SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 12, No. 1, pp. 30–68

Analysis of a Computational Framework for Bayesian Inverse Problems: Ensemble Kalman Updates and MAP Estimators under Mesh Refinement*

Daniel Sanz-Alonso† and Nathan Waniorek‡

Abstract. This paper analyzes a popular computational framework for solving infinite-dimensional Bayesian inverse problems, discretizing the prior and the forward model in a finite-dimensional weighted inner product space. We demonstrate the benefit of working on a weighted space by establishing operator-norm bounds for finite element and graph-based discretizations of Matérn-type priors and deconvolution forward models. For linear-Gaussian inverse problems, we develop a general theory for characterizing the error in the approximation to the posterior. We also embed the computational framework into ensemble Kalman methods and maximum a posteriori (MAP) estimators for nonlinear inverse problems. Our operator-norm bounds for prior discretizations guarantee the scalability and accuracy of these algorithms under mesh refinement.

Key words. Bayesian inverse problem, discretization error, ensemble Kalman methods, MAP estimation

MSC codes. 65M32, 62F15, 68Q25, 35Q62

DOI. 10.1137/23M1567035

1. Introduction.

Bayesian inverse problems on infinite-dimensional Hilbert spaces arise in numerous applications, including medical imaging, seismology, climate science, reservoir modeling, and mechanical engineering [6, 10, 12, 15, 30]. In these and other applications, it is important to reconstruct function input parameters of partial differential equations (PDEs) based on noisy measurement of the PDE solution. This paper analyzes a framework for numerically solving infinite-dimensional Bayesian inverse problems, where the discretization is carried out in a weighted inner product space. The framework, along with a compelling demonstration of its computational benefits, was introduced in [10] for solving PDE-constrained Bayesian inverse problems using finite element discretizations. We develop a rigorous analysis that explains the advantage of working on a weighted space. Our theory accommodates not only finite element discretizations but also graph-based methods from machine learning. For linear-Gaussian inverse problems, we bound the error in the approximation to the posterior. More broadly, we embed the discretization framework into ensemble Kalman methods and *maximum a posteriori* (MAP) estimators for nonlinear inverse problems and study the accuracy and scalability of these algorithms under mesh refinement.

*Received by the editors April 20, 2023; accepted for publication (in revised form) November 13, 2023; published electronically February 2, 2024.
https://doi.org/10.1137/23M1567035
Funding: The work of the first author was supported by NSF DMS-2027056, NSF DMS-2237628, DOE DE-SC0022232, and the BBVA Foundation.
†Department of Statistics, University of Chicago, Chicago, IL 60637 USA (sanzalonso@uchicago.edu).
‡Committee on Computational and Applied Mathematics, University of Chicago, Chicago, IL 60637 USA (waniorek@uchicago.edu).

An overarching theme of this paper is that, by suitably choosing the weighted inner product, one can ensure accurate approximation of the infinite-dimensional Hilbert inner product as the discretization mesh is refined. Following this unifying principle, we analyze the finite element discretizations considered in [10] together with graph-based discretizations. Finite elements are predominant in scientific computing solutions of PDE-constrained inverse problems (see, e.g., [6, 10, 33]), whereas graph-based methods have been used, for instance, to solve inverse problems on manifolds [27, 28, 32] and in machine learning applications in semisupervised learning [4, 25]. For both types of discretization, we obtain operator-norm bounds for important classes of prior and forward models. Specifically, we will use Matérn-type Gaussian priors and deconvolution forward models as guiding examples.

Matérn priors play a central role in Bayesian inverse problems [49], spatial statistics [53], and machine learning [47]. Efficient algorithms for sampling Matérn priors within the computational framework of [10] have recently been investigated in [2]. Here, we obtain new operator-norm bounds for both finite element and graph-based discretizations. The former relies on classical theory, while the latter extends recent results from [51]. Deconvolution forward models arise in image deblurring and heat inversion, among other important applications. Since the seminal paper by Franklin that introduced the formulation of Bayesian inversion in function space [21], heat inversion has been widely adopted as a tractable testbed for theoretical and methodological developments; see, e.g., [25, 23, 27, 55]. Here, we derive operator-norm bounds for both finite element and graph-based discretizations.

For linear-Gaussian inverse problems, operator-norm error bounds for prior and forward model discretizations translate into error bounds in the approximation of the posterior. We formalize this claim under a general assumption, which we verify for our guiding examples of Matérn priors and deconvolution. Additionally, we apply the computational framework in [10] to algorithms for nonlinear inverse problems beyond the Markov chain Monte Carlo method for posterior sampling considered in [46]. Specifically, we investigate (i) ensemble Kalman methods [14, 31], where we show, building on [26], that the effective dimension which determines the required sample size remains bounded along mesh refinements; and (ii) MAP estimation [19, 33], where we show, building on [3], the convergence of MAP estimators under mesh refinement to the MAP estimator of the infinite-dimensional inverse problem. These results complement the vast literature on function-space sampling algorithms (see e.g., [1, 16, 23, 45]) and demonstrate that ensemble Kalman methods and MAP estimation can be scalable and accurate under mesh refinement. We exemplify the new theory for MAP estimation in the nonlinear inverse problem of recovering the initial condition of the Navier–Stokes equations from pointwise observations of the velocity field [18, 17, 43].

The well-posedness of the posterior measure under perturbations is one of the hallmarks of the Bayesian formulation of inverse problems [39, 40, 50, 55]. For a fixed prior measure, the error in the posterior measure caused by discretization of the forward model can be bounded in Hellinger distance [18] and Kullback–Leibler divergence [41]. Posterior stability under perturbations to the prior measure, in addition to perturbations to the likelihood, have been recently investigated using Wasserstein distance [52] and more general integral probability metrics [22]. These results hold even when the prior and the perturbation are mutually singular, as is the case for a discretization of the prior measure. In the linear-Gaussian setting, the Wasserstein distance bounds in [52] yield a stability theory similar to ours; however, bounding the Wasserstein distance between the prior measures requires trace bounds on the covariance operators, whereas our results only necessitate operator-norm bounds.

1.1. Outline and main contributions.

- Section 2 reviews the formulation of a Bayesian inverse problem in a Hilbert space and presents our general discretization framework. We introduce the heat inversion problem with Matérn Gaussian process prior as a model guiding example, and then illustrate how finite element and graph-based methods can be interpreted within our discretization framework.
- Section 3 presents a novel analysis of the error incurred in our discretization framework for linear-Gaussian Bayesian inverse problems. Theorems 3.3 and 3.4 quantify the errors in the discretized posterior mean and covariance operators to their continuum counterparts, up to universal constants. We then verify the assumptions of this general theory for the finite element and graph-based discretizations of the heat inversion problem and thus derive error bounds for these popular discretization schemes.
- Section 4 formulates the ensemble Kalman update within our general discretization framework. The results in this section give nonasymptotic bounds on the ensemble estimation of the posterior mean and covariance in terms of a notion of effective dimension based on spectral decay of the prior covariance operator. We then show that the effective dimension of the discretized prior covariance can be controlled by the effective dimension of the continuum covariance, which is necessarily finite. Consequently, the ensemble approximation will not deteriorate under mesh refinement.
- Section 5 considers posterior measures resulting from nonlinear Bayesian inverse problems and their MAP estimators. We show that with a suitable discretization of the forward model, the MAP estimators of the computationally tractable discretized posterior measures converge to the MAP estimators of the continuum posterior. Finally, we apply the theory to the inverse problem of recovering the initial condition of the Navier–Stokes equations from pointwise observations.
- Section 6 closes the paper with conclusions and directions for future work.

1.2. Notation. $\mathcal{S}^d_+$ denotes the set of $d \times d$ symmetric positive-semidefinite matrices, and $\mathcal{S}^d_{++}$ denotes the set of $d \times d$ symmetric positive-definite matrices. Similarly, $\mathcal{S}^{\mathcal{H}}_+$ denotes the set of symmetric positive-semidefinite trace-class operators from a Hilbert space $\mathcal{H}$ to itself, and $\mathcal{S}^{\mathcal{H}}_{++}$ denotes the set of symmetric positive-definite trace-class operators from $\mathcal{H}$ to itself. Given two normed spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, and a linear mapping $A : X \to Y$, we denote the operator norm of A as $\|A\|_{op} := \sup_{\|x\|_X=1} \|Ax\|_Y$. $B(X,Y)$ denotes the space of all bounded linear operators from X to Y. Given two positive sequences $\{a_n\}$ and $\{b_n\}$, the relation $a_n \lesssim b_n$ denotes that $a_n \le cb_n$ for some constant $c > 0$. $\mathbb{1}_B$ denotes the indicator of the set B. Given a matrix $A \in \mathcal{S}^d_{++}$, we denote the weighted inner product as $\langle \vec{u}, \vec{v}\rangle_A := \vec{u}^T A \vec{v}$, and the corresponding weighted norm is denoted as $\|\vec{u}\|_A := \sqrt{\vec{u}^T A \vec{u}}$. Finally, $\|\cdot\|_2$ denotes the usual Euclidean norm on $\mathbb{R}^d$.

2. Problem setting and computational framework. This section contains necessary background. Subsection 2.1 overviews the formulation of Bayesian inverse problems in Hilbert space. Subsection 2.2 describes the computational framework analyzed in this paper. Finally, subsection 2.3 shows how finite element and graph-based methods for inverse problems in function space can be viewed as particular instances of the general framework.

2.1. Inverse problem in Hilbert space.

Let $\mathcal{H}$ be an infinite-dimensional separable Hilbert space with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$. Consider the inverse problem of recovering an unknown $u\in\mathcal{H}$ from data $y\in\mathbb{R}^{d_y}$ related by

$$y=\mathcal{F}u+\eta, \quad \eta\sim\mathcal{N}(0,\Gamma), \tag{2.1}$$

where $\mathcal{F}:\mathcal{H}\to\mathbb{R}^{d_y}$ is a linear and bounded forward model, and η represents Gaussian observation noise with known covariance matrix $\Gamma\in S^{d_y}_{++}$. Nonlinear forward models will be considered in section 5. The observation model (2.1) implies a Gaussian likelihood function

$$\pi_{\text{like}}(y|u)\propto\exp\left(-\frac{1}{2}\|y-\mathcal{F}u\|^2_{\Gamma^{-1}}\right). \tag{2.2}$$

We adopt a Bayesian approach [33, 50, 57] and let $\mu_0=\mathcal{N}(m_0,\mathcal{C}_0)$ be a Gaussian prior measure on $\mathcal{H}$, where $\mathcal{C}_0:\mathcal{H}\to\mathcal{H}$ is a trace-class covariance operator defined by the requirement that

$$\langle v,\mathcal{C}_0w\rangle_{\mathcal{H}}=\mathbb{E}^{u\sim\mu_0}\Big[\langle v,(u-m_0)\rangle_{\mathcal{H}}\langle(u-m_0),w\rangle_{\mathcal{H}}\Big] \qquad \forall v,w\in\mathcal{H}, \tag{2.3}$$

and $m_0\in\mathcal{H}$ is assumed to belong to the Cameron–Martin space $E=\operatorname{Im}(\mathcal{C}_0^{1/2})\subset\mathcal{H}$. The Bayesian solution to the inverse problem is the conditional law of u given y, which is called the posterior probability measure μ_{post}. Application of Bayes' rule in infinite dimensions [55] characterizes the posterior as a change of measure with respect to the prior, with Radon–Nikodym derivative given by the likelihood,

$$\frac{d\mu_{\text{post}}}{d\mu_0}(u)=\frac{1}{Z}\pi_{\text{like}}(y|u), \tag{2.4}$$

where $Z=\int\pi_{\text{like}}(y|u)\,d\mu_0$ is a normalizing constant. Since the prior is Gaussian, and the forward model is linear and bounded, the posterior is also Gaussian, $\mu_{\text{post}}=\mathcal{N}(m_{\text{post}},\mathcal{C}_{\text{post}})$, where

$$m_{\text{post}}=m_0+\mathcal{C}_0\mathcal{F}^*(\mathcal{F}\mathcal{C}_0\mathcal{F}^*+\Gamma)^{-1}(y-\mathcal{F}m_0), \tag{2.5}$$

$$\mathcal{C}_{\text{post}}=\mathcal{C}_0-\mathcal{C}_0\mathcal{F}^*(\mathcal{F}\mathcal{C}_0\mathcal{F}^*+\Gamma)^{-1}\mathcal{F}\mathcal{C}_0. \tag{2.6}$$

Here, $\mathcal{F}^*$ denotes the adjoint of $\mathcal{F}$, which is the unique map from $\mathbb{R}^{d_y}$ to $\mathcal{H}$ that satisfies

$$\langle\mathcal{F}u,y\rangle=\langle u,\mathcal{F}^*y\rangle_{\mathcal{H}} \qquad \forall u\in\mathcal{H},y\in\mathbb{R}^{d_y}.$$

In subsection 2.2, we will review the computational framework in [10] to approximate the posterior. The idea is to replace the inverse problem (2.1) on the infinite-dimensional Hilbert space $\mathcal{H}$ with an inverse problem on a finite-dimensional weighted inner product space. Under general conditions on the discretization of the forward model and prior measure, the posterior mean and covariance of the finite- and infinite-dimensional inverse problems are close together; this claim will be formalized and rigorously established in section 3.

2.2. Computational framework. This subsection gives an overview of the computational framework presented in [10]. Let $\{\phi_1, \phi_2, \ldots, \phi_n\}$ be a basis for an n-dimensional space $\mathcal{V} \subset \mathcal{H}$. For $u = \sum_{i=1}^n u_i \phi_i \in \mathcal{V}$, we denote by $\vec{u} = (u_1, \ldots, u_n)^T$ the vector of coefficients of u in this basis. We will endow the space of coefficients with a weighted inner product that is naturally inherited from the inner product in $\mathcal{H}$. Specifically, for any $u, v \in \mathcal{V}$, we have

$$\langle u, v\rangle_{\mathcal{H}} = \vec{u}^T M \vec{v} = \langle \vec{u}, \vec{v}\rangle_M,$$

where the matrix $M = (M_{ij})_{i,j=1}^n$ is given by

$$M_{ij} = \langle \phi_i, \phi_j\rangle_{\mathcal{H}}, \quad i, j \in \{1, \ldots, n\}.$$

This observation motivates us to introduce the inner product space $\mathbb{R}^n_M$ defined by vector space $\mathbb{R}^n$ and inner product $\langle \cdot, \cdot\rangle_M$. We remark that for both the finite elements and graph-based discretizations introduced in subsections 2.3.1 and 2.3.2, the corresponding basis $\{\phi_i\}$ will not be orthonormal; consequently, $M \neq I$ and $\mathbb{R}^n_M$ will not agree with the standard Euclidean inner product space. We now aim to replace the inverse problem (2.1) with an inverse problem on the weighted space $\mathbb{R}^n_M$.

To begin, we define a discretization map $P : \mathcal{H} \to \mathbb{R}^n_M$ that assigns to $u \in \mathcal{H}$ the vector $\vec{u}^\star \in \mathbb{R}^n_M$ satisfying

$$Pu = \vec{u}^\star = (u_1^\star, \ldots, u_n^\star)^T = \arg\min_{u_1,\ldots,u_n} \frac{1}{2}\Big\| u - \sum_{i=1}^n u_i \phi_i \Big\|_{\mathcal{H}}. \tag{2.7}$$

In practice, this discretization can be computed by solving the linear system $M\vec{u}^\star = \vec{b}$, where $b_i = \langle u, \phi_i\rangle_{\mathcal{H}}$. Conversely, our theoretical analysis will rely on an extension map $P^* : \mathbb{R}^n_M \to \mathcal{V} \subset \mathcal{H}$ that assigns to a vector $\vec{u} \in \mathbb{R}^n_M$ the element $P^* u \in \mathcal{V}$ defined by

$$P^* \vec{u} = \sum_{i=1}^n u_i \phi_i. \tag{2.8}$$

One can verify that $\langle Pu, \vec{v}\rangle_M = \langle u, P^* \vec{v}\rangle_{\mathcal{H}}$, so that P^* is the adjoint of P, as our notation suggests. Similarly, one can verify that the map $P^*P : \mathcal{H} \to \mathcal{V} \subset \mathcal{H}$ is the orthogonal projection onto $\mathcal{V}$, so that $PP^* : \mathbb{R}^n_M \to \mathbb{R}^n_M$ is the identity map.

We are ready to introduce the inverse problem on the weighted inner product space $\mathbb{R}^n_M$. In analogy with (2.1), we seek to recover $\vec{u} \in \mathbb{R}^n_M$ from data y related by

$$y = F\vec{u} + \eta, \quad \eta \sim \mathcal{N}(0, \Gamma), \tag{2.9}$$

where $F : \mathbb{R}^n_M \to \mathbb{R}^{d_y}$ is a discretized forward model, identified with a matrix $F \in \mathbb{R}^{d_y \times n}$. (Here and henceforth we will abuse notation and identify linear maps and matrices without further notice.) The map F should approximate $\mathcal{F}$ in the sense that $\|\mathcal{F} - FP\|_{op}$ is small; we refer the reader to subsections 3.2.2 and 3.3.2 for examples arising from discretization of PDE-constrained forward models in function space. In analogy with (2.2), the likelihood of observed data y given a coefficient vector $\vec{u} \in \mathbb{R}^n_M$ is given by

$$\pi_{\text{like}}(y|\vec{u}) \propto \exp\left(-\frac{1}{2}\|y - F\vec{u}\|^2_{\Gamma^{-1}}\right).$$

Following the analogy with subsection 2.1, we choose a Gaussian prior $\mu_0^n = \mathcal{N}(\vec{m}_0, C_0)$ on the space $\mathbb{R}^n_M$, with Lebesgue density

$$\pi_0(\vec{u}) \propto \exp\left(-\frac{1}{2}\langle \vec{u} - \vec{m}_0, C_0^{-1}(\vec{u} - \vec{m}_0)\rangle_M\right). \tag{2.10}$$

For the prior mean, we take $\vec{m}_0 = Pm_0$; for the prior covariance, we may take an invertible operator $C_0 : \mathbb{R}^n_M \to \mathbb{R}^n_M$ that well approximates $\mathcal{C}_0$, in the sense that $\|\mathcal{C}_0 - P^* C_0 P\|_{op}$ is small; we refer the reader to subsections 3.2.1 and 3.3.1 for examples arising from discretization of Matérn-type Gaussian processes. Similar to (2.3), the prior covariance C_0 satisfies, by definition, that

$$\langle \vec{v}, C_0 \vec{w}\rangle_M = \mathbb{E}^{\vec{u}\sim\mu_0^n}\left[\langle \vec{v}, (\vec{u} - \vec{m})\rangle_M \langle (\vec{u} - \vec{m}), \vec{w}\rangle_M\right] \qquad \forall \vec{v}, \vec{w} \in \mathbb{R}^n_M.$$

While we view μ_0^n as a Gaussian measure on $\mathbb{R}^n_M$, it is helpful to note that, as a multivariate Gaussian distribution on $\mathbb{R}^n$ equipped with the standard inner product, we have that $\mu_0^n = \mathcal{N}(\vec{m}_0, C_0^E)$, with covariance $C_0^E = C_0 M^{-1}$.

