$ . For any  $z_i$ , Theorem 4.3 implies that there exists  $v_i \in \mathbb{R}^r$  such that  $z_i = W_r v_i$ . We get from eq. (3.3b) and eq. (3.6) that

$$\begin{aligned}\beta_{i+1} u_{i+1} &= A W_r v_i - \alpha_i u_i, \\ \alpha_{i+1} v_{i+1} &= W_r^\top A^\top u_{i+1} - \beta_{i+1} v_i,\end{aligned}$$

where the second equation comes from  $\alpha_{i+1} W_r v_{i+1} = W_r W_r^\top A^\top u_{i+1} - \beta_{i+1} W_r v_i$ . Combining the above two relations with eq. (3.3a), we conclude that the iterative process for generating  $u_i$  and  $v_i$  is the standard GKB process of  $A W_r$  with starting vector  $b$  between the two finite dimensional Hilbert spaces  $(\mathbb{R}^r, \langle \cdot, \cdot \rangle_2)$  and  $(\mathbb{R}^m, \langle \cdot, \cdot \rangle_2)$ . Therefore,  $\{u_i\}_{i=1}^{k+1}$  and  $\{v_i\}_{i=1}^{k+1}$  are two 2-orthonormal bases of the Krylov subspaces

$$\begin{aligned}\mathcal{K}_{k+1}(A W_r (A W_r)^\top, b) &= \text{span}\{(A W_r W_r^\top A^\top)^i b\}_{i=0}^k, \\ \mathcal{K}_{k+1}((A W_r)^\top A W_r, (A W_r)^\top b) &= \text{span}\{(W_r^\top A^\top A W_r)^i W_r^\top A^\top b\}_{i=0}^k,\end{aligned}$$

respectively; see e.g. [13, §10.4]. Then,  $W_r W_r^\top = C_{rkhs}^\dagger$  implies eq. (4.3). Also,  $\{z_i\}_{i=1}^{k+1} = \{W_r v_i\}_{i=1}^{k+1}$  is a  $C_{rkhs}$ -orthonormal basis of  $W_r \mathcal{K}_{k+1}((A W_r)^\top A W_r, (A W_r)^\top b)$  since  $W_r$  is  $C_{rkhs}$ -orthonormal.

Finally,  $\{z_i\}_{i=1}^{k+1}$  are  $C_{rkhs}$  orthogonal by construction, and by using the relation

$$W_r (W_r^\top A^\top A W_r)^i W_r^\top A^\top b = (W_r W_r^\top A^\top A)^i W_r W_r^\top A^\top b = (C_{rkhs}^\dagger A^\top A)^i C_{rkhs}^\dagger A^\top b,$$

we get that  $\{z_i\}_{i=1}^{k+1}$  in the RKHS-restricted Krylov subspace in eq. (4.2). ■

**Proposition 4.5 (gGKB termination number)** Suppose the gGKB in Algorithm 2 terminates at step  $k_t = \max_{k \geq 1} \{\alpha_k \beta_k > 0\}$ . Let the distinct nonzero eigenvalues of  $A C_{rkhs}^\dagger A^\top$  be  $\mu_1 > \dots > \mu_s > 0$  with multiplicities  $m_1, \dots, m_s$ , and the corresponding eigenspaces are  $\mathcal{G}_1, \dots, \mathcal{G}_s$ . Then,  $k_t = q$ , where  $q$  is the number of nonzero elements in  $\{P_{\mathcal{G}_1} b, \dots, P_{\mathcal{G}_s} b\}$ , and

$$q = \dim(\mathcal{K}_\infty(A C_{rkhs}^\dagger A^\top, b)) = \dim(\mathcal{K}_\infty(C_{rkhs}^\dagger A^\top A, C_{rkhs}^\dagger A^\top b)), \quad (4.4)$$

where  $K_\infty(M, v) = \text{span}\{M^i v\}_{i=0}^\infty$  denotes the Krylov subspace of  $\{M, v\}$ . Moreover,  $\sum_{i=1}^s m_i = r$  with  $r$  being the rank of  $A$  and  $k_t = q \leq r$ .

**Proof.** First, we prove eq. (4.4) with  $q$  being the number of nonzero elements in  $\{P_{\mathcal{G}_1} b, \dots, P_{\mathcal{G}_s} b\}$ . Let  $g_j = P_{\mathcal{G}_j} b / \|P_{\mathcal{G}_j} b\|_2$  for  $1 \leq j \leq s$ , and let  $g_1, \dots, g_q \neq \mathbf{0}$  without loss of generality. Note that  $\{g_j\}_{j=1}^q$  are orthonormal. Let  $G_j$  be a matrix with orthonormal columns that span  $\mathcal{G}_j$ . Note that  $P_{\mathcal{G}_j} = G_j G_j^\top$ . By the eigenvalue decomposition  $A C_{rkhs}^\dagger A^\top = \sum_{j=1}^s \mu_j G_j G_j^\top$ , we have

$$w_i := (A C_{rkhs}^\dagger A^\top)^{i-1} b = \sum_{j=1}^s \mu_j^{i-1} G_j G_j^\top b = \sum_{j=1}^q \mu_j^{i-1} \|P_{\mathcal{G}_j} b\|_2 g_j. \quad (4.5)$$

Hence,  $\text{rank}\{w_i\}_{i=1}^\infty \leq q$ . On the other hand, for  $1 \leq k \leq q$ , setting  $\bar{w}_i = \|P_{\mathcal{G}_j} b\|_2 g_j$ , we have  $(w_1, \dots, w_k) = (\bar{w}_1, \dots, \bar{w}_q) T_k$  with

$$T_k = \begin{pmatrix} 1 & \mu_1 & \dots & \mu_1^{k-1} \\ 1 & \mu_2 & \dots & \mu_2^{k-1} \\ \vdots & \vdots & \dots & \vdots \\ 1 & \mu_q & \dots & \mu_q^{k-1} \end{pmatrix} = \begin{pmatrix} T_{k1} \\ T_{k2} \end{pmatrix},$$

where  $T_{k1} \in \mathbb{R}^{k \times k}$  consists of the first  $k$  rows of  $T_k$ , and  $T_{k2} \in \mathbb{R}^{k, q-k}$  consists of the rest rows. Note that  $T_{k1}$  is a Vandermonde matrix and it is nonsingular since  $\mu_i \neq \mu_j$  for  $1 \leq i \neq j \leq k$ . Then,  $T_k$  has full column rank, thereby the rank of  $\{w_i\}_{i=1}^k$  is  $k$  for  $1 \leq k \leq q$ , and  $\text{rank}\{w_i\}_{i=1}^\infty \geq \text{rank}\{w_i\}_{i=1}^q = q$ . Therefore, we have  $\dim(\mathcal{K}_\infty(AC_{rkhs}^\dagger A^\top, b)) = \text{rank}\{w_i\}_{i=1}^\infty = q$ .

