

## STOCHASTIC LEARNING APPROACH FOR BINARY OPTIMIZATION: APPLICATION TO BAYESIAN OPTIMAL DESIGN OF EXPERIMENTS\*

AHMED ATTIA<sup>†</sup>, SVEN LEYFFER<sup>†</sup>, AND TODD S. MUNSON<sup>†</sup>

**Abstract.** We present a novel stochastic approach to binary optimization suited for optimal experimental design (OED) for Bayesian inverse problems governed by mathematical models such as partial differential equations. The OED utility function, namely, the regularized optimality criterion, is cast into a stochastic objective function in the form of an expectation over a multivariate Bernoulli distribution. The probabilistic objective is then solved by using a stochastic optimization routine to find an optimal observational policy. This formulation (a) is generally applicable to binary optimization problems with soft constraints and is ideal for OED and sensor placement problems; (b) does not require differentiability of the original objective function (e.g., a utility function in OED applications) with respect to the design variable, and thus it enables direct employment of sparsity-enforcing penalty functions such as  $\ell_0$ , without needing to utilize a continuation procedure or apply a rounding technique; (c) exhibits much lower computational cost than traditional gradient-based relaxation approaches; and (d) can be applied to both linear and nonlinear OED problems with proper choice of the utility function. The proposed approach is analyzed from an optimization perspective with detailed convergence analysis of the optimization approach and is also analyzed from a machine learning perspective with correspondence to policy gradient reinforcement learning. The approach is demonstrated numerically by using an idealized two-dimensional Bayesian linear inverse problem and validated by extensive numerical experiments carried out for sensor placement in a parameter identification setup.

**Key words.** design of experiments, binary optimization, Bayesian inverse problem, reinforcement learning, stochastic optimization

**AMS subject classifications.** 62K05, 35Q62, 62F15, 35R30, 35Q93, 65C60, 93E35

**DOI.** 10.1137/21M1404363

**1. Introduction.** Optimal experimental design (OED) is the general mathematical formalism for configuring an observational setup. This can refer to the frequency of data collection or the spatiotemporal distribution of observational gridpoints for physical experiments. OED has seen a recent surge of interest in the field of sensor placement for applications of model-constrained Bayesian inverse problems; see, for example, [1, 2, 3, 4, 6, 9, 10, 22, 24, 25, 26, 28, 39, 41, 52].

Many physical phenomena can be simulated by using mathematical models. The accuracy of the mathematical models is limited, however, by the level at which the physics of the true system is captured by the model and by the numerical errors produced, for example, by computer simulations. Model-based simulations are often corrected based on snapshots of reality, such as sensor-based measurements. Such corrections involve first inferring the model parameter from the noisy measurements, a problem referred to as the “inverse problem,” which can be solved by different inversion or data assimilation methods; see e.g., [11, 12, 14, 19, 36]. These observations themselves are noisy but in general follow a known probability distribution. The

---

\*Submitted to the journal’s Computational Methods in Science and Engineering section March 11, 2021; accepted for publication (in revised form) November 19, 2021; published electronically April 13, 2022.

<https://doi.org/10.1137/21M1404363>

**Funding:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357.

<sup>†</sup>Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439 USA ([attia@mcs.anl.gov](mailto:attia@mcs.anl.gov), [leyffer@mcs.anl.gov](mailto:leyffer@mcs.anl.gov), [tmunson@mcs.anl.gov](mailto:tmunson@mcs.anl.gov)).

OED problem is concerned with finding the most informative, and thus optimal, observational grid configuration out of a set of candidates, which, when deployed, will result in a reliable solution of the inverse problem.

OED is most beneficial when there is a limit on the number of sensors that can be used, for example, when sensors are expensive to deploy or operate. In order to solve an OED problem for sensor placement, a binary optimization problem is formulated by assigning a binary design variable  $\zeta_i$  for each candidate sensor location. The objective is often set to a utility function that summarizes the inversion parameter uncertainty or the amount of information gained from the observations. The optimal design is then defined as the optimizer of this objective function. The objective is constrained by the model evolution equations, such as partial differential equations, and also by any regularity or sparsity constraints on the design. Here we focus on sparse and binary designs. A popular choice of a penalty function to enforce a sparse and binary design is the  $\ell_0$  penalty [3].

Solving binary optimization problems is computationally prohibitive, and often the binary problem is replaced with a relaxation. In the relaxation of an OED problem, the design variable is allowed to take any value in the interval  $[0, 1]$  and is often interpreted as an importance weight of the candidate location. Once a solution of the relaxed problem is obtained numerically, for example, by a gradient-based optimization procedure, a rounding procedure is applied to transform the solution of the relaxed problem into a candidate solution of the original problem. In general, however, the numerical solutions of the two problems are not guaranteed to be related or even similar.

Using a gradient-based approach to solve the relaxed OED problem requires many evaluations of the simulation model in order to evaluate the objective function and its gradient. Moreover, this approach requires the penalty function to be differentiable with respect to the design. To this end, since  $\ell_0$  is discontinuous, numerous efforts have been dedicated to approximating the effect of  $\ell_0$  sparsification with other approaches. For example, in [24], an  $\ell_1$  penalty followed by a thresholding method is used; however,  $\ell_1$  is nonsmooth and hence is nondifferentiable. Continuation procedures are used in [3, 33], where a series of OED problems are solved with a sequence of penalty functions that successively approximate  $\ell_0$  sparsification. Another approach that can potentially induce a binary design is the sum-up-rounding algorithm [56], which also provides optimality guarantees.

Here we propose a new efficient approach for directly solving binary optimization problems which is ideally suited for OED and sensor placement applications. In this framework, the objective function, namely, the regularized optimality criterion, is cast into an objective function in the form of an expectation over a multivariate Bernoulli distribution. A stochastic optimization approach is then applied to solve the reformulated optimization problem in order to find an optimal observational policy. Related work on stochastic optimization for OED has been explored in [27], where the utility function, in other words, the optimality criterion, is approximated by using Monte Carlo (MC) estimators. A stochastic optimization procedure is then applied with the gradient of the utility function with respect to the design being approximated by using Robbins–Monro stochastic approximation [43] or sample average approximation [37, 48].

The main differences from the standard OED approach and previous stochastic OED approaches, which highlight our main contributions in this work, are as follows. First, the proposed methodology searches for an optimal probability distribution representing the optimal binary design; hence it does not require relaxation or binarization to follow the optimization step as in the traditional OED formulation.

Second, in our approach the penalized OED criterion is not required to be differentiable with respect to the design. Thus, one can incorporate the  $\ell_0$  regularization norm to enforce sparsification without needing to approximate it with the  $\ell_1$  penalty term or apply a continuation procedure to retrieve a binary optimal design. Third, the proposed stochastic OED formulation does not require formulating the gradient of the objective function with respect to the design, thus implying a massive reduction in computation cost. Fourth, the proposed framework is generally applicable to regularized binary optimization problems with nondifferentiable penalty terms and is not limited to OED applications. Fifth, when used to solve OED binary optimization problems, the proposed approach and solution algorithms are completely independent from the specific choice of the OED optimality criterion and the penalty function and hence can be used with both linear and nonlinear problems.

In this work, for simplicity we provide numerical experiments for Bayesian inverse problems constrained by linear forward operators, and we consider an A-optimal design, that is, a design that minimizes the trace of the posterior covariances of the inversion parameter. The approach proposed, however, applies to other OED optimality criteria, such as D-optimality and Kullback–Leibler (KL) divergence between the prior and posterior [28]. We provide an analysis of the method from a mathematical point of view and an interpretation from a machine learning (ML) perspective. Specifically, the proposed algorithm has a direct link to policy gradient algorithms widely used in neuro-dynamic programming and reinforcement learning [15, 50, 54]. Numerical experiments using a toy example are provided to help with understanding the problem; the proposed algorithm; and the relation between the binary OED problem, the relaxed OED problem, and the proposed formulation. Moreover, extensive numerical experiments are performed for optimal sensor placement for an advection-diffusion problem that simulates the spatiotemporal evolution of the concentration of a contaminant in a bounded domain.

The paper is organized as follows. Section 2 gives a brief description of the Bayesian inverse problem and the standard formulation of an optimal experimental design in the context of Bayesian inversion. In section 3, we describe our proposed approach for solving the binary OED problem and present a detailed analysis of the proposed algorithms. An explanatory example and extensive numerical experiments are given in section 4. Discussion and conclusions are presented in section 5.

**2. OED for Bayesian inversion.** Consider the forward model described by

$$(2.1) \quad \mathbf{y} = \mathbf{F}(\theta) + \delta,$$

where  $\theta \in \mathbb{R}^{N_{\text{state}}}$  is the discretized model parameter and  $\mathbf{y} \in \mathbb{R}^{N_{\text{obs}}}$  is the observation, where  $\delta \in \mathbb{R}^{N_{\text{obs}}}$  is the observation error. Here  $\mathbf{F}$  is the parameter-to-observable map which projects the model parameter  $\theta$  onto the observation space. Assuming Gaussian observational noise  $\delta \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}_{\text{noise}})$ , then the data likelihood takes the form

$$(2.2) \quad \mathcal{L}(\mathbf{y}|\theta) \propto \exp\left(-\frac{1}{2} \|\mathbf{F}(\theta) - \mathbf{y}\|_{\mathbf{\Gamma}_{\text{noise}}^{-1}}^2\right),$$

where  $\mathbf{\Gamma}_{\text{noise}}$  is the observation error covariance matrix (positive definite) and the matrix-weighted norm is defined as  $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^T \mathbf{A} \mathbf{x}$ .

An inverse problem refers to the retrieval of the model parameter  $\theta$  from the noisy observation  $\mathbf{y}$ , conditioned by the model dynamics; see, e.g., [30, 51]. In Bayesian inversion, the goal is to study the probability distribution of  $\theta$  conditioned by the

observation  $\mathbf{y}$  that is the posterior obtained by applying Bayes' theorem,  $\mathbb{P}(\theta|\mathbf{y}) \propto \mathbb{P}(\mathbf{y}|\theta)\mathbb{P}(\theta)$ , where  $\mathbb{P}(\theta)$  is the prior distribution of the inversion parameter  $\theta$ , which in many cases is assumed to be Gaussian  $\theta \sim \mathcal{N}(\theta_{\text{pr}}, \mathbf{\Gamma}_{\text{pr}})$ . If the parameter-to-observable map  $\mathbf{F}$  is linear, the posterior  $\mathbb{P}(\theta|\mathbf{y})$  is Gaussian  $\mathcal{N}(\theta_{\text{post}}^{\mathbf{y}}, \mathbf{\Gamma}_{\text{post}})$  with

$$(2.3) \quad \mathbf{\Gamma}_{\text{post}} = (\mathbf{F}^* \mathbf{\Gamma}_{\text{noise}}^{-1} \mathbf{F} + \mathbf{\Gamma}_{\text{pr}}^{-1})^{-1}, \quad \theta_{\text{post}}^{\mathbf{y}} = \mathbf{\Gamma}_{\text{post}} (\mathbf{\Gamma}_{\text{pr}}^{-1} \theta_{\text{pr}} + \mathbf{F}^* \mathbf{\Gamma}_{\text{noise}}^{-1} \mathbf{y}),$$

where  $\mathbf{F}^*$  is the forward operator adjoint. If the forward operator  $\mathbf{F}$  is nonlinear, however, the posterior distribution is no longer Gaussian. A Gaussian approximation of the posterior can be obtained in this case by linearization of  $\mathbf{F}$  around the maximum a posteriori (MAP) estimate, which is inherently data dependent. For simplicity, we will focus on linear models; however, the approach proposed in this work is not limited to the linear case. Further analysis for nonlinear settings will follow in separate works.

In a Bayesian OED context (see, e.g., [1, 2, 3, 9, 10, 24, 25]), the observation covariance  $\mathbf{\Gamma}_{\text{noise}}$  is replaced with a weighted version  $\mathbf{W}_{\Gamma}(\zeta)$ , parameterized by the design  $\zeta$ , resulting in the following weighted data-likelihood  $\mathcal{L}(\mathbf{y}|\theta; \zeta)$  and posterior covariance  $\mathbf{\Gamma}_{\text{post}}(\zeta)$ .

$$(2.4) \quad \mathcal{L}(\mathbf{y}|\theta; \zeta) \propto \exp\left(-\frac{1}{2} \|\mathbf{F}(\theta) - \mathbf{y}\|_{\mathbf{W}_{\Gamma}(\zeta)}^2\right), \quad \mathbf{\Gamma}_{\text{post}}(\zeta) = (\mathbf{F}^* \mathbf{W}_{\Gamma}(\zeta) \mathbf{F} + \mathbf{\Gamma}_{\text{pr}}^{-1})^{-1}.$$

The exact form of the posterior covariance  $\mathbf{\Gamma}_{\text{post}}(\zeta)$  depends on how  $\mathbf{W}_{\Gamma}(\zeta)$  is formulated. We are interested mainly in binary designs  $\zeta \in \{0, 1\}^{n_s}$ , required, for example, in sensor placement applications; see, for example, [1, 10]. In this case, we assume a set of  $n_s$  candidate sensor locations, and we seek the optimal subset of locations. Note that selecting a subset of the observations is equivalent to applying a projection operator onto the subspace spanned by the activated sensors. This is equivalent to defining the weighed design matrix as  $\mathbf{W}_{\Gamma} := \mathbf{\Gamma}_{\text{noise}}^{-1/2} \mathbf{W} \mathbf{\Gamma}_{\text{noise}}^{-1/2}$ , where  $\mathbf{W} := \text{diag}(\zeta)$  is a diagonal matrix with binary values on its diagonal. The  $i$ th entry of the design is set to 1 to activate a sensor and is set to 0 to turn it off.