Applying Bayes' rule as in (2.4), we can express the posterior as a change of measure with respect to the prior:

$$\frac{\mu^n_{\text{post}}}{\mu_0^n} = \frac{1}{Z_h}\pi_{\text{like}}(y|\vec{u}). \tag{2.11}$$

Since the space $\mathbb{R}^n_M$ is finite-dimensional, we recover the standard Bayes' formula for the posterior distribution with Lebesgue density given by

$$\pi^n_{\text{post}}(\vec{u}|y) \propto \exp\left(-\frac{1}{2}\|y - F\vec{u}\|^2_{\Gamma^{-1}} - \frac{1}{2}\langle \vec{u} - \vec{m}_0, C_0^{-1}(\vec{u} - \vec{m}_0)\rangle_M\right). \tag{2.12}$$

As in the infinite-dimensional setting, the posterior $\mu^n_{\text{post}} = \mathcal{N}(\vec{m}_{\text{post}}, C_{\text{post}})$ is Gaussian with an analogous expressions for its mean $\vec{m}_{\text{post}}$ and covariance C_{post}. A note of caution is that in the weighted space $\mathbb{R}^n_M$, a careful distinction must be made between the adjoint and the matrix transpose. We present a detailed discussion of this distinction in Appendix A. In particular, if the forward model is given by matrix $F \in \mathbb{R}^{d_y \times n}$, then the adjoint map is given by $F^\natural = M^{-1}F^T \in \mathbb{R}^{n \times d_y}$. Thus,

$$\vec{m}_{\text{post}} = \vec{m}_0 + C_0 F^\natural (F C_0 F^\natural + \Gamma)^{-1}(y - F\vec{m}_0), \tag{2.13}$$
$$C_{\text{post}} = C_0 - C_0 F^\natural (F C_0 F^\natural + \Gamma)^{-1} F C_0. \tag{2.14}$$

One can also view the posterior as a Gaussian measure in the standard Euclidean space, in which case the covariance operator is given by $C^E_{\text{post}} = C_{\text{post}} M^{-1}$. For completeness, we include the expressions for the posterior mean and covariance in Euclidean space written in terms of the Euclidean prior covariance as

$$\vec{m}_{\text{post}} = \vec{m}_0 + C_0^E F^T (F C_0^E F^T + \Gamma)^{-1}(y - F\vec{m}_0), \tag{2.15}$$
$$C^E_{\text{post}} = C_0^E - C_0^E F^T (F C_0^E F^T + \Gamma)^{-1} F C_0^E. \tag{2.16}$$

Table 1

Roadmap of the discretizations analyzed in this paper as particular examples of the general computational framework introduced in subsection 2.2.

Domain	Basis	Discretization	Error bounds
$\mathcal{D}$	Lagrange finite elements	subsection 2.3.1	subsection 3.2
$\mathcal{M}$	Random geometric graphs	subsection 2.3.2	subsection 3.3

Notice that both pairs of equations (2.13)–(2.14) and (2.15)–(2.16) are analogous to their infinite-dimensional counterparts in (2.5)–(2.6). However, our theory and examples in the next subsection will demonstrate the advantage of working on the weighted inner product space. Specifically, we will show that finite element and graph discretizations of important classes of prior covariance and the forward model give discretized quantities $F^\natural$ and C_0 that well approximate in operator norm their infinite-dimensional counterparts $\mathcal{F}^*$ and $\mathcal{C}_0$; the same would not be true for the Euclidean analogues F^T and C_0^E.

2.3. Inverse problems in function space. As particular instances of the computational framework reviewed in the previous subsection, here we consider discretizations of inverse problems in function space using finite elements and graphs. To illustrate the ideas, we focus on inverse problems with Matérn-type priors and deconvolution forward models, introduced in Examples 2.1 and 2.2 below. We let $\mathcal{H} = L^2(\Omega)$ be the Hilbert space of square integrable functions on Ω with the usual inner product $\langle \cdot, \cdot \rangle_{L^2(\Omega)}$. For finite element discretizations, we take $\Omega = \mathcal{D}$ to be a sufficiently regular domain $\mathcal{D} \subset \mathbb{R}^d$, while for graph discretizations we take $\Omega = \mathcal{M}$ to be a d-dimensional smooth, connected, compact Riemannian manifold embedded in $\mathbb{R}^D$. These choices of domain Ω reflect popular settings for finite element and graph-based discretizations in Bayesian inverse problems.

We next introduce our examples of prior and forward models, followed by a discussion of their finite element and graph discretizations, both of which will be analyzed in a unified way in section 3.

Example 2.1 (Matérn-type prior). We will consider Gaussian priors $\mu_0 = \mathcal{N}(m_0, \mathcal{C}_0)$ with Matérn covariance operator $\mathcal{C}_0$ and mean m_0 in the Cameron–Martin space $Im(\mathcal{C}_0^{1/2})$. Specifically, for positive integer α, we define $\mathcal{C}_0 = \mathcal{A}^{-\alpha}$, where $\mathcal{A}^{-1} : L^2(\Omega) \to H^2(\Omega)$ is the solution operator for the following elliptic PDE in weak form:

$$\int_\Omega \Theta \nabla u \cdot \nabla p + \int_\Omega bup = \int_\Omega fp \qquad \forall p \in H^1(\Omega). \tag{2.17}$$

That is, for $f \in L^2(\Omega)$, $\mathcal{A}^{-1} f = u$, where u solves (2.17). Here and below, integrals are with respect to the Lebesgue measure if $\Omega = \mathcal{D}$ and with respect to the Riemannian volume form if $\Omega = \mathcal{M}$. In the case $\Omega = \mathcal{D}$, we assume for concreteness Dirichlet boundary conditions and take $\mathcal{H}^1(\Omega) \equiv H_0^1(\mathcal{D})$. In the case $\Omega = \mathcal{M}$, we take $\mathcal{H}^1(\Omega) \equiv H^1(\mathcal{M})$. We emphasize that we view $\mathcal{A}^{-1}$ as a map from the right-hand side of the elliptic PDE to the solution. To ensure that a solution to the PDE exists, is unique, and is sufficiently regular, the coefficients Θ and b must be positive and sufficiently smooth; precise assumptions will be given in our theorem statements. Choosing these coefficients to be spatially varying allows one to encode additional prior information about the unknown, such as anisotropic correlations. The parameter α in the covariance $\mathcal{C}_0 = \mathcal{A}^{-\alpha}$ controls the regularity of prior draws. For simplicity, we restrict our attention to positive integer-valued α, but extensions to fractional values are possible [2, 7]. If the domain Ω is sufficiently regular, α is chosen to be large enough, and Θ, b are positive and sufficiently smooth, then the covariance operator $\mathcal{C}_0$ is trace-class on $L^2(\Omega)$ [55].

Example 2.2 (deconvolution forward model). Consider the heat equation

$$\begin{cases} v_t(x,t) = \Delta v(x,t), & x \in \Omega,\, t > 0, \\ v(x,0) = u(x), & x \in \Omega. \end{cases} \tag{2.18}$$

In the case $\Omega = \mathcal{D}$, we supplement (2.18) with Dirichlet boundary conditions $v(x,t) = 0$, $x \in \partial\mathcal{D}$. We let $\mathcal{G} : L^2(\Omega) \to L^2(\Omega)$ map an input function $u \in L^2(\Omega)$ to the solution to (2.18) at time $t = 1$. It is well known that the heat equation admits a variational characterization, which will lend itself more naturally to the numerical approximations introduced in subsections 2.3.1 and 2.3.2. We consider a linear observation model given by local averages of the solution over a Euclidean ball of radius $\delta > 0$, which approximate pointwise observations:

$$\mathcal{O}v = \left[v^\delta(x_1), \dots, v^\delta(x_{d_y})\right]^T, \quad v^\delta(x_i) := \int_{B_\delta(x_i)\cap\Omega} v(x,1). \tag{2.19}$$

We then write our forward model as

$$\mathcal{F} := \mathcal{O} \circ \mathcal{G}. \tag{2.20}$$

The utility of considering the local average observation map is that it is a bounded linear map from $L^2(\Omega)$ to $\mathbb{R}^{d_y}$, whereas pointwise observations are only bounded from subspaces of $L^2(\Omega)$ with enough regularity to guarantee almost everywhere continuity. While the forward map given by the heat equation is sufficiently smoothing to guarantee the solution is continuous, the graph-based approximations to the forward map that we will consider may only converge in $L^2(\Omega)$. In those cases, it will be more convenient to consider the local average observations, as in [25].

2.3.1. Finite element discretizations. The use of finite elements within the computational framework reviewed in subsection 2.2 was proposed and numerically investigated in [10]. Here, the domain Ω is an open, bounded, and sufficiently regular subset $\mathcal{D} \subset \mathbb{R}^d$. We let $\mathcal{V}$ be a finite element discretization space with basis functions denoted by $\{\phi_j\}_{j=1}^n$. For concreteness, we restrict our attention to linear Lagrange polynomial basis vectors. These basis functions correspond to nodal points $\{x_j\}_{j=1}^n \in \mathbb{R}^d$ such that $\phi_i(x_j) = \delta_{ij}$ for $i, j \in \{1, \dots, n\}$. The domain $\mathcal{D}$ is partitioned by a mesh into elements $e_1, e_2, \dots$ with $h_i := \text{diameter}(e_i)$ and $h := \max h_i$. Two canonical families of Lagrange elements include d-simplexes and d-hypercubes. A d-simplex is the region $e \subset \mathbb{R}^d$ determined by $d+1$ distinct points (for $d = 1$ this is an interval, and for $d = 2$ is a triangle). A d-hypercube is a product of d intervals on the real line, $e = [a_1, b_1] \times \cdots \times [a_d, b_d] \subset \mathbb{R}^d$. As the mesh is refined (that is, as we take $h \to 0$), the dimension of $\mathcal{V}$ grows. To emphasize that we are considering a family of subspaces indexed by the mesh width parameter h, we adopt the notation $\mathcal{V} = \mathcal{V}_h$, as is typical in the finite element literature. The theory we will present in subsection 3.2 holds for uniform (and, more generally, quasi-uniform) mesh refinements.

Example 2.3 (finite element approximation of Matérn covariance). For an operator $\mathcal{B} : L^2(\mathcal{D}) \to L^2(\mathcal{D})$, one can consider the action of this operator restricted to the subspace $\mathcal{V}_h$, given by $\mathcal{B}_h = P^*P\mathcal{B}P^*P : \mathcal{V}_h \to \mathcal{V}_h$, which is a finite-dimensional linear map. The matrix representation B of this operator must satisfy

$$\int_{\mathcal{D}} \phi_i \mathcal{B}\phi_j \, dx = \langle \vec{e}_i, B\vec{e}_j \rangle_M,$$

where $\vec{e}_i$ is the coordinate vector that corresponds to ϕ_i. This implies that $B : \mathbb{R}^n_M \to \mathbb{R}^n_M$ can be written explicitly as

$$B = M^{-1}K, \tag{2.21}$$

where

$$K_{ij} = \int_{\mathcal{D}} \phi_i \mathcal{B}\phi_j \, dx, \quad i, j \in \{1, \dots, n\}. \tag{2.22}$$

In particular, when representing the differential operator $\mathcal{A}$ defined in (2.17), we get that K is given by the stiffness matrix with entries

$$K_{ij} = \int_{\mathcal{D}} \Theta(x)\nabla\phi_i(x) \cdot \nabla\phi_j(x) + b(x)\phi_i(x)\phi_j(x) \, dx, \quad i, j \in \{1, \dots, n\}. \tag{2.23}$$

One can verify that both $A = M^{-1}K$ and $A^{-1} = K^{-1}M$ are self-adjoint in the weighted inner product space. Recall that the coefficients of the Galerkin approximation to the elliptic PDE (2.17) are given by solving the linear system

$$K\vec{u} = \vec{b},$$

where $b_i = \int_{\mathcal{D}} f\phi_i \, dx$ for $1 \le i \le n$, see [44]. Since we have that $\vec{b} = M\vec{f}$ (where $\vec{f} = Pf$), we see that $A^{-1}\vec{f} = K^{-1}M\vec{f}$ is precisely the Galerkin approximation to $\mathcal{A}^{-1}f$. We will show in subsection 3.2.1 that the matrix $C_0 : \mathbb{R}^n_M \to \mathbb{R}^n_M$ given by $C_0 = A^{-\alpha} = (K^{-1}M)^{\alpha}$ well approximates the continuum Matérn covariance operator $\mathcal{C}_0$.

Example 2.4 (finite element approximation of heat forward model). The variational formulation of the heat equation (2.18) naturally leads to a *semidiscrete* Galerkin approximation

$$v_h(x,t) = \sum_{i=1}^{n} A^i(t)\phi_i(x), \tag{2.24}$$

which is required to satisfy that, for all $\phi \in \mathcal{V}_h$,

$$\int_{\mathcal{D}} \frac{\partial v_h}{\partial t}\phi(x) \, dx + \int_{\mathcal{D}} \nabla v_h \cdot \nabla\phi(x) \, dx = 0, \qquad t \in (0,1], \tag{2.25}$$

$$\int_{\mathcal{D}} v_h(x,0)\phi(x) \, dx = \int_{D} u(x)\phi(x) \, dx, \qquad t = 0. \tag{2.26}$$

This leads to a system of differential equations for the time-dependent coefficients $A^i(t)$,

$$\sum_{j=1}^{n} M_{ij}\frac{\partial}{\partial t}A^j(t) + K_{ij}A^j(t) = 0, \qquad i = 1, \dots, n, \tag{2.27}$$

$$\sum_{j=1}^{n} M_{ij} A^j(0) = g_i, \qquad i=1,\dots,n, \tag{2.28}$$

where $M_{ij} = \langle \phi_i, \phi_j \rangle_{L^2(\mathcal{D})}$, $K_{ij} = \langle \phi_i, -\Delta\phi_j \rangle_{L^2(\mathcal{D})}$, and $g_i = \langle u, \phi_i \rangle_{L^2(\mathcal{D})}$. Note that $\vec{A}(t) = \big(A^1(t),\dots,A^n(t)\big)^T \in \mathbb{R}^n_M$. Given the eigenpairs of K and M, this system of differential equations can be solved analytically. However, computing the eigenpairs may be prohibitively expensive in practice, so the solution is often approximated by a time-discretization of the semidiscrete problem. One such classical time discretization of (2.25) is the Crank–Nicolson method: find $v_h^k \in V_h$, for $k=0,\dots,1/\Delta t$ such that, for every $\phi \in \mathcal{V}_h$,

$$\int_{\mathcal{D}} \frac{v_h^k(x) - v_h^{k-1}(x)}{\Delta t}\phi(x)\,dx + \int_{\mathcal{D}} \nabla\left(\frac{v_h^k(x) + v_h^{k-1}(x)}{2}\right)\cdot\nabla\phi(x)\,dx = 0, \quad k=1,\dots,1/\Delta t, \tag{2.29}$$

$$\int_{\mathcal{D}} v_h^0(x)\phi(x)\,dx = \int_{\mathcal{D}} u(x)\phi(x)\,dx. \tag{2.30}$$

This leads to a system of linear equations analogous to (2.27) for the coefficients $A_k^j \approx A^j(k\Delta t)$ for $k=0,\dots,1/\Delta t - 1$:

$$\begin{aligned} \frac{M+K\Delta t}{2}\vec{A}_{k+1} &= \frac{M-K\Delta t}{2}\vec{A}_k, \qquad && k=1,\dots,1/\Delta t - 1,\\ M\vec{A}_0 &= \vec{g}, && k=0. \end{aligned}$$

The second equation is simply the requirement that $\vec{A}_0$ be given by the orthogonal projection onto $\mathcal{V}_h$ as $\vec{A}_0 = Pu$. We write the mapping from the coefficients of the initial condition to the coefficients of the solution at $t=1$ as $G : \mathbb{R}^n_M \to \mathbb{R}^n_M$, where

$$G = \left(\left(\frac{M+K\Delta t}{2}\right)^{-1}\left(\frac{M-K\Delta t}{2}\right)\right)^{\frac{1}{\Delta t}}. \tag{2.31}$$

We denote by $O : \mathbb{R}^n_M \to \mathbb{R}^{d_y}$ the map from coefficients to observations. Our discretized forward model $F : \mathbb{R}^n_M \to \mathbb{R}^{d_y}$ is then given by

$$F := OG. \tag{2.32}$$

We remark that the observation map $\mathcal{O}$ applied to the numerical solution to the heat equation is an integral of a piecewise linear function. Such an integral can be computed exactly via quadrature methods, which are typically tractable in low dimensions (e.g., $d=1,2,3$). Since finite element methods are seldom implemented in higher dimensions in practice, we make the simplifying assumption that the observation map can be evaluated exactly for functions in $\mathcal{V}_h$.

2.3.2. Graph-based discretizations. We will now see how the graph-based discretizations considered in [25] fit into the above framework. Here we let $\Omega = \mathcal{M}$ be a d-dimensional smooth, connected, compact, Riemannian manifold embedded in $\mathbb{R}^D$. We assume further that $\mathcal{M}$ has bounded sectional curvature and Riemannian metric inherited from $\mathbb{R}^D$. Without loss of generality, we let $\mathcal{M}$ be normalized such that $vol(\mathcal{M})=1$. We assume we have access to a point cloud $\mathcal{M}_n=\{x_1,\dots,x_n\}\subset\mathcal{M}$ of n samples from the uniform distribution γ on $\mathcal{M}$. We let $\gamma_n=\frac{1}{n}\sum_{i=1}^n\delta_{x_i}$ denote the empirical measure of $\mathcal{M}_n$. We will show that this setting naturally fits into our computational framework with basis vectors of the form $\phi_i(x)=\mathbb{1}_{U_i}(x)$, where the sets $\{U_i\}_{i=1}^n\subset\mathcal{M}$ form a partition of $\mathcal{M}$. We recall the following result from [24].

Proposition 2.5 (existence of transport maps). *There is a constant c such that, with probability one, there exists a sequence of transport maps $T_n:\mathcal{M}\to\mathcal{M}_n$ such that $\gamma_n=T_{n\sharp}\gamma$ and*

$$\lim_{n\to\infty}\sup\frac{n^{1/d}\sup_{x\in\mathcal{M}}d_{\mathcal{M}}(x,T_n(x))}{(\log n)^{c_d}}\le c, \tag{2.33}$$

where $c_d=3/4$ if $d=2$ and $c_d=1/d$ otherwise.

In the above, $\gamma_n=T_{n\sharp}\gamma$ indicates that $\gamma(T_n^{-1}(U))=\gamma_n(U)$ for all measurable U, and $d_{\mathcal{M}}$ denotes the geodesic distance. We denote the preimage of each singleton as $U_i=T_n^{-1}(\{x_i\})$. The sets U_i form a partition of $\mathcal{M}$, and by the measure preserving property of T_n all have mass $\gamma(U_i)=1/n$. It follows from Proposition 2.5 that we have $U_i\subset B_{\mathcal{M}}(x_i,\varepsilon_n)$, where

$$\varepsilon_n:=\sup_{x\in\mathcal{M}}d_{\mathcal{M}}(x,T_n(x))\lesssim\frac{(\log n)^{c_d}}{n^{1/d}}.$$

These sets will be used to define our basis functions $\phi_i(x)=\mathbb{1}_{U_i}(x)$, which correspond to locally constant interpolations of functions in $L^2(\mathcal{M})$. We can then interpret vectors of function values on the point cloud $\mathcal{M}_n$ as coefficient vectors in $\mathbb{R}^n_M$, where the corresponding mass matrix is $M=\frac{1}{n}I_n$.

