THE UNIVERSITY OF CHICAGO

HIGH-DIMENSIONAL ANALYSIS OF ENSEMBLE KALMAN FILTERS AND HIERARCHICAL BAYESIAN INVERSE PROBLEMS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

COMMITTEE ON COMPUTATIONAL AND APPLIED MATHEMATICS

BY
NATHAN WANIOREK

CHICAGO, ILLINOIS
JUNE 2026

© 2026 by Nathan Waniorek

TABLE OF CONTENTS

LIST OF FIGURES v

LIST OF TABLES vi

ACKNOWLEDGMENTS vii

ABSTRACT viii

1 INTRODUCTION 1
1.1 Bayesian Inverse Problems 1
1.2 Data Assimilation 3
1.3 Low-Dimensional Structure 7
1.3.1 Smoothness 7
1.3.2 Low-Dimensional Dynamics 8
1.3.3 Sparsity 9
1.4 Outline and Main Contributions 10

2 ANALYSIS OF A COMPUTATIONAL FRAMEWORK FOR BAYESIAN INVERSE PROBLEMS: ENSEMBLE KALMAN UPDATES AND MAP ESTIMATORS UNDER MESH REFINEMENT 13
2.1 Introduction 13
2.1.1 Outline and Main Contributions 15
2.1.2 Notation 16
2.2 Problem Setting and Computational Framework 17
2.2.1 Inverse Problem in Hilbert Space 17
2.2.2 Computational Framework 19
2.2.3 Inverse Problems in Function Space 23
2.3 Error Analysis 32
2.3.1 Error Analysis: General Computational Framework 33
2.3.2 Error Analysis: Finite Element Discretizations 38
2.3.3 Error Analysis: Graph-Based Approximations 41
2.4 Sample Size Requirements for Ensemble Kalman Updates 44
2.4.1 Ensemble Approximation to Finite-Dimensional Posterior 44
2.4.2 Ensemble Kalman Approximation: Error Analysis 45
2.4.3 Effective Dimension of Prior Covariance 48
2.5 Convergence of MAP Estimators Under Mesh Refinement 50
2.6 Conclusion and Future Directions 57
2.7 Adjoints in Weighted Space 58
2.8 Proofs Subsection 2.3.2 58
2.9 Proofs Subsection 2.3.3 60
2.10 Proof of Theorem 2.5.7 70

3 LONG-TIME ACCURACY OF ENSEMBLE KALMAN FILTERS FOR CHAOTIC AND MACHINE-LEARNED DYNAMICAL SYSTEMS 74
3.1 Introduction 74
3.1.1 Related work 75
3.1.2 Main contributions and outine 78
3.2 Main results 79
3.2.1 Long-time accuracy of ensemble Kalman filters 79
3.2.2 Long-time accuracy with surrogate models 88
3.3 Filter accuracy via mean-field theory 90
3.4 Filter accuracy with surrogate models via mean-field theory 109
3.5 Numerical experiments 127
3.6 Conclusions 131

4 HIERARCHICAL BAYESIAN INVERSE PROBLEMS: A HIGH-DIMENSIONAL STATISTICS VIEWPOINT 133
4.1 Introduction 133
4.2 Problem Formulation 137
4.3 Unified Theoretical Framework 145
4.3.1 Approximate Decomposability 145
4.3.2 General Bound on the Reconstruction Error 150
4.4 Reconstruction Error for Sparsity-Promoting Models 153
4.4.1 Hierarchical Models for Sparsity 153
4.4.2 Hierarchical Models for Group Sparsity 158
4.4.3 Hierarchical Models for Sparse Representations 161
4.5 Proof of Reconstruction Error for Sparsity-Promoting Models 163
4.5.1 Hierarchical Models for Sparsity 163
4.5.2 Hierarchical Models for Group Sparsity 166
4.5.3 Hierarchical Models for Sparse Representations 168
4.6 Conclusions 169
4.7 Computational Properties of GS-IAS and O-IAS 171
4.8 Approximate Decomposability: Examples 176
4.9 Approximate Decomposability and General Bounds 177

REFERENCES 182

LIST OF FIGURES

3.1 Average filter error over 50 Monte Carlo trials for decreasing observation noise levels, plotted with two standard errors (shaded). 128
3.2 The structure of Ψ^s. The output channels of CNN_1 are divided into three groups of equal length, $\text{CNN}_1^{(1)}$, $\text{CNN}_1^{(2)}$, and $\text{CNN}_1^{(3)}$. The input channels to CNN_2 are a concatenation of $\text{CNN}_1^{(1)}$ and $(\text{CNN}_1^{(1)} \times \text{CNN}_1^{(2)})$, where the multiplication is point-wise. 129
3.3 Visualization of the true state trajectory over time, the ensemble means, the observed data, and the difference between the true state and the ensemble means. 130
3.4 The left plot displays the average filter error over 50 Monte Carlo trials plotted with two standard errors. The right plot displays the average surrogate model forecast error from an exactly known initial condition over 50 Monte Carlo trials plotted with two standard errors. This figure shows that long-time filter accuracy is possible with surrogate models that only give accurate short-term forecasts. . 131

LIST OF TABLES

2.1 Roadmap of the discretizations analyzed in this paper as particular examples of the general computational framework introduced in Subsection 2.2.2. 23

3.1 Summary of the training details, estimated model error, and average filter error for each model. 131

ACKNOWLEDGMENTS

First and foremost, I want to thank my advisor, Daniel Sanz-Alonso. Without his support, guidance, patience, and insight, none of this would have been possible. I could not have asked for a better advisor. I want to thank Nisha Chandramoorthy for being an excellent collaborator and for providing thoughtful feedback on my work. Nisha's questions, comments, and advice have greatly improved my research. Thank you to Pedram Hassanzadeh; I have learned a tremendous amount from our discussions. I am grateful to my brilliant collaborators, Jiaheng Chen and Omar Al-Ghattas, from whom I have learned so much about computational mathematics and statistics.

I am immensely thankful for the friends I have made here at UChicago. I consider myself very lucky to have had the opportunity to spend the past five years with such wonderful people. I will cherish these memories and our friendships forever. Thank you to Luke, Giuliana, Morgan, and Paulina for everything, especially for the reminders that there is a world outside of math. Thank you to my family: Judy, Lyle and Levi. Your support has meant the world to me. Finally, thank you to Russ, Bernice, and Joanne. I could not possibly be where I am today without you.

ABSTRACT

The ensemble Kalman filter (EnKF) is a widely used algorithm for high-dimensional data assimilation tasks, such as numerical weather prediction. Despite its widespread use and empirical success, the theoretical underpinnings of the EnKF are largely not understood. The first part of this thesis aims to provide a rigorous analysis of the EnKF in high/infinite-dimensional state spaces. We first consider discretizations of Bayesian inverse problems on a Hilbert space, where we prove that the statistical complexity of the discretized ensemble Kalman update is independent of the mesh refinement. Our results hold for the important practical settings of finite-element discretizations and graph-based discretizations. We also prove the convergence of discretized MAP estimators to the continuum solutions under mesh-refinement. Next, we establish a filter accuracy result for the EnKF applied to partially-observed nonlinear dynamical systems, addressing an important open problem in the data assimilation literature. Previous EnKF accuracy guarantees have been limited to either fully observed dynamical systems or linear models, making the results inapplicable to realistic systems. We provide a natural condition coupling the nonlinear dynamics and partial observations such that the estimation error of the EnKF remains small in the long-time horizon. We also prove a filter accuracy result for the EnKF with a surrogate dynamics model, validating the use of machine-learned forecast models in ensemble data assimilation. The second part of this thesis analyzes a family of hierarchical Bayesian models originally developed in the computational mathematics community for inverse problems in medical imaging. The main contribution of this work is to bring techniques from the high-dimensional statistics literature to establish the first known reconstruction error guarantees for these models. Our analysis leverages a property of hierarchical Bayesian regularizers that we call approximate decomposability to obtain non-asymptotic bounds on the reconstruction error attained by maximum a posteriori estimators. The new theory explains how hierarchical Bayesian models that exploit sparsity, group sparsity, and sparse representations of the

unknown parameter can achieve accurate reconstructions in high-dimensional settings.

CHAPTER 1

INTRODUCTION

This thesis aims to develop rigorous mathematical and statistical foundations for algorithms to combine complex predictive models with observed data, specifically for methods arising as solutions to Bayesian inverse problems and data assimilation tasks. A theme unifying the problems considered here is that of high-dimensionality, which presents both computational and statistical challenges. Across the work presented in this thesis, we characterize settings where widely-used approaches for inverse problems and data assimilation can perform well despite the challenges presented by high or infinite-dimensional parameter spaces. In doing so, we can better understand the mechanisms by which these methods are successful in practice and the types of intrinsic low-dimensional structure they are able exploit. Throughout this thesis, we will see how methods for inverse problems and filtering can exploit spatial smoothness, low-dimensional dynamics, and sparsity to accurately infer high- or infinite-dimensional parameters from few observations. Rigorously understanding these mechanisms is fundamental to (i) reliably using these methods in practical settings and (ii) to developing novel methodology for settings in which the current algorithms may fail.

1.1 Bayesian Inverse Problems

Inverse problems concern the inference of unknown parameters of interest from indirect observations. To specify an inverse problem, we consider the data generating process

$$y = \mathcal{F}(u) + \eta, \quad u \in \mathcal{U}, \quad y \in \mathcal{Y}. \tag{1.1}$$

Here, u is the unknown of interest living in a parameter space $\mathcal{U}$, y is the observed data living in data space $\mathcal{Y}$, and $\eta \in \mathcal{Y}$ is measurement noise. The forward model $\mathcal{F} : \mathcal{U} \to \mathcal{Y}$ maps parameters to the data space. To goal of the inverse problem is to estimate the parameter u

from noisy measurements y, and quantify the uncertainty in this estimate. Inverse problems of interest are typically ill-posed, meaning that an inverse of $\mathcal{F}$ may not exist, may not be unique, or may not depend continuously on the data [182]. In this thesis, we are primarily concerned with the Bayesian approach to inverse problems [315, 80, 297], in which the inverse problem is regularized by modeling the unknown parameters probabilistically. In the Bayesian setting, the primary object of interest is the posterior distribution of the unknown u given the observed data y. The posterior is specified by Bayes' rule, written informally as

$$\pi(u|y) \propto \pi(y|u)\pi(u). \tag{1.2}$$

In this expression, $\pi(y|u)$ is the likelihood, which is specified via the data generating process in Equation 1.1 by making probabilistic assumptions on the measurement errors, η. The distribution $\pi(u)$ denotes the user-specified prior distribution for the unknown. The prior should be designed to encode all uncertainty in the unknown, u, and impose any known structure about the parameter of interest. The resulting posterior distribution, in its totality, is considered the "solution" to the Bayesian inverse problem. However, one can very seldom obtain closed-form expressions for the posterior distributions arising in the scientific applications motivating this thesis. Often, this is because the forward model $\mathcal{F}$ involves the solution to a nonlinear (partial) differential equation, or because the known structure about the parameters may be best described by a complicated non-Gaussian prior. One notable exception where analytic expressions for the posterior are available is when the forward model $\mathcal{F}$ is linear, and the prior and noise distributions are assumed to be independent Gaussians, $u \sim \mathcal{N}(m_0, \mathcal{C}_0)$ and $\eta \sim \mathcal{N}(0, \Gamma)$. Under these assumptions, the posterior is also Gaussian, with mean and covariance given by

$$\begin{aligned} m_{\text{post}} &= m_0 + \mathcal{C}_0\mathcal{F}^*(\mathcal{F}\mathcal{C}_0\mathcal{F}^* + \Gamma)^{-1}(y - \mathcal{F}m_0), \\ \mathcal{C}_{\text{post}} &= \mathcal{C}_0 - \mathcal{C}_0\mathcal{F}^*(\mathcal{F}\mathcal{C}_0\mathcal{F}^* + \Gamma)^{-1}\mathcal{F}\mathcal{C}_0. \end{aligned} \tag{1.3}$$

The linear-Gaussian Bayesian update will appear frequently throughout this thesis. It will be a major focus of much of Chapter 2, where we consider problems formulated on an infinite-dimensional Hilbert space. In the Hilbert space setting, the Gaussian posterior specified in Equation 1.3 remains difficult to compute despite having an analytic expression. This is because the mappings $\mathcal{F}$ and $\mathcal{C}_0$ typically involve the solution operators to partial differential equations that do not admit closed-form expressions. In Chapters 3 and 4, the linear-Gaussian update will be used as an algorithmic step within various methods for non-Gaussian problems. For a non-Gaussian posterior, one often relies on methods to (approximately) sample from the posterior distribution, such as Markov chain Monte Carlo [63], measure transport [252], or variational inference [46]. Sampling from Bayesian posteriors is an active research area, but is largely not the focus of this thesis.

Often, one is interested in obtaining a single point estimate for the unknown parameters. A common choice is the maximum a posteriori (MAP) estimator, which is the point maximizing the posterior density with respect to a reference measure. Defining the correct notion of MAP estimator in infinite-dimensional problems is a delicate matter, which we discuss in detail in Chapter 2. One of the main contributions of Chapter 2 is demonstrating that the MAP estimators computed for a discretized version of the inverse problem converge to the MAP estimators for the continuum problem under reasonable assumptions on the prior and forward model. Further, in Chapter 4, we will show that for a popular family of sparsity-promoting hierarchical prior models, the MAP estimators can provably be computed efficiently and can accurately reconstruct high-dimensional unknowns from limited observations under sparsity assumptions.

1.2 Data Assimilation

Data assimilation concerns the estimation of the state of a dynamical system from partial observations corrupted by noise. Mathematically, this problem can be formulated as a state-

space model

$$\begin{aligned} u_{j+1} &= \Psi(u_j) + \xi_j, \quad j \geq 0, \\ y_{j+1} &= Hu_{j+1} + \eta_{j+1}, \quad j \geq 0. \end{aligned} \tag{1.4}$$

Here, u_j denotes the state at time j, Ψ denotes the dynamics map, y_j denotes the observed data at time j, and the observation map is given by H. The variables ξ_j and η_j denote the noise in the dynamics and observations at time j, respectively. The *filtering problem* seeks to incorporate the data in an online fashion, estimating at each time j the state given all of the observations up until and including those obtained at the current time. The filtering problem is of great interest in numerical weather prediction, in which one aims to estimate the current state of the atmosphere from sparse, noisy observations arriving sequentially in time. One can also view the filtering problem as a sequence of inverse problems in which the data is obtained in an online manner. Taking a Bayesian perspective, the analogous object to the posterior distribution in inverse problems is known as the filtering distribution. Assuming the initial condition is only known statistically, $u_0 \sim \pi_0(u)$, and making probabilistic assumptions on the noise in the dynamics and observations, the model given in Equation 1.4 specifies a sequence of *filtering distributions*, i.e. the sequence of conditional distributions of the state u_j given all of the observations up until time j,

$$\pi_j(u_j) := \pi(u_j | y_1, \ldots, y_j). \tag{1.5}$$

For generic systems of the form in Equation 1.4, the filtering distribution does not have a closed-form expression, and must be approximated. As was the case with inverse problems, an important exception is in the linear-Gaussian setting. Assume the dynamics are linear, $\Psi(\cdot) = M\cdot$, the noise in the dynamics $\xi_j \sim \mathcal{N}(0, Q)$ and the observations $\eta_j \sim \mathcal{N}(0, \Gamma)$ are i.i.d. Gaussian, and that the initial condition is modeled with by a Gaussian distribution,

$u_0 \sim \mathcal{N}(m_0, C_0)$. Under these assumptions, the filtering distribution is Gaussian at all times j and recovers the celebrated Kalman filter [210], given in Algorithm 1.

Algorithm 1 The Kalman Filter

1: **Initialization**: $\pi_0(u_0) = \mathcal{N}(m_0, C_0)$.

2: For $j = 1, 2, \ldots$ do the following:

3: **Prediction**:

$$\begin{aligned} \mu_j &= M m_{j-1}, \\ \Sigma_j &= M C_{j-1} M^* + Q \end{aligned} \tag{1.6}$$

4: **Analysis**:

$$\begin{aligned} m_j &= \mu_j + \Sigma_j H^* (H \Sigma_j H^* + \Gamma)^{-1} (y_j - \mu_j), \\ C_j &= (I - \Sigma_j H^* (H \Sigma_j H^* + \Gamma)^{-1} H) \Sigma_j \end{aligned} \tag{1.7}$$

5: **Output**: Filtering distributions $\{\mathcal{N}(m_j, C_j)\}_{j=1}^{\infty}$.

Many dynamical systems of interest, particularly those arising in numerical weather prediction, are highly nonlinear. Thus, the Kalman filter is not defined for these systems. In a seminal 1994 paper [147], Geir Evensen proposed the Ensemble Kalman Filter (EnKF) as a generalization of the Kalman filter to nonlinear dynamical systems. The theoretical study of the EnKF is one of the primary objectives of this thesis.

Algorithm 2 The Ensemble Kalman Filter [147]

1: **Initialization**: Draw $u_0^{(n)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(m_0, C_0), \qquad 1 \le n \le N.$

2: For $j = 1, 2, \dots$ do the following:

3: **Prediction**: Compute

$$\hat{u}_j^{(n)} = \Psi\big(u_{j-1}^{(n)}\big) + \xi_j^{(n)}, \qquad \xi_j^{(n)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, Q), \qquad 1 \le n \le N,$$
$$\widehat{\mu}_j = \frac{1}{N}\sum_{n=1}^{N} \hat{u}_j^{(n)}, \qquad \widehat{\Sigma}_j = \frac{1}{N}\sum_{n=1}^{N} \big(\hat{u}_j^{(n)} - \widehat{\mu}_j\big) \otimes \big(\hat{u}_j^{(n)} - \widehat{\mu}_j\big).$$

4: **Analysis**: Update

$$u_j^{(n)} = \hat{u}_j^{(n)} + \Sigma_j H^* (H \Sigma_j H^* + \Gamma)^{-1} (y_j - H \hat{u}_j^{(n)}), \qquad 1 \le n \le N.$$

5: **Output**: Ensemble $\{u_j^{(n)}\}_{n=1}^N, \quad j = 0, 1, 2, \dots$

Comparing Algorithms 1 and 2, it is evident how one might derive the EnKF as a natural generalization of the Kalman filter: first compute a prediction step by propagating an ensemble of state estimates through the dynamics, and then update the ensemble members through a Kalman-type analysis step that uses the empirical covariances of the forecast ensemble. Variants of the EnKF are used operationally in numerical weather prediction, notably at the Meteorological Service of Canada [198]. In atmospheric data assimilation, the EnKF typically operates in state spaces of dimension $d_u = 10^9$ and data dimension $d_y = 10^6$, while propagating only around $N = 100$ ensemble members. This is quite appealing from a computational perspective, since one need not form and store covariance matrices of size $10^9 \times 10^9$, which would be prohibitively expensive. In the linear-Gaussian setting, this algorithm is also quite sensible from a Bayesian perspective. It is known that the EnKF recovers the Kalman filter in the large-ensemble limit [251]. However, classical statistical wisdom would suggest that one should not, in general, be able to accurately estimate a covariance matrix

in dimension 10^9 from only 100 samples, as is the case in operational settings. Further, the systems considered in such geophysical applications are highly nonlinear, and in contrast to the widespread empirical success of the EnKF in applications, it is known that the ensemble Kalman filter is not a consistent estimator of the filtering distribution for nonlinear dynamics models [234, 145]. In Chapters 2 and 3, we will provide theory to understand why the EnKF performs well in practice despite these challenges.

1.3 Low-Dimensional Structure

In this section, we provide relevant background on each of the notions of structure that are relevant to the theory in this thesis: smoothness, low-dimensional dynamics, and sparsity.

1.3.1 Smoothness

Algorithms that exploit the smoothness of a target function are ubiquitous in numerical analysis [181], nonparametric statistics [346], optimization [270], and a wealth of other fields across mathematics and statistics. Here I will briefly present some background on the smoothness of Matérn-type Gaussian fields that are central to results in Chapter 2.

Taking $\mathcal{D} \subset \mathbb{R}^d$ to be a sufficiently regular open domain, we consider the operator $\mathcal{A}^{-1} : L^2(\mathcal{D}) \to H^2(\mathcal{D})$ defined as the solution operator for the an elliptic PDE: for $f \in L^2(\mathcal{D})$, $\mathcal{A}^{-1}f = u$, where u solves the following.

$$\int_{\mathcal{D}} \Theta \nabla u \cdot \nabla p + \int_{\mathcal{D}} bup = \int_{\mathcal{D}} fp, \quad \forall p \in H_0^1(\mathcal{D}),$$

where we assume $\Theta(x) \in C^1(\bar{D})$ and $b(x) \in L^\infty$ are bounded below by positive constants. We then consider a Gaussian prior $\mu_0 = \mathcal{N}(m_0, \mathcal{C}_0)$ with covariance operator defined to be $\mathcal{C}_0 = \mathcal{A}^{-\alpha}$ for some $\alpha > \frac{d}{2}$ and mean function $m_0 \in \mathrm{Im}(\mathcal{C}_0^{1/2})$. In Theorem 2.18 in [127], it is

shown that for a sample $u \sim \mu_0$, we have μ_0-almost surely that $u \in H^t(\mathcal{D})$ for all $t < \alpha - \frac{d}{2}$. This regularity is a result the decay of the spectrum of $\mathcal{C}_0$. We will see in Chapter 2 that the spectral decay of $\mathcal{C}_0$ also governs the statistical complexity of the ensemble Kalman update for problems with such Matérn-type priors. More precisely, the estimation error in the EnKF update is controlled by the *effective dimension* of $\mathcal{C}_0$:

$$r(\mathcal{C}_0) := \frac{\mathrm{Tr}(\mathcal{C}_0)}{\|\mathcal{C}_0\|_{op}}. \tag{1.8}$$

If $\mathcal{C}_0$ is trace-class (as is necessary to samples from μ_0 are almost surely L^2-integrable), then this quantity will be finite. If the eigenvalues of $\mathcal{C}_0$ decay sufficiently rapidly then the effective dimension will be quite small, enabling the EnKF to perform well with a small ensemble size.

1.3.2 Low-Dimensional Dynamics

The accuracy guarantees for the ensemble Kalman filter in Chapter 3 are built on the assumption that the dynamics are—in a particular sense—low-dimensional, and that the observation model observes enough of the "right" part of the dynamics. More precisely, we require that the mapping from the initial condition to the unobserved part of the forecasted state is contracting:

$$\|(I - H^*H)(\Psi(u) - \Psi(v))\| \leq \alpha\|u - v\|,$$

for some $\alpha \in (0, 1)$. In the data assimilation literature, this assumption is referred to as the *squeezing property* [299]. With appropriate partial observation models H, the squeezing property holds for a number of dynamics models of interest, including the short-time flow of the Lorenz systems and the 2D Navier-Stokes equations. Assumptions of this flavor underpin a rich theoretical perspective on data assimilation algorithms, which we broadly refer to as

filter accuracy. Generally speaking, the filter accuracy literature studies how a sufficiently informative observation model can synchronize (in the sense of [279]) the filtering algorithm with the true trajectory such that the algorithm converges to the true state. Early seminal works in this area arose in studying the theory of dissipative PDEs [273], and further works developed theory for more realistic observation models [20, 190] and for algorithms more often used operationally [230]. Precisely defining the notions of "low-dimensionality" that fundamentally govern these problems remains an active area of research [91]. Questions of how to use this perspective to better design algorithms [93] or how to use these ideas in other contexts, such as smoothing or machine learning [71], are currently in their infancy.

1.3.3 Sparsity

Sparsity, in its simplest definition, refers to a vector with only a small number of nonzero entries. Sparsity exploiting methods, broadly speaking, have been the subject of a vast breadth of research across statistics [342], signal processing [250], compressed sensing [89], and inverse problems [332]. In Chapter 4, we are interested in solving a linear inverse problem of the form

$$y = Au + \varepsilon, \tag{1.9}$$

where $u \in \mathbb{R}^d$, $y \in \mathbb{R}^n$, and $\varepsilon \sim \mathcal{N}(0, I_n)$ under various sparsity assumptions on u. Here we will outline several representative types of sparse vectors that will appear in Chapter 4.

- **Hard Sparsity:** We say a vector $u \in \mathbb{R}^d$ is s-sparse if u has at most s nonzero entries.
- ℓ^q **Sparsity:** We say a vector $u \in \mathbb{R}^d$ is ℓ^q-sparse if $\|u\|_q^q = \sum_{i=1}^d |u_i|^q \leq R_q$ for some $q \in [0, 1)$. Note that if $q = 0$ and $R_q = s$, this precisely recovers hard sparsity.
- **Group Sparsity:** Let the components of $u \in \mathbb{R}^d$ be partitioned into groups $\{g_1, \ldots, g_k\}$ where $g_1 \subseteq \{1, \ldots, d\}$ and $g_i \cap g_j = \emptyset$ for $i \neq j$. We say that u is s group-sparse if

at most s groups of indices g_i contain nonzero entries of u. One can also define an analogous notion of group ℓ^q-sparsity.

- **Sparse Representations:** We say that $u \in \mathbb{R}^d$ admits an s-sparse representation if there exists a tight frame $W \in \mathbb{R}^{d \times k}$ with $k \gg d$ such that $\alpha = W^T u \in \mathbb{R}^k$, where α is s-sparse. ℓ^q-sparse representations are naturally defined by assuming α to be ℓ^q-sparse.

There are natural notions of sparsity for a number of other objects such as matrices [70], functions [3], and operators [11], however these will not be discussed in this thesis.

1.4 Outline and Main Contributions

I provide an outline of the upcoming chapters and summarize the main contributions therein.

- Chapter 2 is a transcript of the published article [300]. We consider the formulation of Bayesian inverse problems on infinite-dimensional Hilbert spaces, where the forward model, prior mean function, and prior covariance operator must be discretized to make the problem computationally tractable. In the linear-Gaussian setting, we develop a general theory to characterize how these approximation errors propagate to the approximation to the posterior. We apply this theory to the heat deconvolution problem with both finite-element and graph-based discretizations. The second main contribution of this chapter to characterize the statistical complexity of estimating the posterior via a discretized ensemble approximation, as is done in widely-used ensemble Kalman inversion methods or within the analysis step of the Ensemble Kalman filter itself. We prove that the required ensemble size scales with the *effective dimension* of the continuum prior covariance operator (which is necessarily finite), and not the dimension of the discretization (which is substantially larger than the ensemble size). This result explains why ensemble Kalman methods perform well with a small ensemble size relative to the state dimension. Finally, for nonlinear forward models, we prove

using a Γ-convergence argument that discretized MAP estimators converge to their continuum limits as the discretizations are refined.

- Chapter 3 is a transcript of the published article [302]. We provide a state-of-the-are theory characterizing when the EnKF can be accurate as a state estimator. Previous EnKF accuracy guarantees in the literature have required either the system be fully-observed or that the dynamics be linear, and typically assume the ensemble size to be larger than the state dimension. These stringent assumptions make the previous state theory unrealistic, and unlikely to apply in practical scenarios. In this work, we provide a natural condition coupling the nonlinear dynamics and partial observations such that the estimation error of the EnKF remains small in the long-time horizon. The theory applies to the Lorenz models and the 2D Navier Stokes equations. An important feature of the result is that the required ensemble size is independent of the state dimension. This is the first state-estimation accuracy result for the EnKF as it is used in practice: with nonlinear dynamics, discrete-time partial observations, and a small ensemble size. We also prove a filter accuracy result for the EnKF with a surrogate dynamics model. The theory only requires short-time accuracy of the surrogate model, provided the true dynamics are sufficiently dissipative. This is the first EnKF accuracy result for surrogate dynamics models in any setting.

- Chapter 4 is a transcript of the published article [301]. We consider a popular family of sparsity-promoting hierarchical Bayesian models originally developed in the computational mathematics literature for inverse problems arising in medical imaging. While these models are widely used and have well-studied computational properties, we provide the first reconstruction error guarantees by adopting a new statistical perspective. Specifically, the analysis leverages techniques from the high-dimensional M-estimation literature along with a novel property of hierarchical Bayesian regularizers that we call approximate decomposability to obtain finite-sample bounds on the error attained by

maximum a posteriori estimators in reconstructing structured unknowns. Our theory demonstrates how hierarchical Bayesian models that exploit sparsity, group sparsity, and sparse representations of the unknown can achieve accurate parameter reconstructions in high-dimensional settings where the size of the parameter space vastly exceeds the number of observations.

While not included in this thesis, I have also contributed to the following manuscripts during my PhD:

- In [12], we develop a non-asymptotic estimation theory for sparse covariance operator estimation to explain the success of localization methods in the EnKF. This work extends the theory of sparse covariance matrix estimation to the infinite-dimensional settings in which data assimilation tasks are most naturally formulated. We propose a thresholding based estimator that exploits a natural continuum sparsity structure based on the L^q semi-norm of the covariance function. In the setting where the true covariance operator has a lengthscale parameter, our localized covariance estimator give an exponential improvement in sample complexity over the usual sample covariance when the lengthscale is sufficiently small, as is often the case in large-scale geophysical applications.

- In [11] consider the information-theoretic limits of sparse covariance operator estimation. For both the L^q-sparsity proposed in [12] and a class of study banded operators with kernels that decay rapidly off-the-diagonal, we derive minimax optimal lower bounds. Our bounds use a novel technique that lifts the theory from high-dimensional matrix estimation to the operator setting, and we identify the fundamental dimension-free quantities that determine the sample complexity. We demonstrate that thresholding and tapering estimators achieve the optimal rates for the L^q-sparsity and banded classes, respectively.

CHAPTER 2

ANALYSIS OF A COMPUTATIONAL FRAMEWORK FOR BAYESIAN INVERSE PROBLEMS: ENSEMBLE KALMAN UPDATES AND MAP ESTIMATORS UNDER MESH REFINEMENT

2.1 Introduction

Bayesian inverse problems on infinite-dimensional Hilbert spaces arise in numerous applications, including medical imaging, seismology, climate science, reservoir modeling, and mechanical engineering [81, 66, 202, 42, 113]. 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 to numerically solve 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 [66] to solve 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.

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 [66] together with graph-based discretizations. Finite elements are predominant in scientific computing solution of PDE-constrained inverse problems, see e.g. [209, 66, 42], whereas graph-based methods have been used, for instance, to solve inverse problems on manifolds [188, 187, 207] and in machine learning applications in semi-supervised learning [27, 161]. 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 [293], spatial statistics [312], and machine learning [347]. Efficient algorithms to sample Matérn priors within the computational framework of [66] have recently been investigated in [16]. Here, we obtain new operator-norm bounds for both finite element and graph-based discretizations. The former rely on classical theory, while the latter extend recent results from [303]. 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 [152], heat inversion has been widely adopted as a tractable testbed for theoretical and methodological developments, see e.g. [315, 161, 159, 187]. 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 [66] to algorithms for nonlinear inverse problems beyond the Markov chain Monte Carlo method for posterior sampling considered in [281]. Specifically, we investigate (i) ensemble Kalman methods [204, 98] where we show, building on [13], that the effective dimension which determines the required sample size remains bounded along mesh refinements; and (ii) MAP estimation [209, 126] where we show, building on [19], 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. [120, 7, 274, 159], 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 [121, 119, 269].

The well-posedness of the posterior measure under perturbations is one of the hallmarks of the Bayesian formulation of inverse problems [315, 228, 229, 298]. For a fixed prior measure, the error in the posterior measure caused by discretization of the forward model can be bounded in Hellinger distance [119] and in Kullback-Leibler divergence [253]. Posterior stability under perturbations to the prior measure in addition to perturbations to the likelihood have been recently investigated using Wasserstein distance [311] and more general integral probability metrics [157]. 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 [311] 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.

2.1.1 Outline and Main Contributions

- Section 2.2 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 2.3 presents a novel analysis of the error incurred in our discretization framework for

linear-Gaussian Bayesian inverse problems. Theorems 2.3.3 and 2.3.4 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, thus deriving error bounds for these popular discretization schemes.

- Section 2.4 formulates the ensemble Kalman update within our general discretization framework. The results in this section give non-asymptotic 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 2.5 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 2.6 closes with conclusions and directions for future work.

2.1.2 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 $a_n \leq cb_n$ 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 as $\|\vec{u}\|_A := \sqrt{\vec{u}^T A \vec{u}}$. Finally, $\|\cdot\|_2$ denotes the usual Euclidean norm on $\mathbb{R}^d$.

2.2 Problem Setting and Computational Framework

This section contains necessary background. Subsection 2.2.1 overviews the formulation of Bayesian inverse problems in Hilbert space. Subsection 2.2.2 describes the computational framework analyzed in this paper. Finally, Subsection 2.2.3 shows how finite element and graph-based methods for inverse problems in function space can be viewed as particular instances of the general framework.

2.2.1 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 2.5. 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 [319, 209, 298] 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

$$\langle v, \mathcal{C}_0 w\rangle_{\mathcal{H}} = \mathbb{E}^{u\sim\mu_0}\Big[\big\langle v, (u - m_0)\big\rangle_{\mathcal{H}}\big\langle (u - m_0), w\big\rangle_{\mathcal{H}}\Big], \qquad \forall v, w \in \mathcal{H}, \tag{2.3}$$

and $m_0 \in \mathcal{H}$ is assumed to belong to the Cameron-Martin space $E = \text{Im}(\mathcal{C}_0^{1/2}) \subset \mathcal{H}$. The Bayesian solution to the inverse problem is the conditional law of u given y, called the posterior probability measure μ_{post}. Application of Bayes' rule in infinite dimension [315] 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 2.2.2, we will review the computational framework in [66] 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 2.3.

2.2.2 Computational Framework

This subsection overviews the computational framework presented in [66]. 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 2.2.3 and 2.2.3, 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 with, 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 to Subsections 2.3.2 and 2.3.3 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 $\vec{u} \in \mathbb{R}_M^n$ is given by

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

Following the analogy with Subsection 2.2.1, we choose a Gaussian prior $\mu_0^n = \mathcal{N}(\vec{m}_0, C_0)$ on the space $\mathbb{R}_M^n$, with Lebesgue density

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

For the prior mean, we take $\vec{m}_0 = Pm_0$; for the prior covariance, we may take an invertible operator $C_0 : \mathbb{R}_M^n \to \mathbb{R}_M^n$ that approximates well $\mathcal{C}_0$, in the sense that $\|\mathcal{C}_0 - P^* C_0 P\|_{op}$ is small; we refer to Subsections 2.3.2 and 2.3.3 for examples arising from discretization of Matérn-type Gaussian processes. Similar to (2.3), the prior covariance C_0 satisfies, by definition, that

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

While we view μ_0^n as a Gaussian measure on $\mathbb{R}_M^n$, it is helpful to notice 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:

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

Since the space $\mathbb{R}_M^n$ 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 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 adjoint and matrix transpose. We present a detailed discussion of this distinction in Section 2.7. 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,

$$\vec{m}_{\text{post}} = \vec{m}_0 + C_0F^\natural(FC_0F^\natural + \Gamma)^{-1}(y - F\vec{m}_0), \tag{2.13}$$

$$C_{\text{post}} = C_0 - C_0F^\natural(FC_0F^\natural + \Gamma)^{-1}FC_0. \tag{2.14}$$

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:

$$\vec{m}_{\text{post}} = \vec{m}_0 + C_0^EF^T(FC_0^EF^T + \Gamma)^{-1}(y - F\vec{m}_0), \tag{2.15}$$

$$C^E_{\text{post}} = C_0^E - C_0^EF^T(FC_0^EF^T + \Gamma)^{-1}FC_0^E. \tag{2.16}$$

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 forward model give discretized quantities $F^\natural$ and C_0 that

approximate well in operator norm their infinite-dimensional counterparts $\mathcal{F}^*$ and $\mathcal{C}_0$; the same would not be true for the Euclidean analogs F^T and C_0^E.

2.2.3 Inverse Problems in Function Space

As particular instances of the computational framework overviewed in the previous subsection, here we consider discretization 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 2.2.1 and 2.2.2 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.

Domain	Basis	Discretization	Error bounds
$\mathcal{D}$	Lagrange finite elements	Subsection 2.2.3	Subsection 2.3.2
$\mathcal{M}$	Random geometric graphs	Subsection 2.2.3	Subsection 2.3.3

Table 2.1: Roadmap of the discretizations analyzed in this paper as particular examples of the general computational framework introduced in Subsection 2.2.2.

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 2.3.

Example 2.2.1 (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 [52, 16]. 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 $L^2(\Omega)$ [315].

Example 2.2.2 (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 Subsec-

tions 2.2.3 and 2.2.3. 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:

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

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 [161].

Finite Element Discretizations

The use of finite elements within the computational framework overviewed in Subsection 2.2.2 was proposed and numerically investigated in [66]. 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 $i, j \in \{1, \ldots, n\}$. The domain $\mathcal{D}$ is partitioned by a mesh into elements $e_1, e_2, \ldots$ 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 $d = 2$ 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 2.3.2 holds for uniform (and more generally for quasi-uniform) mesh refinements.

Example 2.2.3 (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

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

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

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

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 $1 \leq i \leq n$, see [271]. 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 2.3.2 that the matrix $C_0 : \mathbb{R}^n_M \to \mathbb{R}^n_M$ given by $C_0 = A^{-\alpha} = (K^{-1}M)^{\alpha}$ approximates well the continuum Matérn covariance operator $\mathcal{C}_0$.

Example 2.2.4 (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 $A^i(t)$:

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

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

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 $\vec{A}(t) =$

$\left(A^1(t), \ldots, A^n(t)\right)^T \in \mathbb{R}^n_M$. 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 $k = 0, \ldots, 1/\Delta t$ such that, for every $\phi \in \mathcal{V}_h$,

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

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

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 $k = 0, \ldots, 1/\Delta t - 1$:

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

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 $\mathcal{V}_h$.

Graph-Based Discretizations

We will now see how the graph-based discretizations considered in [161] 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 $\mathcal{M}_n = \{x_1, \ldots, x_n\} \subset \mathcal{M}$ 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 [158]:

Proposition 2.2.5 (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*

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

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 pre-image 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 2.2.5 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 coefficients vectors in $\mathbb{R}^n_M$, where the corresponding mass matrix is $M = \frac{1}{n} I_n$.

Example 2.2.6 (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 $W_{ij} \geq 0$ 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 $D = diag(d_1, \ldots, d_n)$ where $d_i = \sum_{j=1}^n W_{ij}$, the unnormalized graph Laplacian is given 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 [28, 68]. 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 $B_n = diag\big(b(x_1), \ldots, b(x_n)\big)$, we approximate the differential operator $\mathcal{A}$ as the matrix $A : \mathbb{R}_M^n \to \mathbb{R}_M^n$ 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 $C_0 : \mathbb{R}_M^n \to \mathbb{R}_M^n$ given by $C_0 = A^{-\alpha}$ will be shown to approximate well the continuum Matérn covariance operator $\mathcal{C}_0$ in Subsection 2.3.3.

Example 2.2.7 (Graph Approximation of Heat Forward Model). For our graph-based approximation of the heat forward model, we consider the differential equation

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

We let $G_n : \mathbb{R}_M^n \to \mathbb{R}_M^n$ map $u_n \in \mathbb{R}_M^n$ to the solution $v_n(1)$ to (2.37). Given the eigenpairs of the graph Laplacian, denoted $\big\{(\lambda_n^{(i)}, \psi_n^{(i)}\big\}_{i=1}^n$, we can explicitly write the solution to this 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 exactly know 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 $1 \leq j \leq d_y$, 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 $1 \leq j \leq d_y$, 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

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

2.3 Error Analysis

This section analyzes the computational framework introduced in Section 2.2. We start in Subsection 2.3.1 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 2.2.3, studying finite element discretizations (Subsection 2.3.2) and graph-based methods (Subsection 2.3.3).

2.3.1 Error Analysis: General Computational Framework

The main results of this subsection, Theorems 2.3.3 and Theorem 2.3.4 below, quantify the errors in the mean and covariance approximations

$$\varepsilon_m = \|m_{\text{post}} - P^* \vec{m}_{\text{post}}\|_{\mathcal{H}}, \tag{2.41}$$

$$\varepsilon_C = \|\mathcal{C}_{\text{post}} - P^* C_{\text{post}} P\|_{op}, \tag{2.42}$$

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 $\|\mathcal{C}_0 - P^* C_0 P\|_{op}$; (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 2.3.1. The following holds:

(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 $\|\mathcal{C}_0 - P^* C_0 P\|_{op} \leq r_1 \psi_1(n)$.

(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 $\|\mathcal{F} - FP\|_{op} \leq r_2 \psi_2(n)$.

(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 $u \in \mathcal{H}'$, $\|(I - P^*P)u\|_{\mathcal{H}} \leq r_3 \psi_3(n) \|u\|_{\mathcal{H}'}$.

In Subsections 2.3.2 and 2.3.3, we will verify Assumption 2.3.1 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 2.3.1 (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 (2.41) and (2.42), we will first prove a lemma similar to those in [13] and [224]. Following the approach in these papers, we introduce the Kalman gain operator

$$\mathscr{K} : S_+ \times B(\mathcal{H}, \mathbb{R}^{d_y}) \to B(\mathbb{R}^{d_y}, \mathcal{H}), \qquad \mathscr{K}(\mathcal{C}, \mathcal{F}) = \mathcal{C}\mathcal{F}^*(\mathcal{F}\mathcal{C}\mathcal{F}^* + \Gamma)^{-1}. \tag{2.43}$$

Unlike [13, 224], we view the Kalman gain operator as a function of both the forward map and the covariance, not just the covariance. The following lemma shows that the Kalman gain operator is pointwise continuous.

Lemma 2.3.2 (Continuity of Kalman Gain Update)**.** *Let $\mathscr{K}$ be the Kalman gain operator defined in* (2.43)*. 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:*

$$\|\mathscr{K}(\mathcal{C}_1, \mathcal{F}_1) - \mathscr{K}(\mathcal{C}_2, \mathcal{F}_2)\|_{op} \leq c_1\|\mathcal{C}_1 - \mathcal{C}_2\|_{op} + c_2\|\mathcal{F}_1 - \mathcal{F}_2\|_{op}, \tag{2.44}$$

where

$$\begin{aligned} c_1 &= \|\Gamma^{-1}\|_{op}\|\mathcal{F}_2\|_{op} + \|\Gamma^{-1}\|_{op}^2\|\mathcal{C}_1\|_{op}\|\mathcal{F}_1\|_{op}\|\mathcal{F}_2\|_{op}^2, \\ c_2 &= \|\Gamma^{-1}\|_{op}\|\mathcal{C}_1\|_{op} + \|\Gamma^{-1}\|_{op}^2\|\mathcal{C}_1\|_{op}^2\|\mathcal{F}_1\|_{op}^2 + \|\Gamma^{-1}\|_{op}\|\mathcal{C}_1\|_{op}^2\|\mathcal{F}_1\|_{op}\|\mathcal{F}_2\|_{op}. \end{aligned} \tag{2.45}$$

Proof. Applying Lemma A.8 from [13] (which was stated for matrices, but also holds in our

infinite-dimensional setting), we get that

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

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

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

and

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

Combining the previous three displayed inequalities gives the desired result. □

Theorem 2.3.3 (Mean Approximation Error)**.** *Consider the error ε_m in* (2.41) *between the true posterior mean in* (2.5) *and its finite-dimensional posterior approximation in* (2.13)*. Then, under Assumption 2.3.1 it holds that*

$$\varepsilon_m \leq r_1'\psi_1(n)+r_2'\psi_2(n)+r_3'\psi_3(n), \tag{2.46}$$

where

$$\begin{aligned}r_1' &= 2c_1r_1\|y-\mathcal{F}m_0\|_2, \\ r_2' &= 2c_2r_2\|y-\mathcal{F}m_0\|_2+r_2\|m_0\|_{\mathcal{H}}\|\mathscr{K}(\mathcal{C}_0,\mathcal{F})\|_{op}, \\ r_3' &= r_3\|m_0\|_{\mathcal{H}'},\end{aligned}$$

and c_1, c_2 *are defined in* (2.45).

Proof. By the definition of ε_m in (2.41) and an application of the triangle inequality,

$$\begin{aligned}\varepsilon_m \leq \|m_0 - P^*\vec{m}_0\|_{\mathcal{H}} + \|\mathscr{K}(\mathcal{C}_0, \mathcal{F}) - \mathscr{K}(P^*C_0P, FP)\|_{op}\|y - F\vec{m}_0\|_2 \\ + \|\mathscr{K}(\mathcal{C}_0, \mathcal{F})\|_{op}\|\mathcal{F}m_0 - F\vec{m}_0\|_2.\end{aligned} \tag{2.47}$$

We next bound each of the terms on the right-hand side. By Assumption 2.3.1 (iii) and the fact that $\vec{m}_0 = Pm_0$, we have that

$$\|m_0 - P^*\vec{m}_0\|_{\mathcal{H}} = \|m_0 - P^*Pm_0\|_{\mathcal{H}} \leq r_3\|m_0\|_{\mathcal{H}'}\psi_3(n). \tag{2.48}$$

By Lemma 2.3.2 and Assumption 2.3.1 (i) and (ii), we have that

$$\begin{aligned}\|\mathscr{K}(\mathcal{C}_0, \mathcal{F}) - \mathscr{K}(P^*C_0P, FP)\|_{op} &\leq c_1\|\mathcal{C}_0 - P^*C_0P\|_{op} + c_2\|\mathcal{F} - FP\|_{op} \\ &\leq c_1r_1\psi_1(n) + c_2r_2\psi_2(n).\end{aligned} \tag{2.49}$$

By Assumption 2.3.1 (ii) and (iii) we have that

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

This implies that for n sufficiently large, it holds that

$$\|y - F\vec{m}_0\|_2 \leq 2\|y - \mathcal{F}m_0\|_2. \tag{2.50}$$

Finally, by Assumption 2.3.1 (ii) we have that

$$\|\mathcal{F}m_0 - F\vec{m}_0\|_2 \leq \|\mathcal{F} - FP\|_{op}\|m_0\|_{\mathcal{H}} \leq r_2\|m_0\|_{\mathcal{H}}\psi_2(n). \tag{2.51}$$

Plugging in the bounds (2.48), (2.49), (2.50), and (2.51) into inequality (2.47) gives the desired result. □

Theorem 2.3.4 (Covariance Approximation Error)**.** *Consider the error ε_C in* (2.42) *between the true posterior covariance operator in* (2.6) *and its finite-dimensional posterior approximation in* (2.14)*. Then, under Assumption 2.3.1 it holds that*

$$\varepsilon_C \leq r_1' \psi_1(n) + r_2' \psi_2(n), \tag{2.52}$$

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 in (2.45)*.*

Proof. By the definition of ε_C in (2.42) and an application of the triangle inequality,

$$\begin{aligned} \varepsilon_C \leq \|\mathcal{C}_0 - P^* C_0 P\|_{op} + \|\mathscr{K}(\mathcal{C}_0, \mathcal{F}) - \mathscr{K}(P^* C_0 P, FP)\|_{op} \|FC_0 P\|_{op} \\ + \|\mathscr{K}(\mathcal{C}_0, \mathcal{F})\|_{op} \|\mathcal{F}\mathcal{C}_0 - FC_0 P\|_{op}. \end{aligned} \tag{2.53}$$

By Assumption 2.3.1 (i) we have that

$$\|\mathcal{C}_0 - PC_0 P\|_{op} \leq r_1 \psi_1(n). \tag{2.54}$$

As in the previous proof, applying Lemma 2.3.2 and using Assumption 2.3.1 (i) and (ii) gives

$$\|\mathscr{K}(\mathcal{C}_0, \mathcal{F}) - \mathscr{K}(P^* C_0 P, FP)\|_{op} \leq c_1 r_1 \psi_1(n) + c_2 r_2 \psi_2(n). \tag{2.55}$$

The triangle inequality along with Assumption 2.3.1 (i) and (ii) yield the bound

$$\begin{aligned}\|\mathcal{F}\mathcal{C}_0 - FC_0P\|_{op} &\leq \|FP\|_{op}\|\mathcal{C}_0 - P^*C_0P\|_{op} + \|\mathcal{F} - FP\|_{op}\|\mathcal{C}_0\|_{op} \\ &\leq r_1\|FP\|_{op}\psi_1(n) + r_2\|\mathcal{C}_0\|_{op}\psi_2(n).\end{aligned} \tag{2.56}$$

Assumption 2.3.1 (i) and (ii) guarantee that, for sufficiently large n,

$$\|FP\|_{op} \leq 2\|\mathcal{F}\|_{op}, \text{ and } \|FC_0P\|_{op} \leq 2\|\mathcal{F}\|_{op}\|\mathcal{C}_0\|_{op}. \tag{2.57}$$

Plugging in the bounds (2.54), (2.55), (2.56), and (2.57) into (2.53) gives the desired result.

□

2.3.2 Error Analysis: Finite Element Discretizations

The following result applies Theorems 2.3.3 and 2.3.4 to quantify the errors in the finite element approximations to the posterior mean and covariance operators in the setting considered in Subsection 2.2.3.

Theorem 2.3.5 (Finite Element Posterior Mean and Covariance Approximation)**.** *Consider the discretization proposed in Subsection 2.2.3 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 $\Theta(x) \in C^1(\bar{\mathcal{D}})$, $b(x) \in L^\infty(\mathcal{D})$, and assume that both functions are (almost surely) bounded below by positive constants. The errors in the mean and covariance approximations are bounded by*

$$\begin{aligned}\|m_{\text{post}}(x) - P^*\vec{m}_{\text{post}}\|_{L^2(\mathcal{D})} &\lesssim h^{\min\{2,s\}} + \Delta t^2, \\ \|\mathcal{C}_{\text{post}} - P^*C_0P\|_{op} &\lesssim h^2 + \Delta t^2.\end{aligned}$$

This result follows from Theorems 2.3.3 and 2.3.4 upon verifying that Assumption 2.3.1

hold 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 2.3.1 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.

Finite Element Approximation of Matérn-Type Prior Covariance

Here we show that the finite element approximation to the covariance operator given in Subsection 2.2.3 satisfies Assumption 2.3.1 (i).

Theorem 2.3.6 (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 2.2, and let $A^{-\alpha} : \mathbb{R}^n_M \to \mathbb{R}^n_M$ be the finite-dimensional approximation defined in Subsection 2.2.3 corresponding to a finite element space $\mathcal{V}_h$ with piecewise linear basis functions. Let $\Theta(x) \in C^1(\bar{\mathcal{D}})$, $b(x) \in L^\infty(\mathcal{D})$, and assume that both functions are (almost surely) bounded below by positive constants. Then, there exists a constant c independent of h such that*

$$\|\mathcal{C}_0 - P^* A^{-\alpha} P\|_{op} \leq ch^2. \tag{2.58}$$

The proof can be found in Section 2.8. This shows that the finite-dimensional prior approximation proposed in Subsection 2.2.3 satisfies Assumption 2.3.1 (i) with $\psi_1(h) = h^2$.

