SIAM J. SCI. COMPUT.
Vol. 45, No. 1, pp. B57-B77

AN OFFLINE-ONLINE DECOMPOSITION METHOD FOR EFFICIENT LINEAR BAYESIAN GOAL-ORIENTED OPTIMAL EXPERIMENTAL DESIGN: APPLICATION TO OPTIMAL SENSOR PLACEMENT*

KEYI WU†, PENG CHEN‡, AND OMAR GHATTAS§

Abstract. Bayesian optimal experimental design (OED) plays an important role in minimizing model uncertainty with limited experimental data in a Bayesian framework. In many applications, rather than minimizing the uncertainty in the inference of model parameters, one seeks to minimize the uncertainty of a model-dependent quantity of interest (QoI). This is known as goal-oriented OED (GOOED). Here, we consider GOOED for linear Bayesian inverse problems governed by large-scale models represented by partial differential equations (PDE) that are computationally expensive to solve. In particular, we consider optimal sensor placement by maximizing an expected information gain (EIG) for the QoI. We develop an efficient method to solve such problems by deriving a new formulation of the goal-oriented EIG. Based on this formulation we propose an offline-online decomposition scheme that achieves significant computational reduction by computing all of the PDE-dependent quantities in an offline stage just once, and optimizing the sensor locations in an online stage without solving any PDEs. Moreover, in the offline stage we need only to compute low-rank approximations of two Hessian-related operators. The computational cost of these low-rank approximations, measured by the number of PDE solves, does not depend on the parameter or data dimensions for a large class of elliptic, parabolic, and sufficiently dissipative hyperbolic inverse problem that exhibit dimension-independent rapid spectra decay. We carry out detailed error analysis for the approximate goal-oriented EIG due to the low-rank approximations of the two operators. Furthermore, in the online stage we extend a swapping greedy method to optimize the sensor locations developed in our recent work that is demonstrated to be more efficient than a standard greedy method. We conduct a numerical experiment for a contaminant transport inverse problem with an infinite-dimensional parameter field to demonstrate the efficiency, accuracy, and both data- and parameter-dimension independence of the proposed algorithm.

Key words. optimal experimental design, goal oriented, Bayesian inverse problems, low-rank approximations

MSC codes. 62K05, 35Q62, 62F15, 35R30, 35Q93, 65C60, 90C27

DOI. 10.1137/21M1466542

1. Introduction.

Optimizing the acquisition of data—e.g., what, where, and when to measure, what experiments to run—to maximize information gained from the data is a fundamental and ubiquitous problem across all of the natural and

* Submitted to the journal's Computational Methods in Science and Engineering section December 20, 2021; accepted for publication (in revised form) September 9, 2022; published electronically January 24, 2023.

https://doi.org/10.1137/21M1466542

Funding: This work was supported by the National Science Foundation, Division of Mathematical Sciences under award DMS-2012453; the Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Mathematical Multifaceted Integrated Capability Centers (MMICCS) program under awards DE-SC0019303 and DE-SC0021239; the Simons Foundation under award 560651; and DOD MURI FA9550-21-1-0084.

† Department of Mathematics, The University of Texas at Austin, Austin, TX 78712 USA (keyiwu@utexas.edu).

‡ School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA (peng@cc.gatech.edu).

§ Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX 78712 USA (omar@oden.utexas.edu).

social sciences, engineering, medicine, and technology. Just three important examples include optimal observing system design for ocean climate data [50], optimal sensor placement for early warning of tsunami waves [33], and optimal experimental design (OED) to accelerate MRI imaging [11]. Bayesian OED (BOED)—including formulations such as active learning, Bayesian optimization, and sensor placement—provides a probabilistic framework to maximize the expected information gain (EIG) or mutual information for uncertain parameters or related quantities of interest [21]. However, evaluating the EIG remains prohibitive for large-scale, complex models, due to the need to compute double integrals with respect to both parameter and data distributions. Recently, advances in efficient evaluations of the EIG and design optimization have been achieved using methods based on posterior Laplace approximation for EIG estimation [49], myopic posterior sampling for adaptive goal-oriented BOED [45], EIG estimation by variational inference for BOED [35], BOED for implicit models by neural EIG estimation [47], and sequential BOED with variable cost structure [66].

Interest has intensified in extending BOED to the case of experiments on, or observations of, complex physical systems, since these can be very expensive (e.g., satellite trajectories, subsurface wells, ocean-bottom acoustic sensors). Such physical systems are typically modeled by partial differential equations (PDEs), which are expensive to solve and often contain infinite-dimensional parameter fields and large numbers of design variables. This presents fundamental challenges to conventional BOED methods, which require prohibitively large numbers of PDE solves.

Several different classes of methods have been developed to tackle these computational challenges The authors in [39, 40, 41] exploited sparsity of polynomial chaos approximations of parameter-to-observable (PtO) maps. In [1, 2, 3, 10, 29, 57], the authors explored intrinsic low dimensionality by low-rank approximation of (prior-preconditioned and data-informed) operators [1, 2, 3, 10, 29, 57]. The low-rank properties revealed by Jacobians and Hessians of the PtO map has been exploited for model reduction for sampling and deep learning [6, 12, 22, 53], Bayesian inference [15, 19, 23, 25, 27, 28], optimization under uncertainty [4, 24, 26], and BOED [2, 3, 10, 29, 57, 62]. The authors in [5, 20, 64] developed gradient methods to solve the optimization problem for sensor placement. They relax the binary nature of the sensor location variables to solve the easier continuous optimization problem with gradient-based methods, and then induce the integer solution. The authors of [51] considered sensor placement problems for signal reconstruction with the D-optimality criterion and compared the accuracy and efficiency between convex optimization and QR pivoting with a greedy method to find the optimal design. The authors of [46, 54] used a greedy algorithm to sequentially select observation locations.

In contrast with the previous work, here we focus on *goal-oriented* OED (GOOED) for linear Bayesian inverse problems, in the context of optimal sensor placement. That is, we seek optimal sensor locations that maximize the information gained from the sensors, not about the model parameters, but (often of greater interest) for a posterior model-predictive goal. In particular, we consider linear PtO maps governed by expensive PDEs with high-dimensional uncertain parameters (e.g., infinite-dimensional before discretization). In [10], a gradient-based optimization method is developed to solve the linear GOOED problem to find the optimal sensor locations. However, in each of the possibly very large number of optimization iterations, many PDE evaluations have to be performed, which makes the algorithm prohibitive if each PDE solve is very expensive.

Contributions. We propose an efficient (fast and scalable) method for high-dimensional Bayesian GOOED problems governed by expensive-to-solve PDEs. We

propose a new computational framework with an efficient offline-online decomposition to evaluate the goal-oriented EIG and solve the optimization problem. In the offline stage, all PDE solves are computed to evaluate two Hessian-related operators, while in the online stage the design optimization is performed free of PDE solves. To overcome the lack of scaling with respect to both parameter and data dimensions, we exploit the intrinsic low-dimensionality of the Hessian-related operators and compute low-rank approximations of these operators in the offline stage. Rapid spectral decay of Hessian operators is a manifestation of ill-posedness of inverse problems. It can be proven for model inverse problems governed by elliptic, parabolic, and sufficiently dissipative hyperbolic PDEs [34, 36], and demonstrated numerically for a broad spectrum of inverse problems governed by large-scale models, including for example ice sheet dynamics [42], shape and medium acoustic and electromagnetic scattering [16, 17, 18], global seismic wave propagation [19], mantle convection [61], viscous incompressible flow [63], atmospheric transport [34], ocean dynamics [44], algebraic turbulence modeling [26], poroelasticity [8], infectious disease spread [28], tumor growth modeling [59], joint inversion [30], and subsurface flow [25]. We provide a detailed error analysis for approximate goal-oriented EIG due to the low-rank approximations. By using a randomized algorithm for the low-rank approximations, we require only a small and dimension-independent number of PDE solves for typical ill-posed inverse problems. Furthermore, for the optimization of the sensor locations, we extend a swapping greedy algorithm that first constructs an initial set of sensors using leverage scores, and then swaps the chosen sensors with other candidates until certain convergence criteria are met. The swapping greedy algorithm has the attractive property that we avoid having to differentiate the goal-oriented EIG objective with respect to the design variables. Note that the offline-online decomposition is not restricted to a greedy solution of the optimization problem, but can be used in conjunction with other optimization-based approaches that encourage a binary solution (such as sparsifying penalties). Finally, we demonstrate the efficiency, accuracy, and data and parameter dimension-independence (with respect to the required number of PDE solves) of the proposed algorithm for a contaminant transport inverse problem with an infinite-dimensional parameter field.

Limitations. For our method to be fast and scalable with respect to parameter and data dimensions, we require that the PtO map be approximated with a small and dimension-independent number of PDE solves. As discussed and illustrated above, many ill-posed inverse problems exhibit rapid spectral decay of the PtO map, which motivates a low-rank approximation of the (data misfit) Hessian via matrix-free methods such as randomized SVD. In the present work, we exploit this rapid spectral decay to effect low-rank approximations of two Hessian-like operators. However, other important inverse problems, for example those governed by high Reynolds number flows, advection-dominated transport, or high frequency wave propagation, have slowly decaying Hessian spectra and, as a consequence, our proposed method as formulated here may require a large number of PDE solves. (For that matter, we are not aware of any other OED method for which this class of highly data-informed inverse problems does not present difficulties.) We do note, however, that rapid spectral decay of the PtO map is not a necessary condition for it to be approximated with a small number of PDE solves. In fact, recent and ongoing work seeks to approximate PtO maps having high global rank with a number of PDE solves far smaller than the global rank, using such representations as product convolutions [7] and hierarchical matrices [9]. In future work we aim to extend our proposed method to exploit such representations.

Another limitation of this work is that we assume uncorrelated observational noise to derive our offline-online decomposition and compute error bounds.

We present background on BOED in section 2, propose our computational framework for GOOED in section 3, and report results on experiments in section 4.

2. Background.

2.1. Linear Bayesian inverse problem.

We consider a general linear model

$$\mathbf{y} = \mathbf{F}\mathbf{m} + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\mathbf{y} \in \mathbb{R}^{d_y}$ is a d_y-dimensional observational data vector corrupted by additive Gaussian noise $\boldsymbol{\epsilon} \in \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{n}})$ with zero mean and covariance $\boldsymbol{\Gamma}_{\text{n}} \in \mathbb{R}^{d_y \times d_y}$, $\mathbf{m} \in \mathbb{R}^{d_m}$ is a d_m-dimensional uncertain parameter vector, and $\mathbf{F} : \mathbb{R}^{d_m} \mapsto \mathbb{R}^{d_y}$ is a linear PtO map. As a specific case, $\mathbf{m}$ is a discretization (e.g., by a finite element method) of an infinite-dimensional parameter field in a model described by PDEs, while $\mathbf{F}$ is implicitly given by solving the PDE model. In this case, the parameter dimension is typically very high, $O(10^6 - 10^9)$ for practical applications.