The optimal design  $\zeta^{\text{opt}}$  is the one that optimizes a predefined criterion  $\Psi(\cdot)$  of choice. The most popular optimality criteria in the Bayesian inversion context are A- and D-optimality. Both criteria define the optimal design as the one that minimizes a scalar summary of posterior uncertainty associated with an inversion parameter or state of an inverse problem. Specifically,  $\Psi(\zeta) := \text{Tr}(\mathbf{\Gamma}_{\text{post}}(\zeta))$  for an A-optimal design, and  $\Psi(\zeta) := \log \det(\mathbf{\Gamma}_{\text{post}}(\zeta))$  for a D-optimal design, i.e., the sum of the eigenvalues of  $\mathbf{\Gamma}_{\text{post}}(\zeta)$ , and the sum of the logarithms of the eigenvalues of  $\mathbf{\Gamma}_{\text{post}}(\zeta)$ , respectively.

To prevent experimental designs that are simply dense, one typically adds a regularization term  $\Phi(\zeta)$  that promotes sparsity, for example. Thus, to find an optimal design  $\zeta^{\text{opt}}$ , one needs to solve the following binary optimization problem,

$$(2.5) \quad \zeta^{\text{opt}} = \underset{\zeta \in \{0, 1\}^{n_s}}{\text{arg min}} \mathcal{J}(\zeta) := \Psi(\zeta) + \alpha \Phi(\zeta),$$

where the function  $\Phi(\zeta)$  promotes regularization or sparsity on the design and  $\alpha$  is a user-defined regularization parameter. For example,  $\Phi$  could represent a budget constraint on the form  $\Phi(\zeta) := \|\zeta\|_0 \leq k; k \in \mathbf{Z}_+$  or a sparsifying (possibly discontinuous) function, for example,  $\|\zeta\|_0$  or  $\|\zeta\|_1$ .

Traditional binary optimization approaches [55] for solving the optimization problem are expensive, rendering the exact solution of (2.5) computationally intractable.

The optimization problem (2.6) is a continuous relaxation of (2.5) and is often used in practice as a surrogate for solving (2.5) with a suitable rounding scheme:

$$(2.6) \quad \zeta^{\text{opt}} = \arg \min_{\zeta \in [0,1]^{n_s}} \mathcal{J}(\zeta) := \Psi(\zeta) + \alpha \Phi(\zeta).$$

In sensor placement we seek a sparse design in order to minimize the deployment cost, and thus we would utilize  $\ell_0$  as a penalty function. Solving the relaxed optimization problem (2.6) follows a gradient-based approach, and thus using  $\ell_0$  as a penalty function is replaced with the nonsmooth  $\ell_1$  norm; see [3] for more details. The optimal solution of the relaxed problem (2.6) provides a lower bound [16, section 9.5] on (2.5), and because  $\{0, 1\}^{n_s} \subseteq [0, 1]^{n_s}$  any binary solution  $\zeta^{\text{opt}} \in \{0, 1\}^{n_s}$  of (2.6) is also an optimal solution of (2.5). Given a solution of (2.6) that is not binary, we can obtain a binary solution by rounding; however, there is no guarantee that the resulting binary solution is optimal, except in special cases, such as certain optimal control problems, where we use sum-up-rounding techniques (see, e.g., [44]). In section 3 we present a new approach for directly solving the original OED binary optimization problem (2.5) without the need to relax the design space or to round the relaxations.

**3. Stochastic learning approach for binary optimization.** Our main goal is to find the solution of the original binary regularized optimization problem (2.5) without solving a mixed-integer programming problem or resorting to the relaxation approach widely followed in OED for Bayesian inversion; see section 2. We propose to reformulate and solve the binary optimization problem (2.5) as a stochastic optimization problem defined over the parameters of a probability distribution. Specifically, we assume  $\zeta$  is a random variable that follows a multivariate Bernoulli distribution.

**3.1. Stochastic formulation of the OED binary optimization problem.**

We associate with each candidate sensor location  $x_i$  a probability of activation  $\mathbf{p}_i$ . Specifically, we assume that  $\zeta_i, i = 1, 2, \dots, n_s$ , are independent Bernoulli random variables associated with the candidate sensor locations  $x_i, i = 1, 2, \dots, n_s$ . The activation (success) probabilities of the respective sensors are  $\mathbf{p}_i, i = 1, 2, \dots, n_s$ ; that is,  $\mathbb{P}(\zeta_i = 1) = \mathbf{p}_i, \mathbb{P}(\zeta_i = 0) = 1 - \mathbf{p}_i$ . The probability associated with any observational configuration is then described by the joint probability of all candidate locations. Assuming independent activation variables  $\zeta_i$  does not interfere with the correlation structure of the observational errors, manifested by the observation error covariance matrix  $\mathbf{\Gamma}_{\text{noise}}$ . Conversely, setting  $\zeta_i$  to 0 removes the  $i$ th row and column from the precision matrix  $\mathbf{\Gamma}_{\text{noise}}^{-1}$ , while setting  $\zeta_i$  to 1 keeps the corresponding row and column, respectively. We let  $\mathbb{P}(\zeta|\mathbf{p})$  denote the joint multivariate Bernoulli probability, with the following probability mass function (PMF):

$$(3.1) \quad \mathbb{P}(\zeta|\mathbf{p}) := \prod_{i=1}^{n_s} \mathbf{p}_i^{\zeta_i} (1 - \mathbf{p}_i)^{1-\zeta_i}, \quad \zeta_i \in \{0, 1\}, \quad \mathbf{p}_i \in [0, 1].$$

We then replace the original problem (2.5) with the following stochastic optimization problem:

$$(3.2) \quad \mathbf{p}^{\text{opt}} = \arg \min_{\mathbf{p} \in [0,1]^{n_s}} \Upsilon(\mathbf{p}) := \mathbb{E}_{\zeta \sim \mathbb{P}(\zeta|\mathbf{p})}[\mathcal{J}(\zeta)] = \mathbb{E}_{\zeta \sim \mathbb{P}(\zeta|\mathbf{p})}[\Psi(\zeta) + \alpha \Phi(\zeta)].$$

Because the support of the probability distribution is discrete, the possible values of  $\zeta \in \{0, 1\}^{n_s}$  are countable and can be assigned unique indexes. An index  $k$  is

assigned to each possible realization of  $\zeta := (\zeta_1, \zeta_2, \dots, \zeta_{n_s})$ , based on the values of its components using the relation

$$(3.3) \quad k = 1 + \sum_{i=1}^{n_s} \zeta_i 2^{i-1}, \quad \zeta_i \in \{0, 1\}.$$

Thus, all possible designs are labeled as  $\zeta[k], k = 1, 2, \dots, 2^{n_s}$ . With the indexing scheme (3.3), the optimization problem (3.2) takes the following equivalent form:

$$(3.4) \quad \mathbf{p}^{\text{opt}} = \arg \min_{\mathbf{p} \in [0,1]^{n_s}} \Upsilon(\mathbf{p}) := \sum_{k=1}^{2^{n_s}} \mathcal{J}(\zeta[k]) \mathbb{P}(\zeta[k]|\mathbf{p}).$$

To solve (3.4), one can follow a gradient-based optimization approach to find the optimal parameter  $\mathbf{p}^{\text{opt}}$ . The gradient of the objective in (3.4) with respect to the distribution parameters  $\mathbf{p}$  is

$$(3.5) \quad \nabla_{\mathbf{p}} \Upsilon(\mathbf{p}) = \nabla_{\mathbf{p}} \mathbb{E}_{\zeta \sim \mathbb{P}(\zeta|\mathbf{p})} [\mathcal{J}(\zeta)] = \nabla_{\mathbf{p}} \sum_{\zeta} \mathcal{J}(\zeta) \mathbb{P}(\zeta|\mathbf{p}) = \sum_{k=1}^{2^{n_s}} \mathcal{J}(\zeta[k]) \nabla_{\mathbf{p}} \mathbb{P}(\zeta[k]|\mathbf{p}),$$

where  $\nabla_{\mathbf{p}} \mathbb{P}(\zeta|\mathbf{p})$  is the gradient of the joint Bernoulli PMF (3.1) (see Appendix A for details):

$$(3.6) \quad \nabla_{\mathbf{p}} \mathbb{P}(\zeta[k]|\mathbf{p}) = \sum_{j=1}^{n_s} \frac{\partial \mathbb{P}(\zeta[k]|\mathbf{p})}{\partial \mathbf{p}_j} \Big|_{\zeta=\zeta[k]} = \sum_{j=1}^{n_s} (-1)^{1-\zeta_j[k]} \prod_{\substack{i=1 \\ i \neq j}}^{n_s} \mathbf{p}_i^{\zeta_i[k]} (1 - \mathbf{p}_i)^{1-\zeta_i[k]} \mathbf{e}_j.$$

In Appendix A, we provide a detailed discussion of the multivariate Bernoulli distribution, with identities and lemmas that will be useful for the following discussion. Clearly, (3.4) and (3.5) are not practical formula, because their statement alone requires the evaluation of all possible designs  $\mathcal{J}(\zeta[k])$ , which is equivalent to complete enumeration of (2.5). We show below how we can utilize stochastic gradient approaches to avoid complete enumeration. A practical and efficient solution of (3.2) is discussed in subsection 3.3. We first establish in subsection 3.2 the connection between the solution of the two problems (2.5) and (3.2) and discuss the benefits of the proposed approach.

**3.2. Benefits of the stochastic formulation.** The connections between (2.5) and (3.2) and their respective solutions are summarized by Proposition 3.1 and Lemma 3.2. The relation between the objective functions and their domains and codomains is sketched in Figure 1.

**PROPOSITION 3.1.** *Consider the two functions  $\mathcal{J} : \Omega_{\zeta} := \{0, 1\}^{n_s} \rightarrow \mathbb{R}$  and  $\Upsilon : \Omega_{\mathbf{p}} := [0, 1]^{n_s} \rightarrow \mathbb{R}$ , defined in (2.5) and (3.2), respectively. Let  $\zeta \in \Omega_{\zeta}$  be a random vector following the multivariate Bernoulli distribution (3.1), with parameter  $\mathbf{p} \in \Omega_{\mathbf{p}}$ . Let  $C := \{c \in \mathbb{R} \mid c := \mathcal{J}(\zeta), \zeta \in \Omega_{\zeta}\}$  be the set of objective values corresponding to all possible designs,  $\zeta \in \Omega_{\zeta}$ . Then the following hold:*

- (a)  $\Omega_{\mathbf{p}}$  is the convex hull of  $\Omega_{\zeta}$  in  $\mathbb{R}^{n_s}$ .
- (b) The values of  $\Upsilon$  (3.2) form the convex hull of  $C$  in  $\mathbb{R}$ , denoted by  $\text{conv}(C)$ , with points  $c \in C$  being the extreme points of  $\text{conv}(C)$ ; that is,  $\text{conv}(C) \equiv \{\Upsilon(\mathbf{p}) \mid \mathbf{p} \in \Omega_{\mathbf{p}}\}$ .

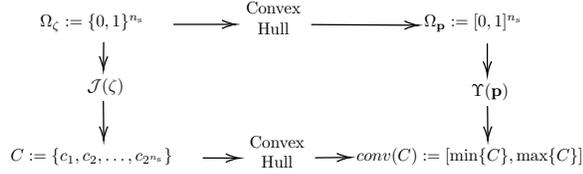


FIG. 1. Relation between problems 2.5 and 3.2 and their respective domains and codomains.

- (c)  $\Upsilon(\mathbf{x}) = \mathcal{J}(\mathbf{x}) \forall \mathbf{x} \in \Omega_\zeta$ .
- (d) For any realization of  $\mathbf{p} \in \Omega_{\mathbf{p}}$ , it holds that  $\min\{C\} \leq \Upsilon(\mathbf{p}) \leq \max\{C\}$ , and  $\Upsilon(\mathbf{p}^{\text{opt}}) = \min\{C\} = \mathcal{J}(\zeta^{\text{opt}})$ .