Finite Element Approximation of Heat Forward Model

We now show that the discretized finite element forward map given in (2.32) satisfies Assumption 2.3.1 (ii).

Theorem 2.3.7 (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*

$$\|\mathcal{F} - FP\|_{op} \leq c_1 h^2 + c_2 \Delta t^2. \tag{2.59}$$

The proof is given in Section 2.8. We see that Assumption 2.3.1 (ii) holds for the Crank-Nicolson discretization with $\psi_2(h) = h^2 + \Delta t^2$.

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 2.3.1 (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 quasiuniform mesh refinement of $\mathcal{D}$ with linear Lagrange basis functions, [271, 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

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

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 $m_0 \in H^s(\mathcal{D})$,

$$\|m_0 - P^*Pm_0\|_{L^2(\mathcal{D})} \leq \|m_0 - U\|_{L^2(\mathcal{D})} \leq ch^{\min\{2,s\}}\|m_0\|_{H^s(\mathcal{D})}. \tag{2.60}$$

We see that Assumption 2.3.1 (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 2.3.1 holds for the finite element discretization, proving Theorem 2.3.5.

2.3.3 Error Analysis: Graph-Based Approximations

The following result applies Theorems 2.3.3 and 2.3.4 to quantify the error in the graph-based approximations to posterior mean and covariance operator in the setting considered in Subsection 2.2.3.

Theorem 2.3.8 (Graph-Based Posterior Mean and Covariance Approximation)**.** *Consider the discretization proposed in Subsection 2.2.3. Assume we are in a realization in which the conclusion of Proposition 2.2.5 holds and we have that $\alpha > (5d+1)/4$. Let $b(x)$ be Lipschitz, $\Theta(x) \in C^1(\mathcal{M})$, 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 2.3.3 and 2.3.4 upon verifying that the graph-based discretization satisfies Assumption 2.3.1 for our model problem. Verifying these assumptions

will be the focus of the following subsections. In all that follows, we assume to be in a realization where the conclusion of Proposition 2.2.5 holds.

Graph-Based Approximation of Matérn-Type Prior Covariance

The following theorem shows that the covariance operator constructed in Section 2.2.3 satisfies Assumption 2.3.1 (ii). The proof can be found in Section 2.9.

Theorem 2.3.9 (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 2.3.10. The proof of Theorem 2.3.9 in Section 2.9 follows that of Theorem D.1 in [303]. 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 2.2.3 satisfies Assumption 2.3.1 (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.

Graph-Based Approximation of Heat Forward Model

We show that the graph-based approximation of the heat forward model satisfies with the surrogate observation operator satisfies Assumption 2.3.1 (iii). The proof can be found in

Section 2.9.

Theorem 2.3.11 (Operator-Norm Error for Graph-Based Forward Map Approximation)**.** *Let $F = \mathcal{O}_n G_n$ be the forward model defined in Subsection 2.2.3. Take $h_n \asymp \sqrt{\frac{(\log n)^{c_d}}{n^{1/d}}}$. Then,*

$$\|\mathcal{F} - P^* F P\|_{op} \lesssim (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}}. \tag{2.61}$$

We have shown that the graph-based finite-dimensional approximation to the forward model proposed in Subsection 2.2.3 satisfies Assumption 2.3.1 (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.

Graph-Based Orthogonal Projection

Finally, we verify that the orthogonal projection of the prior mean function onto the subspace $\mathcal{V} = \text{span}\{\mathbb{1}_{U_i}(x)\}_{i=1}^n$ converges as in Assumption 2.3.1 (iii). Since the mean function must lie in the Cameron-Martin space $E = \text{Im}(\mathcal{C}_0^{1/2})$, we necessarily have that $m_0(x) \in H^1(\mathcal{M})$. The following result from [158, Lemma 12] quantifies the convergence rate of the projected mean function:

Lemma 2.3.12 (Lemma 12 in [158])**.** *There exists a constant $c_{\mathcal{M}}$ independent of n such that, for every $m_0 \in H^1(\mathcal{M})$, we have*

$$\|m_0 - P^* P m_0\|_{L^2(\mathcal{M})} \leq c_{\mathcal{M}} \varepsilon_n \|m_0\|_{H^1(\mathcal{M})} \lesssim \frac{(\log n)^{c_d}}{n^{1/d}} \|m_0\|_{H^1(\mathcal{M})}. \tag{2.62}$$

We see that Assumption 2.3.1 (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 2.3.1 holds for the graph-based discretization, proving Theorem 2.3.8.

2.4 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.

2.4.1 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

$$u^{(j)} = \vec{m}_0 + L\xi^{(j)}, \quad u^{(j)} \sim \mathcal{N}(\vec{m}_0, C_0), \tag{2.63}$$

for $1 \leq j \leq J$, 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

$$v^{(j)} = u^{(j)} + \widehat{C}_0 F^\natural \left(F\widehat{C}_0 F^\natural + \Gamma\right)^{-1} \left(y - Fu^{(j)} + \eta^{(j)}\right), \quad \eta^{(j)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma). \tag{2.64}$$

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

$$\widehat{m}_{\text{post}} = \frac{1}{J}\sum_{j=1}^{J} v^{(j)}, \text{ and } \widehat{C}_{\text{post}} = \frac{1}{J-1}\sum_{J=1}^{J}(v^{(j)} - \widehat{m}_{\text{post}})(v^{(j)} - \widehat{m}_{\text{post}})^T M. \tag{2.65}$$

2.4.2 Ensemble Kalman Approximation: Error Analysis

We now want to derive non-asymptotic expectation bounds for the error in the ensemble approximation to the infinite-dimensional posterior. We consider the errors

$$\widehat{\varepsilon}_m = \mathbb{E}\,\|m_{\text{post}} - P^*\widehat{m}_{\text{post}}\|_{\mathcal{H}}, \tag{2.66}$$

$$\hat{\varepsilon}_C = \mathbb{E}\,\|\mathcal{C}_{\text{post}} - P^*\hat{C}_{\text{post}}P\|_{op}. \tag{2.67}$$

Controlling these errors will amount to controlling the error between the ensemble approximations and the finite-dimensional approximations using the theory from [13] and then applying our results from Section 2.3. 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

$$r_M(C_0) := \frac{\text{Tr}(C_0)}{\|C_0\|_{op}}, \tag{2.68}$$

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 [331]. We will show in Subsection 2.4.3 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 $r(\mathcal{C}_0) = \frac{\text{Tr}(\mathcal{C}_0)}{\|\mathcal{C}_0\|_{op}}$, 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 2.4.1 (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* (2.66)*. Assume that the finite-dimensional approximations satisfy Assumption 2.3.1, that for n sufficiently large $r_M(C_0) \leq cr(\mathcal{C}_0)$ for some constant c independent of n, and that $J \geq r_M(C_0)$. Then, it holds that*

$$\widehat{\varepsilon}_m \leq r_1' \psi_1(n) + r_2' \psi_2(n) + r_3' \psi_3(n) + r_4' \frac{1}{\sqrt{J}}, \tag{2.69}$$

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

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

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

The first term on the right-hand side is deterministic and is bounded by Theorem 2.3.3:

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

For the second term on the right-hand side, we recall [13, Theorem 3.3], which guarantees the following:

$$\mathbb{E}\,\|\widehat{m}_{\text{post}} - \vec{m}_{\text{post}}\|_M \lesssim c_n' \sqrt{\frac{r_M(C_0)}{J}} + c_n'' \sqrt{\frac{r_2(\Gamma)}{J}}, \tag{2.70}$$

where $c_n' = \left(\|C_0\|_{op}^{1/2} \vee \|C_0\|_{op}^2\right) \left(\|F\|_{op} \vee \|F\|_{op}^4\right) \left(\|\Gamma^{-1}\|_{op} \vee \|\Gamma^{-1}\|_{op}^2\right) (1 \vee \|y - F\vec{m}_0\|_2)$, 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 2.3.1 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:

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

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

Theorem 2.4.2 (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* (2.66)*. Assume that the finite-dimensional approximations satisfy Assumption 2.3.1, that for n sufficiently large $r_M(C_0) \leq cr(\mathcal{C}_0)$ for some constant c independent of n, and that $J \geq r_M(C_0)$. Then, it holds that*

$$\widehat{\varepsilon}_C \leq r_1'\psi_1(n) + r_2'\psi_2(n) + r_3'\frac{1}{\sqrt{J}}, \tag{2.71}$$

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 2.4.1. We decompose the error as

$$\mathbb{E}\,\|\mathcal{C}_{\text{post}} - P^*\widehat{C}_{\text{post}}P\|_{op} \leq \mathbb{E}\,\|\mathcal{C}_{\text{post}} - P^*C_{\text{post}}P\|_{op} + \mathbb{E}\,\|C_{\text{post}} - \widehat{C}_{\text{post}}\|_{op}. \tag{2.72}$$

The first term on the right-hand side is deterministic and is controlled by Theorem 2.3.4:

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

For the second term, we use [13, Theorem 3.5], which guarantees that

$$\mathbb{E}\,\|\widehat{C}_{\text{post}} - C_{\text{post}}\|_{op} \lesssim c_n' \sqrt{\frac{r_M(C_0)}{J}} + c_n'' \left(\sqrt{\frac{r_M(C_0)}{J}} \vee \sqrt{\frac{r_2(\Gamma)}{J}} \right), \tag{2.73}$$

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 2.3.1 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:

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

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

The results in [13] 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.

2.4.3 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 2.2.3. To get an upper bound for 2.68, 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 2.4.3 (Effective Dimension Upper Bound)**.** *Let the assumptions of Theorem 2.3.6 hold. Then, for h sufficiently small, there exists a constant τ independent of h such that*

$$r_M(C_0) \leq \frac{\mathrm{Tr}(\mathcal{C}_0)}{\|\mathcal{C}_0\|_{op} - \tau h^2}. \tag{2.74}$$

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

$$r_M(C_0) \leq c r(\mathcal{C}_0). \tag{2.75}$$

Proof. By Theorem 2.3.6 we have that

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

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

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

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

$$\|C_0\|_{op} \geq \|\mathcal{C}_0\|_{op} - \tau h^2. \tag{2.76}$$

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 $\{\lambda^{(i)}\}_{i=1}^\infty$ 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 to [313, 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 $\lambda^{(i)} \leq \lambda_h^{(i)}$. It then follows that

$$\mathrm{Tr}(C_0) = \sum_{i=1}^{n} (\lambda_h^{(i)})^{-\alpha} \leq \sum_{i=1}^{n} (\lambda^{(i)})^{-\alpha} \leq \sum_{i=1}^{\infty} (\lambda^{(i)})^{-\alpha} = \mathrm{Tr}(\mathcal{C}_0). \tag{2.77}$$

Combining (2.76) and (2.77) gives (2.74), completing the proof. □

Remark 2.4.4. 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 Theorem 2.9.1 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 [159].

2.5 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 [19], from which we will show Γ-convergence of the relevant Onsager-Machlup functions and consequently convergence of

the MAP estimators (up to subsequences). Under reasonable conditions on the nonlinear forward model (see Assumption 2.7 in [315]), the posterior can still be characterized as a change of measure with respect to the prior as in (2.4):

$$\frac{d\mu_{\text{post}}}{d\mu_0}(u) = \frac{1}{Z}\exp\left(-\frac{1}{2}\|y - \mathcal{F}(u)\|^2_{\Gamma^{-1}}\right), \tag{2.78}$$

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 maximum a posteriori (MAP) estimator. In infinite dimensions, the notion of a MAP estimator was introduced in [126] as the maximizer of a small ball probability. The theory of MAP estimation in function spaces has been further refined in [222, 227, 217, 19, 226]. 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

$$\lim_{\delta \to 0} \frac{\mu_{\text{post}}\left(B^\delta(u^{\text{MAP}})\right)}{M_\delta} = 1, \quad M_\delta = \sup_{u \in \mathcal{H}} \mu_{\text{post}}\big(B^\delta(u)\big). \tag{2.79}$$

It is shown in [226] that the MAP estimators of (2.78) 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 Onsager-Machlup (OM) functional $I_{\text{post}} : \mathcal{H} \to \mathbb{R}$ given by

$$I_{\text{post}}(u) = \begin{cases} \Phi(u) + I_0(u), & \text{if } u - m_0 \in \text{Im}(\mathcal{C}_0^{1/2}), \\ +\infty, & \text{otherwise,} \end{cases} \tag{2.80}$$

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 (2.80) may not coincide with MAP estimators defined in (2.79), and a weaker notion of MAP estimator is required [194]. The optimization problem of minimizing (2.80) is in general difficult, if not impossible, to solve analytically. The computational framework put forth in Subsection 2.2.2 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:

$$I^{(n)}_{\text{post}}(u) = \begin{cases} \Phi^{(n)}(u) + I^{(n)}_0(u), & \text{if } u - m \in \text{Im}((P^*C_0P)^{1/2}), \\ +\infty, & \text{otherwise,} \end{cases} \tag{2.81}$$

where

$$\Phi^{(n)}(u) = \frac{1}{2}\|y - F(Pu)\|^2_{\Gamma^{-1}}, \quad \text{and} \quad I^{(n)}_0 = \frac{1}{2}\|(P^*C_0P)^{\dagger/2}(u - m_0)\|^2_{\mathcal{H}}.$$

Here $F : \mathbb{R}^n_M \to \mathbb{R}^k$ is a discretized approximation to $\mathcal{F}$ and the operator $(P^*C_0P)^{\dagger/2} = \sum_{i=1}^n (\lambda_n^{(i)})^{-\alpha/2} P^*\psi_n^{(i)} \otimes P^*\psi_n^{(i)}$ is the Moore-Penrose pseudoinverse of $(P^*C_0P)^{1/2}$. We remark that $u - m \in \text{Im}((P^*C_0P)^{1/2})$ if and only if $u, m \in \mathcal{V}$, so we can equivalently write (2.81) as $I^{(n)}_{\text{post}} : \mathbb{R}^n_M \to \mathbb{R}$, with

$$I^{(n)}_{\text{post}}(\vec{u}) = \frac{1}{2}\|y - F(\vec{u})\|^2_{\Gamma^{-1}} + \frac{1}{2}\|C_0^{-1/2}(\vec{u} - \vec{m}_0)\|^2_M, \tag{2.82}$$

which can be minimized by methods from the breadth of literature on finite-dimensional optimization problems. To apply the results in [19], we first review some preliminary definitions and results regarding Γ-convergence.

Definition 2.5.1. 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 $\mathcal{H}$,

$$I(u) \leq \liminf_{n\to\infty} I_n(u_n); \tag{2.83}$$

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

$$I(u) \geq \limsup_{n\to\infty} I_n(u_n). \tag{2.84}$$

Definition 2.5.2. 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 2.5.3. 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 $n \geq N$ 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 is implied by uniform convergence of I_n to I in the case that I is continuous [124]. The following classical result can be found for instance in [56].

Proposition 2.5.4. *Let $I_n, I : \mathcal{H} \to \overline{\mathbb{R}}$ with $\Gamma\text{-}\lim_{n\to\infty} I_n = I$, and $\{I_n\}_{n=1}^{\infty}$ equicoercive. Then,*

$$\min_{\mathcal{H}} I = \lim_{n\to\infty} \inf_{\mathcal{H}} I_n, \tag{2.85}$$

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 2.3.2 and the results in [19] to prove convergence of the finite element discretized MAP estimators as the mesh is refined:

Theorem 2.5.5 (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 2.3.1 (i), and let the orthogonal projection operator P satisfy Assumption 2.3.1 (iii). Assume additionally that the approximate forward model is such that $\Phi^{(n)}$ converges to Φ continuously as $n \to \infty$. Then, the corresponding sequence $\{I^{(n)}_{\text{post}}\}_{n=1}^{\infty}$ of OM functionals given in* (2.82) *satisfies that*

$$\Gamma\text{-}\lim_{n\to\infty} I^{(n)}_{\text{post}} = I_{\text{post}}, \tag{2.86}$$

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. Assumptions 2.3.1 (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 [19] to conclude that $I_0 = \Gamma\text{-}\lim_{n\to\infty} I^{(n)}_0$ and that the sequence $\{I^{(n)}_0\}_{n=1}^{\infty}$ is equicoercive. Then, under the assumption that $\Phi^{(n)} \to \Phi$ continuously as $n \to \infty$, applying Theorem 6.1 in [19] 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 2.5.4 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 Assumptions 2.3.1 (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 [19] shows that if the approximate forward map is of the form $\mathcal{F}(P^*P\cdot)$, then continuous 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 2.5.7 that continuous convergence —as well as Assumptions 2.3.1 (i) and (iii)— can still be verified.

Example 2.5.6 (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 overviews Eulerian data assimilation for the Navier-Stokes equations, an important model problem in numerical weather forecasting. We refer to [121, 119] for a more detailed discussion.

As described in [292, 322], 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

$$\frac{dv}{dt} = \nu \mathsf{A} v + \mathsf{B}(v,v) = f, \quad v(0) = u, \tag{2.87}$$

on the Hilbert space

$$\mathcal{H} = \left\{ v \in L^2_{\text{per}}(\mathcal{D}) : \int_{\mathcal{D}} v \, dx = 0, \nabla \cdot v = 0 \right\} \tag{2.88}$$

equipped with the standard $L^2(\mathcal{D})$ inner product. The Stokes operator A in (2.87) is self-adjoint, positive, 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 [321]. For $u \in \mathcal{H}$ and f sufficiently regular [292, 119], there exists a unique solution to (2.87) such that $v \in L^\infty\big(0,T;\mathcal{H}^{1+s}\big) \subset L^\infty\big(0,T;L^\infty(\mathcal{D})\big)$.

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

$$y_k = v(x_k, t) + \eta_k, \quad k = 1, \dots, K, \tag{2.89}$$

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

$$\mathcal{F}(u) = \Big(v(x_1, t)^T, \dots, v(x_K, t)^T\Big)^T \in \mathbb{R}^{d_y}, \tag{2.90}$$

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 (2.78).

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 2.2.2 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 $C_0 = diag\Big((\lambda^{(1)})^{-\alpha}, \dots, (\lambda^{(n)})^{-\alpha}\Big)$. 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) ordinary differential equation

$$\frac{dv^N}{dt} + \nu \mathsf{A} v^N + P\mathsf{B}(v^N, v^N) = Pf, \quad v^N(0) = Pu. \tag{2.91}$$

We then define

$$F(u) = \Big(v^N(x_1, t), \dots, v^N(x_{d_y}, t)\Big)^T. \tag{2.92}$$

Given these approximations, we have a discretized OM functional exactly of the form (2.82).

For the Eulerian data assimilation problem summarized in Example 2.5.6 and further

detailed in [119], we can verify the assumptions of Theorem 2.5.5 to obtain the following result. The proof can be found in Section 2.10.

Theorem 2.5.7 (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 2.5.6. Then, as $n \to \infty$, the sequence $\{I_{\text{post}}^{(n)}\}_{n=1}^{\infty}$ of discretized OM functionals satisfies that*

$$\Gamma\text{-}\lim_{n\to\infty} I_{\text{post}}^{(n)} = I_{\text{post}}, \tag{2.93}$$

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

2.6 Conclusion and Future Directions

This paper analyzed a computational framework to solve 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 to obtain 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.

2.7 Adjoints in Weighted Space

Let X and Y be two Hilbert spaces. Recall that [316], given $A \in B(X, Y)$, the adjoint of A is the unique map $A^* \in B(Y, X)$ such that

$$\langle Ax, y\rangle_Y = \langle x, A^* y\rangle_X, \qquad \forall x \in X, \forall y \in Y. \tag{2.94}$$

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 (2.94) 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 $F^\natural = M^{-1}F^T$.

2.8 Proofs Subsection 2.3.2

This section contains the proofs of Theorems 2.3.6 and 2.3.7 in Subsection 2.3.2. The proofs rely on classical finite element method convergence results from [271].

Proof of Theorem 2.3.6. 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

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

by [60, Theorem 14.3.3] and boundedness of the PDE solution operator from $L^2(\mathcal{D})$ to $H^2(\mathcal{D})$ (see [146] 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:

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

and the second follows from submultiplicativity 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 2.3.7. 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 Theorem [271, 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. □

2.9 Proofs Subsection 2.3.3

This section contains the proofs of Theorems 2.3.9 and 2.3.11 in Subsection 2.3.3. The proofs rely on existing spectral convergence results from [303]. 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 [303]:

Proposition 2.9.1. *Suppose $k = k_n$ is such that $h_n\sqrt{\lambda^{(k)}} \ll 1$ for large n. Then,*

$$\frac{|\lambda_n^{(k)} - \lambda^{(k)}|}{\lambda^{(k)}} \leq c\left(\frac{\varepsilon_n}{h_n} + h_n\sqrt{\lambda^{(k)}}\right), \tag{2.95}$$

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

Proposition 2.9.2. *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 $j = k_n, \ldots, k_n + \ell - 1$,*

$$\|P^*\psi_n^{(j)} - \psi^{(j)}\|_{L^2(\mathcal{M})} \leq cj^{3/2}\left(\frac{\varepsilon_n}{h_n} + h_n\sqrt{\lambda^{(j)}}\right)^{1/2}, \tag{2.96}$$

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

Proof of Theorem 2.3.9. Let $f \in L^2(\mathcal{M})$. Note that k_n was chosen such that we can apply (2.9.2) and (2.9.1) 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, introduce four intermediate quantities:

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

We will bound the difference between each two consecutive functions. By Weyl's law [90, Theorem 72], we have that

$$
\|u - u^{k_n}\|_{L^2(\mathcal{M})} \leq \left(\sum_{i=k_n+1}^{\infty} (\lambda^{(i)})^{-2\alpha} \right)^{\frac{1}{2}} \lesssim \left(\sum_{i=k_n+1}^{\infty} i^{-\frac{4\alpha}{d}} \right)^{\frac{1}{2}} \lesssim \left(\int_{k_n}^{\infty} x^{-\frac{4\alpha}{d}} \right)^{\frac{1}{2}} \lesssim k_n^{\frac{1}{2} - \frac{2\alpha}{d}}. \tag{2.97}
$$

By (2.9.1) 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

$$
\|u_n - u_n^{k_n}\|_{L^2(\mathcal{M})} = \left(\sum_{i=k_n+1}^{n} (\lambda_n^{(i)})^{-2\alpha} |\langle f, P^* \psi_n^{(i)} \rangle_{L^2(\mathcal{M})}|^2 \right)^{\frac{1}{2}} \leq (\lambda_n^{(k_n)})^{-\alpha} \lesssim k_n^{-\frac{2\alpha}{d}}. \tag{2.98}
$$

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 2.9.1 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

$$\begin{aligned}
\|u_n^{k_n} - \tilde{u}_n^{k_n}\|_{L^2(\mathcal{M})} &\leq \left(\sum_{i=1}^{k_n} \left[(\lambda_n^{(i)})^{-\alpha} - (\lambda^{(i)})^{-\alpha}\right]^2\right)^{\frac{1}{2}} \\
&\lesssim \left(\sum_{i=1}^{k_n} (\lambda^{(i)})^{-2\alpha} \left(\frac{\varepsilon_n}{h_n} + h_n\sqrt{\lambda^{(i)}}\right)^2\right)^{\frac{1}{2}} \\
&\lesssim \left(\frac{\varepsilon_n}{h_n} + h_n\right) \left(\sum_{i=1}^{k_n} (\lambda^{(i)})^{-2\alpha+1}\right)^{\frac{1}{2}} \lesssim \left(\frac{\varepsilon_n}{h_n} + h_n\right),
\end{aligned} \tag{2.99}$$

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 (2.9.2) to deduce that

$$\begin{aligned}
\|\tilde{u}^{k_n} - \tilde{u}_n^{k_n}\|_{L^2(\mathcal{M})} &\lesssim \sum_{i=1}^{k_n} (\lambda^{(i)})^{-\alpha} \|\psi^{(i)} - P^*\psi_n^{(i)}\|_{L^2(\mathcal{M})} \lesssim \sum_{i=1}^{k_n} (\lambda^{(i)})^{-\alpha} i^{\frac{3}{2}} \sqrt{\frac{\varepsilon_n}{h_n} + h_n\sqrt{\lambda^{(i)}}} \\
&\lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n} \sum_{i=1}^{k_n} i^{\frac{3}{2}} (\lambda^{(i)})^{-\alpha+\frac{1}{4}} \lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n} \sum_{i=1}^{k_n} i^{\frac{3}{2} - \frac{2\alpha}{d} + \frac{1}{2d}} \\
&\lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n},
\end{aligned} \tag{2.100}$$

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

$$\begin{aligned} \|u^{k_n} - \tilde{u}^{k_n}\|_{L^2(\mathcal{M})} &= \left\| \sum_{i=1}^{k_n} \langle f, \psi^{(i)} - P^* \psi_n^{(i)} \rangle_{L^2(\mathcal{M})} (\lambda^{(i)})^{-\alpha} \psi^{(i)} \right\|_{L^2(\mathcal{M})} \\ &\lesssim \sum_{i=1}^{k_n} (\lambda^{(i)})^{-\alpha} \|\psi^{(i)} - P^* \psi_n^{(i)}\|_{L^2(\mathcal{M})} \lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n}\,, \end{aligned} \tag{2.101}$$

where the last step follows by the same argument as for the previous expression if $\alpha > \frac{5}{4}d + \frac{1}{4}$. Combining (2.97), (2.98), (2.99), (2.100), and (2.101), 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 2.3.11. Let $u \in L^2(\mathcal{M})$ with $\|u\|_{L^2(\mathcal{M})} \neq 0$. We write

$$v = \mathcal{G}u = \sum_{i=1}^{\infty} \exp\left(-\lambda^{(i)}\right) \langle u, \psi^{(i)} \rangle_{L^2(\mathcal{M})} \psi^{(i)}, \tag{2.102}$$

$$v_n = G_n(u_n) = \sum_{i=1}^{n} \exp\left(-\lambda_n^{(i)}\right) \langle Pu, \psi_n^{(i)} \rangle_{L^2(\mathcal{M})} P^* \psi_n^{(i)}. \tag{2.103}$$

We then consider the quantity

$$\|\mathcal{F}u - FPu\|_2 = \|\mathcal{O}v - \mathcal{O}^n v_n\|_2 \leq \|\mathcal{O}\|_{op} \|v - v_n\|_{L^2(\mathcal{D})} + \|\mathcal{O}v_n \mathcal{O}^n v_n\|_{L^2(\mathcal{M})}. \tag{2.104}$$

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 two consecutive functions, as in the proof of Theorem 2.3.9. By Weyl's law, we have that

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

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

$$
\|v - v^{k_n}\|_{L^2(\mathcal{M})} \leq \left(\int_{k_n}^{\infty} x^{-4\alpha/d} \, dx \right)^{\frac{1}{2}} \lesssim k_n^{\frac{1}{2} - \frac{2\alpha}{d}}. \tag{2.105}
$$

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 (2.9.1) 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

$$\begin{aligned} \|v_n - v_n^{k_n}\|_{L^2(\mathcal{M})} &= \left(\sum_{i=k_n+1}^{n} \exp\left(-\lambda_n^{(i)}\right) |\langle u, P^* \psi_n^{(i)} \rangle_{L^2(\mathcal{M})}|^2 \right)^{\frac{1}{2}} \\ &\leq \exp\left(-\frac{1}{2}\lambda_n^{(k_n)}\right) \lesssim \exp\left(-\frac{1}{2} k_n^{\frac{2}{d}}\right). \end{aligned} \tag{2.106}$$

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

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

From this, we get that

$$\begin{aligned} \|v_n^{k_n} - \tilde{v}_n^{k_n}\|_{L^2(\mathcal{M})} &\leq \left(\sum_{i=1}^{k_n} \left| \exp\left(-\lambda_n^{(i)}\right) - \exp\left(-\lambda^{(i)}\right) \right| \right)^{\frac{1}{2}} \\ &\lesssim \left(\exp\left(-2\lambda^{(i)}\right) \left(\frac{\varepsilon_n}{h_n} + h_n \sqrt{\lambda^{(i)}}\right)^2 \right)^{\frac{1}{2}} \\ &\lesssim \left(\frac{\varepsilon_n}{h_n} + h_n\right) \left(\sum_{i=1}^{k_n} \frac{\lambda^{(i)}}{\exp\left(2\lambda^{(i)}\right)} \right)^{\frac{1}{2}} \lesssim \left(\frac{\varepsilon_n}{h_n} + h_n\right). \end{aligned} \tag{2.107}$$

Next, we use (2.9.2) to deduce that

$$\begin{aligned}
\|\tilde{v}^{k_n} - \tilde{v}_n^{k_n}\|_{L^2(\mathcal{M})} &\lesssim \sum_{i=1}^{k_n} \exp\left(-\lambda^{(i)}\right) \|\psi^{(i)} - P^* \psi_n^{(i)}\|_{L^2(\mathcal{M})} \\
&\lesssim \sum_{i=1}^{k_n} \exp\left(-\lambda^{(i)}\right) i^{\frac{3}{2}} \sqrt{\frac{\varepsilon_n}{h_n} + h_n \sqrt{\lambda^{(i)}}} \\
&\lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n} \sum_{i=1}^{k_n} i^{\frac{3}{2}} \exp\left(-\lambda^{(i)}\right) (\lambda^{(i)})^{\frac{1}{4}} \\
&\lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n} \sum_{i=1}^{k_n} \frac{i^{\frac{3}{2}+\frac{1}{2d}}}{\exp\left(i^{2/d}\right)} \lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n}.
\end{aligned} \tag{2.108}$$

Finally, we use Cauchy-Schwartz to get that

$$\begin{aligned}
\|v^{k_n} - \tilde{v}^{k_n}\|_{L^2(\mathcal{M})} &= \left\| \sum_{i=1}^{k_n} \langle u, \psi^{(i)} - P^* \psi_n^{(i)} \rangle_{L^2(\mathcal{M})} \exp\left(-\lambda^{(i)}\right) \psi^{(i)} \right\|_{L^2(\mathcal{M})} \\
&\lesssim \sum_{i=1}^{k_n} \exp\left(-\lambda^{(i)}\right) \|\psi^{(i)} - P^* \psi_n^{(i)}\|_{L^2(\mathcal{M})} \lesssim \sqrt{\frac{\varepsilon_n}{h_n} + h_n}\ ,
\end{aligned} \tag{2.109}$$

where the last step follows by the same argument as for (2.108). Combining (2.105), (2.106), (2.107), (2.108), and (2.109), and taking $k_n \asymp n^{\frac{2d+1}{4\alpha}}$, we see that the error is dominated by

$$\|v - v_n\|_{op} \lesssim k_n^{\frac{1}{2} - \frac{2\alpha}{d}} + \sqrt{\frac{\varepsilon_n}{h_n} + h_n} \lesssim (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}}, \tag{2.110}$$

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

$$\|\mathcal{G} - P^* G_n P\|_{op} \lesssim k_n^{\frac{1}{2} - \frac{2\alpha}{d}} + \sqrt{\frac{\varepsilon_n}{h_n} + h_n} \lesssim (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}}. \tag{2.111}$$

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

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

where $\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}_n} \in \mathbb{R}^n_M$ is the vector with i^{th} 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 Cauchy-Schwarz to get the bound

$$\begin{aligned}\|\mathcal{O}v_n - \mathcal{O}^n v_n\|_2^2 &= \sum_{k=1}^{d_y}\langle v_n, \mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}}(x) - P^*\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}_n}(x)\rangle_{L^2(\mathcal{M})} \\ &\le \sum_{k=1}^{d_y}\|v_n\|^2_{L^2(\mathcal{M})}\|\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}} - P^*\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}_n}\|^2_{L^2(\mathcal{M})}.\end{aligned} \tag{2.112}$$

From (2.110), we have that

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

Hence, for n large enough it holds that

$$\|v_n\|_{L^2(\mathcal{M})} \leq 2\|\mathcal{G}\|_{op}\|u\|_{L^2(\mathcal{M})}. \tag{2.113}$$

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 remark that we can write

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

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

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

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 $\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}}\left(T_n(x)\right)$ and $\mathbb{1}_{B_\delta(x_k)\cap\mathcal{M}}\left(x\right)$ are not equal. Thus, we simply need to bound the measure of the sets

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

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 Theorem 2.2.5, 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 $\text{length}(\partial(B_\delta(x_k) \cap \mathcal{M})) = C_{\delta,k}$, gives us that

$$\gamma\left(\tilde{E}_k\right) \leq 2\varepsilon_n \text{length}(\partial(B_\delta(x_k) \cap \mathcal{M})) = 2C_{\delta,k}\varepsilon_n. \tag{2.114}$$

The constant $C_{\delta,k}$ depends on δ, $\mathcal{M}$, and k, but not n. This inequality, combined with (2.112) and (2.113) gives us that

$$\|\mathcal{O}v_n - \mathcal{O}^n v_n\|_2^2 \leq 2d_y \|\mathcal{G}\|_{op} C_{\delta,d}\varepsilon_n \lesssim d_y \frac{(\log n)^{c_d}}{n^{1/d}}. \tag{2.115}$$

Plugging (2.110) and (2.115) into (2.104) and taking the supremum of both sides over all $\|u\|_{L^2(\mathcal{M})} = 1$ yields

$$\|\mathcal{F} - FP\|_{op} \lesssim (\log n)^{\frac{c_d}{4}} n^{-\frac{1}{4d}} + \sqrt{d_y}(\log n)^{\frac{c_d}{2}} n^{-\frac{1}{2d}}. \tag{2.116}$$

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

2.10 Proof of Theorem 2.5.7

Proof of Theorem 2.5.7. We want to verify the assumptions of Theorem 2.5.5 for the setting of Eulerian data assimilation described in Example 2.5.6. 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 [119] that

$$\|m_0 - P^*Pm_0\|^2_{L^2(\mathcal{D})} \leq \frac{1}{n^{2\alpha}}\|m_0\|^2_{\mathcal{H}^\alpha(\mathcal{D})},$$

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 $\|u\|_{L^2(\mathcal{D})} = 1$,

$$\begin{aligned}\|\mathcal{C}_0 u - P^*C_0Pu\|^2_{L^2(\mathcal{D})} &= \sum_{i=n+1}^{\infty} (\lambda^{(i)})^{-2\alpha}\langle u, \psi^{(i)}\rangle^2_{L^2(\mathcal{D})} \\ &\leq \sum_{i=n+1}^{\infty} \left(\lambda^{(i)}\right)^{-2\alpha} \\ &\lesssim \sum_{i=n+1}^{\infty} i^{-2\alpha} \lesssim \frac{1}{n^{2\alpha-1}}.\end{aligned}$$

Hence, we have that

$$\|\mathcal{C}_0 - P^*C_0P\|_{op} \lesssim \frac{1}{n^{\alpha-\frac{1}{2}}}, \tag{2.117}$$

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 [121], 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

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

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

$$\left|\Phi(u) - \Phi(u')\right| \leq \frac{1}{2}L\|u - u'\|_{L^2(\mathcal{D})}\left(2\|y\|_{\Gamma^{-1}} + 2\|\mathcal{F}(u)\|_{\Gamma^{-1}} + Lt_0\|u - u'\|_{L^2(\mathcal{D})}\right). \quad (2.118)$$

From Lemma 4.2 in [119], we have that for the Galerkin approximation v^N in (2.91) to the true solution v in (2.87), 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

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

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

$$\left|\Phi(u') - \Phi^{(n)}(u')\right| \leq \frac{1}{2}\tilde{L}(\|u\|_{L^2(\mathcal{D})} + \delta)\psi(n) \tag{2.119}$$

$$\cdot \left(2\|y\|_{\Gamma^{-1}} + 2\|\mathcal{F}(u)\|_{\Gamma^{-1}} + \tilde{L}(\|u\|_{L^2(\mathcal{D})} + \delta)\psi(n)\right). \tag{2.120}$$

From (2.118) 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 (2.119) 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 $u' \in B_\delta(u) \in \mathcal{H}$,

$$\left|\Phi(u) - \Phi^{(n)}(u')\right| < \varepsilon, \tag{2.121}$$

proving continuous convergence of $\Phi^{(n)}$ to Φ as $n \to \infty$. We can then apply Theorem 2.5.5 and complete the proof. □

CHAPTER 3

LONG-TIME ACCURACY OF ENSEMBLE KALMAN FILTERS FOR CHAOTIC AND MACHINE-LEARNED DYNAMICAL SYSTEMS

3.1 Introduction

The filtering problem involves the sequential estimation of the state of a dynamical system from partial and noisy observations. For high-dimensional problems, particularly those arising in the geophysical sciences [211, 272, 197], the ensemble Kalman filter [147] is often the method of choice. The algorithm propagates an ensemble of particles under the dynamics and nudges the ensemble to new observations using empirical covariances in a Kalman style update. In the linear-Gaussian setting, the ensemble Kalman filter converges to the Kalman filter as the ensemble size grows [251, 224], thus fully characterizing the filtering distribution and yielding the optimal estimator of the state [297]. Furthermore, non-asymptotic theory shows that an ensemble size much smaller than the state dimension often suffices to approximate the Kalman filter [13]. For nonlinear dynamics, however, the ensemble Kalman filter does not reproduce the filtering distribution even in the large ensemble limit [234, 145, 233]. Despite this caveat, this paper rigorously justifies the use of ensemble Kalman filters in practical settings with nonlinear dynamics, partial observations, and an ensemble size smaller than the state dimension. We do so by providing conditions on the dynamics and the observations that guarantee the long-time accuracy of the ensemble Kalman filter as a state estimator. Our conditions cover a wide class of partially-observed chaotic dissipative dynamical systems, which includes the Navier-Stokes equations and Lorenz models. Dynamical systems that do not exhibit chaotic behavior are also covered by our theory.

Motivated by recent advancements in machine learning emulation of geophysical systems, this paper additionally studies ensemble Kalman filters that use surrogate models for

the dynamics. Examples of machine learning models for global weather forecasting include ClimaX [268], FourCastNet [278], GraphCast [225], Pangu [38], and FengWu [104], among others. Due to the chaotic nature of the dynamical systems of interest, it is unreasonable to expect these surrogate models to provide accurate state predictions over long time scales, motivating work on learning surrogates that preserve the time-invariant statistical properties of the dynamics [206, 276, 318]. However, in data assimilation, long-time filter accuracy may only need forecasts to be accurate over one assimilation cycle [5]. Consequently, cheap-to-evaluate machine learning surrogate models have the potential to enormously speed up data assimilation algorithms and allow for the generation of much larger ensembles than is possible with traditional numerical approximations of the true dynamics [221]. This paper establishes precise mathematical conditions for long-time accuracy of ensemble Kalman filters with surrogate models. While our theory and numerical examples are motivated by the use of machine-learned surrogate models for chaotic dynamical systems, our theory is also directly applicable to other types of surrogate models and misspecified dynamics.

3.1.1 Related work

Filter accuracy A significant body of work on the accuracy of filtering algorithms builds on the theory of synchronization in dynamical systems [279]. The seminal paper [190] establishes the convergence of discrete-time data assimilation methods for the Lorenz-63 and Navier-Stokes equations with partial but noiseless observations. These results were extended to more realistic observation models and noisy observations in [20, 151, 36], and similar ideas are used to study cycled data assimilation schemes in [260]. Long-time accuracy for the 3DVar algorithm is shown in [61] for the Navier-Stokes equations, [230] for the Lorenz-63 system, and [232] for the Lorenz-96 system. The accuracy of approximate Gaussian filters for the Navier-Stokes equations with model errors was studied numerically in [58], and conditions guaranteeing the long-time accuracy of 3DVar with surrogate models were given in

[5]. A rigorous connection between the determining modes property of the Navier-Stokes equations and the synchronization of filters was made in [91], and the relationship between the nudging filter and synchronization filter for the Navier-Stokes equations was studied in [92].

Ensemble Kalman filters Much of the work on the theoretical foundations of ensemble Kalman filters has focused on the linear-Gaussian setting. In this setting, it is known that the ensemble Kalman filter converges to the Kalman filter in the large ensemble asymptotic [251, 224]. In the nonlinear setting, a similar line of work assumes the *mean-field* perspective [73, 195, 135, 134] and analyzes the large ensemble limits of the ensemble Kalman filter. Recent results using this approach have shown that the mean-field ensemble Kalman filter is near the filtering distribution in the near-Gaussian setting [95] and in the near-linear setting [72]. These results only hold over a finite-time horizon and do not guarantee accurate state estimation. A complementary line of work provides a non-asymptotic analysis of ensemble Kalman updates, explaining how the ensemble Kalman filter can well approximate the mean-field updates with an ensemble size much smaller than the dimension of the state space [13, 300, 10]. This perspective helps explain why localization techniques are so successful within ensemble Kalman filters and extends notions of localization and sparse covariance estimation to the function space setting [12, 11]. However, none of these non-asymptotic analyses hold over an infinite-time horizon, and they do not consider state estimation accuracy. A number of previous works have considered the state estimation accuracy of the ensemble Kalman filter, largely using the theory of synchronization in dynamical systems [279]. The work [214] proves stability and accuracy of the ensemble Kalman filter with covariance inflation, but only in the setting with complete observations of the system. They provide numerical evidence of filter accuracy with partial observations, but do not extend the theorems to the partially-observed setting. Recently, [317] proves accuracy of the ensemble Kalman filter with multiplicative covariance inflation, but again requires full observations and the results are

restricted to finite-dimensional dynamical systems. Notable works in the partially-observed setting include [328, 173, 45]. In [328], the authors prove long-time stability of the ensemble Kalman filter with an adaptive variance inflation scheme, but do not prove filter accuracy. For hyperbolic dynamical systems, [173] proves accuracy of the ensemble Kalman filter with partial observations using a shadowing argument. The work most similar to ours is [45], which proves accuracy of a continuous-time variant of the ensemble Kalman filter for the Navier-Stokes equations in the partially-observed setting. Their algorithm includes variance inflation and a somewhat atypical localization procedure where the unobserved components of the forecast covariance are projected to be zero. In contrast, our results hold in the discrete-time setting and do not require this localization procedure.

Machine learning surrogate models in data assimilation Machine learning methods have made far-reaching impacts on data assimilation; see [49, 110, 22] and [105, Chapter 10] for recent reviews. The use of machine learning surrogates has the potential to rapidly accelerate data assimilation algorithms, due to both the orders of magnitude speedup in simulating the dynamics and the possibility of employing automatic differentiation for the computation of adjoint models [189, 254, 348], although these adjoint models may display unphysical behavior [324]. Within ensemble data assimilation schemes, surrogate dynamics models make it possible to generate much larger ensembles than traditional expensive numerical solvers [102, 176] and can in turn lead to improved estimates of the background covariance [103]. In the context of global weather prediction, case studies in [5] and [220] provide compelling numerical evidence that entirely data-driven surrogates can be used successfully within 3DVar and the ensemble Kalman filter, respectively. For the 3DVar algorithm, [5] provides conditions on the surrogate model that ensure accurate state estimation over long time horizons. An alternative approach is to use machine learning models to learn and correct the model error incurred by a traditional physics-based model [106, 149]. Other approaches learn the state trajectory and model/model error together [50, 57, 109, 175, 185, 267]. Another line

of work uses machine learning models not as surrogates for the dynamics, but to learn other components of the data assimilation process such as the forecast error covariance, the analysis update, or the entire data assimilation procedure [17, 261, 280, 334].

3.1.2 Main contributions and outine

- Section 3.2 formalizes the problem setting and states our main results. Our first main result, Theorem 3.2.2, proves the long-time accuracy of the square-root ensemble Kalman filter with appropriate variance inflation given that the observations are sufficiently informative. The second main result, Theorem 3.2.8, proves the long-time accuracy of the square-root ensemble Kalman filter with a surrogate model, provided the model is sufficiently accurate in the unobserved part of the system.

- Section 3.3 contains the proof of Theorem 3.2.2. We first establish accuracy for a mean-field version of the ensemble Kalman filter in Theorem 3.3.2. Then, we show that the ensemble means from the square-root ensemble Kalman filter remain close to those output by the idealized mean-field version in Theorem 3.3.5.

- Section 3.4 contains the proof of Theorem 3.2.8. As in Section 3.3, we will analyze an idealized mean-field algorithm as an intermediate step in the analysis.

- Section 3.5 contains numerical experiments illustrating our theoretical results for the Lorenz-96 system. In particular, we show that long-time filter accuracy is achievable with machine-learned surrogate models that only provide accurate short-term forecasts.

- Section 3.6 concludes with a summary of our work and presents several open directions for future research.

3.2 Main results

3.2.1 Long-time accuracy of ensemble Kalman filters

In this subsection, we show long-time accuracy for a standard implementation of the ensemble Kalman filter. Our setting covers a wide range of partially-observed dissipative chaotic dynamical systems often used to benchmark filtering algorithms.

Set-up

We study the filtering problem of sequentially estimating a time-evolving hidden signal as new observations become available. Suppose that the signal process $\{u_j\}_{j=1}^{\infty} \subset \mathcal{H}$ lies on a separable Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle, \|\cdot\|)$ and its evolution is governed by a discrete-time dynamical system given by $u_0 \in \mathcal{H}$ and

$$u_j = \Psi(u_{j-1}), \qquad j = 1, 2, \dots \tag{3.1}$$

The map $\Psi : \mathcal{H} \to \mathcal{H}$ is assumed to be known, but the initial condition $u_0 \in \mathcal{H}$ —and hence the signal process— are assumed to be unknown. To estimate the signal, we are given observations $\{y_j\}_{j=1}^{\infty} \subset \mathbb{R}^k$ that comprise finite-dimensional noisy measurements of the state

$$y_j = Hu_j + \varepsilon\eta_j, \qquad j = 1, 2, \dots$$

Here, $H : \mathcal{H} \to \mathbb{R}^k$ is assumed to satisfy $HH^* = I_k$ and we define $P := H^*H : \mathcal{H} \to \mathcal{H}$, which is an orthogonal projection onto a k-dimensional subspace of $\mathcal{H}$. We are interested in the case where $k \ll \dim(\mathcal{H}) \in \mathbb{N} \cup \{\infty\}$ and assume that the observation errors $\{\eta_j\}_{j=1}^{\infty}$ are independent, with $\mathbb{E}[\eta_j] = 0$ and $\mathrm{Cov}[\eta_j] = R \in \mathbb{R}^{k\times k}$. The parameter $\varepsilon > 0$ controls the size of the observation noise, which is often small in applications. Filtering concerns sequentially estimating the state u_j given all the observations $\{y_i\}_{i=1}^{j}$ available at time j.

The ability of filtering algorithms to accurately estimate the signal over a long time window necessarily hinges on control-theoretic conditions on the signal and observation processes. We will adopt the following assumption, also used in [299], on the dynamics and the observations. As discussed in Subsection 3.2.1 below, this assumption is satisfied by important test filtering problems, including partially-observed Lorenz models and Navier-Stokes equations. For a constant $\beta \geq 0$, we define $V : \mathcal{H} \to [0, \infty)$ by

$$V(u) := \Big(\|u\|^2 + \beta \|Pu\|^2 \Big)^{1/2}.$$

The norm V will act as a Lyapunov function in the theory.

Assumption 3.2.1. The following properties hold:

(Absorbing ball property) There exist a constant $r > 0$ such that the ball

$$\mathcal{B} := \{u \in \mathcal{H} : \|u\| \leq r\} \tag{3.2}$$

is absorbing and forward invariant for the dynamical system (3.1).

(Lipschitz continuity on $\mathcal{B}$) There exists $L > 0$ such that, for any $u, v \in \mathcal{B}$,

$$V^2\big(\Psi(u) - \Psi(v)\big) \leq L V^2(u - v). \tag{3.3}$$

(Squeezing property) There exists $\alpha \in (0, 1)$ such that, for any $u, v \in \mathcal{B}$,

$$V^2\Big((I - P)\big(\Psi(u) - \Psi(v)\big)\Big) \leq \alpha V^2 \left(u - v\right). \tag{3.4}$$

The absorbing ball property and Lipschitz continuity on the abosorbing ball are naturally satisfied by dissipative dynamical systems where there is an energy-loss mechanism. In such systems, the bounded invariant sets of the dynamics contain all the information about asymptotic behavior [314, Section 1.8]; therefore, these sets play a key role in the study

of long-time filter accuracy. The squeezing property can be interpreted as a detectability condition that ensures that the dynamics contract in the unobserved part of the system [299].

Square-root ensemble Kalman filter

Ensemble Kalman filters sequentially estimate the state using an ensemble of interacting particles. At time $j = 0$, we generate an *initial ensemble* $\mathsf{U}_0 = \{u_0^{(n)}\}_{n=1}^N$ of N particles; for concreteness, we assume that $u_0^{(n)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(m_0, C_0)$, $1 \le n \le N$, for given mean $m_0 \in \mathcal{H}$ and covariance operator C_0. Then, at each time $j \ge 1$, we estimate the hidden state u_j by the empirical mean $\widehat{m}_j$ of an ensemble $\mathsf{U}_j = \{u_j^{(n)}\}_{n=1}^N$. Particles are updated from time $j-1$ to time j in two steps, referred to as *prediction* and *analysis*, which utilize respectively the dynamical system (3.1) and the observation y_j acquired at time j. Thus, each update can be schematically represented by:

$$\mathsf{U}_{j-1} = \left\{u_{j-1}^{(n)}\right\}_{n=1}^N \xrightarrow[\text{Dynamical System}]{\text{Prediction}} \mathsf{V}_j = \left\{v_j^{(n)}\right\}_{n=1}^N \xrightarrow[\text{Observation}]{\text{Analysis}} \mathsf{U}_j = \left\{u_j^{(n)}\right\}_{n=1}^N . \tag{3.5}$$

In the prediction step, we use the dynamics governing the evolution of the state and set

$$v_j^{(n)} = \Psi\big(P_{\mathcal{B}_V} u_{j-1}^{(n)}\big) + \xi_j^{(n)}, \qquad \xi_j^{(n)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, Q), \qquad 1 \le n \le N. \tag{3.6}$$

Here, we follow [299] and project each particle $u_{j-1}^{(n)}$ into the absorbing ball $\mathcal{B}$ in Assumption 3.2.1; specifically, $P_{\mathcal{B}_V}$ denotes the orthogonal projection onto $\mathcal{B}$ in the V-norm. To promote spread of the prediction ensemble and avoid downward bias of estimates of the covariance of the state [154], we introduce additional randomness $\xi_j^{(n)}$ in the propagation of each particle $u_{j-1}^{(n)}$. This is a standard procedure in data assimilation known as *covariance inflation*; see for instance [148, Chapter 10], [212, Section 4], and the discussion in [22]. We take the covariance inflation operators to be of the form $Q = \gamma P$ for some constant $\gamma > 0$. Since

P is an orthogonal projection onto a k-dimensional subspace, Q is trace-class and positive semi-definite, and hence a valid covariance operator in $\mathcal{H}$. It will be relevant to note that for $\xi \sim \mathcal{N}(0, Q)$, it holds that $\mathrm{Cov}[H\xi] = HQH^* = \gamma HPH^* = \gamma I_{k\times k}$.