Example 2.6 (graph approximation of Matérn covariance). In the manifold setting, we opt to approximate the elliptic differential operator via the graph Laplacian. We define a symmetric weight matrix $W\in\mathbb{R}^{n\times n}$ that has entries $W_{ij}\ge 0$ corresponding to closeness between points in $\mathcal{M}_n$ by

$$W_{ij}=\frac{2(d+2)}{n\nu_d h_n^{d+2}}\mathbb{1}_{\{\|x_i-x_j\|_2<h_n\}}. \tag{2.34}$$

Here ν_d denotes the volume of the d-dimensional unit ball, and h_n denotes the graph connectivity. Defining the degree matrix $D=diag(d_1,\dots,d_n)$, where $d_i=\sum_{j=1}^n W_{ij}$, we give the unnormalized graph Laplacian by $\Delta_n^{un}=D-W$. This operator can be shown to converge pointwise and spectrally to $-\Delta$ (the Laplace–Beltrami operator) in the manifold setting [5, 11]. This construction can also be generalized to elliptic operators with spatially varying coefficients. To do so, we define a nonstationary weight matrix $\tilde{W}$ with entries

$$\tilde{W}_{ij}=W_{ij}\sqrt{\Theta(x_i)\Theta(x_j)}=\frac{2(d+2)}{n\nu_d h_n^{d+2}}\mathbb{1}_{\{\|x_i-x_j\|_2<h_n\}}\sqrt{\Theta(x_i)\Theta(x_j)}. \tag{2.35}$$

We then define $\Delta_n^{\Theta}=\tilde{D}-\tilde{W}$ where $\tilde{D}_{ii}=\sum_{j=1}^n\tilde{W}_{ij}$. Denoting $B_n=diag(b(x_1),\dots,b(x_n))$, we approximate the differential operator $\mathcal{A}$ as the matrix $A:\mathbb{R}^n_M\to\mathbb{R}^n_M$ given by

$$A=\Delta_n^{\Theta}+B_n. \tag{2.36}$$

Both A and A^{-1} are symmetric and self-adjoint in the weighted inner product space. The matrix $C_0 : \mathbb{R}^n_M \to \mathbb{R}^n_M$ given by $C_0 = A^{-\alpha}$ will be shown to well approximate the continuum Matérn covariance operator $\mathcal{C}_0$ in subsection 3.3.1.

Example 2.7 (graph approximation of heat forward model). For our graph-based approximation of the heat forward model, we consider the differential equation

$$\begin{cases} \frac{\partial v_n}{\partial t} = -\Delta_n^{un} v_n, & t > 0, \\ v_n(0) = u_n. \end{cases} \tag{2.37}$$

We let $G_n : \mathbb{R}^n_M \to \mathbb{R}^n_M$ map $u_n \in \mathbb{R}^n_M$ to the solution $v_n(1)$ of (2.37). Given the eigenpairs of the graph Laplacian, denoted $\{(\lambda_n^{(i)}, \psi_n^{(i)}\}_{i=1}^n$, we can explicitly write the solution to this ordinary differential equation (ODE):

$$G_n(u_n) = \sum_{i=1}^n \exp\left(-\lambda_n^{(i)}\right) \langle u_n, \psi_n^{(i)} \rangle_M \psi_n^{(i)}. \tag{2.38}$$

The computations presented above for both the prior covariance discretization and the heat forward model do not require the explicit computation of the sets U_i that partition $\mathcal{M}$ and define our basis vectors. In order to exactly compute the observation map defined in (2.19), we would need to know precisely the sets U_i. To circumvent this challenge, we instead consider a surrogate observation map $\mathcal{O}_n : \mathbb{R}^n_M \to \mathbb{R}^{d_y}$ that only requires knowledge of the point cloud $\mathcal{M}_n$,

$$\mathcal{O}_n v_n = \left[\bar{v}^\delta_{n,1}, \ldots, \bar{v}^\delta_{n,1}\right]^T, \quad \bar{v}^\delta_{n,j} := \frac{1}{n} \sum_{k : x_k \in B_\delta(x_j) \cap \mathcal{M}_n} [v_n(1)]_k, \tag{2.39}$$

for $1 \le j \le d_y$, where $[v_n(1)]_k$ denotes the kth entry of the vector v_n. For the purposes of our analysis, it will be necessary to make the assumption on the boundary of $\mathcal{B}_\delta(x_k) \cap \mathcal{M}$. In particular, we assume that length$\,(\partial(\mathcal{B}_\delta(x_k) \cap \mathcal{M})) = C_{\delta,j}$ for $1 \le j \le d_y$, where $C_{\delta,j}$ is a constant that depends on $\mathcal{M}$, δ, and the observation point x_j.

Our discretized forward model $F : \mathbb{R}^n_M \to \mathbb{R}^{d_y}$ is then given by

$$F := \mathcal{O}_n G_n. \tag{2.40}$$

3. Error analysis.

This section analyzes the computational framework introduced in section 2. We start in subsection 3.1 by establishing error bounds on the infinite-dimensional posterior mean and covariance based on a general assumption. Next, we verify this assumption in the function space setting of subsection 2.3, studying finite element discretizations (subsection 3.2) and graph-based methods (subsection 3.3).

3.1. Error analysis: General computational framework.

The main results of this subsection, Theorems 3.3 and 3.4, quantify the errors in the mean and covariance approximations

$$\varepsilon_m = \|m_{\text{post}} - P^* \vec{m}_{\text{post}}\|_{\mathcal{H}}, \tag{3.1}$$

$$\varepsilon_C = \|\mathcal{C}_{\text{post}} - P^* C_{\text{post}} P\|_{op}, \tag{3.2}$$

where we recall that $P:\mathcal{H}\to\mathbb{R}^n_M$ is the discretization map defined in (2.7) and $P^*:\mathbb{R}^n_M\to\mathcal{V}\subset\mathcal{H}$ is the extension map defined in (2.8). Our bounds on ε_m and ε_C will rely on a general assumption on (i) the closeness of the covariance operator $\mathcal{C}_0:\mathcal{H}\to\mathcal{H}$ and its discrete approximation $C_0:\mathbb{R}^n_M\to\mathbb{R}^n_M$, encoded in an assumption on the operator norm $\|\mathcal{C}_0-P^*C_0P\|_{op}$; (ii) the closeness of the forward model $\mathcal{F}:\mathcal{H}\to\mathbb{R}^{d_y}$ and its approximation $F:\mathbb{R}^n_M\to\mathbb{R}^{d_y}$, encoded in an assumption on the operator norm $\|\mathcal{F}-FP\|_{op}$; and (iii) the closeness of the finite-dimensional space $\mathcal{V}$ and the infinite-dimensional space $\mathcal{H}$, encoded in an assumption that the orthogonal projection operator $P^*P:\mathcal{H}\to\mathcal{V}$ is close to the identity.

Assumption 3.1. The following hold:

(i) (Error in approximation of prior covariance.) There is a constant $r_1\in\mathbb{R}$ and a function $\psi_1:\mathbb{N}\to\mathbb{R}$ with $\lim_{n\to\infty}\psi_1(n)=0$ such that $\|\mathcal{C}_0-P^*C_0P\|_{op}\le r_1\psi_1(n)$.

(ii) (Error in approximation of forward model.) There is a constant $r_2\in\mathbb{R}$ and a function $\psi_2:\mathbb{N}\to\mathbb{R}$ with $\lim_{n\to\infty}\psi_2(n)=0$ such that $\|\mathcal{F}-FP\|_{op}\le r_2\psi_2(n)$.

(iii) (Error in orthogonal projection.) There is a Hilbert space $\mathcal{H}'$ continuously embedded in $\mathcal{H}$, a constant $r_3\in\mathbb{R}$, and a function $\psi_3:\mathbb{N}\to\mathbb{R}$ with $\lim_{n\to\infty}\psi_3(n)=0$ such that, for every $u\in\mathcal{H}'$, $\|(I-P^*P)u\|_{\mathcal{H}}\le r_3\psi_3(n)\|u\|_{\mathcal{H}'}$.

In subsections 3.2 and 3.3, we will verify Assumption 3.1 for the important examples of finite element and graph discretization spaces, Matérn-type prior covariance operators, and deconvolution forward models. In these examples, we will take the space $\mathcal{H}'$ in Assumption 3.1(iii) to be an appropriate Sobolev space contained in the Cameron–Martin space $\mathrm{Im}(\mathcal{C}_0^{1/2})$.

To establish upper bounds on the mean and covariance approximation errors (3.1) and (3.2), we will first prove a lemma similar to the lemmas in [26] and [36]. Following the approach in these papers, we introduce the Kalman gain operator

$$\mathscr{K}:S_+\times B(\mathcal{H},\mathbb{R}^{d_y})\to B(\mathbb{R}^{d_y},\mathcal{H}),\qquad \mathscr{K}(\mathcal{C},\mathcal{F})=\mathcal{C}\mathcal{F}^*(\mathcal{F}\mathcal{C}\mathcal{F}^*+\Gamma)^{-1}. \tag{3.3}$$

Unlike [26, 36], we view the Kalman gain operator as a function of not just the covariance but of both the forward map and the covariance. The following lemma shows that the Kalman gain operator is pointwise continuous.

Lemma 3.2 (continuity of Kalman gain update). *Let $\mathscr{K}$ be the Kalman gain operator defined in (3.3). Let $\mathcal{C}_1,\mathcal{C}_2\in S_+$, let $\mathcal{F}_1,\mathcal{F}_2\in B(\mathcal{H},\mathbb{R}^{d_y})$, and let $\Gamma\in S^{d_y}_{++}$. The following holds:*

$$\|\mathscr{K}(\mathcal{C}_1,\mathcal{F}_1)-\mathscr{K}(\mathcal{C}_2,\mathcal{F}_2)\|_{op}\le c_1\|\mathcal{C}_1-\mathcal{C}_2\|_{op}+c_2\|\mathcal{F}_1-\mathcal{F}_2\|_{op}, \tag{3.4}$$

where

$$\begin{aligned}c_1&=\|\Gamma^{-1}\|_{op}\|\mathcal{F}_2\|_{op}+\|\Gamma^{-1}\|^2_{op}\|\mathcal{C}_1\|_{op}\|\mathcal{F}_1\|_{op}\|\mathcal{F}_2\|^2_{op},\\ c_2&=\|\Gamma^{-1}\|_{op}\|\mathcal{C}_1\|_{op}+\|\Gamma^{-1}\|^2_{op}\|\mathcal{C}_1\|^2_{op}\|\mathcal{F}_1\|^2_{op}+\|\Gamma^{-1}\|_{op}\|\mathcal{C}_1\|^2_{op}\|\mathcal{F}_1\|_{op}\|\mathcal{F}_2\|_{op}.\end{aligned} \tag{3.5}$$

Proof. Applying Lemma A.8 from [26] (which was stated for matrices but also holds in our infinite-dimensional setting), we get that

$$\begin{aligned}\|\mathscr{K}(\mathcal{F}_1,\mathcal{C}_1)-\mathscr{K}(\mathcal{F}_2,\mathcal{C}_2)\|_{op}\le{}&\|\Gamma^{-1}\|_{op}\|\mathcal{C}_1\mathcal{F}_1^*-\mathcal{C}_2\mathcal{F}_2^*\|_{op}\\&+\|\Gamma^{-1}\|^2_{op}\|\mathcal{C}_1\mathcal{F}_1^*\|_{op}\|\mathcal{F}_1\mathcal{C}_1\mathcal{F}_1^*-\mathcal{F}_2\mathcal{C}_2\mathcal{F}_2^*\|_{op}.\end{aligned}$$

We can then bound each of these terms with the triangle inequality as follows:

$$\begin{aligned}\|\mathcal{C}_1\mathcal{F}_1^* - \mathcal{C}_2\mathcal{F}_2^*\|_{op} &= \|\mathcal{C}_1\mathcal{F}_1^* - \mathcal{C}_1\mathcal{F}_2^* + \mathcal{C}_1\mathcal{F}_2^* - \mathcal{C}_2\mathcal{F}_2^*\|_{op}\\ &\le \|\mathcal{C}_1\|_{op}\|\mathcal{F}_1 - \mathcal{F}_2\|_{op} + \|\mathcal{F}_2\|_{op}\|\mathcal{C}_1 - \mathcal{C}_2\|_{op}\end{aligned}$$

and

$$\begin{aligned}\|\mathcal{F}_1\mathcal{C}_1\mathcal{F}_1^* - \mathcal{F}_2\mathcal{C}_2\mathcal{F}_2^*\|_{op} &= \|\mathcal{F}_1\mathcal{C}_1\mathcal{F}_1^* - \mathcal{F}_2\mathcal{C}_1\mathcal{F}_1^* + \mathcal{F}_2\mathcal{C}_1\mathcal{F}_1^* - \mathcal{F}_2\mathcal{C}_2\mathcal{F}_2^*\|_{op}\\ &\le \|\mathcal{C}_1\mathcal{F}_1^*\|_{op}\|\mathcal{F}_1 - \mathcal{F}_2\|_{op} + \|\mathcal{F}_2\|_{op}\|\mathcal{C}_1\mathcal{F}_1^* - \mathcal{C}_2\mathcal{F}_2^*\|_{op}.\end{aligned}$$

Combining these three displayed inequalities gives the desired result. ■

Theorem 3.3 (mean approximation error). *Consider the error ε_m in (3.1) between the true posterior mean in (2.5) and its finite-dimensional posterior approximation in (2.13). Then, under Assumption 3.1 it holds that*

$$\varepsilon_m \le r_1'\psi_1(n) + r_2'\psi_2(n) + r_3'\psi_3(n), \tag{3.6}$$

where

$$\begin{aligned}r_1' &= 2c_1 r_1\|y - \mathcal{F}m_0\|_2,\\ r_2' &= 2c_2 r_2\|y - \mathcal{F}m_0\|_2 + r_2\|m_0\|_{\mathcal{H}}\|\mathscr{K}(\mathcal{C}_0, \mathcal{F})\|_{op},\\ r_3' &= r_3\|m_0\|_{\mathcal{H}'},\end{aligned}$$

and c_1, c_2 are defined as in (3.5).

Proof. By the definition of ε_m in (3.1) and an application of the triangle inequality,

$$\begin{aligned}\varepsilon_m \le \|m_0 - P^*\vec{m}_0\|_{\mathcal{H}} &+ \|\mathscr{K}(\mathcal{C}_0, \mathcal{F}) - \mathscr{K}(P^*C_0P, FP)\|_{op}\|y - F\vec{m}_0\|_2\\ &+ \|\mathscr{K}(\mathcal{C}_0, \mathcal{F})\|_{op}\|\mathcal{F}m_0 - F\vec{m}_0\|_2.\end{aligned} \tag{3.7}$$

We next bound each of the terms on the right-hand side. By Assumption 3.1(iii) and the fact that $\vec{m}_0 = Pm_0$, we have that

$$\|m_0 - P^*\vec{m}_0\|_{\mathcal{H}} = \|m_0 - P^*Pm_0\|_{\mathcal{H}} \le r_3\|m_0\|_{\mathcal{H}'}\psi_3(n). \tag{3.8}$$

By Lemma 3.2 and Assumption 3.1(i) and (ii), we have that

$$\begin{aligned}\|\mathscr{K}(\mathcal{C}_0, \mathcal{F}) - \mathscr{K}(P^*C_0P, FP)\|_{op} &\le c_1\|\mathcal{C}_0 - P^*C_0P\|_{op} + c_2\|\mathcal{F} - FP\|_{op}\\ &\le c_1 r_1\psi_1(n) + c_2 r_2\psi_2(n).\end{aligned} \tag{3.9}$$

By Assumption 3.1(ii) and (iii) we have that

$$\begin{aligned}\|y - F\vec{m}_0\| &\le \|y - \mathcal{F}m_0\| + \|\mathcal{F}\|_{op}\|m_0 - Pm_0\|_{\mathcal{H}} + \|\mathcal{F} - FP\|_{op}\|m_0\|_{\mathcal{H}}\\ &\le \|y - \mathcal{F}m_0\| + \|\mathcal{F}\|_{op}\|m_0\|_{\mathcal{H}'}r_3\psi_3(n) + \|m_0\|_{\mathcal{H}}r_2\psi_2(n).\end{aligned}$$

This implies that for n sufficiently large, it holds that

$$\|y - F\vec{m}_0\|_2 \le 2\|y - \mathcal{F}m_0\|_2. \tag{3.10}$$

Finally, by Assumption 3.1(ii) we have that

$$\|\mathcal{F}m_0 - F\vec{m}_0\|_2 \le \|\mathcal{F} - FP\|_{op}\|m_0\|_{\mathcal{H}} \le r_2\|m_0\|_{\mathcal{H}}\psi_2(n). \tag{3.11}$$

Plugging the bounds (3.8), (3.9), (3.10), and (3.11) into inequality (3.7) gives the desired result. ■

Theorem 3.4 (covariance approximation error). *Consider the error ε_C in (3.2) between the true posterior covariance operator in (2.6) and its finite-dimensional posterior approximation in (2.14). Then, under Assumption 3.1 it holds that*

$$\varepsilon_C \le r_1' \psi_1(n) + r_2' \psi_2(n), \tag{3.12}$$

where

$$\begin{aligned} r_1' &= r_1 + 2c_1 r_1 \|\mathcal{F}\|_{op} \|\mathcal{C}_0\|_{op} + 2r_1 \|\mathcal{F}\|_{op} \|\mathscr{K}(\mathcal{C}_0, \mathcal{F})\|_{op}, \\ r_2' &= 2c_2 r_2 \|\mathcal{F}\mathcal{C}_0\|_{op} + r_2 \|\mathcal{C}_0\|_{op} \|\mathscr{K}(\mathcal{C}_0, \mathcal{F})\|_{op}, \end{aligned}$$

and c_1, c_2 are defined as in (3.5).

Proof. By the definition of ε_C in (3.2) and an application of the triangle inequality,

$$\begin{aligned} \varepsilon_C \le{}& \|\mathcal{C}_0 - P^* C_0 P\|_{op} + \|\mathscr{K}(\mathcal{C}_0, \mathcal{F}) - \mathscr{K}(P^* C_0 P, FP)\|_{op} \|F C_0 P\|_{op} \\ &+ \|\mathscr{K}(\mathcal{C}_0, \mathcal{F})\|_{op} \|\mathcal{F}\mathcal{C}_0 - F C_0 P\|_{op}. \end{aligned} \tag{3.13}$$

By Assumption 3.1(i) we have that

$$\|\mathcal{C}_0 - P C_0 P\|_{op} \le r_1 \psi_1(n). \tag{3.14}$$

As in the previous proof, applying Lemma 3.2 and using Assumption 3.1(i) and (ii) gives

$$\|\mathscr{K}(\mathcal{C}_0, \mathcal{F}) - \mathscr{K}(P^* C_0 P, FP)\|_{op} \le c_1 r_1 \psi_1(n) + c_2 r_2 \psi_2(n). \tag{3.15}$$

The triangle inequality along with Assumption 3.1(i) and (ii) yield the bound

$$\begin{aligned} \|\mathcal{F}\mathcal{C}_0 - F C_0 P\|_{op} &\le \|FP\|_{op} \|\mathcal{C}_0 - P^* C_0 P\|_{op} + \|\mathcal{F} - FP\|_{op} \|\mathcal{C}_0\|_{op} \\ &\le r_1 \|FP\|_{op} \psi_1(n) + r_2 \|\mathcal{C}_0\|_{op} \psi_2(n). \end{aligned} \tag{3.16}$$

Assumption 3.1(i) and (ii) guarantee that, for sufficiently large n,

$$\|FP\|_{op} \le 2\|\mathcal{F}\|_{op} \quad \text{and} \quad \|F C_0 P\|_{op} \le 2\|\mathcal{F}\|_{op} \|\mathcal{C}_0\|_{op}. \tag{3.17}$$

Plugging in the bounds (3.14), (3.15), (3.16), and (3.17) into (3.13) gives the desired result. ∎

3.2. Error analysis: Finite element discretizations. The following result applies Theorems 3.3 and 3.4 to quantify the errors in the finite element approximations to the posterior mean and covariance operators in the setting considered in subsection 2.3.1.

