

HYPERDIFFERENTIAL SENSITIVITY ANALYSIS IN THE CONTEXT OF BAYESIAN INFERENCE APPLIED TO ICE-SHEET PROBLEMS

William Reese,¹ Joseph Hart,² Bart van Bloemen Waanders,^{2,*}
Mauro Perego,² John Jakeman,³ & Arvind K. Saibaba¹

¹Department of Mathematics, P.O. Box 8205, NC State University Raleigh, North Carolina 27695-8205, USA

²Scientific Machine Learning, Sandia National Laboratories, P.O. Box 5800, Albuquerque, New Mexico 87185, USA

³Optimization and Uncertainty Quantification, Sandia National Laboratories, P.O. Box 5800, Albuquerque, New Mexico 87123-1320, USA

*Address all correspondence to: Bart van Bloemen Waanders, Scientific Machine Learning, Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185, E-mail: bartv@sandia.gov

Original Manuscript Submitted: 1/17/2023; Final Draft Received: 8/20/2023

Inverse problems constrained by partial differential equations (PDEs) play a critical role in model development and calibration. In many applications, there are multiple uncertain parameters in a model which must be estimated. Although the Bayesian formulation is attractive for such problems, computational cost and high dimensionality frequently prohibit a thorough exploration of the parametric uncertainty. A common approach is to reduce the dimension by fixing some parameters (which we will call auxiliary parameters) to a best estimate and use techniques from PDE-constrained optimization to approximate properties of the Bayesian posterior distribution. For instance, the maximum a posteriori probability (MAP) and the Laplace approximation of the posterior covariance can be computed. In this article, we propose using hyperdifferential sensitivity analysis (HDSA) to assess the sensitivity of the MAP point to changes in the auxiliary parameters. We establish an interpretation of HDSA as correlations in the posterior distribution. Our proposed framework is demonstrated on the inversion of bedrock topography for the Greenland ice-sheet with uncertainties arising from the basal friction coefficient and climate forcing (ice accumulation rate).

KEY WORDS: hyperdifferential sensitivity analysis, Bayesian inverse problems, inversion

1. INTRODUCTION

Large-scale inverse problems occur in a range of geoscience applications from seismicity to ice-sheet flows [1,2]. In such problems, the quantities of interest typically depend on unknown parameters that describe material properties, source terms, and boundary and initial conditions in the governing partial differential equations (PDEs). The goal is to reconcile the differences between measurements and numerical predictions by estimating or reconstructing the unknown parameters. This is fraught with many challenges such as limited availability and noise in the observed data, the need to suitably regularize the problem, and additional uncertainties present in the model.

We discuss some of these challenges in the context of an ice-sheet model, the driving application for this work. Ice sheets play an important role in the global climate through their effects on sea level rise. Sea level rise may potentially cause severe flooding and weather changes that will negatively impact wildlife, agriculture, and coastal infrastructure [3]. We consider an ice-sheet model with several uncertain parameters: the bedrock topography beneath the ice, accumulation and ablation on the upper surface of the ice, and basal friction on the bottom of the ice-sheet. The bedrock topography is estimated using data acquired from aircraft flyovers but is uncertain because of data sparsity (the paths flown by aircraft) [4,5]. Accumulation and ablation, a forcing term in the model, requires climate

information to be accurately specified [6,7]. The basal friction, a combination of several physical phenomena, is uncertain due to the limited availability of observations beneath the ice-sheet.

Joint inversion [8] is one approach to estimate multiple uncertain parameters that may be interdependent. An example is inversion of electromagnetic and seismic parameters [9]. However, joint inversion is challenging because the parameters can be high-dimensional and a joint regularization term that reflects spatial correlations and imposes regularity [8] is typically required. Some examples of joint regularization are the cross gradient regularization term [10] and color total variation (TV) [11].

Bayesian theory provides a suitable formalism to solve inverse problems with statistical characterizations. In practice, joint Bayesian inversion of *all* unknown parameters is computationally intractable for the applications of interest. A commonly used alternative, which we also adopt in this paper, is to invert for one set of parameters (which we call *inversion parameters*) and fix all remaining parameters (which we call *auxiliary parameters*) to a nominal value. The challenge with this approach is that inferences made about the inversion parameter are relative to the value of the nominal parameters. This implies that if the nominal values change the inversion may yield very different results. Therefore, it remains to quantify the sensitivity of the solution of the inverse problem with respect to auxiliary parameters. To this end, we use hyperdifferential sensitivity analysis (HDSA) to compute postoptimality sensitivities of the maximum *a posteriori* estimate with respect to perturbations of the auxiliary parameters.

Our approach is complementary to the Bayesian approximation error (BAE) method [12–14]. In BAE, the auxiliary parameter are marginalized out of the Bayesian posterior giving an inverse problem to estimate the inversion parameters with a cognizance of the auxiliary parameter uncertainty. In contrast, we show how HDSA provides information about the joint posterior in a neighborhood of the nominal auxiliary parameters thus elucidating dependencies between the parameters. There is also the potential to combine BAE and HDSA, but such analysis is beyond the scope of this article.

Contributions. We build on previous efforts [15–17] and apply HDSA to analyze the sensitivity of the estimated bedrock topography with respect to perturbations of the accumulation/ablation forcing and basal friction for a realistic model of the Greenland ice-sheet. Previous work has applied HDSA for infinite-dimensional inverse problems. This article extends these demonstrations in the complexity of the model being analyzed (a transient nonlinear partial differential equation defined on a subset of Greenland). Furthermore, we introduce a new interpretation of the sensitivities in terms of correlations in the joint Bayesian posterior distribution. This links properties of the computationally intractable joint inversion problem with postoptimality sensitivities which may be computed efficiently. Additionally, we demonstrate that computing HDSA on the likelihood informed subspace (LIS), as introduced in [15], provides the eigenvectors needed to compute approximation posterior samples with the Laplace approximation [2,18]. With this novel observation, we show how postoptimality sensitivities and approximate posterior samples may be computed simultaneously. The resulting approach is applicable to large-scale nonlinear inverse problems with expensive forward models involving systems of PDEs and multiple unknown parameters.

Overview. In Section 2, we provide a brief overview of Bayesian inversion in the context of PDE constrained optimization and HDSA. An interpretation of HDSA as correlations in the joint posterior is established in Section 3. Section 4 demonstrates how sensitivities and approximate posterior samples may be simultaneously computed using the LIS. The computational costs are analyzed in Section 5. In Section 6, we demonstrate this process on an inverse problem for the bedrock topography in a region of Greenland. Concluding remarks are made in Section 7.

2. BACKGROUND

We consider inverse problems to estimate parameters $z \in Z$, where Z may be finite- or infinite-dimensional. Assume that z cannot be observed directly, but rather we have sparse and noisy observations of a state $u \in U$ (U is an appropriate function space) which is related to z by a PDE, c , in the form

$$c(u, z, \theta) = 0, \quad (1)$$

where $\theta \in \Theta$ are uncertain parameters, referred to as auxiliary parameters, needed to define the PDE.

Assume that the PDE is uniquely solvable for each $z \in Z$ and $\theta \in \Theta$ and let $\Psi : Z \times \Theta \rightarrow U$ denote the PDE solution operator, i.e.,

$$c(\Psi(z, \theta), z, \theta) = 0 \quad \text{for all } z \in Z, \theta \in \Theta.$$

Let $\mathcal{F} : Z \times \Theta \rightarrow \mathbb{R}^d$ denote the parameter-to-observable map. Specifically,

$$\mathcal{F}(z, \theta) = \mathcal{O} \circ \Psi(z, \theta),$$

maps parameters z and θ to observations of the PDE solution at d locations via the observation operator $\mathcal{O} : U \rightarrow \mathbb{R}^d$.

We focus on the finite-dimensional inverse problem which arises from the discretization of the PDE. To simplify the exposition, assume that $Z_h \subset Z$ and $\Theta_h \subset \Theta$ are finite-dimensional (a result of the PDE discretization) with bases $\{y_1, y_2, \dots, y_m\} \subset Z_h$ and $\{\phi_1, \phi_2, \dots, \phi_n\} \subset \Theta_h$. We will use $z = (z_1, z_2, \dots, z_m)^T \in \mathbb{R}^m$ and $\theta = (\theta_1, \theta_2, \dots, \theta_n)^T \in \mathbb{R}^n$ to denote coordinate representations of elements in Z_h and Θ_h (which may be function spaces). Let $f : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^d$ denote the discretized parameter-to-observable map. We assume that the data $d \in \mathbb{R}^d$ are related to parameters z and θ as

$$d = f(z, \theta) + \eta, \quad (2)$$

where $\eta \in \mathbb{R}^d$ is noise. The goal of the finite-dimensional inverse problem is to estimate z and θ from d .

2.1 Bayesian Inverse Problems

Inverse problems are frequently ill-posed in the sense that there are many different z 's and θ 's such that $f(z, \theta) \approx d$. This ill-posedness motivates a Bayesian approach to the inverse problem. We review core concepts below and direct the reader to [19,20] for a detailed introduction to Bayesian inverse problems.

Assume that the data d are related to the parameters z and θ as in Eq. (2) where $\eta \sim \mathcal{N}(0, \Gamma_{\text{noise}})$. Throughout, we also assume that the prior distributions of z and θ are independent Gaussians $z \sim \mathcal{N}(\mu_z, \Gamma_z)$ and $\theta \sim \mathcal{N}(\mu_\theta, \Gamma_\theta)$, where both Γ_z and Γ_θ are symmetric positive definite. Applying Bayes rule, the posterior distribution is

$$\pi_{\text{joint}}(z, \theta | d) = \frac{1}{C_{\text{joint}}} \exp\left(-\frac{1}{2}\|f(z, \theta) - d\|_{\Gamma_{\text{noise}}^{-1}}^2 - \frac{1}{2}\|z - \mu_z\|_{\Gamma_z^{-1}}^2 - \frac{1}{2}\|\theta - \mu_\theta\|_{\Gamma_\theta^{-1}}^2\right), \quad (3)$$

where C_{joint} denotes a normalizing constant which is difficult to compute and is unimportant for this discussion. However, jointly estimating z and θ is challenging since the dimensions m and n may both be large and the data may not be sufficiently informative of both parameters. Therefore, one approach in practice is to fix the auxiliary parameters to a nominal value, denoted by $\bar{\theta}$, and estimate z from the data.

From a probabilistic perspective, fixing θ to a nominal value corresponds to conditioning the joint posterior π_{joint} on $\theta = \bar{\theta}$. In other words, analyzing the conditional posterior,

$$\pi_{\text{cond}}(z | \bar{\theta}, d) = \frac{1}{C_{\text{cond}}(\bar{\theta})} \exp\left(-\frac{1}{2}\|f(z, \bar{\theta}) - d\|_{\Gamma_{\text{noise}}^{-1}}^2 - \frac{1}{2}\|z - \mu_z\|_{\Gamma_z^{-1}}^2\right), \quad (4)$$

where the normalizing constant $C_{\text{cond}}(\bar{\theta})$ is a function of the nominal value of the auxiliary parameters. Computing $C_{\text{cond}}(\bar{\theta})$ is challenging as it requires integrating $\pi_{\text{joint}}(z, \theta | d)$ with respect to z ; however, it is unimportant for this discussion.