In the analysis step, we update the prediction ensemble $\mathsf{V}_j = \{v_j^{(n)}\}_{n=1}^N$ into an analysis ensemble $\mathsf{U}_j = \{u_j^{(n)}\}_{n=1}^N$ by imposing that the empirical mean and covariances are consistent with Kalman-type formulae for linear-Gaussian inverse problems. To that end, we define the *Kalman gain* associated to covariance operator Σ by

$$\mathscr{K}(\Sigma) := \Sigma H^*(H\Sigma H^* + \varepsilon^2 R)^{-1}. \tag{3.7}$$

Then, denoting by $(\widehat{\mu}_j, \widehat{\Sigma}_j)$ the empirical mean and covariance of V_j and by $(\widehat{m}_j, \widehat{C}_j)$ the empirical mean and covariance of U_j, we transform the particles so that

$$\begin{aligned} \widehat{m}_j &= \widehat{\mu}_j + \mathscr{K}(\widehat{\Sigma}_j)\left(y_{j+1} - H\widehat{\mu}_j\right), \\ \widehat{C}_j &= \left(I - \mathscr{K}(\widehat{\Sigma}_j)H\right)\widehat{\Sigma}_j. \end{aligned} \tag{3.8}$$

As described in [326] and reviewed in [13], many deterministic maps can be used to transform the prediction ensemble V_j into an analysis ensemble U_j that satisfies the moment constraints in (3.8), leading to different square-root ensemble Kalman filters. Notable examples include the Ensemble Transform Kalman Filter [44] and the Ensemble Adjustment Kalman Filter [14]. Our theory is agnostic to the choice of transformation, and holds for the general formulation of square-root ensemble Kalman filters summarized in Algorithm 3.

Algorithm 3 Square-Root Ensemble Kalman Filter

1: **Initialization**: Draw

$$u_0^{(n)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(m_0, C_0), \qquad 1 \le n \le N.$$

2: For $j = 1, 2, \dots$ do the following:

3: **Prediction**: Compute

$$v_j^{(n)} = \Psi\big(P_{\mathcal{B}_V} u_{j-1}^{(n)}\big) + \xi_j^{(n)}, \qquad \xi_j^{(n)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, Q), \qquad 1 \le n \le N,$$
$$\widehat{\mu}_j = \frac{1}{N}\sum_{n=1}^{N} v_j^{(n)}, \qquad \widehat{\Sigma}_j = \frac{1}{N}\sum_{n=1}^{N} \big(v_j^{(n)} - \widehat{\mu}_j\big) \otimes \big(v_j^{(n)} - \widehat{\mu}_j\big).$$

4: **Analysis**: Transform $\{v_j^{(n)}\}_{n=1}^N \mapsto \{u_j^{(n)}\}_{n=1}^N$ such that

$$\widehat{m}_j := \frac{1}{N}\sum_{n=1}^{N} u_j^{(n)} = \widehat{\mu}_j + \mathscr{K}(\widehat{\Sigma}_j)\left(y_{j+1} - H\widehat{\mu}_j\right),$$
$$\widehat{C}_j := \frac{1}{N-1}\sum_{n=1}^{N} \big(u_j^{(n)} - \widehat{m}_j\big) \otimes \big(u_j^{(n)} - \widehat{m}_j\big) = \big(I - \mathscr{K}(\widehat{\Sigma}_j)H\big)\widehat{\Sigma}_j.$$

5: **Output**: Means and covariances $\{\widehat{m}_j, \widehat{C}_j\}_{j=1}^{\infty}$.

We are ready to state our first main result, which establishes long-time accuracy of the ensemble Kalman filter. We show that both observed and unobserved parts of the state can be estimated at the level ε of the observation noise. The proof will be deferred to Section 3.3.

Theorem 3.2.2. *Suppose that Assumption 3.2.1 holds and that $N \ge 6k$. Then, if the inflation parameter $\gamma > 0$ is sufficiently large, there is a constant C independent of ε such*

that

$$\lim_{j\to\infty} \sup \mathbb{E}\|\widehat{m}_j - u_j\| \leq \mathsf{C}\varepsilon.$$

Remark 3.2.3. Our theory requires the ensemble size N to scale with the dimension k of the observations, but *not* with the state dimension. The assumption that $N \geq 6k$ is a requirement of the results in [262] that we use to lower bound the smallest eigenvalue of $H\widehat{\Sigma}_j H^*$. Instead of performing variance inflation by randomly perturbing each particle in (3.6), one could consider deterministic variance inflation by noiselessly propagating each particle through the dynamics, i.e. setting $v_j^{(n)} = \Psi\big(P_{\mathcal{B}_V} u_{j-1}^{(n)}\big)$, and then deterministically and additively inflating the empirical covariance. Then, our proof techniques suggest that $N \geq k$ suffices. In practical applications, the dimension of the observations is often much larger than the size of the ensemble; for instance $k \sim 10^6$ and $N \sim 10^2$ in operational weather forecasting. As such, there remains a signficant gap between our theory and the observed success of the ensemble Kalman filter in practice. However, previous theoretical guarantees on the accuracy of the ensemble Kalman filter have either considered the large ensemble limit [251, 224, 72], assumed N to be larger than the state dimension [131, 317], or assumed the system to be fully observed [214]. Our result is state of the art in that we do not require such stringent assumptions on the ensemble size or the observation model. Removing the undesirable dependence on the observation dimension remains an area for future work.

Application to partially-observed dissipative chaotic dynamical systems

Here we verify Assumption 3.2.1 for partially-observed signals arising from dissipative systems with quadratic, energy-conserving nonlinearity of the form

$$\frac{du}{dt} + \mathsf{A}u + \mathsf{B}(u,u) = \mathsf{F}, \quad u(0) = u_0 \in \mathcal{H}. \tag{3.9}$$

We denote by $\Delta t > 0$ the time between observations and let Ψ denote the Δt-flow of (3.9); that is, $\Psi(u_0)$ is the value at time Δt of the solution to (3.9) with initial condition $u_0 \in \mathcal{H}$. With this set-up, the state u_j in (3.1) represents the solution to (3.9) at time $t_j := j\Delta t$ with initial condition $u_0 \in \mathcal{H}$. We next verify Assumption 3.2.1 for partially-observed Lorenz-63, Lorenz-96, and Navier-Stokes equations, all of which can be written in the form (3.9). These examples are not new [190, 292, 232, 299], but we review them here to demonstrate the wide applicability of our novel long-time accuracy theory for ensemble Kalman filters.

Example 3.2.4. The Lorenz-63 model[1] with standard parameter values $(10, 8/3, 28)$ can be written in the form (3.9) with $\mathcal{H} := \mathbb{R}^3$,

$$\mathsf{A} := \begin{bmatrix} 10 & -10 & 0 \\ 10 & 1 & 0 \\ 0 & 0 & 8/3 \end{bmatrix}, \qquad \mathsf{B}(u, \tilde{u}) := \begin{bmatrix} 0 \\ (u_1\tilde{u}_3 + u_3\tilde{u}_1)/2 \\ -(u_1\tilde{u}_2 + u_2\tilde{u}_1)/2 \end{bmatrix}, \qquad \mathsf{F} = \begin{bmatrix} 0 \\ 0 \\ -304/3 \end{bmatrix}. \quad (3.10)$$

The Lorenz-63 model satisfies the absorbing ball property, see e.g. [231, Example 2.6]. Moreover, the squeezing property is satisfied observing only the first component of the state: for sufficiently small Δt, the Δt-flow $\Psi : \mathbb{R}^3 \to \mathbb{R}^3$ of the Lorenz equation and the observation map $H = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} : \mathbb{R}^3 \to \mathbb{R}$ satisfy (3.4) with $\beta = 1$. We refer to [299] for a proof.

Example 3.2.5. The Lorenz-96 model can be written in the form (3.9) with $\mathcal{H} := \mathbb{R}^d$ where

1. In these examples, subscripts denote coordinates and coefficient indices rather than time.

we assume $d \in 3\mathbb{N}$,

$$\mathsf{A} := I_{d\times d}, \quad \mathsf{B}(u,\tilde{u}) := -\frac{1}{2}\begin{bmatrix} \tilde{u}_2 u_d + u_2\tilde{u}_d - \tilde{u}_d u_{d-1} - u_d\tilde{u}_{d-1} \\ \vdots \\ \tilde{u}_{i-1}u_{i+1} + u_{i-1}\tilde{u}_{i+1} - \tilde{u}_{i-2}u_{i-1} - u_{i-2}\tilde{u}_{i-1} \\ \vdots \\ \tilde{u}_{d-1}u_1 + u_{d-1}\tilde{u}_1 - \tilde{u}_{d-2}u_{d-1} - u_{d-2}\tilde{u}_{d-1} \end{bmatrix}, \quad \mathsf{F} = \begin{bmatrix} 8 \\ \vdots \\ 8 \end{bmatrix}. \tag{3.11}$$

The Lorenz-96 model satisfies the absorbing ball property, see e.g. [231, Example 2.7]. Moreover, it satisfies the squeezing property observing only two out of every three components of the state. Let $H \in \mathbb{R}^{\frac{2}{3}d\times d}$ denote the identity matrix with every third row removed. Then, for sufficiently small Δt, the Δt-flow $\Psi : \mathbb{R}^d \to \mathbb{R}^d$ of the Lorenz-96 model and the observation map H satisfy (3.4) with $\beta = 1$. We refer to [299] for a proof.

Example 3.2.6. We consider the incompressible Navier-Stokes equations on the 2-dimensional torus, $\mathbb{T}^2 = [0,1) \times [0,1)$. Let $\mathcal{U}_0$ be the space of zero-mean, divergence-free, vector-valued polynomials u from $\mathbb{T}^2$ to $\mathbb{R}^2$, and let $\mathcal{U}$ be the closure of $\mathcal{U}_0$ with respect to the L^2 norm. We define $P_{\mathcal{U}} : \left(L^2(\mathbb{T}^2)\right)^2 \to \mathcal{U}$ to be the Leray-Helmholtz orthogonal projector [321]. Functions in $\mathcal{U}$ can be written as a Fourier series

$$u = \sum_{\ell\in\mathcal{L}} u_\ell e^{i\ell x}, \qquad \mathcal{L} = \Big\{2\pi(n_1,n_2) : n_i \in \mathbb{Z}, (n_1,n_2) \neq (0,0)\Big\}. \tag{3.12}$$

We define $\mathcal{H}$ to be the closure of $\mathcal{U}_0$ with respect to norm defined by

$$\|u\|^2 := \sum_{\ell\in\mathcal{L}} |\ell|^2 |u_\ell|^2,$$

which is equivalent to the Sobolev H^1 norm. The Navier-Stokes equations can be written in

the form (3.9) with

$$\mathsf{A} := -\nu P_{\mathcal{U}} \Delta, \qquad \mathsf{B}(u, \tilde{u}) = P_{\mathcal{U}}(u \cdot \nabla \tilde{u}), \qquad \mathsf{F} = f \in \mathcal{U}. \tag{3.13}$$

Given an initial condition in $\mathcal{H}$, the Navier-Stokes equations have a unique strong solution for any time $\Delta t > 0$ that depends continuously on the initial data in the $\mathcal{H}$ norm [292, Theorem 9.5]. One can show that there exists $\theta = \theta(\nu) > 0$ such that for every $u \in \mathcal{H}$, $\|\mathsf{A}u\|^2 \geq \theta \|u\|^2$ [292, Chapter 12]. Then, the absorbing ball property holds with

$$\mathcal{B} = \left\{ u \in \mathcal{H} : \|u\| \leq r := \sqrt{2} \frac{\|f\|}{\theta} \right\}.$$

We define $\mathcal{L}_\lambda = \{\ell \in \mathcal{L} : |\ell|^2 \leq \lambda\}$ and set $k = \mathrm{card}(\mathcal{L}_\lambda)$. Consider the observation operator $H_\lambda u := (|\ell| u_\ell)_{\ell \in \mathcal{L}_\lambda}$ with corresponding operator $P_\lambda : \mathcal{H} \to \mathcal{H}$ given by

$$P_\lambda u = \sum_{\ell \in \mathcal{L}_\lambda} u_\ell e^{i\ell x},$$

which is an orthogonal projection onto a k-dimensional subspace of $\mathcal{H}$. There exist constants $\alpha = \alpha(r) \in (0, 1)$ and $\lambda_* = \lambda_*(r) > 0$ such that for all $\lambda > \lambda_*$, there exists $\Delta t_* = \Delta t_*(r, \lambda)$ such that, for all $u, v \in \mathcal{B}$ and all $\Delta t \leq \Delta t_*$,

$$\big\|(I - P)\Psi(u) - \Psi(v)\big\| \leq \alpha \big\|u - v\big\|.$$

Hence, the squeezing property is satisfied with $\beta = 0$ by the observation map $H = H_\lambda$ for any $\lambda \geq \lambda_*$. We refer to [299] for a proof of the result as stated here, and to [292, Chapter 14] for additional background on the squeezing property for the Navier-Stokes equations.

3.2.2 Long-time accuracy with surrogate models

Next, we analyze ensemble Kalman filters that utilize a surrogate model for the dynamics.

Set-up

We assume that the signal and observation processes are defined as in Subsection 3.2.1, with dynamics map Ψ and observation matrix H satisfying Assumption 3.2.1. We will study ensemble Kalman filters that utilize a surrogate dynamics model Ψ^s satisfying:

Assumption 3.2.7. The following properties hold:
(Bounded approximation error on $\mathcal{B}$) There exists $\kappa > 0$ such that, for any $u \in \mathcal{B}$,

$$\|\Psi^s(u) - \Psi(u)\| \leq \kappa.$$

(Lipschitz continuity on $\mathcal{B}$) There exists $L_s > 0$ such that, for any $u, v \in \mathcal{B}$,

$$V^2\big(\Psi^s(u) - \Psi^s(v)\big) \leq L_s V^2\big(u - v\big).$$

(Accuracy in unobserved part) There exists small $\delta > 0$ such that

$$\big\|(I - P)\big(\Psi^s(u) - \Psi(u)\big)\big\| \leq \delta.$$

Assumption 3.2.7 controls the error of the surrogate model in the unobserved part of the system. Notice that the surrogate model is only assumed to be accurate over a single discrete time-step, without requiring stability over a long time window. Consequently, our theory will validate the use of ensemble data assimilation with machine-learned surrogate models that provide accurate *short-term* forecasts in the unobserved part of the system.

Square-root ensemble Kalman filter with a surrogate model

We will study the following algorithm, obtained by replacing in Algorithm 3 the true dynamics map Ψ with the surrogate model Ψ^s:

Algorithm 4 Square-Root Ensemble Kalman Filter with a Surrogate Model

1: **Initialization**: Draw

$$u_0^{(n),s} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(m_0, C_0), \qquad 1 \le n \le N.$$

2: For $j = 1, 2, \dots$ do the following:

3: **Prediction**: Compute

$$v_j^{(n),s} = \Psi^s\big(P_{\mathcal{B}_V} u_{j-1}^{(n),s}\big) + \xi_j^{(n)}, \qquad \xi_j^{(n)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, Q), \qquad 1 \le n \le N,$$
$$\widehat{\mu}_j^s = \frac{1}{N}\sum_{n=1}^N v_j^{(n),s}, \qquad \widehat{\Sigma}_j^s = \frac{1}{N}\sum_{n=1}^N \big(v_j^{(n),s} - \widehat{\mu}_j^s\big) \otimes \big(v_j^{(n),s} - \widehat{\mu}_j^s\big).$$

4: **Analysis**: Transform $\{v_j^{(n),s}\}_{n=1}^N \mapsto \{u_j^{(n),s}\}_{n=1}^N$ such that

$$\widehat{m}_j^s := \frac{1}{N}\sum_{n=1}^N u_j^{(n),s} = \widehat{\mu}_j^s + \mathscr{K}(\widehat{\Sigma}_j^s)\left(y_j - H\widehat{\mu}_j^s\right),$$
$$\widehat{C}_j^s := \frac{1}{N-1}\sum_{n=1}^N \big(u_j^{(n),s} - \widehat{m}_j^s\big) \otimes \big(u_j^{(n),s} - \widehat{m}_j^s\big) = \big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\widehat{\Sigma}_j^s.$$

5: **Output**: Means and covariances $\{\widehat{m}_j^s, \widehat{C}_j^s\}_{j=1}^\infty$.

We now state our second main result, which establishes long-time accuracy of ensemble Kalman filters with a surrogate model. We recall that ε represents the level of observation noise and δ the accuracy of the surrogate model. The proof can be found in Section 3.4.

Theorem 3.2.8. *Suppose that Assumptions 3.2.1 and 3.2.7 hold and that $N \ge 6k$. Then, if the inflation parameter $\gamma > 0$ is sufficiently large, there is a constant C^s independent of ε*

and δ such that

$$\limsup_{j\to\infty} \mathbb{E}\|\widehat{m}_j^s - u_j\| \leq \mathsf{C}^s(\varepsilon + \delta).$$

Remark 3.2.9. In both of our main results, the inflation parameter must be taken to be increasing with L, $\frac{1}{1-\alpha^{1/2}}$, β, $\mathrm{Tr}(R)$, N and ε, as well as κ, L_s, and δ in the latter theorem. We conjecture that the dependence on N is an artifact of the proof techniques in Lemma 3.3.6 and could possibly be removed, whereas removing the dependence on the other quantities would require a signicantly different approach than that considered here. Informally, we require the inflation parameter to be large enough such that the smallest eigenvalue of the predictive covariance in the observation space is large enough to sufficiently nudge the ensemble towards the observations. An interested reader can track the exact required dependence on these quantities through the proofs, although we do not claim the dependence to be sharp, nor do we intend to present the theorem as a guideline for choosing the variance inflation parameter in practical applications.

3.3 Filter accuracy via mean-field theory

In this section, we prove the first main result, Theorem 3.2.2. Our analysis will leverage new accuracy bounds for an idealized, non-implementable algorithm that uses population moments, rather than moments computed through an ensemble. Specifically, we consider the following Gaussian projected filter, which arises from a mean-field limit ($N \to \infty$) of the ensemble Kalman filter in Algorithm 3. Notice that this idealized algorithm does not propagate an ensemble of particles, and covariance inflation is enforced by adding a positive definite operator $Q = \gamma P$ to the prediction covariance.

Algorithm 5 Gaussian Projected Filter

1: **Initialization**: m_0, C_0.

2: For $j = 1, 2 \dots$ do the following:

3: **Prediction**:

$$\mu_j = \mathbb{E}_{u\sim\mathcal{N}(m_{j-1},C_{j-1})}\big[\Psi(P_{\mathcal{B}_V}u)\big],$$
$$\Sigma_j = \mathbb{E}_{u\sim\mathcal{N}(m_{j-1},C_{j-1})}\big[\big(\Psi(P_{\mathcal{B}_V}u) - \mu_j\big) \otimes \big(\Psi(P_{\mathcal{B}_V}u) - \mu_j\big)\big].$$

4: **Analysis**:

$$m_j = \mu_j + \mathscr{K}(\Sigma_j + Q)\left(y_j - H\mu_j\right),$$
$$C_j = \big(I - \mathscr{K}(\Sigma_j + Q)H\big)(\Sigma_j + Q).$$

5: **Output**: Means and covariances $\{m_j, C_j\}_{j=1}^{\infty}$.

Our first lemma shows that the squeezing property implies that the trace of the analysis covariance operators generated by Algorithms 3 and 5 are controlled at the noise level ε in the long-time asymptotic.

Lemma 3.3.1. *Under Assumption 3.2.1, the covariances output by Algorithm 5 satisfy*

$$\mathrm{Tr}\left(C_j\right) \leq \alpha^j \mathrm{Tr}\left(C_0\right) + (1+\alpha\beta)\frac{1-\alpha^j}{1-\alpha}\varepsilon^2 \mathrm{Tr}(R), \tag{3.14}$$

and those output by Algorithm 3 satisfy

$$\mathrm{Tr}\big(\widehat{C}_j\big) \leq \alpha^j \mathrm{Tr}\big(\widehat{C}_0\big) + (1+\alpha\beta)\frac{1-\alpha^j}{1-\alpha}\varepsilon^2 \mathrm{Tr}(R). \tag{3.15}$$

Proof. Since P is an orthogonal projection and $Q = \gamma P$, we have

$$\begin{aligned} \mathrm{Tr}(C_j) &= \mathrm{Tr}(PC_jP^*) + \mathrm{Tr}\big((I-P)C_j(I-P)^*\big) \\ &\leq \mathrm{Tr}(PC_jP^*) + \mathrm{Tr}\big((I-P)\Sigma_j(I-P)^*\big). \end{aligned} \tag{3.16}$$

We next bound in turn the two terms in the right-hand side.

For the first term, let $\{\varphi_i\}_{i=1}^k$ be an orthonormal basis for the range of P, which we can extend to a basis for $\mathcal{H}$ via Gram-Schmidt by the separability of $\mathcal{H}$. We then compute that

$$\begin{aligned} \mathrm{Tr}\left(PC_jP^*\right) &= \sum_{i=1}^{\infty} \langle \varphi_i, PC_jP^*\varphi_i\rangle = \sum_{i=1}^{k} \langle \varphi_i, PC_jP^*\varphi_i\rangle \\ &= \sum_{i=1}^{k} \langle H\varphi_i, HC_jH^*H\varphi_i\rangle_{\mathbb{R}^k} = \mathrm{Tr}\left(HC_jH^*\right) \leq \varepsilon^2 \mathrm{Tr}\left(R\right), \end{aligned} \tag{3.17}$$

where we have used that $\{H\varphi_i\}_{i=1}^k$ is an orthonormal basis for $\mathbb{R}^k$ and that $\epsilon^2 R - HC_jH^*$ is positive semi-definite in the last inequality.

For the second term in the right-hand side of (3.16), note that

$$\begin{aligned} \mathrm{Tr}\big((I-P)\Sigma_j(I-P)^*\big) &= \mathbb{E}_{\mathcal{N}(m_{j-1},C_{j-1})}\Big[\big\|(I-P)(\Psi(P_{\mathcal{B}_V}u) - \mu_j)\big\|^2\Big] \\ &\leq \mathbb{E}_{\mathcal{N}(m_{j-1},C_{j-1})}\Big[\big\|(I-P)\big(\Psi(P_{\mathcal{B}_V}u) - \Psi(P_{\mathcal{B}_V}m_{j-1})\big)\big\|^2\Big] \\ &\leq \mathbb{E}_{\mathcal{N}(m_{j-1},C_{j-1})}\Big[\alpha\left(\big\|P_{\mathcal{B}_V}u - P_{\mathcal{B}_V}m_{j-1}\big\|^2 + \beta\big\|P(P_{\mathcal{B}_V}u - P_{\mathcal{B}_V}m_{j-1})\big\|^2\right)\Big] \\ &\leq \mathbb{E}_{\mathcal{N}(m_{j-1},C_{j-1})}\Big[\alpha\left(\|u - m_{j-1}\|^2 + \beta\|P(u-m_{j-1})\|^2\right)\Big] \\ &= \alpha\mathrm{Tr}\left(C_{j-1}\right) + \alpha\beta\mathrm{Tr}\left(PC_{j-1}P^*\right). \end{aligned} \tag{3.18}$$

Combining (3.16), (3.17), and (3.18), we deduce that

$$\begin{aligned}\mathrm{Tr}(C_j) &\leq \varepsilon^2 \mathrm{Tr}(R) + \alpha \mathrm{Tr}(C_{j-1}) + \alpha\beta \mathrm{Tr}(PC_{j-1}P^*) \\ &\leq (1+\alpha\beta)\varepsilon^2 \mathrm{Tr}(R) + \alpha \mathrm{Tr}(C_{j-1}).\end{aligned}$$

Applying the discrete Gronwall lemma [297, Theorem 1.19] yields (3.14).

To prove (3.15), we proceed similarly. Since $Q = \gamma P$, it holds that $(I-P)\xi_j^{(n)} = 0$ almost surely, from which we have

$$\begin{aligned}\mathrm{Tr}\big((I-P)\widehat{\Sigma}_j(I-P)^*\big) &= \frac{1}{N-1}\sum_{n=1}^{N}\Big\|(I-P)(v_j^{(n)} - \widehat{\mu}_j)\Big\|^2 \\ &= \frac{1}{N-1}\sum_{n=1}^{N}\Big\|(I-P)\big(\Psi(P_{\mathcal{B}_V}u_{j-1}^{(n)}) - \frac{1}{N}\sum_{n=1}^{N}\Psi(P_{\mathcal{B}_V}u_{j-1}^{(n)})\big)\Big\|^2 \\ &\leq \frac{1}{N-1}\sum_{n=1}^{N}\Big\|(I-P)\big(\Psi(P_{\mathcal{B}_V}u_{j-1}^{(n)}) - \Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1})\big)\Big\|^2 \\ &\leq \frac{1}{N-1}\sum_{n=1}^{N}\alpha\left(\big\|P_{\mathcal{B}_V}u_{j-1}^{(n)} - P_{\mathcal{B}_V}\widehat{m}_{j-1}\big\|^2 + \beta\big\|P(P_{\mathcal{B}_V}u_{j-1}^{(n)} - P_{\mathcal{B}_V}\widehat{m}_{j-1})\big\|^2\right) \\ &\leq \frac{1}{N-1}\sum_{n=1}^{N}\alpha\left(\big\|u_{j-1}^{(n)} - \widehat{m}_{j-1}\big\|^2 + \beta\big\|P(u_{j-1}^{(n)} - \widehat{m}_{j-1})\big\|^2\right) \\ &= \alpha\mathrm{Tr}\big(\widehat{C}_{j-1}\big) + \alpha\beta\mathrm{Tr}\big(P\widehat{C}_{j-1}P^*\big).\end{aligned}$$

The remainder of the proof proceeds exactly as for the mean-field version. □

The next theorem establishes long-time accuracy of the idealized mean-field Algorithm 5.

Theorem 3.3.2. *Suppose that Assumption 3.2.1 holds. Then, if the inflation parameter*

$\gamma > 0$ *is sufficiently large, there is a constant* C_1 *independent of* ε *such that*

$$\limsup_{j\to\infty} \mathbb{E}\|m_j - u_j\| \leq \mathsf{C}_1\varepsilon.$$

Proof. From Lemma 3.3.1, there exists a time j_* such that for all $j \geq j_*$, $\mathrm{Tr}(C_j) \leq \left(1 + \frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\right)\varepsilon^2$. Throughout the rest of the proof, we assume that $j \geq j_*$. We assume without loss of generality that $u_0 \in \mathcal{B}$, since if $u_0 \notin \mathcal{B}$, there exists a time j such that $u_j \in \mathcal{B}$, and we take $j_* > j$. Next, we write

$$\begin{aligned} u_j &= \Psi(u_{j-1}) \\ &= \big(I - \mathscr{K}(\Sigma_j + Q)H\big)\Psi(u_{j-1}) + \mathscr{K}(\Sigma_j + Q)H\Psi(u_{j-1}), \end{aligned} \tag{3.19}$$

and introduce the auxiliary variable

$$\begin{aligned} w_j &:= \big(I - \mathscr{K}(\Sigma_j + Q)H\big)\Psi(P_{\mathcal{B}_V} w_{j-1}) + \mathscr{K}(\Sigma_j + Q)y_j, \\ &= \big(I - \mathscr{K}(\Sigma_j + Q)H\big)\Psi(P_{\mathcal{B}_V} w_{j-1}) + \mathscr{K}(\Sigma_j + Q)H\Psi(u_{j-1}) + \varepsilon\mathscr{K}(\Sigma_j + Q)H\eta_j. \end{aligned} \tag{3.20}$$

Subtracting (3.19) from (3.20), we obtain that

$$\begin{aligned} w_j - u_j &= \big(I - \mathscr{K}(\Sigma_j + Q)H\big)\big(\Psi(P_{B_V} w_{j-1}) - \Psi(u_{j-1})\big) + \varepsilon\mathscr{K}(\Sigma_j + Q)\eta_j \\ &= \big(I - \mathscr{K}(\Sigma_j + Q)H\big)P\big(\Psi(P_{B_V} w_{j-1}) - \Psi(u_{j-1})\big) \\ &\quad + \big(I - P\big)\big(\Psi(P_{B_V} w_{j-1}) - \Psi(u_{j-1})\big) + \varepsilon\mathscr{K}(\Sigma_j + Q)\eta_j, \end{aligned}$$

where we used that $H(I - P) = 0$. Taking the V norm of both sides and using that $P^2 = P$

yields

$$
\begin{aligned}
V\big(w_j - u_j\big) &\le \big\|P - \mathscr{K}(\Sigma_j + Q)H\big\|_{V,V} V\Big(P\big(\Psi(P_{B_V} w_{j-1}) - \Psi(u_{j-1})\big)\Big) \\
&\quad + V\Big(\big(I - P\big)\big(\Psi(P_{B_V} w_{j-1}) - \Psi(u_{j-1})\big)\Big) + \varepsilon\big\|\mathscr{K}(\Sigma_j + Q)H\big\|_{V,V} V(H^*\eta_j) \\
&\le \big\|P - \mathscr{K}(\Sigma_j + Q)H\big\|_{V,V} L^{1/2} V\big(w_{j-1} - u_{j-1}\big) \\
&\quad + \alpha^{1/2} V\big(w_{j-1} - u_{j-1}\big) + \varepsilon\big\|\mathscr{K}(\Sigma_j + Q)H\big\|_{V,V} V(H^*\eta_j),
\end{aligned}
$$

where the last inequality follows by Assumption 3.2.1. Now, note that

$$
\begin{aligned}
\big\|P - \mathscr{K}(\Sigma_j + Q)H\big\|_{V,V} &= \sup_{V(v)=1} V\Big(\big(P - \mathscr{K}(\Sigma_j + Q)H\big)Pv\Big) \\
&\le \sup_{V(v)=1} \sqrt{1+\beta}\Big(\big\|\big(P - \mathscr{K}(\Sigma_j + Q)H\big)Pv\big\| + \big\|(I - P)\big(P - \mathscr{K}(\Sigma_j + Q)H\big)Pv\big\|\Big) \\
&\le \sup_{V(v)=1} \sqrt{1+\beta}\Big(\big\|(P - \mathscr{K}(\Sigma_j + Q)H)Pv\big\| + \sup_{V(v)=1} \big\|(I - P)\mathscr{K}(\Sigma_j + Q)Hv\big\|\Big).
\end{aligned}
$$

The first of the terms on the right-hand side can be bounded as

$$
\begin{aligned}
\big\|\big(P - \mathscr{K}(\Sigma_j + Q)H\big)Pv\big\| &\le \Big\|I - H(\Sigma_j + Q)H^*\big(H(\Sigma_j + Q)H^* + \varepsilon^2 R\big)^{-1}\Big\|_{op} \|Pv\| \\
&\le \varepsilon^2 \frac{\|R\|_{op}}{\lambda_{\min}(H\Sigma_j H^* + \gamma I)} \|Pv\| \\
&\le \varepsilon^2 \frac{\|R\|_{op}}{\gamma} \|Pv\| \le \varepsilon^2 \frac{\|R\|_{op}}{\gamma} V(v).
\end{aligned}
$$

For the second term, we have

$$\begin{aligned}\big\|(I-P)\mathscr{K}(\Sigma_j+Q)Hv\big\| &\le \big\|(I-P)(\Sigma_j+Q)\big\|_{op}\Big\|\big(H(\Sigma_j+Q)H^*+\varepsilon^2R\big)^{-1}\Big\|_{op}\|Hv\| \\ &\le \frac{\|\Sigma_j\|_{op}}{\lambda_{\min}\big(H(\Sigma_j+Q)H^*\big)}\|Hv\| \\ &\le \frac{L\|C_{j-1}\|_{op}}{\gamma}\|Hv\| \le \frac{L\mathrm{Tr}(C_{j-1})}{\gamma}\|Hv\| \\ &\le \frac{L\left(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\right)\varepsilon^2}{\gamma}\|Hv\| \le \frac{L\left(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\right)\varepsilon^2}{\gamma}V(v).\end{aligned}$$

Consequently,

$$\big\|P-\mathscr{K}(\Sigma_j+Q)H\big\|_{V,V} \le \frac{\varepsilon^2}{\gamma}\sqrt{1+\beta}\bigg(\|R\|_{op}+L\Big(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\Big)\bigg). \tag{3.21}$$

This implies that

$$\|\mathscr{K}(\Sigma_j+Q)H\|_{V,V} \le 1+\frac{\varepsilon^2}{\gamma}\sqrt{1+\beta}\bigg(\|R\|_{op}+L\Big(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\Big)\bigg). \tag{3.22}$$

By taking γ large enough to guarantee that $\big\|P-\mathscr{K}(\Sigma_j+Q)H\big\|_{V,V}L^{1/2} \le \frac{1-\alpha^{1/2}}{2}$ we obtain

$$\begin{aligned}V\big(w_j-u_j\big) &\le \alpha_* V\big(w_{j-1}-u_{j-1}\big) \\ &\quad+\varepsilon\bigg(1+\frac{\varepsilon^2}{\gamma}\sqrt{1+\beta}\bigg(\|R\|_{op}+L\Big(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\Big)\bigg)\bigg)V(H^*\eta_j),\end{aligned}$$

where $\alpha_* := \frac{1+\alpha^{1/2}}{2} < 1$. Taking the expectation of both sides of the inequality we have

$$\begin{aligned}\mathbb{E}V\big(w_j-u_j\big) &\le \alpha_*\mathbb{E}V\big(u_{j-1}-w_{j-1}\big) \\ &\quad+\varepsilon\bigg(1+\frac{\varepsilon^2}{\gamma}\sqrt{1+\beta}\bigg(\|R\|_{op}+L\Big(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\Big)\bigg)+\sqrt{2\mathrm{Tr}(R)}\bigg).\end{aligned}$$

Applying the discrete Gronwall inequality yields

$$\mathbb{E}V\big(w_j - u_j\big) \leq \frac{c_1\varepsilon}{1-\alpha_*}\Big(1-\alpha_*^{j-j_*}\Big) + \alpha_*^{j-j_*}\mathbb{E}V\big(u_{j_*} - w_{j_*}\big), \tag{3.23}$$

where $c_1 = \Big(1 + \frac{\varepsilon^2}{\gamma}\sqrt{1+\beta}\,\Big(\|R\|_{op} + L\,\Big(1 + \frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\Big)\Big) + \sqrt{2\mathrm{Tr}(R)}\Big)$.

We now consider the difference

$$\begin{aligned} m_j - w_j &= \big(I - \mathscr{K}(\Sigma_j + Q)H\big)\big(\mu_j - \Psi(P_{\mathcal{B}_V} w_{j-1})\big) \\ &= \big(I - \mathscr{K}(\Sigma_j + Q)H\big)\big(\mu_j - \Psi(P_{\mathcal{B}_V} m_{j-1})\big) \\ &\quad + \big(I - \mathscr{K}(\Sigma_j + Q)H\big)\big(\Psi(P_{\mathcal{B}_V} m_{j-1}) - \Psi(P_{\mathcal{B}_V} w_{j-1})\big). \end{aligned}$$

Carrying out the same argument as above, we have that

$$\begin{aligned} V\big(m_j - w_j\big) &\leq V\Big(\big(I - \mathscr{K}(\Sigma_j + Q)H\big)\big(\mu_j - \Psi(P_{\mathcal{B}_V} m_{j-1})\big)\Big) + \alpha_* V\big(m_{j-1} - w_{j-1}\big) \\ &\leq \|\,\big(I - \mathscr{K}(\Sigma_j + Q)H\big)\,\|_{V,V} V\big(\mu_j - \Psi(P_{\mathcal{B}_V} m_{j-1})\big) + \alpha_* V\big(m_{j-1} - w_{j-1}\big) \\ &\leq \left(1 + \frac{\varepsilon^2}{\gamma}\sqrt{1+\beta}\left(\|R\|_{op} + L\Big(1 + \frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\Big)\right)\right) V\big(\mu_j - \Psi(P_{\mathcal{B}_V} m_{j-1})\big) \\ &\quad + \alpha_* V\big(m_{j-1} - w_{j-1}\big). \end{aligned}$$

We now bound $V\big(\mu_j - \Psi(P_{\mathcal{B}_V} m_{j-1})\big)$:

$$\begin{aligned}
V\big(\mu_j - \Psi(P_{\mathcal{B}_V} m_{j-1})\big) &= V\left(\mathbb{E}_{u\sim\mathcal{N}(m_{j-1},C_{j-1})}\big[\Psi(P_{\mathcal{B}_V}u) - \Psi(P_{\mathcal{B}_V}m_{j-1})\big]\right) \\
&\le \left(\mathbb{E}_{u\sim\mathcal{N}(0,C_{j-1})}\Big[V\Big(\Psi\big(P_{\mathcal{B}_V}(u-m_{j-1})\big) - \Psi\big(P_{\mathcal{B}_V}m_{j-1}\big)\Big)\Big]\right)^{1/2} \\
&\le L^{1/2}\left(\mathbb{E}_{u\sim\mathcal{N}(0,C_{j-1})}\Big[V\Big(P_{\mathcal{B}_V}(u-m_{j-1}) - P_{\mathcal{B}_V}m_{j-1}\Big)\Big]\right)^{1/2} \\
&\le L^{1/2}\left(\mathbb{E}_{u\sim\mathcal{N}(0,C_{j-1})}\big[V\left(u-m_{j-1}\right)\big]\right)^{1/2} \\
&\le L^{1/2}\big(\mathrm{Tr}(C_{j-1}) + \beta\mathrm{Tr}(HC_{j-1}H^*)\big)^{1/2} \\
&\le \varepsilon L^{1/2}\sqrt{1+\beta}\left(1 + \left(\frac{1+\alpha\beta}{1-\alpha}+1\right)\mathrm{Tr}(R)\right)^{1/2}.
\end{aligned}$$

Consequently,

$$\mathbb{E}V\big(m_j - w_j\big) \le \alpha_*\mathbb{E}V\big(m_{j-1} - w_{j-1}\big) + c_2\varepsilon,$$

where

$$\begin{aligned}
c_2 = L^{1/2}\sqrt{1+\beta}\left(1 + \Big(\frac{1+\alpha\beta}{1-\alpha}+1\Big)\mathrm{Tr}(R)\right)^{1/2} \\
\cdot\left(1 + \frac{\varepsilon^2}{\gamma}\sqrt{1+\beta}\left(\|R\|_{op} + L\Big(1 + \frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\Big)\right)\right).
\end{aligned}$$

Therefore, the discrete Gronwall inequality gives

$$\mathbb{E}V\big(m_j - w_j\big) \le \frac{c_2\varepsilon}{1-\alpha_*}\left(1 - \alpha_*^{j-j_*}\right) + \alpha_*^{j-j_*}\mathbb{E}V\big(m_{j_*} - w_{j_*}\big). \tag{3.24}$$

Combining (3.23) and (3.24), we have

$$\mathbb{E}V\big(m_j - u_j\big) \le \frac{(c_1+c_2)\varepsilon}{1-\alpha_*}\left(1 - \alpha_*^{j-j_*}\right) + \alpha_*^{j-j_*}\mathbb{E}V\big(u_{j_*} - w_{j_*}\big) + \alpha_*^{j-j_*}\mathbb{E}V\big(m_{j_*} - w_{j_*}\big).$$

Since $\alpha_* \in (0,1)$ it follows that, for $c_3 = \frac{c_1+c_2}{1-\alpha^*}$,

$$\limsup_{j\to\infty} \mathbb{E}V\big(m_j - u_j\big) \leq c_3\varepsilon.$$

The equivalence of the $V(\cdot)$ and $\|\cdot\|$ norms concludes the proof. □

Remark 3.3.3. One interpretation of this result is that under Assumption 3.2.1, the predictive covariances given by the mean-field ensemble Kalman filter in Algorithm 5 behave similarly in the long-time asymptotic to the fixed predictive covariances in an optimally tuned 3DVAR algorithm [61]. Similarly to 3DVAR, the ensemble Kalman filter operates by strongly contracting the state estimate towards the observed data, while the unobservable direction is contractive by the squeezing property.

Remark 3.3.4. In the proof of Theorem 3.3.2 we show a bound on $\mathbb{E}V\big(m_j - u_j\big)$ that holds for finite j that is slightly stronger than the long-time asymptotic result presented in the theorem statement. Similar non-asymptotic bounds are also shown in the proofs of Theorems 3.3.5, 3.4.2, and 3.4.3. We refer readers to the proofs of the theorems for specific characterizations of the finite-time bounds.

We now show that the means output by the square-root filter Algorithm 3 and those output by the idealized mean-field Algorithm 5 remain close in the long-time asymptotic. The proof of our first main result, Theorem 3.2.2, follows directly by combining Theorem 3.3.2 and the next result:

Theorem 3.3.5. *Suppose that Assumption 3.2.1 holds and that $N \geq 6k$. Then, if the inflation parameter $\gamma > 0$ is sufficiently large, there is a constant C_2 independent of ε such that*

$$\limsup_{j\to\infty} \mathbb{E}\|\widehat{m}_j - m_j\| \leq \mathsf{C}_2\varepsilon. \tag{3.25}$$

Proof. We assume that γ is chosen to be at least as large as in Theorem 3.3.2. From Theorem 3.3.2 and Lemma 3.3.1, there exists a time j_* such that for all $j > j_*$, $\mathrm{Tr}(\widehat{C}_j) \leq \left(1 + \frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\right)\varepsilon^2$, $\mathrm{Tr}(C_j) \leq \left(1 + \frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\right)\varepsilon^2$, and $\mathbb{E}V(m_j - u_j) \leq \max\{1, 2\mathsf{C}_1\varepsilon\}$. For $j > j_*$, we consider the difference

$$\begin{aligned}\widehat{m}_j - m_j &= \big(I - \mathscr{K}(\widehat{\Sigma}_j)H\big)\widehat{\mu}_j - \big(I - \mathscr{K}(\Sigma_j + Q)H\big)\mu_j + \big(\mathscr{K}(\widehat{\Sigma}_j) - \mathscr{K}(\Sigma_j)\big)y_j \\ &= \big(I - \mathscr{K}(\Sigma_j + Q)H\big)\big(\mu_j - \widehat{\mu_j}\big) + \big(I - \mathscr{K}(\Sigma_j + Q)H - I + \mathscr{K}(\widehat{\Sigma}_j)H\big)\big(P\widehat{\mu}_j - H^* y_j\big).\end{aligned}$$

Taking the V norm of both sides, we have

$$\begin{aligned}V\left(\widehat{m}_j - m_j\right) \leq\ & V\Big(\big(I - \mathscr{K}(\Sigma_j + Q)H\big)\big(\mu_j - \widehat{\mu_j}\big)\Big) \\ & + V\Big(\big(\mathscr{K}(\widehat{\Sigma}_j)H - \mathscr{K}(\Sigma_j + Q)H\big)\big(P\widehat{\mu}_j - H^* y_j\big)\Big).\end{aligned} \tag{3.26}$$

We begin by bounding the first term on the right-hand side of (3.26)

$$\begin{aligned}V\Big(\big(I - \mathscr{K}(\Sigma_j + Q)H\big)\big(\mu_j - \widehat{\mu_j}\big)\Big) \leq\ & V\Big(\big(I - \mathscr{K}(\Sigma_j + Q)H\big)\big(\mu_j - \Psi(P_{B_V} m_{j-1})\big)\Big) \\ & + V\Big(\big(I - \mathscr{K}(\Sigma_j + Q)H\big)\big(\Psi(P_{B_V} m_{j-1}) - \Psi(P_{B_V}\widehat{m}_{j-1})\big)\Big) \\ & + V\left(\big(I - \mathscr{K}(\Sigma_j + Q)H\big)\left(\Psi(P_{B_V}\widehat{m}_{j-1}) - \frac{1}{N}\sum_{n=1}^{N}\Psi\big(P_{B_V} u_{j-1}^{(n)}\big)\right)\right) \\ & + V\left(\big(I - \mathscr{K}(\Sigma_j + Q)H\big)\frac{1}{N}\sum_{n=1}^{N}\xi_j^{(n)}\right).\end{aligned}$$

Recall that $\|I - \mathscr{K}(\Sigma_j + Q)H\|_{V,V} \leq c_1 := 1 + \frac{\varepsilon^2}{\gamma}\sqrt{1+\beta}\left(\|R\|_{op} + L\left(1 + \frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\right)\right)$. By the same arguments as in the proof of Theorem 3.3.2, we have

$$V\Big(\big(I - \mathscr{K}(\Sigma_j + Q)H\big)\big(\mu_j - \Psi(P_{\mathcal{B}_V} m_{j-1})\big)\Big) \leq c_2\varepsilon,$$

and

$$V\Big(\big(I-\mathscr{K}(\Sigma_j+Q)H\big)\big(\Psi(P_{\mathcal{B}_V}m_{j-1})-\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1})\big)\Big) \\ \leq \left(\frac{\varepsilon^2}{\gamma}c_3+\alpha^{1/2}\right)V\left(m_{j-1}-\widehat{m}_{j-1}\right),$$

with $c_2 = c_1 L^{1/2}\Big(1+\Big(1+\frac{1+\alpha\beta}{1-\alpha}\Big)\mathrm{Tr}(R)\Big)^{1/2}$ and $c_3 = \sqrt{1+\beta}\Big(\|R\|_{op}+L\left(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\right)\Big)$. Moreover,

$$\begin{aligned}
&V\left(\big(I-\mathscr{K}(\Sigma_j+Q)H\big)\Big(\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1})-\frac{1}{N}\sum_{n=1}^{N}\Psi(P_{\mathcal{B}_V}u_{j-1}^{(n)})\Big)\right)\\
&\leq c_1 V\left(\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1})-\frac{1}{N}\sum_{n=1}^{N}\Psi(P_{\mathcal{B}_V}u_{j-1}^{(n)})\right)\\
&\leq c_1\left(\frac{L}{N}\sum_{n=1}^{N}V\left(u_{j-1}^{(n)}-\widehat{m}_{j-1}\right)\right)^{1/2}\\
&\leq c_1 L^{1/2}\left(\mathrm{Tr}(\widehat{C}_{j-1})+\beta\mathrm{Tr}(H\widehat{C}_{j-1}H^*)\right)^{1/2}\\
&\leq \varepsilon c_1 L^{1/2}\sqrt{1+\beta}\left(1+\left(\frac{1+\alpha\beta}{1-\alpha}+1\right)\mathrm{Tr}(R)\right)^{1/2}\\
&\leq c_4\varepsilon,
\end{aligned}$$

where $c_4 = c_1 L^{1/2}\sqrt{1+\beta}\left(1+\left(\frac{1+\alpha\beta}{1-\alpha}+1\right)\mathrm{Tr}(R)\right)^{1/2}$. Finally, we have

$$\begin{aligned}
\mathbb{E}V\left(\big(I-\mathscr{K}(\Sigma_j+Q)H\big)\frac{1}{N}\sum_{n=1}^{N}\xi_j^{(n)}\right) &\le \sqrt{1+\beta}\big\|P-\mathscr{K}(\Sigma_j+Q)H\big\|_{V,V}\,\mathbb{E}\Big\|\frac{1}{N}\sum_{n=1}^{N}H\xi_j^{(n)}\Big\| \\
&\le \sqrt{1+\beta}\frac{\varepsilon^2}{\gamma}\Big(\|R\|_{op}+L\Big(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\Big)\Big)\mathbb{E}\Big\|\frac{1}{N}\sum_{n=1}^{N}H\xi_j^{(n)}\Big\| \\
&\overset{\text{(i)}}{\le} \sqrt{1+\beta}\frac{\varepsilon^2}{\gamma}\Big(\|R\|_{op}+L\Big(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\Big)\Big)\sqrt{\frac{\mathrm{Tr}(\gamma I)}{N}} \\
&\overset{\text{(ii)}}{\le} \sqrt{1+\beta}\varepsilon^2\gamma^{-1/2}\Big(\|R\|_{op}+L\Big(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\Big)\Big),
\end{aligned}$$

where (i) uses Lemma B.9 in [10] and (ii) follows from the assumption that $N \ge 6k$. Consequently,

$$\mathbb{E}V\Big(\big(I-\mathscr{K}(\Sigma_j+Q)H\big)\big(\mu_j-\widehat{\mu_j}\big)\Big) \le \left(\frac{\varepsilon^2}{\gamma}c_3+\alpha^{1/2}\right)\mathbb{E}V\big(\widehat{m}_{j-1}-m_{j-1}\big)+c_5\varepsilon, \quad (3.27)$$

where

$$c_5 := c_2+c_4+\varepsilon\gamma^{-1/2}\Big(\|R\|_{op}+\sqrt{1+\beta}L\Big(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\Big)\Big).$$

We now bound the second term in (3.26):

$$\begin{aligned}
&V\Big(\big(\mathscr{K}(\widehat{\Sigma}_j)H-\mathscr{K}(\Sigma_j+Q)H\big)\big(P\widehat{\mu}_j-H^*y_j\big)\Big) \\
&\le V\Big(\big(P-\mathscr{K}(\widehat{\Sigma}_j)H\big)\big(P\widehat{\mu}_j-H^*y_j\big)\Big)+V\Big(\big(P-\mathscr{K}(\Sigma_j+Q)H\big)\big(P\widehat{\mu}_j-H^*y_j\big)\Big) \\
&\le \Big(\|P-\mathscr{K}(\Sigma_j+Q)H\|_{V,V}V+\|P-\mathscr{K}(\widehat{\Sigma}_j)H\|_{V,V}\Big)V\big(P\widehat{\mu}_j-H^*y_j\big) \\
&\le \frac{\varepsilon^2}{\lambda_{\min}(H\Sigma_jH^*+\gamma I)}c_3V\big(P\widehat{\mu}_j-H^*y_j\big)+\frac{\varepsilon^2}{\lambda_{\min}(H\widehat{\Sigma}_jH^*)}c_3V\big(P\widehat{\mu}_j-H^*y_j\big).
\end{aligned}$$