Theorem 3.5 (finite element posterior mean and covariance approximation). *Consider the discretization proposed in subsection 2.3.1 with a finite element space $\mathcal{V}_h$ of linear Lagrange basis vectors and time discretization step of Δt. Assume the prior mean function is chosen to be in the Sobolev space $H^s(\mathcal{D})$. Let $\Theta(x) \in C^1(\bar{\mathcal{D}})$, let $b(x) \in L^\infty(\mathcal{D})$, and assume that both functions are (almost surely) bounded below by positive constants. The errors in the mean and covariance approximations are bounded by*

$$\begin{aligned} \|m_{\text{post}}(x) - P^* \vec{m}_{\text{post}}\|_{L^2(\mathcal{D})} &\lesssim h^{\min\{2,s\}} + \Delta t^2, \\ \|\mathcal{C}_{\text{post}} - P^* C_0 P\|_{op} &\lesssim h^2 + \Delta t^2. \end{aligned}$$

This result follows from Theorems 3.3 and 3.4 upon verifying that Assumption 3.1 holds for the finite element discretizations of our model Bayesian inverse problem. Verifying these assumptions will be the focus of the following three subsections. In so doing, we will view the ψ_i in Assumption 3.1 as functions of the mesh width h, noting that this quantity determines the dimension n of the discretization space. For simplicity, we have restricted our discussion to linear Lagrange basis vectors. As our general theory suggests, higher degree polynomials could be used in the finite element approximations to yield faster convergence rates.

3.2.1. Finite element approximation of Matérn-type prior covariance. Here we show that the finite element approximation to the covariance operator given in subsection 2.3.1 satisfies Assumption 3.1(i).

Theorem 3.6 (operator-norm error for finite element prior covariance approximation). *Let $\mathcal{C}_0 : L^2(\mathcal{D}) \to L^2(\mathcal{D})$ be the prior covariance operator defined in section 2, and let $A^{-\alpha} : \mathbb{R}^n_M \to \mathbb{R}^n_M$ be the finite-dimensional approximation defined in subsection 2.3.1 corresponding to a finite element space $\mathcal{V}_h$ with piecewise linear basis functions. Let $\Theta(x) \in C^1(\bar{\mathcal{D}})$, let $b(x) \in L^\infty(\mathcal{D})$, and assume that both functions are (almost surely) bounded below by positive constants. Then, there exists a constant c independent of h such that*

$$\|\mathcal{C}_0 - P^* A^{-\alpha} P\|_{op} \leq ch^2. \tag{3.18}$$

The proof can be found in Appendix B. This shows that the finite-dimensional prior approximation proposed in subsection 2.3.1 satisfies Assumption 3.1(i) with $\psi_1(h) = h^2$.

3.2.2. Finite element approximation of heat forward model. We now show that the discretized finite element forward map given in (2.32) satisfies Assumption 3.1(ii).

Theorem 3.7 (operator-norm error for finite element forward map approximation). *Let $\mathcal{F} : L^2(\mathcal{D}) \to \mathbb{R}^{d_y}$ be the forward model defined in (2.20), and let $F : \mathbb{R}^n_M \to \mathbb{R}^{d_y}$ be the finite-dimensional approximation defined in (2.32) corresponding to a finite element space $\mathcal{V}_h$ consisting of linear Lagrange basis vectors and a time-discretization step of Δt. Then, there exist constants c_1 and c_2 independent of h and Δt such that*

$$\|\mathcal{F} - FP\|_{op} \leq c_1 h^2 + c_2 \Delta t^2. \tag{3.19}$$

The proof is given in Appendix B. We see that Assumption 3.1(ii) holds for the Crank–Nicolson discretization with $\psi_2(h) = h^2 + \Delta t^2$.

3.2.3. Finite element orthogonal projection. We first verify that the orthogonal projection of the prior mean function onto the finite element subspace $\mathcal{V}_h$ converges as in Assumption 3.1(iii) as the mesh is refined. We let $m_0(x) \in H^s(\mathcal{D})$. Since we require that the mean function lies in the Cameron–Martin space $E = \mathrm{Im}(\mathcal{C}_0^{1/2})$, we must have that $s \geq \alpha$. We assume that $\mathcal{V}_h$ consists of piecewise linear basis functions. Since $s \geq \alpha > d/2$, the Sobolev embedding theorem guarantees that functions in $H^s(\mathcal{D})$ are continuous. Consequently, the prior mean function can be well approximated by piecewise linear interpolants and, in turn, by its orthogonal projection onto $\mathcal{V}_h$. Assuming that the finite element space $\mathcal{V}_h$ is given by a quasi-uniform mesh refinement of $\mathcal{D}$ with linear Lagrange basis functions, [44, Theorem 6.8] guarantees that, for any $m_0 \in H^s(\mathcal{D})$, there exists an element $U \in \mathcal{V}_h$ such that

$$\|m_0 - U\|_{L^2(\mathcal{D})} \leq ch^{\min\{2,s\}}\|m_0\|_{H^s(\mathcal{D})}.$$

Since the orthogonal projection is defined to minimize the $L^2(\mathcal{D})$ error over all functions in $\mathcal{V}_h$, we have that, for any $m_0 \in H^s(\mathcal{D})$,

$$\|m_0 - P^*Pm_0\|_{L^2(\mathcal{D})} \leq \|m_0 - U\|_{L^2(\mathcal{D})} \leq ch^{\min\{2,s\}}\|m_0\|_{H^s(\mathcal{D})}. \tag{3.20}$$

We see that Assumption 3.1(iii) holds for the finite element discretization with $\psi_3(h) = h^{\min\{2,s\}}$ and $\mathcal{H}' = H^s(\mathcal{D})$.

We have now verified that Assumption 3.1 holds for the finite element discretization, proving Theorem 3.5.

3.3. Error analysis: Graph-based approximations. The following result applies Theorems 3.3 and 3.4 to quantify the error in the graph-based approximations to the posterior mean and covariance operator in the setting considered in subsection 2.3.2.

Theorem 3.8 (graph-based posterior mean and covariance approximation). *Consider the discretization proposed in subsection 2.3.2. Assume we are in a realization in which the conclusion of Proposition 2.5 holds and we have that $\alpha > (5d+1)/4$. Let $b(x)$ be Lipschitz, let $\Theta(x) \in C^1(\mathcal{M})$, and suppose that both are bounded below by positive constants. Then, if we take the scaling $h_n \asymp \sqrt{\frac{(\log n)^{c_d}}{n^{1/d}}}$, the errors in the posterior mean and covariance approximations are bounded by*

$$\|m_{\text{post}}(x) - P^*\vec{m}_{\text{post}}\|_{L^2(\mathcal{M})} \lesssim (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}},$$
$$\|\mathcal{C}_{\text{post}} - P^*C_{\text{post}}P\|_{op} \lesssim (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}},$$

where $c_d = 3/4$ if $d = 2$ and $c_d = 1/d$ otherwise.

This result follows from Theorems 3.3 and 3.4 upon verifying that the graph-based discretization satisfies Assumption 3.1 for our model problem. Verifying these assumptions will be the focus of the following subsections. In all that follows, we assume we are in a realization where the conclusion of Proposition 2.5 holds.

3.3.1. Graph-based approximation of Matérn-type prior covariance. The following theorem shows that the covariance operator constructed in section 2.3.2 satisfies Assumption 3.1(ii). The proof can be found in Appendix C.

Theorem 3.9 (operator-norm error for graph-based prior covariance approximation). *Let $b(x)$ be Lipschitz, $\Theta(x) \in C^1(\mathcal{M})$, and suppose that both are bounded below by positive constants. Let $\alpha > (5d+1)/4$, and take $h_n \asymp \sqrt{\frac{(\log n)^{c_d}}{n^{1/d}}}$. Then,*

$$\|\mathcal{C}_0 - P^*(\Delta_n^\Theta + B_n)^{-\alpha}P\|_{op} \lesssim (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}},$$

where $c_d = 3/4$ if $d = 2$ and $c_d = 1/d$ otherwise.

Remark 3.10. The proof of Theorem 3.9 in Appendix C follows that of Theorem D.1 in [51]. However, since we are working with the covariance operator (as opposed to the square root of the covariance operator as is the case when sampling the Matérn fields), we get convergence rates for $\alpha > (5d+1)/4$ as opposed to $\alpha > (5d+1)/2$. The rates of convergence are still the same, since the error is still dominated by the error in approximating the eigenfunctions.

We have shown that the graph-based finite-dimensional approximation to the prior covariance proposed in subsection 2.3.2 satisfies Assumption 3.1(i) with $\psi_1(n) = (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}}$, where $c_d = 3/4$ if $d = 2$ and $c_d = 1/d$ otherwise.

3.3.2. Graph-based approximation of heat forward model.

We show that the graph-based approximation of the heat forward model, together with the surrogate observation operator, satisfies Assumption 3.1(iii). The proof can be found in Appendix C.

Theorem 3.11 (operator-norm error for graph-based forward map approximation). *Let $F = \mathcal{O}_n G_n$ be the forward model defined in subsection 2.3.2. Take $h_n \asymp \sqrt{\frac{(\log n)^{c_d}}{n^{1/d}}}$. Then,*

$$\|\mathcal{F} - P^* F P\|_{op} \lesssim (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}}. \tag{3.21}$$

We have shown that the graph-based finite-dimensional approximation to the forward model proposed in subsection 2.3.2 satisfies Assumption 3.1(ii) with $\psi_2(n) = (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}}$, where $c_d = 3/4$ if $d = 2$ and $c_d = 1/d$ otherwise.

3.3.3. Graph-based orthogonal projection.

Finally, we verify that the orthogonal projection of the prior mean function onto the subspace $\mathcal{V} = \operatorname{span}\{\mathbb{1}_{U_i}(x)\}_{i=1}^n$ converges as in Assumption 3.1(iii). Since the mean function must lie in the Cameron–Martin space $E = \operatorname{Im}(\mathcal{C}_0^{1/2})$, we necessarily have that $m_0(x) \in H^1(\mathcal{M})$. The following result from [24, Lemma 12] quantifies the convergence rate of the projected mean function.

Lemma 3.12 (Lemma 12 in [24]). *There exists a constant $c_{\mathcal{M}}$ independent of n such that, for every $m_0 \in H^1(\mathcal{M})$, we have*

$$\|m_0 - P^* P m_0\|_{L^2(\mathcal{M})} \leq c_{\mathcal{M}} \varepsilon_n \|m_0\|_{H^1(\mathcal{M})} \lesssim \frac{(\log n)^{c_d}}{n^{1/d}} \|m_0\|_{H^1(\mathcal{M})}. \tag{3.22}$$

We see that Assumption 3.1(iii) holds for the graph-based discretization with $\psi_3(n) = \frac{(\log n)^{c_d}}{n^{1/d}}$ and $\mathcal{H}' = H^1(\mathcal{M})$. For large n, this term will be dominated by $\psi_1(n)$ and $\psi_2(n)$.

We have now verified that Assumption 3.1 holds for the graph-based discretization, proving Theorem 3.8.

4. Sample size requirements for ensemble Kalman updates.

In this section, we formulate the ensemble Kalman update in weighted inner product space and show that the effective dimension of the discretized problem is controlled by the effective dimension of the continuum problem, which is finite.

4.1. Ensemble approximation to finite-dimensional posterior.

To implement an ensemble Kalman approximation to the finite-dimensional Gaussian posterior given by (2.12), we need to draw samples from the Gaussian prior distribution, $\mathcal{N}(\vec{m}_0, C_0)$ given by (2.10). To do so, we draw $\xi^{(j)} \sim \mathcal{N}(0, I)$ and compute samples

$$u^{(j)} = \vec{m}_0 + L\xi^{(j)}, \quad u^{(j)} \sim \mathcal{N}(\vec{m}_0, C_0), \tag{4.1}$$

for $1 \le j \le J$, where L is a linear map from $\mathbb{R}^n \to \mathbb{R}^n_M$ such that $C_0 = LL^\sharp = LL^T M$. Recall that given samples $u^{(1)}, \ldots, u^{(J)} \sim \mathcal{N}(0, \mathcal{C})$ from a Gaussian measure in a Hilbert space $\mathcal{H}$, the sample covariance operator is defined as the operator $\widehat{\mathcal{C}} : \mathcal{H} \to \mathcal{H}$,

$$\widehat{\mathcal{C}} u := \frac{1}{J-1} \sum_{j=1}^{J} \langle u^{(j)}, u \rangle_{\mathcal{H}} u^{(j)}, \quad u \in \mathcal{H}.$$

As such, the sample covariance operator on $\mathbb{R}^n_M$ is given by

$$\widehat{C}_0 = \frac{1}{J-1} \sum_{j=1}^{J} (u^{(j)} - \widehat{m}_0)(u^{(j)} - \widehat{m}_0)^T M,$$

where $\widehat{m}_0 = \frac{1}{J} \sum_{j=1}^{J} u^{(j)}$ is the sample mean. The perturbed observation ensemble Kalman update transforms each sample from the prior ensemble via

$$v^{(j)} = u^{(j)} + \widehat{C}_0 F^\natural \left(F \widehat{C}_0 F^\natural + \Gamma \right)^{-1} \left(y - F u^{(j)} + \eta^{(j)} \right), \quad \eta^{(j)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma). \tag{4.2}$$

We then output the sample mean and covariance of the transformed ensemble, given by

$$\widehat{m}_{\text{post}} = \frac{1}{J} \sum_{j=1}^{J} v^{(j)} \quad \text{and} \quad \widehat{C}_{\text{post}} = \frac{1}{J-1} \sum_{J=1}^{J} (v^{(j)} - \widehat{m}_{\text{post}})(v^{(j)} - \widehat{m}_{\text{post}})^T M. \tag{4.3}$$

4.2. Ensemble Kalman approximation: Error analysis. We now want to derive nonasymptotic expectation bounds for the error in the ensemble approximation to the infinite-dimensional posterior. We consider the errors

$$\widehat{\varepsilon}_m = \mathbb{E} \, \| m_{\text{post}} - P^* \widehat{m}_{\text{post}} \|_{\mathcal{H}}, \tag{4.4}$$

$$\hat{\varepsilon}_C = \mathbb{E} \, \| \mathcal{C}_{\text{post}} - P^* \hat{C}_{\text{post}} P \|_{op}. \tag{4.5}$$

Controlling these errors will amount to controlling the error between the ensemble approximations and the finite-dimensional approximations using the theory from [26] and then applying our results from section 3. Since the dimension of the discretized covariance matrices increases as the mesh is refined, we will require bounds that do not have an explicit dependence on the dimension. To do so, we define the *effective dimension* of a mapping $C_0 : \mathbb{R}^n_M \to \mathbb{R}^n_M$ to be

$$r_M(C_0) := \frac{\operatorname{Tr}(C_0)}{\|C_0\|_{op}}, \tag{4.6}$$

where $\| \cdot \|_{op}$ is the operator norm from $\mathbb{R}^n_M$ to $\mathbb{R}^n_M$. When the eigenvalues of C_0 decay rapidly, the effective dimension $r_M(C_0)$ is a better measure of dimension than the nominal state dimension n [60]. We will show in subsection 4.3 that for the finite element discretization, the effective dimension of the discretized operator is controlled by the effective dimension of the continuum operator, defined to be $r(\mathcal{C}_0) = \frac{\mathrm{Tr}(\mathcal{C}_0)}{\|\mathcal{C}_0\|_{op}}$, where $\|\cdot\|_{op}$ denotes the operator norm from $\mathcal{H}$ to $\mathcal{H}$. Since the prior covariance operator must be chosen to be bounded and trace-class, this quantity is necessarily finite. As such, the error bounds that we now derive for the ensemble approximation will not degenerate as the mesh is refined, despite the state dimension increasing arbitrarily.

Theorem 4.1 (mean ensemble approximation error). *Let $\widehat{\varepsilon}_m$ be the expected error between the infinite-dimensional posterior mean and the ensemble posterior mean with an ensemble of size J, as defined in (4.4). Assume that the finite-dimensional approximations satisfy Assumption 3.1; that for n sufficiently large, $r_M(C_0) \le c r(\mathcal{C}_0)$ for some constant c independent of n; and that $J \ge r_M(C_0)$. Then, it holds that*

$$\widehat{\varepsilon}_m \le r_1'\psi_1(n) + r_2'\psi_2(n) + r_3'\psi_3(n) + r_4'\frac{1}{\sqrt{J}}, \tag{4.7}$$

where the constants r_i' are independent of n and J.