*Proof.* (a) This follows from the definition of  $\Omega_\zeta$  and  $\Omega_{\mathbf{p}}$  and the fact that  $\Omega_{\mathbf{p}}$  is the hypercube in  $\mathbb{R}^{n_s}$  whose vertices formulate the set  $\Omega_\zeta$ .  
 (b) For any realization  $\mathbf{p}$ , the function  $\Upsilon$  defines a convex combination of all points in  $C$ , because the coefficients  $\mathbb{P}(\zeta|\mathbf{p})$ ,  $\zeta \in \Omega_\zeta$  are probabilities satisfying that  $0 \leq \mathbb{P}(\zeta|\mathbf{p}) \leq 1$  and  $\sum_{\zeta \in \Omega_\zeta} \mathbb{P}(\zeta|\mathbf{p}) = 1$ .  
 (c) Setting  $\mathbf{p}$  to any  $\mathbf{x} \in \Omega_\zeta$  yields a degenerate multivariate Bernoulli distribution, which is a Dirac measure  $\delta_\zeta(\{\mathbf{x}\})$  defined on the set  $\{\mathbf{x}\}$ . Thus,  $\mathbb{P}(\zeta|\mathbf{p} = \mathbf{x}) = \delta_\zeta(\{\mathbf{x}\})$ , and  $\Upsilon(\mathbf{x}) = \sum_{\zeta \in \Omega_{\mathbf{p}}} \mathbb{P}(\zeta|\mathbf{p} = \mathbf{x}) \mathcal{J}(\zeta) = \sum_{\zeta \in \Omega_{\mathbf{p}}} \delta_\zeta(\{\mathbf{x}\}) \mathcal{J}(\zeta) = \mathcal{J}(\mathbf{x})$ .  
 (d) It follows from Carathéodory’s theorem that for any realization of  $\mathbf{p} \in \Omega_{\mathbf{p}}$ , the value of the objective  $\Upsilon(\mathbf{p})$  falls on a line segment in  $\mathbb{R}$  that connects at most two points in  $C$ . This guarantees that  $\min\{C\} \leq \Upsilon(\mathbf{p}) \leq \max\{C\}$ . Additionally, from points (a) and (b) above, it follows that  $\Upsilon(\mathbf{p}^{\text{opt}}) = \min\{\text{conv}(C)\} = \min\{C\} = \mathcal{J}(\zeta^{\text{opt}})$ .  $\square$

LEMMA 3.2. The optimal solutions of the two problems (2.5) and (3.2) are such that

$$\arg \min_{\zeta \in \Omega_\zeta} \mathcal{J}(\zeta) \subseteq \arg \min_{\mathbf{p} \in \Omega_{\mathbf{p}}} \Upsilon(\mathbf{p}).$$

Moreover, if the optimal solution  $\zeta^{\text{opt}}$  of (2.5) is unique, then  $\mathbf{p}^{\text{opt}} = \zeta^{\text{opt}}$ , where  $\mathbf{p}^{\text{opt}}$  is the unique optimal solution of (3.2).

*Proof.* Proposition 3.1 guarantees that  $\nexists \mathbf{p} \in \Omega_{\mathbf{p}}$ ,  $\Upsilon(\mathbf{p}) < \mathcal{J}(\zeta^{\text{opt}})$ . Additionally, from points (a) and (c) in Proposition 3.1, it follows that  $\arg \min_{\zeta \in \Omega_\zeta} \mathcal{J}(\zeta) \subseteq \arg \min_{\mathbf{p} \in \Omega_{\mathbf{p}}} \Upsilon(\mathbf{p})$ .

Assume  $\zeta^{\text{opt}}$  is the unique optimal solution of (2.5); it follows that  $\mathbf{p}^{\text{opt}} = \zeta^{\text{opt}}$  is an optimal solution of (3.2). Now assume  $\exists \mathbf{p} \in \Omega_{\mathbf{p}}$  such that  $\mathbf{p} \neq \mathbf{p}^{\text{opt}}$ ,  $\Upsilon(\mathbf{p}) = \Upsilon(\mathbf{p}^{\text{opt}}) = \mathcal{J}(\zeta^{\text{opt}})$ . If  $\mathbf{p}$  is the parameter of a degenerate Bernoulli distribution, then  $\mathbf{p} \in \Omega_\zeta$  and  $\mathcal{J}(\mathbf{p}) = \mathcal{J}(\zeta^{\text{opt}})$ , thus contradicting the uniqueness of  $\zeta^{\text{opt}}$ . Conversely, if  $\mathbf{p}$  is the parameter of a nondegenerate multivariate Bernoulli distribution, then there are at least two designs  $\zeta, \eta \in \Omega_\zeta$  with nontrivial probabilities  $\in (0, 1)$ , with  $\mathcal{J}(\zeta) = \mathcal{J}(\eta)$ , again contradicting the uniqueness assumption of  $\zeta^{\text{opt}}$ .  $\square$

The stochastic formalism proposed in subsection 3.1 is beneficial for multiple reasons. First, this probabilistic formulation (3.2) to solving the original optimization problem (2.5) enables converting a binary design domain into a bounded continuous domain, where the optimal solution of the two problems (3.2) and (2.5) coincide. Second, the stochastic formulation (3.2) enables utilizing efficient stochastic optimization algorithms to solve binary optimization problems, which are not applicable for the

deterministic binary optimization problem (2.5). Third, the solution of the stochastic problem (3.2) is an optimal parameter  $\mathbf{p}^{\text{opt}}$  that can be used for sampling binary designs  $\zeta$  by sampling  $\mathbb{P}(\zeta|\mathbf{p}^{\text{opt}})$ , even if only a suboptimal solution is found. Fourth, the stochastic formulation (3.2) requires evaluating the derivative of  $\Upsilon$  with respect to the parameter of the probability distribution  $\mathbf{p}$ , instead of the design  $\zeta$ . Thus, we do not need to evaluate the derivative of  $\mathcal{J}$  with respect to the design  $\zeta$ . In OED for Bayesian inversion, as discussed in section 2, evaluating the gradient (with respect to the design) is very expensive since it requires many forward and backward solves of expensive simulation models. Fifth, because  $\mathcal{J}$  is not required to be differentiable with respect to the design  $\zeta$ , a nonsmooth penalty function  $\Phi$  can be utilized in defining  $\mathcal{J}$ , for example, to enforce sparsity or budget constraints, with the need to approximate this penalty with a smoother penalty approximation. Note that in this case, as explained by (2.5) and (3.2), the penalty is asserted on the design, rather than the distribution parameter  $\mathbf{p}$ .

**3.3. Approximately solving the stochastic optimization problem.** The objective in (3.4) and the gradient (3.5) amount to evaluating a large finite sum, which can be approximated by sampling. The simplest approach for approximately solving (3.4) is to follow an MC sample-based approximation to the objective  $\Upsilon$  and the gradient  $\nabla_{\mathbf{p}}\Upsilon$ , where the sample is drawn from an appropriate probability distribution. However, the objective (3.2) is deterministic with respect to  $\mathbf{p}$ . Nevertheless, one can add randomness by assuming a uniform random variable (an index) over all terms of the objective and the associated gradient. This allows utilizing external and internal sampling-based stochastic optimization algorithms; see, for example, [20, 21, 31, 32, 35, 42, 46, 47]. These algorithms, however, are beyond the scope of this work.

Here, instead, we focus on efficient algorithmic procedures inspired by recent developments in reinforcement learning [15, 50, 54]. Specifically, we view  $\mathbb{P}(\zeta|\mathbf{p})$  as a policy that describes the probability of activating each sensor: The design  $\zeta$  is viewed as the state, and the action entails altering the status (turn on/off) of any combination of sensors given an initial policy. Here, in RL terminology, one action is allowed to reach the optimal policy from the initial policy, that is, only one time step is assumed. Our goal is to find an optimal policy to configure the observational grid. The reward is prescribed by the value of the original objective function, and we seek the optimal design that maximizes the total expected reward. Note that we can replace the maximization with minimization of the negative of the objective function, thus enabling mapping between RL terminology and various OED objectives such as A-optimality. This is further explained below.

An alternative form of the gradient (3.5) can be obtained by using the “kernel trick,” which utilizes the fact that  $\nabla_{\mathbf{p}}\log(\mathbb{P}(\zeta|\mathbf{p})) = \frac{1}{\mathbb{P}(\zeta|\mathbf{p})}\nabla_{\mathbf{p}}\mathbb{P}(\zeta|\mathbf{p})$ , and thus  $\nabla_{\mathbf{p}}\mathbb{P}(\zeta|\mathbf{p}) = \mathbb{P}(\zeta|\mathbf{p})\nabla_{\mathbf{p}}\log(\mathbb{P}(\zeta|\mathbf{p}))$ . Using this identity, we can rewrite the gradient (3.5) as

$$(3.7) \quad \nabla_{\mathbf{p}}\Upsilon(\mathbf{p}) = \sum_{k=1}^{2^{n_s}} \left( \mathcal{J}(\zeta[k]) \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta[k]|\mathbf{p}) \right) \mathbb{P}(\zeta[k]|\mathbf{p}) = \mathbb{E}_{\zeta \sim \mathbb{P}(\zeta|\mathbf{p})} [\mathcal{J}(\zeta) \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})].$$

We assume, without loss of generality, that  $\mathbf{p}$  falls inside the open ball  $(0, 1)^{n_s}$ , and thus both the logarithm and the associated derivative of the log-probability are well defined. If any of the components of  $\mathbf{p}$  attain their bound, then the distribution

becomes degenerate in this direction, and the probability is set to either 0 or 1 and is thus taken out of the formulation since the gradient in that direction is set to 0. This is equivalent to projecting  $\mathbf{p}$  onto a lower-dimensional subspace.

The form of the gradient described by (3.7) is equivalent to (3.5). However, it shows that the gradient can be written as an expectation of gradients. This enables us to approximate the gradient using MC sampling by following a stochastic optimization approach. Specifically, given a sample  $\zeta[j] \sim \mathbb{P}(\zeta|\mathbf{p})$ ,  $j = 1, 2, \dots, N_{\text{ens}}$ , we can use the following MC approximation of the gradient,

$$(3.8) \quad \mathbf{g} := \nabla_{\mathbf{p}} \mathbb{E}_{\mathbb{P}(\zeta|\mathbf{p})}[\mathcal{J}(\zeta)] \approx \widehat{\mathbf{g}} := \frac{1}{N_{\text{ens}}} \sum_{j=1}^{N_{\text{ens}}} \mathcal{J}(\zeta[j]) \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta[j]|\mathbf{p}),$$

where  $\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta[j]|\mathbf{p})$  is the score function of the multivariate Bernoulli distribution (see Appendix A). By combining (3.8) with (A.7), we obtain the following form of the stochastic gradient:

$$(3.9) \quad \widehat{\mathbf{g}} = \frac{1}{N_{\text{ens}}} \sum_{j=1}^{N_{\text{ens}}} \mathcal{J}(\zeta[j]) \sum_{i=1}^{n_s} \left( \frac{\zeta_i[j]}{\mathbf{p}_i} + \frac{\zeta_i[j] - 1}{1 - \mathbf{p}_i} \right) \mathbf{e}_i.$$

To solve the optimization problem (3.2), one can start with an initial policy, in other words, an initial set of parameters  $\mathbf{p}$ , and iteratively follow an approximate descent direction until an approximately *optimal policy* is obtained. Because the optimization problems described here require box constraints, we introduce the projection operator  $P$ , which maps an arbitrary point  $\mathbf{p}$  onto the feasible region described by the box constraints. One can apply projection by truncation (see [38, section 16.7]):

$$(3.10) \quad P(\mathbf{p}_i) := \min\{1, \max\{0, \mathbf{p}_i\}\} \equiv \begin{cases} 0 & \text{if } \mathbf{p}_i < 0, \\ \mathbf{p}_i & \text{if } \mathbf{p}_i \in [0, 1], \\ 1 & \text{if } \mathbf{p}_i > 1, \end{cases}$$

where the projection  $P(\mathbf{p})$  is applied componentwise to the parameters' vector  $\mathbf{p}$ . The projection operator  $P$  is nonexpansive and thus is orthogonal, guaranteeing that for any  $P(\mathbf{p}[1]), P(\mathbf{p}[2]) \in \mathbb{R}^{n_s}$ :  $\|P(\mathbf{p}[1]) - P(\mathbf{p}[2])\| \leq \|\mathbf{p}[1] - \mathbf{p}[2]\|$ . A stochastic steepest-descent step to solve (3.2) is approximated by the stochastic approximation of the gradient (3.9) and is described as

$$(3.11) \quad \mathbf{p}^{(n+1)} = P\left(\mathbf{p}^{(n)} - \eta^{(n)} \widehat{\mathbf{g}}^{(n)}\right),$$

where  $\eta^{(n)}$  is the step size (learning rate) at the  $n$ th iteration. Equation (3.11) describes a stochastic gradient descent approach, using the stochastic approximation (3.9) at each iteration in place of  $g^{(n)}$ , where the sample is generated from  $\mathbb{P}(\zeta|\mathbf{p}^{(n)})$ . One can choose a fixed step size or follow a decreasing schedule to guarantee convergence or can do a line search using the sampled design to approximate the objective function. One can also use approximate second-order information and create a quasi-Newton-like step.