Taking expectations,

$$
\begin{aligned}
&\mathbb{E}V\Big(\big(\mathscr{K}(\widehat{\Sigma}_j)H - \mathscr{K}(\Sigma_j + Q)H\big)\big(P\widehat{\mu}_j - H^* y_j\big)\Big) \\
&\leq \frac{\varepsilon^2}{\gamma} c_3 \mathbb{E}V\big(P\widehat{\mu}_j - H^* y_j\big) + \varepsilon^2 c_3 \mathbb{E}\left[\frac{1}{\lambda_{\min}(H\widehat{\Sigma}_j H^*)} V\left(P\widehat{\mu}_j - H^* y_j\right)\right] \\
&\leq \frac{\varepsilon^2}{\gamma} c_3 \mathbb{E}V\big(P\widehat{\mu}_j - H^* y_j\big) + \varepsilon^2 c_3 \mathbb{E}\left[\mathbb{E}\left[\frac{1}{\lambda_{\min}(H\widehat{\Sigma}_j H)} V\big(P\widehat{\mu}_j - H^* y_j\big)\Big| \{u_{j-1}^{(n)}\}_{n=1}^N\right]\right].
\end{aligned}
$$

Denoting $\mathbb{E}^{\mathsf{U_{j-1}}}[\,\cdot\,] := \mathbb{E}\Big[\,\cdot\,\big|\{u_{j-1}^{(n)}\}_{n=1}^N\Big]$, we then compute

$$
\begin{aligned}
&\mathbb{E}^{\mathsf{U_{j-1}}}\left[\frac{1}{\lambda_{\min}(H\widehat{\Sigma}_j H)} V\left(P\widehat{\mu}_j - H^* y_j\right)\right] \\
&\leq \mathbb{E}^{\mathsf{U_{j-1}}}\left[\frac{1}{\lambda_{\min}(H\widehat{\Sigma}_j H)}\left(V\left(\frac{1}{N}\sum_{n=1}^N P\Psi(P_{\mathcal{B}_V} u_{j-1}^{(n)}) - P\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1})\right)\right.\right. \\
&+ V\Big(P\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}) - P\Psi(P_{\mathcal{B}_V} m_{j-1})\Big) + V\Big(P\Psi(P_{\mathcal{B}_V} m_{j-1}) - P\Psi(u_{j-1})\Big) \\
&\left.\left. + \varepsilon V\left(H^*\eta_j\right) + V\left(\frac{1}{N}\sum_{n=1}^N \xi_j^{(n)}\right)\right)\right] \\
&\overset{\text{(i)}}{\leq} \mathbb{E}^{\mathsf{U_{j-1}}}\left[\frac{1}{\lambda_{\min}(H\widehat{\Sigma}_j H)}\right]\mathbb{E}^{\mathsf{U_{j-1}}}\left[V\left(\frac{1}{N}\sum_{n=1}^N \Psi(P_{\mathcal{B}_V} u_{j-1}^{(n)}) - \Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1})\right)\right. \\
&\left. + V\Big(\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}) - \Psi(P_{\mathcal{B}_V} m_{j-1})\Big) + V\Big(\Psi(P_{\mathcal{B}_V} m_{j-1}) - \Psi(u_{j-1})\Big) + \varepsilon V\left(H^*\eta_j\right)\right] \\
&+ \mathbb{E}^{\mathsf{U_{j-1}}}\left[\frac{1}{\lambda_{\min}(H\widehat{\Sigma}_j H)} V\left(\frac{1}{N}\sum_{n=1}^N \xi_j^{(n)}\right)\right] \\
&\overset{\text{(ii)}}{\leq} \mathbb{E}^{\mathsf{U_{j-1}}}\left[\frac{1}{\lambda_{\min}(H\widehat{\Sigma}_j H)}\right]\mathbb{E}^{\mathsf{U_{j-1}}}\left[V\left(\frac{1}{N}\sum_{n=1}^N \Psi(P_{\mathcal{B}_V} u_{j-1}^{(n)}) - \Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1})\right)\right. \\
&\left. + V\Big(\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}) - \Psi(P_{\mathcal{B}_V} m_{j-1})\Big) + V\Big(\Psi(P_{\mathcal{B}_V} m_{j-1}) - \Psi(u_{j-1})\Big) + \varepsilon V\left(H^*\eta_j\right)\right] \\
&+ \left(\mathbb{E}^{\mathsf{U_{j-1}}}\left[\frac{1}{\lambda_{\min}(H\widehat{\Sigma}_j H)^2}\right]\mathbb{E}^{\mathsf{U_{j-1}}}\left[V^2\left(\frac{1}{N}\sum_{n=1}^N \xi_j^{(n)}\right)\right]\right)^{1/2},
\end{aligned}
$$

where (i) uses conditional independence and (ii) uses Cauchy-Schwarz. If γ is chosen to guarantee that $\frac{10NL\left(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\right)\varepsilon^2}{\gamma k} \leq 1$, Lemma 3.3.6 yields that

$$
\mathbb{E}^{\mathsf{U_{j-1}}}\left[\frac{1}{\lambda_{\min}(H\widehat{\Sigma}_j H)}\right] \leq \frac{2C'}{\gamma} \quad \text{and} \quad \mathbb{E}^{\mathsf{U_{j-1}}}\left[\frac{1}{\lambda_{\min}(H\widehat{\Sigma}_j H)^2}\right]^{1/2} \leq \frac{\sqrt{2}C'}{\gamma}.
$$

Note that both of these upper bounds are independent of $\{u_{j-1}^{(n)}\}_{n=1}^N$. Consequently,

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{E}^{\mathsf{U}_{\mathsf{j}-1}}\Big[\frac{1}{\lambda_{\min}(H\widehat{\Sigma}_j H)}V\left(P\widehat{\mu}_j - H^* y_j\right)\Big]\right] \\
&\leq \frac{2C'}{\gamma}\mathbb{E}V\left(\frac{1}{N}\sum_{n=1}^{N}\Psi(P_{\mathcal{B}_V}u_{j-1}^{(n)}) - \Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1})\right) \\
&+ \mathbb{E}V\Big(\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}) - \Psi(P_{\mathcal{B}_V}m_{j-1})\Big) \\
&+ \mathbb{E}V\Big(\Psi(P_{\mathcal{B}_V}m_{j-1}) - \Psi(u_{j-1})\Big) + \varepsilon\mathbb{E}V\left(H^*\eta_j\right) + \mathbb{E}\left[V^2\Big(\frac{1}{N}\sum_{n=1}^{N}\xi_j^{(n)}\Big)\right]^{1/2} \\
&\leq \frac{2C'}{\gamma}\Bigg\{L^{1/2}\left(\mathrm{Tr}\big(\widehat{C}_{j-1}\big) + \beta\mathrm{Tr}\big(H\widehat{C}_{j-1}H^*\big)\right)^{1/2} \\
&+ L^{1/2}\mathbb{E}V\left(\widehat{m}_{j-1} - m_{j-1}\right) + L^{1/2}\mathbb{E}V\left(m_{j-1} - u_{j-1}\right) \\
&+ \varepsilon\sqrt{1+\beta}\mathrm{Tr}\left(R\right)^{1/2} + \sqrt{1+\beta}\sqrt{\frac{\mathrm{Tr}(\gamma I)}{N}}\Bigg\}.
\end{aligned}
$$

Since $j > j_*$, we have that

$$
\mathbb{E}V\big(\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}) - \Psi(P_{\mathcal{B}_V}m_{j-1})\big) \leq \frac{1}{\gamma}\left(c_6 + c_7\mathbb{E}V(\widehat{m}_{j-1} - m_{j-1})\right) + \frac{c_8}{\gamma^{1/2}}, \tag{3.28}
$$

with $c_6 = 2\varepsilon\sqrt{1+\beta}C'\left(1 + \frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\right)^{1/2} + 2C'L^{1/2}\max\{2\mathsf{C}_1\varepsilon, 1\} + 2C'\varepsilon\sqrt{1+\beta}\mathrm{Tr}(R)^{1/2}$, $c_7 = 2C'L^{1/2}$, and $c_8 = C'\sqrt{\frac{1+\beta}{3}}$. By a similar computation, we have

$$
\mathbb{E}V(P\widehat{\mu}_j - H^* y_j) \leq c_9 + L^{1/2}\mathbb{E}V(\widehat{m}_{j-1} - m_{j-1}) + c_{10}\gamma^{1/2}, \tag{3.29}
$$

with $c_9 = \varepsilon\sqrt{1+\beta}\left(1 + \frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)\right)^{1/2} + L^{1/2}\max\{2\mathsf{C}_1\varepsilon, 1\} + \varepsilon\sqrt{1+\beta}\mathrm{Tr}(R)^{1/2}$ and $c_{10} =$

$\sqrt{\frac{1+\beta}{6}}$. Combining (3.28) and (3.29),

$$\begin{aligned}
&\mathbb{E}V\Big(\big(\mathscr{K}(\widehat{\Sigma}_j)H - \mathscr{K}(\Sigma_j+Q)H\big)\big(P\widehat{\mu}_j - H^*y_j\big)\Big) \\
&\leq \frac{\varepsilon^2}{\gamma}c_3\left(c_9 + L^{1/2}\mathbb{E}V(\widehat{m}_{j-1}-m_{j-1}) + c_{10}\gamma^{1/2} + c_6 + c_7\mathbb{E}V(\widehat{m}_{j-1}-m_{j-1}) + c_8\gamma^{1/2}\right).
\end{aligned} \tag{3.30}$$

Together, (3.27) and (3.29) yield

$$\mathbb{E}V\big(\widehat{m}_j - m_j\big) \leq \big(\frac{\varepsilon^2}{\gamma}c_3\left(1 + L^{1/2} + c_7\right) + \alpha^{1/2}\big)\mathbb{E}V\big(\widehat{m}_{j-1} - m_{j-1}\big) + c_{11}\varepsilon, \tag{3.31}$$

where $c_{11} = c_5 + \frac{\varepsilon}{\gamma}c_3\left(c_9 + c_{10}\gamma^{1/2} + c_6 + c_8\gamma^{1/2}\right)$. Taking γ to be large enough to guarantee that $\frac{\varepsilon^2}{\gamma}c_3\left(1 + L^{1/2} + c_7\right) \leq \frac{1-\alpha^{1/2}}{2}$, we obtain

$$\mathbb{E}V\big(\widehat{m}_j - m_j\big) \leq \alpha_*\mathbb{E}V\big(\widehat{m}_{j-1} - m_{j-1}\big) + c_{11}\varepsilon.$$

Applying the discrete Gronwall inequality and using the equivalence of $V(\cdot)$ and $\|\cdot\|$ yields the desired result. □

The following lemma was used in the proof of Theorem 3.3.2. This lemma is a generalization of Corollary 4 in [262] to the setting where the samples are independent, but have different means and thus are not identically distributed.

Lemma 3.3.6. *Let $\{u^{(n)}\}_{n=1}^N \in \mathcal{H}$ with empirical covariance satisfying $\mathrm{Tr}(\widehat{C}) \leq c\varepsilon^2$. Consider $v^{(n)} = \Psi(P_{\mathcal{B}_V}u^{(n)}) - \frac{1}{N}\sum_{n=1}^N \Psi(P_{\mathcal{B}_V}u^{(n)}) + \xi^{(n)}$ with $\xi^{(n)} \sim \mathcal{N}(0, \gamma P)$. Then, if $N \geq \min\{6k, 12\}$ and $\gamma \geq \frac{10NLc\varepsilon^2}{k}$, the sample covariance $\widehat{\Sigma} = \frac{1}{N}\sum_{n=1}^N v^{(n)} \otimes v^{(n)}$ satisfies*

$$\mathbb{E}\left[\max\{1, \lambda_{\min}(H\widehat{\Sigma}H^*)^{-q}\}\right]^{1/q} \leq \frac{2C'}{\gamma},$$

for any $1 \leq q \leq N/12$, where C' is a universal constant.

Proof. If we can show that for $N \geq 6k$, for every $t \in (0,1)$,

$$\mathbb{P}\left(\lambda_{\min}\Big(\frac{1}{\gamma}H\widehat{\Sigma}H^*\Big) \leq t\right) \leq (C't)^{N/6}, \tag{3.32}$$

for a universal constant C', then the desired result follows exactly as in [262]. Showing that (3.32) holds is tantamount to verifying that the conclusion of Theorem 4 in [262] holds in our modified setting. We begin by showing that each $Hv^{(n)}$ satisfies a small ball condition. We have that for every $\theta \in \mathbb{R}^k$ with $\|\theta\|^2 = 1$

$$\begin{aligned}
\mathbb{P}\left(|\langle \theta, Hv^{(n)}\rangle| \leq t\right) &\leq \sup_{y\in\mathbb{R}} \mathbb{P}\left(|\langle \theta', \gamma^{-1/2}H\xi^{(n)}\rangle - y| \leq t\right) \\
&\overset{\text{(i)}}{\leq} \sup_{y\in\mathbb{R}} \int \mathbf{1}\left\{|z-y| \leq t\right\} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz \\
&\leq \sup_{y\in\mathbb{R}} \int \mathbf{1}\{|z-\gamma| \leq t\} \frac{1}{\sqrt{2\pi}} dz \leq \frac{\sqrt{2}}{\pi} t,
\end{aligned} \tag{3.33}$$

where (i) uses the fact that $z = \langle \theta, \gamma^{-1/2}H\xi^{(n)}\rangle \sim \mathcal{N}(0,1)$ for $\|\theta\|^2 = 1$. Given this, we now argue that the proof of Theorem 4 in [262] goes through essentially as written. As in [262], we replace $\gamma^{-1/2}Hv^{(n)}$ by the truncated vectors $\omega^{(n)} := \max\{1, \frac{\sqrt{k}}{\|\gamma^{-1/2}Hv^{(n)}\|}\}\gamma^{-1/2}Hv^{(n)}$, and define $\widehat{\Sigma}_\omega := \frac{1}{N}\sum_{n=1}^N \omega^{(n)}(\omega^{(n)})^T$. Since $\widehat{\Sigma}_\omega \preceq \widehat{\Sigma}$, we can proceed by establishing a

lower bound for $\lambda_{\min}(\widehat{\Sigma}_\omega)$. For every $\|\theta\| = 1$, $t \in (0, \frac{\pi}{\sqrt{2}})$, and $b \geq 1$,

$$\begin{aligned}
\mathbb{P}\left(|\langle \omega^{(n)}, \theta\rangle| \leq t\right) &\leq \mathbb{P}\left(|\langle \gamma^{-1/2} H v^{(n)}, \theta\rangle| \leq bt\right) + \mathbb{P}\left(\frac{\sqrt{k}}{\|\gamma^{-1/2} H v^{(n)}\|} \leq \frac{1}{b}\right) \\
&\leq \frac{\sqrt{2}}{\pi} bt + \mathbb{P}\left(\|\gamma^{-1/2} H v^{(n)}\| \geq b\sqrt{k}\right) \\
&\overset{\text{(i)}}{\leq} \frac{\sqrt{2}}{\pi} bt + \frac{\mathbb{E}\|\gamma^{-1/2} H v^{(n)}\|^2}{b^2 k} \\
&\leq \frac{\sqrt{2}}{\pi} bt + \frac{k + \gamma^{-1}\|\Psi(P_{\mathcal{B}_V} u^{(n)}) - \frac{1}{N}\sum_{n=1}^{N} \Psi(P_{\mathcal{B}_V} u^{(n)})\|^2}{b^2 k} \\
&\leq \frac{\sqrt{2}}{\pi} bt + \frac{k + \gamma^{-1} NL \mathrm{Tr}(\widehat{C})}{b^2 k} \\
&\leq \frac{\sqrt{2}}{\pi} bt + \frac{k + \gamma^{-1} NLc\varepsilon^2}{b^2 k} \overset{\text{(ii)}}{\leq} \frac{\sqrt{2}}{\pi} bt + \frac{1 + \frac{1}{10}}{b^2},
\end{aligned}$$

where (i) uses Markov's inequality and (ii) uses our assumption on γ. Taking $b = \left(\frac{\sqrt{2}}{\pi} t\right)^{-1/3}$, we have

$$\mathbb{P}\left(|\langle \omega^{(n)}, \theta\rangle| \leq t\right) \leq \left(2 + \frac{1}{10}\right)\left(\frac{\sqrt{2}}{\pi} t\right)^{2/3}, \tag{3.34}$$

for each $1 \leq n \leq N$. Note that this agrees with Equation 69 in [262] with a constant factor $2 + \frac{1}{10}$ in place of the constant factor 2. Lemma 7 in [262] then implies that for every $\lambda > 0$,

$$\mathbb{E}\left[\exp\left(-\lambda \langle \omega^{(n)}, \theta\rangle^2\right)\right] \leq 3\left(\frac{2}{\pi^2}/\lambda\right)^{1/3},$$

for each $1 \leq n \leq N$. Defining

$$Z_n(\theta) = -\lambda \langle \omega^{(n)}, \theta\rangle^2 + \frac{1}{3}\log\left(\frac{\lambda}{C^2}\right) - \log\left(2 + \frac{1}{10}\right),$$

for some fixed $\lambda > 0$, we necessarily have that $\mathbb{E}\big[\exp\big(Z_n(\theta)\big)\big] \leq 1$. Consequently, letting

$$Z(\theta) := \sum_{n=1}^{N} Z_n(\theta) = N\left[-\lambda\langle\widehat{\Sigma}_\omega\theta,\theta\rangle + \frac{1}{3}\log\left(\frac{\lambda}{C^2}\right) - \log\left(2+\frac{1}{10}\right)\right],$$

the fact that $\omega^{(n)}$ are independent given $\{u^{(n)}\}_{n=1}^N$ implies that for every $\|\theta\| = 1$,

$$\mathbb{E}[Z(\theta)] \leq 1. \tag{3.35}$$

The PAC-Bayesian argument only considers $Z(\theta)$, and goes through exactly as in [262]. In the conclusion of the proof, [262] uses the assumption that $\frac{k}{N} \leq \frac{1}{6}$ to argue that $\log 2 + \frac{k}{N}\log\left(\frac{5N}{2k}\right) + \frac{k}{2N} \leq c_0 = 1.3$. Since it also holds that $\log\left(2+\frac{1}{10}\right) + \frac{k}{N}\log\left(\frac{5N}{2k}\right) + \frac{k}{2N} \leq c_0 = 1.3$, the proof concludes as the proof of Theorem 4 in [262] to show that (3.32) holds, and our desired result follows. □

3.4 Filter accuracy with surrogate models via mean-field theory

In this section, we prove the long-time accuracy of an ensemble Kalman filters which employs a surrogate dynamics model satisfying Assumption 3.2.7. We begin by proving an analogue of Lemma 3.3.1 for the covariance operators generated by Algorithms 5 and 6.

Algorithm 6 Gaussian Projected Filter with a Surrogate Model

1: **Initialization**: m_0, C_0.

2: For $j = 1, 2, \ldots$ do the following:

3: **Prediction**:

$$\begin{aligned} \mu_j^s &= \mathbb{E}_{u\sim\mathcal{N}(m_{j-1},C_{j-1})}\big[\Psi^s(P_{\mathcal{B}_V}u)\big], \\ \Sigma_j^s &= \mathbb{E}_{u\sim\mathcal{N}(m_{j-1},C_{j-1})}\big[\big(\Psi^s(P_{\mathcal{B}_V}u)-\mu_j\big)\otimes\big(\Psi^s(P_{\mathcal{B}_V}u)-\mu_j\big)\big]. \end{aligned} \tag{3.36}$$

4: **Analysis**:

$$\begin{aligned} m_j^s &= \mu_j^s + \mathscr{K}(\Sigma_j^s+Q)\left(y_j - H\mu_j^s\right), \\ C_j^s &= \big(I-\mathscr{K}(\Sigma_j^s+Q)H\big)(\Sigma_j^s+Q). \end{aligned} \tag{3.37}$$

5: **Output**: Means and covariances $\{m_j^s, C_j^s\}_{j=1}^{\infty}$.

The next result, analogous to Lemma 3.3.1, shows that the trace of the covariances generated by Algorithms 4 and 6 are controlled at the noise level ε and the surrogate error δ in the long-time asymptotic.

Lemma 3.4.1. *Under Assumptions 3.2.1 and 3.2.7, the covariance operators output by Algorithm 6 satisfy*

$$\mathrm{Tr}\left(C_j^s\right) \le \alpha_*^j \mathrm{Tr}\left(C_0\right) + (1+\alpha_*\beta)\,\frac{1-\alpha_*^j}{1-\alpha_*}\varepsilon^2\mathrm{Tr}\left(R\right) + 4\frac{1+\alpha}{1-\alpha}\frac{1-\alpha_*^j}{1-\alpha_*}\delta^2, \tag{3.38}$$

and those output by Algorithm 4 satisfy

$$\mathrm{Tr}\big(\widehat{C}_j^s\big) \le \alpha_*^j \mathrm{Tr}\big(\widehat{C}_0\big) + (1+\alpha_*\beta)\,\frac{1-\alpha_*^j}{1-\alpha_*}\varepsilon^2\mathrm{Tr}\left(R\right) + 4\frac{1+\alpha}{1-\alpha}\frac{1-\alpha_*^j}{1-\alpha_*}\delta^2, \tag{3.39}$$

where $\alpha_* := \frac{1+\alpha}{2}$.

Proof. Since P is an orthogonal projection and $Q = \gamma P$, we have

$$\begin{aligned} \mathrm{Tr}(C_j^s) &= \mathrm{Tr}(PC_j^sP^*) + \mathrm{Tr}\big((I-P)C_j^s(I-P)^*\big) \\ &\leq \mathrm{Tr}(PC_j^sP^*) + \mathrm{Tr}\big((I-P)\Sigma_j^s(I-P)^*\big). \end{aligned} \tag{3.40}$$

We next bound in turn the two terms in the right-hand side. The first term proceeds exactly as in Lemma 3.3.1. Let $\{\varphi_i\}_{i=1}^k$ be an orthonormal basis for the range of P, which we can extend to a basis for $\mathcal{H}$ via Gram-Schmidt by the separability of $\mathcal{H}$. We then compute that

$$\begin{aligned} \mathrm{Tr}\left(PC_j^sP^*\right) &= \sum_{i=1}^{\infty}\langle\varphi_i, PC_j^sP^*\varphi_i\rangle = \sum_{i=1}^{k}\langle\varphi_i, PC_j^sP^*\varphi_i\rangle \\ &= \sum_{i=1}^{k}\langle H\varphi_i, HC_j^sH^*H\varphi_i\rangle_{\mathbb{R}^k} = \mathrm{Tr}\left(HC_j^sH^*\right) \leq \varepsilon^2\mathrm{Tr}\left(R\right), \end{aligned} \tag{3.41}$$

where we have used that $\{H\varphi_i\}_{i=1}^k$ is an orthonormal basis for $\mathbb{R}^k$ and that $\varepsilon^2 R - HC_j^sH^*$ is positive semi-definite in the last inequality.

For the second term in the right-hand side of (3.40), note that

$$
\begin{aligned}
\mathrm{Tr}\Big((I-P)\Sigma_j^s(I-P)^*\Big) &= \mathbb{E}_{\mathcal{N}(m_{j-1}^s, C_{j-1}^s)}\Big[\big\|(I-P)\big(\Psi^s(P_{\mathcal{B}_V}u)-\mu_j\big)\big\|^2\Big] \\
&\le \mathbb{E}_{\mathcal{N}(m_{j-1}^s, C_{j-1}^s)}\Big[\big\|(I-P)\big(\Psi^s(P_{\mathcal{B}_V}u)-\Psi(P_{\mathcal{B}_V}m_{j-1}^s)\big)\big\|^2\Big] \\
&\le \mathbb{E}_{\mathcal{N}(m_{j-1}^s, C_{j-1}^s)}\Big[\big\|(I-P)\big(\Psi^s(P_{\mathcal{B}_V}u)-\Psi(P_{\mathcal{B}_V}u)\big)\big\|^2 \\
&+ 2\big\|(I-P)\big(\Psi^s(P_{\mathcal{B}_V}u)-\Psi(P_{\mathcal{B}_V}u)\big)\big\|\big\|(I-P)(\Psi(P_{\mathcal{B}_V}u)-\Psi(P_{\mathcal{B}_V}m_{j-1}^s)\big\| \\
&+ \big\|(I-P)\big(\Psi(P_{\mathcal{B}_V}u)-\Psi(P_{\mathcal{B}_V}m_{j-1}^s)\big)\big\|^2\Big] \\
&\overset{\text{(i)}}{\le} \mathbb{E}_{\mathcal{N}(m_{j-1}^s, C_{j-1}^s)}\Big[4\frac{1+\alpha}{1-\alpha}\|(I-P)\big(\Psi^s(P_{\mathcal{B}_v}u)-\Psi(P_{\mathcal{B}_V}u)\big)\big\|^2 \\
&+ \frac{1+\alpha}{2\alpha}\big\|(I-P)\big(\Psi(P_{\mathcal{B}_V}u)-\Psi(P_{\mathcal{B}_V}m_{j-1}^s)\big)\big\|^2\Big] \\
&\overset{\text{(ii)}}{\le} 4\frac{1+\alpha}{1-\alpha}\delta^2 + \mathbb{E}_{\mathcal{N}(m_{j-1}^s, C_{j-1}^s)}\Big[\frac{1+\alpha}{2\alpha}\big\|(I-P)\big(\Psi(P_{\mathcal{B}_V}u)-\Psi(P_{\mathcal{B}_V}m_{j-1}^s)\big)\big\|^2\Big] \\
&\overset{\text{(iii)}}{\le} 4\frac{1+\alpha}{1-\alpha}\delta^2 + \mathbb{E}_{\mathcal{N}(m_{j-1}^s, C_{j-1}^s)}\Big[\frac{1+\alpha}{2}\Big(\big\|u-m_{j-1}^s\big\|^2 + \beta\big\|(I-P)(u-m_{j-1}^s)\big\|^2\Big)\Big] \\
&= 4\frac{1+\alpha}{1-\alpha}\delta^2 + \alpha_*\mathrm{Tr}\left(C_{j-1}^s\right) + \alpha_*\beta\mathrm{Tr}\left(PC_{j-1}^sP^*\right),
\end{aligned}
\tag{3.42}
$$

where (i) uses Young's inequality, (ii) uses Assumption 3.2.7, and (iii) uses Assumption 3.2.1. Combining (3.40), (3.41), and (3.42), we deduce that

$$
\begin{aligned}
\mathrm{Tr}(C_j^s) &\le \varepsilon^2\mathrm{Tr}(R) + 4\frac{1+\alpha}{1-\alpha}\delta^2 + \alpha_*\mathrm{Tr}\left(C_{j-1}^s\right) + \alpha_*\beta\mathrm{Tr}\left(PC_{j-1}^sP^*\right) \\
&\le 4\frac{1+\alpha}{1-\alpha}\delta^2 + (1+\alpha_*\beta)\varepsilon^2\mathrm{Tr}(R) + \alpha_*\mathrm{Tr}(C_{j-1}^s).
\end{aligned}
$$

Applying the discrete Gronwall lemma yields (3.38). The proof of (3.39) is similar. □

The next result establishes long-time accuracy for the idealized mean-field Algorithm 6.

Theorem 3.4.2. *Suppose that Assumptions 3.2.1 and 3.2.7 hold. Then, if the inflation*

parameter $\gamma > 0$ is sufficiently large, there is a constant C_3 independent of ε and δ such that

$$\limsup_{j\to\infty} \mathbb{E}\|m_j - m_j^s\| \le \mathsf{C}_3(\varepsilon + \delta). \tag{3.43}$$

Proof. Assume γ is taken to be at least as large as in Theorem 3.3.2. From Lemma 3.3.1, Lemma 3.4.1 and Theorem 3.3.2 there exists a time j_* such that for all $j \ge j_*$, $\mathrm{Tr}(C_j) \le c_1\varepsilon^2$ with $c_1 := 1 + \frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)$, $\mathrm{Tr}(C_j^s) \le c_2(\varepsilon^2 + \delta^2)$ with $c_2 = 2\max\{(1 + \alpha_*\beta)\frac{1}{1-\alpha_*}\mathrm{Tr}(R), 4\frac{1+\alpha}{1-\alpha}\frac{1}{1-\alpha_*}\}$, and that $\mathbb{E}V(m_j - u_j) \le 2\mathsf{C}_1\varepsilon$. For $j > j^*$, we consider the difference

$$\begin{aligned}
m_j^s - m_j &= \Big(I - \mathscr{K}(\Sigma_j^s + Q)H\Big)\,\mu_j^s + \mathscr{K}(\Sigma_j^s + Q)y_j \\
&\quad - \big(I - \mathscr{K}(\Sigma_j + Q)\big)\,\mu_j - \mathscr{K}(\Sigma_j + Q)y_j \\
&= \Big(I - \mathscr{K}(\Sigma_j^s + Q)H\Big)(\mu_j^s - \mu_j) \\
&\quad + \Big(I - \mathscr{K}(\Sigma_j^s + Q)H - I + \mathscr{K}(\Sigma_j + Q)H\Big)(\mu_j - H^*y_j).
\end{aligned}$$

Taking the V norm of both sides yields

$$\begin{aligned}
V\big(m_j^s - m_j\big) &\le V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H\big)\big(\mu_j^s - \mu_j\big)\Big) \\
&\quad + V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H - I + \mathscr{K}(\Sigma_j + Q)H\big)\big(\mu_j - H^*y_j\big)\Big).
\end{aligned} \tag{3.44}$$

The first of the terms on the right-hand side of (3.44) can be bounded as

$$\begin{aligned}
V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H\big)\big(\mu_j^s - \mu_j\big)\Big) &\le V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H\big)\big(\mu_j^s - \Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)\big)\Big) \\
&\quad + V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H\big)\big(\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s) - \Psi(P_{\mathcal{B}_V}m_{j-1}^s)\big)\Big) \\
&\quad + V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H\big)\big(\Psi(P_{\mathcal{B}_V}m_{j-1}^s) - \Psi(P_{\mathcal{B}_V}m_{j-1})\big)\Big) \\
&\quad + V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H\big)\big(\Psi(P_{\mathcal{B}_V}m_{j-1}) - \mu_j\big)\Big).
\end{aligned}$$

We bound each of these terms in turn. By an analogous argument to that in Theorem 3.3.2,

$$\|P - \mathscr{K}(\Sigma_j^s + Q)H\|_{V,V} \leq \frac{1}{\gamma}\sqrt{1+\beta}\Big(\varepsilon^2\|R\|_{op} + L_s c_1(\varepsilon^2 + \delta^2)\Big) \tag{3.45}$$

and

$$\|I - \mathscr{K}(\Sigma_j^s + Q)H\|_{V,V} \leq 1 + \frac{1}{\gamma}\sqrt{1+\beta}\Big(\varepsilon^2\|R\|_{op} + L_s c_1(\varepsilon^2 + \delta^2)\Big). \tag{3.46}$$

Defining $c_3 := \sqrt{1+\beta}\Big(\varepsilon^2\|R\|_{op} + L_s c_2(\varepsilon^2 + \delta^2)\Big)$, we have

$$\begin{aligned}
V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H\big)\big(\mu_j^s - \Psi^s(m_{j-1}^s)\big)\Big) &\leq \|I - \mathscr{K}(\Sigma_j^s + Q)H\|_{V,V} V\Big(\mu_j^s - \Psi^s(m_{j-1}^s)\Big) \\
&\leq \Big(1 + \frac{c_3}{\gamma}\Big) V\Big(\mu_j^s - \Psi^s(m_{j-1}^s)\Big) \\
&\leq \Big(1 + \frac{c_3}{\gamma}\Big) L_s^{1/2}\Big(\mathrm{Tr}(C_{j-1}^s) + \beta\mathrm{Tr}(HC_{j-1}H^*)\Big)^{1/2} \\
&\leq \Big(1 + \frac{c_3}{\gamma}\Big) L_s^{1/2}\sqrt{1+\beta}\Big(c_2(\varepsilon^2 + \delta^2) + \varepsilon^2\mathrm{Tr}(R)\Big)^{1/2} \\
&\leq c_4(\varepsilon + \delta),
\end{aligned} \tag{3.47}$$

with $c_4 := (1 + \frac{c_3}{\gamma})L_s^{1/2}\sqrt{1+\beta}(c_2 + \mathrm{Tr}(R))^{1/2}$. Next, we have that

$$
\begin{aligned}
&V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H\big)\big(\Psi^s(P_{\mathcal{B}_V} m_{j-1}^s) - \Psi(P_{\mathcal{B}_V} m_{j-1}^s)\big)\Big) \\
&\leq V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H\big)P\big(\Psi^s(P_{\mathcal{B}_V} m_{j-1}^s) - \Psi(P_{\mathcal{B}_V} m_{j-1}^s)\big)\Big) \\
&+ V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H\big)(I-P)\big(\Psi^s(P_{\mathcal{B}_V} m_{j-1}^s) - \Psi(P_{\mathcal{B}_V} m_{j-1}^s)\big)\Big) \\
&\leq \|P - \mathscr{K}(\Sigma_j^s + Q)H\|_{V,V} V\Big(P\big(\Psi^s(P_{\mathcal{B}_V} m_{j-1}^s) - \Psi(P_{\mathcal{B}_V} m_{j-1}^s)\big)\Big) \\
&+ V\Big((I-P)\big(\Psi^s(P_{\mathcal{B}_V} m_{j-1}^s) - \Psi(P_{\mathcal{B}_V} m_{j-1}^s)\big)\Big) \\
&\overset{\text{(i)}}{\leq} \|P - \mathscr{K}(\Sigma_j^s + Q)H\|_{V,V}\sqrt{1+\beta}\kappa + \delta \\
&\leq \frac{\varepsilon^2\|R\|_{op} + L_s c_1(\varepsilon^2 + \delta^2)}{\gamma}(1+\beta)\kappa + \delta \leq c_5(\varepsilon + \delta),
\end{aligned}
\tag{3.48}
$$

where (i) uses Assumption 3.2.7 and $c_5 := 2(1+\beta)\kappa\frac{\|R\|_{op} + L_s c_1}{\gamma}(\varepsilon + \delta) + 1$. Similarly, we compute

$$
\begin{aligned}
&V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H\big)\big(\Psi(P_{\mathcal{B}_V} m_{j-1}^s) - \Psi(P_{\mathcal{B}_V} m_{j-1})\big)\Big) \\
&\leq V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H\big)P\big(\Psi(P_{\mathcal{B}_V} m_{j-1}^s) - \Psi(P_{\mathcal{B}_V} m_{j-1})\big)\Big) \\
&+ V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H\big)(I-P)\big(\Psi(P_{\mathcal{B}_V} m_{j-1}^s) - \Psi(P_{\mathcal{B}_V} m_{j-1})\big)\Big) \\
&\leq \|P - \mathscr{K}(\Sigma_j^s + Q)H\|_{V,V} V\Big(\Psi(P_{\mathcal{B}_V} m_{j-1}^s) - \Psi(P_{\mathcal{B}_V} m_{j-1})\Big) \\
&+ V\Big((I-P)\big(\Psi(P_{\mathcal{B}_V} m_{j-1}^s) - \Psi(P_{\mathcal{B}_V} m_{j-1})\big)\Big) \\
&\overset{\text{(i)}}{\leq} \|P - \mathscr{K}(\Sigma_j^s + Q)H\|_{V,V} L^{1/2} V\Big(m_{j-1}^s - m_{j-1}\Big) + \alpha^{1/2} V\big(m_{j-1}^s - m_{j-1}\big) \\
&\leq \Big(\frac{\sqrt{1+\beta}\varepsilon^2\|R\|_{op} + \sqrt{1+\beta}L_s c_2(\varepsilon^2 + \delta^2)}{\gamma} + \alpha^{1/2}\Big) V\big(m_{j-1}^s - m_{j-1}\big) \\
&= \Big(\frac{c_3}{\gamma} + \alpha^{1/2}\Big) V\big(m_{j-1}^s - m_{j-1}\big),
\end{aligned}
\tag{3.49}
$$

where (i) uses Assumption 3.2.1. Finally, we have that

$$\begin{aligned} V\Big(\big(I-\mathscr{K}(\Sigma_j^s+Q)H\big)(\Psi(P_{\mathcal{B}_V}m_{j-1})-\mu_j)\Big) &\le \Big(1+\frac{c_2}{\gamma}\Big)V\Big(\Psi(P_{\mathcal{B}_V}m_{j-1})-\mu_j\Big) \\ &\le \Big(1+\frac{c_2}{\gamma}\Big)L^{1/2}\Big(\mathrm{Tr}(C_{j-1})+\beta\mathrm{Tr}(HC_{j-1}H^*)\Big)^{1/2} \\ &\le \Big(1+\frac{c_2}{\gamma}\Big)L^{1/2}\sqrt{1+\beta}\Big(c_1\varepsilon^2+\mathrm{Tr}(R)\varepsilon^2\Big)^{1/2}=c_6\varepsilon, \end{aligned} \tag{3.50}$$

with $c_6 := \sqrt{1+\beta}(1+\frac{c_2}{\gamma})L^{1/2}(c_1+\mathrm{Tr}(R))^{1/2}$. Combining (3.47), (3.48), (3.49), and (3.50), we have

$$V\Big(\big(I-\mathscr{K}(\Sigma_j^s+Q)H\big)(\mu_j^s-\mu_j)\Big) \le \Big(\frac{c_3}{\gamma}+\alpha^{1/2}\Big)V\big(m_{j-1}^s-m_{j-1}\big)+c_7(\varepsilon+\delta), \tag{3.51}$$

where $c_7 := c_4+c_5+c_6$. We now bound the second term in (3.44).

$$\begin{aligned} &V\Big(\big(I-\mathscr{K}(\Sigma_j^s+Q)H-I+\mathscr{K}(\Sigma_j+Q)H\big)\big(\mu_j-H^*y_j\big)\Big) \\ &= V\Big(\big(P-\mathscr{K}(\Sigma_j^s+Q)H-P+\mathscr{K}(\Sigma_j^s+Q)H\big)\big(P\mu_j-H^*y_j\big)\Big) \\ &\le \Big(\|P-\mathscr{K}(\Sigma_j^s+Q)H\|_{V,V}+\|P-\mathscr{K}(\Sigma_j+Q)H\|_{V,V}\Big)V\big(P\mu_j-H^*y_j\big) \\ &\le \sqrt{1+\beta}\bigg(\frac{\varepsilon^2\|R\|_{op}+L_sc_1(\varepsilon^2+\delta^2)}{\gamma}+\frac{\varepsilon^2(\|R\|_{op}+L(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R)))}{\gamma}\bigg)V\big(P\mu_j-H^*y_j\big) \\ &\overset{\text{(i)}}{\le} c_8(\varepsilon+\delta)V\big(P\mu_j-H^*y_j\big) \\ &\le c_8(\varepsilon+\delta)\Big[V\Big(\mu_j-\Psi(P_{\mathcal{B}_V}m_{j-1})\Big)+V\Big(\Psi(P_{B_V}m_{j-1})-\Psi(u_{j-1})\Big)+V(H^*\eta)\Big] \\ &\overset{\text{(ii)}}{\le} c_8(\varepsilon+\delta)\Big[L^{1/2}\sqrt{1+\beta}\Big(c_1\varepsilon^2+\mathrm{Tr}(R)\varepsilon^2\Big)^{1/2} \\ &+V\Big(\Psi(P_{B_V}m_{j-1})-\Psi(u_{j-1})\Big)+V(H^*\eta)\Big], \end{aligned}$$

where in (i) we define $c_8 := \sqrt{1+\beta}\frac{2\varepsilon\|R\|_{op}+2L_sc_1(\varepsilon+\delta)+L(1+\frac{1+\alpha\beta}{1-\alpha}\mathrm{Tr}(R))}{\gamma}$, and (ii) uses the

same argument as in (3.50). Taking expectations, we have

$$\begin{aligned}
&\mathbb{E}V\Big(\big(I - \mathscr{K}(\Sigma_j^s + Q)H - I + \mathscr{K}(\Sigma_j + Q)H\big)\big(\mu_j - H^* y_j\big)\Big) \\
&\leq c_8(\varepsilon+\delta)\Big(L^{1/2}\sqrt{1+\beta}\Big(c_1\varepsilon^2 + \mathrm{Tr}(R)\varepsilon^2\Big)^{1/2} + L^{1/2}2\mathsf{C}_1\varepsilon + \varepsilon\sqrt{1+\beta}\mathrm{Tr}(R)^{1/2}\Big) \qquad (3.52)\\
&= c_9(\varepsilon+\delta)
\end{aligned}$$

with $c_9 := c_8\Big(L^{1/2}\sqrt{1+\beta}\Big(c_1\varepsilon^2 + \mathrm{Tr}(R)\varepsilon^2\Big)^{1/2} + L^{1/2}2\mathsf{C}_1\varepsilon + \varepsilon\sqrt{1+\beta}\mathrm{Tr}(R)^{1/2}\Big)$. Combining (3.51) and (3.52),

$$\mathbb{E}V\big(m_j^s - m_j\big) \leq \Big(\frac{c_3}{\gamma} + \alpha^{1/2}\Big)V\big(m_{j-1}^s - m_{j-1}\big) + (c_7 + c_9)(\varepsilon+\delta),$$

and we see that taking $\gamma \geq \frac{2c_3}{1-\alpha^{1/2}}$,

$$\mathbb{E}V\big(m_j^s - m_j\big) \leq \alpha_* V\big(m_{j-1}^s - m_{j-1}\big) + (c_7 + c_9)(\varepsilon+\delta)$$

with $\alpha_* = \frac{1+\alpha^{1/2}}{2}$. The discrete-time Gronwall lemma and equivalence of $V(\cdot)$ and $\|\cdot\|$ completes the proof. □

We now show that the means output by Algorithm 4 and those output by the idealized mean-field Algorithm 6 remain close in the long-time asymptotic. The proof of our second main result, Theorem 3.2.8, follows directly by combining Theorem 3.4.2 with Theorem 3.3.2 and the next result:

Theorem 3.4.3. *Suppose that Assumptions 3.2.1 and 3.2.7 hold and that $N \geq 6k$. Then, if the inflation parameter $\gamma > 0$ is sufficiently large, there is a constant C_4 independent of ε and δ such that*

$$\limsup_{j\to\infty} \mathbb{E}\|\widehat{m}_j^s - m_j^s\| \leq \mathsf{C}_4(\varepsilon+\delta). \qquad (3.53)$$

Proof. Assume that γ is taken to be at least as large as in Theorem 3.4.2. From Lemma 3.4.1 and Theorem 3.4.2 there exists a time j_* such that for $j \geq j_*$, $\mathrm{Tr}(C_j^s) \leq c_1(\varepsilon^2+\delta^2)$ and $\mathrm{Tr}(\widehat{C}_j^s) \leq c_1(\varepsilon^2+\delta^2)$ with $c_1 := 2\max\{\frac{1+\alpha_*\beta}{1-\alpha_*}\mathrm{Tr}(R), 4\frac{1+\alpha}{1-\alpha}\frac{1}{1-\alpha^*}\}$, and that $\mathbb{E}V(m_j^s - u_j) \leq 2(\mathsf{C}_1+\mathsf{C}_3)(\varepsilon+\delta)$. For $j > j^*$, we consider the difference

$$\begin{aligned}\widehat{m}_j^s - m_j^s &= \big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\widehat{\mu}_j^s + \mathscr{K}(\widehat{\Sigma}_j^s)y_j - \big(I - \mathscr{K}(\Sigma_j^s)\big)\mu_j^s - \mathscr{K}(\Sigma_j^s)y_j \\ &= \big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)(\widehat{\mu}_j^s - \mu_j^s) + \big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H - I + \mathscr{K}(\Sigma_j^s)H\big)(\mu_j^s - H^* y_j).\end{aligned}$$

Taking the V norm and expectation of both sides gives

$$\begin{aligned}\mathbb{E}V\big(\widehat{m}_j^s - m_j^s\big) \leq{}& \mathbb{E}V\Big(\big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\big(\widehat{\mu}_j^s - \mu_j^s\big)\Big) \\ &+ \mathbb{E}V\Big(\big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H - I + \mathscr{K}(\Sigma_j^s)H\big)\big(\mu_j^s - H^* y_j\big)\Big).\end{aligned} \tag{3.54}$$

The first term on the right-hand side of (3.54) can be bounded as

$$\begin{aligned}&\mathbb{E}V\Big(\big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\big(\widehat{\mu}_j^s - \mu_j^s\big)\Big) \\ &\leq \mathbb{E}V\Big(\big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(\frac{1}{N}\sum_{n=1}^N \Psi^s(P_{\mathcal{B}_V} u_{j-1}^{(n),s}) - \Psi^s(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)\Big)\Big) \\ &+ \mathbb{E}V\Big(\big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(\Psi^s(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s) - \Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)\Big)\Big) \\ &+ \mathbb{E}V\Big(\big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s) - \Psi(P_{\mathcal{B}_V} m_{j-1}^s)\Big)\Big) \\ &+ \mathbb{E}V\Big(\big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(\Psi(P_{\mathcal{B}_V} m_{j-1}^s) - \Psi^s(P_{\mathcal{B}_V} m_{j-1}^s)\Big)\Big) \\ &+ \mathbb{E}V\Big(\big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(\Psi^s(P_{\mathcal{B}_V} m_{j-1}^s) - \mu_j^s\Big)\Big) \\ &+ \mathbb{E}V\Big(\big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\frac{1}{N}\sum_{n=1}^N \xi^{(n)}\Big).\end{aligned}$$

We bound each of these terms in turn. By a similar argument to that in the proof of Theorem 3.3.2, we have

$$\|P - \mathscr{K}(\widehat{\Sigma}_j^s)H\|_{V,V} \leq \frac{\varepsilon^2 + \delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^s H^*)}\sqrt{1+\beta}\Big(\|R\|_{op} + L_s c_1\Big), \tag{3.55}$$

and

$$\|I - \mathscr{K}(\widehat{\Sigma}_j^s)H\|_{V,V} \leq 1 + \frac{\varepsilon^2 + \delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^s H^*)}\sqrt{1+\beta}\Big(\|R\|_{op} + L_s c_1\Big). \tag{3.56}$$

Since $\widehat{\Sigma}_j^s$ is conditionally independent of $\frac{1}{N}\sum_{n=1}^N \Psi^s(P_{\mathcal{B}_V} u_{j-1}^{(n),s}) - \Psi^s(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)$ given $\mathsf{U}_{\mathsf{j-1}}^{\mathsf{s}}$,

$$\begin{aligned}
&\mathbb{E}V\bigg(\big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(\frac{1}{N}\sum_{n=1}^N \Psi^s(P_{\mathcal{B}_V} u_{j-1}^{(n),s}) - \Psi^s(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)\Big)\bigg) \\
&\leq \mathbb{E}\bigg[\mathbb{E}^{\mathsf{U}_{\mathsf{j-1}}^{\mathsf{s}}}\Big[1 + \frac{\varepsilon^2 + \delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^s H^*)}\sqrt{1+\beta}\Big(\|R\|_{op} + L_s c_1\Big)\Big] \\
&\cdot \mathbb{E}^{\mathsf{U}_{\mathsf{j-1}}^{\mathsf{s}}}\Big[V\Big(\frac{1}{N}\sum_{n=1}^N \Psi^s(P_{\mathcal{B}_V} u_{j-1}^{(n),s}) - \Psi^s(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)\Big)\Big]\bigg] \\
&\leq \mathbb{E}\bigg[\mathbb{E}^{\mathsf{U}_{\mathsf{j-1}}^{\mathsf{s}}}\Big[1 + \frac{\varepsilon^2 + \delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^s H^*)}\sqrt{1+\beta}\Big(\|R\|_{op} + L_s c_1\Big)\Big] \\
&\cdot \mathbb{E}^{\mathsf{U}_{\mathsf{j-1}}^{\mathsf{s}}}\Big[L_s^{1/2}\Big(\mathrm{Tr}(\widehat{C}_{j-1}^s) + \beta\mathrm{Tr}(H\widehat{C}_{j-1}^s H^*)\Big)^{1/2}\Big]\bigg] \\
&\leq \mathbb{E}\bigg[\mathbb{E}^{\mathsf{U}_{\mathsf{j-1}}^{\mathsf{s}}}\Big[1 + \frac{\varepsilon^2 + \delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^s H^*)}\sqrt{1+\beta}\Big(\|R\|_{op} + L_s c_1\Big)\Big] L_s^{1/2}\sqrt{(1+\beta)c_1}(\varepsilon + \delta)\bigg].
\end{aligned} \tag{3.57}$$