Also, we have  $\dim(\mathcal{K}_\infty(C_{rkhs}^\dagger A^\top A, C_{rkhs}^\dagger A^\top b)) = \text{rank}\{C_{rkhs}^\dagger A^\top w_i\}_{i=1}^\infty = q$ , where the first equality follows from

$$(C_{rkhs}^\dagger A^\top A)^{i-1} C_{rkhs}^\dagger A^\top b = C_{rkhs}^\dagger A^\top (AC_{rkhs}^\dagger A^\top)^{i-1} b = C_{rkhs}^\dagger A^\top w_i, \quad (4.6)$$

and the second equality follows from  $\text{rank}\{C_{rkhs}^\dagger A^\top w_i\}_{i=1}^\infty = \text{rank}\{w_i\}_{i=1}^\infty = q$  since  $C_{rkhs}^\dagger A^\top$  is non-singular on  $\text{span}\{w_i\}_{i=1}^\infty$ , which is a subset of  $\text{span}\{\mathcal{G}_l\}_{l=1}^s$  by eq. (4.5). In fact,  $C_{rkhs}^\dagger A^\top$  is non-singular on  $\text{span}\{\mathcal{G}_l\}_{l=1}^s$  because  $\{\mathcal{G}_l\}$  are eigenspaces of  $AC_{rkhs}^\dagger A^\top$  corresponding to the positive eigenvalues.

Furthermore, eq. (4.6) and the non-degeneracy of  $C_{rkhs}^\dagger A^\top$  on  $\text{span}\{w_i\}_{i=1}^\infty$  imply that

$$\begin{aligned} \dim(\mathcal{K}_q(C_{rkhs}^\dagger A^\top A, C_{rkhs}^\dagger A^\top b)) &= \text{rank}(\{C_{rkhs}^\dagger A^\top A\}^{i-1} C_{rkhs}^\dagger A^\top b)_{i=0}^{q-1} \\ &= \text{rank}(\{C_{rkhs}^\dagger A^\top w_i\}_{i=1}^q) = \text{rank}\{w_i\}_{i=1}^q = q. \end{aligned} \quad (4.7)$$

That is, the vectors  $\{C_{rkhs}^\dagger A^\top A\}^{i-1} C_{rkhs}^\dagger A^\top b\}_{i=0}^{q-1}$  are linearly independent.

Next, we prove that  $k_t = q$ . Clearly,  $k_t \leq q$  since by Theorem 4.4  $\{z_i\}_{i=1}^{k_t}$  are orthogonal and they are in  $\mathcal{K}_{k_t}(C_{rkhs}^\dagger A^\top A, C_{rkhs}^\dagger A^\top b) \subset \mathcal{K}_\infty(C_{rkhs}^\dagger A^\top A, C_{rkhs}^\dagger A^\top b)$ , whose dimension is  $q$ . On the other hand, we show next that if  $k_t < q$ , there will be a contradiction; hence, we must have  $k_t = q$ . In fact, eq. (3.3b) and eq. (3.6) imply that, for each  $1 \leq i \leq k_t$ ,

$$\begin{aligned} C_{rkhs}^\dagger A^\top A z_i &= \alpha_i C_{rkhs}^\dagger A^\top u_i + \beta_{i+1} C_{rkhs}^\dagger A^\top u_{i+1} \\ &= \alpha_i(\alpha_i z_i + \beta_i z_{i+1}) + \beta_{i+1}(\alpha_{i+1} z_{i+1} + \beta_{i+1} z_i), \end{aligned}$$

which leads to

$$\alpha_{i+1} \beta_{i+1} z_{i+1} = C_{rkhs}^\dagger A^\top A z_i - (\alpha_i^2 + \beta_{i+1}^2) z_i - \alpha_i \beta_i z_{i-1}.$$

Note that  $\alpha_1 \beta_1 z_1 = C_{rkhs}^\dagger A^\top \beta_1 u_1 = C_{rkhs}^\dagger A^\top b$ . Combining the above two relations and using  $\alpha_k \beta_k > 0$  for all  $k \leq k_t$ , it follows that  $z_k \in \text{span}\{(C_{rkhs}^\dagger A^\top A)^i C_{rkhs}^\dagger A^\top b\}_{i=0}^k$  for all  $k \leq k_t$ . Hence, recursively applying  $z_i = \frac{1}{\alpha_i \beta_i} C_{rkhs}^\dagger A^\top A z_{i-1} - \frac{1}{\alpha_i \beta_i} (\alpha_{i-1}^2 + \beta_i^2) z_{i-1} - \frac{\alpha_{i-1} \beta_{i-1}}{\alpha_i \beta_i} z_{i-2}$  for all  $2 \leq i \leq k_t$ , we can write

$$\alpha_{k_t+1} \beta_{k_t+1} z_{k_t+1} = \sum_{i=0}^{k_t} \xi_i (C_{rkhs}^\dagger A^\top A)^i C_{rkhs}^\dagger A^\top b,$$

with  $\xi_i \in \mathbb{R}$  and in particular,  $\xi_{k_t} = 1/\prod_{i=1}^{k_t} \alpha_i \beta_i \neq 0$ . Now  $\alpha_{k_t+1} \beta_{k_t+1} = 0$  implies that  $\{(C_{rkhs}^\dagger A^\top A)^i C_{rkhs}^\dagger A^\top b\}_{i=0}^{k_t}$  are linearly dependent, contradicting with the fact that they are linearly independent as suggested by eq. (4.7). Therefore, we have  $q = k_t$ .

Lastly, to prove that  $\sum_{i=1}^s m_i = r$ , it suffices to show that  $\text{rank}(AC_{rkhs}^\dagger A^\top) = r$  since the eigenvalues of  $AC_{rkhs}^\dagger A^\top$  are nonnegative. To see that its rank is  $r$ , following the proof of Theorem 4.4, we write  $AC_{rkhs}^\dagger A^\top = AW_r W_r^\top A^\top$ . Since  $AW_r = AV_r \Lambda_r^{1/2}$ , we only need to prove that  $AV_r$  has full column rank. Suppose  $AV_r y = \mathbf{0}$  with  $y \in \mathbb{R}^r$ . By Theorem 2.4 it follows  $A^\top AV_r y = \mathbf{0} \Leftrightarrow BV_r \Lambda_r y = \mathbf{0} \Leftrightarrow y = \mathbf{0}$ . Thus,  $\text{rank}(AC_{rkhs}^\dagger A^\top) = \text{rank}(AW_r) = r$ . ■

## 4.2 Uniqueness of solutions in the iterations

**Proposition 4.6** *If gGKB terminates at step  $k_t$  in eq. (4.1), the iterative solution  $x_{k_t}$  is the unique least squares solution to  $\min_{x \in \mathcal{R}(C_{rkhs})} \|Ax - b\|_2$ .*

**Proof.** Following the proof of Theorem 4.4, the solution to  $\min_{x \in \mathcal{R}(C_{rkh})} \|Ax - b\|_2$  is  $x_\star = W_r y_\star$  with  $y_\star = \operatorname{argmin}_y \|AW_r y - b\|_2$ . Since  $AW_r$  has full column rank, it follows that  $y_\star$  is the unique solution to  $W_r^\top A^\top (AW_r y - b) = \mathbf{0}$ . Note that  $\mathcal{R}(W_r) = \mathcal{R}(C_{rkh})$ . Thus,  $x_\star$  is the solution to  $\min_{x \in \mathcal{R}(C_{rkh})} \|Ax - b\|_2$  if and only if  $P_{\mathcal{R}(C_{rkh})} A^\top (Ax_\star - b) = \mathbf{0}$ .