Proof. By the triangle inequality, we decompose the expected error as

$$\mathbb{E}\,\|m_{\text{post}}(x) - P^*\widehat{m}_{\text{post}(x)}\|_{L^2(\mathcal{D})} \le \mathbb{E}\,\|m_{\text{post}}(x) - P^*\vec{m}_{\text{post}}(x)\|_{L^2} + \mathbb{E}\,\|\widehat{m}_{\text{post}} - \vec{m}_{\text{post}}\|_M.$$

The first term on the right-hand side is deterministic and is bounded by Theorem 3.3:

$$\mathbb{E}\,\|m_{\text{post}}(x) - P^*\vec{m}_{\text{post}}(x)\|_{L^2} = \|m_{\text{post}}(x) - P^*\vec{m}_{\text{post}}(x)\|_{L^2} \lesssim r_1'\psi_1(n) + r_2'\psi_2(n) + r_3'\psi_3(n).$$

For the second term on the right-hand side, we recall [26, Theorem 3.3], which guarantees the following:

$$\mathbb{E}\,\|\widehat{m}_{\text{post}} - \vec{m}_{\text{post}}\|_M \lesssim c_n'\sqrt{\frac{r_M(C_0)}{J}} + c_n''\sqrt{\frac{r_2(\Gamma)}{J}}, \tag{4.8}$$

where $c_n' = (\|C_0\|_{op}^{1/2} \vee \|C_0\|_{op}^2)\left(\|F\|_{op} \vee \|F\|_{op}^4\right)\left(\|\Gamma^{-1}\|_{op} \vee \|\Gamma^{-1}\|_{op}^2\right)(1 \vee \|y - F\vec{m}_0\|_2)$, and $c_n'' = \|F\|_{op}\|\Gamma^{-1}\|_{op}\|\Gamma\|_{op}^{1/2}\|C_0\|_{op}$. The quantity $r_2(\Gamma)$ is the effective dimension of Γ in the usual Euclidean inner product space. The quantities c_n' and c_n'' depend on n; however, under Assumption 3.1 each of the discretized operators appearing in the constants can be made arbitrarily close to their continuum counterparts when n is sufficiently large. In particular, we can take n to be large enough such that the following two bounds hold:

$$\begin{aligned} c_n' &\le c' = 2\left(\|\mathcal{C}_0\|_{op}^{1/2} \vee \|\mathcal{C}_0\|_{op}^2\right)\left(\|\mathcal{F}\|_{op} \vee \|\mathcal{F}\|_{op}^4\right)\left(\|\Gamma^{-1}\|_{op} \vee \|\Gamma^{-1}\|_{op}^2\right)(1 \vee \|y - \mathcal{F}m_0\|_2), \\ c_n'' &\le c'' = 2\|\mathcal{F}\|_{op}\|\Gamma^{-1}\|_{op}\|\Gamma\|_{op}^{1/2}\|\mathcal{C}_0\|_{op}. \end{aligned}$$

From these bounds combined with our assumption that $r_M(C_0) \le c r(\mathcal{C}_0)$, we take $r_4' = \sqrt{c}c'\sqrt{r(\mathcal{C}_0)} + c''\sqrt{r(\Gamma)}$, and the result follows. ∎

Theorem 4.2 (covariance ensemble approximation error). *Let $\widehat{\varepsilon}_C$ be the expected error between the infinite-dimensional posterior covariance and the ensemble posterior covariance with an ensemble of size J, as defined in (4.4). Assume that the finite-dimensional approximations satisfy Assumption 3.1; that for n sufficiently large, $r_M(C_0) \le c r(\mathcal{C}_0)$ for some constant c independent of n; and that $J \ge r_M(C_0)$. Then, it holds that*

$$\widehat{\varepsilon}_C \le r_1' \psi_1(n) + r_2' \psi_2(n) + r_3' \frac{1}{\sqrt{J}}, \tag{4.9}$$

where the constants r_i' are independent of the discretization dimension n and the ensemble size J.

Proof. The proof proceeds similarly to that of Theorem 4.1. We decompose the error as

$$\mathbb{E}\, \|\mathcal{C}_{\text{post}} - P^* \widehat{C}_{\text{post}} P\|_{op} \le \mathbb{E}\, \|\mathcal{C}_{\text{post}} - P^* C_{\text{post}} P\|_{op} + \mathbb{E}\, \|C_{\text{post}} - \widehat{C}_{\text{post}}\|_{op}. \tag{4.10}$$

The first term on the right-hand side is deterministic and is controlled by Theorem 3.4:

$$\mathbb{E}\, \|\mathcal{C}_{\text{post}} - P^* C_{\text{post}} P\|_{op} = \|\mathcal{C}_{\text{post}} - P^* C_{\text{post}} P\|_{op} \lesssim r_1' \psi_1(n) + r_2' \psi_2(n).$$

For the second term, we use [26, Theorem 3.5], which guarantees that

$$\mathbb{E}\, \|\widehat{C}_{\text{post}} - C_{\text{post}}\|_{op} \lesssim c_n' \sqrt{\frac{r_M(C_0)}{J}} + c_n'' \left(\sqrt{\frac{r_M(C_0)}{J}} \vee \sqrt{\frac{r_2(\Gamma)}{J}} \right), \tag{4.11}$$

where

$$\begin{aligned} c_n' &= \left(\|C_0\|_{op} \vee \|C_0\|_{op}^3\right) \left(\|F\|_{op}^2 \vee \|F\|_{op}^4\right) \left(\|\Gamma^{-1}\|_{op} \vee \|\Gamma^{-1}\|_{op}^2\right), \\ c_n'' &= \left(\|F\|_{op} \vee \|F\|_{op}^3\right) \left(\Gamma^{-1}\| \vee \|\Gamma^{-1}\|^2\right) \left(\|C_0\|_{op} \vee \|\Gamma\|_{op}\right) \left(\|C_0\|_{op} \vee \|C_0\|^2\right). \end{aligned}$$

Again, the quantities c_n' and c_n'' depend on n; however, under Assumption 3.1 each of the discretized operators appearing in the constants can be made arbitrarily close to their continuum counterparts when n is sufficiently large. In particular, we take n to be large enough such that the following two bounds hold:

$$\begin{aligned} c_n' &\le c' = 2\left(\|\mathcal{C}_0\|_{op} \vee \|\mathcal{C}_0\|_{op}^3\right) \left(\|\mathcal{F}\|_{op}^2 \vee \|\mathcal{F}\|_{op}^4\right) \left(\|\Gamma^{-1}\|_{op} \vee \|\Gamma^{-1}\|_{op}^2\right), \\ c_n'' &\le c'' = 2\left(\|\mathcal{F}\|_{op} \vee \|\mathcal{F}\|_{op}^3\right) \left(\Gamma^{-1}\| \vee \|\Gamma^{-1}\|^2\right) \left(\|\mathcal{C}_0\|_{op} \vee \|\Gamma\|_{op}\right) \left(\|\mathcal{C}_0\|_{op} \vee \|\mathcal{C}_0\|^2\right). \end{aligned}$$

With these bounds and our assumption that $r_M(C_0) \le c r(\mathcal{C}_0)$, we take $r_3 = \sqrt{c} c' \sqrt{r(\mathcal{C}_0)} + c''(\sqrt{c}\sqrt{r(\mathcal{C}_0)} \vee r_2(\Gamma))$, and we are done. ∎

The results in [26] also prove high probability bounds for the error in the posterior means and covariances, which are stronger than the expectation bounds used here. High probability bounds can be just as easily obtained within our computational framework, but we elect to present the expectation bounds for simplicity of exposition.

4.3. Effective dimension of prior covariance. The results in the previous subsection relied on the assumption that the effective dimension of the discretized covariance operator in the weighted inner product space is bounded above independently of the discretization level for sufficiently large n. We will now show that this assumption does indeed hold for the finite element covariance discretization given in subsection 2.3.1. To get an upper bound for 4.6, we need an upper bound for $\mathrm{Tr}(C_0)$ and a lower bound for $\|C_0\|_M$. The following result uses classical eigenvalue approximation results and our operator-norm convergence result to derive such a bound.

Theorem 4.3 (effective dimension upper bound). *Let the assumptions of Theorem 3.6 hold. Then, for h sufficiently small, there exists a constant τ independent of h such that*

$$r_M(C_0) \le \frac{Tr(\mathcal{C}_0)}{\|\mathcal{C}_0\|_{op} - \tau h^2}. \tag{4.12}$$

Consequently, for any constant $c > 1$, taking $h \le \sqrt{\tau(1-\frac{1}{c})\|\mathcal{C}_0\|_{op}}$ guarantees that

$$r_M(C_0) \le c r(\mathcal{C}_0). \tag{4.13}$$

Proof. By Theorem 3.6 we have that

$$\|\mathcal{C}_0 - P^* C_0 P\|_{op} \le \tau h^2$$

for some constant τ independent of h. Applying the reverse triangle inequality to this inequality gives us that

$$\|P^* C_0 P\|_{op} \ge \|\mathcal{C}_0\|_{op} - \tau h^2.$$

Then, using the fact that the $\mathbb{R}^n_M$ matrix operator norm coincides with the $L^2(\mathcal{D})$ operator norm of any mapping in the weighted inner product space, we get that

$$\|C_0\|_{op} \ge \|\mathcal{C}_0\|_{op} - \tau h^2. \tag{4.14}$$

To upper bound the trace of C_0 we proceed to bound the eigenvalues of C_0. Recall that $C_0 = A^{-\alpha}$. We let $\{\lambda_h^{(i)}\}_{i=1}^n$ denote the eigenvalues of A, and let $\{\lambda^{(i)}\}_{i=1}^\infty$ denote the eigenvalues of the continuum differential operator $\mathcal{A}$, both in ascending order. To characterize these eigenvalues, we can use classical finite element eigenvalue estimates. In particular, we refer the reader to [54, Theorem 6.1], which proves that the eigenvalues of the finite element approximation overestimate the eigenvalues of the continuum differential operator. That is, for each $i = 1, \dots, n$ we have that $\lambda^{(i)} \le \lambda_h^{(i)}$. It then follows that

$$\mathrm{Tr}(C_0) = \sum_{i=1}^n (\lambda_h^{(i)})^{-\alpha} \le \sum_{i=1}^n (\lambda^{(i)})^{-\alpha} \le \sum_{i=1}^\infty (\lambda^{(i)})^{-\alpha} = \mathrm{Tr}(\mathcal{C}_0). \tag{4.15}$$

Combining (4.14) and (4.15) gives (4.12), completing the proof. ∎

Remark 4.4. In the graph-based approximation setting, one must work slightly harder to derive such an upper bound on the effective dimension, as the eigenvalues of the graph

Laplacian are not guaranteed to overestimate the eigenvalues of the Laplace–Beltrami operator. Further, the eigenvalue estimates in Proposition C.1 only hold for eigenvalues that satisfy $h_n\sqrt{\lambda^{(k)}} \ll 1$, which typically only holds up to some $k<n$. As such, it may be most natural to consider a truncated prior covariance that retains only the portion of the spectrum that provably approximates the continuum operator, as is done in [23].

5. Convergence of MAP estimators under mesh refinement.

For many inverse problems of interest, the forward model is not linear. In this section, we will see how our computational framework and analysis can be leveraged to guarantee the convergence of the discretized maximum a posteriori (MAP) estimators to their continuum counterparts in nonlinear Bayesian inverse problems as the mesh is refined. In particular, our framework very naturally fits into the theory developed in [3], from which we will show Γ-convergence of the relevant Onsager–Machlup (OM) functions and consequently convergence of the MAP estimators (up to subsequences). Under reasonable conditions on the nonlinear forward model (see Assumption 2.7 in [55]), the posterior can still be characterized as a change of measure with respect to the prior as in (2.4),

$$\frac{d\mu_{\text{post}}}{d\mu_0}(u) = \frac{1}{Z}\exp\left(-\frac{1}{2}\|y-\mathcal{F}(u)\|^2_{\Gamma^{-1}}\right), \tag{5.1}$$

where $Z = \int_{\mathcal{H}} \exp\left(-\Phi(u)\right) d\mu_0(u)$. For a general $\mathcal{F}$, this posterior measure will no longer be Gaussian. As such, it is difficult to fully characterize the posterior measure. One particularly useful point summary of the posterior measure is a MAP estimator. In infinite dimensions, the notion of a MAP estimator was introduced in [19] as the maximizer of a small ball probability. The theory of MAP estimation in function spaces has been further refined in [3, 34, 35, 37, 38]. For $u \in \mathcal{H}$, we let $B^\delta(u) \subset \mathcal{H}$ be the open ball centered at u with radius δ. A *MAP estimator* (or *strong mode*) for μ_{post} is any point $u^{\text{MAP}} \in \mathcal{H}$ satisfying

$$\lim_{\delta\to 0} \frac{\mu_{\text{post}}\left(B^\delta(u^{\text{MAP}})\right)}{M_\delta} = 1, \qquad M_\delta = \sup_{u\in\mathcal{H}} \mu_{\text{post}}(B^\delta(u)). \tag{5.2}$$

It is shown in [37] that the MAP estimators of (5.1) with a Gaussian prior $\mu_0 = \mathcal{N}(m_0, \mathcal{C}_0)$ in a separable Hilbert space are precisely characterized by the minimizers of the OM functional $I_{\text{post}} : \mathcal{H} \to \mathbb{R}$ given by

$$I_{\text{post}}(u) = \begin{cases} \Phi(u) + I_0(u) & \text{if } u - m_0 \in \operatorname{Im}(\mathcal{C}_0^{1/2}), \\ +\infty & \text{otherwise,} \end{cases} \tag{5.3}$$

where

$$\Phi(u) = \frac{1}{2}\|y-\mathcal{F}(u)\|^2_{\Gamma^{-1}} \quad \text{and} \quad I_0(u) = \frac{1}{2}\|\mathcal{C}_0^{-1/2}(u-m_0)\|^2_{\mathcal{H}}.$$

In more general settings, the minimizers of (5.3) may not coincide with MAP estimators defined as in (5.2), and a weaker notion of MAP estimator is required [29]. The optimization problem of minimizing (5.3) is in general difficult, if not impossible, to solve analytically. The computational framework put forth in subsection 2.2 provides a tractable finite-dimensional optimization problem that provably approximates the infinite-dimensional problem. The discretized posterior measure in (2.11) gives the following OM functional:

$$I^{(n)}_{\text{post}}(u) = \begin{cases} \Phi^{(n)}(u) + I_0^{(n)}(u) & \text{if } u - m \in \operatorname{Im}((P^* C_0 P)^{1/2}), \\ +\infty & \text{otherwise,} \end{cases} \tag{5.4}$$

where

$$\Phi^{(n)}(u) = \frac{1}{2}\|y - F(Pu)\|^2_{\Gamma^{-1}} \quad \text{and} \quad I_0^{(n)} = \frac{1}{2}\|(P^* C_0 P)^{\dagger/2}(u - m_0)\|^2_{\mathcal{H}}.$$

Here $F : \mathbb{R}^n_M \to \mathbb{R}^k$ is a discretized approximation to $\mathcal{F}$, and the operator $(P^* C_0 P)^{\dagger/2} = \sum_{i=1}^n (\lambda_n^{(i)})^{-\alpha/2} P^* \psi_n^{(i)} \otimes P^* \psi_n^{(i)}$ is the Moore–Penrose pseudoinverse of $(P^* C_0 P)^{1/2}$. We remark that $u - m \in \operatorname{Im}((P^* C_0 P)^{1/2})$ if and only if $u, m \in \mathcal{V}$, so we can equivalently write (5.4) as $I^{(n)}_{\text{post}} : \mathbb{R}^n_M \to \mathbb{R}$, with

$$I^{(n)}_{\text{post}}(\vec{u}) = \frac{1}{2}\|y - F(\vec{u})\|^2_{\Gamma^{-1}} + \frac{1}{2}\|C_0^{-1/2}(\vec{u} - \vec{m}_0)\|^2_M, \tag{5.5}$$

which can be minimized by methods from the breadth of literature on finite-dimensional optimization problems. To apply the results in [3], we first review some preliminary definitions and results regarding Γ-convergence.

Definition 5.1. *Let $I, I_n : \mathcal{H} \to \overline{\mathbb{R}}$. We say that I_n Γ-converges to I, or $\Gamma\text{-}\lim_{n\to\infty} I_n = I$, if, for every $u \in \mathcal{H}$, the following two conditions hold:*

(a) (Γ-lim inf Inequality.) *For every sequence $\{u_n\}_{n=1}^\infty$ converging to u in $\mathcal{H}$,*

$$I(u) \le \liminf_{n\to\infty} I_n(u_n). \tag{5.6}$$

(b) (Γ-lim sup Inequality.) *There exists a sequence $\{u_n\}_{n=1}^\infty$ converging to u such that*

$$I(u) \ge \limsup_{n\to\infty} I_n(u_n). \tag{5.7}$$

Definition 5.2. *A sequence of functionals $\{I_n\}_{n=1}^\infty$ is equicoercive if, for all $t \in \mathbb{R}$, there exists a compact subset $K_t \subseteq \mathcal{H}$ such that, for all n, $I_n^{-1}([-\infty, t]) \subseteq K_t$.*

Definition 5.3. *Let $I, I_n : \mathcal{H} \to \overline{\mathbb{R}}$. We say that I_n converges continuously to I if, for every $u \in \mathcal{H}$ and every neighborhood V of $I(u)$ in $\overline{\mathbb{R}}$, there exist $N \in \mathbb{N}$ and a neighborhood U of u such that $n \ge N$ and $u' \in U$ imply that $I_n(u') \in V$.*

Continuous convergence is a stronger notion of convergence than pointwise convergence and Γ-convergence, but it is implied by uniform convergence of I_n to I in the case when I is continuous [42]. The following classical result can be found, for instance, in [8].

Proposition 5.4. *Let $I_n, I : \mathcal{H} \to \overline{\mathbb{R}}$ with $\Gamma\text{-}\lim_{n\to\infty} I_n = I$, and let $\{I_n\}_{n=1}^\infty$ be equicoercive. Then,*

$$\min_{\mathcal{H}} I = \lim_{n\to\infty} \inf_{\mathcal{H}} I_n, \tag{5.8}$$

and if $\{u_n\}_{n=1}^{\infty}$ is a precompact sequence such that $\lim_{n\to\infty} I_n(u_n) = \min_{\mathcal{H}} I$, then every limit of a convergent subsequence of $\{u_n\}_{n=1}^{\infty}$ is a minimizer of I. In particular, if each I_n has a minimizer u_n, then any convergent subsequence of these minimizers is a minimizer of I.

Essentially, this proposition states that Γ-convergence is the correct notion of convergence of functionals to guarantee that their minimizers also converge. We now use our results from subsection 3.2 and the results in [3] to prove convergence of the finite element discretized MAP estimators as the mesh is refined.

Theorem 5.5 (convergence of MAP estimators). *Let the covariance operator $\mathcal{C}_0 : \mathcal{H} \to \mathcal{H}$ and its discrete approximation $C_0 : \mathbb{R}^n_M \to \mathbb{R}^n_M$ satisfy Assumption 3.1, and let the orthogonal projection operator P satisfy Assumption 3.1. Assume additionally that the approximate forward model is such that $\Phi^{(n)}$ converges to Φ continuously as $n \to \infty$. Then, the corresponding sequence $\{I^{(n)}_{post}\}_{n=1}^{\infty}$ of OM functionals given in (5.5) satisfies that*

$$\Gamma\text{-}\lim_{n\to\infty} I^{(n)}_{post} = I_{post}, \tag{5.9}$$

and the cluster points as $n \to \infty$ of the MAP estimators of the posteriors μ^n_{post} are MAP estimators of the continuum posterior μ_{post}.

Proof. Assumption 3.1(i) and (iii) imply that, as $n \to \infty$, P^*C_0P converges to $\mathcal{C}_0$ in operator norm, and P^*Pm_0 converges to m_0 in $\mathcal{H}$. Therefore, we can apply Theorem 5.5 in [3] to conclude that $I_0 = \Gamma\text{-}\lim_{n\to\infty} I_0^{(n)}$ and that the sequence $\{I_0^{(n)}\}_{n=1}^{\infty}$ is equicoercive. Then, under the assumption that $\Phi^{(n)} \to \Phi$ continuously as $n \to \infty$, applying Theorem 6.1 in [3] gives us that $\Gamma\text{-}\lim_{n\to\infty} I^{(n)}_{\text{post}} = I_{\text{post}}$, and the sequence $\{I^{(n)}_{\text{post}}\}_{n=1}^{\infty}$ is also equicoercive since $\Phi^{(n)} \geq 0$ in our setting. Thus, by Proposition 5.4 we conclude that if u_n is a minimizer of $I^{(n)}_{\text{post}}$, $n \geq 1$, then any convergent subsequence of $\{u_n\}_{n=1}^{\infty}$ converges to a minimizer of I_{post}. Since the minimizers of the OM functionals $I^{(n)}_{\text{post}}$ and I_{post} coincide with the MAP estimators of $\mu^{(n)}_{\text{post}}$ and μ_{post}, respectively, we have hence shown that any convergent subsequence of MAP estimators of the discretized posterior converges to a MAP estimator of the continuum posterior. ∎

We have demonstrated how Assumption 3.1(i) and (iii) can be verified for finite element and graph-based discretizations. For nonlinear forward maps, verifying the continuous convergence of the discretized potential may not be straightforward. Lemma B.9 in [3] shows that if the approximate forward map is of the form $\mathcal{F}(P^*P\cdot)$, then the corresponding potentials converge continuously. The following example illustrates an important problem where the approximate forward model is not of this form; nonetheless, we will show in Theorem 5.7 that continuous convergence—as well as Assumption 3.1(i) and (iii)—can still be verified.