Observe that the maximum a posteriori probability (MAP) point of $\pi_{\text{cond}}(z | \bar{\theta}, d)$ is obtained by solving the optimization problem,

$$\min_{z \in \mathbb{R}^m} J(z, \bar{\theta}) := M(z, \bar{\theta}) + R(z), \quad (5)$$

where

$$M(z, \bar{\theta}) := \frac{1}{2}\|f(z, \bar{\theta}) - d\|_{\Gamma_{\text{noise}}^{-1}}^2 \quad \text{and} \quad R(z) := \frac{1}{2}\|z - \mu_z\|_{\Gamma_z^{-1}}^2, \quad (6)$$

are the negative log-likelihood and the negative log of z 's prior PDF (with normalizing constants omitted). In the context of the optimization problem (5), we will also refer to M and R as the misfit and regularization, respectively.

For high-dimensional inverse problems constrained by computationally intensive PDEs, sampling from the posterior distribution is challenging. When f is a linear function, the posterior distribution is Gaussian and its covariance

matrix is given by the inverse Hessian of the negative log of the posterior PDF. For a general nonlinear $\mathbf{f}$, the posterior covariance is non-Gaussian and significantly more challenging to estimate. Computing the MAP point (5) is computationally intensive since it requires many PDE solves [1,2,21]. However, techniques from PDE-constrained optimization may be leveraged to solve (5) at large scales. We use techniques including finite element discretization, matrix-free linear algebra, adjoint-based derivative computation, and parallel computing. The reader is referred to [21–34] for a sampling of the PDE-constrained optimization literature. From the perspective of the Bayesian inverse problem, PDE-constrained optimization is a valuable tool to efficiently compute the MAP point (and possibly approximate the covariance).

2.2 Hyperdifferential Sensitivity Analysis

Postoptimality sensitivity analysis is predicated on employing the implicit function theorem to write the solution of an inverse problem as a function of the auxiliary parameters. This function and its Jacobian are defined within a neighborhood of the nominal value of the auxiliary parameters. Through a combination of tools from postoptimality sensitivity analysis, PDE-constrained optimization, and numerical linear algebra, HDSA provides unique and valuable insights for optimal control and deterministic inverse problems [16,17,35]. This subsection provides essential background to prepare for our extension of HDSA to Bayesian inverse problems. To facilitate our analysis, assume that the objective function $J : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}$ in Eq. (5) is twice continuously differentiable with respect to $(\mathbf{z}, \boldsymbol{\theta})$.

Let $\mathbf{z}^*$ be a local minimum of the objective function with auxiliary parameters fixed to $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}} \in \mathbb{R}^n$; i.e., it is a MAP point of the posterior for $\mathbf{z}$ when conditioning on $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$. A fundamental assumption of post-optimality sensitivity analysis is that $\mathbf{z}^*$ satisfies the well-known first- and second-order optimality conditions:

$$(A1) \quad \nabla_{\mathbf{z}} J(\mathbf{z}^*, \bar{\boldsymbol{\theta}}) = 0,$$

$$(A2) \quad \nabla_{\mathbf{z}, \mathbf{z}} J(\mathbf{z}^*, \bar{\boldsymbol{\theta}}) \text{ is positive definite,}$$

where $\nabla_{\mathbf{z}} J$ and $\nabla_{\mathbf{z}, \mathbf{z}} J$ denote the gradient and Hessian of J with respect to $\mathbf{z}$, respectively. Then applying the implicit function theorem (IFT) to the equation $\nabla_{\mathbf{z}} J(\mathbf{z}, \boldsymbol{\theta}) = 0$, which is satisfied by $(\mathbf{z}^*, \bar{\boldsymbol{\theta}})$, gives the existence of a continuously differentiable operator $\mathcal{G} : N(\bar{\boldsymbol{\theta}}) \rightarrow \mathbb{R}^m$ such that $\nabla_{\mathbf{z}} J(\mathcal{G}(\boldsymbol{\theta}), \boldsymbol{\theta}) = 0$ for all $\boldsymbol{\theta} \in N(\bar{\boldsymbol{\theta}})$. The operator's domain $N(\bar{\boldsymbol{\theta}}) = \{\boldsymbol{\theta} \mid \|\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\|_2 < \epsilon\}$ is a radius $\epsilon > 0$ neighborhood of $\bar{\boldsymbol{\theta}}$ arising from the IFT. Furthermore, the IFT gives the Jacobian of $\mathcal{G}$ evaluated at $\bar{\boldsymbol{\theta}}$,

$$\mathcal{G}'(\bar{\boldsymbol{\theta}}) = -\mathcal{H}^{-1} \mathcal{B}, \quad (7)$$

where $\mathcal{B} = \nabla_{\mathbf{z}, \boldsymbol{\theta}} J(\mathbf{z}^*, \bar{\boldsymbol{\theta}})$ denotes the Jacobian of $\nabla_{\mathbf{z}} J$ with respect to $\boldsymbol{\theta}$, and $\mathcal{H} = \nabla_{\mathbf{z}, \mathbf{z}} J(\mathbf{z}^*, \bar{\boldsymbol{\theta}})$ denotes the Hessian of J with respect to $\mathbf{z}$, each evaluated at $\mathbf{z} = \mathbf{z}^*$ and $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$. Then we may interpret $\mathcal{G}'(\bar{\boldsymbol{\theta}}) \hat{\boldsymbol{\theta}}$ as the change in the optimal solution of Eq. (5) when $\bar{\boldsymbol{\theta}}$ is perturbed in the direction $\hat{\boldsymbol{\theta}}$, i.e., the sensitivity of the MAP point $\mathbf{z}^*$ to the perturbation $\hat{\boldsymbol{\theta}}$.

HDSA efficiently interrogates the postoptimality sensitivity operator $\mathcal{G}'(\bar{\boldsymbol{\theta}})$ by leveraging tools such as adjoint-based derivative calculations and randomized linear algebra. We also note that postoptimality sensitivities and HDSA may be formally developed for infinite-dimensional problems in a full space optimization framework. The reader is directed to [17] for additional details.

3. INTERPRETATION OF SENSITIVITIES AS CORRELATIONS IN JOINT POSTERIOR

Although HDSA was developed in the context of PDE-constrained optimization, we establish a statistical interpretation of the sensitivities by linking them to the Bayesian inverse problem in Eq. (3). As shown in [36], the inverse Hessian of the negative log posterior PDF equals the covariance matrix of the posterior distribution when $\mathbf{f}$ is linear. Though this does not hold for general nonlinear problems, the Laplace approximation linearizes $\mathbf{f}$ around the MAP point and approximates the posterior covariance [18] using the inverse Hessian evaluated at the MAP point. Similar logic is followed to establish an interpretation of the postoptimality sensitivities as a correlation in the joint posterior distribution of $(\mathbf{z}, \boldsymbol{\theta})$. We provide an analytical result for linear inverse problems and then argue for its local validity in nonlinear inverse problems.

Consider the Bayesian inverse problem on the joint distribution of (z, θ) and assume that the parameter-to-observable map $f(z, \theta)$ is linear in (z, θ) ; i.e., $f(z, \theta) = Az + B\theta$ for given matrices $A \in \mathbb{R}^{d \times m}$ and $B \in \mathbb{R}^{d \times n}$.

Since f is linear, and the prior is Gaussian, the joint posterior distribution $\pi_{\text{joint}}(z, \theta|d)$ is Gaussian with covariance matrix Γ_{joint} which satisfies

$$\Gamma_{\text{joint}}^{-1} = \begin{bmatrix} A^T \\ B^T \end{bmatrix} \Gamma_{\text{noise}}^{-1} \begin{bmatrix} A & B \end{bmatrix} + \begin{bmatrix} \Gamma_z^{-1} & \\ & \Gamma_\theta^{-1} \end{bmatrix} = \begin{bmatrix} A^T \Gamma_{\text{noise}}^{-1} A + \Gamma_z^{-1} & A^T \Gamma_{\text{noise}}^{-1} B \\ B^T \Gamma_{\text{noise}}^{-1} A & B^T \Gamma_{\text{noise}}^{-1} B + \Gamma_\theta^{-1} \end{bmatrix}.$$

Theorem 1 provides an interpretation of HDSA in the context of the posterior covariance.

Theorem 1. Assume that the forward model is linear; that is, $f(z, \theta) = Az + B\theta$, and denote joint posterior covariance matrix as

$$\Gamma_{\text{joint}} := \begin{bmatrix} \Gamma_{z,z} & \Gamma_{z,\theta} \\ \Gamma_{z,\theta}^T & \Gamma_{\theta,\theta} \end{bmatrix}.$$

Then the postoptimality sensitivity operator $\mathcal{G}'(\bar{\theta})$ corresponding to the negative log of the conditional posterior (4) satisfies

$$\mathcal{G}'(\bar{\theta}) = \Gamma_{z,\theta} \Gamma_{\theta,\theta}^{-1}.$$

Proof. Define $K := A^T \Gamma_{\text{noise}}^{-1} A + \Gamma_z^{-1}$, $L := A^T \Gamma_{\text{noise}}^{-1} B$, and $M := B^T \Gamma_{\text{noise}}^{-1} B + \Gamma_\theta^{-1}$, which represent the (1, 1), (1, 2), and (2, 2) blocks of $\Gamma_{\text{joint}}^{-1}$, respectively. Using the block inversion formula gives

$$\Gamma_{\text{joint}} = \begin{bmatrix} \Gamma_{z,z} & \Gamma_{z,\theta} \\ \Gamma_{z,\theta}^T & \Gamma_{\theta,\theta} \end{bmatrix} = \begin{bmatrix} K^{-1} + K^{-1} L S^{-1} L^T K^{-1} & -K^{-1} L S^{-1} \\ -S^{-1} L^T K^{-1} & S^{-1} \end{bmatrix},$$

where $S := M^{-1} - L^T K^{-1} L$ is the Schur complement of M . Equating the individual subblocks, we have $\Gamma_{z,\theta} = -K^{-1} L S^{-1}$ and $\Gamma_{\theta,\theta} = S^{-1}$, so that $\Gamma_{z,\theta} \Gamma_{\theta,\theta}^{-1} = -K^{-1} L$.

From calculus we have $\nabla_{z,z} J = A^T \Gamma_{\text{noise}}^{-1} A + \Gamma_z^{-1}$ and $\nabla_{z,\theta} J = A^T \Gamma_{\text{noise}}^{-1} B$, which implies that

$$\mathcal{G}'(\bar{\theta}) = -(A^T \Gamma_{\text{noise}}^{-1} + \Gamma_z^{-1})^{-1} A^T \Gamma_{\text{noise}}^{-1} B = -K^{-1} L = \Gamma_{z,\theta} \Gamma_{\theta,\theta}^{-1}. \quad \square$$

Theorem 1 implies that the sensitivity of the MAP point for z with respect to θ , for linear Bayesian inverse problems with Gaussian noise and priors, corresponds to a scaled version of the posterior covariance between z and θ . In the case where z and θ are scalars, we may relate $\mathcal{G}'(\bar{\theta})$ to the Pearson correlation coefficient. Specially, we measure the magnitude of a change in z , according to the Mahalanobis norm defined by the z posterior, under a unit norm perturbation of θ according to the Mahalanobis norm defined by the θ posterior. That is, for the unit vector $e \in 1/\sqrt{\Gamma_{\theta,\theta}^{-1}} \in \mathbb{R}$

$$\|\mathcal{G}'(\bar{\theta})e\|_{\Gamma_{z,z}^{-1}}^2 = \frac{1}{\sqrt{\Gamma_{\theta,\theta}^{-1}}} \Gamma_{\theta,\theta}^{-1} \Gamma_{z,\theta} \Gamma_{z,z}^{-1} \Gamma_{z,\theta} \Gamma_{\theta,\theta}^{-1} \frac{1}{\sqrt{\Gamma_{\theta,\theta}^{-1}}} = \frac{\Gamma_{z,\theta}^2}{\Gamma_{z,z} \Gamma_{\theta,\theta}},$$

or equivalently,

$$\|\mathcal{G}'(\bar{\theta})e\|_{\Gamma_{z,z}^{-1}} = \frac{\Gamma_{z,\theta}}{\sqrt{\Gamma_{z,z}} \sqrt{\Gamma_{\theta,\theta}}},$$

which is the Pearson correlation coefficient defined by the covariance of z and θ , normalized by the standard deviations of z and θ . This provides a foundation to interpret HDSA in terms of the correlation between z and θ in their joint posterior distribution. Correlations are a considerable challenge in joint inversion, and hence this interpretation of HDSA provides critical posterior information to enable better characterization of uncertainty. For instance, posterior distributions with strong correlations will have poor sample efficiency in Markov chain Monte Carlo or rejection sampling algorithms if they are not informed by the correlation structure. Hence this interpretation of HDSA provides critical posterior information to enable better characterization of uncertainty.