Now we only need to prove  $P_{\mathcal{R}(C_{rkh})} A^\top (Ax_q - b) = \mathbf{0}$  since  $k_t = q$  by Theorem 4.5. Using the property  $P_{\mathcal{R}(C_{rkh})} = C_{rkh} C_{rkh}^\dagger$ , we get from eq. (3.8c)

$$P_{\mathcal{R}(C_{rkh})} A^\top U_{k+1} = C_{rkh} (Z_k B_k^T + \alpha_{k+1} z_{k+1} e_{k+1}^\top).$$

Combining the above relation with eq. (3.10), we have

$$\begin{aligned} P_{\mathcal{R}(C_{rkh})} A^\top (Ax_q - b) &= C_{rkh} (Z_q B_q^T + \alpha_{q+1} z_{q+1} e_{q+1}^\top) (B_q y_q - \beta_1 e_1) \\ &= C_{rkh} [Z_q (B_q^\top B_q y_q - B_q^\top \beta_1 e_1) + \alpha_{q+1} \beta_{q+1} z_{q+1} e_q^\top y_q] \\ &= \alpha_{q+1} \beta_{q+1} C_{rkh} z_{q+1} e_q^\top y_q = \mathbf{0}, \end{aligned}$$

since  $\alpha_{q+1} \beta_{q+1} = 0$  when gGKB terminates. ■

This result shows the necessity for early stopping the iteration to avoid getting a naive solution. The next theorem shows the uniqueness of the solution in each iteration of the algorithm.

**Theorem 4.7 (Uniqueness of solution in each iteration)** *For each iteration with  $k < k_t$ , there exists a unique solution to eq. (3.1). Furthermore, there exists a unique solution to  $\min_{x \in \mathcal{S}_k} \|Ax - b\|_2$ .*

**Proof.** Let  $W_k \in \mathbb{R}^{n \times k}$  that has orthonormal columns and spans  $\mathcal{S}_k$ . For any  $x \in \mathcal{S}_k$ , there is a unique  $y \in \mathbb{R}^k$  such that  $x = W_k y$ . Then the solution to eq. (3.1) should be  $x_k = W_k y_k$ , where  $y_k$  is the solution to

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

By [10, Theorem 2.1], it has a unique solution  $y_k$  iff  $\mathcal{N}(C_{rkh}^{1/2} W_k) \cap \mathcal{N}(AW_k) = \{\mathbf{0}\}$ .

Now we prove  $\mathcal{N}(C_{rkh}^{1/2} W_k) = \{\mathbf{0}\}$ . To this end, suppose  $y \in \mathcal{N}(C_{rkh}^{1/2} W_k)$  and  $x = W_k y$ . Then  $x \in \mathcal{S}_k \cap \mathcal{N}(C_{rkh}^{1/2})$ . Since  $\mathcal{N}(C_{rkh}^{1/2}) = \mathcal{N}(C_{rkh}) = \mathcal{R}(C_{rkh})^\perp$ , we get  $x \in \mathcal{S}_k \cap \mathcal{R}(C_{rkh})^\perp \subseteq \mathcal{R}(C_{rkh}) \cap \mathcal{R}(C_{rkh})^\perp = \{\mathbf{0}\}$ . Therefore,  $x = W_k y = \mathbf{0}$ , which leads to  $y = \mathbf{0}$ .

To prove the uniqueness of a solution to  $\min_{x \in \mathcal{S}_k} \|Ax - b\|_2$ , suppose that there are two minimizers,  $x_1 \neq x_2 \in \mathcal{S}_k$ , and we prove that they must be the same. Let  $x_* = x_1 - x_2$  and we have  $A^\top A x_* = 0$  since the minimizer of  $\mathcal{E}(x) = \|Ax - b\|_2$  must satisfy  $0 = \frac{1}{2} \nabla \mathcal{E}(x) = A^\top A x - A^\top b$ . That is,  $x_* \in \mathcal{N}(A^\top A)$ .

On the other hand, since  $x_* \in \mathcal{S}_k \subset \mathcal{R}(C_{rkh})$  by Theorem 4.3, and note that  $C_{rkh} = B(A^\top A)^\dagger B = BV_r \Lambda_r^{-1} V_r^\top B$ , we have  $B^{-1} x_* \in \mathcal{R}((A^\top A)^\dagger B) \subset \mathcal{N}(A^\top A)^\perp$ .

Combining the above two, we have  $\langle x_*, B^{-1} x_* \rangle = 0$ . But  $B$  is a symmetric positive definite matrix, so we must have  $x_* = 0$ . ■

## 5 Numerical Examples

### 5.1 The Fredholm equation of the first kind

We first examine the iDARR in solving the discrete Fredholm integral equation of the first kind. The tests cover two distinct types of spectral decay: exponential and polynomial. The latter is well-known to be challenging, often occurring in applications such as image deblurring. Additionally, we investigate two scenarios concerning whether the true solution is inside or outside the function space of identifiability (FSOI).

Table 2: Three regularization norms in iterative and direct methods.

| Norms   | $\ x\ _*^2$         | Iterative | Direct |
|---------|---------------------|-----------|--------|
| $l^2$   | $x^\top I_n x$      | IR-l2     | l2     |
| $L^2$   | $x^\top B x$        | IR-L2     | L2     |
| DA-RKHS | $x^\top C_{rkhs} x$ | iDARR     | DARTR  |

**Three norms in iterative and direct methods.** We compare the  $l^2$ ,  $L^2$ , and DA-RKHS norms in iterative and direct methods. The direct methods are based on matrix decomposition. These regularizers are listed in Table 2.

The iterative methods differ primarily in their regularization norms. For the  $l^2$  norm, we use the LSQR method in the **IR TOOLS** package [12], and we stop the iteration when  $\alpha_i$  or  $\beta_i$  becomes negligible to maintain stability. For the  $L^2$  norm, we use gGKB to construct solution subspaces by replacing  $C_{rkhs}$  by the basis matrix  $B$ ; note that this method is equivalent to the LSQR method using  $L = \sqrt{B}$  as a preconditioner in the **IR TOOLS** package.

The direct methods are Tikhonov regularizers using the L-curve method [16].