Algorithm 3.1 describes a stochastic descent approach for solving (2.5) by solving (3.2) followed by a sampling step. As  $\zeta$  is sampled from a multivariate Bernoulli distribution with parameter  $\mathbf{p}$ , in step 5 if  $\mathbf{p}_i \in \{0, 1\}$ , then  $\zeta_i = \mathbf{p}_i$ . Thus the corresponding term in the summation vanishes.

Algorithm 3.1 is a stochastic gradient descent method with convergence guarantees in expectation only; see subsection 3.4. In practice, the convergence test could

**Algorithm 3.1** Stochastic optimization algorithm for binary OED.

**Input:** Initial distribution parameter  $\mathbf{p}^{(0)}$ , step size schedule  $\eta^{(n)}$ ; sample sizes  $N_{\text{ens}}, m$

**Output:**  $\zeta^{\text{opt}}$

- 1: Initialize  $n = 0$
- 2: **while** Not Converged **do**
- 3:     Update  $n \leftarrow n + 1$
- 4:     Sample  $\{\zeta[i]; i = 1, 2, \dots, N_{\text{ens}}\} \sim \mathbb{P}(\zeta|\mathbf{p}^{(n)})$
- 5:     Calculate  $\hat{g}^{(n)} = \frac{1}{N_{\text{ens}}} \sum_{j=1}^{N_{\text{ens}}} (\mathcal{J}(\zeta[j])) \sum_{i=1}^{n_s} \left( \frac{\zeta_i[j]}{\mathbf{p}_i} + \frac{\zeta_i[j]-1}{1-\mathbf{p}_i} \right) \mathbf{e}_i$
- 6:     Update  $\mathbf{p}^{(n+1)} = P(\mathbf{p}^{(n)} - \eta^{(n)}\hat{g}^{(n)})$
- 7: **end while**
- 8: Set  $\mathbf{p}^{\text{opt}} = \mathbf{p}^{(n)}$
- 9: Sample  $\{\zeta[j]; j = 1, 2, \dots, m\} \sim \mathbb{P}(\zeta|\mathbf{p}^{\text{opt}})$ , and calculate  $\mathcal{J}(\zeta[j])$
- 10: **return**  $\zeta^{\text{opt}}$ : the design  $\zeta$  with smallest value of  $\mathcal{J}$  in the sample.

be replaced with a maximum number of iterations. Doing so, however, might not result in degenerate PMF. The output of the algorithm  $\mathbf{p}^{\text{opt}} = \mathbf{p}^{(a)}$  is used for sampling binary designs  $\zeta$ , by sampling  $\mathbb{P}(\zeta|\mathbf{p}^{\text{opt}})$ , even if a suboptimal solution is found. Another potential convergence criterion is the magnitude of the projected gradient  $\|P(g^{(k)})\| \leq \text{pgtol}$ . We also note that Algorithm 3.1 is equivalent to the vanilla policy gradient REINFORCE algorithm [54]. The remainder of section 3 is devoted to addressing convergence guarantees, convergence analysis, and improvements of Algorithm 3.1.

**3.4. Convergence analysis.** Here, we show that the optimal solution of (3.4) is a degenerate multivariate Bernoulli distribution and that the convergence of Algorithm 3.1 to such an optimal distribution in expectation is guaranteed under mild conditions. First, we study the properties of the exact objective defined by (3.4) and the associated exact gradient (3.5). We then address the stochastic approximation and the performance of Algorithm 3.1. For clarity of the discussion, the proofs of the lemmas discussed in subsections 3.4 and 3.5 are provided in Appendix B.

**3.4.1. Analysis of the exact stochastic optimization problem.** Recall that the objective function utilized in (3.4) and the associated gradient take the respective forms

$$(3.12) \quad \Upsilon(\mathbf{p}) = \sum_{k=1}^{2^{n_s}} \mathcal{J}(\zeta[k])\mathbb{P}(\zeta[k]|\mathbf{p}) ; \quad \nabla_{\mathbf{p}} \Upsilon(\mathbf{p}) \equiv \mathbf{g} = \sum_{k=1}^{2^{n_s}} \mathcal{J}(\zeta[k])\nabla_{\mathbf{p}} \mathbb{P}(\zeta[k]|\mathbf{p}) ,$$

where  $\mathbf{p} \in \Omega_{\mathbf{p}} := [0, 1]^{n_s}$ ,  $\mathbb{P}(\zeta|\mathbf{p})$  is the multivariate Bernoulli distribution (3.1), and we use the indexing scheme (3.3). We are viewing the objective  $\Upsilon$  explicitly as a function of the PMF parameter  $\mathbf{p}$  because all combinations of binary designs  $\zeta[k]; k = 1, 2, \dots, 2^{n_s}$  are present in the expectation.

Here, we show that the exact gradient of  $\Upsilon$  is bounded (Theorem 3.3) and that the Hessian of  $\Upsilon$  is bounded; hence, by following a steepest-descent approach, a locally optimal design is obtained. This will set the ground for an analysis of Algorithm 3.1.

**THEOREM 3.3.** *Let  $\zeta \in \Omega_{\zeta} = \{0, 1\}^{n_s}$ , and let  $C = \max_{\zeta \in \Omega_{\zeta}} \{|\mathcal{J}(\zeta)|\}$ . Then the following bounds of the gradient of the stochastic objective  $\Upsilon$  hold,*

$$(3.13a) \quad \|\nabla_{\mathbf{p}} \Upsilon(\mathbf{p})\| \leq \sqrt{n_s |\Omega_\zeta|} C, \quad \mathbf{p} \in \Omega_{\mathbf{p}},$$

$$(3.13b) \quad \|\nabla_{\mathbf{p}} \Upsilon(\mathbf{p}[1]) - \nabla_{\mathbf{p}} \Upsilon(\mathbf{p}[2])\| \leq 2\sqrt{n_s |\Omega_\zeta|} C, \quad \mathbf{p}[1], \mathbf{p}[2] \in \Omega_{\mathbf{p}},$$

where  $|\Omega_\zeta| = 2^{n_s}$  is the cardinality of the design domain. Moreover, the Hessian is bounded by

$$(3.14) \quad \left| \frac{\partial^2 \Upsilon}{\partial \mathbf{p}_i \partial \mathbf{p}_j} \right| \leq \sqrt{|\Omega_\zeta|} C.$$

*Proof.* See Appendix B. □

Theorem 3.3 is especially interesting because it shows that the gradient of  $\Upsilon(\mathbf{p})$  is bounded. On the contrary,  $\mathcal{J}(\zeta)$  does not have bounded gradients for  $\ell_0$  regularization. Additionally, it follows from Theorem 3.3 that the entries of the Hessian are bounded, implying that the Hessian is bounded. Thus the objective  $\Upsilon$  is Lipschitz smooth. Finding the Lipschitz constant, however, depends on the value of the function  $\mathcal{J}$  evaluated at all possible values of  $\zeta$ , or at least the maximum value. This is impossible to know a priori, without any prior knowledge about the problem in hand. Moreover, the bound given above indicates that the Lipschitz constant decreases for higher dimensions, since it is exponentially proportional to the cardinality of the design space  $\Omega_\zeta$ . However, one can estimate the Hessian matrix by using an ensemble of realizations and use it to estimate the Lipschitz constant. This can be helpful for choosing a proper step size for a steepest-descent algorithm. Alternatively, one can use a decreasing step size sequence  $\{\eta^{(k)}\}_{k=0}^\infty$  such that  $\lim_{k \rightarrow \infty} \eta^{(k)} = 0$  and  $\sum_k \eta^{(k)} = \infty$ , which guarantees convergence to a local optimum; see [15, Proposition 4.1].

**3.4.2. Analysis of stochastic steepest-descent algorithm.** Here we show that  $\widehat{\mathbf{g}}(\mathbf{p})$  is an unbiased estimator of the true gradient  $\mathbf{g}(\mathbf{p})$ . This fact, along with Theorem 3.3 and the fact that  $\Upsilon$  is a convex combination, guarantees convergence, in expectation, of Algorithm 3.1 to a locally optimal policy. At each iteration  $n$  of Algorithm 3.1, the gradient is approximated with a sample from the respective conditional distribution  $\mathbb{P}(\zeta|\mathbf{p}^{(n)})$ . First, we note that

$$(3.15) \quad \|\widehat{\mathbf{g}}\| = \left\| \frac{1}{N_{\text{ens}}} \sum_{j=1}^{N_{\text{ens}}} \mathcal{J}(\zeta[j]) \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta[j]|\mathbf{p}) \right\| \leq \frac{1}{N_{\text{ens}}} \sum_{j=1}^{N_{\text{ens}}} |\mathcal{J}(\zeta[j])| \|\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta[j]|\mathbf{p})\|,$$

which shows that the magnitude of  $\widehat{\mathbf{g}}$  is bounded. Moreover, we will show next that  $\widehat{\mathbf{g}}(\mathbf{p})$  is an unbiased estimator of  $\mathbf{g}(\mathbf{p})$  and that the variance of this estimator is bounded.

LEMMA 3.4. *The stochastic estimator  $\widehat{\mathbf{g}}$  defined by (3.8) is unbiased, with sampling total variance  $\text{var}(\widehat{\mathbf{g}})$ , such that*

$$(3.16) \quad \mathbb{E}[\widehat{\mathbf{g}}] = \mathbf{g} = \nabla_{\mathbf{p}} \Upsilon(\mathbf{p}); \quad \text{var}(\widehat{\mathbf{g}}) = \frac{1}{N_{\text{ens}}} \text{var}(\mathcal{J}(\zeta) \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})),$$

where the total variance operator evaluates the trace of the variance-covariance matrix of the random vector. Moreover, the total variance of  $\widehat{\mathbf{g}}$  is bounded, and there exist some positive constants  $K_1, K_2$ , such that

$$(3.17) \quad \mathbb{E}[\widehat{\mathbf{g}}^T \widehat{\mathbf{g}}] = \mathbb{E}\|\widehat{\mathbf{g}}\|^2 \leq K_1 + K_2 \|\mathbf{g}\|^2 = K_1 + K_2 \mathbf{g}^T \mathbf{g}.$$

*Proof.* See Appendix B.  $\square$

The significance of Lemma 3.4 is that (3.17) guarantees that Assumption (d) of [15, Assumptions 4.2] is satisfied. This, along with the fact that  $\hat{\mathbf{g}}$  is unbiased, and given the boundedness of entries of the Hessian (Theorem 3.3), and by complying with the step size requirement, guarantees the convergence of Algorithm 3.1 to a locally optimal policy  $\mathbf{p}^{\text{opt}}$ .

The sample-based approximation of the gradient described by (3.9) exhibits high variance, however, and thus requires a prohibitively large number of samples in order to achieve acceptable convergence behavior to a locally optimal policy. Alternatively, one can use importance sampling [7] or antithetic variates [34] or can add a baseline to the objective, in order to reduce the variability of the estimator. This issue is discussed next (subsection 3.5).

**3.5. Variance reduction: Introducing the baseline.** The formulation of the stochastic estimator  $\hat{\mathbf{g}}$  defined by (3.8) provides limited control over its variability. The implication of Lemma 3.4 is that, while the gradient estimator is unbiased, large samples are required to reduce its variability. Specifically, it follows from (3.16) that  $\text{var}(\hat{g}) = \mathcal{O}(N_{\text{ens}}^{-1})$ ; thus, to reduce the variability of the estimator, one would need to increase the sample size. In general, MC estimators are known to suffer from high variance; thus this estimator, while unbiased, will be impractical especially if  $n_s$  is large and the sample size  $N_{\text{ens}}$  is small. Instead of increasing the sample size, one can reduce the estimator variability by introducing a *baseline*  $b$  to the objective function. The objective function  $\Upsilon$  can be replaced with the following baseline version,

$$(3.18) \quad \Upsilon^b(\mathbf{p}) := \mathbb{E}_{\zeta \sim \mathbb{P}(\zeta|\mathbf{p})}[\mathcal{J}(\zeta) - b],$$

where  $b$  is a baseline assumed to be independent from the parameter  $\mathbf{p}$ . Since  $b$  is independent from  $\Upsilon$  and by linearity of the expectation, it follows that  $\Upsilon^b(\mathbf{p}) = \mathbb{E}_{\zeta \sim \mathbb{P}(\zeta|\mathbf{p})}[\mathcal{J}(\zeta)] - b$ , and thus  $\arg \min_{\mathbf{p}} \Upsilon^b = \arg \min_{\mathbf{p}} \Upsilon$ , and  $\nabla_{\mathbf{p}} \Upsilon^b = \nabla_{\mathbf{p}} \Upsilon - \nabla_{\mathbf{p}} b = \nabla_{\mathbf{p}} \Upsilon$ . By applying the kernel trick again, we can write the gradient of (3.18) as

$$(3.19) \quad \nabla_{\mathbf{p}} \Upsilon^b(\mathbf{p}) = \mathbb{E}_{\zeta \sim \mathbb{P}(\zeta|\mathbf{p})}[(\mathcal{J}(\zeta) - b) \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})],$$

which can be approximated by the sample estimator

$$(3.20) \quad \hat{\mathbf{g}}^b := \frac{1}{N_{\text{ens}}} \sum_{j=1}^{N_{\text{ens}}} (\mathcal{J}(\zeta[j]) - b) \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta[j]|\mathbf{p}) = \hat{\mathbf{g}} - \frac{b}{N_{\text{ens}}} \sum_{j=1}^{N_{\text{ens}}} \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta[j]|\mathbf{p}).$$

Note that (3.19) is also an unbiased estimator since  $\mathbb{E}[\hat{\mathbf{g}}^b] = \mathbb{E}[\hat{\mathbf{g}}] = \mathbf{g}$ . The variance of such an estimator, however, can be controlled by choosing an adequate baseline  $b$ . In subsection 3.5.1, we provide guidance for choosing the baseline. In what follows, however, we keep the baseline as a user-defined parameter, opening the door for other choices.