Following similar principles of the Laplace approximation (e.g., [18]), HDSA may be interpreted as local correlations in the joint posterior distribution for nonlinear inverse problems. As in [18], we consider a linear approximation of $\mathbf{f}$ around $(z^*, \bar{\theta})$, i.e., the linear inverse problem whose parameter-to-observable map is, to a first-order approximation,

$$\mathbf{f}(z, \theta) \approx \mathbf{f}(z^*, \bar{\theta}) + \mathbf{f}'_z(z^*, \bar{\theta})(z - z^*) + \mathbf{f}'_\theta(z^*, \bar{\theta})(\theta - \bar{\theta}),$$

where $\mathbf{f}'_z$ and $\mathbf{f}'_\theta$ denote the Jacobians of $\mathbf{f}$. Applying Theorem 1 to this linearized problem, $\mathcal{G}'(\bar{\theta})$ measures the correlation between z and θ in the region where the linear approximation is valid, i.e., in a neighborhood of $(z^*, \bar{\theta})$. Example 1 below further explores the connection between HDSA and the joint posterior distribution.

Example 1. To illustrate the relationship between HDSA and correlations in the joint posterior distribution, consider the synthetic model problem

$$f(z, \theta_1, \theta_2) = e^{(1/10)z\theta_1} + \theta_2.$$

We generate three data points by evaluating f at the “true” parameters $z^\dagger = 5$ and $\theta^\dagger = (\theta_1, \theta_2)^T = (5, 1)^T$ and contaminate them with Gaussian noise. Taking a joint Gaussian prior on (z, θ_1, θ_2) with mean and covariance matrix

$$\mu = \begin{pmatrix} 5 \\ 5 \\ 1 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} 5^2 & 0 & 0 \\ 0 & 2^2 & 0 \\ 0 & 0 & 0.2^2 \end{pmatrix},$$

respectively, we consider joint inversion on z and θ . Figure 1 displays the objective function J with θ_2 fixed at 1 (left panel) and θ_1 fixed at 5 (right panel) with MAP points of the conditional posteriors (for different θ 's) and hyperdifferential sensitivities indicated by dots.

We observe several trends which illustrate the interpretation of HDSA:

- The joint objective indicates a stronger dependency between z and θ_1 in comparison to z and θ_2 . Corresponding to this, we observe a much larger sensitivity for θ_1 (given by the red-yellow color bar in the left panel) than θ_2 (given by the red-yellow color bar in the right panel).
- The sensitivities vary with θ showing that it is measuring local correlations which vary with θ . We observe the greatest sensitivity for small values of θ_1 which is consistent with the visual observation that the MAP point of $\pi_{\text{cond}}(z|\mathbf{d}; \bar{\theta})$ varies more for small values of θ_1 .

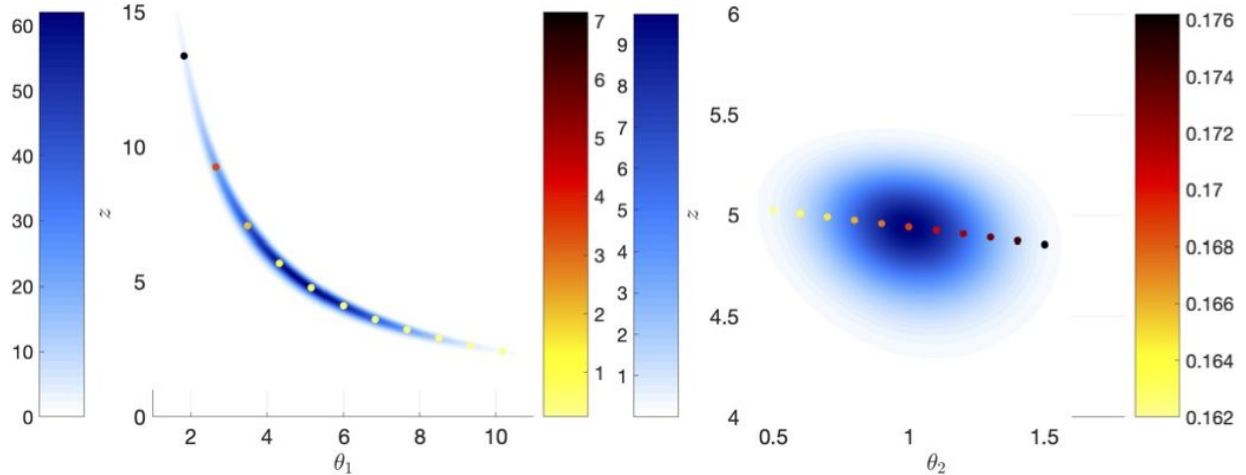


FIG. 1: Objective function J with θ_2 fixed at 1 (left panel) and θ_1 fixed at 5 (right panel). In each panel, the shading (in the blue color scale) indicates the joint objective while the solid dots (in the red-yellow color scale) denote the maximum a posterior probability (MAP) point z^* of $\pi_{\text{cond}}(z|\mathbf{d}; \bar{\theta})$. The color of the dot (in the red-yellow color scale) indicates the hyperdifferential sensitivities for the MAP point with respect to θ_1 (left panel) and θ_2 (right panel).

- There is little variation in the sensitivities in the right panel where the distribution is approximately Gaussian (since local correlations do not vary in Gaussian distributions) and much greater variation in the left panel where the distribution is highly non-Gaussian.

4. HDSEA AND LIKELIHOOD INFORMED SUBSPACES

Having motivated the postoptimality sensitivities for Bayesian inverse problems, this section recalls the computation of sensitivities in the likelihood informed subspace [15] and, demonstrates how this computation may be done concurrently with approximate posterior sampling via the Laplace approximation.

4.1 Computing the Projected Sensitivity Indices

For ill-posed inverse problems, the postoptimality sensitivities may be dominated by the directions which are poorly informed by data. These directions are given by the eigenvectors corresponding to the smallest eigenvalues of $\mathcal{H}$. To compute sensitivities in the directions informed by data, we follow [15] and introduce the projection of the sensitivity operator onto the likelihood informed subspace (LIS).

The LIS was used for dimension reduction in nonlinear Bayesian inverse problems in [37]. It defines the informed directions by the r leading eigenvectors of the generalized eigenvalue problem,

$$\mathcal{H}_M \mathbf{v}_j = \lambda_j \Gamma_z^{-1} \mathbf{v}_j, \quad j = 1, 2, \dots, m, \quad (8)$$

where

$$\mathcal{H}_M = \nabla_{zz} M(\mathbf{z}^*, \bar{\boldsymbol{\theta}}), \quad (9)$$

with M as defined in Eq. (6). Multiplying Eq. (8) by $\mathbf{v}_j^T$ to compute the Rayleigh quotient,

$$\lambda_j = \frac{\mathbf{v}_j^T \mathcal{H}_M \mathbf{v}_j}{\mathbf{v}_j^T \Gamma_z^{-1} \mathbf{v}_j},$$

provides an interpretation of the eigenvalue as a ratio of contributions from the likelihood (since $\mathcal{H}_M$ is the negative log likelihood Hessian) and the prior. The eigenvectors define the corresponding directions; i.e., the eigenvectors corresponding to large eigenvalues specify directions in the parameter space which are informed by the likelihood.

The truncation rank r is application dependent and dictates the computational complexity. For ill-posed problems, r may be small as a result of data sparsity. In this case, one can use efficient Krylov or randomized methods to solve the eigenvalue problem [15]. To solve Eq. (8), we employ the “two pass” randomized algorithm introduced in [38]. We define a projector onto the LIS as

$$\mathcal{P} = \mathbf{V}_r \mathbf{V}_r^T \Gamma_z^{-1}, \quad (10)$$

where the columns of $\mathbf{V}_r$ are given by the r leading eigenvectors in Eq. (8). Then the sensitivities in directions informed by the likelihood, see [15] for details, are given by

$$S_i = \|\mathcal{P} \mathcal{H}^{-1} \mathbf{B} \mathbf{e}_i\|_{\mathbf{W}_Z} = \sqrt{\sum_{k=1}^r \sum_{j=1}^r \left(\frac{\mathbf{v}_j^T \mathbf{B} \mathbf{e}_i}{1 + \lambda_j} \right) \left(\frac{\mathbf{v}_k^T \mathbf{B} \mathbf{e}_i}{1 + \lambda_k} \right) \mathbf{v}_k^T \mathbf{W}_Z \mathbf{v}_j}, \quad i = 1, 2, \dots, n, \quad (11)$$

where $\mathbf{e}_i$ denotes the i th canonical basis vector and $\mathbf{W}_Z$ is a symmetric positive definite matrix defining a weighted inner product, for instance, a mass matrix if z is discretized with finite elements. We interpret S_i as the influence of the i th parameter on the MAP point projected onto the LIS. Equation (11) follows from considering a spectral representation of the $\mathcal{H}$ in combination with the Sherman-Morrison-Woodbury formula.

4.2 Gaussian Approximation of the Posterior

We seek to efficiently approximate the posterior by leveraging the LIS computation in the postoptimality sensitivity computations. If the mapping from the parameter space to observations $\mathbf{f}$ is nonlinear, the posterior distribution is not Gaussian. As a result, the posterior cannot conveniently be described in terms of its mean and covariance. This motivates alternative methods for describing the posterior. State-of-the-art Markov chain Monte Carlo (MCMC) methods may be used to exactly specify the posterior via sampling. However, MCMC methods often require repeated evaluations of the parameter-to-observable map. In the applications that we are interested in, $\mathbf{f}$ maps onto observations of a (possibly) nonlinear PDE. For computational efficiency, we use the Laplace approximation to the posterior distribution (see, e.g., [18]). Specifically, we assume the posterior is approximately Gaussian with mean $\mathbf{z}^*$ and covariance matrix $\mathbf{\Gamma}_{\text{post}} = \mathbf{H}^{-1}$, where $\mathbf{H}$ is the Hessian of the objective function (5) evaluated at $\mathbf{z}^*$. We write the approximate posterior as $\hat{\pi}_{\text{cond}}$ and define it as

$$\hat{\pi}_{\text{cond}}(\mathbf{z}|\mathbf{d};\bar{\boldsymbol{\theta}}) \sim \mathcal{N}(\mathbf{z}^*, \mathbf{\Gamma}_{\text{post}}). \quad (12)$$

Explicit construction of the Hessian/covariance matrix is infeasible as they are only accessible via matrix-vector products. Specifically, computing $\mathbf{H}\mathbf{x}$, for $\mathbf{x} \in \mathbb{R}^m$, requires two linear PDE solves (called the incremental state and incremental adjoint equations), and computing $\mathbf{H}^{-1}\mathbf{x}$, for $\mathbf{x} \in \mathbb{R}^m$, requires an iterative solver needing multiple matrix-vector products with $\mathbf{H}$. Thus we resort to a low-rank approximation of the Hessian to efficiently store and compute with the posterior covariance matrix.