Example 5.6 (Eulerian data assimilation for the Navier–Stokes equations). Eulerian data assimilation is concerned with learning the initial condition of a dynamical system from pointwise observations of the state at fixed spatial locations. This example briefly reviews Eulerian data assimilation for the Navier–Stokes equations, an important model problem in numerical weather forecasting. We refer the reader to [18, 17] for a more detailed discussion.

As described in [48, 59], the Navier–Stokes equations on the two-dimensional torus $\mathcal{D} = \mathbb{T}^2 = [0,1]^2$ can be written as an infinite-dimensional dynamical system

$$\frac{dv}{dt} = \nu \mathsf{A} v + \mathsf{B}(v,v) = f, \quad v(0) = u, \tag{5.10}$$

on the Hilbert space

$$\mathcal{H} = \left\{ v \in L^2_{\text{per}}(\mathcal{D}) : \int_{\mathcal{D}} v\, dx = 0, \nabla \cdot v = 0 \right\} \tag{5.11}$$

equipped with the standard $L^2(\mathcal{D})$ inner product. The Stokes operator A in (5.10) is self-adjoint, positive, and densely defined on $\mathcal{H}$ and has a complete set of eigenpairs, $\{(\lambda^{(i)}, \psi^{(i)})\}_{i=1}^{\infty}$, with increasingly sorted eigenvalues. The energy-conserving quadratic nonlinearity $\mathsf{B}(v,v)$ arises from projection under the Leray operator [58]. For $u \in \mathcal{H}$ and f sufficiently regular [18, 48], there exists a unique solution to (5.10) such that $v \in L^\infty(0,T;\mathcal{H}^{1+s}) \subset L^\infty(0,T;L^\infty(\mathcal{D}))$.

We wish to determine the initial condition u from noisy observations of the velocity field v at time $t > 0$ and fixed spatial locations $x_1, \dots, x_K \in \mathcal{D}$, given by

$$y_k = v(x_k, t) + \eta_k, \quad k = 1, \dots, K, \tag{5.12}$$

where $\eta \sim \mathcal{N}(0, \Gamma)$. This Eulerian data assimilation task can be formulated as an inverse problem with nonlinear forward map

$$\mathcal{F}(u) = \left(v(x_1,t)^T, \dots, v(x_K,t)^T \right)^T \in \mathbb{R}^{d_y}, \tag{5.13}$$

where $d_y = 2K$. We consider a Gaussian prior measure $\mu_0 \sim \mathcal{N}(m_0, \mathcal{C}_0)$, where $\mathcal{C}_0 = \mathsf{A}^{-\alpha}$ with $\alpha > 1$ and $m_0 \in \mathcal{H}^\alpha(\mathcal{D})$. Here, powers of A are defined spectrally. The posterior measure can then be characterized as in (5.1).

We discretize the inverse problem on the space spanned by the first n eigenfunctions $\{\psi^{(i)}\}_{i=1}^n$ of the Stokes operator. Note that since the eigenfunctions are orthonormal, the weighted inner product described in subsection 2.2 coincides with the usual Euclidean inner product. The discretized prior covariance operator $C_0 : \mathbb{R}^n \to \mathbb{R}^n$ is then given by $C_0 = diag\left((\lambda^{(1)})^{-\alpha}, \dots, (\lambda^{(n)})^{-\alpha} \right)$. The discretized forward map is given by a Galerkin approximation to the PDE solution, defined via the projection $P : \mathcal{H} \to \mathbb{R}^n$ given in (2.7). We denote by v^N the solution to the (finite-dimensional) ODE

$$\frac{dv^N}{dt} + \nu \mathsf{A} v^N + P\mathsf{B}(v^N, v^N) = Pf, \quad v^N(0) = Pu. \tag{5.14}$$

We then define

$$F(u) = \left(v^N(x_1,t), \dots, v^N(x_{d_y},t) \right)^T. \tag{5.15}$$

Given these approximations, we have a discretized OM functional exactly in the form of (5.5).

For the Eulerian data assimilation problem summarized in Example 5.6 and further detailed in [18], we can verify the assumptions of Theorem 5.5 to obtain the following result. The proof can be found in Appendix D.

Theorem 5.7 (MAP estimation for Eulerian data assimilation). *Let $f \in L^2(0,T;\mathcal{H}^s)$ with $s > 0$, and consider the Eulerian data assimilation problem summarized in Example 5.6. Then, as $n \to \infty$, the sequence $\{I^{(n)}_{post}\}_{n=1}^{\infty}$ of discretized OM functionals satisfies that*

$$\Gamma\text{-}\lim_{n\to\infty} I^{(n)}_{\text{post}} = I_{\text{post}}, \tag{5.16}$$

and the cluster points of the MAP estimators of the posteriors μ^n_{post} are MAP estimators of the continuum posterior μ_{post}.

6. Conclusion and future directions. This paper analyzed a computational framework for solving infinite-dimensional Bayesian inverse problems. Working on a weighted inner product space, we have studied finite element and graph-based discretizations in a unified framework, using Matérn-type priors and deconvolution forward models as guiding examples. We have established error guarantees for linear inverse problems and analyzed ensemble Kalman algorithms and MAP estimators applicable in nonlinear inverse problems. In future work, our analysis may be extended to other types of discretizations, priors, and forward models. The generality of our presentation will facilitate the integration of numerical analysis for PDEs to obtain similar error guarantees for other priors and forward models. Additionally, our general approach for obtaining error bounds for discretizations of covariance operators and forward maps will also facilitate the analysis under mesh refinement of other algorithms for nonlinear inverse problems.

Appendix A. Adjoints in weighted space. Let X and Y be two Hilbert spaces. Recall that [56], given $A \in B(X,Y)$, the adjoint of A is the unique map $A^* \in B(Y,X)$ such that

$$\langle Ax, y\rangle_Y = \langle x, A^* y\rangle_X \qquad \forall x \in X, \forall y \in Y. \tag{A.1}$$

In $\mathbb{R}^n$ equipped with the usual Euclidean inner product, the adjoint of a matrix coincides with the matrix transpose. However, for linear maps to and from the weighted inner product space $\mathbb{R}^n_M$, the structure of the adjoint specified by (A.1) no longer coincides with the matrix transpose. For the case where $A \in B(\mathbb{R}^n_M, \mathbb{R}^n_M)$, we have that A^* must satisfy

$$\vec{u}^T A^T M \vec{v} = \langle A\vec{u}, \vec{v}\rangle_M = \langle \vec{u}, A^*\vec{v}\rangle_M = \vec{u}^T M A^* \vec{v} \qquad \forall \vec{u}, \vec{v} \in \mathbb{R}^n_M,$$

from which it follows that $A^* = M^{-1}A^T M$. Similarly, for $F \in B(\mathbb{R}^n_M, \mathbb{R}^{d_y})$, $F^\natural$ must satisfy

$$\vec{u}^T F^T y = \langle F\vec{u}, y\rangle = \langle \vec{u}, F^\natural y\rangle_M = \vec{u}^T M F^\natural y \qquad \forall \vec{u} \in \mathbb{R}^n_M,\ \forall y \in \mathbb{R}^{d_y},$$

from which it follows that $F^\natural = M^{-1}F^T$.

Appendix B. Proofs of subsection 3.2. This appendix contains the proofs of Theorems 3.6 and 3.7 in subsection 3.2. The proofs rely on classical finite element method convergence results from [44].

Proof of Theorem 3.6. We let $f \in L^2(\mathcal{D})$ with $\|f\|_{L^2(\mathcal{D})} \neq 0$. We proceed by induction on α. For $\alpha = 1$ we have that

$$\|\mathcal{A}^{-1} f - P^* A^{-1} P f\|_{L^2(\mathcal{D})} \le ch^2 \|\mathcal{A}^{-1} f\|_{H^2(\mathcal{D})} \le ch^2 \|\mathcal{A}^{-1}\|_{op} \|f\|_{L^2(\mathcal{D})}$$

by [9, Theorem 14.3.3] and boundedness of the PDE solution operator from $L^2(\mathcal{D})$ to $H^2(\mathcal{D})$ (see [20, Chapter 6.3]). Dividing both sides by $\|f\|_{L^2(\mathcal{D})}$, we have that

$$\|\mathcal{A}^{-1} - P^* A^{-1} P\|_{op} \leq ch^2$$

for a constant c independent of h, completing the base case. Now, we assume that for some integer $\alpha \geq 1$, we have that

$$\|\mathcal{A}^{-\alpha} - P^* A^{-\alpha} P\|_{op} \leq ch^2.$$

We then consider the quantity $\|\mathcal{A}^{-(\alpha+1)} f - P^* A^{-\alpha-1} P f\|_{L^2(\mathcal{D})}$. Then, by the triangle inequality, we have that

$$\begin{aligned}\|\mathcal{A}^{-\alpha-1} f - P^*(K^{-1}M)^{\alpha+1} P f\|_{L^2(\mathcal{D})} \leq &\|\mathcal{A}^{-\alpha}(\mathcal{A}^{-1} - P^* A^{-1} P) f\|_{L^2(\mathcal{D})} \\ &+ \|(\mathcal{A}^{-\alpha} - P^* A^{-\alpha} P) P^* A^{-1} P f\|_{L^2(\mathcal{D})}.\end{aligned}$$

The first of these terms can be controlled using the boundedness of $\mathcal{A}^{-1}$ as in the proof of the base case,

$$\|\mathcal{A}^{-\alpha}(\mathcal{A}^{-1} - P^* A^{-1} P) f\|_{L^2(\mathcal{D})} \leq \|\mathcal{A}^{-1}\|_{op}^{\alpha} \|\mathcal{A}^{-1} f - P^* A^{-1} P f\|_{L^2(\mathcal{D})} \leq c \|\mathcal{A}^{-1}\|_{op}^{\alpha} h^2 \|f\|_{L^2(\mathcal{D})},$$

and the second follows from the submultiplicativity of the operator norm and the inductive hypothesis,

$$\begin{aligned}\|(\mathcal{A}^{-\alpha} - P^* A^{-\alpha} P) P^* A^{-1} P f\|_{L^2(\mathcal{D})} &\leq \|\mathcal{A}^{-\alpha} - P^* A^{-\alpha} P\|_{op} \|P^* A^{-1} P f\|_{L^2(\mathcal{D})} \\ &\leq c \|\mathcal{A}^{-1}\|_{op} h^2 \|f\|_{L^2(\mathcal{D})}.\end{aligned}$$

We conclude that

$$\|\mathcal{A}^{-\alpha-1} - P^* A^{-\alpha} P\|_{op} \lesssim h^2,$$

completing the proof. ■

Proof of Theorem 3.7. Let $u \in L^2(\mathcal{D})$ with $\|u\|_{L^2(\mathcal{D})} \neq 0$. We denote $v = \mathcal{G}u$ and $v_h = GPu$. We then have that

$$\|\mathcal{F}u - FPu\|_2 \leq \|\mathcal{O}\|_{op} \|v - v_h\|_{L^2(\mathcal{D})} \leq \|\mathcal{O}\|_{op} \left(c_1 h^2 \|v\|_{H^2(\mathcal{D})} + c_2 \Delta t^2 \|Pu\|_{L^2(\mathcal{D})} \right),$$

where the first inequality uses the fact that $\mathcal{O}$ is a bounded linear operator from $L^2(\mathcal{D})$, and the second uses [44, Theorem 9.6], which is a classical finite element error result for the Crank–Nicolson method. Then, since $\mathcal{G} : L^2(\mathcal{D}) \to H^2(\mathcal{D})$ is bounded and $\|P\|_{op} = 1$, we get that

$$\|\mathcal{F}u - FPu\|_2 \leq c_1 \|\mathcal{O}\|_{op} \|\mathcal{G}\|_{op} h^2 \|u\|_{L^2(\mathcal{D})} + c_2 \|\mathcal{O}\|_{op} \Delta t^2 \|u\|_{L^2(\mathcal{D})}.$$

Dividing both sides by $\|u\|_{L^2(\mathcal{D})}$ gives the desired result. ■

Appendix C. Proofs of subsection 3.3. This appendix contains the proofs of Theorems 3.9 and 3.11 in subsection 3.3. The proofs rely on existing spectral convergence results from [51]. We denote by $\{(\lambda_n^{(i)}, \psi_n^{(i)})\}_{i=1}^n$ and $\{(\lambda^{(i)}, \psi^{(i)})\}_{i=1}^\infty$ the eigenpairs of the matrix A defined in (2.36) and the operator $\mathcal{A}$ defined in (2.17), with the eigenvalues in ascending order. We recall the spectral convergence results derived in [51].

Proposition C.1. *Suppose $k = k_n$ is such that $h_n\sqrt{\lambda^{(k)}} \ll 1$ for large n. Then,*

$$\frac{|\lambda_n^{(k)} - \lambda^{(k)}|}{\lambda^{(k)}} \leq c\left(\frac{\varepsilon_n}{h_n} + h_n\sqrt{\lambda^{(k)}}\right), \tag{C.1}$$

where c is a constant depending on $\mathcal{M}$, Θ, and b.

Proposition C.2. *Let λ be an eigenvalue of $\mathcal{A}$ with multiplicity ℓ. Suppose that $h_n\sqrt{\lambda^{(k_n)}} \ll 1$ and $\varepsilon_n \ll h_n$ for n large. Let $\psi_n^{(k_n)}, \ldots, \psi_n^{(k_n+\ell-1)}$ be orthonormal eigenvectors of A associated with eigenvalues $\lambda_n^{(k_n)}, \ldots, \lambda_n^{(k_n+\ell-1)}$. Then, there exist orthonormal eigenfunctions $\psi^{(k_n)}, \ldots, \psi^{(k_n+\ell-1)}$ of $\mathcal{A}$ so that, for $j = k_n, \ldots, k_n+\ell-1$,*

$$\|P^*\psi_n^{(j)} - \psi^{(j)}\|_{L^2(\mathcal{M})} \leq cj^{3/2}\left(\frac{\varepsilon_n}{h_n} + h_n\sqrt{\lambda^{(j)}}\right)^{1/2}, \tag{C.2}$$

where c is a constant depending on $\mathcal{M}$, Θ, and b.

Proof of Theorem 3.9. Let $f \in L^2(\mathcal{M})$. Note that k_n was chosen such that we can apply (C.2) and (C.1) to quantify the errors in the eigenvalues and eigenvectors up to $i = k_n$. We denote

$$u = \mathcal{C}_0 f = \mathcal{A}^{-\alpha} f = \sum_{i=1}^{\infty} (\lambda^{(i)})^{-\alpha} \langle f, \psi^{(i)}\rangle_{L^2(\mathcal{M})} \psi^{(i)},$$
$$u_n = P^*(\Delta_n^\Theta + B_n)^{-\alpha} P f = \sum_{i=1}^{n} (\lambda_n^{(i)})^{-\alpha} \langle Pf, \psi_n^{(i)}\rangle_M P^*\psi_n^{(i)}.$$

We remark that $\langle Pf, \psi_n^{(i)}\rangle_M = \langle f, P^*\psi_n^{(i)}\rangle_{L^2(\mathcal{M})}$, so we can also write

$$u_n = \sum_{i=1}^{n} (\lambda_n^{(i)})^{-\alpha} \langle f, P^*\psi_n^{(i)}\rangle_{L^2(\mathcal{M})} P^*\psi_n^{(i)}.$$

We want to bound $\|u - u_n\|_{L^2(\mathcal{M})}$. To do so, we introduce four intermediate quantities

$$u_n^{k_n} = \sum_{i=1}^{k_n} (\lambda_n^{(i)})^{-\alpha} \langle f, P^*\psi_n^{(i)}\rangle_{L^2(\mathcal{M})} P^*\psi_n^{(i)},$$
$$\tilde{u}_n^{k_n} = \sum_{i=1}^{k_n} (\lambda^{(i)})^{-\alpha} \langle f, P^*\psi_n^{(i)}\rangle_{L^2(\mathcal{M})} P^*\psi_n^{(i)},$$
$$\tilde{u}^{k_n} = \sum_{i=1}^{k_n} (\lambda^{(i)})^{-\alpha} \langle f, P^*\psi_n^{(i)}\rangle_{L^2(\mathcal{M})} \psi^{(i)},$$

$$u^{k_n} = \sum_{i=1}^{k_n} (\lambda^{(i)})^{-\alpha} \langle f, \psi^{(i)} \rangle_{L^2(\mathcal{M})} \psi^{(i)}.$$

We will bound the difference between each pair of consecutive functions. By Weyl's law [13, Theorem 72], we have that

$$\text{(C.3)} \quad \|u - u^{k_n}\|_{L^2(\mathcal{M})} \leq \left(\sum_{i=k_n+1}^{\infty} (\lambda^{(i)})^{-2\alpha} \right)^{\frac{1}{2}} \lesssim \left(\sum_{i=k_n+1}^{\infty} i^{-\frac{4\alpha}{d}} \right)^{\frac{1}{2}} \lesssim \left(\int_{k_n}^{\infty} x^{-\frac{4\alpha}{d}} \right)^{\frac{1}{2}} \lesssim k_n^{\frac{1}{2} - \frac{2\alpha}{d}}.$$

By (C.1) and Weyl's law, we have that $\lambda_n^{(k_n)} \gtrsim \lambda^{(k_n)} \gtrsim k_n^{\frac{2}{d}}$, from which we get

$$\text{(C.4)} \qquad \|u_n - u_n^{k_n}\|_{L^2(\mathcal{M})} = \left(\sum_{i=k_n+1}^{n} (\lambda_n^{(i)})^{-2\alpha} |\langle f, P^* \psi_n^{(i)} \rangle_{L^2(\mathcal{M})}|^2 \right)^{\frac{1}{2}} \leq (\lambda_n^{(k_n)})^{-\alpha} \lesssim k_n^{-\frac{2\alpha}{d}}.$$