**3.5.1. On the choice of the baseline.** We define the optimal baseline to be the one that minimizes the variability in the gradient estimator. We provide the following results, which will help us obtain an optimal baseline.

LEMMA 3.5. Let  $\mathbf{d} = \frac{1}{N_{\text{ens}}} \sum_{j=1}^{N_{\text{ens}}} \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta[j]|\mathbf{p})$ . Then the following identities hold:

$$(3.21) \quad \mathbb{E}[\mathbf{d}] = 0; \quad \text{var}(\mathbf{d}) = \frac{1}{N_{\text{ens}}} \sum_{i=1}^{n_s} \frac{1}{\mathbf{p}_i - \mathbf{p}_i^2}.$$

*Proof.* See Appendix B. □

LEMMA 3.6. The ensemble estimator  $\hat{\mathbf{g}}^b$  described by (3.20) is unbiased, with sampling total variance  $\text{var}(\hat{\mathbf{g}}^b)$ , such that

$$(3.22) \quad \mathbb{E}[\hat{\mathbf{g}}^b] = \mathbf{g} = \nabla_{\mathbf{p}} \Upsilon(\mathbf{p}); \quad \text{var}(\hat{\mathbf{g}}^b) = \text{var}(\hat{\mathbf{g}}) - 2b\mathbb{E}[\hat{\mathbf{g}}^T \mathbf{d}] + \frac{b^2}{N_{\text{ens}}} \sum_{i=1}^{n_s} \frac{1}{\mathbf{p}_i - \mathbf{p}_i^2}.$$

*Proof.* See Appendix B. □

The variance of  $\hat{\mathbf{g}}^b$  is described by Lemma 3.6. We can view (B.15) as a quadratic expression in  $b$  and minimize it over  $b$ . Because  $\sum_{i=1}^{n_s} \frac{1}{\mathbf{p}_i - \mathbf{p}_i^2} > 0$ , the quadratic is convex, and the min in  $b$  is obtained by equating the derivative of the estimator variance (B.15) to zero, which yields the following optimal baseline:

$$(3.23) \quad b^{\text{opt}} = \frac{N_{\text{ens}}}{\sum_{i=1}^{n_s} \frac{1}{\mathbf{p}_i - \mathbf{p}_i^2}} \mathbb{E}[\hat{\mathbf{g}}^T \mathbf{d}].$$

The expectation in (3.23), however, depends on the value of the function  $\Upsilon$  and can be estimated by an ensemble of realizations  $\hat{\mathbf{g}}[j], \mathbf{d}[j], j = 1, 2, \dots, m$ ,

$$(3.24) \quad \mathbb{E}[\hat{\mathbf{g}}^T \mathbf{d}] \approx \frac{1}{b_m} \sum_{e=1}^{b_m} \hat{\mathbf{g}}[e]^T \mathbf{d}[e],$$

where  $\mathbf{g}[e]$  and  $\mathbf{d}[e]$  are realizations of  $\mathbf{g}$  and  $\mathbf{d}$ , respectively. Thus, we propose to estimate the optimal baseline  $b^{\text{opt}}$  as follows. Given a realization  $\mathbf{p}$  of the hyperparameter, a set of  $b_m$  batches each of size  $N_{\text{ens}}$  are sampled from  $\mathbb{P}(\zeta|\mathbf{p})$ , resulting in the multivariate Bernoulli samples  $\{\zeta[e, j]; e = 1, 2, \dots, b_m; j = 1, 2, \dots, N_{\text{ens}}\}$ . The following function is then used to estimate  $b^{\text{opt}}$ :

$$(3.25) \quad b^{\text{opt}} \approx \hat{b}^{\text{opt}} := \frac{\sum_{e=1}^{b_m} \left( \sum_{j=1}^{N_{\text{ens}}} \mathcal{J}(\zeta[e, j]) \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta[e, j]|\mathbf{p}) \right)^T \left( \sum_{j=1}^{N_{\text{ens}}} \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta[e, j]|\mathbf{p}) \right)}{N_{\text{ens}} b_m \sum_{i=1}^{n_s} \frac{1}{\mathbf{p}_i - \mathbf{p}_i^2}}.$$

**3.5.2. Complete algorithm statement.** We conclude this section with an algorithmic description of the stochastic steepest-descent algorithm with the optimal baseline suggested here. Algorithm 3.2 is a modification of Algorithm 3.1, where we only added the baseline (3.25).

Note that in both Algorithms 3.1 and 3.2, the value of  $\mathcal{J}$  is evaluated repeatedly at instances of the binary design  $\zeta$ . With the algorithm progressing, it becomes more likely to revisit previously sampled designs. One should keep track of the sampled designs and the corresponding value of  $\mathcal{J}$ , for example, by utilizing the indexing scheme (3.3), to prevent redundant computations. We remark that as noted in Algorithm 3.1, if  $\mathbf{p}_i \in \{0, 1\}$  in step 6 or step 7 of Algorithm 3.2, then  $\zeta_i = \mathbf{p}_i$ . Thus the corresponding term in the summation vanishes.

---

**Algorithm 3.2** Stochastic optimization for binary OED with the optimal baseline.

---

**Input:** Initial distribution parameter  $\mathbf{p}^{(0)}$ , step size schedule  $\eta^{(n)}$ , sample sizes  $N_{\text{ens}}$ ,  $m$ , baseline batch size  $b_m$

**Output:**  $\zeta^{\text{opt}}$

- 1: initialize  $n = 0$
- 2: **while** Not Converged **do**
- 3:   Update  $n \leftarrow n + 1$
- 4:   Sample  $\{\zeta[j]; j = 1, 2, \dots, N_{\text{ens}}\} \sim \mathbb{P}(\zeta|\mathbf{p}^{(n)})$
- 5:   Calculate  $b = \text{OPTIMALBASELINE}(\mathbf{p}^{(n)}, N_{\text{ens}}, b_m)$
- 6:   Calculate  $\mathbf{g}^{(n)} = \frac{1}{N_{\text{ens}}} \sum_{j=1}^{N_{\text{ens}}} (\mathcal{J}(\zeta[j]) - b) \sum_{i=1}^{n_s} \left( \frac{\zeta_i[j]}{\mathbf{p}_i} + \frac{\zeta_i[j]-1}{1-\mathbf{p}_i} \right) \mathbf{e}_i$
- 7:   Update  $\mathbf{p}^{(n+1)} = P(\mathbf{p}^{(n)} - \eta^{(n)} \mathbf{g}^{(n)})$
- 8: **end while**
- 9: Set  $\mathbf{p}^{\text{opt}} = \mathbf{p}^{(n)}$
- 10: Sample  $\{\zeta[j]; j = 1, 2, \dots, m\} \sim \mathbb{P}(\zeta|\mathbf{p}^{\text{opt}})$ , and calculate  $\mathcal{J}(\zeta[j])$
- 11: **return**  $\zeta^{\text{opt}}$ : the design  $\zeta$  with smallest value of  $\mathcal{J}$  in the sample.
- 12: **function** OPTIMALBASELINE( $\theta, N_{\text{ens}}, b_m$ )
- 13:   Initialize  $b \leftarrow 0$
- 14:   **for**  $e \leftarrow 1$  to  $b_m$  **do**
- 15:     **for**  $j \leftarrow 1$  to  $N_{\text{ens}}$  **do**
- 16:       Sample  $\zeta[j] \sim \mathbb{P}(\zeta|\mathbf{p})$
- 17:       Calculate  $\mathbf{r}[j] = \sum_{i=1}^{n_s} \left( \frac{\zeta_i[j]}{\mathbf{p}_i} + \frac{\zeta_i[j]-1}{1-\mathbf{p}_i} \right) \mathbf{e}_i$
- 18:     **end for**
- 19:     Calculate  $\mathbf{d}[e] = \frac{1}{N_{\text{ens}}} \sum_{j=1}^{N_{\text{ens}}} \mathbf{r}[j]$ , and  $\mathbf{g}[e] = \frac{1}{N_{\text{ens}}} \sum_{j=1}^{N_{\text{ens}}} \mathcal{J}(\zeta[j]) \mathbf{r}[j]$
- 20:     Update  $b \leftarrow b + (\mathbf{g}[e])^\top \mathbf{d}[e]$
- 21:   **end for**
- 22:   Update  $b \leftarrow b N_{\text{ens}} / \left( b_m \sum_{i=1}^{n_s} \frac{1}{\mathbf{p}_i - \mathbf{p}_i^2} \right)$
- 23:   **return**  $b$
- 24: **end function**

---

**3.6. Computational considerations.** Here, we discuss the computational cost of the proposed algorithms and of standard OED approaches in terms of the number of forward model evaluations. We assume  $\mathcal{J}$  is set to the A-optimality criterion, that is, the trace of the posterior covariance of the inversion parameter. This discussion extends easily to other OED optimality criteria.

Standard OED approaches require solving the relaxed OED problem (2.6), which requires evaluating the gradient of the objective  $\mathcal{J}$ , namely, the optimality criterion, with respect to the relaxed design, in addition to evaluating the objective itself  $\mathcal{J}$  for line-search optimization. Formulating the gradient requires one Hessian solve and a forward integration of the model  $\mathbf{F}$  for each entry of the gradient. The Hessian, being the inverse of the posterior covariance of the inversion parameter in the linear (or the linearized) inverse problem, is a function of the relaxed design; see, for example, [9, 17] for details. Hessian solves can be done by using a preconditioned conjugate gradient (CG) method. Each application of the Hessian requires a (linear/linearized) forward and an adjoint model evaluation. If the prior covariance is employed as a preconditioner, and assuming  $r \ll N_{\text{state}}$  is the numerical rank of the prior preconditioned data misfit Hessian (see [17, 29]), then the cost of one Hessian solve is  $\mathcal{O}(r)$

CG iterations, that is,  $\mathcal{O}(2r)$  evaluations of the forward model  $\mathbf{F}$ . To summarize, the cost of evaluating the optimality criterion  $\mathcal{J}$  for a given design  $\zeta$  is  $\mathcal{O}(2r N_{\text{state}})$  forward model solves. Moreover, the cost of evaluating the gradient of  $\mathcal{J}$  with respect to the design is  $\mathcal{O}(2r n_s N_{\text{state}})$  model solves. Note that the cost of the gradient in the relaxed OED formulation reduces to  $\mathcal{O}(2r n_s N_{\text{state}})$  when the observation errors are uncorrelated with identical variances, that is, when  $\mathbf{\Gamma}_{\text{noise}} = \sigma_{\text{obs}}^2 \mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix and  $\sigma^2$  is the variance of the observational errors; see, e.g., [3].

In contrast, with the proposed algorithms, we do not need to evaluate the gradient of  $\mathcal{J}$  with respect to the design. At each iteration of Algorithm 3.1, the function  $\mathcal{J}$  is evaluated for each sampled design to evaluate the stochastic gradient. Assuming the size of the sample used to formulate the stochastic gradient  $\hat{\mathbf{g}}$  is  $N_{\text{ens}}$ , then the cost of each iteration is  $\mathcal{O}(2r N_{\text{ens}} N_{\text{state}})$ . The cost of evaluating the gradient of the multivariate Bernoulli distribution (3.6) is negligible compared with solving the forward model  $\mathbf{F}$ . Note that unlike the case with the relaxed OED formulation, in the proposed framework the design space is binary; and as we will show later, as the optimization algorithm proceeds, it reuses previously sampled designs. Moreover, the value of  $\mathcal{J}$  can be evaluated independently, and thus evaluating the stochastic gradient is embarrassingly parallel.

**4. Numerical experiments.** We start this section with a small illustrative model to clarify the approach proposed and provide additional insight. Next, we present numerical experiments using an advection-diffusion model.