We assume a Gaussian prior $\mathbf{m} \sim \mathcal{N}(\mathbf{m}_{\text{pr}}, \boldsymbol{\Gamma}_{\text{pr}})$ with mean $\mathbf{m}_{\text{pr}}$ and covariance $\boldsymbol{\Gamma}_{\text{pr}}$ for the parameter $\mathbf{m}$ with density

$$\pi_{\text{pr}}(\mathbf{m}) \propto \exp\left(-\frac{1}{2}||\mathbf{m} - \mathbf{m}_{\text{pr}}||^2_{\boldsymbol{\Gamma}_{\text{pr}}^{-1}}\right), \tag{2.2}$$

where $||\mathbf{m}-\mathbf{m}_{\text{pr}}||^2_{\boldsymbol{\Gamma}_{\text{pr}}^{-1}} := (\mathbf{m}-\mathbf{m}_{\text{pr}})^T\boldsymbol{\Gamma}_{\text{pr}}^{-1}(\mathbf{m}-\mathbf{m}_{\text{pr}})$. Then by Bayes' rule the posterior density of $\mathbf{m}$ satisfies

$$\pi_{\text{post}}(\mathbf{m}|\mathbf{y}) \propto \pi_{\text{like}}(\mathbf{y}|\mathbf{m})\pi_{\text{pr}}(\mathbf{m}). \tag{2.3}$$

Here $\pi_{\text{like}}(\mathbf{y}|\mathbf{m})$ is the likelihood function that satisfies

$$\pi_{\text{like}}(y|\mathbf{m}) \propto \exp\left(-\Phi(\mathbf{m}, \mathbf{y})\right) \tag{2.4}$$

under Gaussian noise $\boldsymbol{\epsilon} \in \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{n}})$, where the potential function

$$\Phi(\mathbf{m}, \mathbf{y}) := \frac{1}{2}||\mathbf{F}\mathbf{m} - \mathbf{y}||^2_{\boldsymbol{\Gamma}_{\text{n}}^{-1}}. \tag{2.5}$$

Under the assumption of Gaussian prior and Gaussian noise, the posterior of $\mathbf{m}$ is also Gaussian $\mathcal{N}(\mathbf{m}_{\text{map}}, \boldsymbol{\Gamma}_{\text{post}})$ with mean $\mathbf{m}_{\text{post}} = \boldsymbol{\Gamma}_{\text{post}}(\mathbf{F}^*\boldsymbol{\Gamma}_{\text{n}}^{-1}\mathbf{y} + \boldsymbol{\Gamma}_{\text{pr}}^{-1}\mathbf{m}_{\text{pr}})$ and covariance $\boldsymbol{\Gamma}_{\text{post}} = (\mathbf{H}_m + \boldsymbol{\Gamma}_{\text{pr}}^{-1})^{-1}$, where

$$\mathbf{H}_m = \mathbf{F}^*\boldsymbol{\Gamma}_{\text{n}}^{-1}\mathbf{F} \tag{2.6}$$

is the (data-misfit) Hessian of the potential $\Phi(\mathbf{m}, \mathbf{y})$, and $\mathbf{F}^*$ is the adjoint of $\mathbf{F}$ with respect to the mass matrix-weighted inner product [60]. The action of $\mathbf{F}^*$ involves solution of an adjoint PDE.

2.2. Bayesian optimal experimental design.

2.2.1. Expected information gain.

The EIG is defined as the expected (with respect to data) Kullback–Leibler (KL) divergence between the posterior and the prior distributions,

$$\Psi := \mathbb{E}_{\mathbf{y}}[D_{\text{KL}}(\pi_{\text{post}}(\cdot|\mathbf{y})\|\pi_{\text{pr}})], \tag{2.7}$$

where the KL divergence is defined as

$$D_{\mathrm{KL}}(\pi_{\mathrm{post}}\|\pi_{\mathrm{pr}}) := \int \ln\left(\frac{d\pi_{\mathrm{post}}}{d\pi_{\mathrm{pr}}}\right) d\pi_{\mathrm{post}}. \tag{2.8}$$

For a Bayesian linear inverse problem as formulated in subsection 2.1, the EIG Ψ admits the closed form [1]

$$\Psi = \frac{1}{2}\,\mathrm{logdet}\left(\mathbf{I}_m + \widetilde{\mathbf{H}}_m\right), \tag{2.9}$$

where $\mathbf{I}_m$ is an identity matrix of size $d_m \times d_m$, and $\widetilde{\mathbf{H}}_m := \boldsymbol{\Gamma}_{\mathrm{pr}}^{\frac{1}{2}}\mathbf{H}_m\boldsymbol{\Gamma}_{\mathrm{pr}}^{\frac{1}{2}}$ is the *prior-preconditioned Hessian* that includes both data and prior information.

2.2.2. BOED for sensor placement. We consider an optimal sensor placement problem. Assume we have a collection of d candidate sensors $\{s_i\}_{i=1}^d$. We need to choose a much smaller number $r < d$ of sensors (due to a limited budget or physical constraints) at which data are collected. The OED problem seeks to find the best sensor combination from the candidates. We use a Boolean design matrix $\mathbf{W} \in \mathcal{W} \subset \mathbb{R}^{r\times d}$ to represent sensor placement such that $\mathbf{W}_{ij} = 1$ if the ith sensor is placed at the jth candidate location, i.e.,

$$\mathbf{W}_{ij} \in \{0,1\},\ \sum_{j=1}^{d}\mathbf{W}_{ij} = 1,\ \sum_{i=1}^{r}\mathbf{W}_{ij} \in \{0,1\}. \tag{2.10}$$

We assume that the observational noise for the d candidate sensors is uncorrelated, with covariance

$$\boldsymbol{\Gamma}_{\mathrm{n}}^{d} = \mathrm{diag}(\sigma_1^2,\ldots,\sigma_d^2). \tag{2.11}$$

As a result, for any design $\mathbf{W}$ with the covariance for the observation noise $\boldsymbol{\epsilon}$ as $\boldsymbol{\Gamma}_{\mathrm{n}}(\mathbf{W}) = \mathbf{W}\boldsymbol{\Gamma}_{\mathrm{n}}^{d}\mathbf{W}^T$, we have

$$\boldsymbol{\Gamma}_{\mathrm{n}}^{-1}(\mathbf{W}) = \mathbf{W}(\boldsymbol{\Gamma}_{\mathrm{n}}^{d})^{-1}\mathbf{W}^T. \tag{2.12}$$

Denoting by $\mathbf{F}_d$ the PtO map using all d candidate sensors, we have the design-specific PtO map

$$\mathbf{F}(\mathbf{W}) = \mathbf{W}\mathbf{F}_d \tag{2.13}$$

with its adjoint $\mathbf{F}^* = \mathbf{F}_d^*\mathbf{W}^T$. We can now state the OED problem as find an optimal design $\hat{\mathbf{W}} \in \mathcal{W}$ such that

$$\hat{\mathbf{W}} = \underset{\mathbf{W}\in\mathcal{W}}{\mathrm{argmax}}\,\Psi(\mathbf{W}). \tag{2.14}$$

3. Goal-oriented optimal experimental design. The classical OED problem seeks a design that maximizes the information gain for the parameter vector $\mathbf{m}$. In this work, we consider a GOOED problem that maximizes the information gain of a predicted *quantity of interest* (QoI) $\boldsymbol{\rho} \in \mathbb{R}^p$, which is assumed to be a linear function of the parameter $\mathbf{m}$,

$$\boldsymbol{\rho} = \mathbf{P}\mathbf{m}, \tag{3.1}$$

where $\mathbf{P} : \mathbb{R}^{d_m} \mapsto \mathbb{R}^{d_\rho}$ is a linear map that typically involves solution of the governing PDEs. Due to linearity, the prior distribution of $\boldsymbol{\rho}$ is Gaussian $\mathcal{N}(\boldsymbol{\rho}_{\mathrm{pr}}, \boldsymbol{\Sigma}_{\mathrm{pr}})$ with mean $\boldsymbol{\rho}_{\mathrm{pr}} = \mathbf{P}\mathbf{m}_{\mathrm{pr}}$ and covariance $\boldsymbol{\Sigma}_{\mathrm{pr}} = \mathbf{P}\boldsymbol{\Gamma}_{\mathrm{pr}}\mathbf{P}^*$, where $\mathbf{P}^*$ is the adjoint of $\mathbf{P}$ with respect to the mass matrix-weighted inner product [60]. Moreover, the posterior distribution of $\boldsymbol{\rho}$ is also Gaussian $\mathcal{N}(\boldsymbol{\rho}_{\mathrm{post}}, \boldsymbol{\Sigma}_{\mathrm{post}})$ with mean $\boldsymbol{\rho}_{\mathrm{post}} = \mathbf{P}\mathbf{m}_{\mathrm{post}}$ and covariance $\boldsymbol{\Sigma}_{\mathrm{post}} = \mathbf{P}\boldsymbol{\Gamma}_{\mathrm{post}}\mathbf{P}^*$.

3.1. EIG for GOOED. To construct an expression for EIG for GOOED, we first introduce Proposition 3.1 [58, Lemma 2.2], which relates the observational data $\mathbf{y}$ and the QoI $\boldsymbol{\rho}$. See proof in Appendix A.

PROPOSITION 3.1. *Model* (2.1) *and QoI* (3.1) *lead to*

$$\mathbf{y} = \mathbf{F}\mathbf{P}_\dagger \boldsymbol{\rho} + \boldsymbol{\eta}, \tag{3.2}$$

where $\mathbf{P}_\dagger := \boldsymbol{\Gamma}_{\text{pr}}\mathbf{P}^*\boldsymbol{\Sigma}_{\text{pr}}^{-1}$ *and* $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_\eta)$ *with*

$$\boldsymbol{\Gamma}_\eta := \boldsymbol{\Gamma}_{\text{n}} + \mathbf{F}(\boldsymbol{\Gamma}_{\text{pr}} - \boldsymbol{\Gamma}_{\text{pr}}\mathbf{P}^*\boldsymbol{\Sigma}_{\text{pr}}^{-1}\mathbf{P}\boldsymbol{\Gamma}_{\text{pr}})\mathbf{F}^* \tag{3.3}$$

or, equivalently, $\boldsymbol{\Gamma}_\eta = \mathbb{C}ov[\boldsymbol{\epsilon}] + \mathbb{C}ov[\mathbf{F}(\mathbf{I}_m - \mathbf{P}_\dagger\mathbf{P})\mathbf{m}]$ *with* $\mathbb{C}ov$ *as covariance. Moreover,* $\boldsymbol{\rho}$ *and* $\boldsymbol{\eta}$ *are independent.*

Thus, the EIG for $\boldsymbol{\rho}$ can be obtained analogously to (2.9),

$$\Psi^\rho(\mathbf{W}) = \frac{1}{2}\,\text{logdet}\left(\mathbf{I}_\rho + \widetilde{\mathbf{H}}_m^\rho(\mathbf{W})\right), \tag{3.4}$$

where $\mathbf{I}_\rho$ is an identity matrix of size $d_\rho \times d_\rho$, and $\widetilde{\mathbf{H}}_m^\rho(\mathbf{W}) = \boldsymbol{\Sigma}_{\text{pr}}^{\frac{1}{2}}\mathbf{H}_m^\rho(\mathbf{W})\boldsymbol{\Sigma}_{\text{pr}}^{\frac{1}{2}}$ with $\mathbf{H}_m^\rho(\mathbf{W})$ given by

$$\mathbf{H}_m^\rho(\mathbf{W}) = (\mathbf{F}(\mathbf{W})\mathbf{P}_\dagger)^*\boldsymbol{\Gamma}_\eta^{-1}(\mathbf{W})\mathbf{F}(\mathbf{W})\mathbf{P}_\dagger. \tag{3.5}$$

3.2. Offline-online decomposition for EIG Ψ^ρ. The EIG $\Psi^\rho(\mathbf{W})$ depends on $\mathbf{W}$ through $\mathbf{F}(\mathbf{W}) = \mathbf{W}\mathbf{F}_d$, which involves solution of the governing PDEs. Since $\Psi^\rho(\mathbf{W})$ must be evaluated repeatedly in the course of maximizing EIG, these repeated PDE solves would be prohibitive. To circumvent this problem, we propose an *offline-online decomposition* scheme, where the PDE-governed computation of quantities that are independent of $\mathbf{W}$ is performed offline just once, and the online experimental design optimization is free of any PDE solves. The key result permitting this decomposition is given in the following theorem with proof in Appendix A.

THEOREM 3.2. *For each design* $\mathbf{W} \in \mathcal{W}$*, the goal-oriented EIG* $\Psi^\rho(\mathbf{W})$ *given in* (3.4) *can be computed as*

$$\Psi^\rho(\mathbf{W}) = \frac{1}{2}\text{logdet}\left(\mathbf{I}_r + \mathbf{L}^T\mathbf{W}\mathbf{H}_d^\rho\mathbf{W}^T\mathbf{L}\right), \tag{3.6}$$

where $\mathbf{I}_r$ *is an identity matrix of size* $r \times r$*,* $\mathbf{H}_d^\rho$ *is given by*

$$\mathbf{H}_d^\rho := \mathbf{F}_d\boldsymbol{\Gamma}_{\text{pr}}\mathbf{P}^*\boldsymbol{\Sigma}_{\text{pr}}^{-1}\mathbf{P}\boldsymbol{\Gamma}_{\text{pr}}\mathbf{F}_d^*, \tag{3.7}$$

and $\mathbf{L}$ *is given by the Cholesky factorization* $\boldsymbol{\Gamma}_\eta^{-1} = (\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T)^{-1} = \mathbf{L}\mathbf{L}^T$*.* $\Delta\mathbf{H}_d := \mathbf{H}_d - \mathbf{H}_d^\rho$ *with* $\mathbf{H}_d = \mathbf{F}_d\boldsymbol{\Gamma}_{\text{pr}}\mathbf{F}_d^*$*.*

Note that by Theorem 3.2, we can separate the design matrix $\mathbf{W}$ and PDE-governed operators ($\mathbf{H}_d^\rho$ and $\Delta\mathbf{H}_d$, independent of $\mathbf{W}$) in $\Psi^\rho(\mathbf{W})$. Hence evaluation of $\Psi^\rho(\mathbf{W})$ can be decomposed as follows: (1) construct the PDE-governed matrices $\mathbf{H}_d^\rho$ and $\Delta\mathbf{H}_d$ offline just once; and (2) for each $\mathbf{W}$ in the online optimization process, assemble a small $(r \times r)$ matrix $\boldsymbol{\Gamma}_\eta(\mathbf{W})$ by (A.6), compute a Cholesky factorization $\boldsymbol{\Gamma}_\eta^{-1} = \mathbf{L}\mathbf{L}^T$, and assemble $\Psi^\rho(\mathbf{W})$ by (3.6), which are all free of the expensive PDE solves.

