

Data-informed posterior approximation for Bayesian linear inverse problems

Haibo Li *

Abstract

Computing posterior distributions in large-scale Bayesian linear inverse problems is challenging due to the high dimensionality of the parameter space. In this work, we develop a data-informed framework that shifts the computational focus from the parameter space to the data space. We rigorously characterize an intrinsically low-dimensional data space, establish its isometric embedding into the parameter space, and show that the prior-to-posterior update is confined to a data-informed subspace. This perspective allows posterior inference to be carried out in a reduced data-informed subspace. Based on this formulation, we propose a quotient-space Golub–Kahan bidiagonalization method to construct data-informed Krylov subspaces, and integrate empirical Bayesian inference into the iterative framework, enabling simultaneous hyperparameter estimation and posterior approximation in a matrix-free manner. Numerical experiments on representative problems support the theoretical framework and demonstrate the effectiveness of the resulting method.

Keywords. Bayesian inverse problem, data-informed subspace, posterior approximation, quotient space, Krylov subspace method, Golub–Kahan bidiagonalization

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 2 |
| 2 | Data-informed structure of the parameter space | 4 |
| 3 | Quotient-space Krylov method in the data space | 8 |
| 3.1 | Quotient-space Golub–Kahan bidiagonalization | 9 |
| 3.2 | Data-informed Krylov subspace | 11 |
| 4 | Iterative approximation of the Bayesian posterior | 14 |
| 4.1 | Approximate the Bayesian posterior by Q-GKB | 14 |
| 4.2 | Accuracy of the approximate posterior | 17 |
| 5 | Numerical experiments | 20 |
| 5.1 | One-dimensional Fredholm integral equation | 20 |
| 5.2 | Two-dimensional image deblurring | 23 |
| 5.3 | X-ray computed tomography | 26 |
| 6 | Conclusion | 28 |
| A | Proofs of technical lemmas | 29 |

*School of Mathematics and Statistics & Hubei Key Laboratory of Engineering Modeling and Scientific Computing, Huazhong University of Science and Technology, Wuhan 430074, China. haiboli@hust.edu.cn, haibolee1729@gmail.com

1 Introduction

Inverse problems arise across a diverse spectrum of scientific and engineering disciplines, including image reconstruction, computed tomography, geoscience, data assimilation and so on [17, 20, 3, 23, 38]. The goal is to recover an unknown parameter or function from indirect measurements contaminated by noise. For linear inverse problems, upon discretization of the underlying continuous forward operator, these problems are typically formulated as a high-dimensional linear system:

$$y = Gx + \eta, \quad \eta \sim \mathcal{N}(0, \Gamma), \quad (1.1)$$

where $x \in \mathbb{R}^n$ represents the discretized unknown, $y \in \mathbb{R}^m$ is the observation data, and $G \in \mathbb{R}^{m \times n}$ is the forward matrix. The noise η is assumed to follow a Gaussian distribution with a symmetric positive definite covariance $\Gamma \in \mathbb{R}^{m \times m}$. A fundamental challenge in solving such systems is the ill-posed nature of the inverse problem, which means that there may be multiple solutions that fit the observation data equally well, or the solution is very sensitive with respect to the observation noise.

To address the ill-posedness, two predominant frameworks are widely employed: deterministic Tikhonov regularization approach and probabilistic Bayesian inference approach. The former is based on minimizing a loss function $\ell(Gx, y)$ added by a regularizer $R(x)$ that incorporates prior information about x [46, 11, 16]. In contrast, the Bayesian framework treats x and y as random variables, providing a complete characterization of the solution through the posterior distribution $x | y$ via Bayes' theorem [20, 45]. This approach has gained significant traction due to its ability to provide a rigorous quantification of the underlying uncertainty. The two methodologies are connected through the maximum a posteriori (MAP) estimator, where the mean of the posterior distribution corresponds to the solution of a specific Tikhonov regularization problem [20, 10, 18].

To reconstruct the unknown, assume that x has a Gaussian prior $\mathcal{N}(0, \lambda^{-1}\Sigma)$ independent of the noise η , where Σ is a symmetric positive definite matrix, and $\lambda > 0$ is a hyperparameter governing the prior variance level. In some applications, λ may be known. Quite often, however, it is unknown and we assume that is the case here. According to Bayes' rule, the prior density and the likelihood function induced by the forward model together yield the posterior probability density:

$$\pi(x | y) \propto \pi(x) \pi(y | x) \propto \exp\left(-\frac{1}{2}\|Gx - y\|_{\Gamma^{-1}}^2 - \frac{\lambda}{2}\|x\|_{\Sigma^{-1}}^2\right), \quad (1.2)$$

where $\|x\|_B := (x^\top Bx)^{1/2}$ is the B -norm of x for a positive definite matrix B . Since both the likelihood and the prior are Gaussian, the posterior distribution is also Gaussian, $x | y \sim \mathcal{N}(x_\lambda, C_\lambda)$, with its mean and covariance given by:

$$x_\lambda = C_\lambda G^\top \Gamma^{-1} y, \quad C_\lambda = (G^\top \Gamma^{-1} G + \lambda \Sigma^{-1})^{-1}. \quad (1.3)$$

Specifically, the MAP estimator is the posterior mean, which can be obtained by minimizing the negative log-density of the Bayesian posterior:

$$x_\lambda = \operatorname{argmin}_{x \in \mathbb{R}^n} \{\|Gx - y\|_{\Gamma^{-1}}^2 + \lambda \|x\|_{\Sigma^{-1}}^2\}, \quad (1.4)$$

It is evident from (1.4) that the Tikhonov penalty term is explicitly determined by the prior covariance structure, bridging the probabilistic and deterministic perspectives of the inverse problem.

In practical large-scale applications, the parameter dimension n is often extremely high, typically with $n \gg m$, as considered in this paper. Combined with the inherent ill-posedness of the inverse problem, this poses significant challenges for both stability and computational tractability. In particular, a major computational bottleneck is the inversion of the matrix $G^\top \Gamma^{-1} G + \lambda \Sigma^{-1}$ to obtain the posterior mean and covariance. A variety of approaches have been proposed for dimension reduction of Bayesian inverse problems. Early work uses the Karhunen-Loève expansion [21, 31] to represent the prior distribution in terms of its dominant eigenmodes and truncates the expansion to achieve dimension reduction [27, 33]. However, this strategy relies entirely on the spectral properties of the prior and does not account for the influence of the forward operator or the observation data, which

restricts the choice of prior models and may lead to significant truncation errors. An alternative approach [28] constructs reduced subspaces using model-constrained sampling and greedy strategy, but it may fail to capture posterior variability in directions that are not informed by the data. Another approach [12] approximates the posterior covariance via a low-rank approximation of the Hessian of the log-likelihood. All of the above approaches aim at dimension reduction in the entire parameter space, but do not fully exploit the interaction between the unknown parameter and observation data.

In the common setting of inverse problems, the prior distribution encodes smoothness of the unknown parameter, and the observation data are linked to the parameter through a smoothing forward operator. As a consequence, the prior-to-posterior update is often confined to a relatively low-dimensional subspace of the parameter space. Exploiting this structure can enable efficient low-rank approximations of the posterior mean and covariance, leading to substantial computational savings [32, 1, 2]. In [9, 43], the terminology *likelihood-informed subspace* is proposed to identify such low-dimensional subspaces, which contain the directions along which the posterior differs significantly from the prior. All these methods approximate the posterior covariance through low-rank corrections of the prior covariance, which can provide a more balanced treatment of prior and data information. However, these approaches are still formulated in the parameter space and often rely on explicit factorizations of the prior covariance or its inverse, which can be computationally prohibitive in large-scale settings.

In this paper, we shift the focus from the parameter space to the data space. Endowed with the prior-induced geometry, the parameter space is given by the finite-dimensional Hilbert space $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_{\Sigma^{-1}})$, which can be extremely high-dimensional if n is very large. In contrast, we introduce the *data space* as the quotient Hilbert space $(\mathbb{R}^m/\mathcal{N}(M), \langle \cdot, \cdot \rangle_M)$, with its precise meaning and motivation detailed in Section 2, where $M = G\Sigma G^\top$ is symmetric positive semidefinite and may be rank deficient with $\mathcal{N}(M)$ denoting its null space. The dimension of this space is $\text{rank}(M) = \text{rank}(G)$, and the smoothing nature of G typically leads to rapidly decaying singular values and hence a much smaller effective dimension. We show that the data space admits an isometric embedding into the parameter space, and that the prior-to-posterior update occurs entirely within the image of this embedding, which has the same dimension as the data space. This structure reveals that the informative directions of the posterior are entirely characterized by a low-dimensional, data-informed subspace that is independent of the ambient parameter dimension. While the principle that the prior-to-posterior change may concentrate on a low-dimensional subspace is well recognized in the literature, the present work provides a rigorous and explicit mathematical characterization of this phenomenon through the quotient-space geometry of the data space.

The data-space formulation offers a fundamentally different computational paradigm. In the parameter space, approximating the posterior mean and covariance requires dealing with $G^\top \Gamma^{-1} G$ as well as manipulating the prior covariance Σ , which is often large, dense, and ill-conditioned when defined through covariance kernels [47, 39]. Such operations typically demand matrix factorizations or inversions that are computationally prohibitive in high dimensions. In contrast, the data-space formulation allows the problem to be recast in a low-dimensional space that captures dominant data-informed variations, while eliminating directions that are not identifiable from the data. As a consequence, both the computational complexity and the representation of the posterior are governed by the intrinsic low dimension of the data. To obtain efficient approximations of the posterior, we propose a new Krylov subspace method [29] to iteratively capture dominant data-informed subspaces, as detailed in Section 3. This leads naturally to matrix-free algorithms whose computational cost scales with the dimension of the data rather than the ambient parameter dimension. In addition, we show that the likelihood-informed subspace introduced in [43] arises naturally from the data-space formulation and admits a geometric interpretation within this framework, providing a structural understanding of data-informed subspaces and their connection to the proposed Krylov method.

An additional challenge in Bayesian inverse problems is the selection of the hyperparameter λ . In much of the existing literature, λ is assumed to be known. In practice, however, it is typically determined either by repeated posterior evaluations over multiple candidates or by maximizing the marginal likelihood within an empirical Bayesian inference framework [4, 36, 40]. Both approaches require repeated large-scale computations—either solving ill-posed linear systems or evaluating the marginal likelihood in the high-dimensional parameter space. From the perspective of Tikhonov regularization, the hyperparameter λ can be iteratively estimated within the hybrid regularization framework

[6, 42, 25, 7], where the original problem is projected onto a sequence of low-dimensional Krylov subspaces and the parameter is determined from the reduced problems using parameter-choice rules such as the discrepancy principle or generalized cross-validation [22, 8]. While these approaches significantly reduce computational cost, their convergence behavior remains not fully understood, and they may fail to produce reliable estimates of λ in practice [8, 37, 26]. These limitations motivate the development of scalable methods that enable stable and efficient hyperparameter estimation while simultaneously approximating the posterior, without requiring repeated computations in the parameter space.

Within the data-space formulation, the data-informed Krylov subspaces are constructed in the quotient space $(\mathbb{R}^m/\mathcal{N}(M), \langle \cdot, \cdot \rangle_M)$. To apply the Krylov subspace method and handle the degeneracy of M , one may work in the range space $(\mathcal{R}(M), \langle \cdot, \cdot \rangle_M)$ that is isometrically isomorphic to the quotient space, where $\mathcal{R}(M)$ denotes the range of M . However, this approach requires computing the orthogonal projector MM^\dagger , which is computationally expensive for large m . To overcome this difficulty, we develop a Krylov subspace framework directly in the quotient space, without requiring the computation of projections onto the range of M . Specifically, we propose a quotient-space formulation of the Golub–Kahan bidiagonalization process, referred to as Q-GKB, which constructs a sequence of data-informed Krylov subspaces of increasing dimension in a fully matrix-free manner. Building on this construction, we integrate empirical Bayesian inference into the Q-GKB iteration, enabling simultaneous hyperparameter estimation as well as posterior approximation by restricting the prior-to-posterior update to the data-informed Krylov subspaces.

The contributions of this paper can be summarized as follows:

- We establish a data-space geometric framework for Bayesian linear inverse problems based on the quotient space induced by the semidefinite matrix $M = G\Sigma G^\top$. This yields an isometric correspondence between data-space and parameter-space representations and provides a rigorous characterization of the low-dimensional nature of posterior update.
- We develop a quotient-space Krylov method via a novel formulation of the Golub–Kahan bidiagonalization process, which constructs a sequence of data-informed Krylov subspaces iteratively in a fully matrix-free manner.
- We integrate empirical Bayesian inference into the proposed framework, enabling simultaneous hyperparameter estimation and posterior approximation within low-dimensional Krylov subspaces. We propose an efficient algorithm for computing low-rank posterior approximation with theoretical accuracy guarantees.

To assess the proposed theory and the performance of the method, we present several representative numerical examples of increasing scale and complexity. In particular, the large-scale examples highlight the effectiveness of the matrix-free method within the data-space formulation, where direct access to the posterior distribution is computationally prohibitive.

The remainder of this paper is organized as follows. In [Section 2](#), we develop the data-space geometric framework and establish a structural correspondence between the parameter and data spaces. In [Section 3](#), we present the Q-GKB method for constructing data-informed Krylov subspaces and investigate its properties. In [Section 4](#), we propose a Q-GKB-based empirical Bayesian method that enables simultaneous estimation of the hyperparameter λ and efficient low-rank posterior approximation in a matrix-free manner, and analyze its approximation accuracy. Numerical experiments are presented in [Section 5](#), and the conclusion is provided in [Section 6](#).

2 Data-informed structure of the parameter space

We investigate the properties of the Bayesian posterior and the geometry of the parameter space, which motivate the introduction of a quotient-space structure for the data space. In the remainder of the paper, we use $\text{Im}(\cdot)$ to denote the image of a linear operator, and I to denote the identity matrix or operator, with its order clear from the context. The symbol 0 is used to denote either the real number zero or the zero vector of a linear space, with its intended meaning being clear in each instance.

The following result gives equivalent representations of the posterior mean and covariance. We remark that the same expressions appear in the literature, e.g., [20, 45]. For completeness, we provide the full proof.

Proposition 2.1 *The posterior mean and covariance admit the equivalent representations*

$$x_\lambda = \Sigma G^\top (G \Sigma G^\top + \lambda \Gamma)^{-1} y, \quad (2.1)$$

$$C_\lambda = \lambda^{-1} \Sigma - \lambda^{-2} \Sigma G^\top (\Gamma + \lambda^{-1} G \Sigma G^\top)^{-1} G \Sigma. \quad (2.2)$$

Proof. The covariance formula can be derived by applying the Woodbury identity, which is

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1},$$

for any nonsingular $A \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{m \times m}$, and $U \in \mathbb{R}^{n \times m}$, $V \in \mathbb{R}^{m \times n}$. Therefore,

$$\begin{aligned} C_\lambda &= (\lambda \Sigma^{-1} + G^\top \Gamma^{-1} G)^{-1} = \lambda^{-1} \Sigma - \lambda^{-1} \Sigma G^\top (\Gamma + \lambda^{-1} G \Sigma G^\top)^{-1} G \lambda^{-1} \Sigma \\ &= \lambda^{-1} \Sigma - \lambda^{-2} \Sigma G^\top (\Gamma + \lambda^{-1} G \Sigma G^\top)^{-1} G \Sigma, \end{aligned}$$

which proves the covariance formula. Substituting the above representation into $x_\lambda = C_\lambda G^\top \Gamma^{-1} y$, we have

$$\begin{aligned} x_\lambda &= [\lambda^{-1} \Sigma - \lambda^{-2} \Sigma G^\top (\Gamma + \lambda^{-1} G \Sigma G^\top)^{-1} G \Sigma] G^\top \Gamma^{-1} y \\ &= \lambda^{-1} \Sigma G^\top \Gamma^{-1} y - \lambda^{-2} \Sigma G^\top (\Gamma + \lambda^{-1} G \Sigma G^\top)^{-1} G \Sigma G^\top \Gamma^{-1} y. \end{aligned}$$

Let $M = G \Sigma G^\top \in \mathbb{R}^{m \times m}$. It holds that

$$\begin{aligned} x_\lambda &= \Sigma G^\top [\lambda^{-1} \Gamma^{-1} - \lambda^{-2} (\Gamma + \lambda^{-1} M)^{-1} M \Gamma^{-1}] y \\ &= \lambda^{-1} \Sigma G^\top [I - (\lambda \Gamma + M)^{-1} M] \Gamma^{-1} y \\ &= \lambda^{-1} \Sigma G^\top [\lambda (\lambda \Gamma + M)^{-1} \Gamma] \Gamma^{-1} y = \Sigma G^\top (\lambda \Gamma + M)^{-1} y, \end{aligned}$$

where we have used the identity $I - (\lambda \Gamma + M)^{-1} M = \lambda (\lambda \Gamma + M)^{-1} \Gamma$. This proves the formula for the posterior mean. \square

From (2.1) and (2.2), it is natural to introduce the positive semidefinite matrix

$$M = G \Sigma G^\top \in \mathbb{R}^{m \times m}. \quad (2.3)$$

Define the data-space variable $z_\lambda = (M + \lambda \Gamma)^{-1} y \in \mathbb{R}^m$. Then the posterior mean can be written as $x_\lambda = \Sigma G^\top z_\lambda$. Consider the regularization problem

$$\min_{z \in \mathbb{R}^m} \{\|Mz - y\|_{\Gamma^{-1}}^2 + \lambda z^\top Mz\}. \quad (2.4)$$

If M is not positive definite, this problem may admit multiple solutions. However, it can be shown that any solution coincides with z_λ up to an element of $\mathcal{N}(M)$.

Proposition 2.2 *Consider the regularization problem (2.4) with $\lambda > 0$. Then*

- (a) z_λ is a minimizer of (2.4), and the set of all minimizers is $z_\lambda + \mathcal{N}(M)$;
- (b) if $z \in z_\lambda + \mathcal{N}(M)$, it holds that $\Sigma G^\top z = x_\lambda$. Hence, all minimizers of (2.4) yield the same posterior mean.