**Numerical settings.** We consider the problem of recovering the input signal  $\phi$  in a discretization of Fredholm integral equation in eq. (2.1) with  $s \in [a, b]$  and  $t \in [c, d]$ . The data are discrete noisy observations  $b = (y(t_1), \dots, y(t_m)) \in \mathbb{R}^m$ , where  $t_j = c + j(d - c)/m$  for  $0 \leq j \leq m$ . The task is to estimate the coefficient vector  $x = (\phi(s_1), \dots, \phi(s_n)) \in \mathbb{R}^n$  in a piecewise-constant function  $\phi(s) = \sum_{i=1}^n \phi(s_i) \mathbf{1}_{[s_{i-1}, s_i]}(s)$ , where  $\mathcal{S} := \{s_i\}_{i=1}^n \subset [a, b]$  with  $s_i = a + i\delta$ ,  $\delta = (b - a)/n$ . We obtain the linear system eq. (1.1) with  $A(j, i) = K(t_j, s_i)\delta$  by a Riemann sum approximation of the integral. We set  $(a, b, c, d) = (1, 5, 0, 5)$ ,  $m = 500$ , and take  $n = 100$  except when testing the computational time with a sequence of large values for  $n$ .

The  $\mathbb{R}^m$ -valued noise  $\mathbf{w}$  is Gaussian  $N(0, \sigma^2 \Delta t I_m)$ . We set the standard deviation of the noise to be  $\sigma = \|Ax\| \times nsr$ , where  $nsr$  is the noise-to-signal ratio, and we test our methods with  $nsr = \{0.0625, 0.125, 0.25, 0.5, 1\}$ .

We consider two integral kernels

$$(a) \quad K(t, s) := s^{-2} e^{-st}; \quad (b) \quad K(t, s) := s^{-1} |\sin(st + 1)|. \quad (5.1)$$

which lead to exponential and polynomial decaying spectra, respectively, as shown in Figure 1. The first kernel arises from magnetic resonance relaxometry [4] with  $\phi$  being the distribution of transverse nuclear relaxation times.

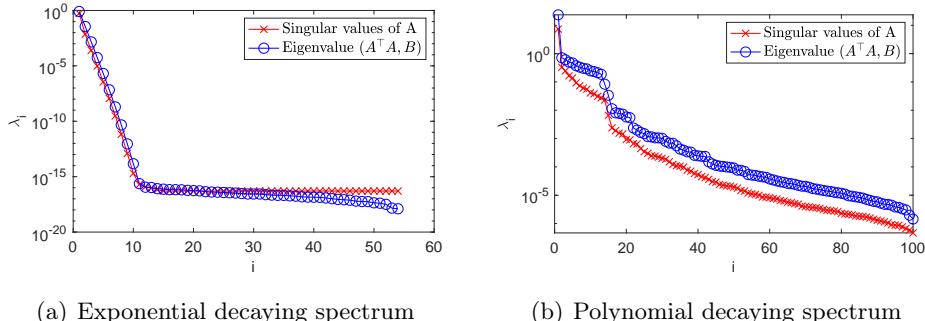


Figure 1: Singular values of  $A$  and generalized eigenvalues of  $(A^\top A, B)$  for kernels in eq. (5.1).

**The DA-RKHS.** By its definition in eq. (2.4), the exploration measure is  $\rho(s_i) = \frac{1}{\gamma} \sum_{j=1}^m |K(t_j, s_i)|\delta$  for  $i = 1, \dots, n$  with  $\gamma$  being the normalizing constant. In other words, it is the normalized column sum

of the absolute values of the matrix  $A$ . The discrete function space  $L^2_\rho(\mathcal{S})$  is equivalent to  $\mathbb{R}^n$  with weight  $\rho = (\rho(s_1), \dots, \rho(s_n))$ , and its norm is  $\langle x, x \rangle_{L^2_\rho} := x^\top \text{diag}(\rho(s_i))x$  for all  $x \in \mathbb{R}^n$ . The basis matrix for the Cartesian basis of  $\mathbb{R}^n$  is  $B = \text{diag}(\rho(s_i))$ , which is also the basis matrix for step functions in the Riemann sum discretization. The DA-RKHS in this discrete setting is  $(\mathcal{N}(A^\top A))^\perp, \langle \cdot, \cdot \rangle_{C_{rkhs}}$ .

**Settings for comparisons.** The comparison consists of two scenarios regarding whether the true solution is inside or outside of the FSOI: (i) the true solution is the second eigenvector of  $\mathcal{L}_{\overline{G}}$ ; thus it is inside the FSOI; and (ii) the true solution is  $\phi(x) = x^2$ , which has significant components outside the FSOI.

For each scenario, we conduct 100 independent simulations, with each simulation comprising five datasets at varying noise levels. The results are presented by a box plot, which illustrates the median, the lower and upper quartiles, any outliers, and the range of values excluding outliers. The key indicator of a regularizer's effectiveness is its ability to produce accurate estimators whose errors decay consistently as the noise level decreases. Since exploring the decay rate in the small noise limit is not the focus of this study, we direct readers to [22, 27] for initial insights into how this rate is influenced by factors such as spectral decay, the smoothness of the true solution, and the choice of regularization strength.

**Results.** We report the results separately according to the spectral decay.

**(i) Exponential decaying spectrum.** Figure 2's top row shows typical estimators of IR-l2, IR-L2, and iDARR and their de-noising of the output signal when  $nsr = 0.5$ . When the true solution is inside the FSOI, the iDARR significantly outperforms the other two in producing a more accurate estimator. However, both IR-L2 and IR-l2 can denoise the data accurately, even though their estimators are largely biased. When the true solution is outside the FSOI, all the regularizers can not capture the true function accurately, but iDARR and IR-L2 clearly outperform the  $l_2$  regularizer. Yet again, all these largely biased estimators can de-noise the data accurately. Thus, this inverse problem is severely ill-defined, and one must restrict the inverse to be in the FSOI.

The 2nd top row of Figure 2 shows the decay of the residual  $\|Ax_k - y\|_2$  as the iteration number increases, as well as the stopping iteration numbers of these regularizers in 100 simulations. The fast decaying residual suggests the need for early stopping, and all three regularizers indeed stop in a few steps. Notably, iDARR consistently stops early at the second iteration for different noise levels, outperforming the other two regularizers in stably detecting the stopping iteration.

The effectiveness of the DA-RKHS regularization becomes particularly evident in the lower two rows of Figure 2, which depict the decaying errors and loss values as the noise-to-signal ratio ( $nsr$ ) decreases in the 100 independent simulations. In both iterative and direct methods, the DA-RKHS norm demonstrates superior performance compared to the  $l^2$  and  $L^2$  norms, consistently delivering more accurate estimators that show a steady decrease in error alongside the noise level. Notably, the values of the corresponding loss functions are similar, underscoring the inherent ill-posedness of the inverse problem. Furthermore, iDARR marginally surpasses the direct method DARTR in producing more precise estimators, particularly when the true solution resides within the FSOI. The performance of iDARR suggests that its early stopping mechanism can reliably determine an optimal regularization level, achieving results that are slightly more refined than those obtained with DARTR using the L-curve method.