Note that $\Delta\mathbf{H}_d \in \mathbb{R}^{d\times d}$ and $\mathbf{H}_d^\rho \in \mathbb{R}^{d\times d}$ are large matrices when we have a large number of candidate sensors $d \gg 1$. Moreover, their explicit construction involves

PDE solves, which may become prohibitive when the parameters are high dimensional, $d_m \gg 1$. Therefore, it is computationally impractical to directly compute and store these matrices. Fortunately, the intrinsic ill-posedness of many high-dimensional inverse problems—data inform only a low-dimensional subspace of parameter space, e.g., [15, 34, 42, 55]—suggests that these matrices are likely of low rank or exhibit rapid spectral decay. We exploit this property and construct low-rank approximations of $\mathbf{H}_d^\rho$ and $\Delta\mathbf{H}_d$ in subsection 3.3.

3.3. Low-rank approximation. $\mathbf{H}_d^\rho$ and $\Delta\mathbf{H}_d$ are given in (3.7) and (A.6) and integrate data, parameter, and QoI information. Noting that $\mathbf{H}_d^\rho$ and $\Delta\mathbf{H}_d$ are both symmetric, we compute their low-rank approximation for given tolerances $\epsilon_\zeta, \epsilon_\lambda > 0$ as

$$\hat{\mathbf{H}}_d^\rho = \mathbf{U}_k \mathbf{Z}_k \mathbf{U}_k^T \quad \text{and} \quad \Delta\hat{\mathbf{H}}_d = \mathbf{V}_l \mathbf{\Lambda}_l \mathbf{V}_l^T, \tag{3.8}$$

where $(\mathbf{U}_k, \mathbf{Z}_k)$ represent the k dominant eigenpairs of $\mathbf{H}_d^\rho$ with $\mathbf{Z}_k = \mathrm{diag}(\zeta_1, \dots, \zeta_k)$ such that

$$\zeta_1 \geq \zeta_2 \geq \cdots \geq \zeta_k \geq \epsilon_\zeta \geq \zeta_{k+1} \cdots \geq \zeta_d \tag{3.9}$$

and (V_l, Λ_l) represent the l dominant eigenpairs of $\Delta\mathbf{H}_d$ with $\mathbf{\Lambda}_l = \mathrm{diag}(\lambda_1, \dots, \lambda_l)$ such that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l \geq \epsilon_\lambda \geq \lambda_{l+1} \geq \cdots \geq \lambda_d. \tag{3.10}$$

With $\hat{\mathbf{\Gamma}}_\eta(\mathbf{W}) := \mathbf{W}(\mathbf{\Gamma}_{\mathrm{n}}^d + \Delta\hat{\mathbf{H}}_d)\mathbf{W}^T$ as an approximation of $\mathbf{\Gamma}_\eta(\mathbf{W})$ in (3.3), we compute the Cholesky factorization $\hat{\mathbf{\Gamma}}_\eta^{-1} = \hat{\mathbf{L}}\hat{\mathbf{L}}^T$. Then we can define an approximate EIG as

$$\hat{\Psi}^\rho(\mathbf{W}) := \frac{1}{2} \mathrm{logdet}\left(\mathbf{I}_r + \hat{\mathbf{L}}^T \mathbf{W} \hat{\mathbf{H}}_d^\rho \mathbf{W}^T \hat{\mathbf{L}}\right). \tag{3.11}$$

The following theorem quantifies the approximation error. See proof in Appendix A

THEOREM 3.3. *For any design* $\mathbf{W} \in \mathcal{W}$*, the error for the goal-oriented EIG* $\Psi^\rho(\mathbf{W})$ *in* (3.6) *by its approximation* $\hat{\Psi}^\rho(\mathbf{W})$ *in* (3.11) *can be bounded by*

$$|\Psi^\rho(\mathbf{W}) - \hat{\Psi}^\rho(\mathbf{W})| \leq \frac{1}{2} \sum_{i=k+1}^{d} \log(1 + \zeta_i/\sigma_{\min}^2) + \frac{1}{2} \sum_{i=l+1}^{k} \log(1 + \lambda_i \zeta_1/\sigma_{\min}^4), \tag{3.12}$$

where $\sigma_{\min}^2 := \min(\sigma_1^2, \dots, \sigma_d^2)$ *as defined in* (2.11).

We remark that with the rapid decay of the eigenvalues $(\zeta_k)_{k\geq 1}$ of $\hat{\mathbf{H}}_d^\rho$ and $(\lambda_l)_{l\geq 1}$ of $\Delta\hat{\mathbf{H}}_d$, the error bound in (3.12) becomes very small. Moreover, the decay rates are often independent of the (candidate) data dimension d and the parameter dimension d_m, as demonstrated in subsection 4.3. We employ a randomized SVD algorithm [37], which requires only $O(k)$ and $O(l)$ PDE solves, respectively. In practice, $k, l \ll d$. More details on the algorithm applied to the example problem in section 4 can be found in Appendix B. This means that an arbitrarily accurate EIG approximation can be constructed with a small number, $O(k + l)$ of PDE solves.

3.4. Swapping greedy optimization. Once the low-rank approximations of $\mathbf{H}_d^\rho$ and $\Delta\mathbf{H}_d$ are constructed per (3.8), we obtain a fast method for evaluating the approximate EIG in (3.11), with a certified approximation error given by Theorem 3.3.

We emphasize that this fast computation does not involve any PDE solves as the designs $\mathbf{W}$ change. We now turn to the (combinatorial) optimization problem of finding the optimal design matrix $\hat{\mathbf{W}}$,

$$\hat{\mathbf{W}} = \arg\max_{\mathbf{W}\in\mathcal{W}} \hat{\Psi}^{\rho}(\mathbf{W}). \tag{3.13}$$

We next introduce a swapping greedy algorithm to solve this problem requiring only evaluation of $\hat{\Psi}^{\rho}(\mathbf{W})$. Maximizing the EIG for linear OED is equivalent to minimizing a D-optimality criterion [10]. This combinatorial optimization problem is NP-hard. Fortunately, a simple greedy algorithm can provide a quasi-optimal solution under proper assumptions (submodularity [48]) with some theoretical guarantees [52]. The standard greedy method sequentially finds the optimal sensors one by one (or batch by batch) [13, 43]. In contrast to the standard greedy algorithm, we extend a swapping greedy algorithm developed for BOED in [62] to solve the GOOED problem. The swapping greedy algorithm is a combination of Fedorov's exchange algorithm [32] and uses leverage scores for a more informative initial guess to accelerate the optimization convergence. Given a current sensor set, it swaps sensors with the remaining sensors to maximize the approximate EIG $\hat{\Psi}^{\rho}(\mathbf{W})$ until convergence. To initialize the chosen sensor set, we take advantage of the low-rank approximation $\hat{\mathbf{H}}_d^{\rho}$ in (3.8), which contains information from the data (through $\mathbf{F}_d$), parameter (through $\mathbf{\Gamma}_{\text{pr}}$), and QoI (through $\mathbf{P}$), as can be seen from (3.7). In particular, the most informative sensors can be revealed by the rows of $\mathbf{U}_k$ with the largest norms, or the leverage scores of $\mathbf{H}_d^{\rho}$ [14]. More specifically, given a budget for selecting r sensors from d candidate locations, we initialize the candidate set $S^0 = \{s_1, \ldots, s_r\}$ such that s_i, $i = 1, \ldots, r$, is the row index corresponding to the ith largest row norm of U_k, i.e.,

$$s_i = \arg\max_{s\in S\setminus S^{i-1}} ||\mathbf{U}_k(s,:)||_2, \quad i = 1, \ldots, r, \tag{3.14}$$

where $\mathbf{U}_k(s,:)$ is the sth row of $\mathbf{U}_k$, $||\cdot||_2$ is the Euclidean norm, and the set $S^{i-1} = \{s_1, \ldots, s_{i-1}\}$ for $i = 2, 3, \ldots$, and $S^0 = \emptyset$. Then, at each step of a loop for $t = 1, \ldots, r$, we swap a sensor s_t from the current chosen sensor set S^{t-1} with one from the candidate set such that the approximate EIG $\hat{\Psi}^{\rho}(\mathbf{W})$ evaluated as in (3.11) can be maximized, i.e., we choose s^* such that

$$s^* = \arg\max_{s\in\{s_t\}\cup(S\setminus S^{t-1})} \hat{\Psi}^{\rho}(\mathbf{W}_s), \tag{3.15}$$

where $\mathbf{W}_s$ is the design matrix corresponding to the sensor choice $S^{t-1}\setminus\{s_t\}\cup\{s\}$. We repeat the loop until a convergence criterion is met, e.g., the chosen S does not change or the difference of the approximate EIG is smaller than a given tolerance ϵ_g. We summarize the swapping algorithm in Algorithm 3.1.

4. Experiments. In this section, we present the results of numerical experiments for GOOED governed by a linear time-dependent PDE model with an infinite-dimensional parameter field and varying numbers of candidate sensors. This problem features the key challenges of (1) an expensive PtO map in the form of a time-dependent PDE solution and (2) high-dimensional parameters and data.

4.1. Model settings. We consider sensor placement for Bayesian inversion of a contaminant source with the goal of maximizing information gain for contaminant concentration on some building surfaces. The transport of the contaminant can

Algorithm 3.1 A swapping greedy algorithm for GOOED

Input: low-rank approximations (3.8), a set $S = \{1, \dots, d\}$ of d candidate sensors, a budget of r sensors to be placed.
Output: the optimal sensor set S^* with r sensors.
Initialize $S^* = \{s_1, \dots, s_r\} \subset S$ according to (3.14).
Set $S^0 = \{\emptyset\}$.
while $S^* \neq S^0$ and $\hat{\Psi}^\rho(\mathbf{W}(S^*)) - \hat{\Psi}^\rho(\mathbf{W}(S^0)) < \epsilon_g$ **do**
  $S^0 \leftarrow S^*$.
  for $t = 1, \dots, r$ **do**
    Choose s^* according to (3.15).
    Update $S^t \leftarrow (S^{t-1} \setminus \{s_t\}) \cup \{s^*\}$.
  end for
  Update $S^* \leftarrow S^r$.
end while
Output: optimal sensor choice S^*.