4.2.1 Efficient Representation of Posterior Covariance

To efficiently represent the inverse Hessian $\mathbf{H}^{-1}$, observe that the Hessian is the sum of the data misfit Hessian $\mathbf{H}_M$ (9) evaluated at $\mathbf{z}^*$ and $\mathbf{\Gamma}_z^{-1}$. In many ill-posed inverse problems of interest, the LIS exhibits fast spectral decay as a symptom of directions in the parameter space poorly informed by the data [39]. Consequently, we can represent the data misfit Hessian with a low rank approximation in the LIS basis.

Let $\mathbf{V}$ and $\mathbf{\Lambda}$ denote the $m \times m$ matrices containing all eigenvectors and eigenvalues from the LIS generalized eigenvalue problem (8). In practice we will truncate after the leading r eigenvalues, but all are retained for the derivation below to provide error analysis. Following the procedure of [40], observe that $\mathbf{H}_M = \mathbf{\Gamma}_z^{-1}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\mathbf{\Gamma}_z^{-1}$ and that the inverse Hessian can be written as

$$\mathbf{H}^{-1} = (\mathbf{\Gamma}_z^{-1} + \mathbf{H}_M)^{-1} = \mathbf{\Gamma}_z - \left(\mathbf{I} + \mathbf{\Gamma}_z^{-1}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \right)^{-1},$$

Applying the Sherman-Morrison-Woodbury formula to invert $\mathbf{I} + \mathbf{\Gamma}_z^{-1}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$, exploiting the orthogonality of $\mathbf{V}$ in the $\mathbf{\Gamma}_z^{-1}$ inner product, and truncating with a rank r approximation, we can write

$$\mathbf{H}^{-1} = (\mathbf{\Gamma}_z^{-1} + \mathbf{H}_M)^{-1} = \mathbf{\Gamma}_z - \mathbf{V}_r \mathbf{D}_r \mathbf{V}_r^T + \mathcal{O}\left(\sum_{i=r+1}^m \frac{\lambda_i}{1 + \lambda_i}\right), \quad (13)$$

where $\mathbf{D}_r = \text{diag}(\lambda_1/(\lambda_1 + 1), \dots, \lambda_r/(\lambda_r + 1))$. For problems admitting a rapid spectral decay, the $\mathcal{O}(\lambda_{r+1})$ truncation error is small and hence the LIS subspace computation provides an efficient way to approximate the inverse Hessian $\mathbf{H}^{-1}$ with a computational cost scaling with the LIS subspace dimension r .

We define the approximate posterior covariance as

$$\mathbf{\Gamma}_{\text{post}} = \mathbf{\Gamma}_z - \mathbf{V}_r \mathbf{D}_r \mathbf{V}_r^T. \quad (14)$$

Equation (14) shows how the eigenpairs calculated for the sensitivity indices can be used for the posterior approximation. This implies that approximate posterior sampling is a computational by-product of computing sensitivities. In the next subsection, we describe the methods for quantifying the uncertainty in the solution through sampling and variance.

4.2.2 Sampling from Approximate Posterior Distribution and Posterior Variance

Recall that the posterior distribution represents the information gained by combining the data (likelihood) and our beliefs (prior). To represent the reduction of uncertainty we draw samples from the prior and posterior and compute their variances (the diagonals of $\mathbf{\Gamma}_z$ and $\mathbf{\Gamma}_{\text{post}}$). A sample z from $\mathcal{N}(\mu_z, \mathbf{\Gamma}_z)$ can be computed via the formula

$$z_{\text{prior}} = \mu_z + Cx, \quad (15)$$

where C is any matrix satisfying $\mathbf{\Gamma}_z = CC^T$ and $x \sim \mathcal{N}(0, I)$. The matrix C can be calculated in a multitude of ways. One method is to compute C as a Cholesky factor. In the case of large-scale problems computing the Cholesky factor is not feasible. To remedy this, we follow the preconditioned Lanczos method for sampling multivariate Gaussian distributions proposed by Chow and Saad in [41]. This method utilizes the Lanczos algorithm to approximate the action of $\mathbf{\Gamma}_z^{1/2}$ onto a vector.

To sample from the Laplace approximation of the posterior $\mathcal{N}(z^*, \mathbf{\Gamma}_{\text{post}})$ we follow the approach described in Section 4.3.2 of [40]. Given a sample z_{prior} from the prior distribution, a sample from $\mathcal{N}(z^*, \mathbf{\Gamma}_{\text{post}})$ is given by

$$z_{\text{post}} = \left(I_m - \mathbf{V}_r \mathbf{S}_r \mathbf{V}_r^T \mathbf{\Gamma}_z^{-1} \right) (z_{\text{prior}} - \mu_z) + z^*, \quad (16)$$

where $\mathbf{S}_r = I_r - (\mathbf{\Lambda}_r + I_r)^{-1/2}$.

Since our approximate posterior distribution is Gaussian with covariance $\mathbf{\Gamma}_{\text{post}} = \mathbf{\Gamma}_z - \mathbf{V}_r \mathbf{D}_r \mathbf{V}_r^T$, the posterior variance is given by

$$\text{diag}(\mathbf{\Gamma}_{\text{post}}) = \text{diag}(\mathbf{\Gamma}_z) - \text{diag}(\mathbf{V}_r \mathbf{D}_r \mathbf{V}_r^T). \quad (17)$$

To approximate $\text{diag}(\mathbf{\Gamma}_z)$, we use the Diag++ algorithm ([42], Algorithm 1). Given a budget of s_D matrix-vector products, Diag++ approximates the diagonal of $\mathbf{\Gamma}_z$ by sketching its range space using $s_D/3$ matrix-vector products, computing the diagonal of a low-rank approximation using $s_D/3$ matrix-vector products, and estimating the diagonal of $\mathbf{\Gamma}_z$ minus its low-rank approximation using $s_D/3$ matrix-vector products. Since $\mathbf{\Gamma}_z$ is the inverse of a differential operator, the main cost of Diag++ is s_D linear solves.

The diagonal of $\mathbf{V}_r \mathbf{D}_r \mathbf{V}_r^T$ is calculated directly via the formula

$$\text{diag}(\mathbf{V}_r \mathbf{D}_r \mathbf{V}_r^T) = \sum_{k=1}^r \left(\frac{\lambda_k}{1 + \lambda_k} \mathbf{v}_k \right) \odot \mathbf{v}_k. \quad (18)$$

5. ALGORITHMS

In this section we show our algorithm for computing HDSA indices and the Gaussian approximation. The algorithm consists of four major components: the MAP point, the generalized eigenvalue problem, the sensitivity indices, and the Gaussian approximation. The MAP point is key to performing HDSA and computing the Gaussian approximation of the posterior (as seen in Section 4). In the subsections below, we summarize the computational details of each major component. Note that this discussion is based on the assumptions for the data d , prior z , and auxiliary parameters θ as stated in Section 2.1.

5.1 Computation of the MAP Point

Recall that the MAP point is the solution of the optimization problem (5). This is a nonlinear least-squares problem and we employ the truncated Newton CG trust region method [34]. This corresponds to the loop in lines 4–8 in Algorithm 1. Given an iterate z_k , we compute the gradient of the objective function at z_k , g_k (which requires a forward and adjoint PDE solve) and check for convergence to a stationary point. If z_k is not a stationary point, we form the local model of the objective function $m_k(s) = J(s) + g_k^T s + (1/2)s^T H_k s$, where H_k is the Hessian of J evaluated at z_k . Then the trust region subproblem is solved using the CG-Steihaug algorithm. Note that we compute exact Hessian vector products using incremental adjoint equations. To calculate z_{k+1} the reduction in the local model

Algorithm 1: Workflow for HDSA in Bayesian inverse problems**Input:** $r_0 \in \mathbb{N}$, $\Delta r \in \mathbb{N}$, $p \in \mathbb{N}$, and $\lambda_{\min} \in \mathbb{R}$.

-
- 1

 - 2 **Solve MAP point** $\min_{\mathbf{z} \in \mathbb{R}^m} J(\mathbf{z}, \bar{\boldsymbol{\theta}}) := M(\mathbf{z}, \bar{\boldsymbol{\theta}}) + R(\mathbf{z})$
 - 3 *Truncated Newton CG trust region procedure:*
 - 4 **for** $k = 0, 1, 2, \dots$ **do**
 - 5 Compute $\mathbf{g}_k = \nabla_{\mathbf{z}} J$ and $\mathbf{H}_k = \nabla_{\mathbf{z}, \mathbf{z}} J$ to form a local model $\mathbf{m}_k$
 - 6 Minimize the local model $\mathbf{m}_k$ subject to the trust region constraint
 - 7 Update trust region radius
 - 8 **end**
 - 9

 - 10 **Compute LIS eigenpairs** $\mathcal{H}_M \mathbf{v}_j = \lambda_j \Gamma_{\mathbf{z}}^{-1} \mathbf{v}_j, j = 1, 2, \dots, r$
 - 11 *Randomized generalized Hermitian eigenvalue procedure:*
 - 12 Set $\lambda_{\text{iter}} = \infty$, Sample a random matrix $\Omega \in \mathbb{R}^{n \times (r_0 - \Delta r + p)}$
 - 13 **while** $\lambda_{\text{iter}} > \lambda_{\min}$ **do**
 - 14 Augment Ω with Δr additional columns
 - 15 Compute:
 - 16 sketch $\mathbf{Y} = \Gamma_{\mathbf{z}} \mathcal{H}_M \Omega$
 - 17 projection $\mathbf{T}$ onto range of $\mathbf{Y}$
 - 18 eigendecomposition $\mathbf{T} = \mathbf{S} \mathbf{\Lambda} \mathbf{S}^T$
 - 19 Set $\lambda_{\text{iter}} = (\mathbf{\Lambda})_{r, r}$
 - 20 **end**
 - 21 Compute $\mathbf{v}_j = \mathbf{Q} \mathbf{s}_j, 1 \leq j \leq r$, where $\mathbf{s}_j$ is the j th column of $\mathbf{S}$
 - 22

 - 23 **Calculate sensitivity indices** $S_i = \|\mathcal{P} \mathcal{H}^{-1} \mathcal{B} \mathbf{e}_i\|_{\mathbf{W}_Z}$ for $1 \leq i \leq n$
 - 24 Compute:
 - 25 $\mathcal{B} \mathbf{e}_i = \nabla_{\mathbf{z}, \boldsymbol{\theta}} J_{\text{post}}(\mathbf{z}^*, \bar{\boldsymbol{\theta}}) \mathbf{e}_i, i = 1, 2, \dots, n$
 - 26 $S_i, i = 1, 2, \dots, n$, using Eq. (11)
 - 27

 - 28 **Laplace approximation** $\hat{\pi}_{\text{post}}(\mathbf{z} | \mathbf{d}; \bar{\boldsymbol{\theta}}) \sim \mathcal{N}(\mathbf{z}^*, \Gamma_{\text{post}}), \quad \Gamma_{\text{post}} = \Gamma_{\mathbf{z}} - \mathbf{V}_r \mathbf{D}_r \mathbf{V}_r^T$
 - 29 *Sampling procedure:* $\mathbf{z}_{\text{post}} = \left(\mathbf{I}_m - \mathbf{V}_r \mathbf{S}_r \mathbf{V}_r^T \Gamma_{\mathbf{z}}^{-1} \right) (\mathbf{z}_{\text{prior}} - \boldsymbol{\mu}_z) + \mathbf{z}^*$, where $\mathbf{z}_{\text{prior}}$ is generated from Eq. (15)
 - 30 *Variance procedure:* Estimate $\text{diag}(\Gamma_{\text{post}})$ using ([42], Algorithm 1) and Eq. (18)
-

is considered and then the trust region is updated. As stated in the Introduction, the MAP is only a point estimate of the unknown and is not sufficient for describing relationships between the unknown and auxiliary parameters. To address this the relationship we employ HDSA to the optimization problem (5).