**(ii) Polynomial decaying spectrum.** Figure 3 illustrates again the superior performance of iDARR over IR-L2 and IR-l2 in the case of polynomial spectral decay. The 2nd top row shows that the slow spectral decay poses a notable challenge to the iterative methods, as the noise level affects their stopping iteration numbers. Also, they all stop early within approximately twelve steps, even though the true solution may lie in a subspace with a higher dimension.

The lower two rows show that iDARR remains effective. It continues to outperform the other two iterative regularizers when the true solution is in the FSOI, and it is marginally surpassed by IR-L2 and IR-l2 when the true solution is outside the FSOI. In both scenarios, the direct method DARTR outperforms all other methods, including iDARR, indicating DARTR is more effective in extracting

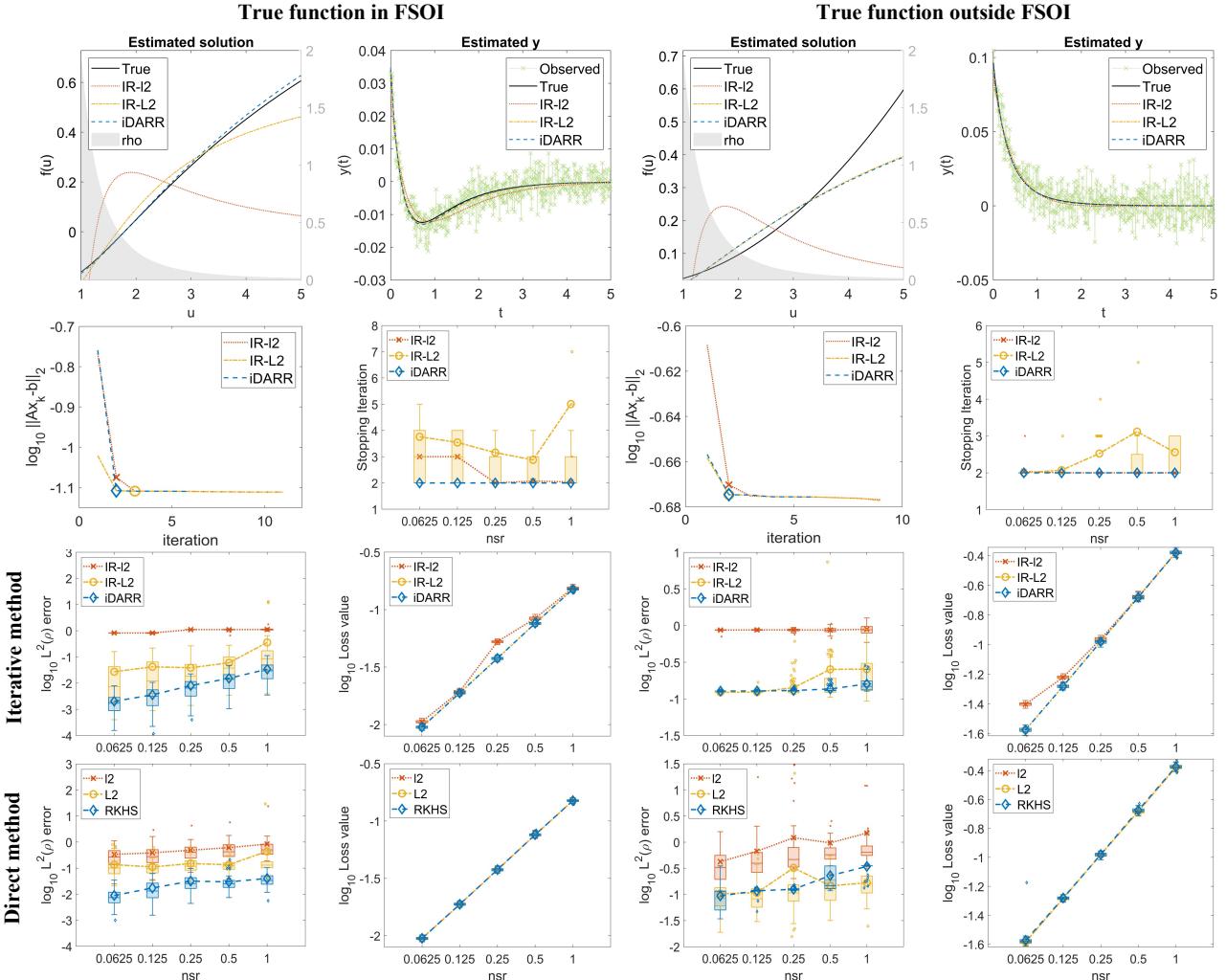


Figure 2: Results in the case of exponentially decaying spectrum. Top-row: typical estimators of IR-l2, IR-L2, and iDARR when  $nsr = 0.5$  and their denoising of the output signal. The 2nd-top row: the residual  $\|Ax_k - b\|_2$  as iteration number  $k$  increases in one realization when  $nsr = 0.5$ , as well as the boxplots of the stopping iteration numbers in the 100 simulations. The lower two rows: boxplots of the estimators'  $L_p^2$  errors and loss function values in the 100 simulations.

information from the spectrum with slow decay.

**Computational Complexity.** The iterative method iDARR is orders of magnitude faster than the direct method DARTR, especially when  $n$  is large. Figure 4 shows their computation time as  $n$  increases in 10 independent simulations, and the results align with the complexity order illustrated in Section 3.3.

In summary, iDARR outperforms IR-L2 and IR-l2 in yielding accurate estimators that consistently decay with the noise level. Its major advantage comes from the DA-RKHS norm that adaptively exploits the information in data and the model.

## 5.2 Image Deblurring

We further test iDARR in 2D image deblurring problems, where the task is to reconstruct images from blurred and noisy observations. The mathematical model of this problem can be expressed in the form of the first-kind Fredholm integral equation in eq. (2.1) with  $s, t \in \mathbb{R}^2$ . The kernel  $K(t, s)$  is a function that specifies how the points in the image are distorted, called the point spread function (PSF). We chose PRblurspeckle from [12] as the blurring operator, which simulates spatially invariant blurring caused by