be modeled by the time-dependent advection-diffusion equation with homogeneous Neumann boundary condition,

$$\begin{aligned} u_t - k\Delta u + \boldsymbol{v} \cdot \nabla u = 0 \text{ in } \mathcal{D} \times (0,T), \\ u(\cdot, 0) = m \text{ in } \mathcal{D}, \\ k\nabla u \cdot n = 0 \text{ on } \partial\mathcal{D} \times (0,T), \end{aligned} \tag{4.1}$$

where $k = 0.001$ is the diffusion coefficient and $T > 0$ is the final time. The domain $\mathcal{D} \subset \mathbb{R}^2$ is open and bounded with boundary $\partial\mathcal{D}$ depicted in Figure 1. The initial condition m is an infinite-dimensional random parameter field in $\mathcal{D}$, which is to be inferred. The velocity field $\boldsymbol{v} \in \mathbb{R}^2$ is obtained as the solution of the steady-state Navier-Stokes equations with Dirichlet boundary condition,

$$\begin{aligned} -\frac{1}{\mathrm{Re}}\Delta \boldsymbol{v} + \nabla q + \boldsymbol{v} \cdot \nabla \boldsymbol{v} = 0 \text{ in } \mathcal{D}, \\ \nabla \cdot \boldsymbol{v} = 0 \text{ in } \mathcal{D}, \\ \boldsymbol{v} = \boldsymbol{g} \text{ on } \partial\mathcal{D}, \end{aligned} \tag{4.2}$$

where q represents the pressure field and the Reynolds number $\mathrm{Re} = 50$. The Dirichlet boundary data $\boldsymbol{g} \in \mathbb{R}^2$ are prescribed as $\boldsymbol{g} = (0, 1)$ on the left wall of the domain, $\boldsymbol{g} = (0, -1)$ on the right wall, and $\boldsymbol{g} = (0, 0)$ elsewhere. We consider a Gaussian prior for the parameter $m \sim \mathcal{N}(m_{\mathrm{pr}}, \mathcal{C}_{\mathrm{pr}})$ with mean m_{pr} and covariance operator $\mathcal{C}_{\mathrm{pr}} = \mathcal{A}^{-2}$, where the elliptic operator $\mathcal{A} = -\gamma\Delta + \delta I$ (with Laplacian Δ and identity I) is equipped with Robin boundary condition $\gamma\nabla m \cdot \mathbf{n} + \beta m$ on $\partial\mathcal{D}$. Here $\gamma, \delta > 0$ control the correlation length and variance of m [31]. In our numerical test, we set $m_{\mathrm{pr}} = 0.25$, $\gamma = 1, \delta = 8$. We synthesize a "true" initial condition $m_{\mathrm{true}} = \min(0.5, \exp(-100\|x - [0.35, 0.7]\|^2)$ as the contaminant source (Figure 1(b)). To solve the PDE model, we use an implicit Euler method for temporal discretization with N_t time steps, and a finite element method for spatial discretization, resulting in a d_m-dimensional discrete parameter $\mathbf{m} \sim \mathcal{N}(\mathbf{m}_{\mathrm{pr}}, \mathbf{\Gamma}_{\mathrm{pr}})$ with $\mathbf{m}_{\mathrm{pr}}, \mathbf{\Gamma}_{\mathrm{pr}}$ denoting finite element discretizations of $m_{\mathrm{pr}}, \mathcal{C}_{\mathrm{pr}}$, respectively.

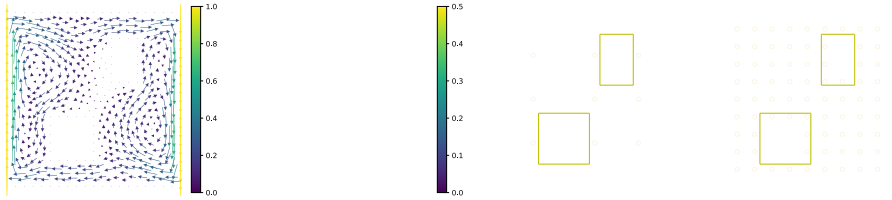
The solution of the PDE for $d_m = 2023$ and $N_t = 40$ at the observation time $T = 0.8$ and d candidate sensor locations are also shown in Figures 1(c) and 1(d), at

which we observe the contaminant concentration u. The linear map $\mathbf{F}$ is defined by the predicted data, i.e., the concentrations at the selected sensors. Finally, we take the QoI as an averaged contaminant concentration at time t_{pred} within a distance $\delta = 0.02$ from the boundaries of either the left, the right, or both buildings, with corresponding QoI maps denoted as $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ at certain prediction times (see Figures 1(c) and 1(d)).

4.2. Numerical results. We first consider the case of a small number of candidate sensors, for which we can use an exhaustive search to find the optimal sensor combination and compare it with the sensors chosen by the standard and swapping greedy algorithms. Specifically, we use a grid of $d = 9$ candidate locations $\{s_i\}_{i=0}^{9}$ ($x_i \in \{0.2, 0.55, 0.8\} \times \{0.25, 0.5, 0.75\}$) as shown in Figure 1(c) with the goal of choosing $r = 2, 3, 4, 5, 6, 7, 8$ sensors for the QoI prediction time $t_{\text{pred}} = 1.0$. We compute the matrices $\mathbf{H}_d^\rho$ and $\Delta\mathbf{H}_d$ (of size 9×9) without low-rank approximation since they are small.

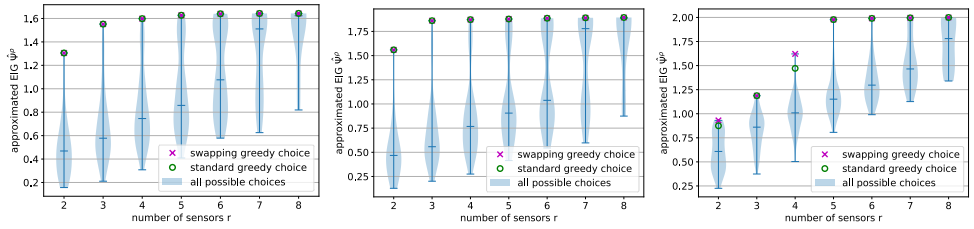
We can see from Figure 2 that for QoI maps $\mathbf{P}_1$ and $\mathbf{P}_2$, both greedy algorithms find the optimal design, while for $\mathbf{P}_3$ with $r = 2, 4$, only swapping greedy finds the optimal design. Moreover, an increase in r leads to diminishing returns, as the gain in information about the QoI from additional sensors saturates. We see that $\sim$3 sensors is sufficient for either building, whereas 5 is sufficient for both.

Next we consider the case of the 75 candidate sensors depicted in Figure 1(d). The total number of possible sensor combinations is $\frac{d!}{r!(d-r)!}$. An exhaustive search across all sensor combinations is not feasible in this case; instead, we compare the best EIG from 200 random designs with those obtained by the greedy algorithms.



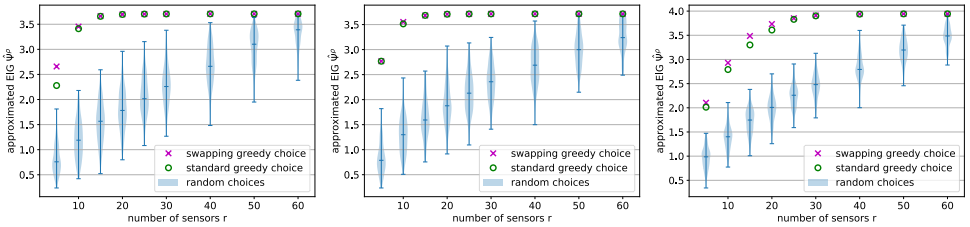
(a) velocity field $\boldsymbol{v}$ (b) initial condition m_{true} (c) 9 candidates (d) 75 candidates

FIG. 1. *The domain $\mathcal{D}$ is $[0,1]^2$ with two rectangular blocks ($[0.25, 0.5] \times [0.15, 0.4], [0.6, 0.75] \times [0.6, 0.85]$) removed. Data of contaminant concentration at time $T = 0.8$, obtained as the solution of (4.1) at the initial condition as shown. The QoI maps ($\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$) are the averaged solution within the lines along the left, right, and both buildings. Candidate sensor locations are shown in circles.*



(a) $\mathbf{P}_1$ (b) $\mathbf{P}_2$ (c) $\mathbf{P}_3$

FIG. 2. *Approximate EIG $\hat{\Psi}^\rho$ at r sensors chosen by the standard and swapping greedy algorithms, and the distribution of $\hat{\Psi}^\rho$ at all possible combinations of 9 candidate sensors. The three plots are for the QoI maps $\mathbf{P}_1, \mathbf{P}_2$, and $\mathbf{P}_3$.*



(a) $\mathbf{P}_1$ (b) $\mathbf{P}_2$ (c) $\mathbf{P}_3$

FIG. 3. *Approximate EIG $\hat{\Psi}^\rho$ for r out of 75 sensors, found by the standard and swapping greedy algorithms, compared with the distribution of $\hat{\Psi}^\rho$ for 200 randomly-chosen sets from the 75. The three plots are for the QoI maps $\mathbf{P}_1, \mathbf{P}_2$, and $\mathbf{P}_3$.*

TABLE 1

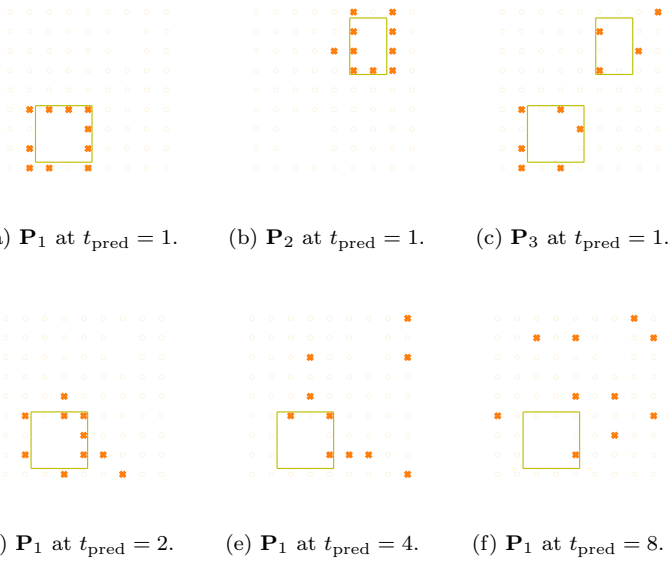
Number of swapping loops (#LOOPS), swaps (# SWAPS), and EIG evaluations (# EIG EVAL) for different numbers of r selected sensors out of 75 candidates. Results are reported for Algorithm 3.1 for the goal $\mathbf{P}_1$.

r	5	10	15	20	25	30	40	50	60
#loops	3	3	3	3	3	2	3	3	3
#swaps	41	73	124	164	190	194	235	199	119
#EIG eval	1050	1950	2700	3300	3750	2700	4200	3750	2700

We seek the r optimal sensors, $r = 5, 10, 15, 20, 25, 30, 40, 50, 60$, from among the 75 candidates. Results are shown in Figure 3. We see that both greedy algorithms find designs with larger EIG than all random choices. Moreover, for small r, the swapping greedy algorithm finds better designs than the standard greedy. For large r, both greedy algorithms can find designs with similar EIG.

To demonstrate the reduction of computational cost achieved by the offline-online decomposition, we report the total number of EIG evaluations, the number of swapping loops, and the number of swaps of the swapping greedy algorithm (Algorithm 3.1) in Table 1 for 75 candidate sensors with different target numbers of sensors. We see that the number of loops at convergence is mostly 3, which does not change with respect to the number of selected sensors. We observe in the experiments that most of the swaps take place in the first loop, followed by a smaller number of swaps in the second loop resulting in slight sensor adjustments. There are no swaps in the last loop, which we require as a convergence criterion. As a result of the offline-online decomposition Theorem 3.2, which relieves the (thousands of) EIG evaluations of expensive PDE solves once the low-rank approximation (3.8) is built, we achieve over 1000X speedup. This is because the PDE solves overwhelmingly dominate the overall cost, and because the offline decomposition is computed at a cost comparable to one direct EIG evaluation by (3.4).

Figure 4 illustrates the effect of the goal of maximizing information gain for the QoIs from optimally placed sensors. Specifically, for the parameter-to-QoI maps $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ that quantify the average contaminant concentration at time $t_{\text{pred}} = 1$ around the left, right, and both blocks, the GOOED finds the sensors depicted in the first row. For $\mathbf{P}_1$ at longer prediction times $t_{\text{pred}} = 1, 2, 4, 8$, we see in the bottom row of Figure 4 that the optimal sensors are no longer placed in the immediate vicinity of the building, but instead are increasingly dispersed to better detect the now more diffused field. Finally, the ability of GOOED to reduce the posterior variance in the initial condition field is depicted in Figure 5 for different goals $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$. Compared



(a) $\mathbf{P}_1$ at $t_{\text{pred}} = 1$. (b) $\mathbf{P}_2$ at $t_{\text{pred}} = 1$. (c) $\mathbf{P}_3$ at $t_{\text{pred}} = 1$.

(d) $\mathbf{P}_1$ at $t_{\text{pred}} = 2$. (e) $\mathbf{P}_1$ at $t_{\text{pred}} = 4$. (f) $\mathbf{P}_1$ at $t_{\text{pred}} = 8$.

FIG. 4. *Sensor locations chosen by the swapping greedy algorithm for* 10 *out of* 75 *candidates for the parameter-to-QoI maps* $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ *at time* $t_{\text{pred}} = 1$ *and also* $\mathbf{P}_1$ *at time* $t_{\text{pred}} = 2, 4, 8$.