Next, we note that since $\lambda_n^{(i)}$ and $\lambda^{(i)}$ are bounded from below by $\min_{x \in \mathcal{M}} b(x) > 0$, and $x^{-\alpha}$ is continuously differentiable away from zero, the mean value theorem guarantees that

$$\left| (\lambda_n^{(i)})^{-\alpha} - (\lambda^{(i)})^{-\alpha} \right| \leq \alpha |\xi|^{-\alpha-1} \left| \lambda_n^{(i)} - \lambda^{(i)} \right|$$

for some ξ between $\lambda_n^{(i)}$ and $\lambda^{(i)}$. This inequality, combined with Proposition C.1, implies that

$$\left| (\lambda_n^{(i)})^{-\alpha} - (\lambda^{(i)})^{-\alpha} \right| \leq \alpha \left(\lambda_n^{(i)} \wedge \lambda^{(i)} \right)^{-\alpha-1} \left| \lambda_n^{(i)} - \lambda^{(i)} \right| \lesssim (\lambda^{(i)})^{-\alpha} \left(\frac{\varepsilon_n}{h_n} + h_n \sqrt{\lambda^{(i)}} \right)$$

for $i = 1, \ldots, k_n$. Using this fact, we get that

(C.5)

$$\|u_n^{k_n} - \tilde{u}_n^{k_n}\|_{L^2(\mathcal{M})} \leq \left(\sum_{i=1}^{k_n} \left[(\lambda_n^{(i)})^{-\alpha} - (\lambda^{(i)})^{-\alpha} \right]^2 \right)^{\frac{1}{2}} \lesssim \left(\sum_{i=1}^{k_n} (\lambda^{(i)})^{-2\alpha} \left(\frac{\varepsilon_n}{h_n} + h_n \sqrt{\lambda^{(i)}} \right)^2 \right)^{\frac{1}{2}}$$
$$\lesssim \left(\frac{\varepsilon_n}{h_n} + h_n \right) \left(\sum_{i=1}^{k_n} (\lambda^{(i)})^{-2\alpha+1} \right)^{\frac{1}{2}} \lesssim \left(\frac{\varepsilon_n}{h_n} + h_n \right),$$

where the last step follows if $\alpha > \frac{d}{4} + \frac{1}{2}$. This is implied by the requirement that $\alpha > (5d+1)/4$. Next, we use (C.2) to deduce that

(C.6)

$$\|\tilde{u}^{k_n} - \tilde{u}_n^{k_n}\|_{L^2(\mathcal{M})} \lesssim \sum_{i=1}^{k_n} (\lambda^{(i)})^{-\alpha} \|\psi^{(i)} - P^* \psi_n^{(i)}\|_{L^2(\mathcal{M})} \lesssim \sum_{i=1}^{k_n} (\lambda^{(i)})^{-\alpha} i^{\frac{3}{2}} \sqrt{\frac{\varepsilon_n}{h_n} + h_n \sqrt{\lambda^{(i)}}}$$
$$\lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n} \sum_{i=1}^{k_n} i^{\frac{3}{2}} (\lambda^{(i)})^{-\alpha + \frac{1}{4}} \lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n} \sum_{i=1}^{k_n} i^{\frac{3}{2} - \frac{2\alpha}{d} + \frac{1}{2d}} \lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n},$$

where the last step follows if $\alpha > \frac{5}{4}d + \frac{1}{4}$. Finally, we use the Cauchy–Schwarz inequality to get that

$$\text{(C.7)} \quad \begin{aligned} \|u^{k_n} - \tilde{u}^{k_n}\|_{L^2(\mathcal{M})} &= \left\| \sum_{i=1}^{k_n} \langle f, \psi^{(i)} - P^* \psi_n^{(i)} \rangle_{L^2(\mathcal{M})} (\lambda^{(i)})^{-\alpha} \psi^{(i)} \right\|_{L^2(\mathcal{M})} \\ &\lesssim \sum_{i=1}^{k_n} (\lambda^{(i)})^{-\alpha} \|\psi^{(i)} - P^* \psi_n^{(i)}\|_{L^2(\mathcal{M})} \lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n}\,, \end{aligned}$$

where the last step follows by the same argument as for the previous expression if $\alpha > \frac{5}{4}d + \frac{1}{4}$. Combining (C.3), (C.4), (C.5), (C.6), and (C.7) and taking $k_n \asymp n^{\frac{2d+1}{4\alpha}}$, we see that the error is dominated by

$$\|\mathcal{C}_0 - P^*(\Delta_n^\Theta + B_n)^{-\alpha} P\|_{op} \lesssim k_n^{\frac{1}{2} - \frac{2\alpha}{d}} + \sqrt{\frac{\varepsilon_n}{h_n} + h_n} \lesssim (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}},$$

as desired. ∎

Proof of Theorem 3.11. Let $u \in L^2(\mathcal{M})$ with $\|u\|_{L^2(\mathcal{M})} \neq 0$. We write

$$\text{(C.8)} \qquad v = \mathcal{G}u = \sum_{i=1}^{\infty} \exp\left(-\lambda^{(i)}\right) \langle u, \psi^{(i)} \rangle_{L^2(\mathcal{M})} \psi^{(i)},$$

$$\text{(C.9)} \qquad v_n = G_n(u_n) = \sum_{i=1}^{n} \exp\left(-\lambda_n^{(i)}\right) \langle Pu, \psi_n^{(i)} \rangle_{L^2(\mathcal{M})} P^* \psi_n^{(i)}.$$

We then consider the quantity

$$\text{(C.10)} \qquad \|\mathcal{F}u - FPu\|_2 = \|\mathcal{O}v - \mathcal{O}^n v_n\|_2 \le \|\mathcal{O}\|_{op} \|v - v_n\|_{L^2(\mathcal{D})} + \|\mathcal{O} v_n \mathcal{O}^n v_n\|_{L^2(\mathcal{M})}.$$

We first want to bound $\|v - v_n\|_{L^2(\mathcal{M})}$. To do so, we introduce the following intermediate quantities:

$$\begin{aligned} v_n^{k_n} &= \sum_{i=1}^{k_n} \exp\left(-\lambda_n^{(i)}\right) \langle u, P^* \psi_n^{(i)} \rangle_{L^2(\mathcal{M})} P^* \psi_n^{(i)}, \\ \tilde{v}_n^{k_n} &= \sum_{i=1}^{k_n} \exp\left(-\lambda^{(i)}\right) \langle u, P^* \psi_n^{(i)} \rangle_{L^2(\mathcal{M})} P^* \psi_n^{(i)}, \\ \tilde{v}^{k_n} &= \sum_{i=1}^{k_n} \exp\left(-\lambda^{(i)}\right) \langle u, P^* \psi_n^{(i)} \rangle_{L^2(\mathcal{M})} \psi^{(i)}, \\ v^{k_n} &= \sum_{i=1}^{k_n} \exp\left(-\lambda^{(i)}\right) \langle u, \psi^{(i)} \rangle_{L^2(\mathcal{M})} \psi^{(i)}. \end{aligned}$$

We proceed to bound the difference between each pair of consecutive functions, as in the proof of Theorem 3.9. By Weyl's law, we have that

$$\|v - v^{k_n}\|_{L^2(\mathcal{M})} \le \left(\sum_{i=k_n+1}^{\infty} \exp\left(-\lambda^{(i)}\right) \right)^{\frac{1}{2}} \lesssim \left(\sum_{i=k_n+1}^{\infty} \exp\left(-i^{\frac{2}{d}}\right) \right)^{\frac{1}{2}} \lesssim \left(\int_{k_n}^{\infty} \exp\left(-x^{\frac{2}{d}}\right) dx \right)^{\frac{1}{2}}.$$

This integral does not have a nice closed form solution. However, for any fixed $\alpha > 0$, it holds that $\exp(-x^{2/d}) < x^{-4\alpha/d}$ for x large enough. Assuming that n is large enough such that this inequality holds for $x = k_n$, we then have the bound

$$\|v - v^{k_n}\|_{L^2(\mathcal{M})} \le \left(\int_{k_n}^{\infty} x^{-4\alpha/d}\, dx\right)^{\frac{1}{2}} \lesssim k_n^{\frac{1}{2}-\frac{2\alpha}{d}}. \tag{C.11}$$

While this bound is not necessarily sharp, the operator-norm error will still be dominated by the error in approximating the eigenfunctions, so the final bound will still be the same. By (C.1) and Weyl's law, we have that $\lambda_n^{(k_n)} \gtrsim \lambda^{(k_n)} \gtrsim k_n^{\frac{2}{d}}$, from which we get

$$\begin{aligned} \|v_n - v_n^{k_n}\|_{L^2(\mathcal{M})} &= \left(\sum_{i=k_n+1}^{n} \exp\left(-\lambda_n^{(i)}\right) |\langle u, P^*\psi_n^{(i)}\rangle_{L^2(\mathcal{M})}|^2\right)^{\frac{1}{2}} \\ &\le \exp\left(-\frac{1}{2}\lambda_n^{(k_n)}\right) \lesssim \exp\left(-\frac{1}{2}k_n^{\frac{2}{d}}\right). \end{aligned} \tag{C.12}$$

Next, we note that since e^{-x} is continuously differentiable, for each $i = 1, \dots, k_n$, we have that

$$\left|\exp(-\lambda_n^{(i)}) - \exp(-\lambda^{(i)})\right| \lesssim \exp\left(-(\lambda_n^{(i)} \wedge \lambda^{(i)})\right) \left|\lambda_n^{(i)} - \lambda^{(i)}\right| \lesssim \exp(-\lambda^{(i)}) \left(\frac{\varepsilon_n}{h_n} + h_n\sqrt{\lambda^{(i)}}\right).$$

From this, we get that

$$\begin{aligned} \|v_n^{k_n} - \tilde{v}_n^{k_n}\|_{L^2(\mathcal{M})} &\le \left(\sum_{i=1}^{k_n} \left|\exp\left(-\lambda_n^{(i)}\right) - \exp\left(-\lambda^{(i)}\right)\right|\right)^{\frac{1}{2}} \\ &\lesssim \left(\exp\left(-2\lambda^{(i)}\right)\left(\frac{\varepsilon_n}{h_n} + h_n\sqrt{\lambda^{(i)}}\right)^2\right)^{\frac{1}{2}} \\ &\lesssim \left(\frac{\varepsilon_n}{h_n} + h_n\right)\left(\sum_{i=1}^{k_n} \frac{\lambda^{(i)}}{\exp\left(2\lambda^{(i)}\right)}\right)^{\frac{1}{2}} \lesssim \left(\frac{\varepsilon_n}{h_n} + h_n\right). \end{aligned} \tag{C.13}$$

Next, we use (C.2) to deduce that

$$\begin{aligned} \|\tilde{v}^{k_n} - \tilde{v}_n^{k_n}\|_{L^2(\mathcal{M})} &\lesssim \sum_{i=1}^{k_n} \exp\left(-\lambda^{(i)}\right) \|\psi^{(i)} - P^*\psi_n^{(i)}\|_{L^2(\mathcal{M})} \\ &\lesssim \sum_{i=1}^{k_n} \exp\left(-\lambda^{(i)}\right) i^{\frac{3}{2}} \sqrt{\frac{\varepsilon_n}{h_n} + h_n\sqrt{\lambda^{(i)}}} \\ &\lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n} \sum_{i=1}^{k_n} i^{\frac{3}{2}} \exp\left(-\lambda^{(i)}\right) (\lambda^{(i)})^{\frac{1}{4}} \\ &\lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n} \sum_{i=1}^{k_n} \frac{i^{\frac{3}{2}+\frac{1}{2d}}}{\exp\left(i^{2/d}\right)} \lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n}. \end{aligned} \tag{C.14}$$

Finally, we use the Cauchy–Schwarz inequality to get that

$$\begin{aligned}\|v^{k_n} - \tilde{v}^{k_n}\|_{L^2(\mathcal{M})} &= \left\| \sum_{i=1}^{k_n} \langle u, \psi^{(i)} - P^* \psi_n^{(i)} \rangle_{L^2(\mathcal{M})} \exp\left(-\lambda^{(i)}\right) \psi^{(i)} \right\|_{L^2(\mathcal{M})} \\ &\lesssim \sum_{i=1}^{k_n} \exp\left(-\lambda^{(i)}\right) \|\psi^{(i)} - P^* \psi_n^{(i)}\|_{L^2(\mathcal{M})} \lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n}\,,\end{aligned} \tag{C.15}$$

where the last step follows by the same argument as for (C.14). Combining (C.11), (C.12), (C.13), (C.14), and (C.15) and taking $k_n \asymp n^{\frac{2d+1}{4\alpha}}$, we see that the error is dominated by

$$\|v - v_n\|_{op} \lesssim k_n^{\frac{1}{2} - \frac{2\alpha}{d}} + \sqrt{\frac{\varepsilon_n}{h_n} + h_n} \lesssim (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}}, \tag{C.16}$$

as desired. Taking the supremum of both sides over all $\|u\|_{L^2(\mathcal{M})} = 1$, we get that

$$\|\mathcal{G} - P^* G_n P\|_{op} \lesssim k_n^{\frac{1}{2} - \frac{2\alpha}{d}} + \sqrt{\frac{\varepsilon_n}{h_n} + h_n} \lesssim (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}}. \tag{C.17}$$

Now we proceed to bound $\|\mathcal{O} v_n - \mathcal{O}^n v_n\|_2$. Recall that, for $k = 1, \dots, d_y$, we have

$$[\mathcal{O} v_n]_k = \int_{B_\delta(x_k) \cap \mathcal{M}} \sum_{i=1}^{n} v_n(x) \mathbb{1}_{U_i}(x) d\gamma = \langle v_n, \mathbb{1}_{B_\delta(x_k) \cap \mathcal{M}}(x) \rangle_{L^2(\mathcal{M})}$$

and

$$[\mathcal{O}^n v_n]_k = \frac{1}{n} \sum_{i=1}^{n} v_n(x_i) \mathbb{1}_{B_\delta(x_k)}(x_i) = \langle P v_n, \mathbb{1}_{B_\delta(x_k) \cap \mathcal{M}_n} \rangle_M,$$

where $\mathbb{1}_{B_\delta(x_k) \cap \mathcal{M}_n} \in \mathbb{R}^n_M$ is the vector with ith entry 1 if $x_i \in B_\delta(x_k)$ for all $x_i \in M_n$. Note that here the points x_i are in the point cloud $\mathcal{M}_n$, whereas the x_k are the points at which the balls of radius δ in the observation operator are centered. By our definition of P, it follows that we can write

$$[\mathcal{O}^n v_n]_k = \langle v_n, P^* \mathbb{1}_{B_\delta(x_k) \cap \mathcal{M}_n}(x) \rangle_{L^2(\mathcal{M})}.$$

Consequently, we can apply the Cauchy–Schwarz inequality to get the bound

$$\begin{aligned}\|\mathcal{O} v_n - \mathcal{O}^n v_n\|_2^2 &= \sum_{k=1}^{d_y} \langle v_n, \mathbb{1}_{B_\delta(x_k) \cap \mathcal{M}}(x) - P^* \mathbb{1}_{B_\delta(x_k) \cap \mathcal{M}_n}(x) \rangle_{L^2(\mathcal{M})} \\ &\leq \sum_{k=1}^{d_y} \|v_n\|_{L^2(\mathcal{M})}^2 \|\mathbb{1}_{B_\delta(x_k) \cap \mathcal{M}} - P^* \mathbb{1}_{B_\delta(x_k) \cap \mathcal{M}_n}\|_{L^2(\mathcal{M})}^2.\end{aligned} \tag{C.18}$$

From (C.16), we have that

$$\|v_n\|_{L^2(\mathcal{M})} \leq \|v\|_{L^2(\mathcal{M})} + c(\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}} \leq \|\mathcal{G}\|_{op} \|u\|_{L^2(\mathcal{M})} + c(\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}}.$$

Hence, for n large enough it holds that

$$\|v_n\|_{L^2(\mathcal{M})} \leq 2\|\mathcal{G}\|_{op}\|u\|_{L^2(\mathcal{M})}. \tag{C.19}$$

It then remains to bound $\|\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}} - P^*\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}_n}\|^2_{L^2(\mathcal{M})}$. To do so, we can write

$$P^*\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}_n}(x) = \sum_{i=1}^n \mathbb{1}_{U_i}(x)\mathbb{1}_{B_\delta(x_k)}(x_i) = \sum_{i=1}^n \mathbb{1}_{U_i}(x)\mathbb{1}_{B_\delta(x_k)}(T_n(x)),$$

since $T_n(x) = x_i$ for all $x \in U_i$. Since the sets U_i partition $\mathcal{M}$, we have that

$$P^*\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}_n}(x) = \mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}}(T_n(x)).$$

We then see that $\|\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}} - P^*\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}_n}\|^2_{L^2(\mathcal{M})}$ is exactly the measure of the set where the indicator functions $\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}}(T_n(x))$ and $\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}}(x)$ are not equal. Thus, we simply need to bound the measure of the sets

$$E_k = \left\{x \in \mathcal{M} : x \in B_\delta(x_k), T_n(x) \notin B_\delta(x_k)\right\} \cup \left\{x \in \mathcal{M} : x \notin B_\delta(x_k), T_n(x) \in B_\delta(x_k)\right\}.$$

We remark that for any $x \in U_i \subset B_\delta(x_k)$, we also then have that $T_n(x) \in U_i \subset B_\delta(x_k)$. For $U_i \subset B_\delta(x_k)$, we have $U_i \subset E_k^C$. Similarly, for any $x \in U_i \subset B_\delta(x_k)^C$, we also have that $T_n(x) \in U_i$, so $x \notin B_\delta(x_k)$ and $T_n(x) \notin B_\delta(x_k)$. Hence, for $U_i \subset B_\delta(x_k)^C$, we have $U_i \subset E_k^C$. Since any U_i that are entirely within $B_\delta(x_k)$ and any U_i that are entirely outside of $B_\delta(x_k)$ are contained in the complement of E_k, and the sets U_i partition $\mathcal{M}$, we must have that E_k is contained in the union of the sets U_i that have nontrivial intersection with both $B_\delta(x_k)$ and $B_\delta(x_k)^C$. That is, E_k is contained in the union of the sets U_i that intersect the boundary of $B_\delta(x_k)\cap\mathcal{M}$,

$$E_k \subset \tilde{E}_k = \bigcup_{i : U_i \cap \partial(B_\delta(x_k)\cap\mathcal{M})} U_i.$$

From Proposition 2.5, we have that $\sup_{x\in\mathcal{M}} d_{\mathcal{M}}(x, T_n(x)) = \varepsilon_n$. That is, the geodesic distance between any two points in a set U_i is at most $2\varepsilon_n$. This fact, combined with the assumption that $\operatorname{length}(\partial(B_\delta(x_k)\cap\mathcal{M})) = C_{\delta,k}$, gives us that

$$\gamma\left(\tilde{E}_k\right) \leq 2\varepsilon_n \operatorname{length}(\partial(B_\delta(x_k)\cap\mathcal{M})) = 2C_{\delta,k}\varepsilon_n. \tag{C.20}$$

The constant $C_{\delta,k}$ depends on δ, $\mathcal{M}$, and k but not n. This inequality, combined with (C.18) and (C.19), gives us that

$$\|\mathcal{O}v_n - \mathcal{O}^n v_n\|_2^2 \leq 2d_y\|\mathcal{G}\|_{op}C_{\delta,d}\varepsilon_n \lesssim d_y \frac{(\log n)^{c_d}}{n^{1/d}}. \tag{C.21}$$

Plugging (C.16) and (C.21) into (C.10) and taking the supremum of both sides over all $\|u\|_{L^2(\mathcal{M})} = 1$ yields

$$\|\mathcal{F} - FP\|_{op} \lesssim (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}} + \sqrt{d_y}(\log n)^{\frac{c_d}{2}} n^{-\frac{1}{2d}}. \tag{C.22}$$

In our setting, where the dimension of the observations is fixed, the error will be dominated by the first of these two terms. ■

Appendix D. Proof of Theorem 5.7.