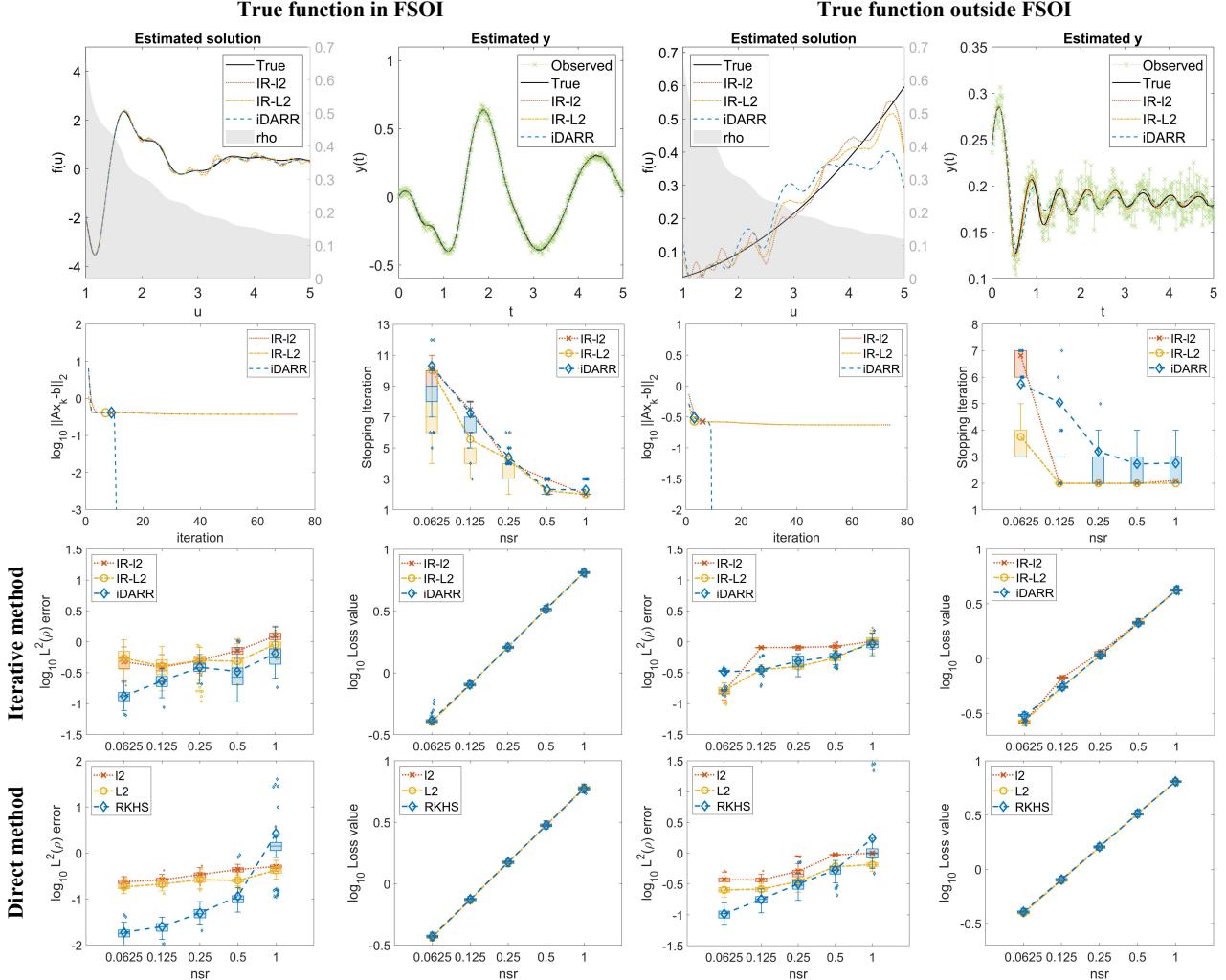


Figure 3: Results in the case of polynomial decaying spectrum. Top-row: typical estimators of IR-I2, IR-L2, and iDARR when  $nsr = 0.0625$  and their denoising of the output signal. The 2nd-top row: the residual  $\|Ax_k - b\|_2$  as iteration number  $k$  increases in one realization when  $nsr = 0.0625$ , as well as the box plots of the stopping iteration numbers the 100 simulations. The lower two rows: box plots of the estimators'  $L^2_p$  errors and loss function values in the 100 simulations.

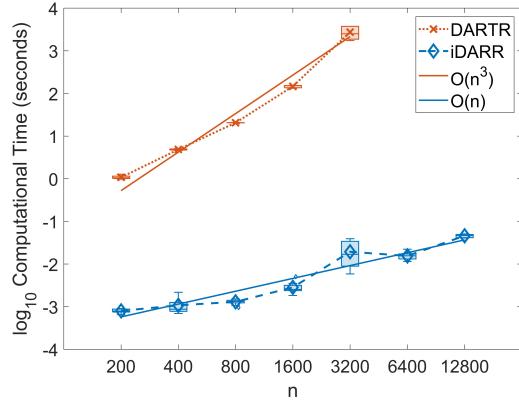


Figure 4: Computational time in 10 simulations with  $m = 500$ .

atmospheric turbulence, and we use zero boundary conditions to construct the matrix  $A$ . For a true

image with  $N \times N$  pixels, the matrix  $A \in \mathbb{R}^{N^2 \times N^2}$  is a `psfMatrix` object. We consider two images with  $256 \times 256$  and  $320 \times 320$  pixels, respectively, and set the noise level to be  $nsr = 0.01$  for both images. The true images, their blurred noisy observations, and corresponding PSFs that define matrices  $A$  are presented in Figure 5.

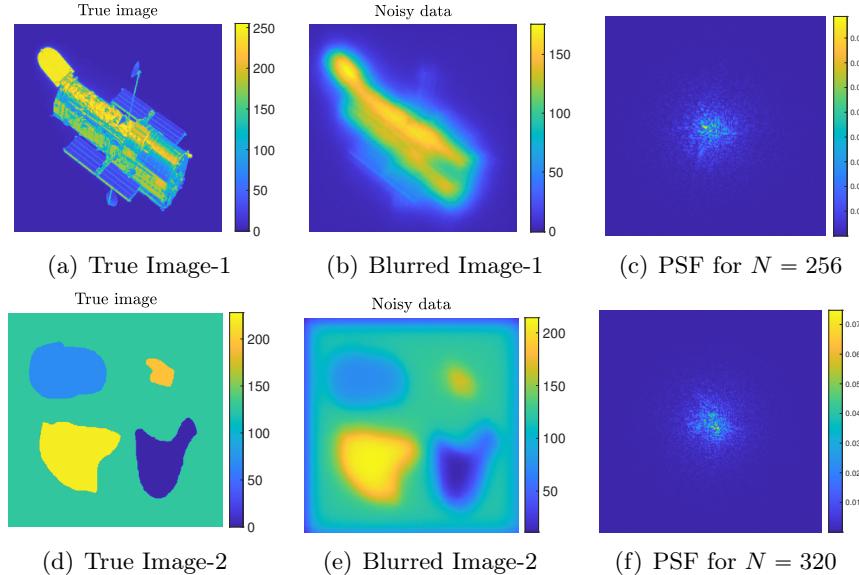


Figure 5: The true images, noisy images blurred by `PRblurspeckle`, and the corresponding PSFs.