Proof. (a) Define $J(z) = \|Mz - y\|_{\Gamma^{-1}}^2 + \lambda z^\top Mz$. Then J is a convex quadratic functional on \mathbb{R}^m . Hence its minimizers are exactly the solutions of

$$\nabla J(z) = 2M\Gamma^{-1}(Mz - y) + 2\lambda Mz = 0.$$

Since z_λ satisfies $(M + \lambda \Gamma)z_\lambda = y$, or equivalently $Mz_\lambda - y = -\lambda \Gamma z_\lambda$, it can be verified that $\nabla J(z_\lambda) = 0$. Next let z be vector such that $\nabla J(z) = 0$. Using $y = (M + \lambda \Gamma)z_\lambda$, we have

$$Mz - y = Mz - (M + \lambda \Gamma)z_\lambda = M(z - z_\lambda) - \lambda \Gamma z_\lambda.$$

Substituting it into $\nabla J(z)$ yields

$$\frac{1}{2}\nabla J(z) = M\Gamma^{-1}(M(z - z_\lambda) - \lambda\Gamma z_\lambda) + \lambda Mz = M\Gamma^{-1}M(z - z_\lambda) + \lambda M(z - z_\lambda).$$

Taking the inner product with $z - z_\lambda$, we obtain

$$\nabla J(z) = 0 \quad \Rightarrow \quad (z - z_\lambda)^\top M\Gamma^{-1}M(z - z_\lambda) + \lambda(z - z_\lambda)^\top M(z - z_\lambda) = 0.$$

Since Γ^{-1} is positive definite and M is positive semidefinite, both terms in the left-hand-side are nonnegative and must therefore vanish. In particular, $(z - z_\lambda)^\top M(z - z_\lambda) = 0$, which implies that $z - z_\lambda \in \mathcal{N}(M)$. Conversely, for any $p \in \mathcal{N}(M)$, we have $\frac{1}{2}\nabla J(z_\lambda + p) = M\Gamma^{-1}Mp + \lambda Mp = 0$. This proves (a).

(b) For any $v \in \mathcal{N}(M)$ we have $0 = v^\top Mv = (G^\top v)^\top \Sigma(G^\top v)$. This implies $G^\top v = 0$, and thereby $\mathcal{N}(M) \subseteq \mathcal{N}(G^\top)$. Conversely, it is obvious that $G^\top v = 0$ implies $G\Sigma G^\top v = 0$, and hence $v \in \mathcal{N}(M)$. Therefore, it holds $\mathcal{N}(M) = \mathcal{N}(G^\top)$. Now for any $p \in \mathcal{N}(M)$, we have

$$\Sigma G^\top(z_\lambda + p) = \Sigma G^\top z_\lambda + \Sigma G^\top p = \Sigma G^\top z_\lambda = x_\lambda.$$

This completes the proof. \square

By [Proposition 2.2](#), the posterior mean lies in the range space $\mathcal{R}(\Sigma G^\top)$. From [\(2.2\)](#), define

$$\Delta = \lambda^{-1}\Sigma - C_\lambda = \lambda^{-2}\Sigma G^\top(\Gamma + \lambda^{-1}M)^{-1}G\Sigma$$

as the prior-to-posterior covariance update. Then $\mathcal{R}(\Delta) \subset \mathcal{R}(\Sigma G^\top)$ and $\mathcal{R}(\Sigma G^\top)^\perp \subset \mathcal{N}(\Delta)$. Hence, the covariance update acts only on $\mathcal{R}(\Sigma G^\top)$ and vanishes on its orthogonal complement, reflecting the principle that the data update the prior only along a subspace of dimension $\text{rank}(G)$. In the parameter space, this phenomenon is described by the terminology *likelihood-informed subspace* (LIS) introduced in [\[9, 43\]](#). Let $H = G^\top \Gamma^{-1}G$, and the generalized eigenpairs of (H, Σ^{-1}) are $\{(\mu_i, w_i)\}_{i=1}^l$ satisfying $\mu_1 \geq \mu_2 \geq \dots \geq \mu_l > 0$, $\mu_i = 0$ for $i > l$, and $\{w_i\}_{i=1}^l$ are mutually Σ^{-1} -orthonormal. It is shown in [\[43\]](#) that the directions along which the posterior changes from the prior are entirely in $\text{span}\{w_i\}_{i=1}^l$. For any $1 \leq r \leq l$, the subspace $\text{span}\{w_i\}_{i=1}^r$ is referred to as the r -dimensional LIS, which can be used to construct low-rank approximation of the posterior.

Motivated by the above structure, it is natural to consider the action of the mapping ΣG^\top on \mathbb{R}^m and identify vectors that differ by elements of $\mathcal{N}(M)$. Accordingly, we define the relation

$$u \sim v \quad \Longleftrightarrow \quad u - v \in \mathcal{N}(M),$$

which is an equivalence relation on \mathbb{R}^m . Denote by $[u]$ the equivalence class of u , which is an element of the quotient linear space $\mathbb{R}^m/\mathcal{N}(M)$ with dimension

$$\dim(\mathbb{R}^m/\mathcal{N}(M)) = m - \dim(\mathcal{N}(M)) = \text{rank}(M) = \text{rank}(G).$$

We refer to the finite-dimensional Hilbert space

$$(\mathbb{R}^m/\mathcal{N}(M), \langle \cdot, \cdot \rangle_M), \quad \langle [u], [v] \rangle_M := u^\top Mv \tag{2.5}$$

as the *data space* of the underlying Bayesian linear inverse problem. The bilinear form $\langle \cdot, \cdot \rangle_M$ is well defined on $\mathbb{R}^m/\mathcal{N}(M)$, since if $u - u' \in \mathcal{N}(M)$ and $v - v' \in \mathcal{N}(M)$, then $u^\top Mv - u'^\top Mv' = (u - u')^\top Mv + u'^\top M(v - v') = 0$. Moreover, if $\langle [u], [u] \rangle_M = 0$, then $u \in \mathcal{N}(M)$, implying that $[u] = 0$. Therefore, $\langle \cdot, \cdot \rangle_M$ defines an inner product on $\mathbb{R}^m/\mathcal{N}(M)$. Notice that $\Gamma + \lambda^{-1}M \in \mathbb{R}^{m \times m}$ is the covariance of the marginal distribution of y . This suggests considering the generalized eigenvalue decomposition of (M, Γ) , which characterizes the associated spectral structure of the data space. Since Γ is positive definite, there exists a set of M -orthonormal generalized eigenvectors corresponding to the nonzero generalized eigenvalues of (M, Γ) . We refer to such eigenpairs as the nonzero M -orthonormal generalized eigenpairs.

The following theorem establishes a precise geometric and spectral correspondence between the data space and parameter space.

Theorem 2.1 (Correspondence between data and parameter spaces) *Define the linear operator*

$$\begin{aligned} \iota : (\mathbb{R}^m / \mathcal{N}(M), \langle \cdot, \cdot \rangle_M) &\longrightarrow (\mathbb{R}^n, \langle \cdot, \cdot \rangle_{\Sigma^{-1}}) \\ [u] &\longmapsto \Sigma G^\top u. \end{aligned} \quad (2.6)$$

Then the following statements hold.

(a) *(Isometric embedding)* The map ι is an isometric embedding, i.e.,

$$\langle \iota([u]), \iota([v]) \rangle_{\Sigma^{-1}} = \langle [u], [v] \rangle_M,$$

and ι is injective.

(b) *(Parameter space decomposition)* It holds that

$$(\mathbb{R}^n, \langle \cdot, \cdot \rangle_{\Sigma^{-1}}) = \text{Im}(\iota) \oplus \mathcal{N}(G),$$

where the two subspaces are Σ^{-1} -orthogonal, and $x_\lambda \in \text{Im}(\iota) = \mathcal{R}(\Sigma G^\top)$.

(c) *(Spectral correspondence)* Let $\{(\hat{\mu}_i, \hat{w}_i)\}_{i=1}^{\hat{l}}$ be all nonzero M -orthonormal generalized eigenpairs of (M, Γ) . Then $l = \hat{l} = \text{rank}(G)$, and $\{\hat{\mu}_i, \iota([\hat{w}_i])\}_{i=1}^{\hat{l}}$ are all nonzero Σ^{-1} -orthonormal eigenpairs of (H, Σ^{-1}) .

Proof. (a) If $[u] = [u']$, then $u - u' \in \mathcal{N}(M)$. From part (a) of [Proposition 2.2](#), $\mathcal{N}(M) = \mathcal{N}(G^\top)$, so $G^\top(u - u') = 0$ and thus $\Sigma G^\top u = \Sigma G^\top u'$. Hence ι is well defined. For the injectivity, suppose $\iota([u]) = \Sigma G^\top u = 0$. Then $u^\top M u = \|\Sigma G^\top u\|_{\Sigma^{-1}}^2 = 0$, so $u \in \mathcal{N}(M)$, i.e., $[u] = 0$. For any $[u], [v] \in \mathbb{R}^m / \mathcal{N}(M)$, it holds

$$\langle \iota([u]), \iota([v]) \rangle_{\Sigma^{-1}} = (\Sigma G^\top u)^\top \Sigma^{-1} (\Sigma G^\top v) = u^\top G \Sigma G^\top v = u^\top M v = \langle [u], [v] \rangle_M.$$

Thus ι is an isometric embedding.

(b) For any $[u] \in \mathbb{R}^m / \mathcal{N}(M)$ and $v \in \mathcal{N}(G)$, it holds

$$\langle \iota([u]), v \rangle_{\Sigma^{-1}} = (\Sigma G^\top u)^\top \Sigma^{-1} v = u^\top G v = 0.$$

Hence $\text{Im}(\iota)$ and $\mathcal{N}(G)$ are Σ^{-1} orthogonal. Since ι is injective, we have

$$\dim(\text{Im}(\iota)) = \dim(\mathbb{R}^m / \mathcal{N}(M)) = \text{rank}(M) = \text{rank}(G) = \dim(\mathcal{R}(\Sigma G^\top)).$$

Therefore, it holds that $\text{Im}(\iota) = \mathcal{R}(\Sigma G^\top)$ and $\dim(\text{Im}(\iota)) + \dim(\mathcal{N}(G)) = n$. This also proves $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_{\Sigma^{-1}}) = \text{Im}(\iota) \oplus \mathcal{N}(G)$. It is obvious that $x_\lambda \in \text{Im}(\iota)$ since $x_\lambda = \Sigma G^\top z_\lambda$.

(c) Since Σ and Γ are symmetric positive definite, we have

$$\text{rank}(M) = \text{rank}(G \Sigma G^\top) = \text{rank}(G), \quad \text{rank}(H) = \text{rank}(G^\top \Gamma^{-1} G) = \text{rank}(G).$$

Hence the pencils (M, Γ) and (H, Σ^{-1}) have the same number of nonzero generalized eigenvalues, equals to $\text{rank}(G)$. Now let $(\hat{\mu}, \hat{w})$ be a nonzero generalized eigenpair of (M, Γ) , i.e., $M \hat{w} = \hat{\mu} \Gamma \hat{w}$ with $\hat{\mu} > 0$. Since $\hat{\mu} \neq 0$, we must have $\hat{w} \notin \mathcal{N}(M)$, and thus $[\hat{w}] \neq 0$ in $\mathbb{R}^m / \mathcal{N}(M)$. Let $w = \iota([\hat{w}]) = \Sigma G^\top \hat{w}$. Because ι is injective, we have $w \neq 0$. Using $M = G \Sigma G^\top$, we have

$$H w = G^\top \Gamma^{-1} G (\Sigma G^\top \hat{w}) = G^\top \Gamma^{-1} M \hat{w} = \hat{\mu} G^\top \hat{w}.$$

Combining with $\Sigma^{-1} w = G^\top \hat{w}$, we obtain $H w = \hat{\mu} \Sigma^{-1} w$. This shows that $(\hat{\mu}, w)$ is a generalized eigenpair of (H, Σ^{-1}) . Conversely, let (μ, w) be a generalized eigenpair of (H, Σ^{-1}) with $\mu > 0$, i.e., $H w = \mu \Sigma^{-1} w$. Let $u = \mu^{-1} \Gamma^{-1} G w \in \mathbb{R}^m$. Then

$$\iota([u]) = \Sigma G^\top u = \mu^{-1} \Sigma G^\top \Gamma^{-1} G w = \mu^{-1} \Sigma H w = w,$$

and

$$Mu = G\Sigma G^\top u = Gw = \mu\Gamma(\mu^{-1}\Gamma^{-1}Gw) = \mu\Gamma u,$$

so (μ, u) is a nonzero generalized eigenpair of (M, Γ) satisfying $w = \iota([u])$. We have therefore established a one-to-one correspondence between the nonzero generalized eigenpairs of (M, Γ) and (H, Σ^{-1}) . Moreover, by part (a), ι maps M -orthonormal vectors to Σ^{-1} -orthonormal vectors. This completes the proof. \square

Remark 2.1 *Theorem 2.1 shows that the data-space formulation provides a complete geometric representation of the data-informed structure of the posterior. In particular, the image $\text{Im}(\iota)$ identifies the data-informed subspace in the parameter space, along which the posterior differs from the prior, while its Σ^{-1} -orthogonal complement $\mathcal{N}(G)$ corresponds to directions that are not informed by the data. Furthermore, the spectral correspondence shows that the eigenstructure in the data space induces the corresponding LIS in the parameter space, thereby providing a geometric characterization of the LIS in terms of the quotient data space.*

In the parameter space, a common approach to approximate the posterior covariance C_λ is based on LIS. We use the Loewner order “ \preceq ” for symmetric matrices, defined by $A \preceq B$ if $B - A$ is symmetric positive definite (SPD) [19]. It is shown in [43] that the optimal approximation to C_λ within the set

$$\mathcal{M}_r = \{\lambda^{-1}\Sigma - KK^\top \succ 0 : \text{rank}(K) \leq r\}$$

with respect to the Förstner distance is achieved by choosing $KK^\top = \sum_{i=1}^r \frac{\mu_i}{\lambda(\lambda + \mu_i)} w_i w_i^\top$, where $\{(\mu_i, w_i)\}_{i=1}^r$ are the leading r generalized eigenpairs of (H, Σ^{-1}) . The resulting approximation can be written as

$$C_\lambda^{(r)} = (H_r + \lambda\Sigma^{-1})^{-1}, \quad (2.7)$$

where

$$H_r = \Sigma^{-1} \left(\sum_{i=1}^r \mu_i w_i w_i^\top \right) \Sigma^{-1}$$

is a rank- r approximation of H in the r -dimensional LIS.

By [Theorem 2.1](#), the data-space geometry induced by M is often intrinsically low-dimensional due to the smoothing nature of the forward operator G . This provides a geometric justification for effective approximation of the posterior structure by identifying dominant subspaces of the data space. In particular, since the nonzero spectral structure of (H, Σ^{-1}) is completely determined by that of (M, Γ) , it is natural to approximate H via the dominant generalized eigenpairs of (M, Γ) . Accordingly, we equivalently can write

$$H_r = \Sigma^{-1} \left(\Sigma G^\top \left(\sum_{i=1}^r \mu_i \hat{w}_i \hat{w}_i^\top \right) G \Sigma \right) \Sigma^{-1} = G^\top \left(\sum_{i=1}^r \mu_i \hat{w}_i \hat{w}_i^\top \right) G, \quad (2.8)$$

where $\{(\mu_i, \hat{w}_i)\}_{i=1}^r$ are the leading r generalized eigenpairs of (M, Γ) . This construction provides a data-space geometric interpretation of LIS-based posterior approximation.

3 Quotient-space Krylov method in the data space

For large-scale problems, a direct eigenvalue decomposition of (M, Γ) is computationally expensive, and forming M explicitly is typically infeasible. Nevertheless, matrix-vector products can be computed efficiently as $Mv = G(\Sigma(G^\top v))$. Moreover, the hyperparameter λ is often unknown in advance and its estimation in the full parameter space is costly. In view of the connection between Bayesian inference and Tikhonov regularization, it is natural to work with the regularization problem (2.4) in the data space, which enables the estimation of λ in an iterative framework. These considerations make the Krylov subspace method particularly attractive, as they rely only on matrix-vector products and naturally capture dominant low-dimensional data-informed subspaces.

3.1 Quotient-space Golub–Kahan bidiagonalization

In the data space, the regularization problem (2.4) can be equivalently formulated as

$$\min_{[z] \in \mathbb{R}^m / \mathcal{N}(M)} \{ \|Mz - y\|_{\Gamma^{-1}}^2 + \lambda \|z\|_M^2 \}. \quad (3.1)$$

To design practical Krylov subspace method for this problem, one possible approach is to represent equivalence classes by elements in the range space $\mathcal{R}(M)$ and restrict the iteration accordingly. However, this requires explicit projections onto $\mathcal{R}(M)$, typically involving the projector MM^\dagger , which is computationally expensive. Instead, we work directly on the quotient space $\mathbb{R}^m / \mathcal{N}(M)$, which removes the degeneracy without requiring explicit projections.

Taking into account the geometry of the data-space and the observation model, we consider the linear operator

$$\begin{aligned} \mathcal{T} : (\mathbb{R}^m / \mathcal{N}(M), \langle \cdot, \cdot \rangle_M) &\longrightarrow (\mathbb{R}^m, \langle \cdot, \cdot \rangle_{\Gamma^{-1}}) \\ [u] &\longmapsto Mu, \end{aligned} \quad (3.2)$$

which is well defined, because $[u] = [v] \Leftrightarrow u - v \in \mathcal{N}(M)$ implies $Mu = Mv$. Its adjoint

$$\mathcal{T}^* : (\mathbb{R}^m, \langle \cdot, \cdot \rangle_{\Gamma^{-1}}) \longrightarrow (\mathbb{R}^m / \mathcal{N}(M), \langle \cdot, \cdot \rangle_M)$$

is characterized by the relation

$$\langle \mathcal{T}[u], p \rangle_{\Gamma^{-1}} = \langle [u], \mathcal{T}^*p \rangle_M, \quad \forall [u] \in \mathbb{R}^m / \mathcal{N}(M), p \in \mathbb{R}^m.$$

The following result provides an explicit expression for the action of \mathcal{T}^* on a vector.