(a) optimal design for $\mathbf{P}_1$ (b) optimal design for $\mathbf{P}_2$ (c) optimal design for $\mathbf{P}_3$ (d) random design

FIG. 5. *Pointwise posterior variance of the parameter at optimal designs for goals* $\mathbf{P}_1$,$\mathbf{P}_2$, $\mathbf{P}_3$, *compared to a random design, for* 10 *sensors. The darker regions represent lower variance.*

to a random design (lower right), the three optimal designs lead to lower variance surrounding regions of interest.

4.3. Scalability with respect to parameter and data dimensions. Here we demonstrate the fast decay of the eigenvalues of $\mathbf{H}_d^\rho$ and $\Delta\mathbf{H}_d$ with respect to the parameter and data dimensions, as exploited by the algorithms of subsection 3.3. For $\mathbf{H}_d^\rho$ defined in (3.7), we have $\text{rank}(\mathbf{H}_d^\rho) \leq \min(p, d)$ with QoI dimension p and data dimension d. In practice, the QoI is often an averaged quantity with small p, so the rank of $\mathbf{H}_d^\rho$ is also small. In our tests we have $\text{rank}(\mathbf{H}_d^\rho) = p = 1$. For $\Delta\mathbf{H}_d = \mathbf{H}_d - \mathbf{H}_d^\rho$ with $\mathbf{H}_d = \mathbf{F}_d\boldsymbol{\Gamma}_{\text{pr}}\mathbf{F}^*$, the spectrum of $\Delta\mathbf{H}_d$ depends on that of $\mathbf{H}_d$, which typically exhibits fast decay due to ill-posedness of inverse problems. As can be observed in the left plot of Figure 6, the eigenvalues of $\Delta\mathbf{H}_d$ decay very rapidly and independently of the parameter dimension, which implies that the required number of PDE solves is small and independent of the parameter dimension while achieving the same absolute accuracy of the approximate EIG by Theorem 3.3. The right plot

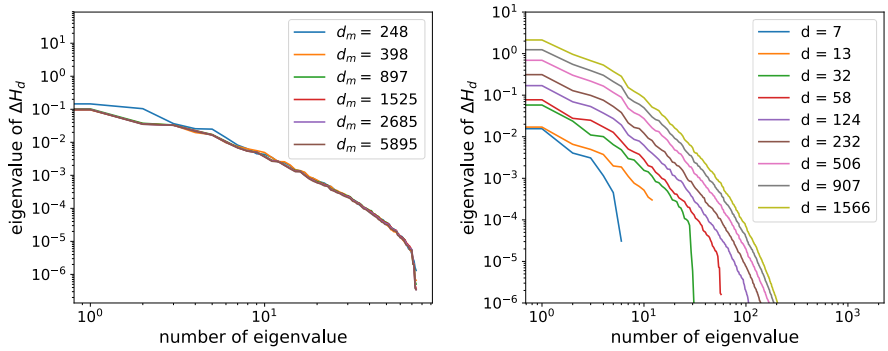


FIG. 6. *Decay of the eigenvalues of* $\Delta\mathbf{H}_d$ *with the increasing parameter dimension (left) and data (candidate sensor locations) dimension (right).*

in Figure 6 also illustrates rapid decay of eigenvalues, with the increasing number of candidate sensors, suggesting that the number of PDE solves is asymptotically independent of the data dimension for the same relative accuracy of the approximate EIG. These plots suggest that $O(100)$ PDE solves are required to accurately capture the information gained about the parameter field and QoI from the data, regardless of the parameter or sensor dimensions, when using randomized SVD (Algorithm B.1).

5. Conclusions. We have developed a fast and scalable computational framework for goal-oriented linear BOED governed by PDEs (or more generally expensive models). Repeated fast evaluation of an (arbitrarily accurate) approximate EIG free of PDE solves as the experimental design changes is made possible by an offline-online decomposition and low-rank approximation of certain operators informed by the parameter, data, and predictive goals of interest. Scalability—as measured by independence of the number of PDE solves from the parameter and data dimensions—is achieved by carefully exploiting the GOOED problem's intrinsic low dimensionality as manifested by the rapid spectral decay of several critical operators. To justify the low-rank approximation of these operators in computing the EIG, we proved an upper bound for the approximation error in terms of the operators' truncated eigenvalues. Moreover, we proposed a new swapping greedy algorithm that is demonstrated to be more effective than the standard greedy algorithm in our experiments. Numerical experiments with optimal sensor placement for Bayesian inference of the initial condition of an advection-diffusion PDE demonstrated over 1000X speedups (measured in PDE solves). Future work includes extension to nonlinear Bayesian GOOED problems with nonlinear PtO maps and nonlinear parameter-to-QoI maps.

Appendix A. Proofs of the main results.

Proof of Proposition 3.1

$$\mathbf{y} = \mathbf{F}\mathbf{m} + \boldsymbol{\epsilon} = \mathbf{F}\mathbf{P}_\dagger\mathbf{P}\mathbf{m} + \mathbf{F}(\mathbf{I} - \mathbf{P}_\dagger\mathbf{P})\mathbf{m} + \boldsymbol{\epsilon} = \mathbf{F}\mathbf{P}_\dagger\boldsymbol{\rho} + \boldsymbol{\eta}, \tag{A.1}$$

where $\mathbf{P}_\dagger := \boldsymbol{\Gamma}_{\text{pr}}\mathbf{P}^*\boldsymbol{\Sigma}_{\text{pr}}^{-1}$ and $\boldsymbol{\eta} := \mathbf{F}(\mathbf{I} - \mathbf{P}_\dagger\mathbf{P})\mathbf{m} + \boldsymbol{\epsilon}$.

Proof of Theorem 3.2. To start with, we introduce the Weinstein–Aronszajn identity in Proposition A.1 which is proven in [56].

PROPOSITION A.1. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be matrices of size* $m \times n$ *and* $n \times m$*, respectively, then*

$$\det(\mathbf{I}_{n\times n} + \mathbf{B}\mathbf{A}) = \det(\mathbf{I}_{m\times m} + \mathbf{A}\mathbf{B}). \tag{A.2}$$

Proof. Since $\mathbf{I}_{m\times m}$ is invertible, the formula for the determinant of a block matrix gives

$$\det\begin{pmatrix}\mathbf{I}_{m\times m} & -\mathbf{A}\\ \mathbf{B} & \mathbf{I}_{n\times n}\end{pmatrix} = \det(\mathbf{I}_{m\times m})\det(\mathbf{I}_{n\times n}-\mathbf{B}\mathbf{I}_{m\times m}^{-1}(-\mathbf{A})) = \det(\mathbf{I}_{n\times n}+\mathbf{B}\mathbf{A}). \tag{A.3}$$

Since $\mathbf{I}_{n\times n}$ is invertible,

$$\det\begin{pmatrix}\mathbf{I}_{m\times m} & -\mathbf{A}\\ \mathbf{B} & \mathbf{I}_{n\times n}\end{pmatrix} = \det(\mathbf{I}_{m\times m}-(-\mathbf{A})\mathbf{I}_{n\times n}^{-1}\mathbf{B})\det(\mathbf{I}_{n\times n}) = \det(\mathbf{I}_{m\times m}+\mathbf{A}\mathbf{B}). \tag{A.4}$$

Thus $\det(\mathbf{I}_{n\times n}+\mathbf{B}\mathbf{A}) = \det(\mathbf{I}_{m\times m}+\mathbf{A}\mathbf{B})$. ∎

We can then reformulate Ψ^ρ in (3.4) as

$$\begin{aligned}\Psi^\rho(\mathbf{W}) &= \frac{1}{2}\log\det(\mathbf{I}_\rho + \widetilde{\mathbf{H}}_m^\rho)\\ &= \frac{1}{2}\log\det(\mathbf{I}_\rho + \boldsymbol{\Sigma}_{\text{pr}}^{\frac{1}{2}}(\mathbf{W}\mathbf{F}_d\mathbf{P}_\dagger)^*\boldsymbol{\Gamma}_\eta^{-1}(\mathbf{W})(\mathbf{W}\mathbf{F}_d\mathbf{P}_\dagger)\boldsymbol{\Sigma}_{\text{pr}}^{\frac{1}{2}}),\end{aligned} \tag{A.5}$$

where

$$\begin{aligned}\boldsymbol{\Gamma}_\eta(\mathbf{W}) &= \boldsymbol{\Gamma}_{\text{n}}(\mathbf{W}) + \mathbf{F}(\mathbf{W})(\boldsymbol{\Gamma}_{\text{pr}} - \boldsymbol{\Gamma}_{\text{pr}}\mathbf{P}^*\boldsymbol{\Sigma}_{\text{pr}}^{-1}\mathbf{P}\boldsymbol{\Gamma}_{\text{pr}})\mathbf{F}^*(\mathbf{W})\\ &= \mathbf{W}\boldsymbol{\Gamma}_{\text{n}}^d\mathbf{W}^T + \mathbf{W}\mathbf{F}_d(\boldsymbol{\Gamma}_{\text{pr}} - \boldsymbol{\Gamma}_{\text{pr}}\mathbf{P}^*\boldsymbol{\Sigma}_{\text{pr}}^{-1}\mathbf{P}\boldsymbol{\Gamma}_{\text{pr}})\mathbf{F}_d^*\mathbf{W}^T\\ &= \mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \mathbf{F}_d(\boldsymbol{\Gamma}_{\text{pr}} - \boldsymbol{\Gamma}_{\text{pr}}\mathbf{P}^*\boldsymbol{\Sigma}_{\text{pr}}^{-1}\mathbf{P}\boldsymbol{\Gamma}_{\text{pr}})\mathbf{F}_d^*)\mathbf{W}^T\\ &= \mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \underbrace{\mathbf{F}_d\boldsymbol{\Gamma}_{\text{pr}}\mathbf{F}_d^*}_{:=\mathbf{H}_d\in\mathbb{R}^{d\times d}} - \underbrace{\mathbf{F}_d\boldsymbol{\Gamma}_{\text{pr}}\mathbf{P}^*\boldsymbol{\Sigma}_{\text{pr}}^{-1}\mathbf{P}\boldsymbol{\Gamma}_{\text{pr}}\mathbf{F}_d^*}_{:=\mathbf{H}_d^\rho\in\mathbb{R}^{d\times d}})\mathbf{W}^T\\ &= \mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \underbrace{\mathbf{H}_d - \mathbf{H}_d^\rho}_{:=\Delta\mathbf{H}_d})\mathbf{W}^T\\ &= \mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T.\end{aligned} \tag{A.6}$$

$\boldsymbol{\Gamma}_{\text{n}}^d$ and $\mathbf{F}_d$ are defined in (2.12) and (2.13).

To this end, we have

$$\begin{aligned}\Psi(\mathbf{W})^\rho &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_\rho + \boldsymbol{\Sigma}_{\text{pr}}^{\frac{1}{2}}(\mathbf{W}\mathbf{F}_d\mathbf{P}_\dagger)^*\boldsymbol{\Gamma}_\eta^{-1}(\mathbf{W})(\mathbf{W}\mathbf{F}_d\mathbf{P}_\dagger)\boldsymbol{\Sigma}_{\text{pr}}^{\frac{1}{2}}\right)\\ &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_\rho + \underbrace{\boldsymbol{\Sigma}_{\text{pr}}^{\frac{1}{2}}(\mathbf{W}\mathbf{F}_d\mathbf{P}_\dagger)^*\mathbf{L}}_{A}\underbrace{\mathbf{L}^T(\mathbf{W}\mathbf{F}_d\mathbf{P}_\dagger)\boldsymbol{\Sigma}_{\text{pr}}^{\frac{1}{2}}}_{B}\right)\\ &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_\rho + \underbrace{\mathbf{L}^T(\mathbf{W}\mathbf{F}_d\mathbf{P}_\dagger)\boldsymbol{\Sigma}_{\text{pr}}^{\frac{1}{2}}}_{B}\underbrace{\boldsymbol{\Sigma}_{\text{pr}}^{\frac{1}{2}}(\mathbf{W}\mathbf{F}_d\mathbf{P}_\dagger)^*\mathbf{L}}_{A}\right)\\ &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_r + \mathbf{L}^T(\mathbf{W}\mathbf{F}_d\mathbf{P}_\dagger)\boldsymbol{\Sigma}_{\text{pr}}(\mathbf{W}\mathbf{F}_d\mathbf{P}_\dagger)^*\mathbf{L}\right)\\ &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_r + \mathbf{L}^T\mathbf{W}\mathbf{F}_d\boldsymbol{\Gamma}_{\text{pr}}\mathbf{P}^*\boldsymbol{\Sigma}_{\text{pr}}^{-1}\boldsymbol{\Sigma}_{\text{pr}}\boldsymbol{\Sigma}_{\text{pr}}^{-1}\mathbf{P}\boldsymbol{\Gamma}_{\text{pr}}\mathbf{F}_d^*\mathbf{W}^T\mathbf{L}\right)\\ &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_r + \mathbf{L}^T\mathbf{W}\mathbf{F}_d\boldsymbol{\Gamma}_{\text{pr}}\mathbf{P}^*\boldsymbol{\Sigma}_{\text{pr}}^{-1}\mathbf{P}\boldsymbol{\Gamma}_{\text{pr}}\mathbf{F}_d^*\mathbf{W}^T\mathbf{L}\right)\\ &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_r + \mathbf{L}^T\mathbf{W}\mathbf{H}_d^\rho\mathbf{W}^T\mathbf{L}\right),\end{aligned} \tag{A.7}$$

where we use the Cholesky decomposition $\boldsymbol{\Gamma}_\eta^{-1} = \mathbf{L}\mathbf{L}^T$ in the second equality, Proposition A.1 in the third, definition of $\mathbf{P}_\dagger$ from (3.2) in the fifth, and the definition of $\mathbf{H}_d^\rho$ from (3.7) in the last.