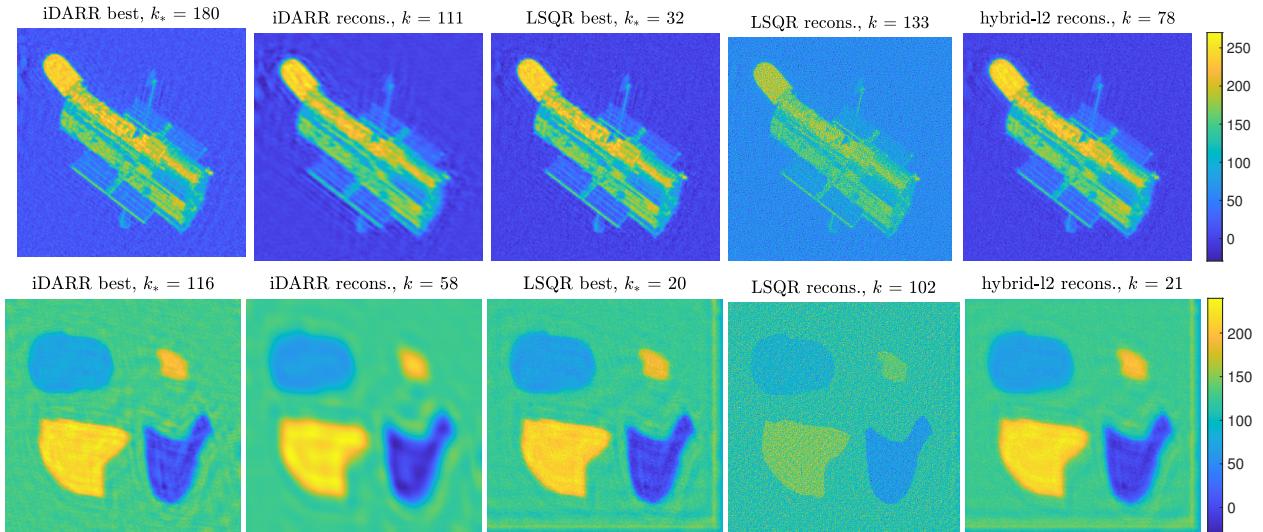


Figure 6: The reconstructed images computed by iDARR, LSQR and hybrid-l2 methods.

Figure 6 shows the reconstructed images computed by iDARR, LSQR, and the hybrid-l2 methods. Here the hybrid-l2 applies an  $l_2$ -norm Tikhonov regularization to the projected problem obtained by LSQR, and it uses the stopping strategy in [12]. The best estimations of for iDARR or LSQR are solutions with  $k_*$  minimizing  $\|x_k - x_{\text{true}}\|_2$ . Their reconstructed solutions are obtained by using the L-curve method for early stopping.

Figure 7 (a)–(f) show the relative errors as the iteration number increases and the selection of the early stopping iterations by the L-curve method in the right two columns.

Notably, Figure 7 reveals that iDARR achieves more accurate reconstructed images than LSQR

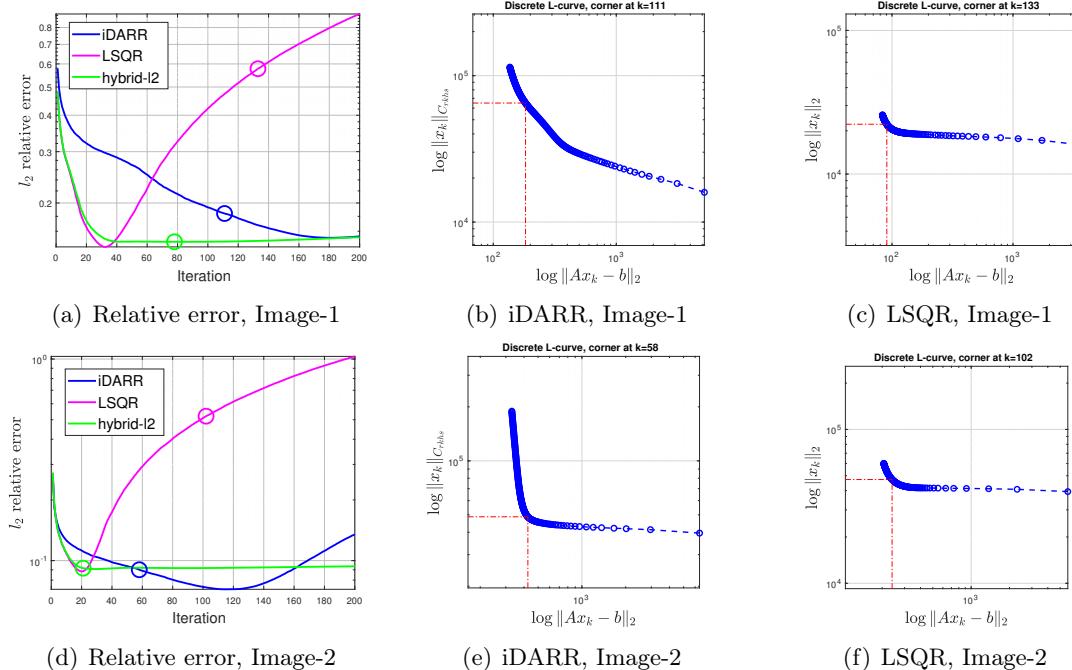


Figure 7: Relative errors in iteration numbers, where the circles mark the early stopping iterations chosen by the L-curve method presented in the right two columns.

for both tests, despite appearing to the contrary in Figure fig:deblur. The LSQR appears prone to stopping late, resulting in lower-quality reconstructions than iDARR. In contrast, iDARR tends to stop earlier than ideal, before achieving the best quality. However, the hybrid-l2 method consistently produces accurate estimators with stable convergence, suggesting potential benefits in developing a hybrid iDARR approach to enhance stability.

The effectiveness of iDARR depends on the alignment of regularities between the convolution kernel and the image, as the DA-RKHS's regularity is tied to the smoothness of the convolution kernel. With the PRblurspeckle featuring a smooth PSF, iDARR obtains a higher accuracy for the smoother Image-2 compared to Image-1, producing reconstructions with smooth edges. An avenue for future exploration involves adjusting the DA-RKHS's smoothness to better align with the smoothness of the data.

## 6 Conclusion and Future Work

We have introduced iDARR, a scalable iterative data-adaptive RKHS regularization method, for solving ill-posed linear inverse problems. It searches for solutions in the subspaces where the true signal can be identified and achieves reliable early stopping via the DA-RKHS norm. A core innovation is a generalized Golub-Kahan bidiagonalization procedure that recursively computes orthonormal bases for a sequence of RKHS-restricted Krylov subspaces. Systematic numerical tests on the Fredholm integral equation show that iDARR outperforms the widely used iterative regularizations using the  $L^2$  and  $l^2$  norms, in the sense that it produces stable accurate solutions consistently converging when the noise level decays. Applications to 2D image de-blurring further show the iDARR outperforms the benchmark of LSQR with the  $l^2$  norm.

**Future Work: Hybrid Methods** The accuracy and stability of the regularized solution hinges on the choice of iteration number for early stopping. While the L-curve criterion is a commonly used tool for determining this number, it can sometimes lead to suboptimal results due to its reliance on identifying a corner in the discrete curve. Hybrid methods are well-recognized alternatives that help stabilize

this semi-convergence issue, as referenced in [8, 20, 32]. One promising approach is to apply Tikhonov regularization to each iteration of the projected problem. The hyperparameter for this process can be determined using the *weighted generalized cross-validation method* (WGCV) as described in [8]. This approach is a focus of our upcoming research project.