Lemma 3.1 *For every $p \in \mathbb{R}^m$, the adjoint of \mathcal{T} is given by*

$$\mathcal{T}^*p = [\Gamma^{-1}p]. \quad (3.3)$$

Proof. By the definition of \mathcal{T}^* , for any $[u] \in \mathbb{R}^m / \mathcal{N}(M)$ and $p \in \mathbb{R}^m$,

$$\langle [u], \mathcal{T}^*p \rangle_M = \langle \mathcal{T}[u], p \rangle_{\Gamma^{-1}} = (Mu)^\top \Gamma^{-1}p = u^\top M\Gamma^{-1}p.$$

By definition of the inner product on $\mathbb{R}^m / \mathcal{N}(M)$, we have

$$u^\top M\Gamma^{-1}p = \langle [u], [\Gamma^{-1}p] \rangle_M = \langle [u], \mathcal{T}^*p \rangle_M,$$

which shows that $\mathcal{T}^*p = [\Gamma^{-1}p]$. In other words, $\Gamma^{-1}p$ is a representative element of the equivalence class \mathcal{T}^*p . \square

We now apply the Golub–Kahan bidiagonalization (GKB) to the operator-vector pair $\{\mathcal{T}, y\}$; see [5] for the GKB process for general linear compact operators. This formulation allows the iteration to be initialized directly from the vector y and carried out entirely within the data-space framework without computing orthogonal projections. The abstract GKB recursions read

$$\begin{cases} \beta_1 u_1 = y, \\ \alpha_i \bar{w}_i = \mathcal{T}^*u_i - \beta_i \bar{w}_{i-1}, \\ \beta_{i+1} u_{i+1} = \mathcal{T} \bar{w}_i - \alpha_i u_i, \quad i = 1, 2, \dots, \end{cases} \quad (3.4)$$

where $\bar{w}_0 := 0$, and α_i and β_i are positive scalars such that $\{u_i\}_i$ are Γ^{-1} -orthonormal and $\{\bar{w}_i\}_i$ are orthonormal in the data space $(\mathbb{R}^m / \mathcal{N}(M), \langle \cdot, \cdot \rangle_M)$.

To obtain a practical matrix-vector implementation, we represent each quotient vector \bar{w}_i by a representative element $v_i \in \mathbb{R}^m$. Then $\|\bar{w}_i\|_M = \sqrt{v_i^\top M v_i}$. By Lemma 3.1, we have $\mathcal{T}^*u_i = [\Gamma^{-1}u_i]$. Using these relations, the abstract iterations reduce to the practical recursions:

$$\begin{cases} \beta_1 u_1 = y, \\ \alpha_i v_i = \Gamma^{-1}u_i - \beta_i v_{i-1}, \\ \beta_{i+1} u_{i+1} = Mv_i - \alpha_i u_i, \quad i = 1, 2, \dots, \end{cases} \quad (3.5)$$

where $v_0 := 0$, and the vectors $\{v_i\}_i$ are M -orthonormal, that is, $v_i^\top M v_j = \delta_{ij}$ with δ_{ij} being the Kronecker symbol. The scalars α_i and β_{i+1} can be computed by evaluating the M - and Γ^{-1} -norms of $\Gamma^{-1}u_i - \beta_i v_{i-1}$ and $Mv_i - \alpha_i u_i$, respectively. This process can be carried out iteratively. We refer to this process as the *quotient-space Golub–Kahan bidiagonalization* (Q-GKB). The pseudocode of Q-GKB is given in [Algorithm 1](#). Importantly, the Q-GKB iteration requires only matrix-vector products with M and Γ^{-1} and does not require computing orthogonal projections onto $\mathcal{R}(M)$. While computations with Γ^{-1} are unavoidable, in many practical settings where η is an uncorrelated Gaussian noise, Γ is diagonal, so Γ^{-1} can be applied efficiently.

Algorithm 1 Quotient-space Golub–Kahan bidiagonalization (Q-GKB)

Input: Semidefinite $M \in \mathbb{R}^{m \times m}$, positive definite $\Gamma \in \mathbb{R}^{m \times m}$, nonzero $y \in \mathbb{R}^m$

```

1: Initialize:  $\beta_1 = (y\Gamma^{-1}y)^{1/2}$ ,  $u_1 = y/\beta_1$ 
2: Compute  $s = \Gamma^{-1}u_1$ ,  $\alpha_1 = (s^\top Ms)^{1/2}$ ,
3: if  $\alpha_1 = 0$  then terminate
4: else
5:    $v_1 = s/\alpha_1$ 
6: end if
7: for  $i = 1, 2, \dots, k$  do
8:    $r = Mv_i - \alpha_i u_i$ ,  $\beta_{i+1} = (r\Gamma^{-1}r)^{1/2}$ 
9:   if  $\beta_{i+1} = 0$  then terminate
10:  else
11:     $u_{i+1} = r/\beta_{i+1}$ 
12:  end if
13:   $s = \Gamma^{-1}u_{i+1} - \beta_{i+1}v_i$ ,  $\alpha_{i+1} = (s^\top Ms)^{1/2}$ 
14:  if  $\alpha_{i+1} = 0$  then terminate
15:  else
16:     $v_{i+1} = s/\alpha_{i+1}$ 
17:  end if
18: end for

```

Output: $\{\alpha_i, \beta_i\}_{i=1}^{k+1}$, $\{u_i, v_i\}_{i=1}^{k+1}$

The Q-GKB iteration terminates in at most $\text{rank}(G)$ steps, since $\{v_i\}_i$ are M -orthonormal and therefore cannot have more than $\text{rank}(G)$ linearly independent vectors. Define the breakdown step as

$$k_b = \min_{k \geq 1} \{k : \alpha_{k+1}\beta_{k+1} = 0\}. \quad (3.6)$$

Before the iteration breakdown, i.e., $1 \leq k \leq k_b$, let

$$U_{k+1} = (u_1, \dots, u_{k+1}) \in \mathbb{R}^{m \times (k+1)}, \quad V_k = (v_1, \dots, v_k) \in \mathbb{R}^{m \times k}, \quad (3.7)$$

and let

$$B_k = \begin{pmatrix} \alpha_1 & & & & & \\ \beta_2 & \alpha_2 & & & & \\ & \beta_3 & \ddots & & & \\ & & & \ddots & & \\ & & & & \alpha_k & \\ & & & & & \beta_{k+1} \end{pmatrix} \in \mathbb{R}^{(k+1) \times k}. \quad (3.8)$$

If $\beta_{k_b+1} = 0$ at step $k = k_b$, then we let $u_{k_b+1} = 0$ in U_{k_b+1} . Now each computed vector v_i is a representative element of \bar{w}_i that is mutually orthonormal in $(\mathbb{R}^m / \mathcal{N}(M), \langle \cdot, \cdot \rangle_M)$. By the property of GKB of linear compact operator, it holds that $V_k^\top M V_k = I$. Moreover, if $k = k_b$ and $\beta_{k_b+1} = 0$, then $U_{k_b}^\top \Gamma^{-1} U_{k_b} = I$, otherwise it holds that $U_{k+1}^\top \Gamma^{-1} U_{k+1} = I$ for any $1 \leq k \leq k_b$.

The following result gives the basic properties of Q-GKB in the data space. We use e_1 and e_{k+1} to denote the first and $(k+1)$ th column of the identity matrix of order $k+1$.

Proposition 3.1 *Assume that Q-GKB runs for k steps with $k \leq k_b$. Then the following properties hold.*

(a) *The matrix-form relations hold:*

$$\begin{cases} \beta_1 U_{k+1} e_1 = y, \\ MV_k = U_{k+1} B_k, \\ \Gamma^{-1} U_{k+1} = V_k B_k^\top + \alpha_{k+1} v_{k+1} e_{k+1}^\top. \end{cases} \quad (3.9)$$

(b) *The vectors $\{u_i\}_{i=1}^k$ form an Γ^{-1} -orthonormal basis of the Krylov subspace*

$$\mathcal{K}_k(M\Gamma^{-1}, y) = \text{span}\{(M\Gamma^{-1})^i y\}_{i=0}^{k-1},$$

and vectors $\{v_i\}_{i=1}^k$ form an M -orthonormal basis of the Krylov subspace

$$\mathcal{K}_k(\Gamma^{-1}M, \Gamma^{-1}y) = \text{span}\{(\Gamma^{-1}M)^i \Gamma^{-1}y\}_{i=0}^{k-1}.$$

Proof. Part (a) is directly obtained by stacking the recurrences in (3.5).

For part (b), from the theory of GKB, it produces two Krylov subspaces,

$$\mathcal{K}_k(\mathcal{T}^*\mathcal{T}, \mathcal{T}^*y) = \text{span}\{(\mathcal{T}^*\mathcal{T})^i \mathcal{T}^*y\}_{i=0}^{k-1} = \text{span}\{[v_i]\}_{i=1}^k$$

and

$$\mathcal{K}_k(\mathcal{T}\mathcal{T}^*, y) = \text{span}\{(\mathcal{T}\mathcal{T}^*)^i y\}_{i=0}^{k-1} = \text{span}\{u_i\}_{i=1}^k,$$

such that $\{u_i\}_{i=1}^k$ and $\{[v_i]\}_{i=1}^k$ are Γ - and M -orthonormal, respectively. From the properties of \mathcal{T} and \mathcal{T}^* , we have

$$\mathcal{T}^*\mathcal{T}[u] = [\Gamma^{-1}Mu], \quad \mathcal{T}\mathcal{T}^*v = M\Gamma^{-1}v$$

for any $[u] \in \mathbb{R}^m/\mathcal{N}(M)$ and $v \in \mathbb{R}^m$. Therefore, we have

$$\mathcal{K}_k(\mathcal{T}^*\mathcal{T}, \mathcal{T}^*y) = \text{span}\{[(\Gamma^{-1}M)^i \Gamma^{-1}y]\}_{i=0}^{k-1}, \quad \mathcal{K}_k(\mathcal{T}\mathcal{T}^*, y) = \text{span}\{(M\Gamma^{-1})^i y\}_{i=0}^{k-1}.$$

This proves the desired property for $\{u_i\}_{i=1}^k$. Moreover, from the recursions (3.5), we have

$$\alpha_1 \beta_1 v_1 = \Gamma^{-1}y, \quad \alpha_{i+1} \beta_{i+1} v_{i+1} = \Gamma^{-1}Mv_i - (\alpha_i^2 + \beta_{i+1}^2)v_i - \alpha_i \beta_i v_{i-1}.$$

Combining these relations with mathematical induction, it can be verified that $v_i \in \mathcal{K}_k(\Gamma^{-1}M, \Gamma^{-1}y)$ for $1 \leq i \leq k$. Since $\{v_i\}_{i=1}^k$ are M -orthonormal, it holds that

$$\dim(\text{span}\{v_i\}_{i=1}^k) = k = \dim(\mathcal{K}_k(\Gamma^{-1}M, \Gamma^{-1}y)).$$

This implies that $\{v_i\}_{i=1}^k$ forms an M -orthonormal basis of $\mathcal{K}_k(\Gamma^{-1}M, \Gamma^{-1}y)$. \square

Proposition 3.1 shows that Q-GKB constructs mutually orthonormal bases of paired Krylov subspaces in the data space, analogous to the classical GKB process. On the other hand, it is not merely a formal modification of standard GKB, but is intrinsically adapted to the quotient-space geometry, enabling the iteration to proceed without explicitly resolving the null space of M .

3.2 Data-informed Krylov subspace

At the k th step of Q-GKB, define $\mathcal{S}_k = \text{span}\{[v_i]\}_{i=1}^k$, which is a k -dimensional subspace of the data space. We approximate the solution of (3.1) by restricting it to \mathcal{S}_k , i.e.,

$$\min_{[z] \in \mathcal{S}_k} \{\|Mz - y\|_{\Gamma^{-1}}^2 + \lambda \| [z] \|_M^2\}, \quad (3.10)$$

The following result shows that the Q-GKB iteration captures the exact posterior mean after finitely many steps.

Theorem 3.1 Let k_b be the breakdown step of Q-GKB defined in (3.6). Then for every $\lambda > 0$, the unique minimizer $[z_\lambda]$ of (3.1) belongs to $\mathcal{S}_{k_b} = \text{span}\{[v_i]\}_{i=1}^{k_b}$.

Proof. Since the objective function of (3.1) can be equivalently written as $\|\mathcal{T}[z] - y\|_{\Gamma^{-1}}^2 + \lambda\|z\|_M^2$, its unique minimizer is the equivalence class

$$[z_\lambda] = (\mathcal{T}^*\mathcal{T} + \lambda I)^{-1}\mathcal{T}^*y.$$

Since \mathcal{S}_{k_b} is the maximal Krylov subspace generated by \mathcal{T}^*y under $\mathcal{T}^*\mathcal{T}$, which means that $\mathcal{S}_{k_b} = \mathcal{K}_\infty(\mathcal{T}^*\mathcal{T}, \mathcal{T}^*y)$, it is invariant under $\mathcal{T}^*\mathcal{T}$ and contains all vectors of the form $p(\mathcal{T}^*\mathcal{T})\mathcal{T}^*y$, where p is a polynomial. Note that $\mathcal{T}_\lambda := \mathcal{T}^*\mathcal{T} + \lambda I$ is a positive self-adjoint operator on $\mathbb{R}^m/\mathcal{N}(M)$. Let $q = \dim(\mathbb{R}^m/\mathcal{N}(M))$ and let $A \in \mathbb{R}^{q \times q}$ be a matrix representation of \mathcal{T}_λ with respect to a basis of $\mathbb{R}^m/\mathcal{N}(M)$. By the Cayley–Hamilton theorem, it holds that $p_A(\mathcal{T}_\lambda) = 0$, where $p_A(\mu) = \det(\mu I - A) = \sum_{i=0}^q c_i \mu^i$ is the characteristic polynomial of A , satisfying that $c_q = 1$ and $c_0 = (-1)^q \det(A) \neq 0$ since A is SPD. It follows that

$$\mathcal{T}_\lambda \sum_{i=1}^l c_i (\mathcal{T}^*\mathcal{T} + \lambda I)^{i-1} = -c_0 I,$$

which implies that \mathcal{T}_λ^{-1} can be represented by a polynomial of $\mathcal{T}^*\mathcal{T}$. Hence there exists a polynomial p_λ such that

$$(\mathcal{T}^*\mathcal{T} + \lambda I)^{-1}\mathcal{T}^*y = p_\lambda(\mathcal{T}^*\mathcal{T})\mathcal{T}^*y \in \mathcal{S}_{k_b}.$$

This proves the claim. \square

Theorem 3.1 implies that the posterior mean x_λ can be approximated within the subspace $\iota(\mathcal{S}_k)$ as k increases. We refer to $\iota(\mathcal{S}_k)$ as the k -dimensional *data-informed Krylov subspace*. Suppose there are t distinct nonzero generalized eigenvalues of (M, Γ) with corresponding M -orthogonal eigenspaces $\mathcal{G}_1, \dots, \mathcal{G}_t$. For any closed subspace \mathcal{G} of \mathbb{R}^m , denote by $P_{\mathcal{G}}$ the M -orthogonal projector onto \mathcal{G} , i.e., $P_{\mathcal{G}}$ satisfies $P_{\mathcal{G}}^2 = P_{\mathcal{G}}$ and $MP_{\mathcal{G}}^\top = P_{\mathcal{G}}M$. The following result shows that, as the iteration proceeds, the subspace $\iota(\mathcal{S}_k)$ progressively captures the informative directions of the dominant data-informed subspace.

Theorem 3.2 Assume that Q-GKB runs for k steps with $k \leq k_b$, and the nonzero 2-orthonormal eigenpairs of $B_k^\top B_k$ are $\left\{(\theta_i^{(k)}, s_i^{(k)})\right\}_{i=1}^k$. Then the following statements hold.

- (a) The pair $(\theta_i^{(k)}, p_i^{(k)}) := (\theta_i^{(k)}, \iota([V_k s_i^{(k)}]))$ gradually approximates a nonzero generalized eigenpair of (H, Σ^{-1}) as k increases, and the approximation becomes exact at $k = k_b$.
- (b) At the breakdown step, $\{p_i^{(k_b)}\}_{i=1}^{k_b}$ are Σ^{-1} -orthonormal generalized eigenvectors of (H, Σ^{-1}) . Moreover, each $p_i^{(k_b)}$ is the image under ι of a generalized eigenvector in one of those \mathcal{G}_j for which $P_{\mathcal{G}_j}\Gamma^{-1}y \neq 0$.

Proof. (a) Using the matrix form relations of Q-GKB, we have

$$\Gamma^{-1}MV_k = \Gamma^{-1}U_{k+1}B_k = (V_k B_k^\top + \alpha_{k+1}v_{k+1}e_{k+1}^\top)B_k.$$

A simple calculation leads to

$$\begin{aligned} MV_k s_i^{(k)} - \theta_i^{(k)} \Gamma V_k s_i^{(k)} &= \Gamma(V_k B_k^\top + \alpha_{k+1}v_{k+1}e_{k+1}^\top)B_k s_i^{(k)} - \theta_i^{(k)} \Gamma V_k s_i^{(k)} \\ &= \Gamma V_k (B_k^\top B_k s_i^{(k)} - \theta_i^{(k)} s_i^{(k)}) + \alpha_{k+1} \beta_{k+1} \Gamma v_{k+1} e_k^\top s_i^{(k)} \\ &= \alpha_{k+1} \beta_{k+1} \Gamma v_{k+1} e_k^\top s_i^{(k)}. \end{aligned}$$

Therefore, the pair $(\theta_i^{(k)}, V_k s_i^{(k)})$ gradually approximates a nonzero generalized eigenpair of (M, Γ) as k increases, and at the breakdown step, $(\theta_i^{(k_b)}, V_k s_i^{(k_b)})$ is an exact eigenpair since $\alpha_{k_b+1} \beta_{k_b+1} = 0$.

By [Theorem 2.1](#), $\left(\theta_i^{(k)}, \iota\left([V_k s_i^{(k)}]\right)\right)$ gradually approximates a nonzero eigenpair of (H, Σ^{-1}) , which becomes an exact one at the breakdown step.