5.2 Computing the Eigenpairs

Solving the Hermitian generalized eigenvalue problem (HGEVP) (8) is important to computing the LIS for HDSA and the Gaussian approximation of the posterior. In this subsection, we elaborate upon the calculations of the HGEVP that correspond to lines 12–21 in Algorithm 1.

The randomized generalized Hermitian eigenvalue procedure is a modification of the “double pass” algorithm [38]. Line 16 sketches the range of $\Gamma_{\mathbf{z}} \mathcal{H}_M$ by multiplying it by a collection of random vectors Ω . Next, a $\Gamma_{\mathbf{z}}^{-1}$ orthogonal basis for the range of $\mathbf{Y}$ is computed using dense linear algebra. The second round of matrix-vector products with $\mathcal{H}_M$ computes the low-rank projection onto the range space of $\mathbf{Y}$. Dense linear algebra is done to compute the

eigendecomposition of the projection $\mathbf{T}$. If the smallest eigenvalue is above the target eigenvalue threshold then the loop continues by augmenting the columns of $\mathbf{\Omega}$, or else the loop ends. The desired eigenvectors are obtained through $\mathbf{v}_j = \mathbf{Q}\mathbf{s}_j$. The sketch dominates the computational cost of the procedure and requires a $2(r + p)$ matrix-vector product with $\mathbf{H}_M$ where r is the final target rank and p is the oversampling parameter. Note that several of the computations in the randomized generalized Hermitian eigenvalue procedure can be easily parallelized. The details of this procedure are given in [15].

5.3 Computing the Sensitivity Indices

The sensitivity indices calculation corresponds to lines 25 and 26. Since we have the eigenpairs of Eq. (8), the dominant cost of the sensitivity calculation is the action of $\mathbf{B}$ onto basis vectors $\mathbf{e}_i$. The matrix $\mathbf{B}$ corresponds to the Jacobian of $\nabla_{\mathbf{z}} J(\mathbf{z}^*, \bar{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\theta}$, so each $\mathbf{B}\mathbf{e}_i$ is two additional PDE solves. Therefore, computing $\{\mathbf{B}\mathbf{e}_i\}_{i=1}^n$ costs $2n$ PDE solves where n is the auxiliary parameter dimension.

5.4 Computing the Laplace Approximation

Approximating the posterior utilizes the calculation of the eigenpairs in Eq. (8). To sample from the posterior we draw from the prior using the Lanczos method as explained in Section 4.2.1. Using this method without a preconditioner, the cost of approximating prior samples $\mathbf{\Gamma}_{\mathbf{z}}^{1/2}\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is a standard normal Gaussian sample, is N_{lanczos} matrix-vector products with $\mathbf{\Gamma}_{\mathbf{z}}$. For the variance, the dominant cost is the Diag++ algorithm requiring s_D matrix-vector products with $\mathbf{\Gamma}_{\mathbf{z}}$.

6. APPLICATION TO ICE-SHEET BEDROCK INVERSION

In this section of the paper, we apply the numerical methods and algorithms described in Sections 4 and 5 to an application in ice-sheet inversion. The goal is to use surface velocity measurements on the Greenland ice-sheet to compute an estimate of the uncertain bedrock topography beneath the ice, quantify the influence of auxiliary parameters on the solution, and provide a statistical characterization. To accomplish this we first introduce the shallow ice-sheet model in Section 6.1 and the inverse problem that we are solving in Section 6.2. The inverse problem formulation is followed by the MAP estimate results in Section 6.3. The MAP estimate drives the HDSA calculation which we apply to the inversion of the bedrock with respect to the log basal friction (a spatial coefficient representing the interaction between the land ice and the bedrock) and forcing. This is detailed in Section 6.5. In Section 6.6 we provide results for posterior samples and variance.

6.1 Ice-Sheet Model

Developing high-fidelity models for ice-sheets is important for global climate modeling and prediction of sea level rise. Simulation of ice-sheets such as Antarctica and Greenland requires extensive computational resources to solve nonlinear partial differential equations (PDEs) on fine spatial meshes. A high-fidelity model for ice-sheet dynamics is typically derived by considering the ice-sheet to behave as a viscous shear-thinning fluid in a low Reynolds-number flow. This results in the nonlinear Stokes equation in three spatial dimensions coupled with equations for the temperature distribution in the ice and the thickness of the ice. From these fundamental equations, a variety of assumptions are made to yield models of different physics fidelities, a comprehensive overview of which is beyond the scope of this article. To demonstrate HDSA in this article we adopt the shallow ice approximation (SIA) [43,44] with the isothermal assumption [45].

6.2 Mathematical Formulation of Inverse Problem

We consider the following inverse problem for the discretized bedrock topography $\mathbf{b}$ of the ice-sheet,

$$\min_{\mathbf{b} \in \mathbb{R}^m} J(\mathbf{b}, \bar{\boldsymbol{\theta}}) = \frac{1}{2} \|\mathbf{f}(\mathbf{b}, \bar{\boldsymbol{\theta}}) - \mathbf{d}\|_{\mathbf{\Gamma}_{\text{noise}}^{-1}}^2 + \frac{1}{2} \|\mathbf{b} - \mathbf{b}_0\|_{\mathbf{\Gamma}_{\mathbf{b}}^{-1}}^2, \quad (19)$$

where $\mathbf{b}_0$, $\mathbf{\Gamma}_b$, $\bar{\boldsymbol{\theta}}$, and $\mathbf{d}$ correspond to the prior bedrock topography mean and covariance, the nominal value of the auxiliary parameters, and the data, respectively. The forward operator $\mathbf{f}(\mathbf{b}, \bar{\boldsymbol{\theta}})$ is the mapping from the bedrock topography and the auxiliary parameters (log basal friction and forcing) to surface velocity in Eq. (22).

Shallow ice approximation. The inverse problem (19) is constrained by the SIA equation

$$\begin{aligned} \frac{\partial s}{\partial t} - \nabla \cdot (Q(s) \nabla s) &= h_{\text{flux}}(\boldsymbol{\theta}) & \text{on } \Omega \times (0, T], \\ \nabla s \cdot \mathbf{n} &= 0 & \text{on } \partial\Omega \times (0, T], \\ s &= s_0 & \text{on } \Omega \times \{0\}, \end{aligned} \quad (20)$$

$$Q(s) = e^{-\gamma(\boldsymbol{\theta})} \rho g (s - b)^2 + \frac{2A\rho^3 g^3}{5} (s - b)^5 \|\nabla s\|^2, \quad (21)$$

where s is the surface height and $Q(s)$ is a velocity field derived from the Stokes velocity model. The scalar parameter $\rho = 910$ (kg/m³) is the density of ice, $g = 9.81$ (m/s²) is the acceleration of gravity, and $A = 10^{-16}$ (Pa⁻³/s) is a flow rate factor.

Parameter to observable map. The parameters b , γ , and h_{flux} are inputs to the shallow ice approximation (20) which predicts the surface height s . The observation operator maps the surface height (the PDE solution) to the ice-sheet surface velocity which is modeled via

$$\mathbf{v}(b, \gamma, h_{\text{flux}}) = -\frac{1}{2} A \rho^3 g^3 (s(b, \gamma, h_{\text{flux}}) - b)^4 \|\nabla s(b, \gamma, h_{\text{flux}})\|^2 \nabla s(b, \gamma, h_{\text{flux}}). \quad (22)$$

We refer the reader to [11,27,35] for additional details concerning the SIA model.

Domain. We consider the inversion of bedrock topography $\mathbf{b}$ in a 550×450 km region of Greenland specified in Fig. 2. The surface velocity $\mathbf{v}$ is calculated as a function of the surface height [the PDE (20) solution] via (22) and compared with velocity observations from satellites.

Auxiliary parameters. The log basal friction γ and forcing h_{flux} are given by

$$\gamma(\boldsymbol{\theta}) = \tilde{\gamma} \delta_1(\boldsymbol{\theta}) \quad \text{and} \quad h_{\text{flux}}(\boldsymbol{\theta}) = \tilde{h}_{\text{flux}} \delta_2(\boldsymbol{\theta}), \quad (23)$$

where

$$\delta_i(\boldsymbol{\theta}) = 1 + 0.2 \sum_{j=1}^n \theta_{((i-1)n+j)} \phi_j, \quad i = 1, 2, \quad (24)$$

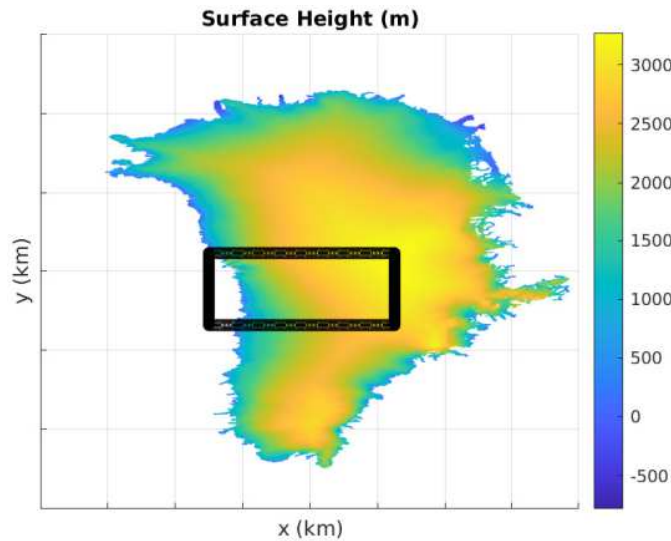


FIG. 2: Ice surface height data for the Greenland ice-sheet. The black box indicates the region analyzed in this article.

are parameterized perturbations; $\{\phi_j\}_{j=1}^n$ are linear finite element basis functions defined on a 31×31 mesh of the domain Ω . This implies that $n = 961$ so there are 1922 auxiliary parameters in total. The nominal parameters are $\bar{\theta} = 0 \in \mathbb{R}^{1922}$ yielding $\gamma(0) = \tilde{\gamma}$ and $h_{\text{flux}}(0) = \tilde{h}$ (magnitudes are shown in Fig. 3). For simplicity, we provide a summary of relevant variables and constants for the inverse problem in Table 1.

Prior parameters. We assume a Gaussian prior with mean zero and covariance Γ_b given by the matrix representation of the operator $(-\beta\Delta + \alpha\mathcal{I})^{-2}$ with constants $\beta = 10^{-2}$ and $\alpha = 9 \times 10^{-7}$. The Laplacian operator Δ and identity operator $\mathcal{I}$ are equipped with zero Neumann boundary conditions. The parameter β controls the smoothness of prior samples and α controls the variance of the prior samples.

Synthetic data generation. To facilitate a numerical demonstration, data are generated from the model using “true” bedrock topography, which is subsequently considered unknown as we solve and analyze the inverse problem to reconstruct it. The bedrock, ice thickness, accumulation/ablation, and log basal friction are taken from [46]. To aid in numerical performance we presmooth the log basal friction and bedrock topography using a local averaging technique. The accumulation/ablation forcing term from [46] is stationary. The forcing term is depicted in the top left panel of Fig. 3.