Proof of Theorem 3.3. We first introduce necessary properties that are proven in [62] for Proposition A.2, [5] for Proposition A.3 and [65] for Proposition A.4.

PROPOSITION A.2. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be matrices of size* $m\times n$ *and* $n\times m$*, respectively, then* $\mathbf{AB}$ *and* $\mathbf{BA}$ *have the same nonzero eigenvalues.*

PROPOSITION A.3. *Let* $\mathbf{A},\mathbf{B}\in\mathbb{C}^{n\times n}$ *be Hermitian positive semidefinite with* $\mathbf{A}\geq\mathbf{B}$ *(i.e.,* $\mathbf{A}-\mathbf{B}$ *is Hermitian positive semidefinite), then*

$$0\leq\log\det(\mathbf{I}+\mathbf{A})-\log\det(\mathbf{I}+\mathbf{B})\leq\log\det(\mathbf{I}+\mathbf{A}-\mathbf{B}). \tag{A.8}$$

PROPOSITION A.4. *Let* $f:\mathbb{R}_+\to\mathbb{R}$ *be a continuous function that is differentiable on* $\mathbb{R}_+$ *(with* $x\geq 0$ *for* $x\in\mathbb{R}_+$*). If the function* $x\mapsto xf'(x)$ *is monotonically increasing on* $\mathbb{R}_+$*, then for any matrices* $\mathbf{A},\mathbf{B}\in\mathbb{R}^{n\times m}$*, it holds that*

$$\sum_{i=1}^{n}f(\upsilon_i(\mathbf{AB}^T))\leq\sum_{i=1}^{n}f(\upsilon_i(\mathbf{A})\upsilon_i(\mathbf{B})), \tag{A.9}$$

where $\upsilon_i(\cdot)$ *denotes the singular values of matrices sorted in nonincreasing order.*

LEMMA A.5. *Let* $\mathbf{A}\in\mathbb{R}^{n\times m},\mathbf{B}\in\mathbb{R}^{m\times m}$*.* $\mathbf{A}^T\mathbf{A}$ *and* $\mathbf{B}$ *are Hermitian positive semidefinite, then*

$$\log\det(\mathbf{I}+\mathbf{ABA}^T)\leq\sum_{i=1}^{m}\log(1+\upsilon_i(\mathbf{A}^T\mathbf{A})\upsilon_i(\mathbf{B})). \tag{A.10}$$

Proof. Since $\log\det(\mathbf{I}+\mathbf{ABA}^T)=\sum_{i=1}^{n}\log(1+\upsilon_i(\mathbf{ABA}^T))=\sum_{i=1}^{n}\log(1+\upsilon_i^2(\mathbf{AB}^{1/2}))$, letting $f(x)=\log(1+x^2)$, which satisfies Proposition A.4, we have

$$\sum_{i=1}^{n}\log(1+\upsilon_i^2(\mathbf{AB}^{1/2}))\leq\sum_{i=1}^{n}\log(1+\upsilon_i^2(\mathbf{A})\upsilon_i^2(\mathbf{B}^{1/2}))=\sum_{i=1}^{m}\log(1+\upsilon_i(\mathbf{A}^T\mathbf{A})\upsilon_i(\mathbf{B})). \tag{A.11}$$

∎

Denote the eigenvalue decompositions of $\mathbf{H}_d^\rho$ and $\Delta\mathbf{H}_d$ as

$$\mathbf{H}_d^\rho=\mathbf{U}_k\mathbf{Z}_k\mathbf{U}_k^T+\mathbf{U}_\perp\mathbf{Z}_\perp\mathbf{U}_\perp^T,\ \text{and}\ \Delta\mathbf{H}_d=\mathbf{V}_l\mathbf{\Lambda}_l\mathbf{V}_l^T+\mathbf{V}_\perp\mathbf{\Lambda}_\perp\mathbf{V}_\perp^T, \tag{A.12}$$

where $(\mathbf{Z}_k,\mathbf{U}_k),(\mathbf{V}_l,\mathbf{\Lambda}_l)$ represent the dominant eigenpairs, and $(\mathbf{Z}_\perp,\mathbf{U}_\perp)$, $(\mathbf{V}_\perp,\Lambda_\perp)$ represent the remaining eigenpairs. By the triangle inequality, we have

$$\begin{aligned}
&|\Psi^\rho(\mathbf{W})-\hat{\Psi}^\rho(\mathbf{W})|\\
&=|\frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{r\times r}+\mathbf{L}^T\mathbf{W}\mathbf{H}_d^\rho\mathbf{W}^T\mathbf{L}\right)-\frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{r\times r}+\hat{\mathbf{L}}^T\mathbf{W}\hat{\mathbf{H}}_d^\rho\mathbf{W}^T\hat{\mathbf{L}}\right)|\\
&\leq\underbrace{|\frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{r\times r}+\mathbf{L}^T\mathbf{W}\mathbf{H}_d^\rho\mathbf{W}^T\mathbf{L}\right)-\frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{r\times r}+\mathbf{L}^T\mathbf{W}\hat{\mathbf{H}}_d^\rho\mathbf{W}^T\mathbf{L}\right)|}_{(a)}\\
&\quad+\underbrace{|\frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{r\times r}+\mathbf{L}^T\mathbf{W}\hat{\mathbf{H}}_d^\rho\mathbf{W}^T\mathbf{L}\right)-\frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{r\times r}+\hat{\mathbf{L}}^T\mathbf{W}\hat{\mathbf{H}}_d^\rho\mathbf{W}^T\hat{\mathbf{L}}\right)|}_{(b)}.
\end{aligned} \tag{A.13}$$

We first look at (a). By Proposition A.3 and noting that $(\mathbf{H}_d^\rho - \hat{\mathbf{H}}_d^\rho) = \mathbf{U}_\perp \mathbf{Z}_\perp \mathbf{U}_\perp^T$ is Hermitian positive semidefinite, we have

$$\begin{aligned}(\text{a}) &\le \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{r\times r} + \mathbf{L}^T\mathbf{W}\mathbf{H}_d^\rho\mathbf{W}^T\mathbf{L} - \mathbf{L}^T\mathbf{W}\hat{\mathbf{H}}_d^\rho\mathbf{W}^T\mathbf{L}\right)\\ &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{r\times r} + \mathbf{L}^T\mathbf{W}(\mathbf{H}_d^\rho - \hat{\mathbf{H}}_d^\rho)\mathbf{W}^T\mathbf{L}\right)\\ &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{r\times r} + \mathbf{L}^T\mathbf{W}\mathbf{U}_\perp\mathbf{Z}_\perp\mathbf{U}_\perp^T\mathbf{W}^T\mathbf{L}\right).\end{aligned} \tag{A.14}$$

Then applying Proposition A.1, we have

$$\begin{aligned}(\text{a}) &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{r\times r} + \mathbf{L}^T\mathbf{W}\mathbf{U}_\perp\mathbf{Z}_\perp^{1/2}\mathbf{Z}_\perp^{1/2}\mathbf{U}_\perp^T\mathbf{W}^T\mathbf{L}\right)\\ &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{(d-k)\times(d-k)} + \mathbf{Z}_\perp^{1/2}\mathbf{U}_\perp^T\mathbf{W}^T\mathbf{L}\mathbf{L}^T\mathbf{W}\mathbf{U}_\perp\mathbf{Z}_\perp^{1/2}\right)\\ &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{(d-k)\times(d-k)} + \mathbf{Z}_\perp^{1/2}\mathbf{U}_\perp^T\mathbf{W}^T(\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T)^{-1}\mathbf{W}\mathbf{U}_\perp\mathbf{Z}_\perp^{1/2}\right).\end{aligned} \tag{A.15}$$

Applying Lemmas A.1 and A.5, let $\mathbf{A} = \mathbf{Z}_\perp^{1/2}\mathbf{U}_\perp^T\mathbf{W}^T, \mathbf{B} = (\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T)^{-1}$, we have

$$\begin{aligned}(\text{a}) &\le \frac{1}{2}\sum_i \log(1 + \upsilon_i(\mathbf{W}\mathbf{U}_\perp\mathbf{Z}_\perp^{1/2}\mathbf{Z}_\perp^{1/2}\mathbf{U}_\perp^T\mathbf{W}^T)\upsilon_i((\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T)^{-1}))\\ &= \frac{1}{2}\sum_i \log(1 + \upsilon_i(\mathbf{W}\mathbf{U}_\perp\mathbf{Z}_\perp\mathbf{U}_\perp^T\mathbf{W}^T)\upsilon_i((\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T)^{-1})).\end{aligned} \tag{A.16}$$

By Proposition 3.1, $\Delta\mathbf{H}_d = \mathbb{C}ov[\mathbf{F}_d(\mathbf{I} - \mathbf{P}_\dagger\mathbf{P})\mathbf{m}]$, is a covariance matrix, thus is positive semidefinite. The smallest eigenvalue of $\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d$ is greater than the smallest eigenvalue of $\boldsymbol{\Gamma}_{\text{n}}^d$. Hence $\upsilon_i(\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T)) \ge \sigma_{\min}^2$, i.e., $\upsilon_i((\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T)^{-1}) \le 1/\sigma_{\min}^2$. Note that $\upsilon_i(\mathbf{W}\mathbf{U}_\perp\mathbf{Z}_\perp\mathbf{U}_\perp^T\mathbf{W}^T) \le \upsilon_i(\mathbf{U}_\perp\mathbf{Z}_\perp\mathbf{U}_\perp^T) = \zeta_i$. Thus we have

$$(\text{a}) \le \frac{1}{2}\sum_{i=k+1}^{d} \log(1 + \zeta_i/\sigma_{\min}^2). \tag{A.17}$$