(b) Without loss of generality, we suppose that only the first q elements in $\{P_{\mathcal{G}_1}\Gamma^{-1}y, \dots, P_{\mathcal{G}_t}\Gamma^{-1}y\}$ are nonzero. Noticing that $\{V_{k_b} s_i^{(k_b)}\}_{i=1}^{k_b}$ are mutually M -orthonormal, by [Theorem 2.1](#) (a) and (c), we only need to prove that the k_b vectors $\hat{p}_i^{(k_b)} := V_{k_b} s_i^{(k)}$ belong separately to the M -orthogonal subspaces $\mathcal{G}_1, \dots, \mathcal{G}_q$. Since $\theta_i^{(k_b)} > 0$ have different values and \mathcal{G}_i are mutually M -orthogonal, these $\hat{p}_i^{(k_b)}$ must belong to different subspaces among $\{\mathcal{G}_i\}_{i=1}^t$. Therefore, we only need to prove $P_{\mathcal{G}}\hat{p}_i^{(k_b)} = \hat{p}_i^{(k_b)}$ for each $1 \leq i \leq q$, where $\mathcal{G} = \mathcal{G}_1 \oplus \dots \oplus \mathcal{G}_q$ and $P_{\mathcal{G}}$ is the M -orthogonal projector onto \mathcal{G} . Suppose W_i is an M -orthonormal matrix whose columns span \mathcal{G}_i . Then $MW_i = \mu_i \Gamma W_i$ and $P_{\mathcal{G}_i} = W_i W_i^\top M$, which leads to

$$\Gamma^{-1} M P_{\mathcal{G}_i} = \Gamma^{-1} M W_i W_i^\top M = \mu_i W_i W_i^\top M = \mu_i P_{\mathcal{G}_i}.$$

Therefore, it holds that

$$\Gamma^{-1} M x = \Gamma^{-1} M \sum_{i=1}^t P_{\mathcal{G}_i} x = \sum_{i=1}^t \mu_i P_{\mathcal{G}_i} x, \quad \forall x \in \mathbb{R}^m.$$

For any $j \geq 0$, it follows that

$$\tilde{w}_i := (\Gamma^{-1} M)^j \Gamma^{-1} y = \left(\sum_{i=1}^t \mu_i P_{\mathcal{G}_i} \right)^j \Gamma^{-1} y = \sum_{i=1}^t \mu_i^j P_{\mathcal{G}_i} \Gamma^{-1} y = \sum_{i=1}^q \mu_i^j P_{\mathcal{G}_i} \Gamma^{-1} y.$$

By [Proposition 3.1](#), we have $\hat{p}_i^{(k_b)} \in \mathcal{K}_{k_b}(\Gamma^{-1} M, \Gamma^{-1} y) = \text{span}\{\tilde{w}_i\}_{i=0}^{k_b-1}$, and

$$(\tilde{w}_0, \dots, \tilde{w}_{k_b-1}) = (P_{\mathcal{G}_1} \Gamma^{-1} y, \dots, P_{\mathcal{G}_q} \Gamma^{-1} y) \begin{pmatrix} 1 & \mu_1 & \dots & \mu_1^{k_b-1} \\ 1 & \mu_2 & \dots & \mu_2^{k_b-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mu_q & \dots & \mu_q^{k_b-1} \end{pmatrix} =: \tilde{G} T_{k_b},$$

where $\tilde{G} = (P_{\mathcal{G}_1} \Gamma^{-1} y, \dots, P_{\mathcal{G}_q} \Gamma^{-1} y)$. Since the q -by- q leading part of T_{k_b} is a Vandermonde matrix with $\mu_i \neq \mu_j$, the rank of T_{k_b} is at most q . Thus the rank of $\tilde{W} := (w_0, \dots, w_{k_b-1})$ is at most q , which is achieved if $k_b = q$. This implies that the expansion of $\mathcal{K}_k(\Gamma^{-1} M, \Gamma^{-1} y)$ ends at $k = q$, hence $k_b = q$. Since $T_{k_b} = T_q$ is nonsingular, it holds that $\mathcal{R}(\tilde{W}) = \mathcal{R}(\tilde{G})$, and we can write $\hat{p}_i^{(k_b)}$ as $\hat{p}_i^{(k_b)} = \tilde{G} c$ with a nonzero $c \in \mathbb{R}^q$. Now we immediately obtain $P_{\mathcal{G}} \hat{p}_i^{(k_b)} = \hat{p}_i^{(k_b)}$, which is the desired result. \square

By [Theorem 3.2](#), the subspace $\iota(\mathcal{S}_k)$ progressively captures one component from each generalized eigenspace \mathcal{G}_j for which $P_{\mathcal{G}_j} \Gamma^{-1} y \neq 0$. Moreover, those components associated with larger generalized eigenvalues are approximated more rapidly, due to the well-known extremal eigenvalue convergence properties of Lanczos-type methods [\[41\]](#). Since the observation data takes the form $y = Gx + \eta$ with $\eta \sim \mathcal{N}(0, \Gamma)$, we have $\Gamma^{-1} y = \Gamma^{-1} Gx + \Gamma^{-1} \eta$ where $\Gamma^{-1} \eta \sim \mathcal{N}(0, \Gamma^{-1})$. Thus, we have $P_{\mathcal{G}_i} \Gamma^{-1} y \neq 0$ almost surely for all $i = 1, \dots, t$, as the Gaussian distribution has full support. For ill-posed problems, the matrix $M = G \Sigma G^\top$ is typically severely ill-conditioned, and the generalized eigenvalues of (M, Γ) tend to decay gradually toward zero without noticeable gaps. In such cases, eigenvalue multiplicities are typically small (often equal to one in practice). This indicates that, in typical settings, the Q-GKB iteration progressively captures all directions of the dominant data-informed subspaces through $\iota(\mathcal{S}_k)$.

Since $V_k s_i^{(k)}$ progressively approximates the dominant generalized eigenvectors of (M, Γ) , this motivates the following low-rank approximation of H , analogous to [\(2.8\)](#). After k steps of Q-GKB, define the rank- k symmetric positive semidefinite matrix

$$\hat{H}_k = G^\top \left(\sum_{i=1}^k \theta_i^{(k)} (V_k s_i^{(k)})(V_k s_i^{(k)})^\top \right) G = G^\top V_k B_k^\top B_k V_k^\top G. \quad (3.11)$$

By [Theorem 3.2](#) and its proof, if all nonzero generalized eigenvalues of (M, Γ) are simple and $P_{\mathcal{G}_i} \Gamma^{-1} y \neq 0$ for $i = 1, \dots, l$, while $P_{\mathcal{G}_i} \Gamma^{-1} y = 0$ for $i > k$, then the breakdown occurs at $k_b = l$ and $\widehat{H}_l = H_l$. Replacing $H = G^\top \Gamma^{-1} G$ in [\(1.3\)](#) by \widehat{H}_k , we define the *approximate Gaussian posterior* with mean and covariance given by

$$\widehat{x}_\lambda^{(k)} = \widehat{C}_\lambda^{(k)} G^\top \Gamma^{-1} y, \quad \widehat{C}_\lambda^{(k)} = (\widehat{H}_k + \lambda \Sigma^{-1})^{-1}. \quad (3.12)$$

In the next section, we show how to simultaneously estimate the hyperparameter λ and compute the approximate posterior mean and covariance efficiently within the Q-GKB iteration.

4 Iterative approximation of the Bayesian posterior

The goal of this section is to construct efficient approximations of the Bayesian posterior based on Q-GKB iteration, and to analyze the accuracy of these approximations. We integrate empirical Bayesian inference into the Q-GKB iteration to obtain an efficient estimate of λ within the data-informed Krylov subspaces. Given $\lambda = \lambda_k$, we then compute the approximations $\widehat{x}_\lambda^{(k)}$ and $\widehat{C}_\lambda^{(k)}$ using byproducts of the Q-GKB iteration.

4.1 Approximate the Bayesian posterior by Q-GKB

In the empirical Bayesian (EB) inference, the hyperparameter λ is treated as a random variable and correspondingly, the prior of x is conditioned on λ . Suppose the prior of λ is $\pi(\lambda)$. Since $x | \lambda \sim \mathcal{N}(0, \lambda^{-1} \Sigma)$ and $y | (x, \lambda) \sim \mathcal{N}(Gx, \Gamma)$, using Bayes' formula, we have

$$\pi(x, \lambda | y) = \frac{\pi(y | x, \lambda) \pi(x | \lambda) \pi(\lambda)}{\pi(y)} \propto \frac{\pi(\lambda) \lambda^{n/2} \exp\left(-\frac{1}{2} \|y - Gx\|_{\Gamma^{-1}}^2 - \frac{\lambda}{2} \|x\|_{\Sigma^{-1}}^2\right)}{\det(\Gamma)^{1/2} \det(\Sigma)^{1/2}}.$$

Integrating out x , the marginal posterior of λ is

$$\begin{aligned} \pi(\lambda | y) &= \int_{\mathbb{R}^n} \pi(x, \lambda | y) dx \\ &\propto \pi(\lambda) \lambda^{n/2} \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2} \|y - Gx\|_{\Gamma^{-1}}^2 - \frac{\lambda}{2} \|x\|_{\Sigma^{-1}}^2\right) dx \\ &\propto \pi(\lambda) \det(\Gamma + \lambda^{-1} G \Sigma G^\top)^{-1/2} \exp\left(-\frac{1}{2} y^\top (\Gamma + \lambda^{-1} G \Sigma G^\top)^{-1} y\right), \end{aligned}$$

see e.g., [\[14, 24\]](#). Notice that $\pi(\lambda | y) = \frac{\pi(y | \lambda) \pi(\lambda)}{\pi(y)} \propto \pi(\lambda) \pi(y | \lambda)$. In the EB method, the hyperparameter λ is typically estimated by maximizing the marginal likelihood $\pi(y | \lambda)$, rather than the posterior $\pi(\lambda | y)$. Ignoring constant terms, minimizing the negative log-marginal likelihood $-\log \pi(y | \lambda)$ is equivalent to solving the following optimization problem:

$$\min_{\lambda > 0} \mathcal{L}(\lambda) := \log \det(\Gamma + \lambda^{-1} M) + y^\top (\Gamma + \lambda^{-1} M)^{-1} y.$$

Although this problem is formulated in the data space, evaluating $\mathcal{L}(\lambda)$ remains computationally prohibitive for large m due to the cost of determinant and matrix inversion operations. To address this, we integrate the EB approach into the Q-GKB iteration, enabling an efficient, adaptive estimation of λ within low-dimensional data-informed Krylov subspaces.

After k steps of Q-GKB, define the symmetric positive semidefinite matrix

$$M_k = M V_k V_k^\top, \quad (4.1)$$

which has rank at most k . For the M -orthogonal projector

$$\begin{aligned} \Pi_k : \mathbb{R}^m / \mathcal{N}(M) &\longrightarrow \mathcal{S}_k \\ [z] &\longmapsto \sum_{i=1}^k \langle [z], [v_i] \rangle_M [v_i], \end{aligned}$$

it holds that

$$\mathcal{T}(\Pi_k[z]) = \mathcal{T}([V_k V_k^\top M z]) = M V_k V_k^\top M z = M_k z.$$

This implies that M_k is the compression of M onto the Krylov subspace \mathcal{S}_k . Accordingly, M_k provides a natural low-rank approximation of M . Using the relation $M V_k = U_{k+1} B_k$, we obtain

$$M_k = M V_k V_k^\top M = U_{k+1} B_k B_k^\top U_{k+1}^\top. \quad (4.2)$$

Thus M_k can be formed using only the small scale matrices B_k and U_{k+1} generated by Q-GKB. Replacing M by M_k in $\mathcal{L}(\lambda)$ yields the k th approximate negative log-marginal likelihood:

$$\mathcal{L}^{(k)}(\lambda) = \log \det(\Gamma + \lambda^{-1} M_k) + y^\top (\Gamma + \lambda^{-1} M_k)^{-1} y. \quad (4.3)$$

The following result shows that the approximate marginal likelihood can be computed efficiently using only the singular value decomposition (SVD) of B_k .

Proposition 4.1 *Suppose that the compact form SVD of B_k is $B_k = P S Q^\top$, where $P \in \mathbb{R}^{(k+1) \times k}$ has 2-orthonormal columns and $S = \text{diag}(s_1, \dots, s_k)$ with $s_i > 0$. Then*

$$\mathcal{L}^{(k)}(\lambda) = \log \det(\Gamma) + \sum_{i=1}^k \log \left(1 + \frac{s_i^2}{\lambda} \right) + \beta_1^2 \left(1 - \sum_{i=1}^k \frac{s_i^2}{s_i^2 + \lambda} p_{1i}^2 \right), \quad (4.4)$$

where (p_{11}, \dots, p_{1k}) is the first row of P .

Proof. For the determinant term in $\mathcal{L}^{(k)}(\lambda)$, we write

$$\begin{aligned} \Gamma + \lambda^{-1} M_k &= \Gamma^{1/2} \left(I + \lambda^{-1} \Gamma^{-1/2} U_{k+1} B_k B_k^\top U_{k+1}^\top \Gamma^{-1/2} \right) \Gamma^{1/2} \\ &= \Gamma^{1/2} \left(I + \lambda^{-1} W B_k B_k^\top W^\top \right) \Gamma^{1/2}, \end{aligned}$$

where $W := \Gamma^{-1/2} U_{k+1}$ is column 2-orthonormal. Therefore, it holds that

$$\det(\Gamma + \lambda^{-1} M_k) = \det(\Gamma) \det \left(I + \lambda^{-1} W B_k B_k^\top W^\top \right) = \det(\Gamma) \det \left(I + \lambda^{-1} B_k^\top B_k \right),$$

where we have used the determinant identity $\det(I + AB) = \det(I + BA)$. Since the eigenvalues of $B_k^\top B_k$ are s_1^2, \dots, s_k^2 , we obtain

$$\det(\Gamma + \lambda^{-1} M_k) = \det(\Gamma) \prod_{i=1}^k \left(1 + \frac{s_i^2}{\lambda} \right).$$

For the quadratic term in $\mathcal{L}^{(k)}(\lambda)$, using $\Gamma^{-1/2} y = \Gamma^{-1/2} \beta_1 U_{k+1} e_1 = \beta_1 W e_1$, we have

$$\begin{aligned} y^\top (\Gamma + \lambda^{-1} M_k)^{-1} y &= y^\top \Gamma^{-1/2} \left(I + \lambda^{-1} W B_k B_k^\top W^\top \right)^{-1} \Gamma^{-1/2} y \\ &= \beta_1^2 e_1^\top W^\top \left(I + \lambda^{-1} W B_k B_k^\top W^\top \right)^{-1} W e_1 \\ &= \beta_1^2 e_1^\top \left[I - B_k (B_k^\top B_k + \lambda I)^{-1} B_k^\top \right] e_1 \\ &= \beta_1^2 \left[1 - e_1^\top P \text{diag} \left(\frac{s_1^2}{s_1^2 + \lambda}, \dots, \frac{s_k^2}{s_k^2 + \lambda} \right) P^\top e_1 \right] \end{aligned}$$

where we have used the identity $(I + \lambda^{-1} B B^\top)^{-1} = I - B (B^\top B + \lambda I)^{-1} B^\top$ and the SVD of B_k . Combining the above two terms leads to the desired result. \square

At the k th iteration, we define $\lambda_k = \text{argmin}_{\lambda > 0} \mathcal{L}^{(k)}(\lambda)$. By [Proposition 4.1](#), the term $\log \det(\Gamma)$ does not affect the minimization of $\mathcal{L}^{(k)}(\lambda)$ and can therefore be omitted. Hence, the dominant computational cost for determining λ_k is the SVD of B_k , which requires only $O(k^3)$ flops.

By [Theorem 3.1](#), we approximate z_λ in \mathcal{S}_k by solving [\(3.10\)](#) with $\lambda = \lambda_k$. Since any $[z] \in \mathcal{S}_k$ can be written as $[z] = [V_k \xi]$ for some $\xi \in \mathbb{R}^k$ and $Mz = MV_k \xi$, using [\(3.9\)](#) and the Γ^{-1} -orthonormality of U_{k+1} , we obtain

$$\|Mz - y\|_{\Gamma^{-1}} = \|U_{k+1}(B_k \xi - \beta_1 e_1)\|_{\Gamma^{-1}} = \|B_k \xi - \beta_1 e_1\|_2.$$

Moreover, using the M -orthonormality of V_k , we have

$$\|[z]\|_M^2 = \|[V_k \xi]\|_M^2 = \xi^\top V_k^\top M V_k \xi = \|\xi\|_2^2.$$

Therefore, [\(3.10\)](#) is equivalent to

$$\min_{\xi \in \mathbb{R}^k} \{\|B_k \xi - \beta_1 e_1\|_2^2 + \lambda_k \|\xi\|_2^2\}, \quad (4.5)$$

which has a unique minimizer $\xi_{\lambda_k}^{(k)} \in \mathbb{R}^k$. Consequently, the k th approximation to z_λ is $[z_{\lambda_k}^{(k)}] = [V_k \xi_{\lambda_k}^{(k)}]$, and corresponding iteratively regularized solution is

$$x_{\lambda_k}^{(k)} = \Sigma G^\top V_k \xi_{\lambda_k}^{(k)}, \quad \xi_{\lambda_k}^{(k)} = (B_k^\top B_k + \lambda_k I)^{-1} B_k^\top \beta_1 e_1. \quad (4.6)$$

The following result provides an efficient computation of $\hat{x}_\lambda^{(k)}$ and $\hat{C}_\lambda^{(k)}$ defined in [\(3.12\)](#), and shows that the approximate posterior mean in the data-informed Krylov subspace coincides with the iteratively regularized solution. In what follows, we write λ in place of λ_k at the k th step to emphasize that the results hold for any value of λ .

Proposition 4.2 *At the k th step, the approximate posterior mean coincides with the iteratively regularized solution, namely,*

$$\hat{x}_\lambda^{(k)} = x_\lambda^{(k)} = \Sigma G^\top V_k (B_k^\top B_k + \lambda I)^{-1} B_k^\top \beta_1 e_1, \quad (4.7)$$

and the approximate covariance admits the representation

$$\hat{C}_\lambda^{(k)} = \lambda^{-1} \Sigma - \lambda^{-1} \Sigma G^\top V_k (\lambda I + B_k^\top B_k)^{-1} B_k^\top B_k V_k^\top G \Sigma. \quad (4.8)$$

Proof. Applying the Woodbury identity to the expression of $\hat{C}_\lambda^{(k)}$, we have

$$\hat{C}_\lambda^{(k)} = \lambda^{-1} \Sigma - \lambda^{-2} \Sigma G^\top V_k \left((B_k^\top B_k)^{-1} + \lambda^{-1} V_k^\top G \Sigma G^\top V_k \right)^{-1} V_k^\top G \Sigma.$$

Using $V_k^\top G \Sigma G^\top V_k = V_k^\top M V_k = I$ and $(T_k^{-1} + \lambda^{-1} I)^{-1} = \lambda (\lambda I + T_k)^{-1} T_k$ with $T_k := B_k^\top B_k$, we immediately obtain [\(4.8\)](#).