Choosing γ to guarantee that $\frac{10NL_s c_1(\varepsilon^2+\delta^2)}{\gamma k} \leq 1$, Lemma 3.3.6 gives us that

$$\begin{aligned}
&\mathbb{E}^{\mathsf{U}_{\mathsf{j-1}}}\Big[1 + \frac{\varepsilon^2 + \delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^s H^*)}\sqrt{1+\beta}\Big(\|R\|_{op} + L_s c_1\Big)\Big] \\
&\qquad \leq \bigg(1 + \frac{2C'(\varepsilon^2 + \delta^2)}{\gamma}\sqrt{1+\beta}\Big(\|R\|_{op} + L_s c_1\Big)\bigg),
\end{aligned}$$

and consequently

$$\begin{aligned}
&\mathbb{E}V\Big(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(\frac{1}{N}\sum_{n=1}^{N}\Psi^s(P_{\mathcal{B}_V}u_{j-1}^{(n),s})-\Psi^s(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)\Big)\Big)\\
&\leq\Big(1+\frac{2C'(\varepsilon^2+\delta^2)}{\gamma}\sqrt{1+\beta}\Big(\|R\|_{op}+L_sc_2\Big)\Big)L_s^{1/2}\sqrt{(1+\beta)c_1}(\varepsilon+\delta)=c_2(\varepsilon+\delta),
\end{aligned}\tag{3.58}$$

where $c_2:=\Big(1+\frac{2C'(\varepsilon^2+\delta^2)}{\gamma}\sqrt{1+\beta}\Big(\|R\|_{op}+L_sc_1\big)\Big)L_s^{1/2}\sqrt{(1+\beta)c_1}$. Similarly, for the second term we have

$$\begin{aligned}
&\mathbb{E}V\Big(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(\Psi^s(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)-\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)\Big)\Big)\\
&\leq\mathbb{E}V\Big(\big(P-\mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(P\Psi^s(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)-P\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)\Big)\Big)\\
&+\mathbb{E}V\Big((I-P)\Psi^s(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)-(I-P)\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)\Big)\\
&\leq\mathbb{E}\Big[\mathbb{E}^{\mathsf{U}_{j-1}^s}\Big[\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}\sqrt{1+\beta}\Big(\|R\|_{op}+L_sc_1\Big)\Big]\\
&\cdot\mathbb{E}^{\mathsf{U}_{j-1}^s}\Big[V\Big(\Psi^s(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)-\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)\Big)\Big]\Big]\\
&+\mathbb{E}V\Big((I-P)\Psi^s(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)-(I-P)\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s)\Big)\\
&\leq\mathbb{E}\Big[\mathbb{E}^{\mathsf{U}_{j-1}^s}\Big[\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}\sqrt{1+\beta}\Big(\|R\|_{op}+L_sc_1\Big)\Big]\sqrt{1+\beta}\kappa\Big]+\sqrt{1+\beta}\delta\\
&\leq\frac{2(1+\beta)\kappa C'(\varepsilon^2+\delta^2)}{\gamma}\Big(\|R\|_{op}+L_sc_1\Big)+\sqrt{1+\beta}\delta\leq c_3(\varepsilon+\delta),
\end{aligned}\tag{3.59}$$

where $c_3 = \frac{2(1+\beta)\kappa C'(\varepsilon+\delta)}{\gamma}(\|R\|_{op} + L_s c_1) + \sqrt{1+\beta}$. For the third term,

$$
\begin{aligned}
&\mathbb{E}V\bigg(\big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s) - \Psi(P_{\mathcal{B}_V}m_{j-1}^s)\Big)\bigg)\\
&\le \mathbb{E}V\bigg(\big(P - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s) - \Psi(P_{\mathcal{B}_V}m_{j-1}^s)\Big)\bigg)\\
&+ \mathbb{E}V\bigg((I-P)\Big(\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s) - \Psi(P_{\mathcal{B}_V}m_{j-1}^s)\Big)\bigg)\\
&\le \mathbb{E}\bigg[\mathbb{E}^{\mathsf{U}_{j-1}^s}\Big[\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^s H^*)}\sqrt{1+\beta}\Big(\|R\|_{op} + L_s c_1\Big)\Big]\\
&\cdot \mathbb{E}^{\mathsf{U}_{j-1}^s}\Big[\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s) - \Psi(P_{\mathcal{B}_V}m_{j-1}^s)\Big]\bigg]\\
&+ \alpha^{1/2}\mathbb{E}V\big(\widehat{m}_{j-1}^s - m_{j-1}^s\big)\\
&\le \frac{2C'(\varepsilon^2+\delta^2)}{\gamma}\sqrt{1+\beta}\Big(\|R\|_{op} + L_s c_1\Big)L^{1/2}\mathbb{E}V\big(\widehat{m}_{j-1}^s - m_{j-1}^s\big)\\
&+ \alpha^{1/2}\mathbb{E}V\big(\widehat{m}_{j-1}^s - m_{j-1}^s\big)\\
&= \Big(\frac{c_4}{\gamma} + \alpha^{1/2}\Big)\mathbb{E}V\big(\widehat{m}_{j-1}^s - m_{j-1}^s\big),
\end{aligned}
$$

with $c_4 = 2C'(\varepsilon^2+\delta^2)\sqrt{1+\beta}\Big(\|R\|_{op} + L_s c_1\Big)L^{1/2}$. Taking $\gamma \ge \frac{2c_4}{1-\alpha^{1/2}}$ yields

$$
\mathbb{E}V\bigg(\big(I - \mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(\Psi(P_{\mathcal{B}_V}\widehat{m}_{j-1}^s) - \Psi(P_{\mathcal{B}_V}m_{j-1}^s)\Big)\bigg) \le \alpha_*\mathbb{E}V\big(\widehat{m}_{j-1}^s - m_{j-1}^s\big), \quad (3.60)
$$

with $\alpha_* := \frac{1+\alpha^{1/2}}{2} \in (0,1)$. For the fourth term, we have

$$
\begin{aligned}
&\mathbb{E}V\Big(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(\Psi(P_{\mathcal{B}_V}m_{j-1}^s)-\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)\Big)\Big)\\
&\le \mathbb{E}V\Big(\big(P-\mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(P\Psi(P_{\mathcal{B}_V}m_{j-1}^s)-P\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)\Big)\Big)\\
&+\mathbb{E}V\Big((I-P)\Psi(P_{\mathcal{B}_V}m_{j-1}^s)-(I-P)\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)\Big)\\
&\le \mathbb{E}\Big[\mathbb{E}^{\mathsf{U}_{j-1}^s}\Big[\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}\sqrt{1+\beta}\Big(\|R\|_{op}+L_sc_1\Big)\Big]\\
&\cdot\mathbb{E}^{\mathsf{U}_{j-1}^s}\Big[V\Big(\Psi(P_{\mathcal{B}_V}m_{j-1}^s)-\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)\Big)\Big]\Big]\\
&+\mathbb{E}V\Big((I-P)\Psi(P_{\mathcal{B}_V}m_{j-1}^s)-(I-P)\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)\Big)\\
&\le \mathbb{E}\Big[\mathbb{E}^{\mathsf{U}_{j-1}^s}\Big[\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}\sqrt{1+\beta}\Big(\|R\|_{op}+L_sc_1\Big)\Big]\sqrt{1+\beta}\kappa\Big]+\sqrt{1+\beta}\delta\\
&\le \frac{2(1+\beta)\kappa C'(\varepsilon^2+\delta^2)}{\gamma}\Big(\|R\|_{op}+L_sc_1\Big)+\sqrt{1+\beta}\delta\le c_3(\varepsilon+\delta).
\end{aligned}
\tag{3.61}
$$

Then, we bound the fifth term as

$$
\begin{aligned}
&\mathbb{E}V\Big(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H\big)\Big(\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)-\mu_j^s\Big)\Big)\\
&\le \mathbb{E}\Big[\mathbb{E}^{\mathsf{U}_{j-1}^s}\Big[1+\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}\sqrt{1+\beta}\Big(\|R\|_{op}+L_sc_1\Big)\Big]\mathbb{E}^{\mathsf{U}_{j-1}^s}\Big[V\Big(\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)-\mu_j^s\Big)\Big]\Big]\\
&\le \Big(1+\frac{2C'(\varepsilon^2+\delta^2)}{\gamma}\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)\Big)\mathbb{E}\Big[V\Big(\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)-\mu_j^s\Big)\Big]\\
&\le \Big(1+\frac{2C'(\varepsilon^2+\delta^2)}{\gamma}\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)\Big)L_s^{1/2}\mathbb{E}\Big[\big(\mathrm{Tr}(C_{j-1}^s)+\beta\mathrm{Tr}(HC_{j-1}^sH^*)\big)^{1/2}\Big]\\
&\le \Big(1+\frac{2C'(\varepsilon^2+\delta^2)}{\gamma}\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)\Big)L_s^{1/2}\sqrt{(1+\beta)c_1}\varepsilon\le c_5(\varepsilon+\delta),
\end{aligned}
\tag{3.62}
$$

where $c_5=\Big(1+\frac{2C'(\varepsilon^2+\delta^2)}{\gamma}\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)\Big)L_s^{1/2}\sqrt{(1+\beta)c_1}$. For the last term, we

have that

$$\mathbb{E}V\bigg(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H\big)\frac{1}{N}\sum_{n=1}^{N}\xi^{(n)}\bigg)=\mathbb{E}V\bigg(\big(P-\mathscr{K}(\widehat{\Sigma}_j^s)H\big)\frac{1}{N}\sum_{n=1}^{N}\xi^{(n)}\bigg)$$
$$\leq\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)\mathbb{E}\bigg[\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}V\bigg(\frac{1}{N}\sum_{n=1}^{N}\xi^{(n)}\bigg)\bigg].$$

Applying the Cauchy-Schwarz inequality and Lemma 3.3.6, it follows that

$$\begin{aligned}&\mathbb{E}V\bigg(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H\big)\frac{1}{N}\sum_{n=1}^{N}\xi^{(n)}\bigg)\\&\leq\sqrt{1+\beta}\big(\|R\|_{op}+L_sc_1\big)(\varepsilon^2+\delta^2)\mathbb{E}\bigg[\frac{1}{\lambda_{\min}(\widehat{H\Sigma}_j^sH^*)^2}\bigg]^{1/2}\mathbb{E}\bigg[V^2\big(\frac{1}{N}\sum_{n=1}^{N}\xi_j^{(n)}\big)\bigg]^{1/2}\\&\leq\sqrt{1+\beta}\big(\|R\|_{op}+L_sc_1\big)(\varepsilon^2+\delta^2)\frac{2C'}{\gamma}\mathbb{E}\bigg[V^2\big(\frac{1}{N}\sum_{n=1}^{N}\xi_j^{(n)}\big)\bigg]^{1/2}.\end{aligned}$$

Using the fact that $\mathbb{E}V^2\big(\frac{1}{N}\sum_{n=1}^{N}\xi_j^{(n)}\big)=(1+\beta)\mathrm{Tr}(\frac{\gamma}{N}I_k)$ and the assumption that $N\geq 6k$, we have that

$$\begin{aligned}\mathbb{E}V\bigg(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H\big)\frac{1}{N}\sum_{n=1}^{N}\xi^{(n)}\bigg)&\leq\sqrt{1+\beta}\big(\|R\|_{op}+L_sc_1\big)(\varepsilon^2+\delta^2)\frac{2C'}{\gamma}\big(\frac{(1+\beta)\gamma k}{N}\big)^{1/2}\\&\leq(1+\beta)\big(\|R\|_{op}+L_sc_1\big)(\varepsilon^2+\delta^2)\frac{2C'}{\gamma^{1/2}}+\leq c_6(\varepsilon+\delta),\end{aligned}\tag{3.63}$$

where we define $c_6:=(1+\beta)\big(\|R\|_{op}+L_sc_1\big)(\varepsilon+\delta)\frac{2C'}{\gamma^{1/2}}$. It remains to bound the second

term in the right-hand side of (3.54):

$$\begin{aligned}
&\mathbb{E}V\Big(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H-I+\mathscr{K}(\Sigma_j^s)H\big)\big(\mu_j^s-H^*y_j\big)\Big)\\
&\le \mathbb{E}V\Big(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H-I+\mathscr{K}(\Sigma_j^s)H\big)\big(\mu_j^s-\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)\big)\Big)\\
&+\mathbb{E}V\Big(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H-I+\mathscr{K}(\Sigma_j^s)H\big)\big(\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)-H^*y_j\big)\Big).
\end{aligned}$$

We now bound each of these terms. Choosing $\gamma \ge \frac{10NL_sc_2(\varepsilon^2+\delta^2)}{k}$, we have

$$\begin{aligned}
&\mathbb{E}V\Big(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H-I+\mathscr{K}(\Sigma_j^s)H\big)\big(\mu_j^s-\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)\big)\Big)\\
&=\mathbb{E}V\Big(\big(P-\mathscr{K}(\widehat{\Sigma}_j^s)H-P+\mathscr{K}(\Sigma_j^s)H\big)\big(\mu_j^s-\Psi^s(m_{j-1}^s)\big)\Big)\\
&\le \mathbb{E}\big[\big(\|P-\mathscr{K}(\widehat{\Sigma}_j^s)H\|_{V,V}+\|P-(\Sigma_j^s)H\|_{V,V}\big)V\big(\mu_j^s-\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)\big)\big]\\
&\le \mathbb{E}\big[\big(\|P-\mathscr{K}(\widehat{\Sigma}_j^s)H\|_{V,V}+\|P-(\Sigma_j^s)H\|_{V,V}\big)L_s^{1/2}\big(\mathrm{Tr}(C_{j-1}^s)+\beta\mathrm{Tr}(HC_{j-1}^sH^*)\big)^{1/2}\big]\\
&\le \mathbb{E}\big[\big(\|P-\mathscr{K}(\widehat{\Sigma}_j^s)H\|_{V,V}+\|P-(\Sigma_j^s)H\|_{V,V}\big)\big]L_s^{1/2}\sqrt{(1+\beta)c_1}(\varepsilon+\delta)\\
&\le \sqrt{1+\beta}(\|R\|_{op}+L_sc_1)\Big(\mathbb{E}\big[\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}\big]+\frac{\varepsilon^2+\delta^2}{\gamma}\Big)L_s^{1/2}\sqrt{(1+\beta)c_1}(\varepsilon+\delta)\\
&\overset{\text{(i)}}{\le} \sqrt{1+\beta}(\|R\|_{op}+L_sc_1)(2C'+1)\frac{\varepsilon^2+\delta^2}{\gamma}L_s^{1/2}\sqrt{(1+\beta)c_1}(\varepsilon+\delta)\le c_7(\varepsilon+\delta),
\end{aligned} \tag{3.64}$$

where (i) uses Lemma 3.3.6 and $c_7 := (1+\beta)(\|R\|_{op}+L_sc_1)(2C'+1)\frac{\varepsilon^2+\delta^2}{\gamma}L_s^{1/2}\sqrt{2c_1}$. To

bound the second term, the triangle inequality and 3.45 yield

$$
\begin{aligned}
&\mathbb{E}V\Big(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H-I+\mathscr{K}(\Sigma_j^s)H\big)\big(\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)-H^*y_j\big)\Big)\\
&=\mathbb{E}V\Big(\big(P-\mathscr{K}(\widehat{\Sigma}_j^s)H-P+\mathscr{K}(\Sigma_j^s)H\big)\big(P\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)-P\Psi(u_{j-1})+H^*\eta_j\big)\Big)\\
&\leq\mathbb{E}\big[\big(\|P-\mathscr{K}(\widehat{\Sigma}_j^s)\|_{V,V}+\|P-\mathscr{K}(\Sigma_j^s)\|_{V,V}\big)V\big(P\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)-P\Psi(u_{j-1})+H^*\eta_j\big)\big]\\
&\leq\mathbb{E}\big[\big(\|P-\mathscr{K}(\widehat{\Sigma}_j^s)\|_{V,V}+\|P-\mathscr{K}(\Sigma_j^s)\|_{V,V}\big)V\big(P\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)-P\Psi(P_{\mathcal{B}_V}m_{j-1}^s)\big)\big]\\
&+\mathbb{E}\big[\big(\|P-\mathscr{K}(\widehat{\Sigma}_j^s)\|_{V,V}+\|P-\mathscr{K}(\Sigma_j^s)\|_{V,V}\big)V\big(P\Psi(P_{\mathcal{B}_V}m_{j-1}^s)-P\Psi(u_{j-1})+H^*\eta_j\big)\big]\\
&\leq\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)\mathbb{E}\big[\big(\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}+\frac{\varepsilon^2+\delta^2}{\gamma}\big)\\
&\cdot V\big(P\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)-P\Psi(P_{\mathcal{B}_V}m_{j-1}^s)\big)\big]\\
&+\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)\mathbb{E}\big[\big(\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}+\frac{\varepsilon^2+\delta^2}{\gamma}\big)\\
&\cdot V\big(P\Psi(P_{\mathcal{B}_V}m_{j-1}^s)-P\Psi(u_{j-1})+H^*\eta_j\big)\big]\\
&\leq\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)\mathbb{E}\big[\big(\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}+\frac{\varepsilon^2+\delta^2}{\gamma}\big)\kappa\big]\\
&+\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)\mathbb{E}\big[\big(\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}+\frac{\varepsilon^2+\delta^2}{\gamma}\big)\\
&\cdot V\big(\Psi(P_{\mathcal{B}_V}m_{j-1}^s)-\Psi(u_{j-1})+H^*\eta_j\big)\big],
\end{aligned}
$$

where the last inequality uses Assumption 3.2.7. Next, by conditional independence and

Lemma 3.3.6,

$$
\begin{aligned}
&\mathbb{E}V\Big(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H-I+\mathscr{K}(\Sigma_j^s)H\big)\big(\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)-H^*y_j\big)\Big)\\
&\leq \sqrt{1+\beta}(\|R\|_{op}+L_sc_1)\big(\mathbb{E}\big[\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}\big]+\frac{\varepsilon^2+\delta^2}{\gamma}\big)\kappa\\
&+\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)\mathbb{E}\Big[\Big(\mathbb{E}^{\mathsf{U}_{j-1}^s}\big[\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}\big]+\frac{\varepsilon^2+\delta^2}{\gamma}\Big)\\
&\cdot\mathbb{E}^{\mathsf{U}_{j-1}^s}\big[V\big(\Psi(P_{\mathcal{B}_V}m_{j-1}^s)-\Psi(u_{j-1})\big)\big]\Big]\\
&+\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)\Big(\mathbb{E}\big[\frac{\varepsilon^2+\delta^2}{\lambda_{\min}(H\widehat{\Sigma}_j^sH^*)}\big]+\frac{\varepsilon^2+\delta^2}{\gamma}\Big)\mathbb{E}V\big(H^*\eta_j\big)\\
&\leq \sqrt{1+\beta}(\|R\|_{op}+L_sc_1)(2C'+1)\frac{\varepsilon^2+\delta^2}{\gamma}\sqrt{1+\beta}\kappa\\
&+\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)(2C'+1)\frac{\varepsilon^2+\delta^2}{\gamma}\mathbb{E}\big[V\big(m_{j-1}^s-u_{j-1}\big)\big]\\
&+\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)(2C'+1)\frac{\varepsilon^2+\delta^2}{\gamma}\sqrt{1+\beta}\mathrm{Tr}(R)^{1/2}\varepsilon
\end{aligned}
$$

where the last inequality also uses the fact that Ψ is Lipschitz on $\mathcal{B}_V$ and that $\mathbb{E}V(H^*\eta_j)\leq\sqrt{1+\beta}\mathrm{Tr}(R)^{1/2}\varepsilon$. Since $j>j^*$, we have that $\mathbb{E}V(m_{j-1}^s-u_{j-1})\leq 2(\mathsf{C}_1+\mathsf{C}_3)(\varepsilon+\delta)$, and thus

$$
\begin{aligned}
&\mathbb{E}V\Big(\big(I-\mathscr{K}(\widehat{\Sigma}_j^s)H-I+\mathscr{K}(\Sigma_j^s)H\big)\big(\Psi^s(P_{\mathcal{B}_V}m_{j-1}^s)-H^*y_j\big)\Big)\\
&\leq \sqrt{1+\beta}(\|R\|_{op}+L_sc_1)(2C'+1)\big(\sqrt{1+\beta}\kappa\frac{\varepsilon+\delta}{\gamma}+2\frac{(\varepsilon+\delta)^2}{\gamma}(\mathsf{C}_1+\mathsf{C}_3)\\
&+\frac{\varepsilon+\delta}{\gamma}\sqrt{1+\beta}\mathrm{Tr}(R)^{1/2}\varepsilon\big)(\varepsilon+\delta)\\
&=c_8(\varepsilon+\delta),
\end{aligned}
\tag{3.65}
$$

where we define $c_8=\sqrt{1+\beta}(\|R\|_{op}+L_sc_1)(2C'+1)\big(\sqrt{1+\beta}\kappa\frac{\varepsilon+\delta}{\gamma}+2\frac{(\varepsilon+\delta)^2}{\gamma}(\mathsf{C}_1+\mathsf{C}_3)+\frac{\varepsilon+\delta}{\gamma}\sqrt{1+\beta}\mathrm{Tr}(R)^{1/2}\varepsilon\big)$. Combining (3.58), (3.59), (3.60), (3.61), (3.62), (3.63), (3.64), and

(3.65) we have

$$\mathbb{E}V\left(\widehat{m}_j^s - m_j^s\right) \leq \alpha_* \mathbb{E}V\left(\widehat{m}_{j-1}^s - m_{j-1}^s\right) + (c_2 + 2c_3 + c_5 + c_6 + c_7 + c_8)(\varepsilon + \delta). \quad (3.66)$$

The discrete Gronwall inequality and the equivalence of $V(\cdot)$ and $\|\cdot\|$ yield the desired result. □

3.5 Numerical experiments

In this section, we numerically verify our theoretical results with the Lorenz-96 system given in Example 3.2.5. Throughout our experiments, we take the state dimension to be $d = 60$ and the observation time-step to be $\Delta t = 0.1$. As described in Example 3.2.5, we take our observation operator to be the identity matrix with every third row removed, and thus $k = 40$.

To illustrate the filter accuracy result given in Theorem 3.2.2, we compare the performance of Algorithm 3 for decreasing noise levels. For 50 Monte Carlo trials, we initialize the true signal at a random Gaussian vector $u_0 \sim \mathcal{N}(0, 10I)$ and evolve the true state until time $T = 25$. We generate observations y_j at noise levels $\varepsilon = (1, 10^{-1}, 10^{-2}, 10^{-3})$. Taking $N = 50$ ensemble members, we initialize the ensemble as $\{u_0^{(n)}\}_{n=1}^N \sim \mathcal{N}(0, I)$. We then run the Ensemble Transform Kalman Filter with variance inflation parameter $\gamma = 1$. In Figure 3.1, we plot the average error between the analysis mean and the true signal, $\|\widehat{m}_j - u_j\|$, over the 50 Monte Carlo trials. As predicted by our theory, the errors are proportional to the noise level, ϵ.

To illustrate Theorem 3.2.8, we compare the performance of Algorithm 4 with surrogate models of increasing fidelity. To generate training data for our surrogate models, we sample 10^3 standard normal random vectors and integrate the Lorenz-96 system until time $T = 100$. This generates 10^6 input-output pairs of the dynamics map that we take to be our training

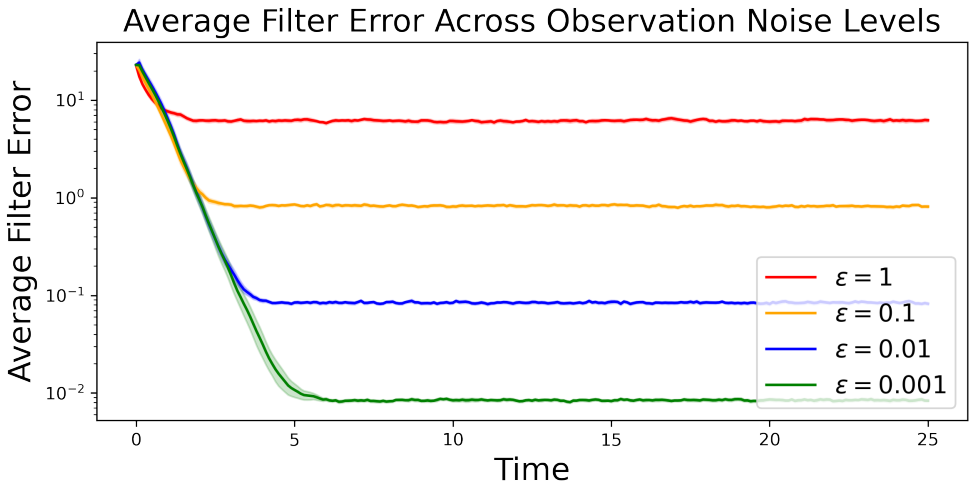
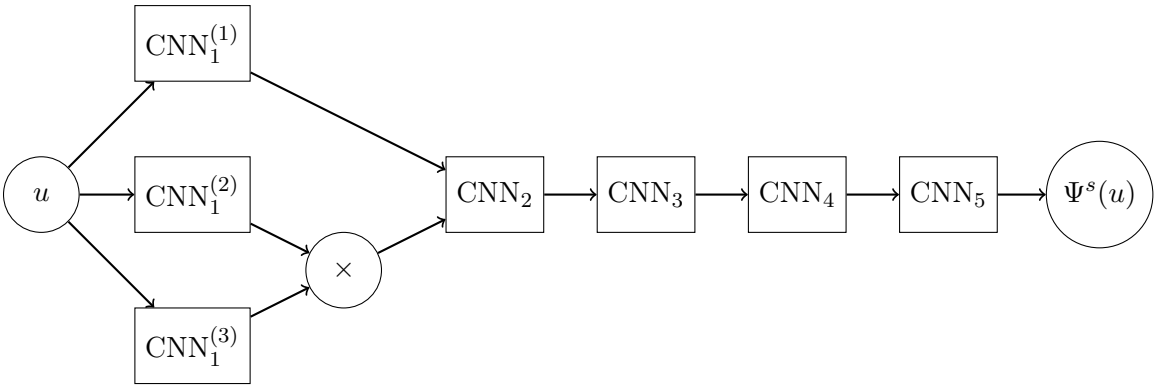


Figure 3.1: Average filter error over 50 Monte Carlo trials for decreasing observation noise levels, plotted with two standard errors (shaded).

data set. We consider a convolutional neural network with architecture similar to that used in [57] with 30,003 total trainable parameters. The structure of the neural network is detailed in Figure 3.2. The neural network weights are initialized randomly using Pytorch's default initialization. To train the neural network, we use the mean squared error loss function with the Adam [216] optimizer with a learning rate of 0.0003, momentum 0.1, and a batch size of 2000. We train three neural networks using different amounts of training data, meant to represent a low-fidelity, medium-fidelity, and high-fidelity machine model. The size of the training set and number of epochs for each model are summarized in Table 3.1. To estimate the model error in the unobserved components of the system, δ, defined in Assumption 3.2.7, we select 10^3 points uniformly at random from the attractor by numerically integrating a long run of the dynamics and compute the maximum error of each of the models over the 10^3 points. The estimated model errors are reported in Table 3.1.

In Figure 3.3, we visualize the performance of the Ensemble Transform Kalman filter with surrogate models, implemented with $N = 50$ ensemble members and variance inflation parameter $\gamma = 10$. For a single realization of the true signal, we plot the evolution of the true state until time $T = 25$, along with the observed data, the surrogate model filter means,



(a) Network structure.

	#in	#out	kernel size
CNN_1	1	72(24× 3)	5, circular
CNN_2	48(24× 2)	37	5, circular
CNN_3	37	37	5, circular
CNN_4	37	37	5, circular
CNN_5	37	1	1

(b) Network details.

Figure 3.2: The structure of Ψ^s. The output channels of CNN_1 are divided into three groups of equal length, $\mathrm{CNN}_1^{(1)}$, $\mathrm{CNN}_1^{(2)}$, and $\mathrm{CNN}_1^{(3)}$. The input channels to CNN_2 are a concatenation of $\mathrm{CNN}_1^{(1)}$ and $(\mathrm{CNN}_1^{(1)} \times \mathrm{CNN}_1^{(2)})$, where the multiplication is point-wise.

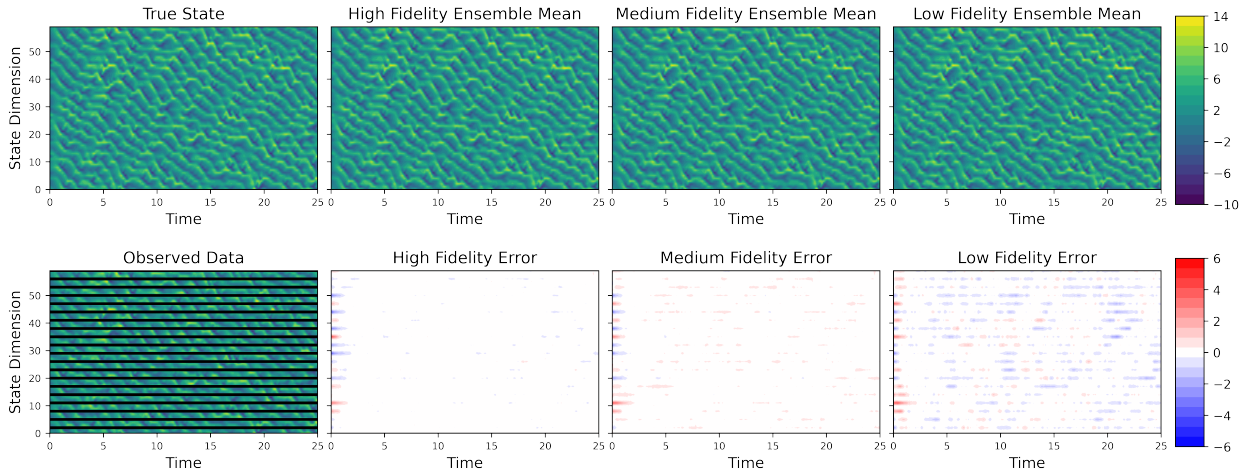


Figure 3.3: Visualization of the true state trajectory over time, the ensemble means, the observed data, and the difference between the true state and the ensemble means.

and the difference between the true state and the filter estimates. Visually, the ensemble means output by each of the filters appear to track the true signal quite well. The error plots reveal that, as predicted by our theory, filtering with the low fidelity surrogate model incurs the largest error, and the high fidelity error incurs the smallest error.

Next, for 50 Monte Carlo trials, we initialize the true signal at a random Gaussian vector $u_0 \sim \mathcal{N}(0, 10I)$ and we initialize the ensemble as $\{u_0^{(n)}\}_{n=1}^N \sim \mathcal{N}(0, I)$. We then run the Ensemble Transform Kalman Filter with $N = 50$ ensemble members and variance inflation parameter $\gamma = 10$. In the first plot in Figure 3.4, we display the average error in the analysis forecast mean $\widehat{m}_j$ from the true state u_j. We observe that the average filter error is proportional to the model error in the unobserved components, as predicted by our theory. The filter errors averaged over times $T = 10$ to $T = 25$ are reported in Table 3.1.

We then consider the forecast error from the true dynamics incurred by applying the surrogate model sequentially from an exactly known initial condition to emulate the dynamics over moderate time-scales. In the second plot in Figure 3.4, we display the average error over 50 Monte Carlo trials with initial conditions drawn from $u_0 \sim \mathcal{N}(0, 10I)$. By time $T = 4$, the error in all surrogate model predictions is large. We see that despite the surrogate models

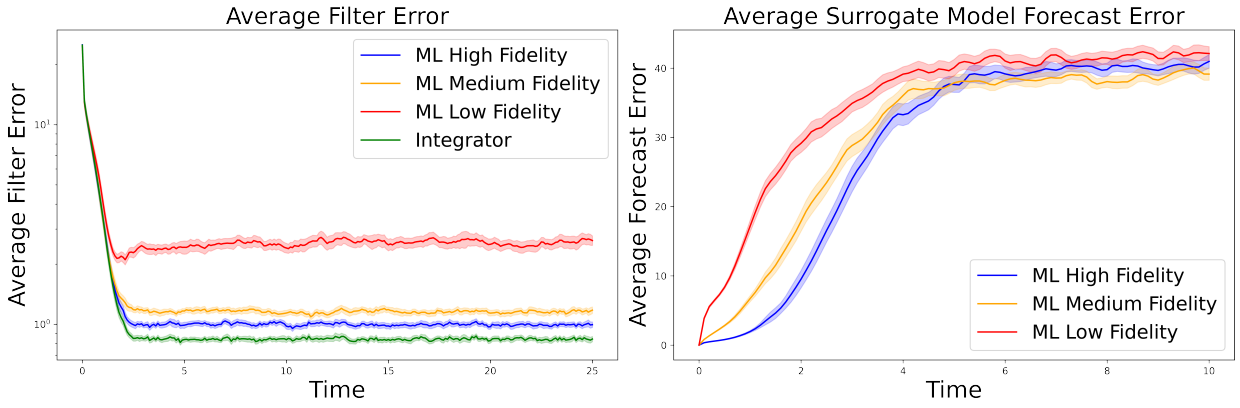


Figure 3.4: The left plot displays the average filter error over 50 Monte Carlo trials plotted with two standard errors. The right plot displays the average surrogate model forecast error from an exactly known initial condition over 50 Monte Carlo trials plotted with two standard errors. This figure shows that long-time filter accuracy is possible with surrogate models that only give accurate short-term forecasts.

being unable to accurately forecast the dynamics beyond short timescales, they can still be used within the ensemble Kalman filter to accurately track the signal given sufficiently informative observations.

Fidelity	Training Samples	Epochs	Model Error	Filter Error
Low	10^3	20	$\delta \approx 2.16$	2.59
Medium	10^4	50	$\delta \approx 1.02$	1.16
High	10^6	200	$\delta \approx 0.35$	0.99

Table 3.1: Summary of the training details, estimated model error, and average filter error for each model.

3.6 Conclusions

This paper established long-time accuracy of ensemble Kalman filters, introducing conditions on the dynamics and the observations under which the estimation error remains small in the long-time horizon. We have shown that these conditions can be verified for many partially-

observed dynamical systems, such as the Lorenz models and the 2-dimensional Navier-Stokes equations. In addition, we have rigorously established and empirically demonstrated that ensemble Kalman filters can provide long-time accurate state estimation even when employing machine-learned surrogate models that only give accurate short-term forecasts.

There are several natural research directions that stem from this work. First, our analysis focuses on linear observations; the nonlinear case could be studied by working on an extended state-space, as in [14]. Second, it would be interesting to extend the theory beyond the case of autonomous dynamical systems and fixed observation operators considered here. Third, our work does not account for covariance localization, which can improve the empirical performance of the algorithm and reduce the required ensemble size. Lastly, there are numerous opportunities to leverage the ideas in this paper to co-learn the state, the dynamics, and the filtering algorithms in data assimilation tasks across many application domains.

CHAPTER 4

HIERARCHICAL BAYESIAN INVERSE PROBLEMS: A HIGH-DIMENSIONAL STATISTICS VIEWPOINT

4.1 Introduction

The problem of estimating a high-dimensional parameter from a few noisy and indirect measurements is of central importance in science and engineering [137, 82, 297] and arises, for instance, in genomics [40], spatial statistics [163], imaging [129, 248], and data assimilation [231]. High-dimensionality —which in many inverse problems results from the unknown parameter being a fine discretization of a function [66, 315]— presents both statistical and computational challenges. From a statistical perspective, consistent estimation of a d-dimensional unknown from $n \ll d$ measurements is only possible when the problem exhibits some underlying low-dimensional structure [65, 342]. From a computational perspective, the scalability of optimization and sampling algorithms hinges on adequately exploiting such structure [120, 300, 159, 7].

In the Bayesian approach to inverse problems, low-dimensional structure can be imposed on the reconstruction through the choice of prior distribution [297, 209]. Hierarchical modelling [164] provides an expressive approach to prior design. The idea is to model the parameters in the prior as random variables distributed according to a hyperprior, thus specifying a joint prior distribution for the unknown of interest and the prior parameters. Conditioning on the observed data yields the posterior distribution, which represents the Bayesian solution to the inverse problem [315]. We consider a popular family of sparsity-promoting hierarchical models with a mean-zero Gaussian prior and gamma hyperpriors on the prior variances. This model was originally proposed for inverse problems in medical imaging, such as cerebral source localization [74] and positron emission tomography [25]. Heuristically, sparsity is promoted in the following way: the gamma hyperprior imposes the

belief that most of the variances are small except for a few outliers; consequently, the entries of prior draws will be nearly zero except for those with outlier variances. Beyond the original motivation in imaging, similar hierarchical models have been applied to deconvolution problems [79], dictionary learning [284, 345, 48], multiple measurement inverse problems [171], adaptive meshing [47], and identification of dynamical systems [9].

Hierarchical Bayesian models with gamma hyperpriors have been extensively studied from a computational viewpoint. For linear inverse problems, the posterior mode —known as the *maximum a posteriori* (MAP) estimator— can be efficiently computed using an iterative alternating scheme that converges linearly [75]. Path-following methods for MAP estimation were developed in [308]. For nonlinear inverse problems, the MAP estimator can be approximated using ensemble Kalman methods [215]. Recently, algorithms for sampling [83] and variational inference [9] have been proposed. While this body of work compellingly showcases the computational benefits afforded by these hierarchical models, reconstruction error bounds are not yet available.

In this paper, we present a new statistical analysis of a family of hierarchical Bayesian models with gamma hyperpriors, obtaining the first error bounds on the reconstruction error of MAP estimators. Our bounds are *non-asymptotic*: they are applicable for any dimension and number of measurements. We view MAP estimators as particular instances of regularized M-estimators, that is, as being defined by minimizing an objective comprising a loss term and a regularization term. M-estimators play a central role in the statistical literature [336]. A canonical example is the Lasso estimator [325], which minimizes the sum of a least-squares loss and an ℓ_1 penalty that enforces sparsity in the reconstruction of the unknown [41]. Other widely used M-estimators include the group Lasso [23], graphical Lasso [205], sparse generalized linear models [337], and nuclear norm penalization for matrix completion [86, 287]. In the variational inverse problems literature, M-estimators often arise from regularization methods [69, 29, 59]. The similarities between regularized M-estimation

and MAP estimation are immediate, and in fact many M-estimators can be derived as MAP estimators of certain posterior distributions. For instance, the Lasso corresponds to the MAP estimator of a posterior derived from a linear forward model with additive Gaussian noise and a Laplace prior [277]. The frequentist analysis of M-estimators has long been an area of active research [201, 288, 191, 307, 244, 8]. Our new theory for hierarchical Bayesian models builds on the general framework for M-estimation developed in [265], which identifies two key properties of the loss and the penalty under which accurate reconstructions are possible in high-dimensional settings: *restricted strong convexity* of the loss and *decomposability* of the regularizer. As we shall see, decomposability does not hold for the penalties that stem from the hierarchical Bayesian models for inverse problems that we consider. To tackle this challenge, we will introduce a weaker notion of *approximate decomposability* under which we obtain non-asymptotic bounds on the deviation between the MAP estimator and the true unknown generating the data.

Outline and Main Contributions

Section 4.2 formalizes the inverse problem of interest and introduces three guiding examples of hierarchical Bayesian models that promote sparsity, group sparsity, and sparse representations. These examples will showcase the broad applicability of our general theory, and will also contribute to making the presentation accessible to a wide audience. Section 4.3 sets forth our unified theoretical framework to establish reconstruction bounds leveraging approximate decomposability of hierarchical Bayesian regularizers. Section 4.4 applies the general theory to our three guiding examples. To streamline the presentation, we defer proofs to Section 4.5 and several appendices. We close in Section 4.6 with conclusions and questions for future research.

This paper contributes to bridge the fields of inverse problems and high-dimensional statistics. In addition to providing a unified and accessible perspective of recent literature

on inverse problems and statistical M-estimation theory, this paper makes the following original contributions:

- Theorem 4.2.6 proves linear convergence of an iterative alternating scheme to compute the MAP estimator for a hierarchical model that promotes group sparsity. Example 4.2.7 proposes a novel hierarchical model to promote sparse representations in an overcomplete dictionary. We derive an iterative alternating scheme to compute the MAP estimator, and we prove its linear convergence in Theorem 4.2.8.

- Definition 4.3.1 introduces the notion of approximate decomposability, and Lemma 4.3.8 highlights an important consequence of this property. Our main result, Theorem 4.3.11, establishes reconstruction error bounds for M-estimators with approximately decomposable regularizers. Our theory generalizes the one in [265], which only holds for exactly decomposable regularizers and is not applicable to the hierarchical models we consider.

- Theorems 4.4.1 and 4.4.3 provide the first known reconstruction error bounds for MAP estimation using sparsity-promoting conditionally Gaussian models with gamma hyperpriors under both well-specified and misspecified settings. For Gaussian noise models and forward maps with Gaussian columns, we sharply characterize the reconstruction errors under well-specified and misspecified settings in Corollaries 4.4.2 and 4.4.4. Our results show that with an appropriate choice of hyperparameters, the MAP estimator achieves the minimax rate for high-dimensional sparse linear regression. Analogous results for group sparsity are established in Theorem 4.4.5 and Corollary 4.4.6, and for sparse representations in Theorem 4.4.7 and Corollary 4.4.8.

4.2 Problem Formulation

Consider the inverse problem of reconstructing an unknown $u \in \mathbb{R}^d$ from data $y \in \mathbb{R}^n$ related by

$$y = Au + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I). \tag{4.1}$$

Here, $A \in \mathbb{R}^{n \times d}$ is a given forward map and ε denotes measurement noise. We assume white noise without loss of generality —our theory can be readily extended to correlated Gaussian noise using a whitening transformation. We are interested in high-dimensional settings where $d \gg n$. In this underdetermined regime, accurate reconstruction hinges on low-dimensional structure that arises, for instance, when the unknown is sparse or admits a sparse representation.

In the Bayesian approach to inverse problems, inference is based on the posterior distribution, which combines the Gaussian likelihood $y|u \sim \mathcal{N}(Au, I)$ implied by (4.1) with a prior distribution that encodes structural beliefs on the unknown. We focus on sparsity-promoting hierarchical priors, where the level of sparsity is controlled by an auxiliary parameter $\theta \in \mathbb{R}^k$. Specifically, we consider priors of the form

$$\pi(u, \theta) = \pi(u|\theta)\pi_\eta(\theta),$$

where the hyperprior π_η on θ depends on a parameter η that will play an important role in the theory. Combining the likelihood and the hierarchical prior, we obtain the posterior distribution

$$\pi(u, \theta|y) \propto \pi(y|u)\pi(u|\theta)\pi_\eta(\theta) \propto \exp\big(-\mathsf{J}(u, \theta)\big), \tag{4.2}$$

where J denotes (up to an additive constant) the negative log-density of the posterior. The

maximum a posteriori (MAP) estimator for (u, θ) is then given by

$$(\hat{u}, \hat{\theta}) = \arg\max_{u,\theta} \pi(u, \theta|y) = \arg\min_{u,\theta} \mathsf{J}(u, \theta). \tag{4.3}$$

This paper conducts a high-dimensional analysis of MAP estimators for hierarchical Bayesian models under the frequentist assumption that the data are generated from a fixed true value $u^\star$ of the unknown, so that

$$y = Au^\star + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I). \tag{4.4}$$

Our goal is to derive bounds on the reconstruction error $\hat{\Delta} := \hat{u} - u^\star$ between the MAP estimator $\hat{u}$ and the true unknown parameter $u^\star$. In a unified framework, we will analyze several hierarchical models that satisfy the following two key properties:

Property 4.2.1. The MAP estimator $(\hat{u}, \hat{\theta})$ can be computed via an alternating optimization scheme of the form:

1. Initialize θ^0, $\ell = 0$.

2. Iterate until convergence

   (i) u-update: $u^{\ell+1} := \arg\min_u \mathsf{J}(u, \theta^\ell)$.

   (ii) θ-update: $\theta^{\ell+1} := \arg\min_\theta \mathsf{J}(u^{\ell+1}, \theta)$.

   (iii) $\ell \to \ell + 1$.

We require that the updates have closed form solutions and the alternating optimization scheme exhibits at least linear convergence to the MAP estimator. □

Property 4.2.2. If $(\hat{u}, \hat{\theta})$ is the MAP estimator, then $\hat{u}$ is the minimizer of an objective $\mathsf{F} : \mathbb{R}^d \to \mathbb{R}$ of the form

$$\mathsf{F}(u) := \frac{1}{2n}\|y - Au\|_2^2 + \lambda \mathsf{R}_\eta(u), \tag{4.5}$$

where $\lambda \mathsf{R}_\eta(\cdot)$ is a convex regularization term determined by the hierarchical prior. $\square$

Previous work has largely focused on Property 4.2.1, motivating hierarchical Bayesian models by the simplicity and effectiveness of alternating optimization algorithms for MAP estimation [84]. On the other hand, in this paper we focus on Property 4.2.2 and show that many regularizers R_η arising from hierarchical Bayesian models satisfy an approximate decomposability property which enables accurate reconstructions under appropriate model assumptions. We next provide three important examples of hierarchical models that satisfy Properties 4.2.1 and 4.2.2. In Section 4.4, we will apply to these three examples the unified theory introduced in Section 4.3.

Example 4.2.3 (Hierarchical Models for Sparsity and the IAS Algorithm). First, we consider a popular hierarchical Bayesian model with gamma hyperpriors to reconstruct sparse parameters, given by

$$\begin{aligned} u|\theta &\sim \mathcal{N}\Big(0, \frac{\sqrt{2}}{\lambda n} D_\theta\Big), \qquad D_\theta = \mathrm{diag}(\theta), \\ \theta_j &\overset{\text{i.i.d}}{\sim} \mathrm{Gamma}\left(\frac{\sqrt{2}\theta_j^*}{\lambda n}, \frac{3}{2} + \frac{\sqrt{2}}{2}\lambda n \eta\right), \quad 1 \le j \le d, \end{aligned} \tag{4.6}$$

with scale parameter $\frac{\sqrt{2}\theta_j^*}{\lambda n} > 0$ and shape parameter $\frac{3}{2} + \frac{\sqrt{2}}{2}\lambda n \eta > 0$. Here, the auxiliary parameter $\theta \in \mathbb{R}^k$, with $k := d$, controls the variance of the coordinates of u, and thereby the level of sparsity imposed on the reconstruction. The hierarchical Bayesian model with gamma hyperpriors in (4.6) was proposed and investigated in [75, 84] as a computationally efficient alternative to ℓ_1 penalization methods for sparse inverse problems. From (4.6), we obtain the joint prior

$$\pi(u, \theta) \propto \exp\left(-\frac{\lambda n}{2\sqrt{2}} \sum_{j=1}^d \frac{u_j^2}{\theta_j} - \frac{\lambda n}{\sqrt{2}} \sum_{j=1}^d \left[\frac{\theta_j}{\theta_j^*} - \eta \log \frac{\theta_j}{\theta_j^*}\right]\right).$$

We assume for the simplicity of our subsequent analysis that the scale parameters are chosen

such that $\theta_j^* = \frac{n}{\|A_j\|_2^2}$, where $A_j \in \mathbb{R}^n$ denotes the jth column of the forward map. We then make the transformation $\big(\theta_j, u_j, A\big) \mapsto \left(\frac{\theta_j}{\theta_j^*}, \frac{u_j}{\sqrt{\theta_j^*}}, AD_{\theta^*}^{1/2}\right)$. Notice that our assumption on the scale parameters is therefore tantamount to assuming that the forward map is column-normalized so that $\frac{\|A_j\|_2}{\sqrt{n}} = 1$ and that $\theta_j^* = 1$ for each $1 \leq j \leq d$. As such, we consider a posterior with negative log-density given, up to an additive constant, by

$$\mathsf{J}(u,\theta) := \underbrace{\frac{1}{2n}\|y - Au\|_2^2 + \overbrace{\frac{\lambda}{2\sqrt{2}}\sum_{j=1}^{d}\frac{u_j^2}{\theta_j} + \frac{\lambda}{\sqrt{2}}\sum_{j=1}^{d}\big(\theta_j - \eta\log\theta_j\big)}^{(b)}}_{(a)} \ . \tag{4.7}$$

The iterative alternating scheme (IAS) computes the MAP estimator for this objective by minimizing the least-squares objective (a) for the u-update and minimizing (b) for the θ-update. Given (u^ℓ, θ^ℓ) both updates have a closed form solution:

$$\begin{aligned} u^{\ell+1} &:= \left(\frac{A^TA}{n} + \frac{\lambda}{\sqrt{2}}D_{\theta^\ell}^{-1}\right)^{-1}\frac{A^Ty}{n}, \\ \theta_j^{\ell+1} &:= \frac{\eta}{2} + \sqrt{\frac{\eta^2}{4} + \frac{(u_j^{\ell+1})^2}{2}}, \qquad 1 \leq j \leq d. \end{aligned} \tag{4.8}$$

For large problems, the u-update can be efficiently computed using conjugate gradient with early stopping based on Morozov's discrepancy principle [75, 77, 84]. The closed form expression for the θ-update was derived in [84]. The IAS algorithm (4.8) was introduced in [75, 84] and has been widely studied, see e.g. [75, 77, 84, 78, 79]. The following known result verifies Property 4.2.2 and summarizes other key computational properties.

Theorem 4.2.4. *For $\eta > 0$, the objective function J in* (4.7) *is strictly convex over $\mathbb{R}^d \times \mathbb{R}_+^d$ and thus admits a unique MAP estimator $(\hat{u}, \hat{\theta})$. Moreover, the u-iterates u^ℓ of the IAS algorithm* (4.8) *converge at least linearly to $\hat{u}$ in the Mahalanobis norm $\|\cdot\|_{D_{\hat{\theta}}}$. Furthermore,*

$\hat{u}$ is the global minimizer of F *in* (4.5) *with*

$$\mathsf{R}_\eta(u) := \frac{1}{\sqrt{2}} \sum_{j=1}^{d} \left[\frac{u_j^2}{2 f_j(u)} + f_j(u) - \eta \log f_j(u) \right], \qquad f_j(u) := \frac{\eta}{2} + \sqrt{\frac{\eta^2}{4} + \frac{u_j^2}{2}}\,. \tag{4.9}$$

As illustrated by Theorem 4.2.4, previous works focus on the convexity of J and the convergence of IAS iterates u^ℓ to the MAP estimator $\hat{u}$. In this work we focus instead on understanding the convergence of the MAP estimator $\hat{u}$ to the true parameter $u^\star$ generating the data.