Proof of Theorem 5.7. We want to verify the assumptions of Theorem 5.5 for the setting of Eulerian data assimilation described in Example 5.6. First, we verify that the discretized prior mean converges to the continuum prior mean in $\mathcal{H}$. Since $m_0 \in \mathcal{H}^\alpha(\mathcal{D})$, we have from equation (4.7) in [18] that

$$\|m_0 - P^* P m_0\|^2_{L^2(\mathcal{D})} \leq \frac{1}{n^{2\alpha}} \|m_0\|^2_{\mathcal{H}^\alpha(\mathcal{D})},$$

which verifies convergence of the discretized prior mean. Next, we verify operator norm convergence of the discretized prior covariance. For any $u \in \mathcal{H}$ with $\|u\|_{L^2(\mathcal{D})} = 1$,

$$\|\mathcal{C}_0 u - P^* C_0 P u\|^2_{L^2(\mathcal{D})} = \sum_{i=n+1}^{\infty} (\lambda^{(i)})^{-2\alpha} \langle u, \psi^{(i)} \rangle^2_{L^2(\mathcal{D})} \leq \sum_{i=n+1}^{\infty} \left(\lambda^{(i)}\right)^{-2\alpha} \lesssim \sum_{i=n+1}^{\infty} i^{-2\alpha} \lesssim \frac{1}{n^{2\alpha-1}}.$$

Hence, we have that

$$\|\mathcal{C}_0 - P^* C_0 P\|_{op} \lesssim \frac{1}{n^{\alpha-\frac{1}{2}}}, \tag{D.1}$$

and since $\alpha > 1$, we have shown operator-norm convergence of the discretized prior covariance.

It remains to show that the discretized potential $\Phi^{(n)}$ converges continuously to Φ as $n \to \infty$. That is, we want to show that for every $u \in \mathcal{H}$ and every $\varepsilon > 0$, there exist $N \in \mathbb{N}$ and $\delta > 0$ such that if $n \geq N$ and $u' \in B_\delta(u)$, then $\left|\Phi^{(n)}(u') - \Phi(u)\right| < \varepsilon$. Fix $u \in \mathcal{H}$ and $\varepsilon > 0$. From Lemma 3.2 in [17], we have that for any $u, u' \in \mathcal{H}$ and $f \in L^2(0, T; \mathcal{H})$, there exists a constant $L(\|u\|_{L^2(\mathcal{D})}, \|u'\|_{L^2(\mathcal{D})}, |f|_0, t_0, \Gamma)$ such that

$$\|\mathcal{F}(u) - \mathcal{F}(u')\|_{\Gamma^{-1}} \leq L\|u - u'\|_{L^2(\mathcal{D})},$$

provided that $t > t_0 > 0$. We then see that

$$\begin{aligned}
\Phi(u) - \Phi(u') &= \frac{1}{2}\left(\|y - \mathcal{F}(u)\|_{\Gamma^{-1}} - \|y - \mathcal{F}(u')\|_{\Gamma^{-1}}\right)\left(\|y - \mathcal{F}(u)\|_{\Gamma^{-1}} + \|y - \mathcal{F}(u')\|_{\Gamma^{-1}}\right) \\
&\leq \frac{1}{2}\|\mathcal{F}(u) - \mathcal{F}(u')\|_{\Gamma^{-1}}\left(\|y - \mathcal{F}(u)\|_{\Gamma^{-1}} + \|y - \mathcal{F}(u')\|_{\Gamma^{-1}}\right) \\
&\leq \frac{1}{2}\|\mathcal{F}(u) - \mathcal{F}(u')\|_{\Gamma^{-1}}\left(2\|y\|_{\Gamma^{-1}} + \|\mathcal{F}(u)\|_{\Gamma^{-1}} + \|\mathcal{F}(u')\|_{\Gamma^{-1}}\right) \\
&\leq \frac{1}{2}L\|u - u'\|_{L^2(\mathcal{D})}\left(2\|y\|_{\Gamma^{-1}} + 2\|\mathcal{F}(u)\|_{\Gamma^{-1}} + L\|u - u'\|_{L^2(\mathcal{D})}\right).
\end{aligned}$$

Exactly the same argument can be carried out for $\Phi(u') - \Phi(u)$ and yields the same upper bound, from which we conclude that

$$\left|\Phi(u) - \Phi(u')\right| \leq \frac{1}{2}L\|u - u'\|_{L^2(\mathcal{D})}\left(2\|y\|_{\Gamma^{-1}} + 2\|\mathcal{F}(u)\|_{\Gamma^{-1}} + Lt_0\|u - u'\|_{L^2(\mathcal{D})}\right). \tag{D.2}$$

From Lemma 4.2 in [18], we have that for the Galerkin approximation v^N in (5.14) to the true solution v in (5.10), there exists a constant $\tilde{L}\left(\|u\|_{L^2(\mathcal{D})}, t_0, \Gamma\right)$ such that, for any $t > t_0 > 0$,

$$\|v(t) - v^N(t)\|_{L^\infty(\mathcal{D})} \leq \tilde{L}(\|u\|_{L^2(\mathcal{D})})\psi(n),$$

where $\psi(n) \to 0$ as $n \to \infty$, and $\tilde{L}(\|u\|_{L^2(\mathcal{D})})$ is continuous and increasing in $\|u\|_{L^2(\mathcal{D})}$. It follows from this inequality that

$$\|\mathcal{F}(u') - F(u')\|_{\Gamma^{-1}} \leq \tilde{L}(\|u'\|_{L^2(\mathcal{D})})\psi(n) \leq \tilde{L}(\|u\|_{L^2(\mathcal{D})} + \delta)\psi(n).$$

Then, we observe that

$$\begin{aligned}\Phi(u') - \Phi^{(n)}(u') &= \frac{1}{2}\left(\|y - \mathcal{F}(u')\|_{\Gamma^{-1}} - \|y - F(u')\|_{\Gamma^{-1}}\right)\left(\|y - \mathcal{F}(u')\|_{\Gamma^{-1}} + \|y - F(u')\|_{\Gamma^{-1}}\right)\\ &\leq \frac{1}{2}\|\mathcal{F}(u') - F(u')\|_{\Gamma^{-1}}\left(\|y - \mathcal{F}(u')\|_{\Gamma^{-1}} + \|y - F(u')\|_{\Gamma^{-1}}\right)\\ &\leq \frac{1}{2}\|\mathcal{F}(u') - F(u')\|_{\Gamma^{-1}}\left(2\|y\|_{\Gamma^{-1}} + \|\mathcal{F}(u')\|_{\Gamma^{-1}} + \|F(u')\|_{\Gamma^{-1}}\right)\\ &\leq \frac{1}{2}\tilde{L}(\|u\|_{L^2(\mathcal{D})} + \delta)\psi(n)\left(2\|y\|_{\Gamma^{-1}} + 2\|\mathcal{F}(u)\|_{\Gamma^{-1}} + \tilde{L}(\|u\|_{L^2(\mathcal{D})} + \delta)\psi(n)\right).\end{aligned}$$

Again, the same argument carried out for $\Phi^{(n)}(u') - \Phi(u')$ yields the same upper bound, and we conclude that

(D.3)
$$\left|\Phi(u') - \Phi^{(n)}(u')\right| \leq \frac{1}{2}\tilde{L}(\|u\|_{L^2(\mathcal{D})} + \delta)\psi(n)\left(2\|y\|_{\Gamma^{-1}} + 2\|\mathcal{F}(u)\|_{\Gamma^{-1}} + \tilde{L}(\|u\|_{L^2(\mathcal{D})} + \delta)\psi(n)\right).$$

From (D.2) we see that we can pick δ small enough such that $|\Phi(u) - \Phi(u')| < \frac{\varepsilon}{2}$ for any $u' \in B_\delta(u) \subset \mathcal{H}$, and from (D.3) we can pick N large enough such that for any $n \geq N$, we have $|\Phi(u') - \Phi^{(n)}(u')| < \frac{\varepsilon}{2}$ for any $u' \in B_\delta(u) \subset \mathcal{H}$. Consequently, there exist N and δ such that for any $n \geq N$ and $u' \in B_\delta(u) \in \mathcal{H}$,

$$\left|\Phi(u) - \Phi^{(n)}(u')\right| < \varepsilon, \tag{D.4}$$

proving continuous convergence of $\Phi^{(n)}$ to Φ as $n \to \infty$. We can then apply Theorem 5.5 and complete the proof. ■

Acknowledgment. The authors are grateful to Yuming Chen for his generous feedback on a previous version of this manuscript.

REFERENCES

[1] S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, and A. M. Stuart, *Importance sampling: Intrinsic dimension and computational cost*, Stat. Sci., 32 (2017), pp. 405–431.

[2] H. Antil and A. K. Saibaba, *Efficient algorithms for Bayesian inverse problems with Whittle–Matérn priors*, SIAM J. Sci. Comput., (2023), pp. S176–S198, https://doi.org/10.1137/22M1494397.

[3] B. Ayanbayev, I. Klebanov, H. C. Lie, and T. J. Sullivan, *Γ-convergence of Onsager–Machlup functionals:* I. *With applications to maximum a posteriori estimation in Bayesian inverse problems*, Inverse Problems, 38 (2021), 025005.

[4] M. Belkin and P. Niyogi, *Semi-supervised learning on Riemannian manifolds*, Mach. Learn., 56 (2004), pp. 209–239.

[5] M. Belkin and P. Niyogi, *Towards a theoretical foundation for Laplacian-based manifold methods*, in Proceedings of the International Conference on Computational Learning Theory, Springer, 2005, pp. 486–500.

[6] D. Bigoni, Y. Chen, N. Garcia Trillos, Y. Marzouk, and D. Sanz-Alonso, *Data-driven forward discretizations for Bayesian inversion*, Inverse Problems, 36 (2020), 105008.

[7] D. Bolin and K. Kirchner, *The rational SPDE approach for Gaussian random fields with general smoothness*, J. Comput. Graph. Statist., 29 (2020), pp. 274–285.

[8] A. Braides, *A handbook of* Γ*-convergence*, in Handbook of Differential Equations: Stationary Partial Differential Equations, Vol. 3, Elsevier, 2006, pp. 101–213.

[9] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, Texts Appl. Math. 15, Springer, 2008.

[10] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler, *A computational framework for infinite-dimensional Bayesian inverse problems Part* I*: The linearized case, with application to global seismic inversion*, SIAM J. Sci. Comput., 35 (2013), pp. A2494–A2523, https://doi.org/10.1137/12089586X.

[11] D. Burago, S. Ivanov, and Y. Kurylev, *A graph discretization of the Laplace–Beltrami operator*, J. Spectr. Theory, 4 (2015), pp. 675–714.

[12] D. Calvetti and E. Somersalo, *Hypermodels in the Bayesian imaging framework*, Inverse Problems, 24 (2008), 034013.

[13] Y. Canzani, *Analysis on Manifolds via the Laplacian*, lecture notes, https://www.math.mcgill.ca/toth/spectral%20geometry.pdf, 2013.

[14] N. K. Chada, Y. Chen, and D. Sanz-Alonso, *Iterative ensemble Kalman methods: A unified perspective with some new variants*, Found. Data Sci., 3 (2021), pp. 331–369.

[15] E. Cleary, A. Garbuno-Inigo, S. Lan, T. Schneider, and A. M. Stuart, *Calibrate, emulate, sample*, J. Comput. Phys., 424 (2021), 109716.

[16] S. Cotter, G. Roberts, A. Stuart, and D. White, *MCMC methods for functions: Modifying old algorithms to make them faster*, Statist. Sci., 28 (2013), pp. 424–446.

[17] S. L. Cotter, M. Dashti, J. C. Robinson, and A. M. Stuart, *Bayesian inverse problems for functions and applications to fluid mechanics*, Inverse Problems, 25 (2009), 115008.

[18] S. L. Cotter, M. Dashti, and A. M. Stuart, *Approximation of Bayesian inverse problems for PDEs*, SIAM J. Numer. Anal., 48 (2010), pp. 322–345, https://doi.org/10.1137/090770734.

[19] M. Dashti, K. Law, A. M. Stuart, and J. Voss, *MAP estimators and their consistency in Bayesian nonparametric inverse problems*, Inverse Problems, 29 (2013), 095017.

[20] L. C. Evans, *Partial Differential Equations*, Grad. Stud. Math. 19, American Mathematical Soc., 2010.

[21] J. Franklin, *Well-posed stochastic extensions of ill-posed linear problems*, J. Math. Anal. Appl., 31 (1970), pp. 682–716.

[22] A. Garbuno-Inigo, T. Helin, F. Hoffmann, and B. Hosseini, *Bayesian Posterior Perturbation Analysis with Integral Probability Metrics*, preprint, arXiv:2303.01512, 2023.

[23] N. Garcia Trillos, Z. Kaplan, T. Samakhoana, and D. Sanz-Alonso, *On the consistency of graph-based Bayesian semi-supervised learning and the scalability of sampling algorithms*, J. Mach. Learn. Res., 21 (2020), pp. 1–47.

[24] N. García Trillos, M. Gerlach, M. Hein, and D. Slepčev, *Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace–Beltrami operator*, Found. Comput. Math., 20 (2020), pp. 827–887.

[25] N. García Trillos and D. Sanz-Alonso, *Continuum limits of posteriors in graph Bayesian inverse problems*, SIAM J. Math. Anal., 50 (2018), pp. 4020–4040, https://doi.org/10.1137/17M1138005.

[26] O. Al Ghattas and D. Sanz-Alonso, *Non-asymptotic Analysis of Ensemble Kalman Updates: Effective Dimension and Localization*, preprint, arXiv:2208.03246, 2022.

[27] J. Harlim, S. W. Jiang, H. Kim, and D. Sanz-Alonso, *Graph-based prior and forward models for inverse problems on manifolds with boundaries*, Inverse Problems, 38 (2022), 035006.

[28] J. Harlim, D. Sanz-Alonso, and R. Yang, *Kernel methods for Bayesian elliptic inverse problems on manifolds*, SIAM/ASA J. Uncertain. Quantif., 8 (2020), pp. 1414–1445, https://doi.org/10.1137/19M1295222.

[29] T. Helin and M. Burger, *Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems*, Inverse Problems, 31 (2015), 085009.

[30] M. A. Iglesias, *Iterative regularization for ensemble data assimilation in reservoir models*, Comput. Geosci., 19 (2015), pp. 177–212.

[31] M. A. Iglesias, K. J. H. Law, and A. M. Stuart, *Ensemble Kalman methods for inverse problems*, Inverse Problems, 29 (2013), 045001.

[32] S. W. Jiang and J. Harlim, *Ghost point diffusion maps for solving elliptic PDEs on manifolds with classical boundary conditions*, Comm. Pure Appl. Math., 76 (2023), pp. 337–405.

[33] J. Kaipo and E. Somersalo, *Statistical and Computational Inverse Problems*, Appl. Math. Sci. 160 Springer-Verlag, 2005.

[34] I. Klebanov and P. Wacker, *Maximum a posteriori estimators in ℓ^2 are well-defined for diagonal Gaussian priors*, Inverse Problems, 39 (2023), 065009.

[35] R. Kretschmann, *Nonparametric Bayesian Inverse Problems with Laplacian Noise*, Ph.D. thesis, Universität Duisburg-Essen, 2019.

[36] E. Kwiatkowski and J. Mandel, *Convergence of the square root ensemble Kalman filter in the large ensemble limit*, SIAM/ASA J. Uncertain. Quantif., 3 (2015), pp. 1–17, https://doi.org/10.1137/140965363.

[37] H. Lambley, *Strong Maximum a Posteriori Estimation in Banach Spaces with Gaussian Priors*, preprint, arXiv:2304.13622, 2023.

[38] H. Lambley and T. Sullivan, *An Order-Theoretic Perspective on Modes and Maximum a Posteriori Estimation in Bayesian Inverse Problems*, preprint, arXiv:2209.11517, 2022.

[39] J. Latz, *On the well-posedness of Bayesian inverse problems*, SIAM/ASA J. Uncertain. Quantif., 8 (2020), pp. 451–482, https://doi.org/10.1137/19M1247176.

[40] J. Latz, *Bayesian inverse problems are usually well-posed*, SIAM Rev., 65 (2023), pp. 831–865, https://doi.org/10.1137/23M1556435.

[41] Y. Marzouk and D. Xiu, *A stochastic collocation approach to Bayesian inference in inverse problems*, Commun. Comput. Phys., 6 (2009), pp. 826–847.

[42] G. Dal Maso, *An Introduction to Γ-Convergence*, Progr. Nonlinear Differential Equations Appl. 8, Birkhäuser Boston, 1993.

[43] R. Nickl and E. S. Titi, *On Posterior Consistency of Data Assimilation with Gaussian Process Priors: The 2D Navier-Stokes Equations*, preprint, arXiv:2307.08136, 2023.

[44] J. T. Oden and J. N. Reddy, *An Introduction to the Mathematical Theory of Finite Elements*, Courier Corporation, 2012.

[45] M. Ottobre, N. S. Pillai, F. J. Pinski, and A. M. Stuart, *A Function space HMC algorithm with second order Langevin diffusion limit*, Bernoulli, 22 (2015), pp. 60–106.

[46] N. Petra, J. Martin, G. Stadler, and O. Ghattas, *A computational framework for infinite-dimensional Bayesian inverse problems, Part II: Stochastic Newton MCMC with application to ice sheet flow inverse problems*, SIAM J. Sci. Comput., 36 (2014), pp. A1525–A1555, https://doi.org/10.1137/130934805.

[47] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, Adapt. Comput. Mach. Learn. 2, MIT Press, 2006.

[48] J. C. Robinson, *Infinite-Dimensional Dynamical Systems: An Introduction to Dissipative Parabolic PDEs and the Theory of Global Attractors*, Cambridge Texts Appl. Math. 28, Cambridge University Press, 2001.

[49] L. Roininen, J. Huttunen, and S. Lasanen, *Whittle-Matérn priors for Bayesian statistical inversion with applications in electrical impedance tomography*, Inverse Probl. Imaging, 8 (2014), pp. 561–586.

[50] D. Sanz-Alonso, A. Stuart, and A. Taeb, *Inverse Problems and Data Assimilation*, London Math. Soc. Stud. Texts 107, Cambridge University Press, 2023.

[51] D. Sanz-Alonso and R. Yang, *The SPDE approach to Matérn fields: Graph representations*, Statist. Sci., 37 (2022), pp. 519–540.

[52] B. Sprungk, *On the local Lipschitz stability of Bayesian inverse problems*, Inverse Problems, 36 (2020), 055015.

[53] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Ser. Statist., Springer, 1999.

[54] G. Strang and G. J. Fix, *An Analysis of the Finite Element Method*, Prentice-Hall Series in Automatic Computation, Prentice-Hall, 1973.

[55] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.

[56] V. S. Sunder, *Functional Analysis: Spectral Theory*, Texts and Readings in Math. 13, Hindustan Book Agency, 1997.
[57] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, 2005, https://epubs.siam.org/doi/book/10.1137/1.9780898717921.
[58] R. Temam, *Navier–Stokes Equations and Nonlinear Functional Analysis*, CBMS-NSF Reg. Conf. Ser. Appl. Math. 66, SIAM, 1995.
[59] R. Temam, *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, Appl. Math. Sci., 68, Springer-Verlag, 1997.
[60] J. A. Tropp, *An introduction to matrix concentration inequalities*, Found. Trends Mach. Learn., 8 (2015), pp. 1–230.