Now we show $\hat{x}_\lambda^{(k)} = x_\lambda^{(k)}$. Using [\(4.8\)](#), we have

$$\hat{x}_\lambda^{(k)} = \lambda^{-1} \Sigma G^\top \Gamma^{-1} y - \lambda^{-1} \Sigma G^\top V_k (\lambda I + T_k)^{-1} T_k V_k^\top M \Gamma^{-1} y.$$

Since $M V_k = U_{k+1} B_k$, we have $V_k^\top M = B_k^\top U_{k+1}^\top$. From $\beta_1 U_{k+1} e_1 = y$, we have $U_{k+1}^\top \Gamma^{-1} y = \beta_1 e_1$. Therefore, $V_k^\top M \Gamma^{-1} y = B_k^\top U_{k+1}^\top \Gamma^{-1} y = B_k^\top \beta_1 e_1$. On the other hand, from [\(3.9\)](#), we have $\Gamma^{-1} y = \beta_1 \Gamma^{-1} U_{k+1} e_1 = V_k B_k^\top \beta_1 e_1$. It follows that $\lambda^{-1} \Sigma G^\top \Gamma^{-1} y = \lambda^{-1} \Sigma G^\top V_k B_k^\top \beta_1 e_1$. Substituting these into the expression for $\hat{x}_\lambda^{(k)}$ gives

$$\begin{aligned} \hat{x}_\lambda^{(k)} &= \lambda^{-1} \Sigma G^\top V_k B_k^\top \beta_1 e_1 - \lambda^{-1} \Sigma G^\top V_k (\lambda I + T_k)^{-1} T_k B_k^\top \beta_1 e_1 \\ &= \Sigma G^\top V_k [\lambda^{-1} I - \lambda^{-1} (\lambda I + T_k)^{-1} T_k] B_k^\top \beta_1 e_1 \\ &= \Sigma G^\top V_k (\lambda I + T_k)^{-1} B_k^\top \beta_1 e_1, \end{aligned}$$

where we used $I - (\lambda I + T_k)^{-1} T_k = \lambda (\lambda I + T_k)^{-1}$. This completes the proof. \square

Algorithm 2 Q-GKB based posterior approximation

Input: Matrix $G \in \mathbb{R}^{m \times n}$, SPD matrices $\Sigma \in \mathbb{R}^{n \times n}$, $\Gamma \in \mathbb{R}^{m \times m}$, vector $y \in \mathbb{R}^m$

- 1: Compute $\beta_1, \alpha_1, u_1, v_1$ by Q-GKB
- 2: **for** $k = 1, 2, \dots, K$ **do**
- 3: Compute $\beta_{k+1}, \alpha_{k+1}, u_{k+1}, v_{k+1}$ by Q-GKB; form B_k and V_k
- 4: (Terminate the iteration if β_{k+1} or α_{k+1} is extremely small)
- 5: Compute the SVD of B_k ; form $\mathcal{L}^{(k)}(\lambda)$ as (4.4) ▷ Discard $\log \det(\Gamma)$
- 6: Compute $\lambda_k = \operatorname{argmin}_{\lambda > 0} \mathcal{L}^{(k)}(\lambda)$
- 7: Compute $\hat{x}_{\lambda_k}^{(k)}$ and $\hat{C}_{\lambda_k}^{(k)}$ by (4.7) and (4.8)
- 8: **end for**

Output: $\hat{x}_{\lambda_K}^{(K)}, \hat{C}_{\lambda_K}^{(K)}$

By this result, the posterior mean can be approximated iteratively as the Q-GKB iteration proceeds, while the approximate posterior covariance can be computed efficiently using only the inversion of small-scale matrices. The resulting algorithm is summarized in Algorithm 2. As a stopping criterion, we can choose a tolerance $\text{tol} > 0$ and terminate the iteration when the relative change in λ satisfies $|\lambda_k - \lambda_{k+1}|/\lambda_k \leq \text{tol}$.

We remark that, for large-scale problems, the prior covariance matrix Σ need not be formed explicitly in the algorithm. Instead, matrix-vector products of the form Σv can be computed efficiently. In particular, when the prior is given by a stationary or translation invariant kernels, the product Σv can be evaluated efficiently using the fast Fourier transform (FFT) [35, 47]. We will illustrate such a case through a numerical example.

4.2 Accuracy of the approximate posterior

We quantify the accuracy of the approximate posterior distribution constructed via the Q-GKB iteration. The exact posterior and its k th approximation are given by

$$\pi = \mathcal{N}(x_\lambda, C_\lambda), \quad \hat{\pi}_k = \mathcal{N}(\hat{x}_\lambda^{(k)}, \hat{C}_\lambda^{(k)}). \quad (4.9)$$

To measure the difference between the exact covariance and its approximation, we employ the Förstner distance. Denote by $\operatorname{Tr}(\cdot)$ the trace of a matrix. For two d -by- d SPD matrices A and B , the Förstner distance is defined by

$$d_F^2(A, B) = \operatorname{Tr} \left[\log^2(A^{-1/2} B A^{-1/2}) \right] = \sum_{i=1}^d \log^2(\sigma_i),$$

where $\{\sigma_i\}_{i=1}^d$ are the positive generalized eigenvalues of (A, B) [13]. It has been shown in [13, 43] that the Förstner distance is particularly well suited for comparing covariance matrices. To measure the difference between two probability distributions, we use the Kullback–Leibler (KL) divergence. Let $\pi_1 = \mathcal{N}(m_1, \Sigma_1)$ and $\pi_2 = \mathcal{N}(m_2, \Sigma_2)$ be two Gaussian distributions on \mathbb{R}^d with positive definite covariances. The KL divergence between them is defined as

$$D_{\text{KL}}(\pi_1 \parallel \pi_2) = \frac{1}{2} \left[\operatorname{Tr}(\Sigma_2^{-1} \Sigma_1) - d - \log \det(\Sigma_2^{-1} \Sigma_1) + (m_2 - m_1)^\top \Sigma_2^{-1} (m_2 - m_1) \right].$$

Before establishing error bounds for the posterior approximation, we introduce two auxiliary lemmas that will be used in the subsequent analysis. We use $\|\cdot\|_F$ to denote the Frobenius norm of a matrix. The proofs are postponed to Section A.

Lemma 4.1 *Let $A = \Sigma^{1/2} H \Sigma^{1/2}$, $\hat{A}_k = \Sigma^{1/2} \hat{H}_k \Sigma^{1/2}$ and $\hat{A}_0 = 0$, and define*

$$\zeta_k = \operatorname{Tr}(A - \hat{A}_k), \quad \gamma_k = \|A - \hat{A}_k\|_F. \quad (4.10)$$

Then it hold that $\zeta_k, \gamma_k \geq 0$, and

$$\begin{cases} \zeta_{k+1} = \zeta_k - (\alpha_{k+1}^2 + \beta_{k+2}^2), & \zeta_0 = \text{Tr}(H\Sigma), \\ \gamma_{k+1}^2 = \gamma_k^2 - 2\alpha_{k+1}^2\beta_{k+1}^2 - (\alpha_{k+1}^2 + \beta_{k+2}^2)^2, & \gamma_0^2 = \text{Tr}(H\Sigma H\Sigma). \end{cases} \quad (4.11)$$

Lemma 4.2 Let A_1 and A_2 be two SPD matrixe of the same order, then

$$\|(I + A_1)^{-1} - (I + A_2)^{-1}\|_2 \leq \frac{\|A_1 - A_2\|_2}{1 + \|A_1 - A_2\|_2}.$$

We are now ready to quantify the discrepancy between the exact posterior and its approximation. For clarity, we again write λ in place of λ_k to emphasize that the result hold for any value of λ .

Theorem 4.1 Assume that Q-GKB runs for k steps with $k \leq k_b$, then we have

$$d_F(C_\lambda, \widehat{C}_\lambda^{(k)}) \leq \frac{\gamma_k}{\lambda}, \quad (4.12)$$

and

$$D_{\text{KL}}(\widehat{\pi}_k \|\pi) \leq \frac{1}{2\lambda} \left(\zeta_k + \frac{\alpha_1^2 \beta_1^2 \gamma_k^2}{\lambda(\lambda + \gamma_k)} \right). \quad (4.13)$$

Proof. We first give the upper bound for $d_F(C_\lambda, \widehat{C}_\lambda^{(k)})$. Rewriting C_λ and $\widehat{C}_\lambda^{(k)}$ as

$$C_\lambda = \Sigma^{1/2}(A + \lambda I)^{-1}\Sigma^{1/2}, \quad \widehat{C}_\lambda^{(k)} = \Sigma^{1/2}(\widehat{A}_k + \lambda I)^{-1}\Sigma^{1/2},$$

where $A = \Sigma^{1/2}H\Sigma^{1/2}$ and $\widehat{A}_k = \Sigma^{1/2}\widehat{H}_k\Sigma^{1/2}$, and using the property

$$d_F(A, B) = d_F(A^{-1}, B^{-1}) = d_F(NAN^\top, NBN^\top)$$

for any nonsingular matrix N [13], we get $d_F(C_\lambda, \widehat{C}_\lambda^{(k)}) = d_F(\lambda I + A, \lambda I + \widehat{A}_k)$. By definition of the F6rstner distance, we have

$$d_F^2(\lambda I + A, \lambda I + \widehat{A}_k) = \sum_{i=1}^m \log^2(\sigma_i(D_k)),$$

where $D_k = (\lambda I + A)(\lambda I + \widehat{A}_k)^{-1} = (I + \lambda^{-1}A)(I + \lambda^{-1}\widehat{A}_k)^{-1}$, and $\{\sigma_i(D_k)\}_i$ are the eigenvalues of D_k that equal to the generalized eigenvalue of $(\lambda I + A, \lambda I + \widehat{A}_k)$. Using the inequality $(\log t)^2 \leq t + t^{-1} - 2$ for $t > 0$, we obtain

$$d_F^2(\lambda I + A, \lambda I + \widehat{A}_k) \leq \text{Tr}(D_k) + \text{Tr}(D_k^{-1}) - 2m.$$

Now let $E_k = A - \widehat{A}_k$. Then it holds that

$$D_k - I = E_k(\lambda I + \widehat{A}_k)^{-1}, \quad D_k^{-1} - I = -E_k(\lambda I + A)^{-1}.$$

It follows that

$$\begin{aligned} \text{Tr}(D_k) + \text{Tr}(D_k^{-1}) - 2n &= \text{Tr}(E_k(\lambda I + \widehat{A}_k)^{-1}) - \text{Tr}(E_k(\lambda I + A)^{-1}) \\ &= \text{Tr}\left(E_k(\lambda I + \widehat{A}_k)^{-1}E_k(\lambda I + A)^{-1}\right), \end{aligned}$$

where we have used the identity

$$(\lambda I + \widehat{A}_k)^{-1} - (\lambda I + A)^{-1} = (\lambda I + \widehat{A}_k)^{-1}E_k(\lambda I + A)^{-1}.$$

Since $\|(\lambda I + A)^{-1}\|_2 \leq \lambda^{-1}$ and $\|(\lambda I + \widehat{A}_k)^{-1}\|_2 \leq \lambda^{-1}$, using [Lemma 4.1](#) we obtain

$$\text{Tr}\left(E_k(\lambda I + \widehat{A}_k)^{-1}E_k(\lambda I + A)^{-1}\right) \leq \|E_k(\lambda I + \widehat{A}_k)^{-1}\|_F \|E_k(\lambda I + A)^{-1}\|_F \leq \frac{\|E_k\|_F^2}{\lambda^2} = \frac{\gamma_k^2}{\lambda^2}.$$

This is the desired bound for $d_F(C_\lambda, \widehat{C}_\lambda^{(k)})$.

The KL divergence between $\widehat{\pi}_k$ and π is

$$D_{\text{KL}}(\widehat{\pi}_k \| \pi) = \frac{1}{2} \left[\underbrace{\text{Tr}(C_\lambda^{-1} \widehat{C}_\lambda^{(k)}) - n - \log \det(C_\lambda^{-1} \widehat{C}_\lambda^{(k)})}_I + \underbrace{\|x_\lambda - \widehat{x}_\lambda^{(k)}\|_{C_\lambda^{-1}}^2}_II \right].$$

In what follows, we bound the terms I and II separately. Noticing that $C_\lambda^{-1} \widehat{C}_\lambda^{(k)} = \Sigma^{-1/2} (A + \lambda I) (\widehat{A}_k + \lambda I)^{-1} \Sigma^{1/2}$, we have

$$\begin{aligned} I &= \text{Tr}((A + \lambda I)(\widehat{A}_k + \lambda I)^{-1}) - n - \log \det((A + \lambda I)(\widehat{A}_k + \lambda I)^{-1}) \\ &= \text{Tr}(E_k(\widehat{A}_k + \lambda I)^{-1}) - \left(\log \det(A + \lambda I) - \log \det(\widehat{A}_k + \lambda I) \right), \end{aligned}$$

where $E_k = A - \widehat{A}_k$. From the proof of [Lemma 4.1](#), we have $\widehat{A}_k = Q_k T_k Q_k^\top$ and $T_k := B_k^\top B_k = Q_k^\top A Q_k$, where $Q_k \in \mathbb{R}^{m \times k}$ have orthonormal columns. Combining these relations with the Woodbury identity, we get

$$(\widehat{A}_k + \lambda I)^{-1} = \lambda^{-1} I - \lambda^{-1} Q_k T_k (T_k + \lambda I)^{-1} Q_k^\top,$$

which leads to

$$\text{Tr}(E_k(\widehat{A}_k + \lambda I)^{-1}) = \lambda^{-1} \text{Tr}(E_k) - \lambda^{-1} \text{Tr}(Q_k^\top E_k Q_k T_k (T_k + \lambda I)^{-1}) = \lambda^{-1} \text{Tr}(E_k),$$

since $Q_k^\top E_k Q_k = Q_k^\top A Q_k - Q_k^\top \widehat{A}_k Q_k = T_k - T_k = 0$. Therefore, we get

$$\text{Tr}((A + \lambda I)(\widehat{A}_k + \lambda I)^{-1}) - n = \frac{\zeta_k}{\lambda}.$$

Next, let $\theta_1 \geq \dots \geq \theta_n \geq 0$ be the eigenvalues of A , and $\hat{\theta}_1 \geq \dots \geq \hat{\theta}_k \geq 0$ be the eigenvalues of T_k . Noticing that $T_k = Q_k^\top A Q_k$, the Cauchy interlacing theorem [\[19\]](#) implies that $\theta_i \geq \hat{\theta}_i$ for $i = 1, \dots, k$. Since the eigenvalues of $\widehat{A}_k = Q_k T_k Q_k^\top$ are $\{\hat{\theta}_1, \dots, \hat{\theta}_k, 0, \dots, 0\}$, we have

$$\det(\widehat{A}_k + \lambda I) = \lambda^{n-k} \prod_{i=1}^k (\lambda + \hat{\theta}_i) \leq \prod_{i=1}^n (\lambda + \theta_i) = \det(A + \lambda I).$$

Hence $\log \det(A + \lambda I) - \log \det(\widehat{A}_k + \lambda I) \geq 0$. Therefore, we obtain $I \leq \frac{\zeta_k}{\lambda}$. For the term II, by the formulas for the exact and approximate posterior means, we have

$$x_\lambda - \widehat{x}_\lambda^{(k)} = \Sigma^{1/2} \left[(A + \lambda I)^{-1} - (\widehat{A}_k + \lambda I)^{-1} \right] \Sigma^{1/2} G^\top \Gamma^{-1} y.$$

Let $F_k = (A + \lambda I)^{-1} - (\widehat{A}_k + \lambda I)^{-1}$ and $\hat{y} = \Sigma^{1/2} G^\top \Gamma^{-1} y$. Using $C_\lambda^{-1} = \Sigma^{-1/2} (A + \lambda I) \Sigma^{-1/2}$, we have

$$\|x_\lambda - \widehat{x}_\lambda^{(k)}\|_{C_\lambda^{-1}}^2 = \hat{y}^\top F_k (A + \lambda I) F_k \hat{y} \leq \|F_k\|_2 \|(A + \lambda I) F_k\|_2 \|\hat{y}\|_2^2.$$

Using $\Gamma^{-1} y = V_k B_k^\top \beta_1 e_1 = \alpha_1 \beta_1 v_1$ and $v_1^\top M v_1 = 1$, we get $\|\hat{y}\|_2^2 = (\Gamma^{-1} y)^\top M (\Gamma^{-1} y) = \alpha_1^2 \beta_1^2$. Using [Lemma 4.2](#), we get

$$\|F_k\|_2 = \lambda^{-1} \|(I + \lambda^{-1} A)^{-1} - (I + \lambda^{-1} \widehat{A}_k)^{-1}\|_2 \leq \frac{1}{\lambda^2} \frac{\|E_k\|_2}{1 + \lambda^{-1} \|E_k\|_2} \leq \frac{\gamma_k}{\lambda(\lambda + \gamma_k)},$$

where the last “ \leq ” uses that $\|E\|_2 \leq \|E\|_F = \gamma_k$ and $\frac{t}{1+\lambda^{-1}t}$ is monotonically increasing for $t \geq 0$. Noticing that

$$(A + \lambda I) F_k = (A + \lambda I)(A + \lambda I)^{-1} \left[(\widehat{A}_k + \lambda I) - (A + \lambda I) \right] (\widehat{A}_k + \lambda I)^{-1} = -E_k (\widehat{A}_k + \lambda I)^{-1},$$

we get $\|(A + \lambda I) F_k\|_2 \leq \|E_k\|_2 \|(\widehat{A}_k + \lambda I)^{-1}\|_2 \leq \frac{\gamma_k}{\lambda}$ since \widehat{A}_k is SPD. Therefore, we obtain

$$II = \|x_\lambda - \widehat{x}_\lambda^{(k)}\|_{C_\lambda^{-1}}^2 \leq \frac{\alpha_1^2 \beta_1^2 \gamma_k^2}{\lambda^2 (\lambda + \gamma_k)}. \quad (4.14)$$

Combining the upper bounds for I and II, we finally obtain the desired bound. \square

This result shows that the accuracy of the approximate posterior is bounded by the computable quantities ζ_k and γ_k , which can be updated iteratively via the recurrences (4.11) at negligible cost. In practice, these bounds can be more reliable than the directly computed true errors in some cases, since the latter may be contaminated by numerical ill-conditioning of the covariance matrices. Hence, the bounds can serve not only as theoretical guarantees, but also as practical and reliable indicators of posterior approximation accuracy.

5 Numerical experiments

We present three representative numerical examples of increasing scale and complexity to illustrate the proposed theory and demonstrate the performance of the method. From a one-dimensional integral equation to large-scale CT reconstruction, we show how the method adapts to different problem settings. In particular, the large-scale example highlights the effectiveness of the matrix-free formulation, where direct access to the posterior distribution is computationally prohibitive.

All numerical experiments are implemented in MATLAB R2025b, and our methods can be executed on a standard personal laptop. The source code is publicly available at <https://github.com/HaiboLi99/DS-Bayes>.