Example 4.2.5 (Hierarchical Models for Group Sparsity and the GS-IAS Algorithm). The second hierarchical Bayesian model we consider promotes group sparsity. A particular case of the model where each group consists of precisely three indices was proposed in [75], and the general setting of arbitrary group sizes was later formulated in [48]. We assume the a priori knowledge that components of the unknown u are partitioned into groups $\{g_1, \ldots, g_k\}$ where $|g_j| = p_j$. The model also postulates a conditionally Gaussian prior on each group and the variance of each group is modeled via a gamma hyperprior,

$$\begin{aligned} u_{g_j} | \theta_j &\sim \mathcal{N}\Big(0, \frac{\sqrt{2}\theta_j}{\lambda n} C_j\Big), \qquad C_j \succ 0, \\ \theta_j &\sim \text{Gamma}\left(\frac{\sqrt{2}\theta_j^*}{\lambda n}, \frac{p_j + 2}{2} + \frac{\sqrt{2}}{2} \lambda n \eta \right), \quad 1 \le j \le k, \end{aligned} \tag{4.10}$$

with scale parameter $\frac{\sqrt{2}\theta_j^*}{\lambda n} > 0$ and shape parameter $\frac{p_j+2}{2} + \frac{\sqrt{2}}{2}\lambda n \eta$. Here, the auxiliary parameter $\theta \in \mathbb{R}^k$ controls the level of group sparsity in the reconstructed solution. From (4.10), we have the joint prior distribution

$$\pi(u, \theta) \propto \exp\left(-\frac{\lambda n}{2\sqrt{2}} \sum_{j=1}^{k} \frac{\|u_{g_j}\|_{C_j}^2}{\theta_j} - \frac{\lambda n}{\sqrt{2}} \sum_{j=1}^{k} \left[\frac{\theta_j}{\theta_j^*} - \eta \log \frac{\theta_j}{\theta_j^*} \right] \right).$$

To simplify our analysis, we assume that the scale parameters are chosen such that $\theta_j^* =$

$\frac{n}{\|A_{g_j}\|_{C_j,2}^2}$, where $A_{g_j} \in \mathbb{R}^{n\times p_j}$ denotes the matrix obtained by keeping the columns from A in group g_j, and $\|A_{g_j}\|_{C_j,2} = \max_{\|v\|_{C_j}=1} \|A_{g_j}\|_2$. Then, via the transformation $(\theta_j, u_{g_j}, A_{g_j}) \mapsto \left(\frac{\theta_j}{\theta_j^*}, \frac{u_j}{\sqrt{\theta_j^*}}, A_{g_j}\sqrt{\theta_j^*}\right)$ we see that our assumption on the scale parameters is equivalent to assuming the forward map is block-normalized so that $\frac{\|A_{g_j}\|_{C_j,2}}{\sqrt{n}} = 1$ with $\theta_j^* = 1$. Applying Bayes rule, we consider a posterior with negative log-density given, up to an additive constant, by

$$\mathsf{J}(u,\theta) := \underbrace{\frac{1}{2n}\|y - Au\|_2^2 + \overbrace{\frac{\lambda}{2\sqrt{2}}\|u\|_{D_\theta}^2 + \frac{\lambda}{\sqrt{2}}\sum_{j=1}^{k}\left(\theta_j - \eta\log\theta_j\right)}^{(b)}}_{(a)}, \tag{4.11}$$

where D_θ is the block-diagonal matrix given by

$$D_\theta := \begin{bmatrix} \theta_1 C_1 & & \\ & \ddots & \\ & & \theta_k C_k \end{bmatrix}.$$

To compute the MAP estimator for this objective, we again use an alternating optimization scheme, minimizing the least-squares objective (a) for the u-update and minimizing (b) for the θ-update. Given (u^ℓ, θ^ℓ), these updates have closed form solutions:

$$\begin{aligned} u^{\ell+1} &:= \left(\frac{A^TA}{n} + \frac{\lambda}{\sqrt{2}}D_{\theta^\ell}^{-1}\right)\frac{A^Ty}{n}, \\ \theta_j^{\ell+1} &:= \frac{\eta}{2} + \sqrt{\frac{\eta^2}{4} + \frac{\|u_{g_j}\|_{C_j}^2}{2}}, \qquad 1 \le j \le k. \end{aligned} \tag{4.12}$$

The group-sparse iterative alternating scheme (GS-IAS) described by (4.12) was derived in [75] and it was shown that the iterates converge to the minimizer of 4.11, although no rate of convergence was derived. This objective and the GS-IAS algorithm admit a number of computationally convenient properties, summarized in the following theorem.

Theorem 4.2.6. *For $\eta > 0$, the objective function* J *in* (4.11) *is strictly convex over* $\mathbb{R}^d \times \mathbb{R}^k_+$, *and thus admits a unique MAP estimator* $(\hat{u}, \hat{\theta})$. *Moreover, the u-iterates* u^ℓ *of the GS-IAS algorithm* (4.12) *converge at least linearly to* $\hat{u}$ *in the Mahalanobis norm* $\|\cdot\|_{D_{\hat{\theta}}}$. *Furthermore,* $\hat{u}$ *is the global minimizer of* F *in* (4.5) *with*

$$\mathsf{R}_\eta(u) := \frac{1}{\sqrt{2}} \sum_{j=1}^{k} \left[\frac{\|u_{g_j}\|^2_{C_j}}{2 f_j(u)} + f_j(u) - \eta \log f_j(u) \right], \qquad f_j(u) := \frac{\eta}{2} + \sqrt{\frac{\eta^2}{4} + \frac{\|u_{g_j}\|^2_{C_j}}{2}}. \tag{4.13}$$

For completeness, we include a proof of Theorem 4.2.6 in Appendix 4.7.

Example 4.2.7 (Hierarchical Models for Sparse Representations and the O-IAS Algorithm). Our final motivating example is a novel hierarchical Bayesian model to promote the sparse representation of the unknown in some redundant dictionary, $W \in \mathbb{R}^{d \times k}$. In this model, we assume the dictionary to be overcomplete, meaning that $k \gg d$. For simplicity, we assume that W is a *tight frame* [344] normalized so that $WW^T = I$. Relevant examples of such dictionaries include Gabor frames [250], curvelet frames [88], and wavelet frames [294]. To encode our a priori information that the coefficients of u in W are sparse, we place a Gaussian prior on $W^T u$ and gamma hyperpriors on the variances,

$$\begin{aligned} W^T u | \theta &\sim \mathcal{N}\Big(0, \frac{\sqrt{2}}{\lambda n} D_\theta\Big), \qquad D_\theta = \operatorname{diag}(\theta), \\ \theta_j &\sim \operatorname{Gamma}\left(\frac{\sqrt{2}\theta_j^*}{\lambda n}, \frac{3}{2} + \frac{\sqrt{2}}{2} \lambda n \eta \right), \quad 1 \le j \le k, \end{aligned} \tag{4.14}$$

with scale parameters $\frac{\sqrt{2}\theta_j^*}{\lambda n}$ and shape parameters $\frac{3}{2} + \frac{\sqrt{2}}{2}\lambda n \eta$ for $\eta, \lambda > 0$. Here, the auxiliary parameter $\theta \in \mathbb{R}^k$ controls the variance of the coefficients of u in the dictionary W, and consequently the level of sparsity imposed upon the coefficients. From (4.14) we have

the joint prior distribution

$$\pi(W^T u, \theta) \propto \exp\left(-\frac{\lambda n}{2\sqrt{2}} \sum_{j=1}^{k} \frac{(W^T u)_j^2}{\theta_j} - \frac{\lambda n}{\sqrt{2}} \sum_{j=1}^{k} \left[\frac{\theta_j}{\theta_j^*} - \eta \log \frac{\theta_j}{\theta_j^*} \right] \right).$$

For the convenience of our analysis in the sequel, we assume the scale parameters are chosen such that $\theta_j^* = \frac{n}{\|(AW)_j\|_2^2}$. Making $\left(\theta_j, (W^T u)_j, (AW)_j\right) \mapsto \left(\frac{\theta_j}{\theta_j^*}, \frac{(W^T u)_j}{\sqrt{\theta_j^*}}, (AW)_j \sqrt{\theta_j^*} \right)$, we see that our assumption on the scale parameters is equivalent to assuming that the forward map is normalized so that $\frac{\|(AW)_j\|_2}{\sqrt{n}} = 1$ with $\theta_j^* = 1$. Consequently, we consider a posterior with negative log-density given, up to an additive constant, by

$$\mathsf{J}(u, \theta) := \underbrace{\frac{1}{2n}\|y - Au\|_2^2 + \overbrace{\frac{\lambda}{2\sqrt{2}} \sum_{j=1}^{k} \frac{(W^T u)_j^2}{\theta_j}}^{(b)} + \frac{\lambda}{\sqrt{2}} \sum_{j=1}^{k} \left(\theta_j - \eta \log \theta_j \right)}_{(a)} . \tag{4.15}$$

To compute the MAP estimator, we use the following overcomplete iterative alternating scheme (O-IAS), which minimizes (a) for the u-update and (b) for the θ-update:

$$\begin{aligned} u^{\ell+1} &:= \left(\frac{A^T A}{n} + \frac{\lambda}{\sqrt{2}} W D_{\theta^\ell}^{-1} W^T \right) \frac{A^T y}{n}, \\ \theta_j^{\ell+1} &:= \frac{\eta}{2} + \sqrt{\frac{\eta^2}{4} + \frac{(W^T u)_j^2}{2}}, \qquad 1 \le j \le k. \end{aligned} \tag{4.16}$$

The objective (4.15) and the accompanying O-IAS algorithm (4.16) admit a number of key computationally convenient properties.

Theorem 4.2.8. *For $\eta > 0$, the objective function J in (4.15) is strictly convex over $\mathbb{R}^d \times \mathbb{R}_+^k$, and thus admits a unique MAP estimator $(\hat{u}, \hat{\theta})$. Moreover, the u-iterates u^ℓ of the O-IAS algorithm converge at least linearly to $\hat{u}$ in the Mahalanobis norm $\|\cdot\|_{(W D_{\hat{\theta}}^{-1} W^T)^{-1}}$.*

Furthermore, $\hat{u}$ is the global minimizer of F in (4.5) with

$$\mathsf{R}_\eta(u) = \frac{1}{\sqrt{2}} \sum_{j=1}^{k} \left[\frac{(W^T u)_j}{2 f_j(u)} + f_j(u) - \eta \log f_j(u) \right], \qquad f_j(u) = \frac{\eta}{2} + \sqrt{\frac{\eta^2}{4} + \frac{(W^T u)_j^2}{2}}. \tag{4.17}$$

For completeness, we include a proof of Theorem 4.2.8 in Appendix 4.7.

4.3 Unified Theoretical Framework

This section develops a framework to obtain reconstruction bounds for hierarchical Bayesian models under appropriate model assumptions. In Subsection 4.3.1 we introduce the notion of approximate decomposabilility, and show that this property is satisfied by the regularizers in the examples in Section 4.2. Then, in Subsection 4.3.2 we present the main result, Theorem 4.3.11.

4.3.1 Approximate Decomposability

We introduce a *model subspace* $\mathcal{M} \subset \mathbb{R}^d$ to capture the sparsity constraints of the problem. We refer to the orthogonal complement $\mathcal{M}^\perp$ of the model subspace as the *perturbation subspace*. Following [265, 342], we say that a regularizer $\mathsf{R} : \mathbb{R}^d \to \mathbb{R}^+$ is decomposable with respect to $\mathcal{M}$ if

$$\mathsf{R}(u + v) = \mathsf{R}(u) + \mathsf{R}(v), \qquad \forall u \in \mathcal{M}, \forall v \in \mathcal{M}^\perp. \tag{4.18}$$

For norm-based regularizers, it necessarily holds that $\mathsf{R}(u + v) \leq \mathsf{R}(u) + \mathsf{R}(v)$. Decomposability requires that equality holds for perturbations $v \in \mathcal{M}^\perp$ away from vectors $u \in \mathcal{M}$, thus ensuring that deviations from the model subspace along the perturbation subspace are heavily penalized. The regularizers R_η discussed in Section 4.2 determined by the hierar-

chical priors are not decomposable over a natural model subspace. However, in the limit as $\eta \to 0$, each of these regularizers approach a decomposable regularizer. Motivated by this observation, we introduce the following notion of *approximate decomposability*.

Definition 4.3.1. A family of regularizers $\mathsf{R}_\eta : \mathbb{R}^d \to \mathbb{R}^+$ parameterized by $\eta > 0$ is approximately decomposable with respect to $\mathcal{M}$ if there exists a norm-based regularizer R decomposable with respect to $\mathcal{M}$ and non-negative functions $c_i^L(\eta)$ and $c_i^U(\eta)$, $i = 1, 2$, such that, for any $u \in \mathbb{R}^d$,

$$\Big(1 - c_1^L(\eta)\Big)\,\mathsf{R}(u) - c_2^L(\eta) \leq \mathsf{R}_\eta(u) \leq \Big(1 + c_1^U(\eta)\Big)\,\mathsf{R}(u) + c_2^U(\eta), \tag{4.19}$$

where $c_i^L(\eta) \to 0$ and $c_i^U(\eta) \to 0$ as $\eta \to 0$.

The functions c_i^L and c_i^U quantify how fast the regularizer $\mathsf{R}_\eta(u)$ converges to the decomposable regularizer $\mathsf{R}(u)$ as η is taken to be close to zero. We remark that other generalizations of decomposability, such as *weakly decomposable norms* [338] and *atomic norms* [100, 87], have previously been studied. Our notion of approximate decomposability differs most significantly from these other generalizations in that it does not require the regularizer itself be a norm, merely that the regularizer be close to a norm for small η. We now discuss the choice of model subspace and establish approximate decomposability for the three example models in Section 4.2.

Example 4.3.2 (Approximate Decomposability: Hierarchical Models for Sparsity). The natural modelling assumption for the IAS algorithm is that of sparsity. Let $S \subset \{1, \ldots, d\}$ be a subset of indices with cardinality $|S| = s$. We define the model subspace

$$\mathcal{M}(S) := \Big\{u \in \mathbb{R}^d : u_j = 0 \text{ for all } j \notin S\Big\}. \tag{4.20}$$

Any vector in the model subspace $\mathcal{M}(S)$ has at most s nonzero entries. In this setting, the

perturbation subspace is the space of vectors with support contained entirely outside of S:

$$\mathcal{M}^{\perp}(S) := \left\{ u \in \mathbb{R}^d : u_j = 0 \text{ for all } j \in S \right\}.$$

One can verify that, for any subset of indices S, the ℓ_1 norm $\mathsf{R}(u) = \|u\|_1$ is decomposable with respect to $\mathcal{M}(S)$. The following lemma shows that the regularizer R_η defined in (4.9) is approximately decomposable with respect to $\mathcal{M}(S)$.

Lemma 4.3.3. *Consider, for* $0 < \eta < \frac{1}{2}$, *the family of regularizers* R_η *in* (4.9). *For any* $u \in \mathbb{R}^d$,

$$\left(1 - \frac{\eta}{2}\right)\|u\|_1 - \frac{d\eta}{\sqrt{2}} \leq \mathsf{R}_\eta(u) \leq \|u\|_1 + \frac{d\eta}{\sqrt{2}}(1 - \log \eta). \tag{4.21}$$

The proof can be found in Appendix 4.8. In the notation of Definition 4.3.1, approximate decomposability holds with $c_1^L(\eta) = \frac{\eta}{2}$, $c_2^L(\eta) = \frac{d\eta}{\sqrt{2}}$, $c_1^U(\eta) = 0$, and $c_2^U(\eta) = \frac{d\eta}{\sqrt{2}}(1 - \log \eta)$.

Example 4.3.4 (Approximate Decomposability: Hierarchical Models for Group Sparsity). We partition the index set into groups $\mathcal{G} = \{g_1, \dots g_k\}$ where each $g_j \subset \{1, \dots, d\}$, $g_i \cap g_j = \emptyset$ for $i \neq j$, and $\cup_{j=1}^k g_j = \{1, \dots, d\}$. We denote the size of each group as $|g_j| = p_j$. For any set of group indices $S_{\mathcal{G}} \subset \mathcal{G}$, we define the model subspace

$$\mathcal{M}(S_{\mathcal{G}}) := \left\{ u \in \mathbb{R}^d : u_g = 0 \text{ for all } g \notin S_{\mathcal{G}} \right\}.$$

This is precisely the subspace of vectors supported only on groups in the index set $S_{\mathcal{G}}$. Conversely, the perturbation subspace is the space of vectors with support contained in groups not in $S_{\mathcal{G}}$:

$$\mathcal{M}^{\perp}(S_{\mathcal{G}}) := \left\{ u \in \mathbb{R}^d : u_g = 0 \text{ for all } g \in S_{\mathcal{G}} \right\}.$$

A decomposable regularizer for this model subspace is the *group sparse norm* [351], given by

$$\mathsf{R}(u) = \sum_{j=1}^{k} \|u_{g_j}\|_{C_j},$$

where $C_j \in \mathbb{R}^{p_j \times p_j}$ is a positive definite matrix. Indeed, we see that, for any $u \in \mathcal{M}(S_{\mathcal{G}})$ and $v \in \mathcal{M}^{\perp}(S_{\mathcal{G}})$, it holds that

$$\mathsf{R}(u+v) = \sum_{g_j \in \mathcal{G}} \|u_{g_j}\|_{C_j} + \sum_{g_j \notin \mathcal{G}} \|v_{g_j}\|_{C_j} = \mathsf{R}(u) + \mathsf{R}(v).$$

The following lemma shows that the regularizer R_η defined in (4.13) is approximately decomposable with respect to $\mathcal{M}(S_{\mathcal{G}})$.

Lemma 4.3.5. *Consider, for* $0 < \eta < \frac{1}{2}$, *the family of regularizers* R_η *in* (4.13). *For any* $u \in \mathbb{R}^d$,

$$\Big(1 - \frac{\eta}{2}\Big) \sum_{j=1}^{k} \|u_{g_j}\|_{C_j} - \frac{k\eta}{\sqrt{2}} \leq \mathsf{R}_\eta(u) \leq \sum_{j=1}^{k} \|u_{g_j}\|_{C_j} + \frac{k\eta}{\sqrt{2}}(1 - \log \eta).$$

The proof can be found in Appendix 4.8. In the notation of Definition 4.3.1, approximate decomposability holds with $c_1^L(\eta) = \frac{\eta}{2}$, $c_2^L(\eta) = \frac{k\eta}{\sqrt{2}}$, $c_1^U(\eta) = 0$, and $c_2^U(\eta) = \frac{k\eta}{\sqrt{2}}(1 - \log \eta)$.

Example 4.3.6 (Approximate Decomposability: Hierarchical Models for Sparse Representations). We consider the situation where the data generating parameter itself is not sparse, but admits a sparse representation in an overcomplete dictionary, $W \in \mathbb{R}^{d \times k}$, where $k \gg d$. As is common in the compressed sensing literature [89], we assume that W is a tight frame such that $WW^T = I$. We let $S_W \subset \{1, \ldots, k\}$ be a subset of indices of cardinality $|S_W| = s$. We then define the model subspace

$$\mathcal{M}(S_W) := \Big\{ u \in \mathbb{R}^d : (W^T u)_j = 0 \text{ for all } j \notin S_W \Big\}. \tag{4.22}$$

This is precisely the subspace of vectors whose representation in W has support contained in S_W. The perturbation subspace is then given by

$$\mathcal{M}^{\perp}(S_W) := \left\{ u \in \mathbb{R}^d : (W^T u)_j = 0 \text{ for all } j \in S_W \right\}.$$

A decomposable regularizer for this model subspace is given by

$$\mathsf{R}(u) = \|W^T u\|_1,$$

since, for any $u \in \mathcal{M}(S_W)$ and $v \in \mathcal{M}^{\perp}(S_W)$, it holds that

$$\mathsf{R}(u+v) = \|W^T u + W^T v\|_1 = \|W^T u\|_1 + \|W^T v\|_1 = \mathsf{R}(u) + \mathsf{R}(v).$$

The following lemma shows that the regularizer R_η defined in (4.17) is approximately decomposable with respect to $\mathcal{M}(S_W)$.

Lemma 4.3.7. *Consider, for* $0 < \eta < \frac{1}{2}$, *the family of regularizers* R_η *in* (4.17). *For any* $u \in \mathbb{R}^d$,

$$\left(1 - \frac{\eta}{2}\right)\|W^T u\|_1 - \frac{k\eta}{\sqrt{2}} \leq \mathsf{R}_\eta(u) \leq \|W^T u\|_1 + \frac{k\eta}{\sqrt{2}}(1 - \log \eta).$$

The proof can be found in Appendix 4.8. In the notation of Definition 4.3.1, approximate decomposability holds with $c_1^L(\eta) = \frac{\eta}{2}$, $c_2^L(\eta) = \frac{k\eta}{\sqrt{2}}$, $c_1^U(\eta) = 0$, and $c_2^U(\eta) = \frac{k\eta}{\sqrt{2}}(1 - \log \eta)$.

We now show how the notion of approximate decomposability can be leveraged to control the error vector $\hat{\Delta} = \hat{u} - u^\star$. To do so, we recall that the dual norm of R with respect to the Euclidean inner product is defined by

$$\mathsf{R}^*(u) := \sup_{v \in \mathbb{R}^d \setminus \{0\}} \frac{\langle u, v \rangle}{\mathsf{R}(v)} = \sup_{\mathsf{R}(v) \leq 1} \langle u, v \rangle.$$

In the following lemma, we show that if the regularization parameter λ is taken to be large

enough relative to the dual norm of the noise vector, then the reconstruction error lies in a specific set. For vector $u \in \mathbb{R}^d$ and subspace $\mathcal{M} \subset \mathbb{R}^d$, we denote by $u_{\mathcal{M}}$ the Euclidean projection of u onto $\mathcal{M}$.

Lemma 4.3.8. *Let $\hat{u}$ be a minimizer of the objective*

$$\mathsf{F}(u) = \frac{1}{2n}\|y - Au\|_2^2 + \lambda \mathsf{R}_\eta(u), \tag{4.23}$$

where R_η is an approximately decomposable regularizer with respect to $\mathcal{M}$. If $\lambda \geq \frac{2}{n}\mathsf{R}^*\left(A^T\varepsilon\right)$ and η is small enough so that $c_1^L(\eta) < \frac{1}{4}$, then the error $\hat{\Delta} = \hat{u} - u^\star$ lies in the set

$$\mathbb{C}_\eta(\mathcal{M}) := \left\{\Delta \in \mathbb{R}^d : \mathsf{R}(\Delta_{\mathcal{M}^\perp}) \leq 7\mathsf{R}(\Delta_{\mathcal{M}}) + 8\mathsf{R}(u^\star_{\mathcal{M}^\perp}) + 4c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) + 4c_2(\eta)\right\},$$

where $c_i(\eta) := c_i^L(\eta) + c_i^U(\eta)$.

The proof can be found in Appendix 4.9. The fact that the error lies in the set $\mathbb{C}_\eta(\mathcal{M})$ allows us to control the components of the error outside the model subspace by the components of the error within the model subspace along with terms that get smaller as the true vector is assumed to lie within the model subspace and as η is taken to be smaller. In the next subsection we leverage this lemma to get a bound on the reconstruction error.

4.3.2 General Bound on the Reconstruction Error

To bound the reconstruction error, we will need to quantify the cost of translating between bounds on a vector in $\mathcal{M}$ given by $\|\cdot\|_2$ and bounds given by $\mathsf{R}(\cdot)$. To that end, we will rely on the notion of *subspace Lipschitz constant*.

Definition 4.3.9. For a subspace $\mathcal{M} \subset \mathbb{R}^d$, the subspace Lipschitz constant with respect

to $(\mathsf{R}, \|\cdot\|_2)$ is given by

$$\Psi(\mathcal{M}) := \sup_{u \in \mathcal{M}\setminus\{0\}} \frac{\mathsf{R}(u)}{\|u\|_2}.$$

We will also need a condition on the forward map. In the high-dimensional setting, the likelihood term $\|y - Au\|_2^2$, necessarily has $d - n$ directions in which it has no curvature up to second order. Consequently, we need an assumption to guarantee that in the relevant directions, chiefly the directions given by vectors in $\mathbb{C}_\eta(\mathcal{M})$, the likelihood has significant curvature at $u^\star$. One such assumption is that of *restricted strong convexity*.

Definition 4.3.10. For a given regularizer R, a linear map $A : \mathbb{R}^d \to \mathbb{R}^n$ satisfies the restricted strong convexity (RSC) condition with curvature κ and tolerance τ^2 if

$$\frac{\|A\Delta\|_2^2}{2n} \geq \frac{\kappa}{2}\|\Delta\|_2^2 - \tau^2 \mathsf{R}^2(\Delta), \qquad \forall \Delta \in \mathbb{R}^d. \tag{4.24}$$

If the matrix A were not rank degenerate, then the RSC condition would hold with $\tau = 0$, which recovers the usual definition of strong convexity. In the high-dimensional setting, this cannot be the case, and a nonzero tolerance term is required. In general, we expect the tolerance to be decreasing in the number n of observations.

Given these definitions, we now state our main result, which generalizes the bounds in [342, Chapter 9] to hold for approximately decomposable regularizers. The proof can be found in Appendix 4.9.

Theorem 4.3.11. *Let $u^\star$ be the true parameter generating the data via* (4.4) *and let $\hat{u}$ be the minimizer of the objective* (4.23)*. Assume that R_η is approximately decomposable with respect to $\mathcal{M}$ and that η is small enough such that $c_1^L(\eta) < \frac{1}{4}$ and that the forward map A satisfies the RSC condition with curvature κ and tolerance τ^2. Then, if $\tau\Psi(\mathcal{M}) \leq \frac{\sqrt{\kappa}}{32}$ and*

$\lambda \geq 2\mathsf{R}^*(\frac{1}{n}A^T\varepsilon)$, *we have the bound*

$$\|\hat{u} - u^\star\|_2^2 \lesssim \frac{\lambda^2}{\kappa^2}\Psi^2(\mathcal{M}) + \frac{\tau^2}{\kappa}\mathsf{R}(u^\star_{\mathcal{M}^\perp})^2 + \frac{\lambda}{\kappa}\mathsf{R}(u^\star_{\mathcal{M}^\perp}) + \mathcal{E}_\eta, \tag{4.25}$$

where

$$\mathcal{E}_\eta := \frac{\tau^2}{\kappa}\Big(c_1(\eta)^2\mathsf{R}(u^\star_{\mathcal{M}})^2 + c_2(\eta)^2\Big) + \frac{\lambda}{\kappa}\Big(c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) + c_2(\eta)\Big),$$

and $c_i(\eta) := c_i^L(\eta) + c_i^U(\eta)$, $i = 1, 2$.

We emphasize that this bound is deterministic, and makes no specific distributional assumptions on the noise or forward map. The first three terms are precisely the terms that appear in the bound proven in [342] for exactly decomposable regularizers. The first term,

$$\mathcal{E}_{\text{est}} := \frac{\lambda^2}{\kappa^2}\Psi^2(\mathcal{M}),$$

can be thought of as the statistical estimation error which captures the inherent difficulty of estimating a parameter in $\mathcal{M}$. The second and third terms together,

$$\mathcal{E}_{\text{app}} := \frac{\tau^2}{\kappa}\mathsf{R}(u^\star_{\mathcal{M}^\perp})^2 + \frac{\lambda}{\kappa}\mathsf{R}(u^\star_{\mathcal{M}^\perp}),$$

represent the approximation error. These terms capture the modelling error incurred when the data-generating parameter does not lie exactly in $\mathcal{M}$, thus quantifying the amount of model misspecification. Notice that if $u^\star \in \mathcal{M}$, then these terms vanish. The final term,

$$\mathcal{E}_\eta := \frac{\tau^2}{\kappa}\Big(c_1(\eta)^2\mathsf{R}(u^\star_{\mathcal{M}})^2 + c_2(\eta)^2\Big) + \frac{\lambda}{\kappa}\Big(c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) + c_2(\eta)\Big),$$

is novel to our bound, and captures the error incurred by the approximate decomposability of the regularizer. This term can be made arbitrarily small with η, which is user-specified.

We remark that restricted strong convexity and approximate decomposability are two of many choices for conditions on the forward map and the regularizer to guarantee reconstruction bounds. In the variational inverse problems literature, the *source condition* and *range condition* have been used to guarantee the reconstruction error is bounded by the observation noise level with proper choice of regularization parameter [69, 29]. These conditions, however, are not natural to verify for the regularizers resulting from the hierarchical models studied here and the results typically assume bounded noise. Such conditions also may not capture the low-dimensional structure of the problem that the hierarchical models are designed to exploit. To this end, [179] considers a *restricted injectivity condition* that weakens the range condition and exploits sparsity. However, these results also assume bounded noise and are thus not immediately applicable to the setting considered here.

4.4 Reconstruction Error for Sparsity-Promoting Models

In this section we apply the unified theoretical framework set forth in Theorem 4.3.11 to each of the examples of hierarchical models discussed in Section 4.2. The results in this section provide the first known reconstruction error bounds for these models. Proofs will be deferred to Section 4.5.

4.4.1 Hierarchical Models for Sparsity

Here we apply our general theory to the sparsity-promoting hierarchical model presented in Example 4.2.3 under both a hard-sparsity assumption and a weaker ℓ^q-sparsity assumption. For both of these types of sparsity, we first prove a deterministic result guaranteeing that under the RSC condition, the MAP estimator will be close to the true data generating parameter as long as we regularize enough relative to the noise level. We emphasize that these results do not make specific assumptions on the noise distribution or the forward map, besides the RSC condition. Applying the deterministic results to the probabilistic setting

of Gaussian noise and Gaussian forward maps, we derive sharp rates in terms of only the dimension, sample size, and the parameter η. These results hold with high probability.

The first assumption we consider is that of hard sparsity. We assume that the true parameter generating the data has support of cardinality s, where $s \ll d$. The following result holds for any s-sparse $u^\star$ so long as the forward map satisfies the RSC condition and the regularization parameter is taken to be large enough.

Theorem 4.4.1. *Let $u^\star$ be the true parameter generating data via (4.1) and suppose that $supp(u^\star) \subset S$ where $|S| = s$. Let $(\hat{u}, \hat{\theta})$ be the unique minimizer of the objective function (4.7) with $0 < \eta < 1/2$. Assume that $\lambda \geq \frac{2\|A^T\varepsilon\|_\infty}{n}$ and that A satisfies the RSC condition (4.24) with curvature κ and tolerance τ^2. Then, if $\tau\sqrt{s} \leq \frac{\sqrt{\kappa}}{32}$ it holds that*

$$\|\hat{u} - u^\star\|_2^2 \lesssim s\frac{\lambda^2}{\kappa^2} + \mathcal{E}_\eta, \tag{4.26}$$

where

$$\mathcal{E}_\eta := \frac{\tau^2}{\kappa}\left(\eta^2\|u_S^\star\|_1^2 + \frac{d^2\eta^2}{2}(2 - \log\eta)^2\right) + \frac{\lambda}{\kappa}\left(\eta\|u_S^\star\|_1 + \frac{d\eta}{\sqrt{2}}(2 - \log\eta)\right).$$

We emphasize that this theorem is deterministic. It makes no specific distributional assumptions on the noise or forward map, only requiring that the regularization parameter is large enough relative to the noise and that the forward map satisfies the RSC condition. The bound illustrates that, as long as the condition $\lambda \geq \frac{2\|A^T\varepsilon\|_\infty}{n}$ is satisfied, we would like to take the regularization parameter λ, the RSC tolerance τ, and the parameter η to be as small as possible. We remark that η is user supplied and can be made arbitrarily small, but doing so may slow down the convergence of the IAS algorithm. The other terms in the bound, λ and τ, inherently depend on the noise and the forward map. As an illustrative example, we explicitly characterize λ and τ under the assumptions of Gaussian noise and a

Gaussian forward map in the following corollary:

Corollary 4.4.2. *Let the forward map $A \in \mathbb{R}^{n\times d}$ have rows distributed i.i.d. according to $\mathcal{N}(0,\Sigma)$, and assume the columns of A are normalized such that $\frac{\|A_j\|_2}{\sqrt{n}} = 1$ for all $j = 1,\dots,d$. Assume the noise is given by $\varepsilon \sim \mathcal{N}(0,I)$ and that $(\hat{u},\hat{\theta})$ is the unique minimizer of the IAS objective functional (4.7) with $\lambda = 4\sqrt{\frac{\log d}{n}}$ and $0 < \eta < 1/2$. If $u^\star$ is s-sparse, then there exists a constant c_Σ depending only on Σ, and universal constants c and c' such that if $n > c_\Sigma s \log d$ it holds with probability at least $1 - c\exp(-c'n\lambda^2)$ that*

$$\|\hat{u} - u^\star\|_2^2 \lesssim c_\Sigma \frac{s\log d}{n} + \min\left\{\varphi(\eta), \varphi(\eta)^2\right\},$$

where

$$\varphi(\eta) := \sqrt{\frac{\log d}{n}}\left(\eta\|u_S^\star\|_1 + \frac{d\eta}{\sqrt{2}}(2 - \log\eta)\right).$$

Notice that the term $\varphi(\eta)$ can be made arbitrarily small by choosing a small enough η. Taking η small enough to match the first term in the bound, the MAP estimator for the hierarchical model in Example 4.2.3 satisfies the non-asymptotic bound

$$\|\hat{u} - u^\star\|_2 \lesssim \sqrt{c_\Sigma \frac{s\log d}{n}},$$

which agrees with known non-asymptotic bounds for the ℓ_1-penalized least squares estimator [41, 257]. The assumptions of Gaussianity in this corollary are largely for theoretical convenience, and are not strictly necessary. Verifying the RSC condition for the forwards maps of interest in the inverse problems community remains a direction of further research.

Instead of assuming the true parameter is exactly s-sparse, one can make the more general

assumption of ℓ^q sparsity. That is, we assume

$$u^\star \in \mathbb{B}_q(R_q) := \left\{ u \in \mathbb{R}^d : \sum_{i=1}^{d} |u_i|^q \leq R_q \right\},$$

where $R_q > 0$ controls the level of sparsity. Notice that if $q = 0$ and $R_q = s$, we recover the hard-sparsity assumption. For $q > 0$ this assumption enforces a decay rate on the entries of $u^\star$. Under ℓ^q sparsity, we choose our model subspace to be vectors with support S_δ, where

$$S_\delta := \left\{ j \in \{1, \ldots, d\} : |u_j^\star| > \delta \right\}. \tag{4.27}$$

This model subspace is well defined for any positive δ. We will see that our bound on the reconstruction error depends on δ. Note that we define δ only as a theoretical tool, since the model subspace is not involved methodologically when computing the MAP estimator, $\hat{u}$. We emphasize that in the setting of ℓ^q sparsity, we have that $u^\star_{S_\delta^C} \neq 0$ in general, and consequently the model error terms in the general bounds are non-negligible. We now present a deterministic result analogous to Theorem 4.4.1 under the ℓ^q-sparsity assumption.

Theorem 4.4.3. *Let $u^\star$ be the true parameter generating data via* (4.1) *and suppose that $u^\star \in \mathbb{B}_q(R_q)$. Let $(\hat{u}, \hat{\theta})$ be the unique minimizer of the objective function* (4.7) *with $0 < \eta < 1/2$. Assume that $\lambda \geq \frac{2\|A^T\varepsilon\|_\infty}{n}$ and that A satisfies the RSC condition* (4.24) *with curvature κ and tolerance τ^2. Then, for any $\delta > 0$, if $\tau\sqrt{R_q\delta^{-q}} \leq \frac{\sqrt{\kappa}}{32}$ it holds that*

$$\|\hat{u} - u^\star\|_2^2 \lesssim \frac{\lambda^2}{\kappa^2} R_q \delta^{-q} + \frac{\tau^2}{\kappa} R_q \delta^{1-q} + \frac{\lambda}{\kappa} R_q \delta^{1-q} + \mathcal{E}_\eta,$$

where

$$\mathcal{E}_\eta := \frac{\tau^2}{\kappa}\Big(\eta^2 R_q^{2/q} + d^2\eta^2(2 - \log\eta)^2\Big) + \frac{\lambda}{\kappa}\Big(\eta R_q^{1/q} + d\eta(2 - \log\eta)\Big).$$

Since this deterministic error bound holds for any positive δ, Theorem 4.4.3 provides a

family of bounds parameterized by δ. Under noise and forward map assumptions that yield particular values for λ, κ, and τ, one can optimize δ to yield the sharpest bound. We do so in the following corollary under the assumptions of Gaussian noise and a Gaussian forward map.

Corollary 4.4.4. *Let the forward map $A \in \mathbb{R}^{n\times d}$ have rows distributed i.i.d. according to $\mathcal{N}(0,\Sigma)$, and assume the columns of A are normalized such that $\frac{\|A_j\|_2}{\sqrt{n}} = 1$ for all $j = 1,\dots,d$. Assume the noise is given by $\varepsilon \sim \mathcal{N}(0,I)$ and that $(\hat{u},\hat{\theta})$ is the unique minimizer of the IAS objective functional (4.7) with $\lambda = 4\sqrt{\frac{\log d}{n}}$ and $0<\eta<1/2$. If $u^\star \in \mathbb{B}_q(R_q)$, then there exists a constant c_Σ depending only on Σ, and universal constants c and c' such that if $n > c_\Sigma \log d R_q^{1-q/2}$ it holds with probability at least $1 - c\exp(-c'n\lambda^2)$ that*

$$\|\hat{u}-u^\star\|_2^2 \lesssim c_\Sigma R_q\left(\frac{\log d}{n}\right)^{1-\frac{q}{2}} + c_\Sigma \min\left\{\varphi(\eta),\varphi(\eta)^2\right\},$$

where

$$\varphi(\eta) := \sqrt{\frac{\log d}{n}}\left(\eta R_q^{\frac{1}{q}} + \frac{d\eta}{\sqrt{2}}(2-\log\eta)\right).$$

This corollary suggests that η should be taken small enough so that $\varphi(\eta)$ matches the rate of $R_q\left(\frac{\log d}{n}\right)^{1-\frac{q}{2}}$. In particular, for any η such that

$$\eta\log\frac{1}{\eta} \le \min\left\{\frac{R_q}{d}\left(\frac{\log d}{n}\right)^{\frac{1-q}{2}}, R_q^{1-\frac{1}{q}}\left(\frac{\log d}{n}\right)^{\frac{1-q}{2}}\right\},$$

the MAP estimator of the hierarchical Bayesian model in Example 4.2.3 achieves up to constant factors the minimax rate for estimation under ℓ^q sparsity [286], given by

$$\|\hat{u}-u^\star\|_2^2 \lesssim R_q\left(\frac{\log d}{n}\right)^{1-\frac{q}{2}}.$$

That is, with this choice of η, the MAP estimator achieves the best possible reconstruction error.

4.4.2 Hierarchical Models for Group Sparsity

In this subsection we derive non-asymptotic bounds on the reconstruction error for the group-sparsity-promoting hierarchical model considered in Example 4.2.5. We prove reconstruction bounds under a weak ℓ^q-group-sparsity assumption. We suppose that

$$u^\star \in \mathbb{B}_q(R_q, \mathcal{G}) := \left\{ u \in \mathbb{R}^d : \sum_{j=1}^{k} \|u_{g_j}\|_{C_j}^q \leq R_q \right\}.$$

We emphasize that in the case where $q = 0$ and $R_q = s$, we recover the hard-group-sparsity assumption that $u_g = 0$ for all $g \in S_{\mathcal{G}}$ with $|S_{\mathcal{G}}| = s$. We consider the set of group indices with norm larger than $\delta > 0$, defined by

$$S_{\mathcal{G},\delta} := \left\{ g_j \in \mathcal{G} : \|u^\star_{g_j}\|_{C_j} > \delta \right\}. \tag{4.28}$$

We now present a deterministic result that holds for any positive δ so long as the regularization parameter is taken to be large enough relative to the noise and the forward map satisfies an RSC condition.

Theorem 4.4.5. *Let $u^\star$ be the true parameter generating the data via (4.4) and suppose that $u^\star \in \mathbb{B}_q(R_q, \mathcal{G})$. Let $(\hat{u}, \hat{\theta})$ be the minimizer of the objective function (4.11) with $0 < \eta < 1/2$. Assume that $\lambda \geq \max_{g_j \in \mathcal{G}} \frac{2}{n} \|(A^T \varepsilon)_{g_j}\|_{C_j^{-1}}$, and that A satisfies the RSC condition (4.24) with curvature κ and tolerance τ^2. Then, for any $\delta > 0$, if $\tau \sqrt{R_q \delta^{-q}} \leq \frac{\sqrt{\kappa}}{32}$ it holds that*

$$\|\hat{u} - u^\star\|_2^2 \lesssim \frac{\lambda^2}{\kappa^2} R_q \delta^{-q} + \frac{\tau^2}{\kappa} R_q \delta^{1-q} + \frac{\lambda}{\kappa} R_q \delta^{1-q} + \mathcal{E}_\eta,$$

where

$$\mathcal{E}_\eta := \frac{\tau^2}{\kappa}\left(\eta^2 R_q^{\frac{2}{q}} + k^2\eta^2(2-\log\eta)^2\right) + \frac{\lambda}{\kappa}\left(\eta R_q^{\frac{1}{q}} + k\eta(2-\log\eta)\right).$$

We remark that if $q = 0$ and $R_q = s$, this theorem recovers bounds for the hard-sparsity assumption that $u^\star \in \mathcal{M}(S_\mathcal{G})$:

$$\|\hat{u} - u^\star\|_2^2 \lesssim s\frac{\lambda^2}{\kappa^2} + \mathcal{E}_\eta,$$

where

$$\mathcal{E}_\eta := \frac{\tau^2}{\kappa}\left(\eta^2\left(\sum_{j\in S_\mathcal{G}} \|u^\star_{g_j}\|_{C_j}\right)^2 + k^2\eta^2(2-\log\eta)^2\right) + \frac{\lambda}{\kappa}\left(\eta\sum_{j\in S_\mathcal{G}} \|u^\star_{g_j}\|_{C_j} + k\eta(2-\log\eta)\right).$$

As was the case for the models discussed in the previous section, under distributional assumptions on the noise and forward map, we can use this deterministic bound to sharply characterize non-asymptotic bounds on the reconstruction error with high probability. For simplicity of presentation, we assume that $C_j = I$ for $1 \le j \le k$. This assumption is straightforward to remove, but doing so is notationally cumbersome. We require a block-normalization condition on A that generalizes the column-normalization assumption used in the previous subsection. Given a group $g_j \in \mathcal{G}$ of size p_j, we consider the submatrix $A_{g_j} : \mathbb{R}^{n\times p_j}$, and require that

$$\frac{\|A_{g_j}\|_{op}}{\sqrt{n}} = 1, \quad j \in \{1,\dots,k\}. \tag{4.29}$$

As remarked in Example 4.2.5, this condition can be guaranteed by appropriately choosing the model hyperparameters. For the following corollary, we introduce the notation $p_{\max} := \max_{j\in\{1,\dots,k\}} p_j$.

Corollary 4.4.6. *Let the forward map $A \in \mathbb{R}^{n\times d}$ have rows distributed i.i.d. according to $\mathcal{N}(0,\Sigma)$ and assume that the block normalization condition (4.29) holds. Assume that the noise is given by $\varepsilon \sim \mathcal{N}(0,I)$ and that $(\hat{u},\hat{\theta})$ is the unique minimizer of the GS-IAS objective functional (4.11) with $\lambda = \left(\sqrt{\frac{p_{\max}}{n}} + \sqrt{\frac{\log k}{n}}\right)$ and $0 < \eta < 1/2$. If $u^\star \in \mathbb{B}_q(R_q,\mathcal{G})$, then there exists a constant c_Σ depending only on Σ and such that if $\sqrt{n} > c_\Sigma R_q^{1-q/2}\left(\sqrt{p_{\max}} + \sqrt{\log k}\right)$ it holds with probability at least $1-2/k^2$ that*

$$\|\hat{u}-u^\star\|_2^2 \lesssim c_\Sigma R_q\left(\frac{p_{\max}}{n}+\frac{\log k}{n}\right)^{1-\frac{q}{2}} + \min\left\{\varphi(\eta),\varphi(\eta)^2\right\},$$

where

$$\varphi(\eta) := \sqrt{\frac{p_{\max}+\log k}{n}}\left(\eta R_q^{\frac{1}{q}} + k\eta(2-\log\eta)\right).$$

If $q=0$ and $R_q = s$, we recover a bound under a hard-group-sparsity assumption:

$$\|\hat{u}-u^\star\|_2^2 \lesssim \frac{sp_{\max}}{n}+\frac{s\log k}{n} + \max\left\{\varphi(\eta),\varphi(\eta)^2\right\},$$

where

$$\varphi(\eta) := \sqrt{\frac{p_{\max}+\log k}{n}}\left(\eta\sum_{j=1}^{k}\|u^\star_{g_j}\|_{C_j} + k\eta(2-\log\eta)\right).$$

In both the ℓ^q-group-sparsity and hard-group-sparsity settings, by taking η to be small enough to match the estimation rate in the first half of the bound, we see that the hierarchical model in Example 4.2.5 and the GS-IAS algorithm can achieve the known bounds for reconstructing group sparse vectors [265, 247, 199].

4.4.3 Hierarchical Models for Sparse Representations

Finally, we apply our theory to derive non-asymptotic reconstruction error bounds for the hierarchical model promoting sparse representations described in Example 4.2.7. We assume that the data-generating unknown $u^\star$ has a representation in a tight frame $W \in \mathbb{R}^{d\times k}$, where k may be much larger than d. We assume this representation lies in an ℓ^q ball:

$$u^\star \in \mathbb{B}_q(R_q, W) := \left\{ u \in \mathbb{R}^d : \sum_{j=1}^{k} |(W^T u)_j|^q \leq R_q \right\}.$$

As in the previous examples, we recover the hard-sparsity assumption that the support of $W^T u^\star$ has cardinality s in the case where $q = 0$ and $R_q = s$. We define our model subspace as in (4.22), taking the index set to be

$$S_{W,\delta} := \left\{ j \in \{1, \dots, k\} : (W^T u^\star)_j > \delta \right\}. \tag{4.30}$$

We show a deterministic bound on the reconstruction error that holds for any positive δ provided that the regularization parameter is large enough relative to the noise and that the forward map satisfies the RSC condition.

Theorem 4.4.7. *Let $u^\star$ be the true parameter generating the data via* (4.4) *and suppose that $u^\star \in \mathbb{B}_q(R_q, W)$. Let $(\hat{u}, \hat{\theta})$ be the minimizer of the objective function* (4.15) *with $0 < \eta < 1/2$. Assume that $\lambda \geq \frac{2\|W^T A^T \varepsilon\|_\infty}{n}$ and that A satisfies the RSC condition* (4.24) *with curvature κ and tolerance τ^2. Then, for any $\delta > 0$, if $\tau\sqrt{R_q \delta^{-q}} \leq \frac{\kappa}{32}$ it holds that*

$$\|\hat{u} - u^\star\|_2^2 \lesssim \frac{\lambda^2}{\kappa^2} R_q \delta^{-q} + \frac{\tau^2}{\kappa} R_q \delta^{1-q} + \frac{\lambda}{\kappa} R_q \delta^{1-q} + \mathcal{E}_\eta,$$

where

$$\mathcal{E}_\eta := \frac{\tau^2}{\kappa}\left(\eta^2 R_q^{2/q} + k^2\eta^2(2-\log(\eta)^2\right) + \frac{\lambda}{\kappa}\left(\eta R_q^{1/q} + k\eta(2-\log\eta)\right).$$

When $q = 0$ and $R_q = s$ this result recovers a bound under the hard-sparsity assumption that $\mathrm{supp}(W^T u^\star) = S$ with $|S| = s$:

$$\|\hat{u} - u^\star\|_2^2 \lesssim s\frac{\lambda^2}{\kappa^2} + \mathcal{E}_\eta,$$

where

$$\mathcal{E}_\eta := \frac{\tau^2}{\kappa}\left(\eta^2\|(W^T u^\star)_S\|_1^2 + k^2\eta^2(2-\log\eta)^2\right) + \frac{\lambda}{\kappa}\left(\eta\|(W^T u^\star)_S\|_1 + k\eta(2-\log\eta)\right).$$

Once again, by making assumptions on the distribution of the noise and forward map, we can derive sharp non-asymptotic bounds on the reconstruction error with high probability.

Corollary 4.4.8. *Let the forward map $A \in \mathbb{R}^{n\times d}$ have rows distributed i.i.d. according to $\mathcal{N}(0,\Sigma)$ and assume that A is normalized such that $\frac{\|(AW)_j\|_2}{\sqrt{n}} = 1$ for all $j = 1,\dots,k$. Assume the noise is given by $\varepsilon \sim \mathcal{N}(0, I)$ and that $(\hat{u}, \hat{\theta})$ is the minimizer of the O-IAS objective functional (4.15) with $\lambda = 4\sqrt{\frac{\log k}{n}}$ and $0 < \eta < 1/2$. If $u^\star \in \mathbb{B}_Q(R_q, Q)$, then there exists a constant c_Σ depending only on Σ and universal constants c and c' such that if $n > c_\Sigma R_q^{1-q/2}\log k$ it holds with probability at least $1 - c\exp(c' n\lambda^2)$ that*

$$\|\hat{u} - u^\star\|_2^2 \lesssim R_q\left(\frac{\log k}{n}\right)^{1-\frac{q}{2}} + \min\left\{\varphi(\eta), \varphi(\eta)^2\right\},$$

where

$$\varphi(\eta) := \sqrt{\frac{\log k}{n}}\left(\eta R_q^{1/q} + k\eta(2-\log\eta)\right).$$

The corresponding hard-sparsity bound under the assumption that $q = 0$ and $R_q = s$ is

$$\|\hat{u} - u^\star\|_2^2 \lesssim \frac{s \log k}{n} + \max\left\{\varphi(\eta), \varphi(\eta)^2\right\},$$

where

$$\varphi(\eta) := \sqrt{\frac{\log k}{n}} \left(\eta \|W^T u^\star\|_1 + k\eta(2 - \log \eta)\right).$$

4.5 Proof of Reconstruction Error for Sparsity-Promoting Models

This section contains the proofs of all the results in Section 4.4.