Then we turn to the second part (b), with Propositions A.1 and A.3, we have

$$\begin{aligned}(\text{b}) &= \Big| -\frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{r\times r} + \mathbf{L}^T\mathbf{W}\mathbf{U}_k\mathbf{Z}_k\mathbf{U}_k^T\mathbf{W}^T\mathbf{L}\right)\\ &\quad + \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{r\times r} + \hat{\mathbf{L}}^T\mathbf{W}\mathbf{U}_k\mathbf{Z}_k\mathbf{U}_k^T\mathbf{W}^T\hat{\mathbf{L}}\right)\Big|\\ &= \frac{1}{2}\Big| -\operatorname{logdet}\left(\mathbf{I}_{k\times k} + \mathbf{Z}_k^{1/2}\mathbf{U}_k^T\mathbf{W}^T\mathbf{L}\mathbf{L}^T\mathbf{W}\mathbf{U}_k\mathbf{Z}_k^{1/2}\right)\\ &\quad + \operatorname{logdet}\left(\mathbf{I}_{k\times k} + \mathbf{Z}_k^{1/2}\mathbf{U}_k^T\mathbf{W}^T\hat{\mathbf{L}}\hat{\mathbf{L}}^T\mathbf{W}\mathbf{U}_k\mathbf{Z}_k^{1/2}\right)\Big|\\ &\le \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{k\times k} + \mathbf{Z}_k^{1/2}\mathbf{U}_k^T\mathbf{W}^T\hat{\mathbf{L}}\hat{\mathbf{L}}^T\mathbf{W}\mathbf{U}_k\mathbf{Z}_k^{1/2} - \mathbf{Z}_k^{1/2}\mathbf{U}_k^T\mathbf{W}^T\mathbf{L}\mathbf{L}^T\mathbf{W}\mathbf{U}_k\mathbf{Z}_k^{1/2}\right)\\ &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{k\times k} + \mathbf{Z}_k^{1/2}\mathbf{U}_k^T\mathbf{W}^T\underbrace{(\hat{\mathbf{L}}\hat{\mathbf{L}}^T - \mathbf{L}\mathbf{L}^T)}_{(c)}\mathbf{W}\mathbf{U}_k\mathbf{Z}_k^{1/2}\right),\end{aligned} \tag{A.18}$$

where

$$\text{(A.19)}\qquad \text{(c)} = \hat{\mathbf{L}}\hat{\mathbf{L}}^T - \mathbf{L}\mathbf{L}^T = (\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\hat{\mathbf{H}}_d)\mathbf{W}^T)^{-1} - (\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T)^{-1})$$

Noting that $(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}$ by Hua's identity [38], let $\mathbf{A} = \mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\hat{\mathbf{H}}_d)\mathbf{W}^T, \mathbf{B} = \mathbf{W}(\Delta\mathbf{H}_d - \Delta\hat{\mathbf{H}}_d)\mathbf{W}^T = \mathbf{W}\mathbf{V}_\perp\Lambda_\perp\mathbf{V}_\perp^T\mathbf{W}^T$, we have

$$\begin{aligned}\text{(A.20)}\qquad (\mathbf{A}+\mathbf{B})^{-1} &= (\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T)^{-1} = (\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\hat{\mathbf{H}}_d)\mathbf{W}^T)^{-1} \\ &\quad - (\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\hat{\mathbf{H}}_d)\mathbf{W}^T)^{-1}\mathbf{W}\mathbf{V}_\perp\boldsymbol{\Lambda}_\perp\mathbf{V}_\perp^T\mathbf{W}^T(\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T)^{-1}.\end{aligned}$$

Then, we can see that

$$\begin{aligned}\text{(A.21)}\qquad \text{(c)} &= \underbrace{(\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\hat{\mathbf{H}}_d)\mathbf{W}^T)^{-1}}_{\mathbf{C}_1}\mathbf{W}\mathbf{V}_\perp\boldsymbol{\Lambda}_\perp\mathbf{V}_\perp^T\mathbf{W}^T\underbrace{(\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T)^{-1}}_{\mathbf{C}_2} \\ &:= \mathbf{C}_1\mathbf{W}\mathbf{V}_\perp\boldsymbol{\Lambda}_\perp\mathbf{V}_\perp^T\mathbf{W}^T\mathbf{C}_2.\end{aligned}$$

Thus,

$$\begin{aligned}\text{(A.22)}\qquad \text{(b)} &\le \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{k\times k} + \mathbf{Z}_k^{1/2}\mathbf{U}_k^T\mathbf{W}^T\mathbf{C}_1\mathbf{W}\mathbf{V}_\perp\boldsymbol{\Lambda}_\perp\mathbf{V}_\perp^T\mathbf{W}^T\mathbf{C}_2\mathbf{W}\mathbf{U}_k\mathbf{Z}_k^{1/2}\right) \\ &= \frac{1}{2}\operatorname{logdet}\left(\mathbf{I}_{(d-l)\times(d-l)} + \boldsymbol{\Lambda}_\perp^{1/2}\mathbf{V}_\perp^T\mathbf{W}^T\mathbf{C}_2\mathbf{W}\mathbf{U}_k\mathbf{Z}_k\mathbf{U}_k^T\mathbf{W}^T\mathbf{C}_1\mathbf{W}\mathbf{V}_\perp\boldsymbol{\Lambda}_\perp^{1/2}\right).\end{aligned}$$

Applying Lemma A.5, we have

$$\begin{aligned}\text{(A.23)}\qquad \text{(b)} &\le \frac{1}{2}\sum_i \log(1 + \upsilon_i(\mathbf{W}\mathbf{V}_\perp\boldsymbol{\Lambda}_\perp\mathbf{V}_\perp^T\mathbf{W}^T)\upsilon_i(\mathbf{C}_2\mathbf{W}\mathbf{U}_k\mathbf{Z}_k\mathbf{U}_k^T\mathbf{W}^T\mathbf{C}_1)) \\ &\le \frac{1}{2}\sum_{i=l+1}^{k}\log(1 + \lambda_i\zeta_1/\sigma_{\min}^4),\end{aligned}$$

where we have used

$$\begin{aligned}\text{(A.24)}\qquad &\upsilon_i(\mathbf{C}_2\mathbf{W}\mathbf{U}_k\mathbf{Z}_k\mathbf{U}_k^T\mathbf{W}^T\mathbf{C}_1) \\ &= \upsilon_i((\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T)^{-1}\mathbf{W}\mathbf{U}_k\mathbf{Z}_k\mathbf{U}_k^T\mathbf{W}^T(\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\hat{\mathbf{H}}_d)\mathbf{W}^T)^{-1}) \\ &\le \upsilon_1((\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\mathbf{H}_d)\mathbf{W}^T)^{-1})\upsilon_1(\mathbf{W}\mathbf{U}_k\mathbf{Z}_k\mathbf{U}_k^T\mathbf{W}^T)\upsilon_1((\mathbf{W}(\boldsymbol{\Gamma}_{\text{n}}^d + \Delta\hat{\mathbf{H}}_d)\mathbf{W}^T)^{-1}) \\ &\le \zeta_1/\sigma_{\min}^4\end{aligned}$$

for $i \le k$ in the last inequality. Note that it vanishes for $i > k$ as Z_k has rank not larger than k. Combining (A.17) and (A.23),

$$\text{(A.25)}\qquad |\Psi^\rho(\mathbf{W}) - \hat{\Psi}^\rho(\mathbf{W})| \le \text{(a)} + \text{(b)} \le \frac{1}{2}\sum_{i=k+1}^{d}\log(1 + \zeta_i/\sigma_{\min}^2) + \frac{1}{2}\sum_{i=l+1}^{k}\log(1 + \lambda_i\zeta_1/\sigma_{\min}^4).$$

Appendix B. Low-rank approximation. To compute the low-rank approximations of $\Delta\mathbf{H}_d$ and $\mathbf{H}_d^\rho$ as described in subsection 3.3, we present the randomized SVD algorithm for these two quantities. Recall the explicit forms of $\Delta\mathbf{H}_d$ and $\mathbf{H}_d^\rho$ as

$$\text{(B.1)}\qquad \mathbf{H}_d^\rho = \mathbf{F}_d\boldsymbol{\Gamma}_{\text{pr}}\mathbf{P}^*\boldsymbol{\Sigma}_{\text{pr}}^{-1}\mathbf{P}\boldsymbol{\Gamma}_{\text{pr}}\mathbf{F}_d^*, \Delta\mathbf{H}_d = \mathbf{F}_d\boldsymbol{\Gamma}_{\text{pr}}\mathbf{F}_d^* - \mathbf{F}_d\boldsymbol{\Gamma}_{\text{pr}}\mathbf{P}^*\boldsymbol{\Sigma}_{\text{pr}}^{-1}\mathbf{P}\boldsymbol{\Gamma}_{\text{pr}}\mathbf{F}_d^*.$$

Algorithm B.1 Randomized SVD to compute $\mathbf{H}$ with low rank k

1: Generate independent and identically distributed Gaussian matrix $\boldsymbol{\Omega} \in \mathbb{R}^{d\times(k+p)}$ with an oversampling parameter p very small (e.g., $p = 10$).
2: Compute $\boldsymbol{Y} = \mathbf{H}\boldsymbol{\Omega}$.
3: Compute the QR factorization $\mathbf{Y} = \mathbf{QR}$ satisfying $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$.
4: Compute $\mathbf{B} = \mathbf{Q}^T\mathbf{HQ}$.
5: Solve an eigenvalue problem for $\mathbf{B}$ such that $\mathbf{B} = \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}^T$.
6: Form $\mathbf{U}_k = \mathbf{QZ}[1:k]$ and $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}[1:k, 1:k]$.

We see that this is a matrix-free eigensolver. Steps 2 and 4 represent $\Delta\mathbf{H}_d$ action on $2(l+p)$ vectors and $\mathbf{H}_d^\rho$ action on $2(k+p)$ vectors. In terms of the total actions, it requires $2(2l+k+2p)$ forward operator $\mathbf{F}$ and $2(l+k+2p)$ of its adjoint $\mathbf{F}^*$, $2(k+l+2p)$ prediction operator $\mathbf{P}$ and its adjoint $\mathbf{P}^*$.

For the contaminant problem given in subsection 4.1, the concentration field $u(x,t)$ is given by

$$\begin{aligned} u_t - k\Delta u + \boldsymbol{v}\cdot\nabla u = 0 \text{ in } \mathcal{D}\times(0,T),\\ u(\cdot,0) = m \text{ in } \mathcal{D},\\ k\nabla u\cdot\boldsymbol{n} = 0 \text{ on } \partial\mathcal{D}\times(0,T); \end{aligned} \tag{B.2}$$

we can form the PtO map $\mathbf{F}$ with $\mathbf{Fm}$ as the discretized value of $\mathcal{B}u(m)$, where $\mathcal{B}$ is the pointwise observation operator. The adjoint problem is a terminal value problem which can be solved backwards in time by the equation:

$$\begin{aligned} -p_t - \nabla\cdot(p\boldsymbol{v}) - k\Delta p = \mathcal{B}^*\mathbf{y} \text{ in } \mathcal{D}\times(0,T),\\ p(\cdot,T) = 0 \text{ in } \mathcal{D},\\ (p\boldsymbol{v} + k\nabla p)\cdot\boldsymbol{n} = 0 \text{ on } \partial\mathcal{D}\times(0,T). \end{aligned} \tag{B.3}$$

Then we can define the adjoint of the PtO map $\mathbf{F}^*$ with $\mathbf{F}^*\mathbf{y}$ as the discretized value of $p(x,0)$ for any $\mathbf{y}$.

REFERENCES

[1] A. Alexanderian, P. J. Gloor, and O. Ghattas, *On Bayesian A-and D-optimal experimental designs in infinite dimensions*, Bayesian Anal., 11 (2016), pp. 671–695, https://doi.org/10.1214/15-BA969.

[2] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas, *A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized ℓ_0-sparsification*, SIAM J. Sci. Comput., 36 (2014), pp. A2122–A2148, https://doi.org/10.1137/130933381.

[3] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas, *A fast and scalable method for A-optimal design of experiments for infinite-dimensional Bayesian nonlinear inverse problems*, SIAM J. Sci. Comput., 38 (2016), pp. A243–A272, https://doi.org/10.1137/140992564.

[4] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas, *Mean-variance risk-averse optimal control of systems governed by PDEs with random parameter fields using quadratic approximations*, SIAM/ASA J. Uncertain. Quantif., 5 (2017), pp. 1166–1192, https://doi.org/10.1137/16M106306X.

[5] A. Alexanderian and A. K. Saibaba, *Efficient D-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems*, SIAM J. Sci. Comput., 40 (2018), pp. A2956–A2985, https://doi.org/10.1137/17M115712X.

[6] N. Alger, P. Chen, and O. Ghattas, *Tensor train construction from tensor actions, with application to compression of large high order derivative tensors*, SIAM J. Sci. Comput., 42 (2020), pp. A3516–A3539.

[7] N. Alger, V. Rao, A. Meyers, T. Bui-Thanh, and O. Ghattas, *Scalable matrix-free adaptive product-convolution approximation for locally translation-invariant operators*, SIAM J. Sci. Comput., 41 (2019), pp. A2296–A2328.

[8] A. Alghamdi, M. Hesse, J. Chen, U. Villa, and O. Ghattas, *Bayesian poroelastic aquifer characterization from InSAR surface deformation data. Part* II*: Quantifying the uncertainty*, Water Resour. Res., 57 (2021), e2021WR029775, https://doi.org/10.1029/2021WR029775.