5.1 One-dimensional Fredholm integral equation

In the first example, we consider the one-dimensional Fredholm integral equation of the first kind:

$$g(s) = \int_a^b \phi(s, t)x(t)dt, \quad \phi(s, t) = \exp(-|s - t|/l), \quad (5.1)$$

where we set $a = -\pi/2$, $b = \pi/2$, and set $l = 10$ in the integral kernel ϕ . For the discretization, we use $n = 5000$ uniform grids on $[-\pi/2, \pi/2]$ and use $m = 3000$ uniform grids on $[-\pi/2, \pi/2]$ for the observation. This leads to the forward matrix $G \in \mathbb{R}^{3000 \times 5000}$. The unknown x is modeled as a Gaussian process on $[-\pi/2, \pi/2]$ with covariance induced by the Gaussian kernel

$$K(t_1, t_2) = \sigma^2 \exp\left(-\frac{(t_1 - t_2)^2}{2l^2}\right), \quad (5.2)$$

where l and σ^2 denote the correlation length and marginal variance, respectively. In the experiment, we set $l = 0.4$ and $\sigma = 0.2$, and draw a random sample from this Gaussian process on the uniform grid as the ground truth x_{true} . We generate noisy data y by adding a white Gaussian noise to Gx_{true} with noise level $\|\eta\|_2/\|Gx_{\text{true}}\|_2 = 0.005$. The true solution and noisy observation are shown in Figure 5.4.

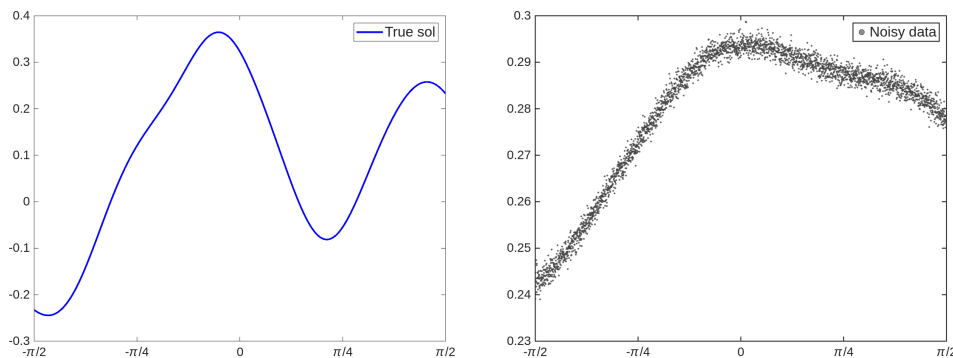


Figure 5.1: Illustration of the true solution and noisy observation data.

To reconstruct the unknown x , we construct the covariance matrix Σ using the Gaussian kernel with $l = 0.4$. The exact value of σ is supposed to be unknown, hence we set $\sigma = 1$ to form Σ . Given the

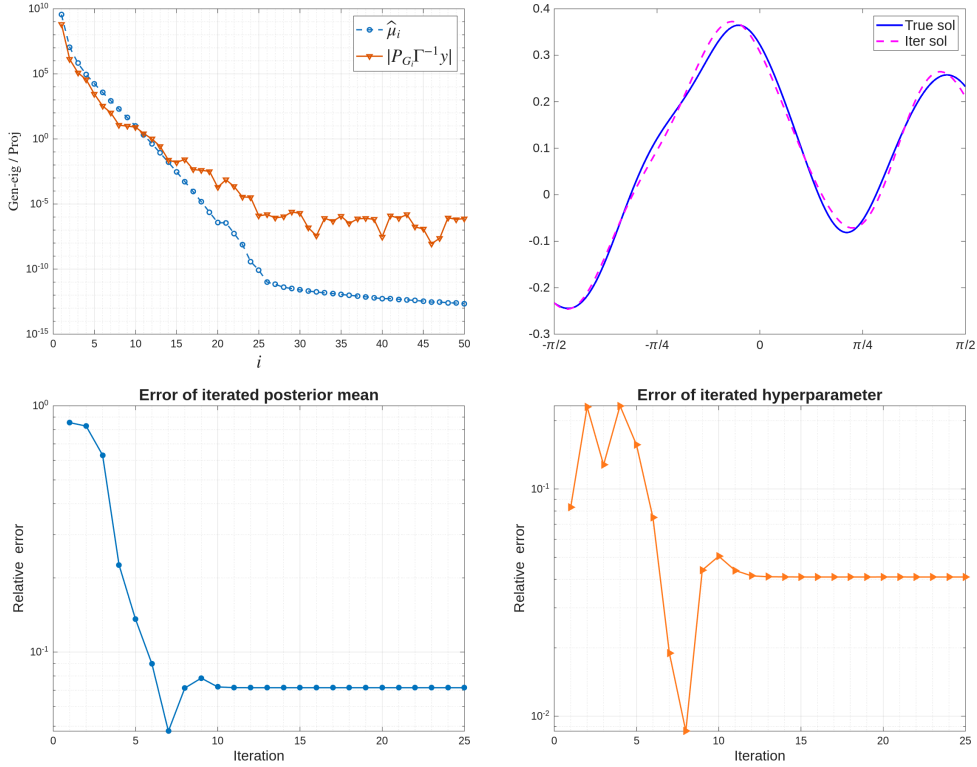


Figure 5.2: Eigenvalue decay of (M, Γ) , reconstructed solution, and convergence behavior of the Q-GKB method for the first example.

relationship $\sigma^2 = 1/\lambda$, the Q-GKB based empirical Bayesian inference (QGKB-EB) method iteratively estimates the optimal hyperparameter λ , thereby obtaining an accurate estimate of σ , which is then be used to compute the posterior distribution.

In Figure 5.2, we illustrate the eigenvalue decay of (M, Γ) and the quantities $P_{G_i}\Gamma^{-1}y$, which support Theorem 3.2. The relative error of the iterated hyperparameter is computed as $|\sigma_k - \sigma|/\sigma$, where $\sigma_k = 1/\sqrt{\lambda_k}$. All generalized eigenvalues $\hat{\mu}_i$ are simple and decay rapidly to zero without significant gaps, and all projected quantities $P_{G_i}\Gamma^{-1}y$ are nonzero. Consequently, the Q-GKB method quickly capture all eigenspaces associated with nonzero generalized eigenvalues as k increases. This behavior is due to the exponential decay of the eigenvalues of the forward operator induced by the integral kernel ϕ . From the top-left subfigure, if we regard $\hat{\mu}_i$ as negligible whenever it is smaller than 10^{-10} , then by Theorem 2.1, the effective dimension of the data space is approximately 25, which is much smaller than that of the parameter space. As a consequence, the bottom subfigures demonstrate rapid convergence of the Q-GKB iteration: the relative errors of both the reconstructed solution and the hyperparameter stabilize after approximately 15 iterations. The reconstructed solution at $k = 25$ is shown to closely match the true solution.

In Figure 5.3, we show the Förstner distance $d_F(C_\lambda, \hat{C}_\lambda^{(k)})$ and the KL divergence $D_{\text{KL}}(\hat{\pi}_k, \pi)$ together with their upper bounds established in Theorem 4.1, where at each iteration we set $\lambda = \lambda_k$. As the Q-GKB iteration proceeds, both quantities and their upper bounds exhibit a clear decreasing trend. In the late stage of the iteration, the computed values of $d_F(C_\lambda, \hat{C}_\lambda^{(k)})$ and $D_{\text{KL}}(\hat{\pi}_k, \pi)$ appear to stagnate at a level around 10^{-5} . This behavior is due to numerical instability in evaluating the exact posterior covariance, which becomes increasingly ill-conditioned. As a result, the computed Förstner distance and KL divergence reach a saturation level determined by numerical precision. This saturation should be interpreted as a numerical precision barrier rather than a failure of the theoretical bounds. In contrast, the upper bounds derived from ζ_k and γ_k continue to capture the true decay

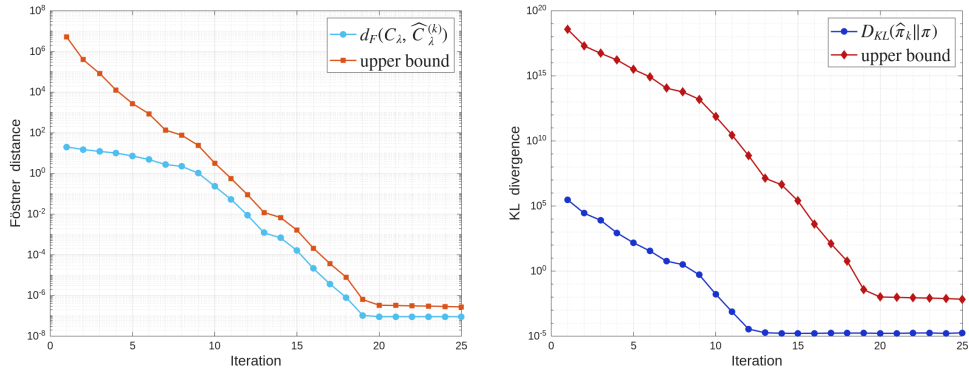


Figure 5.3: The Förstner distance and the KL divergence between the exact and approximate posterior, and their upper bounds.

of the posterior approximation error, albeit at a very slow rate. In practice, these computable upper bounds can serve as reliable stopping criteria for iterative posterior approximation.

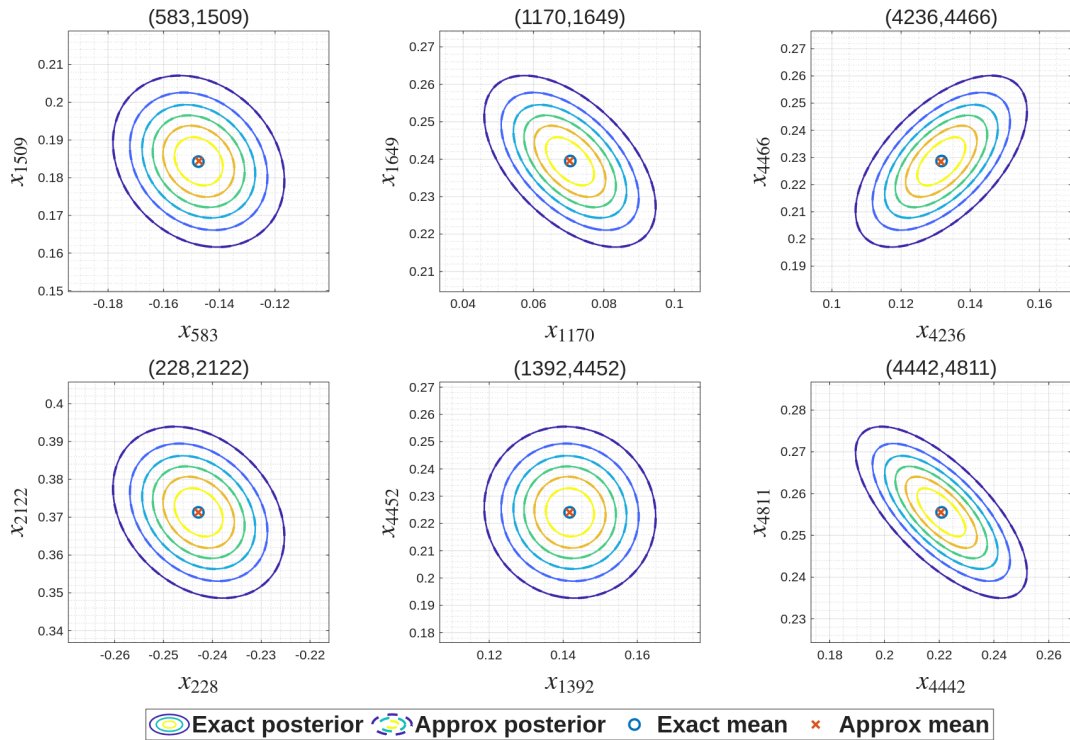


Figure 5.4: Contour plots of the marginal distributions of six randomly selected x_i - x_j pairs for both the exact and approximate posteriors.

To further assess the quality of the posterior approximation at $k = 25$, we compare the exact and approximate posteriors through their two-dimensional marginal distributions. In Figure 5.4, we depict contour plots of six randomly selected x_i - x_j pairs. For the exact posterior, we set $\lambda = \lambda_{25}$ and compute its mean and covariance using (2.1) and (2.2). The contours of the exact and approximate marginals are visually indistinguishable across all selected pairs, indicating an excellent agreement between the two distributions. This observation is consistent with the small KL divergence $D_{\text{KL}}(\hat{\pi}_k, \pi)$ at $k = 25$. Indeed, an error on the order of 10^{-5} implies that the two distributions are nearly indistinguishable

in practice. These results confirm that the Q-GKB approximation accurately captures the posterior structure in a low-dimensional data-informed Krylov subspace.

5.2 Two-dimensional image deblurring

In the second example, we consider a two-dimensional image deblurring problem on the domain $D = [0, 1]^2$. Let $x : D \rightarrow \mathbb{R}$ denote the unknown image, and $b : D \rightarrow \mathbb{R}$ denote the observed blurred data. The forward model is given by the two-dimensional Fredholm integral equation of the first kind:

$$b(\mathbf{s}) = \iint_D f_{\text{psf}}(\mathbf{s}, \mathbf{t}) x(\mathbf{t}) dt, \quad \mathbf{s} \in D, \quad (5.3)$$

where f_{psf} is a Gaussian point spread function (PSF) defined as

$$f_{\text{psf}}(s, t) = \exp\left(-\frac{\|s - t\|_2^2}{l_{\text{blur}}}\right), \quad (5.4)$$

with $l_{\text{blur}} > 0$ controlling the size of the blur spreading. In the experiment we set $l_{\text{blur}} = 0.01$. We discretize the domain D using n_1 uniform grids on each dimension such that the true image has $n_1 \times n_1$ pixels; the blurred observations are collected on a coarser $m_1 \times m_1$ grid, where we take $n_1 = 256$ and $m_1 = 128$. This leads to a discrete forward matrix $G \in \mathbb{R}^{16384 \times 65536}$.

The unknown image x is modeled as a two-dimensional Gaussian random field with a Matérn-type covariance structure. We adopt a separable Matérn kernel of the form

$$\text{Cov}(x(\mathbf{s}), x(\mathbf{t})) = K_1(s_1, t_1) K_1(s_2, t_2), \quad \mathbf{s} = (s_1, s_2), \mathbf{t} = (t_1, t_2). \quad (5.5)$$

Here, K_1 denotes the one-dimensional Matérn kernel

$$K_1(s, t) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}|s-t|}{\rho}\right)^\nu B_\nu\left(\frac{\sqrt{2\nu}|s-t|}{\rho}\right) \quad (5.6)$$

with smoothness parameter $\nu > 0$, correlation length $\rho > 0$, and marginal variance σ^2 , where Γ is the gamma function and B_ν is the modified Bessel function of the second kind [47]. In the experiment we set $\nu = 3$, $\rho = 0.1$, and $\sigma = 1$, and a random sample drawn from this Gaussian random field on the $n_1 \times n_1$ grid is used as the ground truth. A white Gaussian noise with relative noise level 0.01 is then added to the exact observation Gx_{true} to generate the noisy data. The true image and the corresponding noisy observation are shown in Figure 5.5.

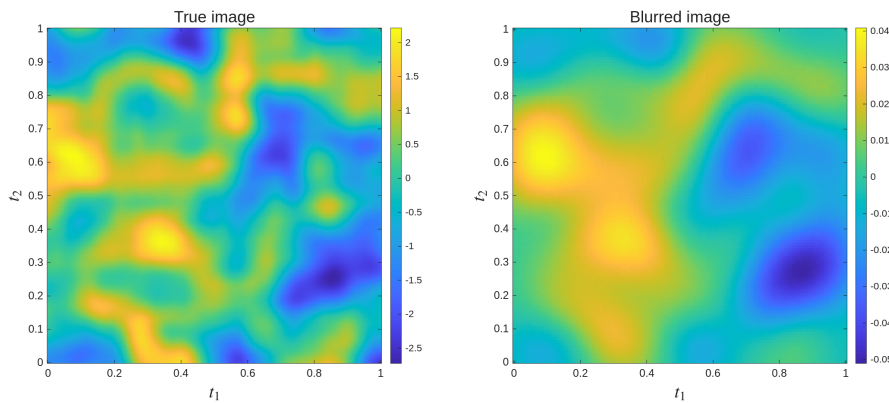


Figure 5.5: Illustration of the true image and noisy blurred image.

Similar to the first example, the prior covariance matrix Σ is constructed using the separable Matérn kernel with $\nu = 3$, $\rho = 0.1$, and $\sigma = 1$. This leads to Kronecker structure $\Sigma = \Sigma_1 \otimes \Sigma_1$,

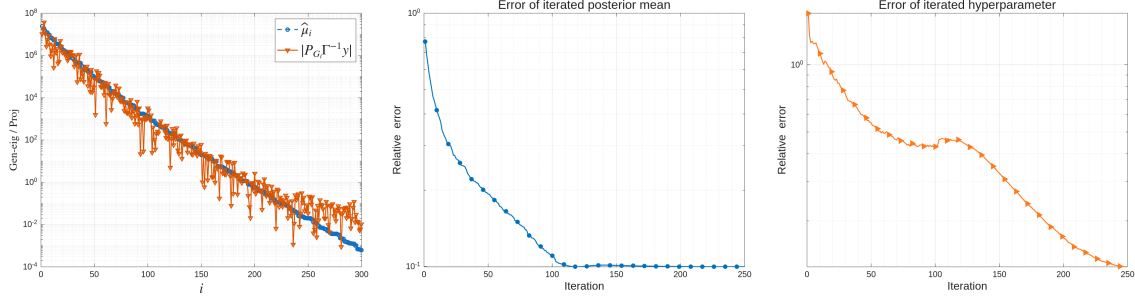


Figure 5.6: Eigenvalue decay of (M, Γ) and convergence behavior of the Q-GKB method for the first example.