We generate data by solving Eq. (20) on a 101×101 mesh with 121 time steps from $t = 0$ to $t = T = 10$ years and evaluating the surface velocity given in Eq. (22). This is a spatial resolution of 5.45 kilometers and a time resolution of 0.083 years. To avoid an “inverse crime,” the data are interpolated onto a 71×71 mesh with 61 time steps, and 5% Gaussian noise is added. The noise covariance is taken as $\Gamma_{\text{noise}} = \sigma_{\text{noise}}^2 \mathbf{I}$ with $\sigma_{\text{noise}} = 50$. We solve the inverse problem on a 71×71 mesh with 61 times steps over 10 years.

6.3 Results for MAP Estimate

We compute the MAP point by solving Eq. (19) for $\bar{\theta} = 0$ using a truncated conjugate gradient trust region algorithm as outlined in lines 4–8 in Algorithm 1. Exact gradients and Hessian vector products are computed using adjoint and incremental adjoint equations. The initial guess for the bedrock topography was given by a highly smoothed version of the “true” bedrock topography.

In Table 2 the optimization history for the computation of the MAP point is given. It terminates upon achieving a gradient norm less than 10^{-7} . We see that the gradient norm has reduced nine orders of magnitude and that the step size decreases as the iterations increase, indicating the convergence of the optimizer.

TABLE 1: Table of relevant variables and constants for the inverse problem (19)–(24)

Constants	Values
β	10^{-2}
α	9×10^{-7}
Γ_b	Matrix representation of $(-\gamma\Delta + \alpha)^{-2}$
Γ_{noise}	$(50)^2 \mathbf{I}$
d	Surface velocity data (m/s)
A	Flow rate factor 10^{-16} (Pa^{-3}/s)
ρ	Density of ice 910 (kg/m^3)
g	Acceleration of gravity 9.81 (m/s^2)
γ	Log basal friction
h_{flux}	Forcing
Variables	Values
b	Bedrock topography (m)
v	Surface velocity (m/s)
s	Surface height (m)

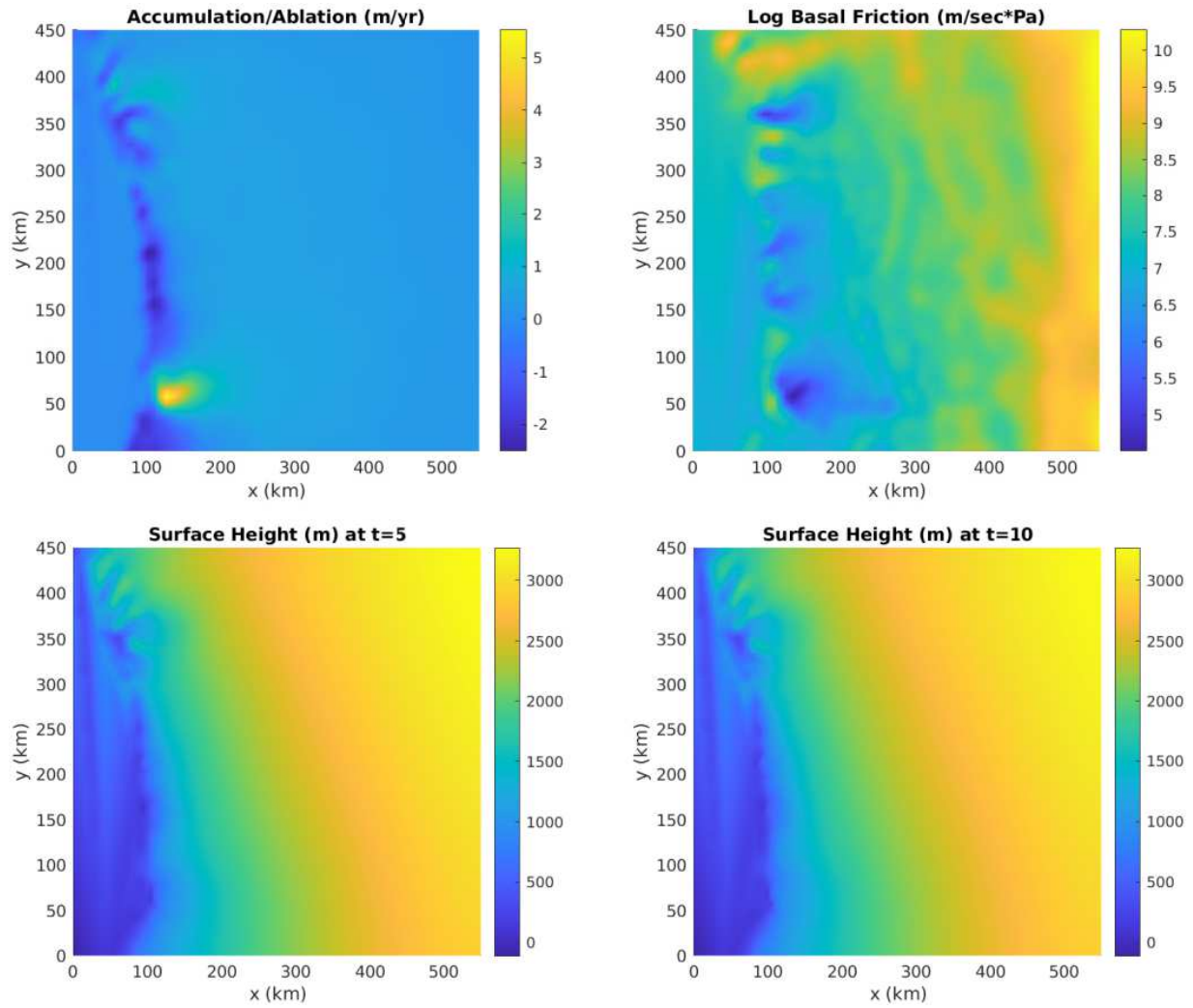


FIG. 3: Top left: accumulation/ablation forcing term; top right: log basal friction; bottom: surface height at time $t = 5$ (left) and $t = 10$ (right)

TABLE 2: Iteration history for calculating the MAP point

Iteration	Objective	Gradient norm	Step size
0	5.83×10^4	2.66×10^1	N/A
2	1.65×10^4	1.0352×10^1	2.5×10^3
4	6.36×10^3	6.089	2.5×10^3
10	3.82×10^3	8.55×10^{-1}	6.10×10^2
25	3.22×10^3	9.72×10^{-8}	7.36×10^{-2}

Visualizations of the initial guess, MAP estimate, true bedrock topography, and difference are provided in Fig. 4. The MAP estimate captures a majority of the features present in the true bedrock topography and the largest deviations are in the mountain in the northwest corner which is reflected in the diff plot. The bedrock is rough in that region and due to the smoothing prior, the reconstruction fails to capture some of the fine-scale features. It is important to note that computing the MAP point provides an estimate of the unknown bedrock topography, but it does not address

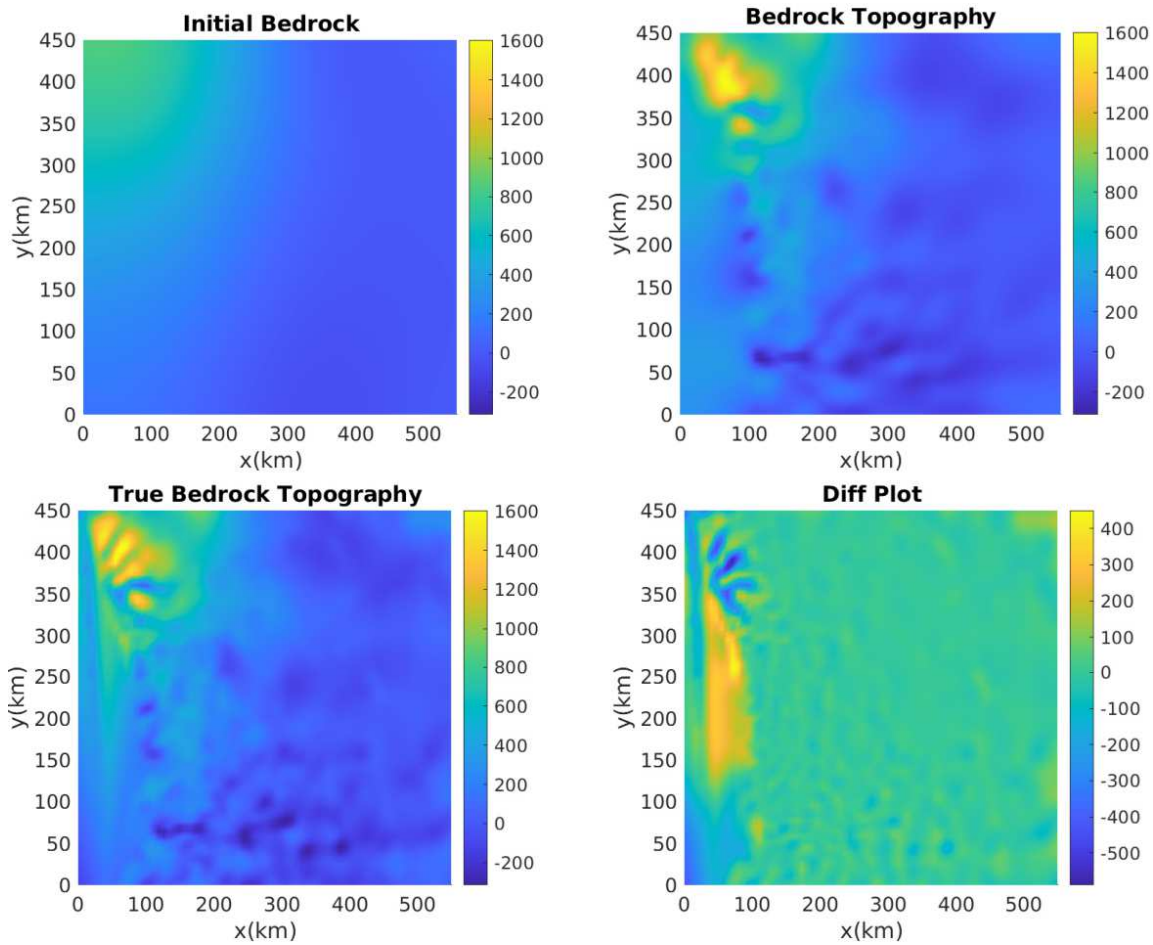


FIG. 4: Top left: initial bedrock topography, top right: MAP point, bottom left: true bedrock topography, bottom right: the difference between true and estimated bedrock topography

how the uncertainty of the auxiliary parameters (log basal friction and forcing) will influence the uncertainty in the bedrock topography. The influence of uncertainty in the log basal friction and the forcing is explored through HDSA.

6.4 Result for the LIS

The LIS eigenpairs were calculated using lines 12–21 in Algorithm 1 with an initial target rank $r_0 = 230$, rank increment $\Delta r = 256$, oversampling parameter $p = 26$, and minimum of eigenvalue threshold $\lambda_{\min} = 0.1$. The literature recommends $p \sim 20$, but we chose p , r_0 , and Δr to leverage parallelism with 16 compute nodes (16 cores per node). Our motivation for choosing $\lambda_{\min} = 0.1$ was to ensure that the decay of the eigenvalues was sufficient for capturing parameter directions influenced by the likelihood and prior. The decay was achieved within two iterations of the for loop in the randomized generalized Hermitian eigenvalue procedure. Therefore we computed a total of 742 eigenvalues which are shown in Fig. 5. To calculate the sensitivity indices and approximate the posterior covariance we used all 742 eigenvalues.

6.5 Results for Sensitivity Indices

Sensitivity indices for log basal friction and forcing are computed as outlined in lines 25 and 26 of Algorithm 1 and plotted in Fig. 6. We make the following observations about the sensitivity indices.