4.5.1 Hierarchical Models for Sparsity

Proof of Theorem 4.4.1. The result follows from applying Theorem 4.3.11. From Lemma 4.3.3 we have that approximate decomposability holds with $c_1(\eta) = \frac{\eta}{2}$ and $c_2(\eta) = \frac{d\eta}{\sqrt{2}}(2 - \log \eta)$. For the model subspace $\mathcal{M}(S)$ defined in (4.20), the subspace Lipschitz constant is given by

$$\Psi\big(\mathcal{M}(S)\big) = \sup_{u \in \mathcal{M}(S)} \frac{\|u\|_1}{\|u\|_2} = \sqrt{s}.$$

The dual of the one norm is the infinity norm, so the assumption that $\lambda \geq \frac{2\|A^T \varepsilon\|_\infty}{n}$ satisfies the condition of Theorem 4.3.11. Since we have assumed $u^\star \in \mathcal{M}(S)$, we have that $u^\star_{\mathcal{M}^\perp} = 0$, and thus $\mathsf{R}(u^\star_{\mathcal{M}^\perp}) = 0$. Plugging each of these pieces into the bound (4.25) gives the desired result. □

Proof of Corollary 4.4.2. First we verify that the given choice of λ satisfies the assumption on Theorem 4.4.1 with high probability. It is shown in [265, Corollary 2] that under the column normalization assumption, for any ε that is (sub-)Gaussian with identity covariance,

it holds that

$$\mathbb{P}\left(\frac{\|A^T\varepsilon\|_\infty}{n} \geq t\right) \leq 2\exp\left(-\frac{nt^2}{2} + \log d\right).$$

Taking $t^2 = \frac{4\log d}{n}$, we conclude that the choice of $\lambda = 4\sqrt{\frac{\log d}{n}}$ satisfies the assumption of Theorem 4.4.1 with probability $1 - c_1\exp(-c_2 n\lambda^2)$. Next we characterize the curvature and tolerance in the RSC condition for a Gaussian forward map. By [342, Theorem 7.16], for any matrix $A \in \mathbb{R}^{n\times d}$ with rows i.i.d. from $\mathcal{N}(0,\Sigma)$ there exist positive constants $c_1 < 1 < c_2$ such that, with probability at least $1 - \frac{e^{-n/32}}{1-e^{-n/32}}$,

$$\frac{\|A\Delta\|_2^2}{2n} \geq c_1\lambda_{\min}(\Sigma)\|\Delta\|_2^2 - c_2\rho(\Sigma)^2\frac{\log d}{n}\|\Delta\|_1^2, \qquad \forall\Delta \in \mathbb{R}^d,$$

where $\rho(\Sigma)^2 := \max_{1\leq j\leq d}\Sigma_{jj}$. In other words, the RSC condition holds with curvature $\kappa = c_1\lambda_{\min}(\Sigma)$ and tolerance $\tau^2 = c_2\rho(\Sigma)^2\frac{\log d}{n}$. Consequently, to guarantee that $\tau\sqrt{s} \leq \frac{\sqrt{\kappa}}{32}$ as assumed in Theorem 4.4.1, we require that

$$\sqrt{n} \geq 32\sqrt{\frac{\frac{c_1}{c_2}\rho(\Sigma)^2 s\log d}{\lambda_{\min}(\Sigma)}}.$$

Given this requirement on n we have satisfied the hypotheses of Theorem 4.4.1. Consequently, we plug in our characterizations of λ, κ, and τ into (4.26) and the desired result follows. □

Proof of Theorem 4.4.3. From Lemma 4.3.3 we have that approximate decomposability holds with $c_1(\eta) = \frac{\eta}{2}$ and $c_2(\eta) = \frac{d\eta}{\sqrt{2}}(2 - \log\eta)$. For the model subspace $\mathcal{M}(S_\delta)$ where S_δ is given by (4.27), the subspace Lipschitz constant is then given by

$$\Psi\big(\mathcal{M}(S_\delta)\big) = |S_\delta|^{1/2}.$$

As such, we need to characterize the cardinality of S_δ. We remark that

$$R_q \geq \sum_{j=1}^{d} |u_j^\star|^q \geq \sum_{j \in S_\delta} |u_j^\star|^q \geq \delta^q |S_\delta|.$$

It follows that

$$|S_\delta| \leq R_q \delta^{-q}.$$

Under ℓ^q sparsity, we do not have that $u^\star_{S_\delta^C} = 0$. However, we have the bound

$$\|u^\star_{S_\delta^C}\|_1 = \sum_{j \in S_\delta^c} |u_j^\star|^q |u_j^\star|^{1-q} \leq R_q \delta^{1-q},$$

since $|u_j^\star| \leq \delta$ for all $j \in S_\delta^C$. Finally, since $q < 1$, we have that

$$\|u^\star_{S_\delta}\|_1 \leq R_q^{1/q}.$$

Hence, we apply Theorem 4.3.11 and the desired result follows. □

Proof of Corollary 4.4.4. Exactly as in the proof of Corollary 4.4.2, the choice of $\lambda = 4\sqrt{\frac{\log d}{n}}$ satisfies the assumption that $\lambda \geq \frac{2\|A^T\varepsilon\|_\infty}{n}$ with probability $1 - c_1 \exp(-c_2 n\lambda^2)$. Again, exactly as in Corollary 4.4.2, the RSC condition holds with curvature $\kappa = c_1 \lambda_{\min}(\Sigma)$ and tolerance $\tau^2 = c_2 \rho(\Sigma)^2 \frac{\log d}{n}$ with probability at least $1 - \frac{e^{-n/32}}{1-e^{-n/32}}$. We pick δ in the definition of S_δ to be $\delta = \frac{\lambda}{\lambda_{\min}(\Sigma)}$ and plugging each of the above equations into Theorem 4.4.3 yields the desired result. □

4.5.2 Hierarchical Models for Group Sparsity

Proof of Theorem 4.4.5. From Lemma 4.3.5 we have that approximate decomposability holds with $c_1(\eta) = \frac{\eta}{2}$ and $c_2(\eta) = \frac{k\eta}{\sqrt{2}}(2 - \log \eta)$. For the model subspace $\mathcal{M}(S_{\mathcal{G},\delta})$, where $S_{\mathcal{G},\delta}$ is given in (4.28), the subspace Lipschitz constant is given by

$$\Psi\big(\mathcal{M}(S_{\mathcal{G},\delta})\big) = \sup_{u \in \mathcal{M}(S_{\mathcal{G},\delta}) \setminus \{0\}} \frac{\sum_{j=1}^{k} \|u_{g_j}\|_{C_j}}{\|u\|_2} = \frac{1}{\sqrt{|S_{\mathcal{G},\delta}|}} \sum_{j \in S_{\mathcal{G},j}} \sqrt{\frac{1}{\lambda_{\min}(C_j)}}. \tag{4.31}$$

Note that if $C_j = I$ for all j, this reduces to $\Psi\big(\mathcal{M}(S_{\mathcal{G},\delta})\big) = \sqrt{|S_{\mathcal{G},\delta}|}$. We can upper bound the subspace Lipschitz constant as

$$\Psi\big(\mathcal{M}(S_{\mathcal{G},\delta})\big) \leq \sqrt{\frac{1}{\lambda_{\min}(C_{j^*})} \big|S_{\mathcal{G},\delta}\big|}, \tag{4.32}$$

where $j^* = \arg\max_{j \in S_{\mathcal{G},\delta}} \frac{1}{\lambda_{\min}(C_j)}$. If the matrices C_j are all equal, then the upper bound (4.32) is an equality. We will use this upper bound in what follows, but we remark that if the spectrum of the matrices C_j varies drastically between groups of variables, one may be able to derive sharper bounds by working with (4.31) directly. The dual norm of our group sparse norm is given by

$$\mathsf{R}^*(u) := \sup_{v \neq 0} \frac{\langle u, v \rangle}{\sum_{j=1}^{k} \|v_{g_j}\|_{C_j}} = \max_{g_j \in \mathcal{G}} \frac{2}{n} \|u_{g_j}\|_{C_j^{-1}},$$

and we see that our assumption on λ satisfies the condition of Theorem 4.3.11. As in the proof of Theorem 4.4.3, one can show that

$$\begin{aligned} \left|S_{\mathcal{G},\delta}\right| &\leq R_q \delta^{-q}, \\ \mathsf{R}\big(u^\star_{\mathcal{M}(S^C_{\mathcal{G},\delta})}\big) &= \sum_{j \in S^C_{\mathcal{G},\delta}} \|u^\star_{g_j}\|_{C_j} \leq R_q \delta^{1-q}, \\ \mathsf{R}\big(u^\star_{\mathcal{M}(S_{\mathcal{G},\delta})}\big) &= \sum_{j \in \mathcal{M}(S_{\mathcal{G},\delta})} \|u^\star_{g_j}\|_{C_j} \leq R_q^{\frac{1}{q}}. \end{aligned}$$

These bounds along with Theorem 4.3.11 give the desired result. □

Proof of Corollary 4.4.6. We first want to verify that the choice of $\lambda = 2\left(\sqrt{\frac{p_{\max}}{n}} + \sqrt{\frac{\log k}{n}}\right)$ satisfies

$$\lambda \geq \max_{g_j \in \mathcal{G}} \frac{2}{n} \|(A^T \varepsilon)_{g_j}\|_2 \tag{4.33}$$

with high probability. We claim that if A satisfies the block-normalization condition (4.29) and $\varepsilon \sim \mathcal{N}(0, I)$, then we have

$$\mathbb{P}\left[\frac{1}{n} \max_{1 \leq j \leq k} \|X^T_{g_j} \varepsilon\|_2 \geq 2\left(\sqrt{\frac{p_{\max}}{n}} + \sqrt{\frac{\log k}{n}}\right)\right] \leq 2\exp\left(2 - \log k\right). \tag{4.34}$$

For a fixed group $g_j \in \mathcal{G}$ of size p_j, we consider the submatrix $A_{g_j} \in \mathbb{R}^{n \times p_j}$. We characterize the deviation above the mean using the same argument as [266] to get that, for all $t > 0$,

$$\mathbb{P}\left[\frac{1}{n}\|X^T_{g_j}\varepsilon\|_2 \geq \mathbb{E}\left[\frac{1}{n}\|X^T_{g_j}\varepsilon\|_2\right] + t\right] \leq 2\exp\left(-\frac{nt^2}{2}\right). \tag{4.35}$$

The expectation in this bound can be bounded as follows

$$\mathbb{E}\left[\frac{1}{n}\|X^T_{g_j}\varepsilon\|_2\right] \overset{\text{(i)}}{\leq} \sqrt{\frac{2}{n}}\mathbb{E}\left[\sqrt{\varepsilon^T\varepsilon}\right] \overset{\text{(ii)}}{\leq} \sqrt{\frac{2p_{\max}}{n}}, \tag{4.36}$$

where for (i) we use the same Gaussian comparison principle argument as [266] and for (ii) we use Jensen's inequality and that $\varepsilon \sim \mathcal{N}(0, I)$. Combining (4.35) and (4.36) gives that, for all $t > 0$,

$$\mathbb{P}\left[\frac{1}{n}\|X_{g_j}^T \varepsilon\|_2 \geq \sqrt{\frac{2p_{\max}}{n}} + t\right] \leq 2\exp\left(-\frac{nt^2}{2}\right).$$

Applying a union bound over all $j \in \{1, \ldots, k\}$ and setting $t^2 = \frac{4\log k}{n}$, we deduce that (4.34) holds as claimed. Hence, our choice of λ satisfies the bound (4.33) with probability at least $1 - 2/k^2$. From [266, Proposition 1], we conclude that the RSC condition holds with curvature $\kappa = \frac{1}{4}\lambda_{\min}(\Sigma)$ and tolerance $\tau^2 = 9\max_{g_j \in \mathcal{G}} \|(\Sigma_{g_j}^{1/2}\|_{op}\left(\sqrt{\frac{p_{\max}}{n}} + \sqrt{\frac{\log k}{n}}\right)^2$. Applying Theorem 4.4.5 with $\delta \asymp \lambda$ gives the desired result. □

4.5.3 Hierarchical Models for Sparse Representations

Proof of Theorem 4.4.7. From Lemma 4.3.7 we have that approximate decomposability holds with $c_1(\eta) = \frac{\eta}{2}$ and $c_2(\eta) = \frac{k\eta}{\sqrt{2}}(2 - \log\eta)$. For the model subspace $\mathcal{M}(S_{W,\delta})$, where $S_{\mathcal{G},\delta}$ is given in (4.30), we have that the subspace Lipschitz constant is given by

$$\Psi\big(\mathcal{M}(S_{W,\delta})\big) = \sup_{u \in \mathcal{M}(S_{W,\delta})\setminus\{0\}} \frac{\|W^T u\|_1}{\|u\|_2} = \sqrt{|S_{W,\delta}|}.$$

The dual of $\|W^T \cdot\|_1$ is given by $\|W^T \cdot\|_\infty$, so our assumption on λ satisfies the condition of Theorem 4.3.11. By the same arguments as in the previous settings, we get the bounds

$$\begin{aligned}
|S_{W,\delta}| &\leq R_q \delta^{-q}, \\
R(u^\star_{\mathcal{M}(S^C_{W,\delta})}) &= \|W^T u^\star_{\mathcal{M}(S^C_{W,\delta})}\|_1 \leq R_q \delta^{1-q}, \\
R(u^\star_{\mathcal{M}(S_{W,\delta})}) &= \|u^\star_{\mathcal{M}(S_{W,\delta})}\|_1 \leq R_q^{1/q}.
\end{aligned}$$

These bounds along with Theorem 4.3.11 give the desired result. □

Proof of Corollary 4.4.8. By the normalization assumption and the fact that ε is Gaussian, we have that, for each $j = 1, \ldots, k$,

$$\mathbb{P}\Big[\big|\langle (AW)_j, \varepsilon\rangle\big| \geq t\Big] \leq 2\exp\left(-\frac{nt^2}{2}\right).$$

A union bound then gives that

$$\mathbb{P}\left[\frac{\|W^T A^T \varepsilon\|_\infty}{n} \geq t\right] \leq 2\exp\left(-\frac{nt^2}{2}\right),$$

and taking $t^2 = \frac{4\log k}{n}$ we see that the choice of $\lambda = 4\sqrt{\frac{\log k}{n}}$ satisfies the assumption of Theorem 4.4.7. The proof of [285, Theorem 1] can be adapted to show that, with probability at least $1 - \frac{e^{-n/32}}{1-e^{-n/32}}$,

$$\frac{\|A\Delta\|_2}{\sqrt{n}} \geq \frac{1}{4}\lambda_{\min}(\Sigma) - 9\rho(W^T\Sigma W)^2\frac{\log k}{n}\|W^T\Delta\|_1^2, \qquad \forall \Delta \in \mathbb{R}^d,$$

where $\rho(W^T\Sigma W)^2 := \max_{1\leq j\leq k}(W^T\Sigma W)_{jj}$. Hence, the RSC condition holds with curvature $\kappa = \frac{1}{4}\lambda_{\min}(\Sigma)$ and tolerance $\tau^2 = 9\rho(W^T\Sigma W)^2\frac{\log k}{n}$. Picking $\delta \asymp \lambda$ and applying Theorem 4.4.7 yields the desired result. □

4.6 Conclusions

This paper has put forth a unified perspective on sparsity-promoting conditionally Gaussian hypermodels as M-estimators, bridging the fields of inverse problems and high-dimensional statistics. We have introduced the notion of approximate decomposability, which quantifies how close the regularization term resulting from a Bayesian hierarchical model is to a decomposable regularizer. We have then generalized the theory in [265] to the setting of

approximately decomposable regularizers, and proved the first known reconstruction error bounds for MAP estimators of conditionally Gaussian hypermodels. In particular, we have obtained bounds for widely-used hierarchical models to reconstruct sparse and group-sparse parameters, as well as for a novel model which promotes sparse representations. Our bounds suggest that the minimax rate can be achieved provided that the hyperpameter η is taken sufficiently small.

This work opens a number of important questions for future research. As noted above, from a statistical perspective, the hyperparameter η should be taken as small as possible to promote sparsity. However, from a computational perspective, the positivity of η enforces the strict convexity of the objective function, and taking η to be too small may degrade the convergence of the IAS algorithm used to compute the MAP estimators. As such, an important avenue for research is to investigate the statistical-computational tradeoff in the choice of η, as well as computational comparisons between the Bayesian hypermodels and frequentist M-estimators that also admit efficient and well-studied optimization algorithms such as ISTA [136], FISTA [26], and ADMM [54].

The connection between conditionally Gaussian hypermodels and M-estimators can be leveraged to motivate the design of new hierarchical Bayesian models and algorithms inspired by M-estimator theory. We conjecture that hierarchical models can be developed for tasks such as low-rank matrix estimation, sparse generalized linear models, and sparsity-promoting dictionary learning, and that these models will inherit the same convenient computational and statistical properties as models considered in this paper. Likewise, our theory could be extended to the broader family of generalized gamma hyperpriors, for which greedy algorithms for efficient MAP estimation have been developed in [79, 78].

The analysis in this work relies heavily on the forward map satisfying the RSC condition (4.24). This condition holds for Gaussian forward maps, which may not be a realistic assumption for many inverse problems of interest. As such, verifying the RSC condition

(possibly after preconditioning) for important forward maps in inverse problems, such as deconvolution or tomography [209], is an important direction for further research.

The hierarchical models considered here are formulated in the high-dimensional yet finite-dimensional setting. Extensions of these models to the infinite-dimensional function space setting and analysis thereof remains an avenue for future research. Techniques from recent work [4, 2, 12] on extending results from finite-dimensional compressed sensing and high-dimensional statistics to the function space setting may be adapted to study infinite-dimensional formulations of the models considered here. Further, this work considers hierarchical models that use a centered parameterization. Non-centered parameterizations can facilitate the scalability of sampling algorithms in high-dimensional problems [350, 108], and MAP estimation is typically not invariant under changes in parameterization [139]. It remains an open direction for future work to extend the models considered here to non-centered parameterizations and subsequently develop reconstruction error bounds in that setting.

4.7 Computational Properties of GS-IAS and O-IAS

Proof of Theorem 4.2.6. Lemma 2.2 in [75] shows that the objective function (4.11) is strictly convex over $\mathbb{R}^d \times \mathbb{R}^k_+$ in the special case where each group is of size $p_j = 3$. The proof holds as written for non-overlapping groups of arbitrary sizes. We denote the Hessian of the objective function (4.11) as

$$H(u,\theta) = \begin{bmatrix} \nabla^2_{uu}\mathsf{J}(u,\theta) & \nabla^2_{u\theta}\mathsf{J}(u,\theta) \\ \nabla^2_{\theta u}\mathsf{J}(u,\theta) & \nabla^2_{\theta\theta}\mathsf{J}(u,\theta) \end{bmatrix}.$$

Let $e^\ell := u^\ell - \hat{u}$. To show linear convergence, we apply Lemma 4.1 in [84], which gives

$$e^{\ell+1} = \nabla^2_{uu}\mathsf{J}(\hat{u},\hat{\theta})^{-1}\nabla^2_{u\theta}\mathsf{J}(\hat{u},\hat{\theta})\nabla^2_{\theta\theta}\mathsf{J}(\hat{u},\hat{\theta})^{-1}\nabla^2_{\theta u}\mathsf{J}(\hat{u},\hat{\theta})e^\ell + o(\|e^\ell\|^2),$$

where each of the blocks of the Hessian are evaluated at the MAP estimator $(\hat{u}, \hat{\theta})$. Denoting by $\mathrm{diag}(B_j)$ the block-diagonal matrix with blocks $B_j, 1 \leq j \leq k$, we thus obtain that

$$\begin{aligned} e^{\ell+1} &= \left(\frac{A^T A}{n} + \lambda D_{\hat{\theta}}^{-1} \right)^{-1} \mathrm{diag}(\lambda \hat{\theta}_j C_j^{-1} \hat{u}_{g_j}) \times \\ &\qquad \times \mathrm{diag}(\lambda \hat{\theta}_j^{-3} \|\hat{u}_{g_j}\|_{C_j}^2 + \hat{\theta}_j^{-2} \eta)^{-1} \mathrm{diag}(\lambda \hat{\theta}_j C_j^{-1} \hat{u}_{g_j}) e^{\ell} + o(\|e^{\ell}\|^2) \\ &= D_{\hat{\theta}}^{1/2} \left(\frac{D_{\hat{\theta}}^{1/2} A^T A D_{\hat{\theta}}^{1/2}}{n\lambda} + I \right)^{-1} \mathrm{diag}(\hat{\theta}_j^{-3/2} C_j^{-1/2} \hat{u}_{g_j}) \times \\ &\qquad \times \mathrm{diag}(\hat{\theta}_j^{-3} \|\hat{u}_{g_j}\|_{C_j}^2 + \hat{\theta}_j^{-2} \eta)^{-1} \mathrm{diag}(\hat{\theta}_j C_j^{-1} \hat{u}_{g_j}) e^{\ell} + o(\|e^{\ell}\|^2). \end{aligned}$$

Multiplying both sides by $D_{\theta}^{-1/2}$, taking norms, and neglecting higher order terms, we get that

$$\|e^{\ell+1}\|_{D_{\hat{\theta}}} \leq \left\| \mathrm{diag}\left(\frac{C_j^{1/2} \hat{u}_{g_j} \hat{u}_{g_j}^T C_j^{1/2}}{\|\hat{u}_{g_j}\|_{C_j}^2 + \hat{\theta}_j \eta} \right) \right\|_{op} \|e^{\ell}\|_{D_{\hat{\theta}}}.$$

Since for every $1 \leq j \leq k$ it holds that $\|C_j^{1/2} \hat{u}_{g_j} \hat{u}_{g_j}^T C_j^{1/2}\|_{op} \leq \|\hat{u}_{g_j}\|_{C_j}^2$ and that $\hat{\theta}_j \eta > 0$, we have linear convergence in the Mahalanobis norm.

To show that $\hat{u}$ is the global minimizer of (4.13), we refer to Theorem 2.1 in [75], which proves the result for the special case of each group having size $p_j = 3$. Their proof still holds in our more general setting. □

Proof of Theorem 4.2.8. We first show that the O-IAS objective function $\mathsf{J}(z)$, where $z = (u, \theta) \in \mathbb{R}^d \times \mathbb{R}_+^k$, given in (4.15) is strictly convex over $\mathbb{R}^d \times \mathbb{R}_+^k$. We let $\tilde{\mathsf{J}}(\tilde{z})$ denote the IAS objective given in (4.7) over $\tilde{z} = (\tilde{u}, \theta) \in \mathbb{R}^k \times \mathbb{R}_+^k$, with the forward map $\tilde{A} = AW$. From [84], we know that $\tilde{\mathsf{J}}(\tilde{z})$ is strictly convex over $\mathbb{R}^k \times \mathbb{R}_+^k$. We then observe that $\mathsf{J}(z) = \tilde{\mathsf{J}}(C\tilde{z})$,

where

$$C = \begin{bmatrix} W^T & 0 \\ 0 & I \end{bmatrix} \in \mathbb{R}^{2k \times d+k}.$$

Since this matrix has trivial nullspace and strict convexity is preserved under such a transformation, we conclude that $\mathsf{J}(z)$ is strictly convex. Since $\mathsf{J}(z)$ tends to infinity as $\theta_j \to 0$, $\theta_j \to \infty$, and $\|u\| \to \infty$, a minimizer $(\hat{u}, \hat{\theta})$ exists and is unique. As in the proof of Theorem 4.2.6, we let $e^\ell = u^\ell - \hat{u}$ and apply Lemma 4.1 in [84], which gives

$$e^{\ell+1} = \nabla^2_{uu}\mathsf{J}(\hat{u},\hat{\theta})^{-1}\nabla^2_{u\theta}\mathsf{J}(\hat{u},\hat{\theta})\nabla^2_{\theta\theta}\mathsf{J}(\hat{u},\hat{\theta})^{-1}\nabla^2_{\theta u}\mathsf{J}(\hat{u},\hat{\theta})e^\ell + o(\|e^\ell\|^2),$$

where each of the blocks of the Hessian are evaluated at the MAP estimator $(\hat{u}, \hat{\theta})$. Plugging in the expressions of the Hessian blocks, we get that

$$\begin{aligned} e^{\ell+1} = &\left(\frac{A^TA}{n} + \lambda W D_{\hat{\theta}}^{-1} W^T\right)^{-1} W \times \\ &\times \operatorname{diag}(\lambda W^T\hat{u}/\hat{\theta}^2)\left(\operatorname{diag}(\lambda(W^T\hat{u})^2/\hat{\theta}^3 + \eta/\hat{\theta}^2)\right)^{-1}\operatorname{diag}(\lambda W^T\hat{u}/\hat{\theta}^2)W^Te^\ell + o(\|e^\ell\|^2). \end{aligned}$$

Multiplying both sides by $\operatorname{diag}(1/\hat{\theta}^{1/2})W^T$, we get that

$$\begin{aligned} \operatorname{diag}(\hat{\theta}^{-1/2})W^Te^{\ell+1} = \operatorname{diag}(\hat{\theta}^{-1/2})W^T&\left(\frac{A^TA}{n} + \lambda W D_{\hat{\theta}}^{-1} W^T\right)^{-1} W \times \\ &\times \operatorname{diag}\left(\frac{\lambda(W^T\hat{u})^2}{(W^T\hat{u})^2\hat{\theta}^{1/2} + \eta\hat{\theta}^{3/2}}\right)\operatorname{diag}(\hat{\theta}^{-1/2})W^Te^\ell + o(\|e^\ell\|^2). \end{aligned}$$

Taking norms of both sides, we see that if

$$\mu = \left\|\operatorname{diag}(\hat{\theta}^{-1/2})W^T\left(\frac{A^TA}{n} + \lambda W D_{\hat{\theta}}^{-1} W^T\right)^{-1} W\operatorname{diag}\left(\frac{\lambda(W^T\hat{u})^2}{(W^T\hat{u})^2\hat{\theta}^{1/2} + \eta\hat{\theta}^{3/2}}\right)\right\| < 1,$$

then we have linear convergence in the Mahalanobis norm. Letting $W\text{diag}(1/\hat{\theta})W^T = ZZ$, where $Z \in \mathbb{R}^{d\times d}$ is the unique matrix square root of $W\text{diag}(1/\hat{\theta})W^T$, we have that

$$\begin{aligned}
\mu &= \left\| \text{diag}(1/\hat{\theta}^{1/2})W^T Z^{-1} \left(\frac{Z^{-1}A^T A Z^{-1}}{\lambda n} + I \right)^{-1} Z^{-1} W \text{diag}\left(\frac{(W^T\hat{u})^2}{(W^T\hat{u})^2\hat{\theta}^{1/2} + \eta\hat{\theta}^{3/2}} \right) \right\| \\
&\leq \left\| \text{diag}(1/\hat{\theta}^{1/2})W^T Z^{-1} Z^{-1} W \text{diag}\left(\frac{(W^T\hat{u})^2}{(W^T\hat{u})^2\hat{\theta}^{1/2} + \eta\hat{\theta}^{3/2}} \right) \right\| \\
&= \left\| \text{diag}(1/\hat{\theta}^{1/2})W^T \left(W\text{diag}(1/\hat{\theta})W^T \right)^{-1} W \text{diag}\left(\frac{(W^T\hat{u})^2}{(W^T\hat{u})^2\hat{\theta}^{1/2} + \eta\hat{\theta}^{3/2}} \right) \right\|.
\end{aligned}$$

Since $WW^T = I$, the SVD of W must have the form

$$W = U \begin{bmatrix} I & 0 \end{bmatrix} V^T = U\tilde{V}^T,$$

where $\tilde{V} \in \mathbb{R}^{k\times d}$ is the matrix containing the first d columns of V. Plugging this expression into our upper bound for μ, we get that

$$\begin{aligned}
\mu &\leq \left\| \text{diag}(1/\hat{\theta}^{1/2})\tilde{V}U^T \left(U\tilde{V}^T \text{diag}(1/\hat{\theta})\tilde{V}U^T \right)^{-1} U\tilde{V}^T \text{diag}\left(\frac{(W^T\hat{u})^2}{(W^T\hat{u})^2\hat{\theta}^{1/2} + \eta\hat{\theta}^{3/2}} \right) \right\| \\
&= \left\| \text{diag}(1/\hat{\theta}^{1/2})\tilde{V} \left(\tilde{V}^T \text{diag}(1/\hat{\theta})\tilde{V} \right)^{-1} \tilde{V}^T \text{diag}\left(\frac{(W^T\hat{u})^2}{(W^T\hat{u})^2\hat{\theta}^{1/2} + \eta\hat{\theta}^{3/2}} \right) \right\|.
\end{aligned}$$

Making the substitution $V' = \text{diag}(1/\hat{\theta}^{1/2})$, we get that

$$\mu \leq \left\| V' \left(V'^T V' \right)^{-1} V'^T \text{diag}\left(\frac{(W^T\hat{u})^2}{(W^T\hat{u})^2 + \eta\hat{\theta}} \right) \right\|.$$

Observing that $V' \left(V'^T V' \right)^{-1} V'^T$ is an orthogonal projector onto the range of V', we have

that

$$\mu \leq \left\| \text{diag}\left(\frac{(W^T\hat{u})^2}{(W^T\hat{u})^2 + \eta\hat{\theta}} \right) \right\| < 1.$$

Consequently, we have shown linear convergence of the O-IAS algorithm in the Mahalanobis norm.

We will now show that the minimizer of (4.15) over u and θ can be written as the minimizer of $\mathsf{J}\big(u, f(u)\big)$, where

$$f_j(u) = \left(\frac{\eta}{2} + \sqrt{\frac{\eta^2}{4} + \frac{(W^T u)_j}{2}} \right).$$

Since the minimizer $(\hat{u}, \hat{\theta})$ is a critical point, it follows that

$$\begin{aligned} \nabla_u \mathsf{J}(\hat{u}, \hat{\theta}) &= \left(\frac{A^T A}{n} + \lambda W \text{diag}(1/\hat{\theta}) W^T \right) \hat{u} - \frac{A^T y}{n} = 0, \\ \nabla_\theta \mathsf{J}(\hat{u}, \hat{\theta}) &= \sum_{j=1}^{k} \left(\frac{(W^T\hat{u})_j^2}{2\hat{\theta}_j^2} - \frac{\eta}{\hat{\theta}_j} - 1 \right) e_j = 0. \end{aligned} \tag{4.37}$$

Given $\hat{u}$, we solve (4.37) for $\hat{\theta}$ in terms of $\hat{u}$. Some algebra shows that the solution is given by $\hat{\theta} = f(\hat{u})$, and thus $\hat{u}$ is a critical point of

$$\mathsf{F}(u) = \mathsf{J}\big(u, f(u)\big). \tag{4.38}$$

If $\tilde{u}$ is a critical point of (4.38), letting $\tilde{\theta} = f(\tilde{u})$, we get that $(\tilde{u}, \tilde{\theta})$ is a critical point of $\mathsf{J}(u, \theta)$ and by uniqueness of the minima we have that $\hat{u} = \tilde{u}$, completing the proof. □

Proof of Lemma 4.3.3. By the triangle inequality, it holds that

$$\frac{|u_j|}{\sqrt{2}} \le f_j(u) \le \eta + \frac{|u_j|}{\sqrt{2}}, \quad \text{and} \quad \eta \le f_j(u). \tag{4.39}$$

Using these inequalities, we deduce that

$$\begin{aligned}
\mathsf{R}_\eta(u) &= \frac{1}{2\sqrt{2}} \sum_{j=1}^{d} \left[\frac{u_j^2}{f_j(u)} + 2f_j(u) - 2\eta \log f_j(u) \right] \\
&\le \frac{1}{2\sqrt{2}} \sum_{j=1}^{d} \left[\sqrt{2}|u_j| + 2\eta + \sqrt{2}|u_j| - 2\eta \log(\eta) \right] \le \|u\|_1 + \frac{d\eta}{\sqrt{2}}(1 - \log \eta),
\end{aligned}$$

proving the upper bound in (4.21). For the lower bound, we have

$$\begin{aligned}
\mathsf{R}_\eta(u) &\overset{\text{(i)}}{\ge} \frac{1}{2\sqrt{2}} \sum_{j=1}^{d} \left[\frac{u_j^2}{\eta + \frac{|u_j|}{\sqrt{2}}} + \sqrt{2}|u_j| - 2\eta \log\left(\eta + \frac{|u_j|}{\sqrt{2}} \right) \right] \\
&\overset{\text{(ii)}}{\ge} \frac{1}{2\sqrt{2}} \sum_{j=1}^{d} \left[\frac{u_j^2}{\eta + \frac{|u_j|}{\sqrt{2}}} + \sqrt{2}|u_j| - 2\eta \frac{|u_j|}{\sqrt{2}} \right] \\
&\overset{\text{(iii)}}{\ge} \frac{1}{2\sqrt{2}} \sum_{j=1}^{d} \left[\sqrt{2}|u_j| - 2\eta + \sqrt{2}|u_j| - \sqrt{2}\eta|u_j| \right] = \left(1 - \frac{\eta}{2}\right)\|u\|_1 - \frac{d\eta}{\sqrt{2}},
\end{aligned}$$

where (i) follows by the inequalities in (4.39), (ii) follows using that for $0 < \eta < 1$ it holds that $\log\left(\eta + \frac{|u_j|}{\sqrt{2}}\right) \le \frac{|u_j|}{\sqrt{2}}$, and (iii) follows using that $\frac{u_j^2}{\eta + \frac{|u_j|}{\sqrt{2}}} \ge \sqrt{2}|u_j| - 2\eta$. □

Proof of Lemma 4.3.5. Similar to the proof of Lemma 4.3.3, replacing $|u_j|$ with $\|u_{g_j}\|_{C_j}$. □

Proof of Lemma 4.3.7. The statement follows from Lemma 4.3.3 applied to $W^T u \in \mathbb{R}^k$. □

4.9 Approximate Decomposability and General Bounds

In this appendix we prove the key property of approximately decomposable regularizers in Lemma 4.3.8 and the general reconstruction bound in Theorem 4.3.11. To control the reconstruction error $\hat{\Delta} = \hat{u} - u^\star$ between the minimizer of the objective F in (4.23) and the true parameter $u^\star$ generating the data according to (4.4), we introduce a function $\mathcal{F} : \mathbb{R}^d \to \mathbb{R}$ given by

$$\begin{aligned} \mathcal{F}(\Delta) :&= \mathsf{F}(u^\star + \Delta) - \mathsf{F}(u^\star) \\ &= \frac{1}{2n}\Big[\|y - A(u^\star + \Delta)\|_2^2 - \|y - Au^\star\|_2^2\Big] + \lambda\Big[\mathsf{R}(u^\star + \Delta) - \mathsf{R}(u^\star)\Big]. \end{aligned} \tag{4.40}$$

By construction, $\mathcal{F}(0) = 0$. Moreover, for $\hat{\Delta} = \hat{u} - u^\star$ we have the *basic inequality*

$$\mathcal{F}(\hat{\Delta}) = \mathsf{F}(\hat{u}) - \mathsf{F}(u^\star) \leq 0, \tag{4.41}$$

since $\hat{u}$ is the minimizer of F. We will leverage this inequality to control the reconstruction error.

Proof of Lemma 4.3.8. We first prove a lower bound on the difference of the regularizer evaluated at the minimizer and at the true parameter. Using approximate decomposability of R_η, reverse triangle inequality, and decomposability of R over $\mathcal{M}$, it follows that

$$\begin{aligned} \mathsf{R}_\eta(\hat{u}) &\geq \Big(1 - c_1^L(\eta)\Big)\, \mathsf{R}(u^\star + \hat{\Delta}) - c_2^L(\eta) \\ &\geq \Big(1 - c_1^L(\eta)\Big)\, \Big(\mathsf{R}(u^\star_{\mathcal{M}} + \hat{\Delta}_{\mathcal{M}^\perp}) - \mathsf{R}(u^\star_{\mathcal{M}^\perp}) - \mathsf{R}(\hat{\Delta}_{\mathcal{M}})\Big) - c_2^L(\eta) \\ &\geq \Big(1 - c_1^L(\eta)\Big)\, \Big(\mathsf{R}(u^\star_{\mathcal{M}}) + \mathsf{R}(\hat{\Delta}_{\mathcal{M}^\perp}) - \mathsf{R}(u^\star_{\mathcal{M}^\perp}) - \mathsf{R}(\hat{\Delta}_{\mathcal{M}})\Big) - c_2^L(\eta). \end{aligned} \tag{4.42}$$

Similarly, using approximate decomposability of R_η and decomposability of R, we obtain

$$\mathsf{R}_\eta(u^\star) \leq \Big(1 + c_1^U(\eta)\Big)\, \Big(\mathsf{R}(u^\star_{\mathcal{M}}) + \mathsf{R}(u^\star_{\mathcal{M}^\perp})\Big) + c_2^U(\eta). \tag{4.43}$$

Combining (4.42) and (4.43) gives

$$\mathsf{R}_\eta(\hat{u}) - \mathsf{R}_\eta(u^\star) \geq \left(1 - c_1^L(\eta)\right)\left(\mathsf{R}(\hat{\Delta}_{\mathcal{M}^\perp}) - \mathsf{R}(\hat{\Delta}_{\mathcal{M}})\right) - c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) - 2\mathsf{R}(u^\star_{\mathcal{M}^\perp}) - c_2(\eta), \tag{4.44}$$

where $c_i(\eta) = c_i^L(\eta) + c_i^U(\eta)$ and we used that $c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}^\perp}) \geq 0$. We next show a lower bound on the difference of the likelihood evaluated at the minimizer and at the true parameter. By direct calculation, it holds that

$$\frac{1}{2n}\left[\|y - A\hat{u}\|_2^2 - \|y - Au^\star\|_2^2\right] = \frac{1}{2n}\|A\hat{\Delta}\|_2^2 - \frac{1}{n}\langle A^T\varepsilon, \hat{\Delta}\rangle \geq -\frac{1}{n}\left|\langle A^T\varepsilon, \hat{\Delta}\rangle\right|.$$

Then, using Hölder's inequality with the regularizer and its dual, we get that

$$\begin{aligned}\frac{1}{2n}\left[\|y - A\hat{u}\|_2^2 - \|y - Au^\star\|_2^2\right] &\geq -\frac{1}{n}\mathsf{R}^*(A^T\varepsilon)\mathsf{R}(\hat{\Delta}) = -\frac{1}{n}\mathsf{R}^*(A^T\varepsilon)\left(\mathsf{R}(\hat{\Delta}_{\mathcal{M}}) + \mathsf{R}(\hat{\Delta}_{\mathcal{M}^\perp})\right) \\ &\geq -\frac{\lambda}{2}\left(\mathsf{R}(\hat{\Delta}_{\mathcal{M}}) + \mathsf{R}(\hat{\Delta}_{\mathcal{M}^\perp})\right),\end{aligned} \tag{4.45}$$

where the last inequality follows by our assumption on λ. Finally, (4.44) and (4.45) combined with the basic inequality (4.41) give

$$\begin{aligned}0 \geq \mathcal{F}(\hat{\Delta}) \geq \lambda\Bigg(&\left(1 - c_1^L(\eta)\right)\left(\mathsf{R}(\hat{\Delta}_{\mathcal{M}^\perp}) - \mathsf{R}(\hat{\Delta}_{\mathcal{M}})\right) - c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) - 2\mathsf{R}(u^\star_{\mathcal{M}^\perp}) - c_2(\eta)\Bigg) \\ &- \frac{\lambda}{2}\left(\mathsf{R}(\hat{\Delta}_{\mathcal{M}}) + \mathsf{R}(\hat{\Delta}_{\mathcal{M}^\perp})\right),\end{aligned}$$

which after rearranging implies that

$$\left(\frac{1}{2} - c_1^L(\eta)\right)\mathsf{R}(\hat{\Delta}_{\mathcal{M}^\perp}) \leq \left(\frac{3}{2} + c_1^L(\eta)\right)\mathsf{R}(\hat{\Delta}_{\mathcal{M}}) + 2\mathsf{R}(u^\star_{\mathcal{M}^\perp}) + c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) + c_2(\eta).$$

The desired result follows from the assumption that $c_1^L(\eta) < \frac{1}{4}$. □

Letting $\delta > 0$ be an error radius that we will specify later, we define the set

$$\mathbb{K}(\delta) := \mathbb{C}_\eta(\mathcal{M}) \cap \{\|\Delta\|_2 = \delta\}.$$

The following result from [266, Lemma 4] demonstrates that it is sufficient to show a positive lower bound for the function $\mathcal{F}$ given in (4.40) over the set $\mathbb{K}(\delta)$.

Lemma 4.9.1. *If $\mathcal{F}(\Delta) > 0$ for all vectors $\Delta \in \mathbb{K}(\delta)$, then $\|\hat{\Delta}\|_2 \leq \delta$.*

Proof. The proof in [266] still holds essentially as written in our slightly modified setting. In particular, the proof does not require that the regularizers in $\mathcal{F}(\Delta)$ be decomposable. The regularizers in our definition of $\mathbb{C}_\eta$ are decomposable with respect to $\mathcal{M}$, and thus the argument this set is "star shaped" (see [266] or [342, Chapter 9] for a definition and discussion of this property) holds as written. □

We are now ready to prove Theorem 4.3.11.

Proof of Theorem 4.3.11. Fix a radius $\delta > 0$, which we will specify at the conclusion of the proof. By Lemma 4.9.1, we proceed by deriving a positive lower bound for $\mathcal{F}(\Delta)$ over $\mathbb{K}(\delta)$, which will yield a bound on $\|\hat{\Delta}\|_2^2$. By definition of $\mathcal{F}$, the RSC condition, and the inequality (4.44)

$$\begin{aligned}
\mathcal{F}(\Delta) &= \frac{1}{2n}\Big[\|y - A(u^\star + \Delta)\|_2^2 - \|y - Au^\star\|_2^2\Big] + \lambda\left[\mathsf{R}(u^\star + \Delta) - \mathsf{R}(u^\star)\right] \\
&\geq \langle A^T\varepsilon, \Delta\rangle + \frac{\kappa}{2}\|\Delta\|_2^2 - \tau^2\mathsf{R}^2(\Delta) + \lambda\left[\mathsf{R}(u^\star + \Delta) - \mathsf{R}(u^\star)\right] \\
&\geq \langle A^T\varepsilon, \Delta\rangle + \frac{\kappa}{2}\|\Delta\|_2^2 - \tau^2\mathsf{R}^2(\Delta) \\
&\quad + \lambda\Big[\Big(1 - c_1^L(\eta)\Big)\Big(\mathsf{R}(\hat{\Delta}_{\mathcal{M}^\perp}) - \mathsf{R}(\hat{\Delta}_{\mathcal{M}})\Big) - c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) - 2\mathsf{R}(u^\star_{\mathcal{M}^\perp}) - c_2(\eta)\Big].
\end{aligned}$$

Applying Hölder's inequality and using the assumption on λ, we have that $|\langle A^T\varepsilon, \Delta\rangle| \leq$

$\frac{\lambda}{2}\mathsf{R}(\Delta)$, which gives

$$\begin{aligned}
\mathcal{F}(\Delta) &\geq \frac{\kappa}{2}\|\Delta\|_2^2 - \tau^2\mathsf{R}^2(\Delta) - \frac{\lambda}{2}\mathsf{R}(\Delta) \\
&+ \lambda\Big[\Big(1 - c_1^L(\eta)\Big)\Big(\mathsf{R}(\hat{\Delta}_{\mathcal{M}^\perp}) - \mathsf{R}(\hat{\Delta}_{\mathcal{M}})\Big) - c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) - 2\mathsf{R}(u^\star_{\mathcal{M}^\perp}) - c_2(\eta)\Big] \\
&= \frac{\kappa}{2}\|\Delta\|_2^2 - \tau^2\mathsf{R}^2(\Delta) + \lambda\Bigg[\left(1 - c_1^L(\eta) - \frac{1}{2}\right)\mathsf{R}(\hat{\Delta}_{\mathcal{M}^\perp}) \\
&- \left(1 - c_1^L(\eta) + \frac{1}{2}\right)\mathsf{R}(\hat{\Delta}_{\mathcal{M}}) - c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) - 2\mathsf{R}(u^\star_{\mathcal{M}^\perp}) - c_2(\eta)\Bigg],
\end{aligned}$$

where the equality follows from the decomposability of R. Notice that, for any $\Delta \in \mathbb{C}_\eta(\mathcal{M})$,

$$\begin{aligned}
\mathsf{R}^2(\Delta) = \big(\mathsf{R}(\Delta_{\mathcal{M}}) + \mathsf{R}(\Delta_{\mathcal{M}^\perp})\big)^2 &\leq 16\big(2\Psi(\mathcal{M})\|\Delta\|_2 + 2\mathsf{R}(u^\star_{\mathcal{M}^\perp}) + c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) + c_2(\eta)\big)^2 \\
&\leq 64\big[4\Psi^2(\mathcal{M})\|\Delta\|_2^2 + 4\mathsf{R}^2(u^\star_{\mathcal{M}^\perp}) + c_1^2(\eta)\mathsf{R}^2(u^\star_{\mathcal{M}}) + c_2^2(\eta)\big].
\end{aligned}$$

Plugging this bound into the previous inequality, we get

$$\begin{aligned}
\mathcal{F}(\Delta) &\geq \frac{\kappa}{2}\|\Delta\|_2^2 - 64\tau^2\big[4\Psi^2(\mathcal{M})\|\Delta\|_2^2 + 4\mathsf{R}^2(u^\star_{\mathcal{M}^\perp}) + c_1^2(\eta)\mathsf{R}^2(u^\star_{\mathcal{M}}) + c_2^2(\eta)\big] \\
&\quad - \lambda\left[\left(1 - c_1^L(\eta) + \frac{1}{2}\right)\mathsf{R}(\hat{\Delta}_{\mathcal{M}}) + c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) + 2\mathsf{R}(u^\star_{\mathcal{M}^\perp}) + c_2(\eta)\right] \\
&\geq \frac{\kappa}{2}\|\Delta\|_2^2 - 64\tau^2\big[4\Psi^2(\mathcal{M})\|\Delta\|_2^2 + 4\mathsf{R}^2(u^\star_{\mathcal{M}^\perp}) + c_1^2(\eta)\mathsf{R}^2(u^\star_{\mathcal{M}}) + c_2^2(\eta)\big] \\
&\quad - \lambda\left[\left(1 - c_1^L(\eta) + \frac{1}{2}\right)\Psi(\mathcal{M})\|\Delta\|_2 + c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) + 2\mathsf{R}(u^\star_{\mathcal{M}^\perp}) + c_2(\eta)\right].
\end{aligned}$$

By our assumption that $\tau\Psi(\mathcal{M}) \leq \frac{\sqrt{\kappa}}{32}$, it follows that

$$\begin{aligned}
\mathcal{F}(\Delta) &\geq \frac{\kappa}{4}\|\Delta\|_2^2 - 64\tau^2\left[4\mathsf{R}^2(u^\star_{\mathcal{M}^\perp}) + c_1^2(\eta)\mathsf{R}^2(u^\star_{\mathcal{M}}) + c_2^2(\eta)\right] \\
&\quad - \lambda\left[\left(1 - c_1^L(\eta) + \frac{1}{2}\right)\Psi(\mathcal{M})\|\Delta\|_2 + c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) + 2\mathsf{R}(u^\star_{\mathcal{M}^\perp}) + c_2(\eta)\right].
\end{aligned}$$

The right-hand side of this inequality is a quadratic polynomial in $\|\Delta\|_2$ that is positive for

$\|\Delta\|$ sufficiently large. In particular, the quadratic formula shows that taking

$$\delta = 72\frac{\lambda^2}{\kappa^2}\Psi^2(\mathcal{M}) \\ + \frac{8}{\kappa}\left(64\tau^2\left[4\mathsf{R}^2(u^\star_{\mathcal{M}^\perp}) + c_1^2(\eta)\mathsf{R}^2(u^\star_{\mathcal{M}}) + c_2^2(\eta)\right] + \lambda\left[c_1(\eta)\mathsf{R}(u^\star_{\mathcal{M}}) + 2\mathsf{R}(u^\star_{\mathcal{M}^\perp}) + c_2(\eta)\right]\right)$$

guarantees a positive lower bound on $\mathcal{F}(\Delta)$ over $\mathbb{K}(\delta)$, completing the proof. $\square$

REFERENCES

[1] H. Abarbanel, *Predicting The Future: Completing Models Of Observed Complex Systems*, Springer, 2013.

[2] B. Adcock, V. Antun, and A. C. Hansen, *Uniform recovery in infinite-dimensional compressed sensing and applications to structured binary sampling*, Applied and Computational Harmonic Analysis, 55 (2021), pp. 1–40.

[3] B. Adcock, S. Brugiapaglia, and C. G. Webster, *Sparse polynomial approximation of high-dimensional functions*, vol. 25, SIAM, 2022.

[4] B. Adcock and A. C. Hansen, *Generalized sampling and infinite-dimensional compressed sensing*, Foundations of Computational Mathematics, 16 (2016), pp. 1263–1323.

[5] M. Adrian, D. Sanz-Alonso, and R. Willett, *Data assimilation with machine learning surrogate models: A case study with fourcastnet*, Artificial Intelligence for the Earth Systems, 4 (2025), p. e240050.

[6] S. Agapiou, J. M. Bardsley, O. Papaspiliopoulos, and A. M. Stuart, *Analysis of the Gibbs sampler for hierarchical inverse problems*, SIAM/ASA Journal on Uncertainty Quantification, 2 (2014), pp. 511–544.

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

[8] A. Agarwal, S. Negahban, and M. J. Wainwright, *Fast global convergence rates of gradient methods for high-dimensional statistical recovery*, Advances in Neural Information Processing Systems, 23 (2010).

[9] S. Agrawal, H. Kim, D. Sanz-Alonso, and A. Strang, *A variational inference approach to inverse problems with gamma hyperpriors*, SIAM/ASA Journal on Uncertainty Quantification, 10 (2022), pp. 1533–1559.

[10] O. Al-Ghattas, J. Bao, and D. Sanz-Alonso, *Ensemble Kalman filters with resampling*, SIAM/ASA Journal on Uncertainty Quantification, 12 (2024), pp. 411–441.

[11] O. Al-Ghattas, J. Chen, D. Sanz-Alonso, and N. Waniorek, *Optimal estimation of structured covariance operators*, arXiv preprint arXiv:2408.02109, (2024).

[12] ——, *Covariance operator estimation: Sparsity, lengthscale, and ensemble kalman filters*, Bernoulli, 31 (2025), pp. 2377–2402.

[13] O. Al-Ghattas and D. Sanz-Alonso, *Non-asymptotic analysis of ensemble Kalman updates: effective dimension and localization*, Information and Inference: A Journal of the IMA, 13 (2024), p. iaad043.

[14] J. L. Anderson, *An ensemble adjustment Kalman filter for data assimilation*, Monthly Weather Review, 129 (2001), pp. 2884–2903.

[15] ——, *An adaptive covariance inflation error correction algorithm for ensemble filters*, Tellus A: Dynamic meteorology and oceanography, 59 (2007), pp. 210–224.

[16] H. Antil and A. K. Saibaba, *Efficient algorithms for Bayesian inverse problems with Whittle–Matérn priors*, SIAM Journal on Scientific Computing, (2023), pp. S176–S198.

[17] R. Arcucci, J. Zhu, S. Hu, and Y.-K. Guo, *Deep data assimilation: integrating deep learning with data assimilation*, Applied Sciences, 11 (2021), p. 1114.

[18] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, *Solving inverse problems using data-driven models*, Acta Numerica, 28 (2019), pp. 1–174.

[19] 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), p. 025005.

[20] A. Azouani, E. Olson, and E. S. Titi, *Continuous data assimilation using general interpolant observables*, Journal of Nonlinear Science, 24 (2014), pp. 277–304.

[21] I. Babuska, R. Tempone, and G. E. Zouraris, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM Journal on Numerical Analysis, 42 (2004), pp. 800–825.

[22] E. Bach, R. Baptista, D. Sanz-Alonso, and A. Stuart, *Inverse Problems and Data Assimilation: A Machine Learning Approach*, arXiv preprint arXiv:2410.10523, (2024).

[23] F. R. Bach, *Consistency of the group Lasso and multiple kernel learning*, Journal of Machine Learning Research, 9 (2008), pp. 1179–1225.

[24] J. F. Bard, *Practical Bilevel Optimization: Algorithms and Applications*, vol. 30, Springer Science & Business Media, 2013.

[25] J. M. Bardsley, D. Calvetti, and E. Somersalo, *Hierarchical regularization for edge-preserving reconstruction of PET images*, Inverse Problems, 26 (2010), p. 035010.

[26] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202.

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

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

[29] M. Benning and M. Burger, *Modern regularization methods for inverse problems*, Acta Numerica, 27 (2018), pp. 1–111.

[30] K. Bergemann and S. Reich, *An ensemble kalman-bucy filter for continuous data assimilation*, Meteorologische Zeitschrift, 127 (2012), pp. 1417–1440.