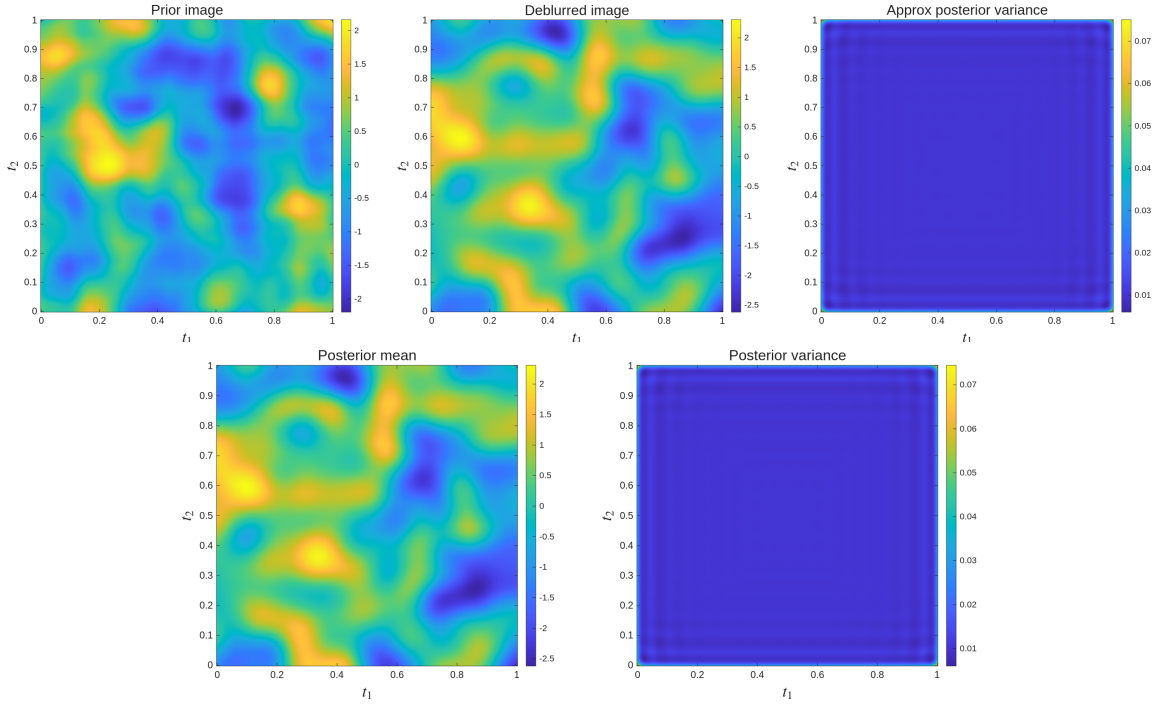


Figure 5.7: Illustration of the prior image, deblurred image at $k = 250$ and the image associated with exact posterior mean, and the corresponding variance of each pixel.

where $\Sigma_1 \in \mathbb{R}^{256 \times 256}$ is the covariance matrix induced by K_1 on a uniform grid of 256 points in $[0, 1]$. While Σ is initially formed with $\sigma = 1$, the marginal variance is treated as unknown and is iteratively estimated via the QGKB-EB method.

In Figure 5.6, we present the eigenvalue decay of the matrix pair (M, Γ) and the projected quantities $P_{\mathcal{G}_i} \Gamma^{-1} y$ for the second example. We find that all generalized eigenvalues $\hat{\mu}_i$ have a multiplicity of one and decay gradually toward zero without significant spectral gaps. Furthermore, since all projected quantities $P_{\mathcal{G}_i} \Gamma^{-1} y$ are non-zero, the Q-GKB method is capable of capturing all eigenspaces associated with nonzero generalized eigenvalues as k increases. With $l_{\text{blur}} = 0.01$, this example is a moderately ill-posed problem, which has a relatively slower decaying rate of the eigenvalues of the forward operator compared to the first example. Consequently, the generalized eigenvalues of (M, Γ) decay at a moderate rate, resulting in an effective data-space dimension of several hundreds. Accordingly, the convergence of the relative errors for the reconstructed solution stagnates after about 150 iterations, while the hyperparameter continues to evolve before eventually stabilizing after about 250 iterations.

We use Figure 5.7 to illustrate the effectiveness of the QGKB-EB method for the image deblurring

problem. For this large-scale case, computing the KL divergence to quantify approximation discrepancy is computationally prohibitive; instead, we examine the pixel-wise variance, represented by the diagonal elements of the posterior covariance matrix. The prior image shown is a random sample drawn from the prior distribution. The deblurred image and its corresponding variance represent the approximate posterior mean and covariance computed at $k = 250$. For comparison, the exact posterior mean and variance are computed using $\lambda = \lambda_{250}$ according to (2.1) and (2.2). Comparing the approximate and exact posteriors, we observe that the reconstructed image progressively converges to an accurate approximation of the posterior mean as the iteration proceeds. Furthermore, the approximate variance aligns closely with the exact posterior variance, thereby demonstrating the effectiveness of the Q-GKB-EB method for posterior approximation within the data-space formulation.

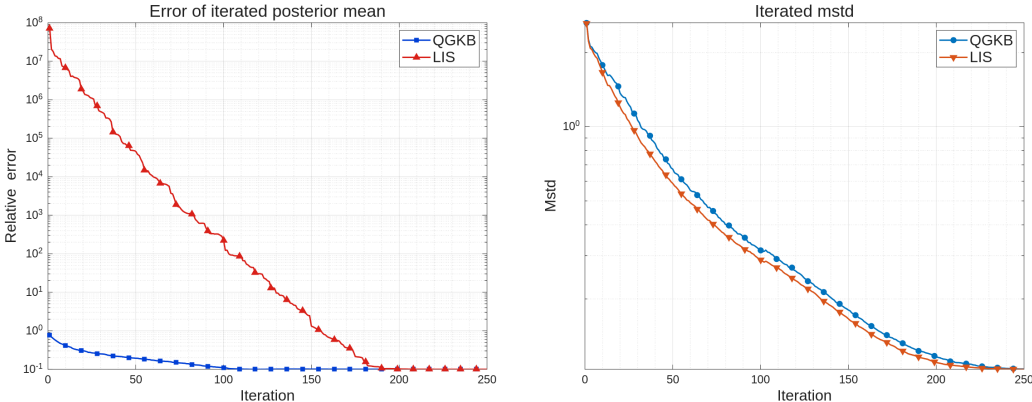


Figure 5.8: The relative error of the approximate posterior mean and the corresponding mean standard variation computed by the Q-GKB method and LIS method.

To further evaluate approximation accuracy and computational efficiency, we compare the proposed Q-GKB method with the direct LIS approach, which serves as a baseline. As shown in [43, 44], the LIS method provides an optimal low-rank approximation of the posterior covariance when the prior-to-posterior update is restricted to positive semidefinite matrices of rank at most k . For a fair comparison, the LIS method uses the same hyperparameter λ_k estimated by the QGKB-EB approach at iteration k . To quantify the convergence behavior of the covariance approximation, we define the mean standard deviation (mstd) as

$$\text{mstd} = \sqrt{\frac{1}{n} \sum_{i=1}^n \text{var}_i}, \tag{5.7}$$

where var_i denotes the posterior variance at the i th pixel. To compute a k -dimensional LIS, we follow [43]. Specifically, we first compute the Cholesky factorization $\Sigma = SS^\top$, and then compute the leading k eigenpairs of $S^\top HS$ using the MATLAB built-in function `eigs`, which yields the dominant LIS directions. From Figure 5.8, we have two important observations: first, in terms of the posterior mean, the Q-GKB method converges significantly faster than the LIS approach; second, for the covariance approximation measured by mstd, the LIS method exhibits slightly faster convergence. This difference can be explained by the intrinsic mechanisms of the two approaches: the Q-GKB iteration, initialized from the data vector y , prioritizes directions that are most relevant for reconstructing the posterior mean; in contrast, the LIS method targets the globally dominant eigendirections of the data-informed subspace, leading to a more efficient approximation of the posterior covariance. Nevertheless, as k becomes sufficiently large to capture the dominant data-informed subspaces, both methods achieve nearly identical accuracy in approximating the posterior distribution. This observation is consistent with the theoretical characterization of data-informed Krylov subspaces and their relation to the LIS method.

Before comparing the running time of the two methods, it is important to emphasize that for a 65536×65536 dense prior covariance Σ , the LIS method is computationally prohibitive on a standard

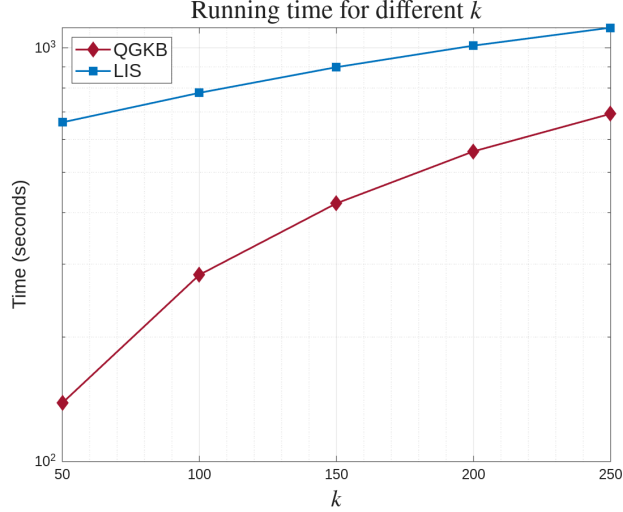


Figure 5.9: Comparison of the running time for the Q-GKB method and LIS method as the iteration number or LIS dimension increases.

personal laptop due to the memory requirements for storing Σ , let alone computing its Cholesky factorization. Thus, for a fair comparison of the running time, we run both the LIS and Q-GKB method in a high performance computing (HPC) cluster. As shown in Figure 5.9, the total wall-clock time of Q-GKB remains significantly lower than that of LIS across all tested subspace dimensions k , even when using MATLAB’s highly optimized `eigs` implementation. The dominant computational cost of LIS arises from the Cholesky factorization of Σ , which scales as $\mathcal{O}(n^3)$ and becomes prohibitively expensive at this resolution.

5.3 X-ray computed tomography

In the third example, we consider a two-dimensional X-ray computed tomography (CT) problem, where the goal is to reconstruct an unknown attenuation function $f : \Omega \rightarrow \mathbb{R}$ supported on a domain $\Omega \subset \mathbb{R}^2$ from a finite collection of line-integral measurements corrupted by noise. The mathematical model of X-ray CT is described by the Beer–Lambert law: when a monochromatic X-ray beam of initial intensity I_0 traverses a medium along a line ℓ , the transmitted intensity satisfies

$$I = I_0 \exp\left(-\int_{\ell} f(\mathbf{x}) ds\right), \quad (5.8)$$

so that the measured log-attenuation along ℓ equals the line integral of f . The corresponding measurement is thus $-\log(I/I_0)$. The forward operator can be described by the Radon transform

$$\mathcal{R}[f](\phi, s) = \int_{-\infty}^{\infty} f(s\boldsymbol{\theta} + t\boldsymbol{\theta}^{\perp}) dt, \quad \phi \in [0, \pi), \quad s \in \mathbb{R}, \quad (5.9)$$

where $\boldsymbol{\theta} = (\cos \phi, \sin \phi)$ is the unit vector in the projection direction and $\boldsymbol{\theta}^{\perp} = (-\sin \phi, \cos \phi)$ is its orthogonal complement. The pair (ϕ, s) parametrizes all oriented lines in \mathbb{R}^2 , where ϕ is the projection angle and s is the signed distance from the origin to the line. We refer to [3, 34] for further details.

The discrete linear system is constructed by discretizing (5.9) on a uniform $N \times N$ pixel grid of $\Omega = [-1, 1]^2$, yielding a vector $x \in \mathbb{R}^n$ with $n = N^2$, where the j th element x_j represents the attenuation coefficient in the j th pixel. Measurements are collected at n_{ϕ} equally spaced projection angles $\phi_i \in [0, \pi)$ and n_s the the number of rays per angle, giving $m = n_{\phi}n_s$ measurements assembled into a vector belonging to \mathbb{R}^m . The forward matrix $G \in \mathbb{R}^{m \times n}$ has entries $G_{ij} = \text{length}(\text{ray}_i \cap \text{pixel}_j)$, i.e., the intersection length of the i th ray with the j th pixel, computed via the standard ray-driven projection

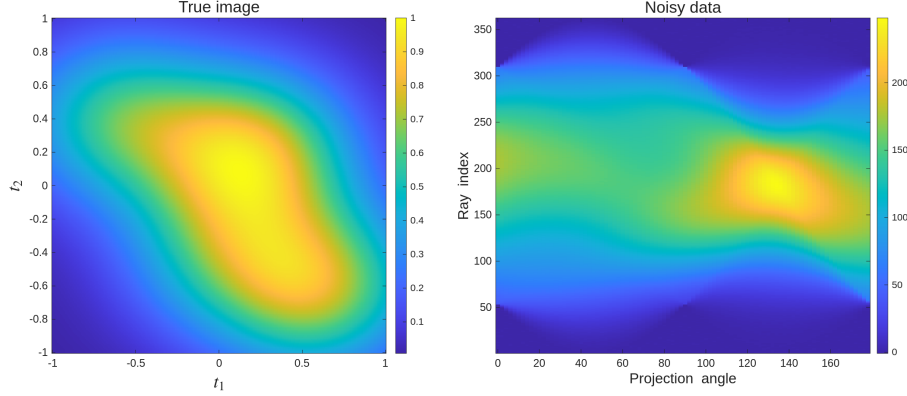


Figure 5.10: Illustration of the true image and noisy observation data.

model [34]. We set $N = 256$, and choose $\{\phi_i\}_{i=1}^{n_\phi} = \{0^\circ, 2^\circ, 4^\circ, \dots, 178^\circ\}$ with $n_s = \text{round}(\sqrt{2}N)$, where $\text{round}(a)$ denotes the nearest integer to a . With this choice, we have $G \in \mathbb{R}^{32580 \times 65536}$. The ground-truth x_{true} is taken from [15], and we generate noisy data y by adding a white Gaussian noise to Gx_{true} with noise level being 0.002. The true image and the corresponding noisy observation are shown in Figure 5.10.

To reconstruct the true image, we place a Gaussian prior on the unknown f with covariance induced by the Matérn kernel

$$K(\mathbf{x}, \mathbf{y}) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} \|\mathbf{x} - \mathbf{y}\|_2}{\rho} \right)^\nu B_\nu \left(\frac{\sqrt{2\nu} \|\mathbf{x} - \mathbf{y}\|_2}{\rho} \right), \quad (5.10)$$

and we set $\nu = 5/2$, $\rho = 2$ and $\sigma = 1$ to construct the matrix Σ . The value of σ is iteratively updated by the QGKB-HB method via the relation $\sigma^2 = 1/\lambda$. To ensure computational tractability for the $256^2 \times 256^2$ covariance matrix Σ , we implement the matrix-vector multiplication Σv in a matrix-free way. Given the stationarity of the Matérn kernel, the covariance operator on a uniform grid can be computed as a circular convolution via the Fast Fourier Transform (FFT). To mitigate periodic wrap-around artifacts, we employ a circulant embedding strategy: the input $N \times N$ image is first augmented to size $2N \times 2N$ through zero-padding; then the covariance action is performed as point-wise multiplication in the spectral domain; the final result is recovered by extracting the $N \times N$ block corresponding to its original domain. This approach reduces the computational complexity of Σv from $O(N^4)$ to $O(N^2 \log N)$, enabling efficient reconstruction on high-resolution grids without explicitly forming the dense covariance matrix. For further implementation details, see, e.g., [48, 35, 30]

In this experiment, we assess the performance of the proposed Q-GKB posterior approximation in a genuinely large-scale setting. In particular, we emphasize its computational scalability, stability, and its ability to accurately capture both the posterior mean and uncertainty. In Figure 5.11, we illustrate the evolution of the approximate posterior mean and covariance as the Q-GKB iteration proceeds from $k = 1$ to $k = 300$. Representative reconstructions and variance fields (i.e., pixel-wise variance) are shown at $k = 40, 120, 200$, and 280 , with a consistent color scale as in Figure 5.12. The corresponding mstd values are also computed following (5.7). From these results, we observe that the posterior mean converges rapidly: already at $k = 40$, the reconstruction achieves high visual quality. In contrast, the posterior covariance requires a larger subspace dimension to stabilize. For small k , the covariance estimate remains relatively large, indicating that the dominant data-informed directions have not yet been fully captured. As k increases, the covariance approximation gradually stabilizes, and the mstd decreases accordingly, indicating that the essential directions of the data-informed subspace are progressively captured. At $k = 280$, both the posterior mean and covariance have essentially converged. This behavior is consistent with the theoretical characterization of data-informed Krylov subspaces and their role in posterior approximation.

The prior image, the reconstructed image, and the corresponding variance field are shown in Figure 5.12, where the variance level is at the order of 10^{-6} , indicating very low uncertainty in the recon-

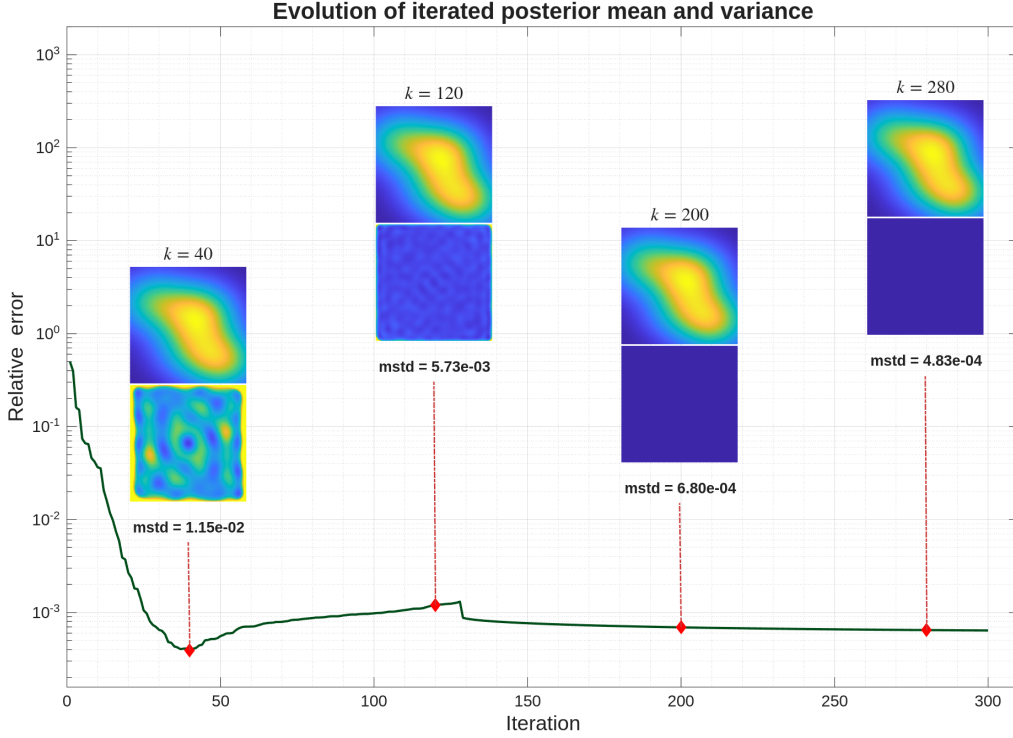


Figure 5.11: Evolution of the approximate posterior mean and variance during the Q-GKB iteration. The top and bottom images show the approximate mean and the corresponding pixel-wise variance, respectively.