**4.1. Results for a two-dimensional problem.** Here we discuss an idealized problem following the definition of the linear forward and inverse problem described in section 2. Python code for this set of experiments is available from [8]. We define the forward operator  $\mathbf{F}$  as a short wide matrix that projects model space into observation space. Moreover, we specify prior and observation covariance matrices and formulate the posterior covariance matrix  $\mathbf{\Gamma}_{\text{post}}$  and the objective  $\mathcal{J}$  accordingly. The forward operator and prior and observation noise covariances are

$$(4.1) \quad \mathbf{F} := \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}; \quad \mathbf{\Gamma}_{\text{pr}} := \text{diag}(4, 1, 0.25, 1); \quad \mathbf{\Gamma}_{\text{noise}} := \text{diag}(0.25, 1),$$

which result in the following form of the objective  $\mathcal{J} = \text{Tr}(\mathbf{\Gamma}_{\text{post}}(\zeta))$ :

$$(4.2) \quad \mathcal{J}(\zeta) = \text{Tr} \left( \left[ \begin{array}{cccc} \zeta_1 + 0.25 & \zeta_1 & 0 & 0 \\ \zeta_1 & \zeta_1 + 1 & 0 & 0 \\ 0 & 0 & 0.25\zeta_2 + 4 & 0.25\zeta_2 \\ 0 & 0 & 0.25\zeta_2 & 0.25\zeta_2 + 1 \end{array} \right]^{-1} \right) = \frac{8\zeta_1 + 5}{5\zeta_1 + 1} + \frac{2\zeta_2 + 20}{\zeta_2 + 16}.$$

The forward operator  $\mathbf{F}$  in (4.1) describes a simulation at four model gridpoints, with two candidate observational sensors. The first sensor measures the average value simulated at the first two model gridpoints, and the second sensor measures the average value simulated at the last two gridpoints. An A-optimal OED problem here seeks an optimal design  $\zeta \in \{0, 1\}^2$  that minimizes (4.2) and thus guides activation/deactivation of each of the two sensors.

Figure 2 (left) shows the surface of the two objective functions  $\mathcal{J}(\zeta)$  (4.2) and  $\Upsilon(\mathbf{p})$  (3.4), respectively. The objective functions are evaluated on a regular grid of 15 values equally spaced in each direction. In the same plot we also display the progress

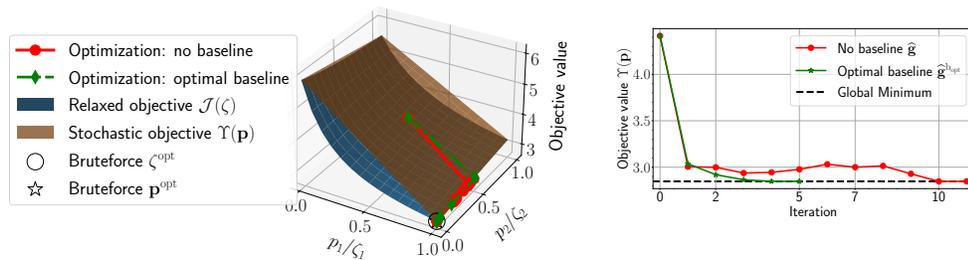


FIG. 2. Left: Surface plot of the relaxed-OED objective function  $\mathcal{J}(\zeta)$  (4.2) and the objective function  $\Upsilon(\mathbf{p})$  of the corresponding stochastic OED problem (3.4). In each direction 15 equally spaced points are taken to create the surface plots. Iterations of both Algorithms 3.1 and 3.2 are shown on the surface plot. Right: Value of the objective function  $\Upsilon$  evaluated at each iteration of the algorithm until convergence. Brute force results are obtained by searching over all 4 possible values of the binary design. The initial parameter  $\mathbf{p}^{(0)}$  of the optimizer is set to  $(0.5, 0.5)^\top$ , and the algorithm terminates when the magnitude of the projected gradient ( $\text{pgtol}$ ) is lower than  $10^{-8}$ .

of both Algorithms 3.1 and 3.2, where in the latter we utilize the optimal baseline  $b^{\text{opt}}$  described in subsection 3.5.1. In both cases, we set the learning rate to 0.25. The value of the objective function  $\Upsilon$  evaluated at each iteration of the optimizer is shown in Figure 2 (right). We note that, as explained in subsection 3.2, the values of  $\mathcal{J}(\zeta)$  and  $\Upsilon(\mathbf{p})$  coincide at the extremal points of the domain  $[0, 1]^{n_s}$ . Moreover, unlike the surface of the stochastic objective  $\Upsilon$ , the surface of the original objective function  $\mathcal{J}$ , evaluated at the relaxed design, flattens out for values of  $\mathbf{p}_1$  greater than 0.5. This behavior makes applying a sparsification procedure challenging when associated with traditional OED approaches.

Figure 2 shows that the optimization algorithms move quickly toward a lower-dimensional space corresponding to lower values of the objective  $\Upsilon$  and then move slowly toward a local optimum. We also note that since both candidate parameter values  $(0, 1)^\top$  and  $(1, 1)^\top$  have similar objective values, the optimizer moves slowly between them, since the value of the gradient in this direction is close to zero. We note, however, that Algorithm 3.2 outperforms Algorithm 3.1 and converges faster, in this case to the global minimum. If the optimizer is terminated before convergence (say, after the first iteration, where  $\mathbf{p}_1$  is set to 1 here), the returned value of  $\mathbf{p}_2$  is in the interval  $(0, 1)$ , which allows sampling estimates of  $\zeta_1^{\text{opt}}$  from  $\{0, 1\}$ , and the decision can be made based on the value of  $\mathcal{J}$  or other decisions, such as budget constraints. Alternatively, one could modify (4.2) by adding a regularization term to enforce desired constraints. This will be explained further in subsection 4.2.

**4.2. Experimental setup for an advection-diffusion problem.** In this subsection we demonstrate the effectiveness of our proposed approach using an advection-diffusion model simulation that has been used extensively in the literature; see, for example, [3, 5, 9, 10, 40].

The advection-diffusion model simulates the spatiotemporal evolution of a contaminant field  $u = u(\mathbf{x}, t)$  in a closed domain  $\mathcal{D}$ . Given a set of candidate locations to deploy sensors to measure the contaminant concentration, we seek the optimal subset of sensors that once deployed would enable inferring the initial distribution of the contaminant with minimum uncertainty. To this end, we seek the optimal subset of candidate sensors that minimized the A-optimality criterion, that is, the trace of the posterior covariance matrix. We carry out numerical experiments in a setup where

only 14 candidate sensor locations are considered inside the domain  $\mathcal{D}$ . The number of possible combinations of active sensors in this case is  $2^{14} = 16,384$ . Despite being large, this allows us to carry out a brute force search. The purpose of the brute force search here is to study the behavior of the proposed methodology and its capability in exploring the design space and utilizing any constraints properly, while seeking the optimal design. In particular, we can compare the quality of our solution with the global minimum in this case.

*Model setup: Advection-diffusion.* We use the same model setup. Specifically, the contaminant field  $u = u(\mathbf{x}, t)$  is governed by the advection-diffusion equation

$$(4.3) \quad \begin{aligned} u_t - \kappa \Delta u + \mathbf{v} \cdot \nabla u &= 0 && \text{in } \mathcal{D} \times [0, T], \\ u(x, 0) &= \theta && \text{in } \mathcal{D}, \\ \kappa \nabla u \cdot \mathbf{n} &= 0 && \text{on } \partial \mathcal{D} \times [0, T], \end{aligned}$$

where  $\kappa > 0$  is the diffusivity,  $T$  is the simulation final time, and  $\mathbf{v}$  is the velocity field. The spatial domain here is  $\mathcal{D} = [0, 1]^2$ , with two rectangular regions inside the domain simulating two buildings where the flow is not allowed to enter. Here,  $\partial \mathcal{D}$  refers to the boundary of the domain, which includes both the external boundary and the walls of the two buildings. The velocity field  $\mathbf{v}$  is assumed to be known and is obtained by solving a steady Navier–Stokes equation, with the side walls driving the flow; see [40] for further details. We utilize the package HippyLib [53] to solve the advection-diffusion problem (4.3) numerically. To create a synthetic simulation, we use the initial distribution of the contaminant shown in Figure 3 (left) as the ground truth.

*Observational setup.* We consider a set of uniformly distributed candidate sensor locations (spatial observational gridpoints). Specifically, we consider  $n_s = 14$  candidate sensor locations as described by Figure 3 (right), and we assume that the sensor locations do not change over time. An observation vector  $\mathbf{y}$  represents the concentration of the contaminant at the sensor locations, at a set of predefined time instances  $\{t_1, \dots, t_{n_t}\} \subset [0, T := 4]$ . The observation times are  $t_1 + s\Delta t$ , with initial observation time  $t_1 = 1$ ;  $\Delta t = 0.2$  is the model simulation timestep; and  $s = 0, 1, \dots, 20$ . The result is  $n_t = 16$  observation time instances, and  $N_{\text{obs}} = n_s \times n_t$  is the observation space dimension.

The observation error distribution is  $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma}_{\text{noise}})$ , with  $\mathbf{\Gamma}_{\text{noise}} \in \mathbb{R}^{N_{\text{obs}} \times N_{\text{obs}}}$  describing spatiotemporal correlations of observational errors. For simplicity, we assume that observation errors are uncorrelated, with fixed standard deviation; that is, the observation error covariance matrix takes the form  $\mathbf{\Gamma}_{\text{noise}} = \sigma_{\text{obs}}^2 \mathbf{I}$ , where  $\mathbf{I} \in \mathbb{R}^{N_{\text{obs}} \times N_{\text{obs}}}$  is

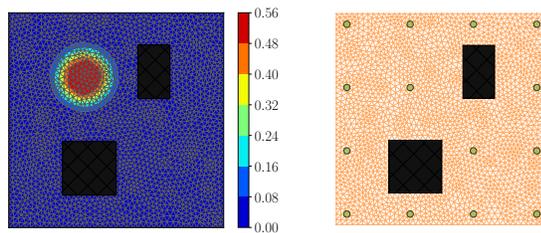


FIG. 3. *Advection-diffusion model domain, discretization, candidate sensor locations, and the ground truth, in other words, the true initial condition. Left: The physical domain  $\mathcal{D}$  including outer boundary and the two buildings, the model grid discretization, and the true model parameter. Right: Candidate observational sensor locations.*

the identity matrix. Here, we set the observation error variances to  $\sigma_{\text{obs}} = 2.482 \times 10^{-2}$ . This specific value is obtained by considering a noise level of 5% of the maximum value of the contaminant concentration captured at all observation points, by running a simulation over  $[0, T]$ , using the the true model parameter; see Figure 3 (left).

*Forward operator, adjoint operator, and the prior.* The forward operator  $\mathbf{F}$  maps the model parameter  $\theta$ , here the model initial condition, to the observation space. Specifically,  $\mathbf{F}$  represents a forward simulation over the interval  $[0, T]$  followed by applying an observation operator (here, a restriction operator), to extract concentrations at sensor locations at observation time instances. Here,  $\mathbf{F}$  is linear, and the adjoint is defined as  $\mathbf{F}^* := \mathbf{M}^{-1} \mathbf{F}^\top$ , where  $\mathbf{M}$  is the finite-element mass matrix. The prior distribution of the parameter  $\theta$  is modeled by a Gaussian distribution  $\mathcal{N}(\theta_{\text{pr}}, \mathbf{\Gamma}_{\text{pr}})$ , where  $\mathbf{\Gamma}_{\text{pr}}$  is a discretization of  $\mathcal{A}^{-2}$ , with  $\mathcal{A}$  being a Laplacian; see [17] for further details.

**4.3. The OED optimization problem.** Now we define the design space and formulate the OED optimization problem. To find the best subset of candidate sensor locations, we assign a binary design variable  $\zeta_i$  to each candidate sensor location  $x_i$ , where  $i = 1, 2, \dots, n_s$ , and hence  $\zeta \in \{0, 1\}^{n_s}$ . We aim to find a binary A-optimal design, that is, the minimizer of the trace of the posterior covariance matrix. Moreover, to promote sparsity of the design, we employ an  $\ell_0$  penalty term  $\Phi$ . We thus define the objective function  $\mathcal{J}$  for this problem as

$$(4.4) \quad \mathcal{J}(\zeta) = \text{Tr} \left( \left( \mathbf{M}^{-1} \mathbf{F}^\top \mathbf{\Gamma}_{\text{noise}}^{-1/2} \text{diag}(\zeta) \mathbf{\Gamma}_{\text{noise}}^{-1/2} \mathbf{F} + \mathbf{\Gamma}_{\text{pr}}^{-1} \right)^{-1} \right) + \alpha \Phi(\zeta),$$

where  $\alpha$  is the user-defined penalty parameter. This parameter controls the level of sparsity that we desire to impose on the design. Specifically, we set  $\Phi(\zeta) := \|\zeta\|_0$  to impose sparsity. On the other hand, if we have a specific budget  $\lambda$ , it would be more reasonable to define the penalty function as  $\Phi(\zeta) := \alpha \left| \|\zeta\|_0 - \lambda \right| = \alpha \left| \sum_{i=1}^{n_s} \zeta_i - \lambda \right|$ . We will discuss these two cases in the following and in the numerical experiments. The stochastic optimization problem (3.2) is formulated given the definition of  $\mathcal{J}$  in (4.4) as

$$(4.5) \quad \mathbf{p}^{\text{opt}} = \arg \min_{\mathbf{p} \in [0, 1]^{n_s}} \mathbb{E}_{\zeta \sim \mathbb{P}(\zeta|\mathbf{p})} \left[ \text{Tr} \left( \left( \mathbf{M}^{-1} \mathbf{F}^\top \mathbf{\Gamma}_{\text{noise}}^{-1/2} \text{diag}(\zeta) \mathbf{\Gamma}_{\text{noise}}^{-1/2} \mathbf{F} + \mathbf{\Gamma}_{\text{pr}}^{-1} \right)^{-1} \right) + \alpha \Phi(\zeta) \right],$$

where  $\mathbb{P}(\zeta|\mathbf{p})$  is the multivariate Bernoulli distribution with PMF given by (3.1).