## References

- [1] Simon R Arridge, Marta M Betcke, and Lauri Harhanen. Iterated preconditioned LSQR method for inverse problems on unstructured grids. *Inverse Probl.*, 30(7):075009, 2014.
- [2] Xianglan Bai, Guang-Xin Huang, Xiao-Jun Lei, Lothar Reichel, and Feng Yin. A novel modified TRSVD method for large-scale linear discrete ill-posed problems. *Appl. Numer. Math.*, 164:72–88, 2021.
- [3] Frank Bauer, Sergei Pereverzev, and Lorenzo Rosasco. On regularization algorithms in learning theory. *Journal of complexity*, 23(1):52–72, 2007.
- [4] Chuan Bi, M. Yvonne Ou, Mustapha Bouhrara, and Richard G. Spencer. Span of regularization for solution of inverse problems with application to magnetic resonance relaxometry of the brain. *Scientific Reports*, 12(1):20194, 2022.
- [5] Åke Björck. A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations. *BIT Numer. Math.*, 28(3):659–670, 1988.
- [6] Noe Angelo Caruso and Paolo Novati. Convergence analysis of LSQR for compact operator equations. *Linear Algebra and its Applications*, 583:146–164, 2019.
- [7] Zhiming Chen, Wenlong Zhang, and Jun Zou. Stochastic convergence of regularized solutions and their finite element approximations to inverse source problems. *SIAM Journal on Numerical Analysis*, 60(2):751–780, 2022.
- [8] Julianne Chung, James G Nagy, and Dianne P O’Leary. A weighted-GCV method for Lanczos-hybrid regularization. *Electr. Trans. Numer. Anal.*, 28(29):149–167, 2008.
- [9] Felipe Cucker and Ding Xuan Zhou. *Learning theory: an approximation theory viewpoint*, volume 24. Cambridge University Press, Cambridge, 2007.
- [10] Lars Eldén. A weighted pseudoinverse, generalized singular values, and constrained least squares problems. *BIT Numer. Math.*, 22:487–502, 1982.
- [11] Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.
- [12] Silvia Gazzola, Per Christian Hansen, and James G Nagy. Ir tools: a matlab package of iterative regularization methods and large-scale test problems. *Numerical Algorithms*, 81(3):773–811, 2019.
- [13] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 4th edition, 2013.
- [14] Jacques Salomon Hadamard. *Lectures on Cauchy’s problem in linear partial differential equations*, volume 18. Yale University Press, 1923.
- [15] Per Christian Hansen. *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. SIAM, 1998.
- [16] Per Christian Hansen. The L-curve and its use in the numerical treatment of inverse problems. In *in Computational Inverse Problems in Electrophysiology*, ed. P. Johnston, *Advances in Computational Bioengineering*, pages 119–142. WIT Press, 2000.
- [17] Guangxin Huang, Yuanyuan Liu, and Feng Yin. Tikhonov regularization with MTRSVD method for solving large-scale discrete ill-posed problems. *J. Comput. Appl. Math.*, 405:113969, 2022.

- [18] Zhongxiao Jia and Yanfei Yang. A joint bidiagonalization based iterative algorithm for large scale general-form Tikhonov regularization. *Appl. Numer. Math.*, 157:159–177, 2020.
- [19] Misha E. Kilmer, Per Christian Hansen, and Malena I. Espanol. A projection-based approach to general-form Tikhonov regularization. *SIAM J. Sci. Comput.*, 29(1):315–330, 2007.
- [20] Misha E Kilmer and Dianne P O’Leary. Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM J. Matrix Anal. Appl.*, 22(4):1204–1221, 2001.
- [21] Jörg Lampe, Lothar Reichel, and Heinrich Voss. Large-scale Tikhonov regularization via reduction by orthogonal projection. *Linear Algebra Appl.*, 436(8):2845–2865, 2012.
- [22] Quanjun Lang and Fei Lu. Small noise analysis for Tikhonov and RKHS regularizations. *arXiv preprint arXiv:2305.11055*, 2023.
- [23] Gongsheng Li and Zuhair Nashed. A modified Tikhonov regularization for linear operator equations. *Numerical functional analysis and optimization*, 26(4-5):543–563, 2005.
- [24] Haibo Li. A preconditioned Krylov subspace method for linear inverse problems with general-form Tikhonov regularization. *arXiv:2308.06577v1*, 2023.
- [25] Fei Lu, Qingci An, and Yue Yu. Nonparametric learning of kernels in nonlocal operators. *Journal of Peridynamics and Nonlocal Modeling*, pages 1–24, 2023.
- [26] Fei Lu, Quanjun Lang, and Qingci An. Data adaptive RKHS Tikhonov regularization for learning kernels in operators. *Proceedings of Mathematical and Scientific Machine Learning, PMLR 190:158-172*, 2022.
- [27] Fei Lu and Miao-Jung Yvonne Ou. An adaptive RKHS regularization for Fredholm integral equations. *arXiv preprint arXiv:2303.13737*, 2023.
- [28] M Zuhair Nashed and Grace Wahba. Generalized inverses in reproducing kernel spaces: an approach to regularization of linear operator equations. *SIAM Journal on Mathematical Analysis*, 5(6):974–987, 1974.
- [29] Dianne P O’Leary and John A Simmons. A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems. *SIAM J. Sci. Statist. Comput.*, 2(4):474–489, 1981.
- [30] Christopher C Paige and Michael A Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Software*, 8:43–71, 1982.
- [31] Lothar Reichel, Fiorella Sgallari, and Qiang Ye. Tikhonov regularization based on generalized Krylov subspace methods. *Appl. Numer. Math.*, 62(9):1215–1228, 2012.
- [32] R A Renaut, S Vatankhah, and V E Ardesta. Hybrid and iteratively reweighted regularization by unbiased predictive risk and weighted GCV for projected systems. *SIAM J. Sci. Statist. Comput.*, 39(2):B221–B243, 2017.
- [33] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.
- [34] Robert Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.
- [35] Andrei Nikolajevits Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Math.*, 4:1035–1038, 1963.
- [36] Grace Wahba. Convergence rates of certain approximate solutions to fredholm integral equations of the first kind. *Journal of Approximation Theory*, 7(2):167–185, 1973.
- [37] Grace Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM journal on numerical analysis*, 14(4):651–667, 1977.
- [38] Yimin Wei, Pengpeng Xie, and Liping Zhang. Tikhonov regularization and randomized GSVD. *SIAM J. Matrix Anal. Appl.*, 37(2):649–675, 2016.

- [39] Hua Xiang and Jun Zou. Regularization with randomized svd for large-scale discrete inverse problems. *Inverse Problems*, 29(8):085008, 2013.
- [40] Hua Xiang and Jun Zou. Randomized algorithms for large-scale inverse problems with general Tikhonov regularizations. *Inverse Probl.*, 31(8):085008, 2015.
- [41] Ye Zhang and Chuchu Chen. Stochastic asymptotical regularization for linear inverse problems. *Inverse Problems*, 39(1):015007, 2022.