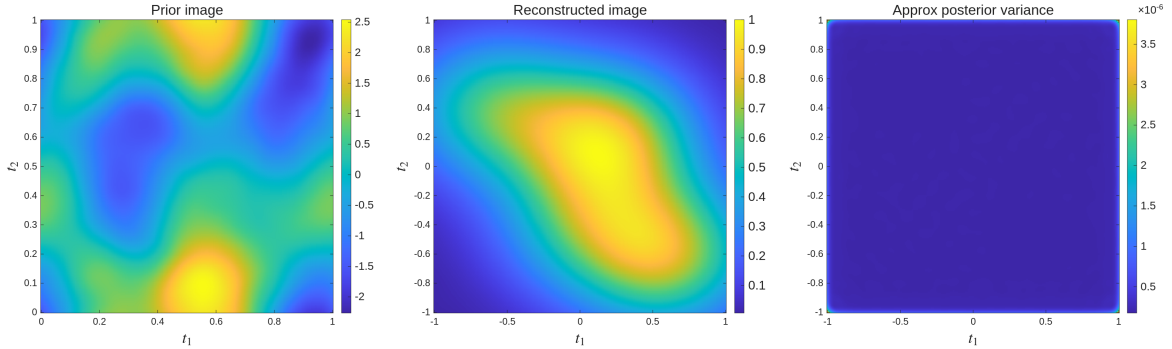


Figure 5.12: Illustration of the prior image, reconstructed image, and corresponding pixel-wise variance at $k = 300$.

struction. Importantly, for this large-scale CT problem, computing the exact posterior covariance, or even its diagonal, is computationally prohibitive under the FFT-based Matérn prior. In contrast, the proposed Q-GKB framework enables efficient matrix-free computation of both the posterior mean and covariance approximation. These results demonstrate that the method remains numerically stable, scalable, and capable of capturing meaningful posterior uncertainty in high-dimensional settings.

6 Conclusion

We have developed a data-informed framework for posterior approximation in large-scale Bayesian linear inverse problems. By shifting the perspective from the high-dimensional parameter space to the

data space, we establish an isometric embedding from the data space to the parameter space, revealing an intrinsic low-dimensional geometric structure of the posterior update induced by the data. This structure naturally leads to a quotient-space formulation of the prior-to-posterior update, in which the informative directions are determined independently of the ambient parameter dimension. Building on this geometric insight, we propose a quotient-space Golub–Kahan bidiagonalization method to construct data-informed Krylov subspaces. By integrating empirical Bayesian inference into the iterative framework, the method enables simultaneous hyperparameter estimation and posterior approximation in a matrix-free manner. Numerical experiments on problems of increasing scale validate the proposed framework and support the theoretical findings.

The data-space geometric perspective offers a foundation for understanding low-dimensional structures in high-dimensional Bayesian inverse problems and provides a new path for posterior inference. It also opens several directions for future research. In particular, extending the quotient data space framework to inverse problems with infinite-dimensional parameter spaces and to nonlinear inverse problems, as well as integrating it with more general prior models and sampling-based inference methods, presents valuable directions.

A Proofs of technical lemmas

Proof of Lemma 4.1. Define $Q_k = \Sigma^{1/2}G^\top V_k$, $T_k = B_k^\top B_k$ and $P_k = Q_k Q_k^\top$. Then the columns of Q_k are orthonormal and P_k is the orthogonal projector onto $\mathcal{R}(Q_k)$, and we have

$$\begin{aligned} AQ_k &= \Sigma^{1/2}G^\top \Gamma^{-1}MV_k = \Sigma^{1/2}G^\top \Gamma^{-1}U_{k+1}B_k \\ &= \Sigma^{1/2}G^\top (V_k B_k^\top + \alpha_{k+1}v_{k+1}e_{k+1}^\top)B_k \\ &= Q_k T_k + \alpha_{k+1}\beta_{k+1}q_{k+1}e_k^\top, \end{aligned} \quad (\text{A.1})$$

where $q_i = \Sigma^{1/2}G^\top v_i$. This implies that $Q_k^\top AQ_k = T_k$, and hence

$$\widehat{A}_k = \Sigma^{1/2}\widehat{H}_k\Sigma^{1/2} = \Sigma^{1/2}G^\top V_k B_k^\top B_k V_k^\top G\Sigma^{1/2} = Q_k T_k Q_k^\top = P_k A P_k. \quad (\text{A.2})$$

Since \widehat{A}_k and T_k share the same nonzero eigenvalues, the Cauchy interlacing theorem implies that $\text{Tr}(A) \geq \text{Tr}(T_k) = \text{Tr}(\widehat{A}_k)$, leading to $\zeta_k > 0$. Using

$$T_{k+1} = B_{k+1}^\top B_{k+1} = \begin{pmatrix} T_k & \alpha_{k+1}\beta_{k+1}e_k \\ \alpha_{k+1}\beta_{k+1}e_k^\top & \alpha_{k+1}^2 + \beta_{k+2}^2 \end{pmatrix},$$

a direct computation gives

$$\widehat{A}_{k+1} - \widehat{A}_k = \alpha_{k+1}\beta_{k+1}(q_k q_{k+1}^\top + q_{k+1} q_k^\top) + (\alpha_{k+1}^2 + \beta_{k+2}^2)q_{k+1} q_{k+1}^\top.$$

Therefore, we have

$$\text{Tr}(\widehat{A}_{k+1} - \widehat{A}_k) = \alpha_{k+1}^2 + \beta_{k+2}^2, \quad \|\widehat{A}_{k+1} - \widehat{A}_k\|_F^2 = 2\alpha_{k+1}^2\beta_{k+1}^2 + (\alpha_{k+1}^2 + \beta_{k+2}^2)^2.$$

This leads to that

$$\zeta_{k+1} = \text{Tr}(A - \widehat{A}_{k+1}) = \zeta_k - \text{Tr}(\widehat{A}_{k+1} - \widehat{A}_k) = \zeta_k - (\alpha_{k+1}^2 + \beta_{k+2}^2),$$

where $\zeta_0 = \text{Tr}(A) = \text{Tr}(H\Sigma)$.

For the recurrence of γ_k , notice that

$$\gamma_k^2 = \|A - \widehat{A}_k\|_F^2 = \|(I - P_k)A + P_k A(I - P_k)\|_F^2 = \|(I - P_k)A\|_F^2 + \|P_k A(I - P_k)\|_F^2,$$

since the two terms are orthogonal in the trace inner product. Using (A.1), we get

$$P_k A(I - P_k) = Q_k Q_k^\top A(I - P_k) = \alpha_{k+1}\beta_{k+1}Q_k e_k q_{k+1}^\top = \alpha_{k+1}\beta_{k+1}q_k q_{k+1}^\top,$$

hence $\|P_k A(I - P_k)\|_F^2 = \alpha_{k+1}^2 \beta_{k+1}^2$. Using $I - P_k = (I - P_{k+1}) + q_{k+1} q_{k+1}^\top$ and notice that the two terms are orthogonal in the trace inner product, we have

$$\|(I - P_k)A\|_F^2 = \|(I - P_{k+1})A\|_F^2 + \|q_{k+1} q_{k+1}^\top A\|_F^2 = \|(I - P_{k+1})A\|_F^2 + \|Aq_{k+1}\|_2^2.$$

From the Q-GKB recurrence, we have

$$\Gamma^{-1} M v_{k+1} = \alpha_{k+1} \beta_{k+1} v_k + (\alpha_{k+1}^2 + \beta_{k+2}^2) v_{k+1} + \alpha_{k+2} \beta_{k+2} v_{k+2},$$

multiplying by $\Sigma^{1/2} G^\top$ gives

$$Aq_{k+1} = \alpha_{k+1} \beta_{k+1} q_k + (\alpha_{k+1}^2 + \beta_{k+2}^2) q_{k+1} + \alpha_{k+2} \beta_{k+2} q_{k+2}.$$

This leads to $\|Aq_{k+1}\|_2^2 = \alpha_{k+1}^2 \beta_{k+1}^2 + (\alpha_{k+1}^2 + \beta_{k+2}^2)^2 + \alpha_{k+2}^2 \beta_{k+2}^2$. Therefore, we obtain

$$\begin{aligned} \gamma_k^2 - \gamma_{k+1}^2 &= \|(I - P_k)A\|_F^2 - \|(I - P_{k+1})A\|_F^2 + \alpha_{k+1}^2 \beta_{k+1}^2 - \alpha_{k+2}^2 \beta_{k+2}^2 \\ &= \|Aq_{k+1}\|_2^2 + \alpha_{k+1}^2 \beta_{k+1}^2 - \alpha_{k+2}^2 \beta_{k+2}^2 \\ &= (\alpha_{k+1}^2 + \beta_{k+2}^2)^2 + 2\alpha_{k+1}^2 \beta_{k+1}^2. \end{aligned}$$

A simple calculation leads to $\gamma_0^2 = \text{Tr}(A^2) = \text{Tr}(H\Sigma H\Sigma)$. This completes the proof. \square

Proof of Lemma 4.2. Let $\delta = \|A_1 - A_2\|_2$. Then $-\delta I \preceq A_1 - A_2 \preceq \delta I$, or equivalently,

$$A_2 - \delta I \preceq A_1 \preceq A_2 + \delta I, \quad A_1 - \delta I \preceq A_2 \preceq A_1 + \delta I.$$

Let $R_i = (I + A_i)^{-1}$ for $i = 1, 2$. Since $f(t) = (1 + t)^{-1}$ is operator monotone decreasing on $[0, \infty)$, using $A_1 \preceq A_2 + \delta I$ and $A_2 \preceq A_1 + \delta I$ we obtain

$$R_1 \succeq (I + A_2 + \delta I)^{-1}, \quad R_2 \succeq (I + A_1 + \delta I)^{-1}.$$

Hence $R_1 - R_2 \preceq R_1 - (I + A_1 + \delta I)^{-1}$. Noticing that

$$R_1 - (I + A_1 + \delta I)^{-1} = \delta(I + A_1)^{-1}(I + A_1 + \delta I)^{-1},$$

and using $0 \prec (I + A_1)^{-1} \preceq I$ and $0 \prec (I + A_1 + \delta I)^{-1} \preceq \frac{\delta}{1 + \delta} I$, we get $R_1 - R_2 \preceq \frac{\delta}{1 + \delta} I$. By symmetry, exchanging A_1 and A_2 yields $R_2 - R_1 \preceq \frac{\delta}{1 + \delta} I$. Therefore, we get

$$-\frac{\delta}{1 + \delta} I \preceq R_1 - R_2 \preceq \frac{\delta}{1 + \delta} I \quad \Rightarrow \quad \|R_1 - R_2\|_2 \leq \frac{\delta}{1 + \delta}.$$

This completes the proof. \square

Acknowledgments

The author thanks Prof. Fei Lu for many insightful discussions.

References

- [1] Tan Bui-Thanh, Carsten Burstedde, Omar Ghattas, James Martin, Georg Stadler, and Lucas C Wilcox. Extreme-scale UQ for Bayesian inverse problems governed by PDEs. In *SC'12: Proceedings of the international conference on high performance computing, networking, storage and analysis*, pages 1–11. IEEE, 2012.
- [2] Tan Bui-Thanh, Omar Ghattas, James Martin, and Georg Stadler. A computational framework for infinite-dimensional Bayesian inverse problems part I: The linearized case, with application to global seismic inversion. *SIAM Journal on Scientific Computing*, 35(6):A2494–A2523, 2013.

- [3] Thorsten Buzug. *Computed Tomography*. Springer, 2008.
- [4] Bradley P Carlin and Thomas A Louis. *Bayes and empirical Bayes methods for data analysis*. London: Chapman and Hall, 1997.
- [5] Noe Angelo Caruso and Paolo Novati. Convergence analysis of LSQR for compact operator equations. *Linear Algebra and its Applications*, 583:146–164, 2019.
- [6] J Chung and A K Saibaba. Generalized hybrid iterative methods for large-scale Bayesian inverse problems. *SIAM Journal on Scientific Computing*, 39(5):S24–S46, 2017.
- [7] Julianne Chung and Silvia Gazzola. Computational methods for large-scale inverse problems: a survey on hybrid projection methods. *SIAM Review*, 66(2):205–284, 2024.
- [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] Tiangang Cui, James Martin, Youssef M Marzouk, Antti Solonen, and Alessio Spantini. Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems*, 30(11):114015, 2014.
- [10] Masoumeh Dashti, Kody JH Law, Andrew M Stuart, and Jochen Voss. Map estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems*, 29(9):095017, 2013.
- [11] H. W. Engl, M Hanke, and A Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, 2000.
- [12] H Pearl Flath, Lucas C Wilcox, Volkan Akçelik, Judith Hill, Bart van Bloemen Waanders, and Omar Ghattas. Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial hessian approximations. *SIAM Journal on Scientific Computing*, 33(1):407–432, 2011.
- [13] Wolfgang Förstner and Boudewijn Moonen. A metric for covariance matrices. In *Geodesy—the Challenge of the 3rd Millennium*, pages 299–309. Springer, 2003.
- [14] Joel N Franklin. Well-posed stochastic extensions of ill-posed linear problems. *Journal of mathematical analysis and applications*, 31(3):682–716, 1970.
- [15] 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.
- [16] Per Christian Hansen. *Discrete inverse problems: insight and algorithms*. SIAM, 2010.
- [17] Per Christian Hansen, J G Nagy, and D P O’Leary. *Deblurring Images: Matrices, Spectra and Filtering*. SIAM, Philadelphia, 2006.
- [18] Tapio Helin and Martin Burger. Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Problems*, 31(8):085009, 2015.
- [19] Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, 2012.
- [20] Jari Kaipio and Erkki Somersalo. *Statistical and Computational Inverse Problems*. Springer, 2006.
- [21] Kari Karhunen. Über lineare methoden in der wahrscheinlichkeitsrechnung. *Ann Acad Sci Fennicae*, 37:1, 1947.
- [22] Misha E Kilmer and Dianne P O’Leary. Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM Journal on Matrix Analysis and Applications*, 22(4):1204–1221, 2001.
- [23] Konstantinos Zygalakis Kody Law, Andrew Stuart. *Data Assimilation: A Mathematical Introduction*. Springer, 2015.
- [24] Markku S Lehtinen, Lassi Paivarinta, and Erkki Somersalo. Linear inverse problems for generalised random variables. *Inverse problems*, 5(4):599–612, 1989.
- [25] Haibo Li. A preconditioned Krylov subspace method for linear inverse problems with general-form Tikhonov regularization. *SIAM Journal on Scientific Computing*, 46(4):A2607–A2633, 2024.

- [26] Haibo Li. Projected Newton method for large-scale Bayesian linear inverse problems. *SIAM Journal on Optimization*, 35(3):1439–1468, 2025.
- [27] Wei Li and Olaf A Cirpka. Efficient geostatistical inverse methods for structured and unstructured grids. *Water resources research*, 42(6), 2006.
- [28] Chad Lieberman, Karen Willcox, and Omar Ghattas. Parameter and state model reduction for large-scale statistical inverse problems. *SIAM Journal on Scientific Computing*, 32(5):2523–2542, 2010.
- [29] Jörg Liesen and Zdenek Strakos. *Krylov subspace methods: principles and analysis*. Numerical Mathematics and Scie, 2013.
- [30] Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4):423–498, 2011.
- [31] M. Loève. *Probability Theory II*. Springer New York, NY, 1978.
- [32] James Martin, Lucas C Wilcox, Carsten Burstedde, and Omar Ghattas. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing*, 34(3):A1460–A1487, 2012.
- [33] Youssef M Marzouk and Habib N Najm. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6):1862–1902, 2009.
- [34] Frank Natterer. *The Mathematics of Computerized Tomography*. SIAM, 2001.
- [35] Wolfgang Nowak, Sascha Tenkleve, and Olaf A Cirpka. Efficient computation of linearized cross-covariance and auto-covariance matrices of interdependent quantities. *Mathematical geology*, 35(1):53–66, 2003.
- [36] Sonia Petrone, Judith Rousseau, and Catia Scricciolo. Bayes and empirical Bayes: do they merge? *Biometrika*, pages 285–302, 2014.
- [37] 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 Journal on Scientific Computing*, 39(2):B221–B243, 2017.
- [38] Mathias Richter. *Inverse Problems: Basics, Theory and Applications in Geophysics*. Springer, 2016.
- [39] Lassi Roininen, Janne MJ Huttunen, and Sari Lasanen. Whittle-Matérn priors for bayesian statistical inversion with applications in electrical impedance tomography. *Inverse Probl. Imag.*, 8(2), 2014.
- [40] Judith Rousseau and Botond Szabo. Asymptotic behaviour of the empirical Bayes posteriors associated to maximum marginal likelihood estimator. *Annals of Statistics*, 2017.
- [41] Yousef Saad. *Numerical methods for large eigenvalue problems: revised edition*. SIAM, 2011.
- [42] Arvind K Saibaba, Julianne Chung, and Katrina Petroske. Efficient Krylov subspace methods for uncertainty quantification in large Bayesian linear inverse problems. *Numer. Linear Algebra Appl.*, 27(5):e2325, 2020.
- [43] Alessio Spantini, Antti Solonen, Tiangang Cui, James Martin, Luis Tenorio, and Youssef Marzouk. Optimal low-rank approximations of Bayesian linear inverse problems. *SIAM Journal on Scientific Computing*, 37(6):A2451–A2487, 2015.
- [44] Alessio Spantini, Tiangang Cui, Karen Willcox, Luis Tenorio, and Youssef Marzouk. Goal-oriented optimal approximations of bayesian linear inverse problems. *SIAM Journal on Scientific Computing*, 39(5):S167–S196, 2017.
- [45] Andrew M Stuart. Inverse problems: a Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.
- [46] A N Tikhonov and V Y Arsenin. *Solutions of Ill-Posed Problems*. Washington, DC, 1977.

- [47] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006.
- [48] Andrew TA Wood and Grace Chan. Simulation of stationary Gaussian processes in $[0, 1]^d$. *Journal of computational and graphical statistics*, 3(4):409–432, 1994.