**4.4. Numerical results with advection-diffusion model.** The main goal here is to study the behavior of the proposed Algorithm 3.2 compared with the global solution of (4.5). Solution by enumeration (brute force) is carried out for all  $2^{14} = 16,384$  possible designs, and the corresponding value of  $\mathcal{J}$  is recorded to identify the global solution of (4.5). In addition, we run Algorithm 3.2 with the maximum number of iterations set to 20. We choose this tight number to test the performance of the stochastic optimization algorithm upon early termination. We choose the learning rate  $\eta = 0.25$  and set the gradient tolerance PGTOL to  $10^{-8}$ . Each sensor is equipped with an initial probability 0.5. This is employed by choosing the initial parameter  $\mathbf{p}^{(0)}$  of the stochastic optimization algorithm to  $\mathbf{p}^{(0)} = (0.5, \dots, 0.5)^\top$ . In all experiments, we set the sample size for estimating the stochastic gradient to  $N_{\text{ens}} = 32$  and the number of epochs for the optimal baseline to  $b_m = 10$ .

The optimization algorithm returns samples from the multivariate Bernoulli distribution associated with the parameter  $\mathbf{p}^{\text{opt}}$  at the final step. Then, it picks  $\zeta^{\text{opt}}$  as

the sampled design associated with the smallest value of  $\mathcal{J}$ . We assume that the optimization procedure samples 10 designs upon termination, from the final distribution. Note that all samples will be identical if the probability distribution is degenerate. In the numerical results discussed next, we show not only the final optimal design returned by the optimization procedure but also the sampled designs.

**4.4.1. Results without penalty term.** We start with numerical results obtained by setting the penalty parameter  $\alpha=0$ . Figure 4 shows the results of the brute force search, along with results returned by Algorithms 3.1 and 3.2. Specifically, the value of  $\mathcal{J} := \text{Tr}(\mathbf{\Gamma}_{\text{post}}(\zeta))$  is evaluated at each possible design  $\zeta$  and is shown on the y-axis. Candidate binary designs are grouped on the x-axis by the number of entries set to 1, that is, the number of active sensors. In this setup, we have access to the value of  $\mathcal{J}$  corresponding to all possible designs, and thus we can in fact evaluate  $\Upsilon(\mathbf{p}) \equiv \mathbb{E}_{\zeta \sim \mathbb{P}(\zeta|\mathbf{p})}[\mathcal{J}(\zeta)]$  exactly, for any choice of the parameter  $\mathbf{p}$ . Of course, such an action is impossible in practice; however, we are interested in understanding the behavior of the optimization algorithm. Here, without any constraints on the number of sensors, the global optimal minimum is attained by  $\zeta^{\text{opt}} = \mathbf{1} \in \mathbb{R}^{n_s}$ , that is, by activating all sensors. However, we note that increasing the number of sensors, say, more than 9, would add little to information gain from data. The reason is the similarity of the values of  $\mathcal{J}$  for all designs with more than 9 active sensors.

In Figure 4, results of the brute force search are shown as light-blue dots. The designs sampled from the final distribution obtained by Algorithm 3.1 are marked as red x marks, which in this case are identical, showing that the final probability distribution is degenerate. The samples generated by Algorithm 3.2 with the optimal baseline estimate (3.23) are marked as green tripod-up marks; in this case, as indicated by the three unique designs in the sample, the resulting policy is nondegenerate and successfully explores the space near the global optimum. Note that both Algorithms 3.1 and 3.2 result in probability distributions (defined by  $\mathbf{p}^{\text{opt}}$ ) associated with small values. However, Algorithm 3.2 with the optimal baseline (3.23) outperforms Algorithm 3.1 and generates designs with significantly smaller objective values. Specifically, as shown in Figure 4, the objective value  $\mathcal{J}$  evaluated at the designs generated by Algorithm 3.2 fall within 1% of the global optimum.

Evaluating the optimal baseline estimate, however, requires additional evaluations of  $\mathcal{J}$ . We monitor the number of additional evaluations of  $\mathcal{J}$  carried out at each iteration of the optimizer. We keep track of the values of  $\mathcal{J}$  for each sampled design  $\zeta$  during the course of the algorithm. By doing so, we avoid any computational redundancy due to recalculating the objective function multiple times for the same design. Figure 5 shows the behavior of the optimization procedures over consecutive

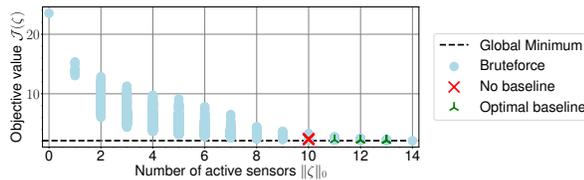


FIG. 4. Results of the policy gradient procedures (Algorithms 3.1, 3.2), compared with the brute force search of all candidate binary designs. No penalty is used here; that is, we set the sparsity penalty parameter to  $\alpha=0$ . Candidate designs are grouped by the number of active sensors, on the x-axis, with the corresponding value of  $\mathcal{J}$  displayed on the y-axis. Brute force results are shown as light-blue dots.

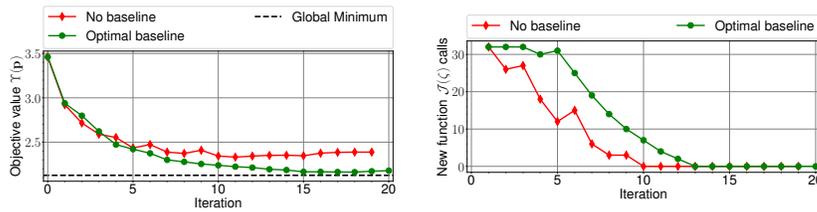


FIG. 5. Behavior of the policy gradient procedures (Algorithms 3.1, 3.2) over consecutive iterations. No penalty is used; we set the penalty parameter  $\alpha=0$ . Left: The value of the objective  $\Upsilon$  at each iteration of the optimization procedures. Right: Number of new function  $\mathcal{J}$  evaluations carried out.

iterations. Specifically, Figure 5 (left) shows the value of  $\Upsilon$  evaluated at the  $k$ th step, and Figure 5 (right) shows the number of new function evaluations carried out at each step. In both cases, we show results obtained by Algorithms 3.1 and 3.2.

Figure 5 (left) suggests that Algorithm 3.1 moves steadily toward a local minimum and converges in a small number of iterations (see Figure 5 (right)) but it fails to explore the space near the global optimum. This action is expected because of sampling error and the high variability of the estimator. Conversely, Algorithm 3.2 with optimal baseline estimate (3.23) moves steadily toward the global minimum. The results in Figure 5 (right) indicate that the computational cost, explained by the number of function evaluations, is not significantly different.

**4.4.2. Results with sparsity and fixed-budget constraints.** To study the behavior of the optimization procedures in the presence of  $\ell_0$  sparsity constraints, we set the penalty function to  $\Phi(\zeta) := \|\zeta\|_0$  and the regularization penalty parameter to  $\alpha = 1.0$ . Here we do not concern ourselves with the choice of  $\alpha$ , and we leave it for the user to tune based on the application at hand and the required level of sparsity. Furthermore, to study the behavior of the optimization algorithms in the presence of an exact budget constraint  $\|\zeta\|_0 = \lambda$ , we carry out the same procedure, with the penalty function set to  $\Phi(\zeta) := \|\zeta\|_0 - \lambda$  and the regularization penalty parameter set to  $\alpha = 1.0$ , and we set the budget to  $\lambda = 8$  sensors. Results are shown in Figures 6 and 7, respectively.

In the case of sparsity constraint, as suggested by Figure 6 (left), there is a unique global optimum design with only three active sensors. Both Algorithms 3.1 and 3.2 result in degenerate probability distributions. The global optimal design, however, is attained by utilizing the optimal baseline estimate as shown in Figure 6. In the case of a fixed-budget constraint, as suggested by Figure 6 (right), the two algorithms

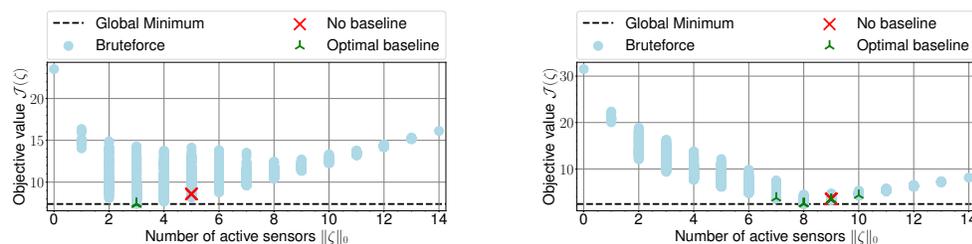


FIG. 6. Similar to Figure 4. Left: We use a sparsity constraint, defined by  $\Phi(\zeta) = \|\zeta\|_0$ . Right: We use budget constraint, defined by  $\Phi(\zeta) := \|\zeta\|_0 - \lambda$ , where  $\lambda = 8$ . In both cases, we set the sparsity penalty parameter to  $\alpha=1.0$ .

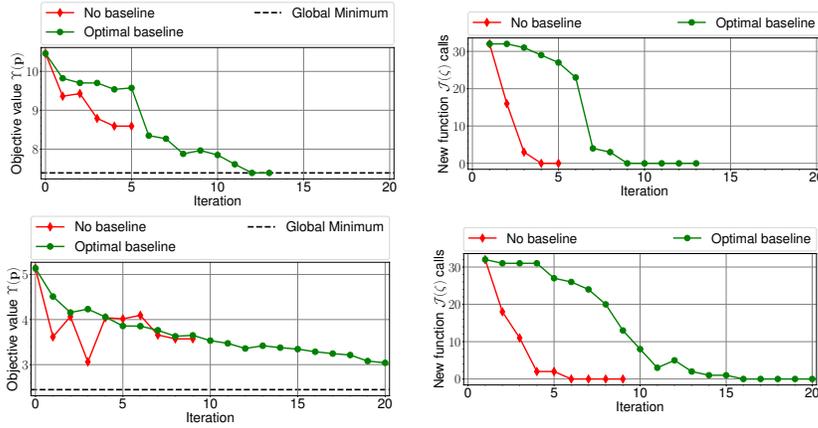


FIG. 7. Similar to Figure 5. Top: We use a sparsity constraint, defined by  $\Phi(\zeta) = \|\zeta\|_0$ . Bottom: We use budget constraint, defined by  $\Phi(\zeta) := \|\zeta\|_0 - \lambda$ , where  $\lambda = 8$ . In both cases, we set the sparsity penalty parameter to  $\alpha = 1.0$ .

(Algorithms 3.1 and 3.2) produce similar policies, but the result of Algorithm 3.2 with optimal baseline estimate (3.23) is significantly better as it produces designs that satisfy the budget constraint. The evolution of the optimization algorithms and the associated computational costs explained by the number of objective function evaluations are shown in Figure 7.

The performance of both Algorithms 3.1 and 3.2 is consistent with and without sparsity constraints. By including the optimal baseline estimate (3.23) in Algorithm 3.2, at slight additional computational cost, the global optimum is more likely to be discovered by the algorithm.

**4.4.3. The initial policy and the optimality gap.** Both Algorithms 3.1 and 3.2 require setting an initial policy by choosing the initial parameter  $\mathbf{p}^{(0)}$  of the optimizer. In the results above, we set  $\mathbf{p}^{(0)} := (0.5, \dots, 0.5)^T$ . Here, we show the results of carrying out the optimization procedure with different initial policies. Specifically, we choose 10 different randomly sampled initial policies (each  $\mathbf{p}_i$  is randomly sampled for a uniform distribution  $\mathcal{U}[0, 1]$ ). For each initial policy, we run Algorithms 3.1 and 3.2. The resulting policy is then used to sample 10 different designs. Figure 8 shows results obtained by setting the sparsity penalty parameter to

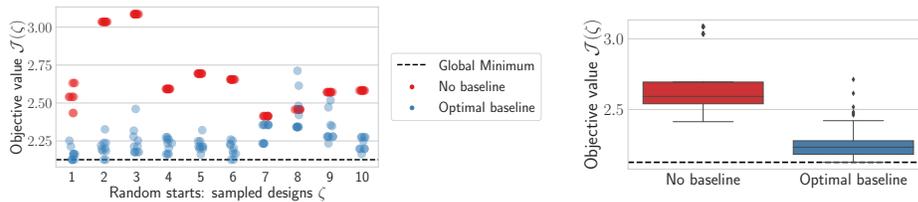


FIG. 8. Results of the policy gradient procedures (Algorithms 3.1 and 3.2). The optimization procedures are run 10 times from different (random) initial policies, and 10 designs are sampled from the policy returned by each algorithm in each experiment. Left: The 10 experiments are grouped by the initial policy on the x-axis. Right: The results of the 10 experiments, each with different initial policy, are grouped by the algorithm used, over all experiments. Here, the penalty parameter is set to  $\alpha = 0$ .