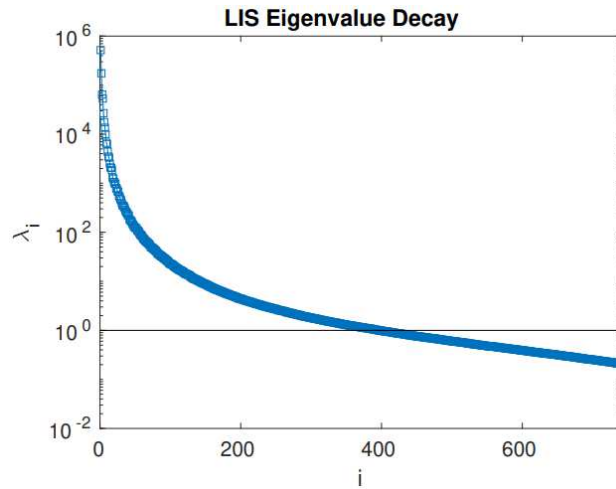


FIG. 5: Logarithmic plot of spectrum for Eq. (8). The line $\lambda = 1$ corresponds to an equal contribution from the likelihood and prior.

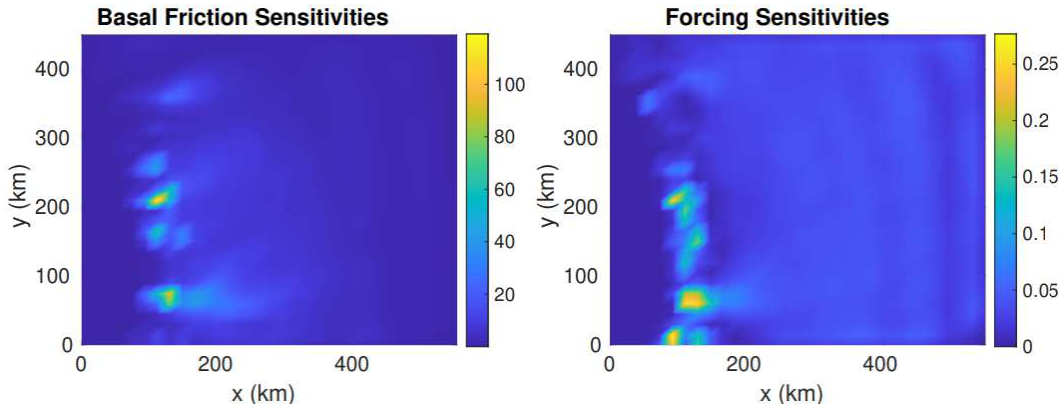


FIG. 6: (Left) Sensitivity heat map for the log basal friction. (Right) Sensitivity heat map for the forcing.

- The magnitude of the log basal friction indices is much greater than the forcing; i.e., changes in log basal friction are more impactful on bedrock topography estimation than the forcing. This is consistent with physical intuition since the log basal friction is representative of “slipping” between the ice-sheet and the underlying bedrock which is known to have a more significant effect over time horizons of 10–20 years [47].
- The interpretation of high-sensitivity regions in the log basal friction is that perturbing the log basal friction in those regions will lead to the largest changes in the bedrock topography estimate.
- The largest log basal sensitivities correspond to the largest differences in the reconstruction of the bedrock topography which demonstrates the joint correlation between z and θ (Theorem 1).
- In the context of joint Bayesian inversion of bedrock topography and log basal friction, the sensitivity indices correspond to correlations in the posterior. Correlations in the joint posterior distribution identify regions where inversion is difficult and may require more data collection.

6.6 Results for Posterior Samples and Variance

As a result of the eigenvectors calculations in the LIS, we compute approximate posterior samples and a variance estimate as described in lines 28–30 of Algorithm 1. We plot samples from the prior and posterior distribution in

Fig. 7. The prior variance computation uses the Diag++ algorithm, ([42], Algorithm 1), with parameters $s_D = 2100$ matrix-vector products. The approximation to the posterior variance is computed using the prior variance and Eq. (17) with $r = 742$. The approximate prior and posterior variances are plotted in Fig. 8; by contrasting the two plots one can gain insight into the reduction in uncertainty, from the prior to the posterior, obtained by solving the inverse problem.

The posterior samples capture most of the heterogeneous structure in the domain including the mountain in the top left corner. In addition, the posterior samples have a similar spatial pattern to the MAP. The well-captured regions correspond to areas where we have more confidence in our estimate of bedrock topography.

These observations are reflected in the plot of the posterior variance (Fig. 8, right panel). In the bottom left corner and the top left corner of the posterior variance plot, the uncertainty is much higher implying that the data are not as informative as the prior in those areas. This observation is also reflected in the inconsistent structure of the posterior samples in these regions. The highly uncertain areas also correspond to the highest-error regions in the diff plot in Fig. 4.

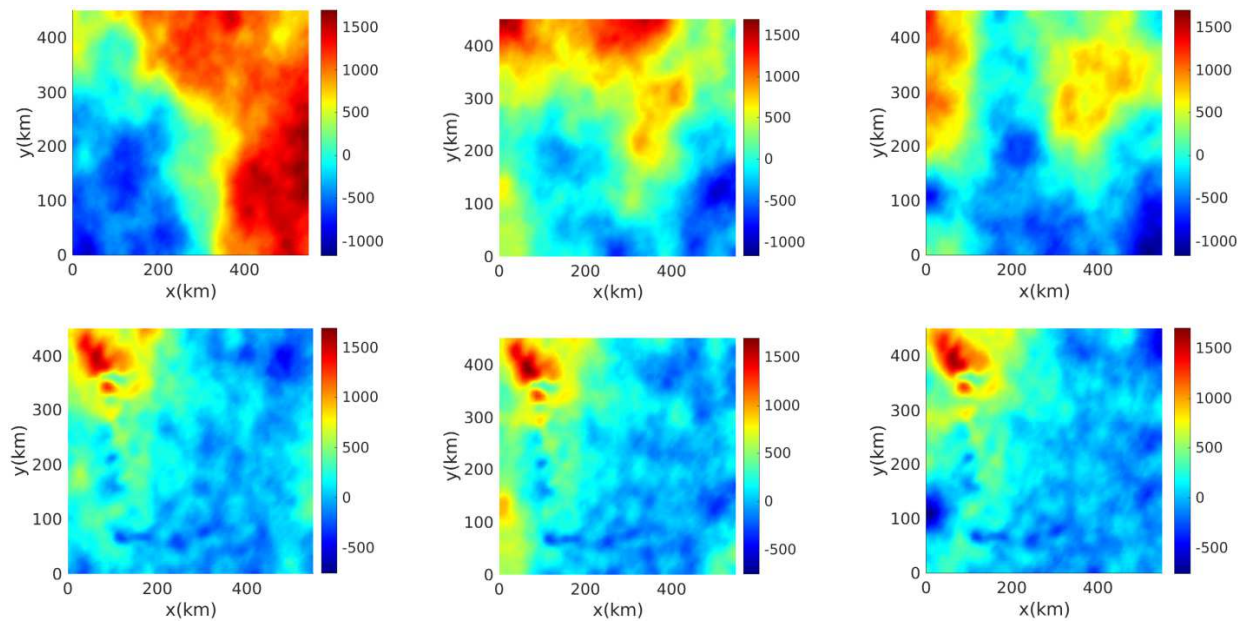


FIG. 7: Samples from the prior distribution, bottom: samples from the posterior distribution

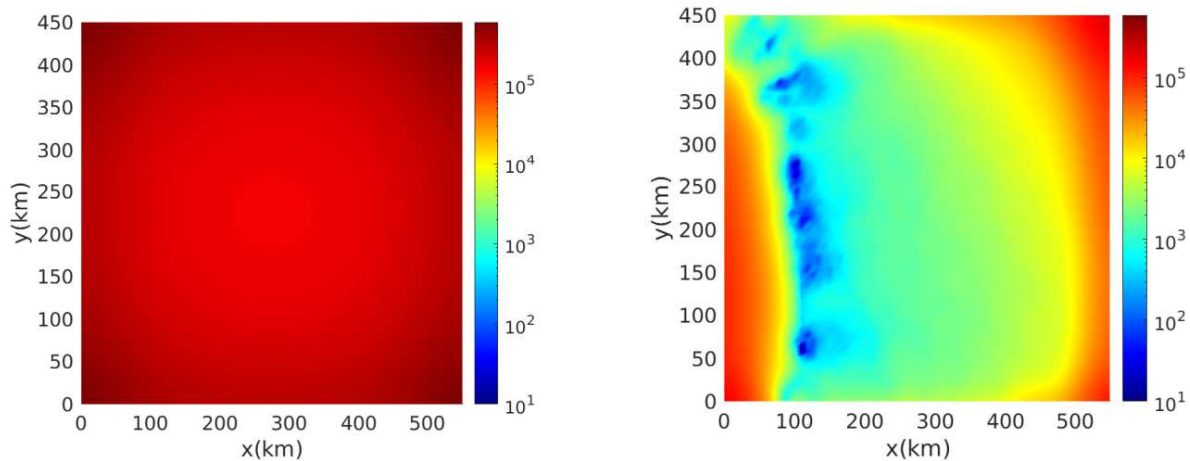


FIG. 8: Left: prior pointwise variance, right: posterior pointwise variance

7. CONCLUSIONS

This article presents a new algorithm that combines hyperdifferential sensitivity analysis (HDSA) and approximate sampling from the posterior distribution for Bayesian inverse problems. HDSA determines the influence of uncertainties on the solution of large-scale inverse problems and is shown to have a Bayesian interpretation in terms of correlations in the joint posterior distribution. By projecting onto the likelihood informed subspace (LIS), the sensitivities are computed efficiently and samples from the Laplace approximation of the posterior distribution are computed as a by-product of the LIS computation. Motivated by large-scale problems and by the added complexity of multiple sources of uncertainty, we demonstrate our methods on a nonlinear dynamic ice-sheet model whose parameters are based on realistic values from the Greenland ice-sheet. As a proof of concept, we generate synthetic data to invert on bedrock topography and demonstrate HDSA with respect to basal friction and surface forcing.

We conclude that basal friction is much more important than the surface forcing in the context of bedrock topography inversion. Although this makes intuitive sense from a physics perspective, HDSA quantifies the differences exactly and provides a spatial characterization of sensitivities for basal friction and surface forcing. From a Bayesian perspective, this implies that the bedrock topography and basal friction are highly correlated in their joint posterior. With these conclusions in hand, one may consider the implementation of higher-fidelity basal friction modeling or reprioritization of the data acquisition program.

HDSA insight is enabled through several computational foundations. Adjoint-based derivatives are used to calculate gradients and Hessians as part of a trust region Newton-CG optimization algorithm, which is predicated on parallel numerical linear algebra. A randomized generalized Hermitian eigenvalue solver is used to efficiently compute the LIS projector asynchronously.

Inversion using observed data (instead of synthetically generated data) can create solver convergence issues making the interpretation of HDSA questionable. However, projecting the sensitivities onto the LIS subspace provides a clear interpretation as discussed in [15].

ACKNOWLEDGMENTS

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC (NTESS), a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration (DOE/NNSA) under contract DE-NA0003525. This written work is authored by an employee of NTESS. The employee, not NTESS, owns the right, title and interest in and to the written work and is responsible for its contents. Any subjective views or opinions that might be expressed in the written work do not necessarily represent the views of the U.S. Government. The publisher acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this written work or allow others to do so, for U.S. Government purposes. The DOE will provide public access to results of federally sponsored research in accordance with the DOE Public Access Plan.