[9] I. Ambartsumyan, W. Boukaram, T. Bui-Thanh, O. Ghattas, D. Keyes, G. Stadler, G. Turkiyyah, and S. Zampini, *Hierarchical matrix approximations of Hessians arising in inverse problems governed by PDEs*, SIAM J. Sci. Comput., 42 (2020), pp. A3397–A3426.

[10] A. Attia, A. Alexanderian, and A. K. Saibaba, *Goal-oriented optimal design of experiments for large-scale Bayesian linear inverse problems*, Inverse Problems, 34 (2018), 095009.

[11] T. Bakker, H. van Hoof, and M. Welling, *Experimental design for MRI by greedy policy search*, in Advances in Neural Information Processing Systems, vol. 33, Curran Associates, Red Hook, NJ, 2020.

[12] O. Bashir, K. Willcox, O. Ghattas, B. van Bloemen Waanders, and J. Hill, *Hessian-based model reduction for large-scale systems with initial condition inputs*, Internat. J. Numer. Methods Engrg., 73 (2008), pp. 844–868.

[13] A. A. Bian, J. M. Buhmann, A. Krause, and S. Tschiatschek, *Guarantees for greedy maximization of non-submodular functions with applications*, J. Mach. Learn. Res. 70 (2017), pp. 498–507.

[14] C. Boutsidis, M. W. Mahoney, and P. Drineas, *An improved approximation algorithm for the column subset selection problem*, in ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, 2009, pp. 968–977.

[15] T. Bui-Thanh, C. Burstedde, O. Ghattas, J. Martin, G. Stadler, and L. C. Wilcox, *Extreme-scale UQ for Bayesian inverse problems governed by PDEs*, in SC12: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, IEEE, Piscataway, NJ, 2012

[16] T. Bui-Thanh and O. Ghattas, *Analysis of the Hessian for inverse scattering problems. Part*I*: Inverse shape scattering of acoustic waves*, Inverse Problems, 28 (2012), 055001, https://doi.org/10.1088/0266-5611/28/5/055001.

[17] T. Bui-Thanh and O. Ghattas, *Analysis of the Hessian for inverse scattering problems. Part*II*: Inverse medium scattering of acoustic waves*, Inverse Problems, 28 (2012), 055002, https://doi.org/10.1088/0266-5611/28/5/055002.

[18] T. Bui-Thanh and O. Ghattas, *Analysis of the Hessian for inverse scattering problems. Part*III*: Inverse medium scattering of electromagnetic waves*, Inverse Probl. Imaging, 7 (2013), pp. 1139–1155.

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

[20] G. Capellari, E. Chatzi, and S. Mariani, *Optimal sensor placement through Bayesian experimental design: Effect of measurement noise and number of sensors*, Multidiscip. Digital Publ. Inst. Proc., 1 (2016), 41.

[21] K. Chaloner and I. Verdinelli, *Bayesian experimental design: A review*, Statist. Sci., 10 (1995), pp. 273–304.

[22] P. Chen and O. Ghattas, *Hessian-based sampling for high-dimensional model reduction*, Int. J. Uncertain. Quantif., 9 (2019) pp. 103–121.

[23] P. Chen and O. Ghattas, *Projected Stein variational gradient descent*, in Advances in Neural Information Processing Systems, vol. 33, Curran Associates, Red Hook, NY, 2020.

[24] P. Chen, M. Haberman, and O. Ghattas, *Optimal design of acoustic metamaterial cloaks under uncertainty*, J. Comput. Phys., 431 (2021), 110114.

[25] P. Chen, U. Villa, and O. Ghattas, *Hessian-based adaptive sparse quadrature for infinite-dimensional Bayesian inverse problems*, Comput. Methods Appl. Mech. Engrg., 327 (2017), pp. 147–172, https://doi.org/10.1016/j.cma.2017.08.016.

[26] P. Chen, U. Villa, and O. Ghattas, *Taylor approximation and variance reduction for PDE-constrained optimal control under uncertainty*, J. Comput. Phys., 385 (2019), pp. 163–186.

[27] P. Chen, K. Wu, J. Chen, T. O'Leary-Roseberry, and O. Ghattas, *Projected Stein variational Newton: A fast and scalable Bayesian inference method in high dimensions*, in Advances in Neural Information Processing Systems, vol. 32, Curran Associates, Red Hook, NY, 2019.

[28] P. Chen, K. Wu, and O. Ghattas, *Bayesian inference of heterogeneous epidemic models: Application to COVID-19 spread accounting for long-term care facilities*, Comput. Methods Appl. Mech. Engrg., 385 (2021), 114020.
[29] B. Crestel, A. Alexanderian, G. Stadler, and O. Ghattas, *A-optimal encoding weights for nonlinear inverse problems, with application to the Helmholtz inverse problem*, Inverse Problems, 33 (2017), 074008, http://iopscience.iop.org/article/10.1088/1361-6420/aa6d8e.
[30] B. Crestel, G. Stadler, and O. Ghattas, *A comparative study of regularizations for joint inverse problems*, Inverse Problems, 35 (2018), 024003.
[31] Y. Daon and G. Stadler, *Mitigating the influence of boundary conditions on covariance operators derived from elliptic PDEs*, Inverse Probl. Imaging, 12 (2018), pp. 1083–1102.
[32] V. V. Fedorov, *Theory of Optimal Experiments*, Academic, New York, 1972.
[33] A. R. Ferrolino, J. E. C. Lope, and R. G. Mendoza, *Optimal location of sensors for early detection of tsunami waves*, in International Conference on Computational Science, Springer, Cham, Switzerland, 2020, pp. 562–575.
[34] H. P. Flath, L. C. Wilcox, V. Akçelik, J. Hill, B. van Bloemen Waanders, and O. Ghattas, *Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations*, SIAM J. Sci. Comput., 33 (2011), pp. 407–432, https://doi.org/10.1137/090780717.
[35] A. Foster, M. Jankowiak, E. Bingham, P. Horsfall, Y. W. Teh, T. Rainforth, and N. Goodman, *Variational Bayesian optimal experimental design*, in Advances in Neural Information Processing Systems, vol. 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett, eds., Curran Associates, Red Hook, NY, 2019, pp. 14036–14047, https://proceedings.neurips.cc/paper/2019/file/d55cbf210f175f4a37916eafe6c04f0d-Paper.pdf.
[36] O. Ghattas and K. Willcox, *Learning physics-based models from data: Perspectives from inverse problems and model reduction*, Acta Numer., 30 (2021), pp. 445–554, https://doi.org/10.1017/S0962492921000064.
[37] N. Halko, P. G. Martinsson, and J. A. Tropp, *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, SIAM Rev., 53 (2011), pp. 217–288.
[38] L.-K. Hua, *Inequalities involving determinants*, Trans. Amer. Math. Soc. Ser. II, 32 (1964), pp. 265–272.
[39] X. Huan and Y. M. Marzouk, *Simulation-based optimal Bayesian experimental design for nonlinear systems*, J. Comput. Phys., 232 (2013), pp. 288–317, https://doi.org/10.1016/j.jcp.2012.08.013.
[40] X. Huan and Y. M. Marzouk, *Gradient-based stochastic optimization methods in Bayesian experimental design*, Int. J. Uncertain. Quantif., 4 (2014), pp. 479–510.
[41] X. Huan and Y. M. Marzouk, *Sequential Bayesian Optimal Experimental Design via Approximate Dynamic Programming*, preprint, https://doi.org/arXiv:1604.08320, 2016.
[42] T. Isaac, N. Petra, G. Stadler, and O. Ghattas, *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 (2015), pp. 348–368, https://doi.org/10.1016/j.jcp.2015.04.047.
[43] J. Jagalur-Mohan and Y. Marzouk, *Batch greedy maximization of non-submodular functions: Guarantees and applications to experimental design*, J. Mach. Learn. Res., 22 (2021), pp. 1–62.
[44] A. G. Kalmikov and P. Heimbach, *A Hessian-based method for uncertainty quantification in global ocean state estimation*, SIAM J. Sci. Comput., 36 (2014), pp. S267–S295.
[45] K. Kandasamy, W. Neiswanger, R. Zhang, A. Krishnamurthy, J. Schneider, and B. Poczos, *Myopic posterior sampling for adaptive goal oriented design of experiments*, Proc. Int. Mach. Learn. (PMLR), 97 (2019), pp. 3222–3232.
[46] M. R. Khodja, M. D. Prange, and H. A. Djikpesse, *Guided Bayesian optimal experimental design*, Inverse Problems, 26 (2010), 055008, https://doi.org/10.1088/0266-5611/26/5/055008.
[47] S. Kleinegesse and M. U. Gutmann, *Bayesian experimental design for implicit models by mutual information neural estimation*, Proc. Int. Mach. Learn. (PMLR), 89 (2020), pp. 5316–5326.
[48] A. Krause and D. Golovin, *Submodular function maximization*, Tractability, 3 (2014), pp. 71–104.
[49] Q. Long, M. Scavino, R. Tempone, and S. Wang, *Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations*, Comput. Methods Appl. Mech. Engrg, 259 (2013), pp. 24–39.

[50] N. Loose and P. Heimbach, *Leveraging uncertainty quantification to design ocean climate observing systems*, J. Adv. Model. Earth Syst., 13 (2021), e2020MS002386.

[51] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, *Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns*, IEEE Control Syst. Mag., 38 (2018), pp. 63–86, https://doi.org/10.1109/MCS.2018.2810460.

[52] G. Nemhauser, L. Wolsey, and M. Fisher, *An analysis of approximations for maximizing submodular set functions*—I, Math. Program., 14 (1978), pp. 265–294, https://doi.org/10.1007/BF01588971.

[53] T. O'Leary-Roseberry, U. Villa, P. Chen, and O. Ghattas, *Derivative-informed projected neural networks for high-dimensional parametric maps governed by PDEs*, Comput. Methods Appl. Mech. Engrg., 388 (2022), 114199.

[54] C. Papadimitriou, *Optimal sensor placement for parametric identification of structural systems*, J. Sound Vib., 278 (2004), pp. 923–947, https://doi.org/10.1016/j.jsv.2003.10.063.

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

[56] C. Pozrikidis, *An Introduction to Grids, Graphs, and Networks*, Oxford University Press, Oxford, 2014.

[57] A. K. Saibaba, A. Alexanderian, and I. C. Ipsen, *Randomized matrix-free trace and log-determinant estimators*, Numer. Math., 137 (2017), pp. 353–395.

[58] A. Spantini, T. Cui, K. Willcox, L. Tenorio, and Y. Marzouk, *Goal-oriented optimal approximations of Bayesian linear inverse problems*, SIAM J. Sci. Comput., 39 (2017), pp. S167–S196.

[59] S. Subramanian, K. Scheufele, M. Mehl, and G. Biros, *Where did the tumor start? An inverse solver with sparse localization for tumor growth models*, Inverse Problems, 36 (2020), 045006, https://doi.org/10.1088/1361-6420/ab649c.

[60] U. Villa, N. Petra, and O. Ghattas, *hIPPYlib*, *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 (2021), 16.

[61] J. Worthen, G. Stadler, N. Petra, M. Gurnis, and O. Ghattas, *Towards adjoint-based inversion for rheological parameters in nonlinear viscous mantle flow*, Phys. Earth Planet. Inter., 234 (2014), pp. 23–34, https://doi.org/10.1016/j.pepi.2014.06.006.

[62] K. Wu, P. Chen, and O. Ghattas, *A fast and scalable computational framework for large-scale and high-dimensional Bayesian optimal experimental design,* SIAM J. Sci. Comput., to appear.

[63] S. Yang, G. Stadler, R. Moser, and O. Ghattas, *A shape Hessian-based boundary roughness analysis of Navier–Stokes flow*, SIAM J. Appl. Math., 71 (2011), pp. 333–355, https://doi.org/10.1137/100796789.

[64] J. Yu, V. M. Zavala, and M. Anitescu, *A scalable design of experiments framework for optimal sensor placement*, J. Process Control, 67 (2018), pp. 44–55.

[65] M.-C. Yue, *A Matrix Generalization of the Hardy-Littlewood-Pólya Rearrangement Inequality and Its Applications*, preprint, https://doi.org/arXiv:2006.08144, 2020.

[66] S. Zheng, D. Hayden, J. Pacheco, and J. W. Fisher III, *Sequential Bayesian experimental design with variable cost structure*, in Advances in Neural Information Processing Systems, vol. 33, Curran Associates, Red Hook, NY, 2020.