[31] T. Berry and J. Harlim, *Variable bandwidth diffusion kernels*, Applied and Computational Harmonic Analysis, 40 (2016), pp. 68–96.

[32] T. Berry and T. Sauer, *Local kernels and the geometric structure of data*, Applied and Computational Harmonic Analysis, 40 (2016), pp. 439–469.

[33] M. Bertalmio, L.-T. Cheng, S. Osher, and G. Sapiro, *Variational problems and partial differential equations on implicit surfaces*, Journal of Computational Physics, 174 (2001), pp. 759–780.

[34] A. L. Bertozzi, X. Luo, A. M. Stuart, and K. C. Zygalakis, *Uncertainty quantification in graph-based classification of high dimensional data*, SIAM/ASA Journal on Uncertainty Quantification, 6 (2018), pp. 568–595.

[35] A. Beskos, A. Jasra, K. Law, R. Tempone, and Y. Zhou, *Multilevel Sequential Monte Carlo Samplers*, Stochastic Processes and their Applications, 127 (1994), pp. 1417–1440.

[36] H. Bessaih, E. Olson, and E. S. Titi, *Continuous data assimilation with stochastically noisy data*, Nonlinearity, 28 (2015), p. 729.

[37] J. Bezdek, R. Hathaway, R. Howard, C. Wilson, and M. Windham, *Local convergence analysis of a grouped variable version of coordinate descent*, Journal of Optimization Theory and Applications, 54 (1987), pp. 471–477.

[38] K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, *Accurate medium-range global weather forecasting with 3d neural networks*, Nature, 619 (2023), pp. 533–538.

[39] P. Bickel, B. Li, and T. Bengtsson, *Pushing the limits of contemporary statistics: Contributions in honor of jayanta k. ghosh: Sharp failure rates for the bootstrap particle filter in high dimensions*, Institute of Mathematical Statistics, (1994), pp. 318–329.

[40] P. J. Bickel, J. B. Brown, H. Huang, and Q. Li, *An overview of recent developments in genomics and associated statistical methods*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 367 (2009), pp. 4313–4337.

[41] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, *Simultaneous analysis of Lasso and Dantzig selector*, The Annals of Statistics, 37 (2009), pp. 1705–1732.

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

[43] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[44] C. H. Bishop, B. J. Etherton, and S. J. Majumdar, *Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects*, Monthly Weather Review, 129 (2001), pp. 420–436.

[45] A. Biswas and M. Branicki, *A unified framework for the analysis of accuracy and stability of a class of approximate gaussian filters for the navier–stokes equations*, Nonlinearity, 37 (2024), p. 125013.

[46] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, *Variational inference: A review for statisticians*, Journal of the American Statistical Association, 112 (2017), pp. 859–877.

[47] A. Bocchinfuso, D. Calvetti, and E. Somersalo, *Adaptive anisotropic Bayesian meshing for inverse problems*, Inverse Problems, 40 (2024), p. 045008.

[48] A. Bocchinfuso, D. Calvetti, and E. Somersalo, *Bayesian sparsity and class sparsity priors for dictionary learning and coding*, Journal of Computational Mathematics and Data Science, 11 (2024), p. 100094.

[49] M. Bocquet, *Surrogate modeling for the climate sciences dynamics with machine learning and data assimilation*, Frontiers in Applied Mathematics and Statistics, 9 (2023), p. 1133226.

[50] M. Bocquet, J. Brajard, A. Carrassi, and L. Bertino, *Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization*, Foundations of Data Science, 2 (2020), pp. 55–80.

[51] V. I. Bogachev, *Gaussian Measures*, no. 62, American Mathematical Soc., 1998.

[52] D. Bolin and K. Kirchner, *The rational SPDE approach for Gaussian random fields with general smoothness*, Journal of Computational and Graphical Statistics, 29 (2020), pp. 274–285.

[53] A. Bonito, J. M. Cascón, K. Mekchay, P. Morin, and R. H. Nochetto, *High-order afem for the laplace–beltrami operator: Convergence rates*, Foundations of Computational Mathematics, 16 (2016), pp. 1473–1539.

[54] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al., *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends® in Machine learning, 3 (2011), pp. 1–122.

[55] A. Bracco, J. Brajard, H. A. Dijkstra, P. Hassanzadeh, C. Lessig, and C. Monteleoni, *Machine learning for the physics of climate*, Nature Reviews Physics, (2024), pp. 1–15.

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

[57] J. Brajard, A. Carrassi, M. Bocquet, and L. Bertino, *Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model*, Journal of Computational Science, 44 (2020), p. 101171.

[58] M. Branicki, A. J. Majda, and K. J. Law, *Accuracy of some approximate Gaussian filters for the Navier–Stokes equation in the presence of model error*, Multiscale Modeling & Simulation, 16 (2018), pp. 1756–1794.

[59] K. Bredies, M. Carioni, M. Holler, Y. Korolev, and C.-B. Schönlieb, *A sparse optimization approach to infinite infimal convolution regularization*, Numerische Mathematik, 157 (2025), pp. 41–96.

[60] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, vol. 3, Springer.

[61] C. E. Brett, K. F. Lam, K. Law, D. McCormick, M. R. Scott, and A. Stuart, *Accuracy and stability of filters for dissipative PDEs*, Physica D: Nonlinear Phenomena, 245 (2013), pp. 34–45.

[62] J. Bröcker, *Existence and uniqueness for four-dimensional variational data assimilation in discrete time*, SIAM Journal on Applied Dynamical Systems, 16 (2013), pp. 361–374.

[63] S. Brooks, A. Gelman, G. Jones, and X. Meng, *Handbook of Markov chain Monte Carlo*, CRC press, (2011).

[64] S. L. Brunton, J. L. Proctor, and J. N. Kutz, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proceedings of the National Academy of Sciences, 113 (2016), pp. 3932–3937.

[65] P. Bühlmann and S. van De Geer, *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer Science & Business Media, 2011.

[66] 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 Journal on Scientific Computing, 35 (2013), pp. A2494–A2523.

[67] F. Bunea and L. Xiao, *On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fpca*, Bernoulli, (2015), pp. 1200–1230.

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

[69] M. Burger, *Variational regularization in inverse problems and machine learning*, arXiv preprint arXiv:2112.04591, (2021), p. 24.

[70] T. Cai and W. Liu, *Adaptive thresholding for sparse covariance matrix estimation*, Journal of the American Statistical Association, 106 (2011), pp. 672–684.

[71] E. Calvello, E. Carlson, N. Kovachki, M. N. Manta, and A. M. Stuart, *Operator learning for smoothing and forecasting*, arXiv preprint arXiv:2603.20359, (2026).

[72] E. Calvello, P. Monmarché, A. M. Stuart, and U. Vaes, *Accuracy of the ensemble kalman filter in the near-linear setting*, SIAM Journal on Numerical Analysis, 64 (2026), pp. 391–429.

[73] E. Calvello, S. Reich, and A. M. Stuart, *Ensemble kalman methods: A mean-field perspective*, Acta Numerica, 34 (2025), pp. 123–291.

[74] D. Calvetti, H. Hakula, S. Pursiainen, and E. Somersalo, *Conditionally Gaussian hypermodels for cerebral source localization*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 879–909.

[75] D. Calvetti, A. Pascarella, F. Pitolli, E. Somersalo, and B. Vantaggi, *A hierarchical Krylov–Bayes iterative inverse solver for MEG with physiological preconditioning*, Inverse Problems, 31 (2015), p. 125005.

[76] ——, *Brain activity mapping from MEG data via a hierarchical Bayesian algorithm with automatic depth weighting*, Brain topography, 32 (2019), pp. 363–393.

[77] D. Calvetti, F. Pitolli, E. Somersalo, and B. Vantaggi, *Bayes meets Krylov: Statistically inspired preconditioners for CGLS*, SIAM Review, 60 (2018), pp. 429–461.

[78] D. Calvetti, M. Pragliola, and E. Somersalo, *Sparsity promoting hybrid solvers for hierarchical Bayesian inverse problems*, SIAM Journal on Scientific Computing, 42 (2020), pp. A3761–A3784.

[79] D. Calvetti, M. Pragliola, E. Somersalo, and A. Strang, *Sparse reconstructions from few noisy data: analysis of hierarchical Bayesian models with generalized gamma hyperpriors*, Inverse Problems, 36 (2020), p. 025010.

[80] D. Calvetti and E. Somersalo, *An Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing*, vol. 2, Springer Science & Business Media, 2007.

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

[82] ——, *Bayesian Scientific Computing*, vol. 215, Springer Nature, 2023.

[83] ——, *Computationally efficient sampling methods for sparsity promoting hierarchical Bayesian models*, SIAM/ASA Journal on Uncertainty Quantification, 12 (2024), pp. 524–548.

[84] D. Calvetti, E. Somersalo, and A. Strang, *Hierachical Bayesian models and sparsity: L^2-magic*, Inverse problems, 35 (2019), p. 035003.

[85] F. Camacho and A. Demlow, *L2 and pointwise a posteriori error estimates for fem for elliptic pdes on surfaces*, IMA Journal of Numerical Analysis, 35 (2015), pp. 1199–1227.

[86] E. Candes and B. Recht, *Exact matrix completion via convex optimization*, Communications of the ACM, 55 (2012), pp. 111–119.

[87] E. Candes and B. Recht, *Simple bounds for recovering low-complexity models*, Mathematical Programming, 141 (2013), pp. 577–589.

[88] E. J. Candes and D. L. Donoho, *New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 57 (2004), pp. 219–266.

[89] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, *Compressed sensing with coherent and redundant dictionaries*, Applied and Computational Harmonic Analysis, 31 (2011), pp. 59–73.

[90] Y. Canzani, *Analysis on manifolds via the Laplacian*, Lecture Notes available at: http://www. math. harvard. edu/canzani/docs/Laplacian. pdf, (2013).

[91] E. Carlson, A. Farhat, V. R. Martinez, and C. Victor, *Determining modes, synchronization, and intertwinement*, arXiv preprint arXiv:2408.01064, (2024).

[92] ——, *On the infinite-nudging limit of the nudging filter for continuous data assimilation*, arXiv preprint arXiv:2408.02646, (2024).

[93] E. Carlson, A. Larios, and E. S. Titi, *Super-exponential convergence rate of a nonlinear continuous data assimilation algorithm: The 2d navier–stokes equation paradigm*, Journal of Nonlinear Science, 34 (2024), p. 37.

[94] A. Carrassi, M. Bocquet, L. Bertino, and G. Geir, *Data assimilation in the geosciences: An overview of methods, issues, and perspectives*, Wiley Interdisciplinary Reviews: Climate Change, 9 (2018).

[95] J. Carrillo, F. Hoffmann, A. Stuart, and U. Vaes, *The mean-field ensemble Kalman filter: near-Gaussian setting*, SIAM Journal on Numerical Analysis, 62 (2024), pp. 2549–2587.

[96] C. M. Carvalho, N. G. Polson, and J. G. Scott, *Handling sparsity via the horseshoe*, in Artificial Intelligence and Statistics, PMLR, 2009, pp. 73–80.

[97] N. Chada, M. Iglesias, L. Roininen, and A. M. Stuart, *Parameterizations for ensemble Kalman inversion*, Inverse Problems, 34 (2017).

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

[99] N. K. Chada, M. A. Iglesias, L. Roininen, and A. M. Stuart, *Parameterizations for ensemble Kalman inversion*, Inverse Problems, 34 (2018), p. 055009.

[100] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, *The convex geometry of linear inverse problems*, Foundations of Computational Mathematics, 12 (2012), pp. 805–849.

[101] M. A. Chaplain, M. Ganesh, I. G. Graham, and G. Lolas, *Mathematical modelling of solid tumour growth: applications of pre-pattern formation*, in Morphogenesis and Pattern Formation in Biological Systems, Springer, 2003, pp. 283–293.

[102] A. Chattopadhyay, M. Mustafa, P. Hassanzadeh, E. Bach, and K. Kashinath, *Towards physics-inspired data-driven weather forecasting: integrating data assimilation with a deep spatial-transformer-based U-NET in a case study with ERA5*, Geoscientific Model Development, 15 (2022), pp. 2221–2237.

[103] A. Chattopadhyay, E. Nabizadeh, E. Bach, and P. Hassanzadeh, *Deep learning-enhanced ensemble-based data assimilation for high-dimensional nonlinear dynamical systems*, Journal of Computational Physics, 477 (2023), p. 111918.

[104] K. Chen, T. Han, J. Gong, L. Bai, F. Ling, J.-J. Luo, X. Chen, L. Ma, T. Zhang, R. Su, et al., *Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead*, arXiv preprint arXiv:2304.02948, (2023).

[105] N. Chen, *Stochastic Methods for Modeling and Predicting Complex Dynamical Systems: Uncertainty Quantification, State Estimation, and Reduced-Order Models*, Springer Nature, 2023.

[106] N. Chen and Y. Li, *BAMCAFE: A Bayesian machine learning advanced forecast ensemble method for complex turbulent systems with partial observations*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 31 (2021).

[107] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Review, 43 (2001), pp. 129–159.

[108] V. Chen, M. M. Dunlop, O. Papaspiliopoulos, and A. M. Stuart, *Dimension-robust mcmc in bayesian inverse problems*, arXiv preprint arXiv:1803.03344, (2018).

[109] Y. Chen, D. Sanz-Alonso, and R. Willett, *Autodifferentiable ensemble Kalman filters*, SIAM Journal on Mathematics of Data Science, 4 (2022), pp. 801–833.

[110] S. Cheng, C. Quilodrán-Casas, S. Ouala, A. Farchi, C. Liu, P. Tandeo, R. Fablet, D. Lucor, B. Iooss, J. Brajard, et al., *Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review*, IEEE/CAA Journal of Automatica Sinica, 10 (2023), pp. 1361–1387.

[111] N. Chopin, O. Papaspiliopoulos, et al., *An introduction to sequential Monte Carlo*, vol. 4, Springer, 2020.

[112] I. Chueshov, *A squeezing property and its applications to a description of long-time behaviour in the three-dimensional viscous primitive equations*, Proceedings of the Royal Society of Edinburgh Section A: Mathematics, 144 (2014), pp. 711–729.

[113] E. Cleary, A. Garbuno-Inigo, S. Lan, T. Schneider, and A. M. Stuart, *Calibrate, emulate, sample*, Journal of Computational Physics, 424 (2021), p. 109716.

[114] R. Cohen, A.and Devore and C. Schwab, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic pdes*, Analysis and Applications, 9 (2011), pp. 11–47.

[115] R. R. Coifman and S. Lafon, *Diffusion maps*, Applied and computational harmonic analysis, 21 (2006), pp. 5–30.

[116] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, *Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps*, Proceedings of the National Academy of Sciences of the United States of America, 102 (2005), pp. 7426–7431.

[117] R. R. Coifman, Y. Shkolnisky, F. J. Sigworth, and A. Singer, *Graph laplacian tomography from unknown random projections*, IEEE Transactions on Image Processing, 17 (2008), pp. 1891–1899.

[118] P. Constantin and C. Foias, *Navier-Stokes Equations*, University of Chicago Press, 2020.

[119] S. Cotter, M. Dashti, and A. M. Stuart, *Approximation of Bayesian inverse problems for PDE's*, SIAM Journal on Numerical Analysis, 48 (2010), pp. 322–345.

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

[121] 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), p. 115008.

[122] D. Crisan, P. Moral, and T. Lyons, *Discrete filtering using branching and interacting particle systems*, Université de Toulouse. Laboratoire de Statistique et Probabilités [LSP], (1998).

[123] G. Da Prato and J. Zabczyk, *Stochastic equations in infinite dimensions*, Cambridge university press, 2014.

[124] G. Dal Maso, *An introduction to* Γ*-convergence*, vol. 8, Springer Science & Business Media, 2012.

[125] P. Damlen, J. Wakefield, and S. Walker, *Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61 (1999), pp. 331–344.

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

[127] M. Dashti and A. M. Stuart, *Bayesian approach to inverse problems*, Handbook of Uncertainty Quantification, (2017), pp. 311–428.

[128] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, *Iteratively reweighted least squares minimization for sparse recovery*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 63 (2010), pp. 1–38.

[129] L. David, *Hyperspectral image data analysis as a high dimensional signal processing problem*, IEEE Signal Processing Magazine, 19 (2002), pp. 17–28.

[130] N. de Freitas, P. Højen-Sørensen, M. I. Jordan, and S. Russell, *Variational MCMC*, in Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, UAI'01, San Francisco, CA, USA, 2001, Morgan Kaufmann Publishers Inc., p. 120–127.

[131] J. De Wiljes, S. Reich, and W. Stannat, *Long-time stability and accuracy of the ensemble kalman–bucy filter for fully observed processes and small measurement noise*, SIAM Journal on Applied Dynamical Systems, 17 (2018), pp. 1152–1181.

[132] G. P. Dehaene, *Computing the quality of the Laplace approximation*, Neural Information Processing Systems, (2017).

[133] A. V. der Vaart, *Asymptotic Statistics*, Cambridge University Press, 1998.

[134] Z. Ding and Q. Li, *Ensemble Kalman inversion: Mean-field limit and convergence analysis*, Statistics and Computing, 31 (2021), pp. 1–21.

[135] ——, *Ensemble Kalman sampler: Mean-field limit and convergence analysis*, SIAM Journal on Mathematical Analysis, 53 (2021), pp. 1546–1578.

[136] D. L. Donoho, *De-noising by soft-thresholding*, IEEE Transactions on Information Theory, 41 (1995), pp. 613–627.

[137] ——, *High-dimensional data analysis: The curses and blessings of dimensionality*, AMS Math Challenges Lecture, 1 (2000), p. 32.

[138] A. Doucet, S. Godsill, and C. Andrieu, *On sequential monte carlo sampling methods for bayesian filtering*, Statistics and computing, 10 (2000), pp. 197–208.

[139] M. M. Dunlop, T. Helin, and A. M. Stuart, *Hyperparameter estimation in bayesian map estimation: parameterizations and consistency*, The SMAI Journal of Computational Mathematics, 6 (2020), pp. 69–100.

[140] M. M. Dunlop, M. A. Iglesias, and A. M. Stuart, *Hierarchical bayesian level set inversion*, Statistics and Computing, 27 (2017), pp. 1555–1584.

[141] G. Dziuk and C. M. Elliott, *Finite element methods for surface PDEs*, Acta Numerica, 22 (2013), pp. 289–396.

[142] C. Eilks and C. M. Elliott, *Numerical simulation of dealloying by surface dissolution via the evolving surface finite element method*, Journal of Computational Physics, 227 (2008), pp. 9727–9741.

[143] C. M. Elliott and B. Stinner, *A surface phase field model for two-phase biological membranes*, SIAM Journal on Applied Mathematics, 70 (2010), pp. 2904–2928.

[144] H. England, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, Springer Science and Business Media, 1996.

[145] O. G. Ernst, B. Sprungk, and H.-J. Starkloff, *Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems*, SIAM/ASA Journal on Uncertainty Quantification, 3 (2015), pp. 823–851.

[146] L. C. Evans, *Partial Differential Equations*, vol. 19, American Mathematical Soc., 2010.

[147] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, Journal of Geophysical Research: Oceans, 99 (1994), pp. 10143–10162.

[148] G. Evensen, F. C. Vossepoel, and P. J. Van Leeuwen, *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*, Springer Nature, 2022.

[149] A. Farchi, P. Laloyaux, M. Bonavita, and M. Bocquet, *Using machine learning to correct model error in data assimilation and forecast applications*, Quarterly Journal of the Royal Meteorological Society, 147 (2021), pp. 3067–3084.

[150] X. Fernique, *Intégrabilité des vecteurs gaussiens*, CR Acad. Sci. Paris Serie A, 270 (1970), pp. 1698–1699.

[151] C. Foias, C. F. Mondaini, and E. S. Titi, *A discrete data assimilation scheme for the solutions of the two-dimensional Navier–Stokes equations and their statistics*, SIAM Journal on Applied Dynamical Systems, 15 (2016), pp. 2109–2142.

[152] J. Franklin, *Well-posed stochastic extensions of ill-posed linear problems*, Journal of Mathematical Analysis and Applications, 31 (1970), pp. 682–716.

[153] P. Frauenfelder, C. Schwab, and R. A. Todor, *Finite elements for elliptic problems with stochastic coefficients*, Computer methods in applied mechanics and engineering, 194 (2005), pp. 205–228.

[154] R. Furrer and T. Bengtsson, *Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants*, Journal of Multivariate Analysis, 98 (2007), pp. 227–255.

[155] D. Gamerman and H. Lopes, *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*, CRC Press, (2006).

[156] T. Gao, S. Z. Kovalsky, and I. Daubechies, *Gaussian process landmarking on manifolds*, SIAM Journal on Mathematics of Data Science, 1 (2019), pp. 208–236.

[157] A. Garbuno-Inigo, T. Helin, F. Hoffmann, and B. Hosseini, *Bayesian posterior perturbation analysis with integral probability metrics*, arXiv preprint arXiv:2303.01512, (2023).

[158] 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*, Foundations of Computational Mathematics, 20 (2020), pp. 827–887.

[159] N. García 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, The Journal of Machine Learning Research, 21 (2020), pp. 1022–1068.

[160] N. Garcia Trillos and D. Sanz-Alonso, *The Bayesian formulation and well-posedness of fractional elliptic inverse problems*, Inverse Problems, 33 (2017), p. 065006.

[161] ——, *Continuum limits of posteriors in graph Bayesian inverse problems*, SIAM Journal on Mathematical Analysis, 50 (2018), pp. 4020–4040.

[162] N. Garcia Trillos and D. Sanz-Alonso, *The Bayesian update: variational formulations and gradient flows*, Bayesian Analysis, (2018).

[163] A. E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes, *Handbook of Spatial Statistics*, CRC Press, 2010.

[164] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman and Hall/CRC, 1995.

[165] C. J. Geoga, M. Anitescu, and M. L. Stein, *Scalable Gaussian process computations using hierarchical matrices*, Journal of Computational and Graphical Statistics, (2019), pp. 1–11.

[166] A. Gibbs and F. Su, *On choosing and bounding probability metrics*, International Statistical Review, 70 (2002), pp. 419–435.

[167] F. Gilani and J. Harlim, *Approximating solutions of linear elliptic PDE's on a smooth manifold using local kernel*, Journal of Computational Physics, 395 (2019), pp. 563–582.

[168] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Springer, 2015.

[169] M. Giles, *Multilevel Monte Carlo methods*, Acta Numerica, 24 (2015), pp. 259–328.

[170] E. Giné and V. Koltchinskii, *Empirical graph Laplacian approximation of Laplace–Beltrami operators: Large sample results*, in High dimensional probability, Institute of Mathematical Statistics, 2006, pp. 238–259.

[171] J. Glaubitz and A. Gelb, *Leveraging joint sparsity in hierarchical Bayesian learning*, SIAM/ASA Journal on Uncertainty Quantification, 12 (2024), pp. 442–472.

[172] M. S. Gockenbach, *Understanding and implementing the finite element method*, vol. 97, Siam, 2006.

[173] C. González-Tokman and B. R. Hunt, *Ensemble data assimilation for hyperbolic systems*, Physica D: Nonlinear Phenomena, 243 (2013), pp. 128–142.

[174] I. F. Gorodnitsky and B. D. Rao, *Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm*, IEEE Transactions on signal processing, 45 (1997), pp. 600–616.

[175] G. A. Gottwald and S. Reich, *Combining machine learning and data assimilation to forecast dynamical systems from noisy partial observations*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 31 (2021).

[176] ——, *Supervised learning from noisy observations: Combining machine-learning techniques with data assimilation*, Physica D: Nonlinear Phenomena, 423 (2021), p. 132911.

[177] M. Grasmair, M. Haltmeier, and O. Scherzer, *Sparse regularization with lq penalty term*, Inverse Problems, 24 (2008), p. 055020.

[178] ——, *The residual method for regularizing ill-posed problems*, Applied Mathematics and Computation, 218 (2011), pp. 2693–2710.

[179] M. Grasmair, O. Scherzer, and M. Haltmeier, *Necessary and sufficient conditions for linear convergence of ℓ_1-regularization*, Communications on Pure and Applied Mathematics, 64 (2011), pp. 161–182.

[180] P. J. Green, *Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives*, Journal of the Royal Statistical Society: Series B (Methodological), 46 (1984), pp. 149–170.

[181] B.-Y. Guo, *Spectral methods and their applications*, World Scientific, 1998.

[182] J. Hadamard, *Lectures on Cauchy's problem in linear partial differential equations*, Courier Corporation, 2014.

[183] M. Hairer, *An introduction to stochastic pdes*, arXiv preprint arXiv:0907.4178, (2009).

[184] M. Hairer, A. M. Stuart, J. Voss, and P. Wiberg, *Analysis of SPDEs arising in path sampling. Part I: The Gaussian case*, Communications in Mathematical Sciences, 3 (2013), pp. 587–603.

[185] F. Hamilton, T. Berry, and T. Sauer, *Ensemble Kalman filtering without a model*, Physical Review X, 6 (2016), p. 011021.

[186] A. Harbey, *Forecasting, structural time series models and the Kalman filter*, Cambridge university press, 1964.

[187] 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), p. 035006.

[188] J. Harlim, D. Sanz-Alonso, and R. Yang, *Kernel methods for Bayesian elliptic inverse problems on manifolds*, SIAM/ASA Journal on Uncertainty Quantification, 8 (2020), pp. 1414–1445.

[189] S. Hatfield, M. Chantry, P. Dueben, P. Lopez, A. Geer, and T. Palmer, *Building tangent-linear and adjoint models for data assimilation with neural networks*, Journal of Advances in Modeling Earth Systems, 13 (2021), p. e2021MS002521.

[190] K. Hayden, E. Olson, and E. Titi, *Discrete data assimilation in the Lorenz and 2D Navier–Stokes equations*, Physica D: Nonlinear Phenomena, 240 (2011), pp. 1416–1425.

[191] C. Heesterman and R. D. Gill, *A central limit theorem for M-estimators by the von Mises method*, Statistica Neerlandica, 46 (1992), pp. 165–177.

[192] M. Hein and J.-Y. Audibert, *Intrinsic dimensionality estimation of submanifolds in r d*, in Proceedings of the 22nd international conference on Machine learning, ACM, 2005, pp. 289–296.

[193] M. Hein, J.-Y. Audibert, and U. Von Luxburg, *From graphs to manifolds–weak and strong pointwise consistency of graph Laplacians*, in International Conference on Computational Learning Theory, Springer, 2005, pp. 470–485.

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

[195] M. Herty and G. Visconti, *Kinetic methods for inverse problems*, Kinetic & Related Models, 12 (2019).

[196] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, *Stochastic variational inference*, The Journal of Machine Learning Research, 14 (2013), pp. 1303–1347.

[197] P. L. Houtekamer and H. L. Mitchell, *Data assimilation using an ensemble Kalman filter technique*, Monthly Weather Review, 126 (1998), pp. 796–811.

[198] P. L. Houtekamer, H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, *Atmospheric data assimilation with an ensemble kalman filter: Results with real observations*, Monthly weather review, 133 (2005), pp. 604–620.

[199] J. Huang and T. Zhang, *The benefit of group sparsity*, The Annals of Statistics, 38 (2010), pp. 1978–2004.

[200] L. Huang, L. Gianinazzi, Y. Yu, P. D. Dueben, and T. Hoefler, *Diffda: a diffusion model for weather-scale data assimilation*, arXiv preprint arXiv:2401.05932, (2024).

[201] P. J. Huber, L. M. LeCam, and J. Neyman, *Proceedings of the fifth berkeley symposium on mathematical statistics and probability*, Chap. The behavior of maximum likelihood estimation under nonstandard conditions, (1967).

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

[203] M. A. Iglesias and C. Dawson, *The representer method for state and parameter estimation in single-phase darcy flow*, Computer Methods in Applied Mechanics and Engineering, 196 (2007), pp. 4577–4596.

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

[205] L. Jacob, G. Obozinski, and J.-P. Vert, *Group Lasso with overlap and graph Lasso*, in Proceedings of the 26th Annual International Conference on Machine Learning, 2009, pp. 433–440.

[206] R. Jiang, P. Y. Lu, E. Orlova, and R. Willett, *Training neural operators to preserve invariant measures of chaotic attractors*, Advances in Neural Information Processing Systems, 36 (2024).

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

[208] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, *An introduction to variational methods for graphical models*, Machine Learning, 37 (1999), pp. 183–233.

[209] J. Kaipo and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer Science & Business Media, 160 (2006).

[210] R. E. Kalman, *A new approach to linear filtering and prediction problems*, (1960).

[211] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*, vol. 341, Cambridge University Press, 2003.

[212] M. Katzfuss, J. R. Stroud, and C. K. Wikle, *Understanding the ensemble Kalman filter*, The American Statistician, 70 (2016), pp. 350–357.

[213] D. Kelly and A. M. Stuart, *Ergodicity and accuracy of optimal particle filters for Bayesian data assimilation*, arXiv preprint arXiv:1611.08761, (2014).

[214] D. T. Kelly, K. J. Law, and A. M. Stuart, *Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time*, Nonlinearity, 27 (2014), p. 2579.

[215] H. Kim, D. Sanz-Alonso, and A. Strang, *Hierarchical ensemble Kalman methods with sparsity-promoting generalized gamma hyperpriors*, Foundations of Data Science, 5 (2023), pp. 366–388.

[216] D. P. Kingma, *Adam: A method for stochastic optimization*, arXiv preprint arXiv:1412.6980, (2014).

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

[218] B. Knapik, A. van der Vaart, and J. van Zanten, *Bayesian inverse problems with Gaussian priors*, Annals of Statistics, 39 (2011), pp. 2626–2657.

[219] V. Koltchinskii and K. Lounici, *Concentration inequalities and moment bounds for sample covariance operators*, Bernoulli, (2017), pp. 110–133.

[220] S. Kotsuki, K. Shiraishi, and A. Okazaki, *Integrating Ensemble Kalman Filter with AI-based Weather Prediction Model ClimaX*, arXiv preprint arXiv:2407.17781, (2024).

[221] V. Krasnopolsky and E. M. Center, *Using machine learning for data assimilation, model physics, and post-processing model outputs*, (2023).

[222] R. Kretschmann, *Nonparametric Bayesian inverse problems with Laplacian noise*, PhD thesis, Dissertation, Duisburg, Essen, Universität Duisburg-Essen, 2019.

[223] A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, and D. M. Blei, *Automatic differentiation variational inference*, The Journal of Machine Learning Research, 18 (2017), pp. 430–474.

[224] E. Kwiatkowski and J. Mandel, *Convergence of the square root ensemble Kalman filter in the large ensemble limit*, SIAM/ASA Journal on Uncertainty Quantification, 3 (2015), pp. 1–17.

[225] R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, et al., *Graphcast: Learning skillful medium-range global weather forecasting*, arXiv preprint arXiv:2212.12794, (2022).

[226] H. Lambley, *Strong maximum a posteriori estimation in banach spaces with gaussian priors*, Inverse Problems, 39 (2023), p. 125010.

[227] H. Lambley and T. J. Sullivan, *An order-theoretic perspective on modes and maximum a posteriori estimation in bayesian inverse problems*, SIAM/ASA Journal on Uncertainty Quantification, 11 (2023), pp. 1195–1224.

[228] J. Latz, *On the well-posedness of Bayesian inverse problems*, SIAM/ASA Journal on Uncertainty Quantification, 8 (2020), pp. 451–482.

[229] ——, *Bayesian inverse problems are usually well-posed*, SIAM Review, 65 (2023), pp. 831–865.

[230] K. Law, A. Shukla, and A. Stuart, *Analysis of the 3dvar filter for thepartially observed lorenz'63 model*, Discrete and Continuous Dynamical Systems, 34 (2013), pp. 1061–1078.

[231] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation*, vol. 214, Springer, 2015.

[232] K. J. Law, D. Sanz-Alonso, A. Shukla, and A. M. Stuart, *Filter accuracy for the Lorenz 96 model: Fixed versus adaptive observation operators*, Physica D: Nonlinear Phenomena, 325 (2016), pp. 1–13.

[233] K. J. Law and A. M. Stuart, *Evaluating data assimilation algorithms*, Monthly Weather Review, 140 (2012), pp. 3757–3782.

[234] F. Le Gland, V. Monbet, and V.-D. Tran, *Large sample asymptotics for the ensemble Kalman filter*, PhD thesis, INRIA, 2009.

[235] P. V. Leeuwen, *Nonlinear data assimilation in geosciences: an extremely efficient particle filter*, Quarterly Journal of the Royal Meteorological Society, 136 (2010), pp. 1991–1999.

[236] P. V. Leeuwen, Y. Cheng, and S. Reich, *Nonlinear data assimilation*, Springer, 2015.

[237] A. J. Lemonte and G. M. Cordeiro, *The exponentiated generalized inverse Gaussian distribution*, Statistics & Probability Letters, 81 (2011), pp. 506–517.

[238] Z. Li and Z. Shi, *A convergent point integral method for isotropic elliptic equations on a point cloud*, Multiscale Modeling & Simulation, 14 (2016), pp. 874–905.

[239] Z. Li, Z. Shi, and J. Sun, *Point integral method for solving poisson-type equations on manifolds from point clouds with convergence guarantees*, Communications in Computational Physics, 22 (2017), pp. 228–258.

[240] C. Lieberman, C. Willcox, and O. Ghattas, *Parameter and state model reduction for large-scale statistical inverse problems*, SIAM Journal on Scientific Computing, 32 (2010), pp. 2535–2542.

[241] F. Lindgren, H. Rue, and J. Lindström, *An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73 (2011), pp. 423–498.

[242] T. Lindvall, *Lectures on the Coupling Method*, Springer, 2002.

[243] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer Science & Business Media, 2008.

[244] P.-L. Loh and M. J. Wainwright, *Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima*, Advances in Neural Information Processing Systems, 26 (2013).

[245] R. J. Lorentzen, K. M. Flornes, G. Nævdal, et al., *History matching channelized reservoirs using the ensemble kalman filter*, SPE Journal, 17 (2012), pp. 137–151.

[246] E. N. Lorenz, *Deterministic nonperiodic flow*, Journal of Atmospheric Sciences, 20 (1963), pp. 130–141.

[247] K. Lounici, M. Pontil, A. Tsybakov, and S. Van De Geer, *Taking advantage of sparsity in multi-task learning*, in COLT 2009-The 22nd Conference on Learning Theory, 2009.

[248] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, *Compressed sensing MRI*, IEEE Signal Processing Magazine, 25 (2008), pp. 72–82.

[249] A. J. Majda and X. T. Tong, *Performance of ensemble Kalman filters in large dimensions*, Communications on Pure and Applied Mathematics, 71 (2018), pp. 892–937.

[250] S. Mallat, *A Wavelet Tour of Signal Processing*, Elsevier, 1999.

[251] J. Mandel, L. Cobb, and J. D. Beezley, *On the convergence of the ensemble Kalman filter*, Applications of Mathematics, 56 (2011), pp. 533–541.

[252] Y. Marzouk, T. Moselhy, M. Parno, and A. Spantini, *Sampling via measure transport: an introduction*, Handbook of Uncertainty Quantification, (2016), pp. 1–41.

[253] Y. Marzouk and D. Xiu, *A stochastic collocation approach to Bayesian inference in inverse problems*, Communications in Computational Physics, 6 (2009), pp. 826–847.

[254] R. Maulik, V. Rao, J. Wang, G. Mengaldo, E. Constantinescu, B. Lusch, P. Balaprakash, I. Foster, and R. Kotamarthi, *Efficient high-dimensional variational data assimilation with machine-learned reduced-order models*, Geoscientific Model Development, 15 (2022), pp. 3433–3445.

[255] A. Maurais and Y. Marzouk, *Sampling in unit time with kernel fisher-rao flow*, arXiv preprint arXiv:2401.03892, (2024).

[256] D. McLaughlin and L. R. Townley, *A reassessment of the groundwater inverse problem*, Water Resources Research, 32 (1996), pp. 1131–1161.

[257] N. Meinshausen and B. Yu, *Lasso-type recovery of sparse representations for high-dimensional data*, The Annals of Statistics, 37 (2009), pp. 246–270.

[258] F. Mémoli, G. Sapiro, and P. Thompson, *Implicit brain imaging*, NeuroImage, 23 (2004), pp. S179–S188.

[259] M. Mercker and A. Marciniak-Czochra, *Bud-neck scaffolding as a possible driving force in escrt-induced membrane budding*, Biophysical journal, 108 (2015), pp. 833–843.

[260] A. J. Moodey, A. S. Lawless, R. W. Potthast, and P. J. Van Leeuwen, *Nonlinear error dynamics for cycled data assimilation methods*, Inverse Problems, 29 (2013), p. 025002.

[261] C. Mou, L. M. Smith, and N. Chen, *Combining stochastic parameterized reduced-order models with machine learning for data assimilation and uncertainty quantification with partial observations*, Journal of Advances in Modeling Earth Systems, 15 (2023), p. e2022MS003597.

[262] J. Mourtada, *Exact minimax risk for linear least squares, and the lower tail of sample covariance matrices*, The Annals of Statistics, 50 (2022), pp. 2157–2178.

[263] J. L. Mueller and S. Siltanen, *Linear and nonlinear inverse problems with practical applications*, vol. 10, Siam, 2012.

[264] C. Naesseth, S. Linderman, R. Ranganath, and D. Blei, *Variational sequential Monte Carlo*, in International conference on artificial intelligence and statistics, PMLR, 2018, pp. 968–977.

[265] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, *A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers*, Statistical Science, 27 (2012), pp. 538–557.

[266] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, *Supplement to 'A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers'*, Statistical Science, (2012).

[267] D. Nguyen, S. Ouala, L. Drumetz, and R. Fablet, *Em-like learning chaotic dynamics from noisy and partial observations*, arXiv preprint arXiv:1903.10335, (2019).

[268] T. Nguyen, J. Brandstetter, A. Kapoor, J. K. Gupta, and A. Grover, *Climax: A foundation model for weather and climate*, arXiv preprint arXiv:2301.10343, (2023).

[269] R. Nickl and E. S. Titi, *On posterior consistency of data assimilation with gaussian process priors: The 2d-navier–stokes equations*, The Annals of Statistics, 52 (2024), pp. 1825–1844.

[270] J. Nocedal and S. J. Wright, *Numerical optimization*, Springer, 2006.

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

[272] D. S. Oliver, A. C. Reynolds, and N. Liu, *Inverse theory for petroleum reservoir characterization and history matching*, 2008.

[273] E. Olson and E. S. Titi, *Determining modes for continuous data assimilation in 2d turbulence*, Journal of statistical physics, 113 (2003), pp. 799–840.

[274] 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.

[275] O. Papaspiliopoulos, G. O. Roberts, and M. Sköld, *A general framework for the parametrization of hierarchical models*, Statistical Science, (2007), pp. 59–73.

[276] J. Park, N. T. Yang, and N. Chandramoorthy, *When are dynamical systems learned from time series data statistically accurate?*, Advances in Neural Information Processing Systems, 37 (2024), pp. 43975–44008.

[277] T. Park and G. Casella, *The Bayesian Lasso*, Journal of the American Statistical Association, 103 (2008), pp. 681–686.

[278] J. Pathak, S. Subramanian, P. Harrington, S. Raja, A. Chattopadhyay, M. Mardani, T. Kurth, D. Hall, Z. Li, K. Azizzadenesheli, et al., *FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators*, arXiv preprint arXiv:2202.11214, (2022).

[279] L. M. Pecora and T. L. Carroll, *Synchronization in chaotic systems*, Physical Review Letters, 64 (1990), p. 821.

[280] S. G. Penny, T. A. Smith, T.-C. Chen, J. A. Platt, H.-Y. Lin, M. Goodliff, and H. D. Abarbanel, *Integrating recurrent neural networks with data assimilation for scalable data-driven state estimation*, Journal of Advances in Modeling Earth Systems, 14 (2022), p. e2021MS002843.

[281] 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 Journal on Scientific Computing, 36 (2014), pp. A1525–A1555.

[282] J. Ping, D. Zhang, et al., *History matching of channelized reservoirs with vector-based level-set parameterization*, Spe Journal, 19 (2014), pp. 514–529.

[283] C. Piret, *The orthogonal gradients method: A radial basis functions method for solving partial differential equations on arbitrary surfaces*, Journal of Computational Physics, 231 (2012), pp. 4662–4675.

[284] M. Pragliola, D. Calvetti, and E. Somersalo, *Overcomplete representation in a hierarchical Bayesian framework*, Inverse Problems and Imaging, 16 (2022), pp. 19–38.

[285] G. Raskutti, M. J. Wainwright, and B. Yu, *Restricted eigenvalue properties for correlated Gaussian designs*, The Journal of Machine Learning Research, 11 (2010), pp. 2241–2259.

[286] ——, *Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls*, IEEE Transactions on Information Theory, 57 (2011), pp. 6976–6994.

[287] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Review, 52 (2010), pp. 471–501.

[288] J. A. Reeds, *On the definition of von Mises functionals*, Ph. D thesis, Harvard University, (1976).

[289] S. Reich, *A dynamical systems framework for intermittent data assimilation*, BIT Numerical Mathematics, 51 (2017), pp. 235–249.

[290] H. E. Robbins, *An empirical Bayes approach to statistics*, in Breakthroughs in Statistics, Springer, 1992, pp. 388–394.

[291] G. O. Roberts and S. K. Sahu, *Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 59 (1997), pp. 291–317.

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

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

[294] A. Ron and Z. Shen, *Affine systems in $L^2(\mathbb{R}^d)$: the analysis of the analysis operator*, Journal of Functional Analysis, 148 (1997), pp. 408–447.

[295] S. J. Ruuth and B. Merriman, *A simple embedding method for solving partial differential equations on surfaces*, Journal of Computational Physics, 227 (2008), pp. 1943–1961.

[296] D. Sanz-Alonso, *Importance sampling and necessary sample size: An information theory approach*, SIAM/ASA Journal on Uncertainty Quantification, 6 (2018), pp. 867–879.

[297] D. Sanz-Alonso, A. Stuart, and A. Taeb, *Inverse Problems and Data Assimilation*, vol. 107, Cambridge University Press, 2023.

[298] ——, *Inverse Problems and Data Assimilation*, vol. 107, Cambridge University Press, 2023.

[299] D. Sanz-Alonso and A. M. Stuart, *Long-time asymptotics of the filtering distribution for partially observed chaotic dynamical systems*, SIAM/ASA Journal on Uncertainty Quantification, 3 (2015), pp. 1200–1220.

[300] D. Sanz-Alonso and N. Waniorek, *Analysis of a computational framework for Bayesian inverse problems: Ensemble Kalman updates and MAP estimators under mesh refinement*, SIAM/ASA Journal on Uncertainty Quantification, 12 (2024), pp. 30–68.

[301] ——, *Hierarchical bayesian inverse problems: A high-dimensional statistics viewpoint*, SIAM Review, 67 (2025), pp. 543–575.

[302] ——, *Long-time accuracy of ensemble kalman filters for chaotic dynamical systems and machine-learned dynamical systems*, SIAM Journal on Applied Dynamical Systems, 24 (2025), pp. 2246–2286.

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

[304] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational calculus for non-convex regularization*, Variational Methods in Imaging, (2009), pp. 159–183.

[305] C. Schillings, B. Sprungk, and P. Wacker, *On the convergence of the Laplace approximation and noise-level-robustness of Laplace-based Monte Carlo methods for Bayesian inverse problems*, Numerische Mathematik, 145 (2020), pp. 915–971.

[306] C. Schwab and J. Zech, *Deep learning in high dimension: Neural network expression rates for generalized polynomial chaos expansions in UQ*, Analysis and Applications, 17 (2019), pp. 19–55.

[307] Q.-M. Shao and Z.-S. Zhang, *Berry–Esseen bounds for multivariate nonlinear statistics with applications to M-estimators and stochastic gradient descent algorithms*, Bernoulli, 28 (2022), pp. 1548–1576.

[308] Z. Si, Y. Liu, and A. Strang, *Path-following methods for Maximum a Posteriori estimators in Bayesian hierarchical models: How estimates depend on hyperparameters*, SIAM Journal on Optimization, 34 (2024), pp. 2201–2230.

[309] A. Singer, *From graph to manifold Laplacian: The convergence rate*, Applied and Computational Harmonic Analysis, 21 (2006), pp. 128–134.

[310] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, *Obstacles to high-dimensional particle filtering*, Monthly Weather Review, 136 (2016), pp. 4629–4640.

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

[312] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Science & Business Media, 2012.

[313] G. Strang and G. J. Fix, *An Analysis of the Finite Element Method*, Englewood Cliffs, N. J., Prentice-Hall, Inc., 1973.

[314] A. Stuart and A. R. Humphries, *Dynamical Systems and Numerical Analysis*, vol. 2, Cambridge University Press, 1998.

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

[316] V. S. Sunder, *Functional analysis: spectral theory*, vol. 13, Springer, 1996.

[317] K. Takeda and T. Sakajo, *Uniform error bounds of the ensemble transform kalman filter for chaotic dynamics with multiplicative covariance inflation*, SIAM/ASA Journal on Uncertainty Quantification, 12 (2024), pp. 1315–1335.

[318] S. Tang, T. Sapsis, and N. Azizan, *Learning chaotic dynamics with embedded dissipativity*, arXiv preprint arXiv:2410.00976, (2024).

[319] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, 2015.

[320] ——, *Towards adjoint-based inversion for rheological parameters in nonlinear viscous mantle flow*, Physics of the Earth and Planetary Interiors, 234 (2015), pp. 23–34.

[321] R. Temam, *Navier–Stokes Equations and Nonlinear Functional Analysis*, SIAM, 1995.

[322] ——, *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, vol. 68, Springer Science & Business Media, 2012.

[323] E. H. Thiede, D. Giannakis, A. R. Dinner, and J. Weare, *Galerkin approximation of dynamical quantities using trajectory data*, The Journal of Chemical Physics, 150 (2019), p. 244111.

[324] X. Tian, D. Holdaway, and D. Kleist, *Exploring the use of machine learning weather models in data assimilation*, arXiv preprint arXiv:2411.14677, (2024).

[325] R. Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society: Series B (Methodological), 58 (1996), pp. 267–288.

[326] M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, *Ensemble square root filters*, Monthly Weather Review, 131 (2003), pp. 1485–1490.

[327] S. Tokdar, S. Kass, and R. Kass, *Importance sampling: a review*, Wiley Interdisciplinary Reviews: Computational Statistics, 2 (2010), pp. 54–60.

[328] X. Tong, A. Majda, and D. Kelly, *Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation*, Nonlinearity, 29 (2015), pp. 54–60.

[329] X. T. Tong, A. J. Majda, and D. Kelly, *Nonlinear stability and ergodicity of ensemble based Kalman filters*, Nonlinearity, 29 (2016), p. 657.

[330] F. Tonolini, J. Radford, A. Turpin, D. Faccio, and R. Murray-Smith, *Variational inference for computational imaging inverse problems*, Journal of Machine Learning Research, 21 (2020), pp. 1–46.

[331] J. A. Tropp, *An introduction to matrix concentration inequalities*, Foundations and Trends® in Machine Learning, 8 (2015), pp. 1–230.

[332] J. A. Tropp and S. J. Wright, *Computational methods for sparse solution of linear inverse problems*, Proceedings of the IEEE, 98 (2010), pp. 948–958.

[333] P. Tseng, *Convergence of a block coordinate descent method for nondifferentiable minimization*, Journal of optimization theory and applications, 109 (2001), pp. 475–494.

[334] T. Tsuyuki and R. Tamura, *Nonlinear data assimilation by deep learning embedded in an ensemble Kalman filter*, Journal of the Meteorological Society of Japan. Ser. II, 100 (2022), pp. 533–553.

[335] A. Tsybakov, *Introduction to Nonparametric Estimation*, Springer Series in Statistics, Springer New York, 2008.

[336] S. van de Geer, *Empirical Processes in M-Estimation*, vol. 6, Cambridge University Press, 2000.

[337] S. van de Geer, *High-dimensional generalized linear models and the Lasso*, The Annals of Statistics, 36 (2008), pp. 614–645.

[338] S. van de Geer, *Weakly decomposable regularization penalties and structured sparsity*, Scandinavian Journal of Statistics, 41 (2014), pp. 72–86.

[339] R. van Handel, *Probability in high dimension*.

[340] R. Vershynin, *High-dimensional Probability: An Introduction with Applications in Data Science*, vol. 47, Cambridge University Press, 2018.

[341] U. Von Luxburg, *A tutorial on spectral clustering*, Statistics and computing, 17 (2007), pp. 395–416.

[342] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, vol. 48, Cambridge University Press, 2019.

[343] M. J. Wainwright and M. I. Jordan, *Graphical Models, Exponential Families, and Variational Inference*, Now Publishers Inc, 2008.

[344] S. F. Waldron, *An Introduction to Finite Tight Frames*, Springer, 2018.

[345] N. Waniorek, D. Calvetti, and E. Somersalo, *Bayesian hierarchical dictionary learning*, Inverse Problems, 39 (2023), p. 024006.

[346] L. Wasserman, *All of Nonparametric Statistics*, Springer Science & Business Media, 2006.

[347] C. K. I. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*, vol. 2, MIT Press Cambridge, MA, 2006.

[348] Y. Xiao, L. Bai, W. Xue, K. Chen, T. Han, and W. Ouyang, *FengWu-4DVar: Coupling the data-driven weather forecasting model with 4D variational assimilation*, arXiv preprint arXiv:2312.12455, (2023).

[349] J.-J. Xu, Z. Li, J. Lowengrub, and H. Zhao, *A level-set method for interfacial flows with surfactant*, Journal of Computational Physics, 212 (2006), pp. 590–616.

[350] Y. Yu and X.-L. Meng, *To center or not to center: That is not the question—an Ancillarity–Sufficiency Interweaving Strategy (ASIS) for boosting MCMC efficiency*, Journal of Computational and Graphical Statistics, 20 (2011), pp. 531–570.

[351] M. Yuan and Y. Lin, *Model selection and estimation in regression with grouped variables*, Journal of the Royal Statistical Society Series B: Statistical Methodology, 68 (2006), pp. 49–67.

[352] L. Zelnik-Manor and P. Perona, *Self-tuning spectral clustering*, in Advances in neural information processing systems, 2005, pp. 1601–1608.

[353] C. Zhang, J. Bütepage, H. Kjellström, and S. Mandt, *Advances in variational inference*, IEEE transactions on pattern analysis and machine intelligence, 41 (2018), pp. 2008–2026.

[354] H. Zou, T. Hastie, and R. Tibshirani, *On the "degrees of freedom" of the lasso*, The Annals of Statistics, 35 (2007), pp. 2173–2192.