Support for Jakeman and Perego was provided through the Scientific Discovery through Advanced Computing (SciDAC) program funded by the US Department of Energy (DOE), Office of Science, Advanced Scientific Computing Research and Biological and Environmental Research Programs.

Saibaba and Reese were partially supported by the National Science Foundation through the award DMS-1845406.

Hart and van Bloemen Waanders were supported by the US Department of Energy, Office of Advanced Scientific Computing Research, Field Work Proposal 20-023231.

REFERENCES

1. Bui-Thanh, T., Ghattas, O., Martin, J., and Stadler, G., A Computational Framework for Infinite-Dimensional Bayesian Inverse Problems. Part 1: The Linearized Case, with Application to Global Seismic Inversion, *SIAM J. Sci. Comput.*, **35**:2494–2523, 2013.

2. Petra, N., Martin, J., Stadler, G., and Ghattas, O., A Computational Framework for Infinite-Dimensional Bayesian Inverse Problems, Part 2: Stochastic Newton MCMC with Application to Ice Sheet Flow Inverse Problems, *SIAM J. Sci. Comput.*, **36**:1525–1555, 2014.
3. Oppenheimer, M., Glavovic, B., Hinkel, J., van de Wal, R., Magnan, A., Abd-Elgawad, A., Cai, R., Cifuentes-Jara, M., DeConto, R., Ghosh, T., Hay, J., Isla, F., Marzeion, B., Meyssignac, B., and Sebesvari, Z., Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities, in *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, Cambridge: Cambridge University Press, pp. 321–445, 2019.
4. van Pelt, W.J.J., Oerlemans, J., Reijmer, C.H., Pettersson, R., Pohjola, V.A., Isaksson, E., and Divine, D., An Iterative Inverse Method to Estimate Basal Topography and Initialize Ice Flow Models, *Cryosphere*, **7**(3):987–1006, 2013.
5. Bonan, B., Nodet, M., Ritz, C., and Peyaud, V., An ETKF Approach for Initial State and Parameter Estimation in Ice Sheet Modelling, *Nonlinear Processes Geophys.*, **21**(2):569–582, 2014.
6. Charbit, S., Dumas, C., Kageyama, M., Roche, D.M., and Ritz, C., Influence of Ablation-Related Processes in the Build-Up of Simulated Northern Hemisphere Ice Sheets during the Last Glacial Cycle, *Cryosphere*, **7**(2):681–698, 2013.
7. Quiquet, A., Punge, H.J., Ritz, C., Fettweis, X., Gallée, H., Kageyama, M., Krinner, G., Salas y Mélia, D., and Sjolte, J., Sensitivity of a Greenland Ice Sheet Model to Atmospheric Forcing Fields, *Cryosphere*, **6**(5):999–1018, 2012.
8. Crestel, B., Stadler, G., and Ghattas, O., A Comparative Study of Structural Similarity and Regularization for Joint Inverse Problems Governed by PDEs, *Inverse Probl.*, **35**(2):024003, 2018.
9. Abubakar, A., Gao, G., Habashy, T.M., and Liu, J., Joint Inversion Approaches for Geophysical Electromagnetic and Elastic Full-Waveform Data, *Inverse Probl.*, **28**(5):055016, 2012.
10. Gallardo, L.A. and Meju, M.A., Characterization of Heterogeneous Near-Surface Materials by Joint 2D Inversion of DC Resistivity and Seismic Data, *Geophys. Res. Lett.*, **30**(13):1658, 2003.
11. Blomgren, P. and Chan, T., Color TV: Total Variation Methods for Restoration of Vector-Valued Images, *IEEE Trans. Image Proc.*, **7**(3):304–309, 1998.
12. Kaipio, J.P. and Kolehmainen, V., *Approximate Marginalization over Modelling Errors and Uncertainties in Inverse Problems*, Oxford: Oxford University Press, 2013.
13. Babaniyi, O., Nicholson, R., Villa, U., and Petra, N., Inferring the Basal Sliding Coefficient Field for the Stokes Ice Sheet Model under Rheological Uncertainty, *Cryosphere*, **15**(4):1731–1750, 2021.
14. Hartland, T., Perego, M., and Petra, N., Towards Inversion of the Basal Sliding Coefficient for the Humboldt Glacier in an Uncertain Ice Sheet Model, Tech. Rep., Sandia National Laboratories, 2021.
15. Hart, J. and van Bloemen Waanders, B., Enabling Hyper-Differential Sensitivity Analysis for Ill-Posed Inverse Problems, *SIAM J. Sci. Comput.*, **45**(4):A1947–A1970, 2023.
16. Sunseri, I., Hart, J., van Bloemen Waanders, B., and Alexanderian, A., Hyper-Differential Sensitivity Analysis for Inverse Problems Constrained by Partial Differential Equations, *Inverse Probl.*, **36**(12):125001, 2020.
17. Hart, J., van Bloemen Waanders, B., and Herzog, R., Hyper-Differential Sensitivity Analysis of Uncertain Parameters in PDE-Constrained Optimization, *Int. J. Uncertainty Quantif.*, **10**(3):225–248, 2020.
18. Isaac, T., Petra, N., Stadler, G., and Ghattas, O., Scalable and Efficient Algorithms for the Propagation of Uncertainty from Data through Inference to Prediction for Large-Scale Problems, with Application to Flow of the Antarctic Ice Sheet, *J. Comput. Phys.*, **296**:348–368, 2015.
19. Stuart, A.M., Inverse Problems: A Bayesian Perspective, *Acta Numer.*, **19**:451–559, 2010.
20. Kaipio, J.P. and Somersalo, E., *Statistical and Computational Inverse Problems*, Vol. 160, New York, NY: Springer, 2005.
21. Biegler, L., Biro, G., Ghattas, O., Heinkenschloss, M., Keyes, D., Mallick, B., Marzouk, Y., Tenorio, L., van Bloemen Waanders, B., and Willcox, K., Eds., *Large-Scale Inverse Problems and Quantification of Uncertainty*, New York, NY: John Wiley and Sons, 2011.
22. Vogel, C.R., Sparse Matrix Computations Arising in Distributed Parameter Identification, *SIAM J. Matrix Anal. Appl.*, **20**(4):1027–1037, 1999.
23. Ascher, U.M. and Haber, E., Grid Refinement and Scaling for Distributed Parameter Estimation Problems, *Inverse Probl.*, **17**:571–590, 2001.
24. Haber, E. and Ascher, U.M., Preconditioned All-At-Once Methods for Large, Sparse Parameter Estimation Problems, *Inverse Probl.*, **17**(6):1847–1864, 2001.

25. Vogel, C.R., *Computational Methods for Inverse Problems*, Philadelphia: SIAM, 2002.
26. Biegler, L.T., Ghattas, O., Heinkenschloss, M., and van Bloemen Waanders, B., Eds., *Large-Scale PDE-Constrained Optimization*, Vol. 30, Berlin: Springer, 2003.
27. Biros, G. and Ghattas, O., Parallel Lagrange-Newton-Krylov-Schur Methods for PDE-Constrained Optimization. Parts I–II, *SIAM J. Sci. Comput.*, **27**:687–738, 2005.
28. Laird, C.D., Biegler, L.T., van Bloemen Waanders, B., and Bartlett, R.A., Time Dependent Contaminant Source Determination for Municipal Water Networks Using Large Scale Optimization, *J. Water Resour. Plann. Manag.*, **131**:125–134, 2005.
29. Hintermüller, M. and Vicente, L.N., Space Mapping for Optimal Control of Partial Differential Equations, *SIAM J. Optim.*, **15**(4):1002–1025, 2005.
30. Hazra, S.B. and Schulz, V., Simultaneous Pseudo-Timestepping for Aerodynamic Shape Optimization Problems with State Constraints, *SIAM J. Sci. Comput.*, **28**:1078–1099, 2006.
31. Biegler, L.T., Ghattas, O., Heinkenschloss, M., Keyes, D., and van Bloemen Waanders, B., Eds., *Real-Time PDE-Constrained Optimization*, Vol. 3, Philadelphia: SIAM, 2007.
32. Borzi, A., High-Order Discretization and Multigrid Solution of Elliptic Nonlinear Constrained Optimal Control Problems, *J. Comput. Appl. Math.*, **200**:67–85, 2007.
33. Hinze, M., Pinnau, R., Ulbrich, M., and Ulbrich, S., *Optimization with PDE Constraints*, Berlin: Springer, 2009.
34. Kouri, D.P. and Ridzal, D., *Inexact Trust-Region Methods for PDE-Constrained Optimization*, New York, NY: Springer, pp. 83–121, 2018.
35. Saibaba, A.K., Hart, J., and van Bloemen Waanders, B., Randomized Algorithms for Generalized Singular Value Decomposition with Application to Sensitivity Analysis, *Numer. Linear Algebra Appl.*, **28**(4):e2364, 2021.
36. Bui-Thanh, T., A Gentle Tutorial on Statistical Inversion Using the Bayesian Paradigm, Tech. Rep., The University of Texas at Austin, 2012.
37. Cui, T., Martin, J., Marzouk, Y.M., Solonen, A., and Spantini, A., Likelihood-Informed Dimension Reduction for Nonlinear Inverse Problems, *Inverse Probl.*, **30**(11):114015, 2014.
38. Saibaba, A.K., Lee, J., and Kitanidis, P.K., Randomized Algorithms for Generalized Hermitian Eigenvalue Problems with Application to Computing Karhunen-Loève Expansion, *Numer. Linear Algebra Appl.*, **23**(2):314–339, 2016.
39. Flath, H.P., Wilcox, L.C., Akçelik, V., Hill, J., van Bloemen Waanders, B., and Ghattas, O., Fast Algorithms for Bayesian Uncertainty Quantification in Large-Scale Linear Inverse Problems Based on Low-Rank Partial Hessian Approximations, *SIAM J. Sci. Comput.*, **33**(1):407–432, 2011.
40. Villa, U., Petra, N., and Ghattas, O., HIPPLYlib: An Extensible Software Framework for Large-Scale Inverse Problems Governed by PDEs. Part I: Deterministic Inversion and Linearized Bayesian Inference, *ACM Trans. Math. Software*, **47**(2):1–34, 2021.
41. Chow, E. and Saad, Y., Preconditioned Krylov Subspace Methods for Sampling Multivariate Gaussian Distributions, *SIAM J. Sci. Comput.*, **36**(2):588–608, 2014.
42. Baston, R.A. and Nakatsukasa, Y., Stochastic Diagonal Estimation: Probabilistic Bounds and an Improved Algorithm, *Comput. Sci. Data Struct. Algorithms*, arXiv:2201.10684, 2022.
43. Morland, L.W. and Johnson, I.R., Steady Motion of Ice Sheets, *J. Glaciol.*, **25**(92):229–246, 1980.
44. Hutter, K., *Theoretical Glaciology*, Dordrecht, the Netherlands: Springer, 1983.
45. Bueler, E., Lingle, C.S., Kallen-Brown, J.A., Covey, D.N., and Bowman, L.N., Exact Solutions and Verification of Numerical Models for Isothermal Ice Sheets, *J. Glaciol.*, **51**(173):291–306, 2005.
46. Tezaur, I.K., Perego, M., Salinger, A.G., Tuminaro, R.S., and Price, S.F., Albany/FELIX: A Parallel, Scalable and Robust, Finite Element, First-Order Stokes Approximation Ice Sheet Solver Built for Advanced Analysis, *Geosci. Model Dev.*, **8**(4):1197–1220, 2015.
47. Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E.M., Brunner, L., Knutti, R., and Hawkins, E., Partitioning Climate Projection Uncertainty with Multiple Large Ensembles and CMIP5/6, *Earth Syst. Dyn.*, **11**:491–508, 2020.