FIG. 9. Similar to Figure 8. Here we set the sparsity penalty parameter to  $\alpha = 1.0$  and use the sparsity-enforcing penalty defined by  $\Phi(\zeta) := \|\zeta\|_0$ .

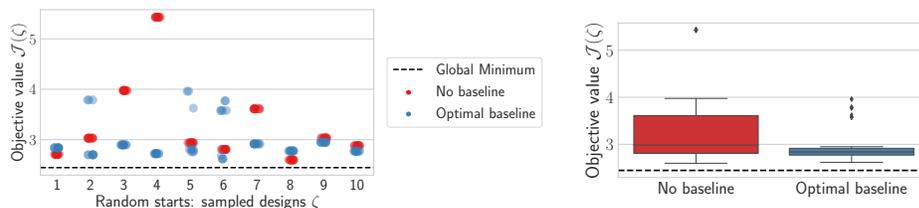


FIG. 10. Similar to Figure 9. Here we set the sparsity penalty parameter to  $\alpha = 1.0$  and use the penalty constraint defined by  $\Phi(\zeta) := \left| \|\zeta\|_0 - \lambda \right|$ , where  $\lambda = 8$ .

$\alpha = 0$ . Similarly, Figures 9 and 10 show results obtained in the presence of sparsity and fixed-budget constraints, respectively. These results suggest that Algorithm 3.2 exhibits much lower variability than Algorithm 3.1 and tends to significantly decrease the optimality gap, and hence is more reliable for exploring the probability space near the global optimal policy. We note, however, that global optimality is not guaranteed by the policy gradient algorithms presented in this work.

**4.4.4. Performance and scalability.** The cardinality of the design space  $|\Omega_\zeta|$  grows exponentially with respect to the number of candidate sensors  $n_s$ , which may bring significant variance for the stochastic optimization with a large  $n_s$  as shown in Theorem 3.3. The proposed formula for variance reduction via optimal baseline estimation (3.25), however, is developed to reduce the variance numerically. Here we investigate the scalability of the proposed approach compared to relaxed OED formulations (see, e.g., [3, 10]), by numerically analyzing the performance for an increasing number of candidate sensors  $n_s$ .

In this set of experiments, we use the same setup as in subsection 4.4.2, that is, we assume an exact budget of  $\lambda = 8$  sensors and enforce it by setting the penalty function  $\Phi(\zeta) := \left| \|\zeta\|_0 - \lambda \right|$  and the penalty parameter  $\alpha = 1$ . The budget  $\lambda = 8$  is enforced in the relaxed approach by using an  $\ell_1$  penalty function with the penalty parameter value tuned by an L-curve approach: using 20 values of  $\alpha$  to approximate the L-curve and selecting the best value; see [9] for details.

We start by analyzing the accuracy and the performance of the proposed approach. The total variance of the gradient estimator is estimated using an ensemble of 32 gradient evaluations. That is, we evaluate the gradients ((3.8) and (3.20), respectively) using 32 different initial random seeds and calculate the sum of the sample variance of these gradient approximations. To analyze the redundancy (reutilization of function  $\mathcal{J}$  evaluations), we study the total redundancy ratio defined as  $1 - \frac{\text{Number of unique sampled designs}}{\text{Total number of sampled designs}}$ . More computational savings are achieved for a redundancy rate closer to 1. As before, and for a fair comparison, we fix the maximum number of iterations of the optimization algorithm to 20 iterations.

Figure 11 shows the total variance and the redundancy incurred by the gradient estimators (3.8) and (3.20) with optimal baseline estimate (3.25), respectively, for an increasing number of candidate sensor locations  $n_s$ . Figure 11 (left) shows that the variance reduction—here the baseline (3.25)—is effective even for large numbers of candidate sensors  $n_s \approx 120$  and in fact is necessary for the success of the proposed approach. Figure 11 (right) shows that in this approach, previously evaluated  $\mathcal{J}$  values are effectively utilized, thus considerably reducing the computational cost.

Next, we study the accuracy and the computational cost of the proposed approach compared to relaxed OED. Figure 12 (left) shows the value of the objective  $\mathcal{J}(\zeta)$  evaluated at the optimal design obtained by each approach, for an increasing number of candidate sensors. Figure 12 (right) shows the total number of model evaluations (number of  $\mathbf{F}$  calls) incurred by the optimization procedures. While the relaxed OED approach can produce designs with reasonably small values of the objective  $\mathcal{J}$ , it is much more computationally demanding, and it is outperformed by the proposed approach.

We conclude this section by showing the values of the activation probability  $\mathbf{p}$  and the sampled designs  $\zeta$  generated by the policy gradient procedures (Algorithms 3.1 and 3.2) in the present set of experiments. Figure 13 shows the optimal parameter value obtained by Algorithm 3.1 (top) and the designs sampled from the corresponding optimal observational policy (bottom) for an increasing number of candidate sensor (as shown in Figure 12). Corresponding results obtained by Algorithm 3.2 are shown in Figure 14. These results show that (a) the policies returned by the optimization procedures are not always degenerate (not all activation probabilities are set to 0 or 1), and in this case an estimate of the optimal design is generated by comparing the values of  $\mathcal{J}(\zeta)$  associated with the sampled designs; (b) the baseline introduced to reduce the variability of the stochastic gradient estimate is crucial for the perfor-

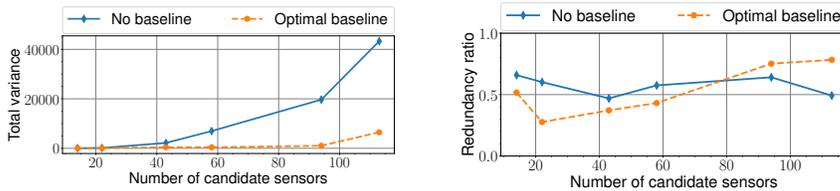


FIG. 11. Results of (Algorithms 3.1, 3.2) for an increasing number of candidate sensor locations. Left: Total variance of the gradient estimators averaged over optimization iterations. Right: Total redundancy ratio.

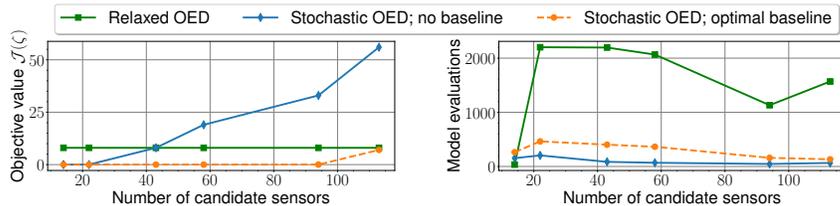


FIG. 12. Left: Value of the  $\ell_0$ -weighted objective function  $\mathcal{J}(\zeta)$  evaluated at the optimal design  $\zeta^{\text{opt}}$  obtained by (a) applying relaxed OED followed by thresholding (here we choose the highest  $\lambda=8$  weights); (b) the policy gradient approach, that is, Algorithms 3.1 and 3.2, respectively. Right: Number of forward model  $\mathbf{F}$  evaluations.

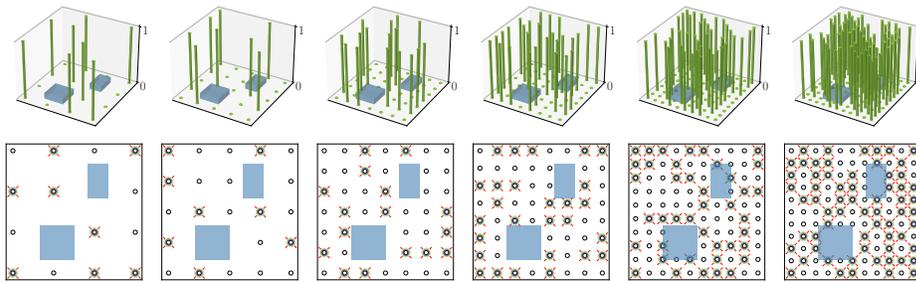


FIG. 13. Values of the optimal activation parameters/probabilities  $\mathbf{p}$  (top) and the sampled designs (bottom) obtained by the policy gradient procedure Algorithm 3.1 for an increasing number of candidate sensor locations (as shown in Figure 12). The designs sampled from the optimal policy are marked by circles, and the sampled optimal design (the design achieving the smallest value of  $\mathcal{J}$  in the sample) is highlighted by red cross marks.

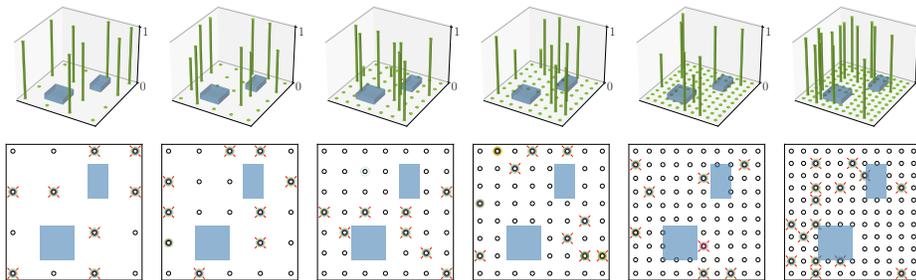


FIG. 14. Similar to Figure 13. Here, we show a result for the policy gradient procedure Algorithm 3.2.

mance of the optimization routine especially for large cardinality, i.e., large  $n_s$ ); (c) Algorithm 3.1 achieves reasonable behavior for small cardinality, but, its performance degrades for increasing cardinality, which is also noted from Figure 12 (left); and (d) the optimal baseline estimate (3.25) enables Algorithm 3.2 to produce a policy that reduces the value of  $\mathcal{J}$  by both reducing the OED criterion and by better respecting the budget constraint imposed by the  $\ell_0$  penalty. Note that for  $n_s \approx 120$ , Algorithm 3.2 still achieves considerably better performance than Algorithm 3.1, but it converged to a suboptimal policy. This can be enhanced by optimizing the learning rate, by incorporating antithetic variates [34], and by utilizing importance sampling [7]. These approaches will be further investigated in future works.

**4.4.5. Results with various learning rates.** We conclude this section of experiments with results obtained by different learning rates. We use the setup in subsection 4.4.2 and assume an exact budget of  $\lambda = 8$  sensors and enforce it by setting the penalty function  $\Phi(\zeta) := |||\zeta||_0 - \lambda|$  and the penalty parameter  $\alpha = 1$ . Figure 15 shows results obtained by varying the learning rate  $\eta$  in Algorithm 3.2, with the optimal baseline estimate (3.23), and note that similar behavior was observed for the other settings used earlier in the paper.

Figure 15 (left) shows the value of the stochastic objective function corresponding to the parameter  $\mathbf{p}^{(k)}$  at the  $k$ th iteration of the optimization algorithm for various choices of the learning rate. Figure 15 (right) shows the number of new calls to the function  $\mathcal{J}$  made at each iteration. By increasing the learning rate  $\eta$ , the algorithm

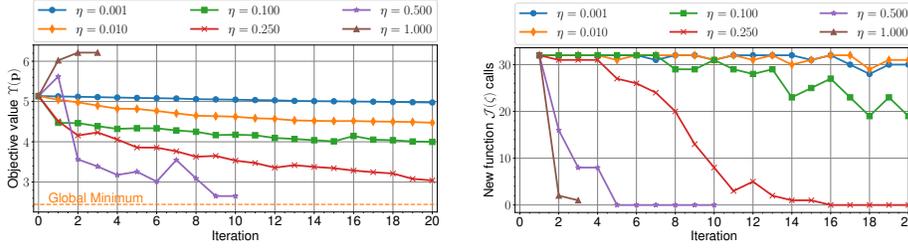


FIG. 15. Similar to Figure 7. Here, we set the sparsity penalty parameter to  $\alpha = 1.0$  and use the budget constraint defined by  $\Phi(\zeta) := \|\zeta\|_0 - \lambda$ , where  $\lambda = 8$ .

tends to converge quickly and explore the space of probability distributions near the global optimal policy very quickly. However, this action is also associated with the risk of divergence. Among the tested values,  $\eta = 0.5$  is the best learning rate.

In general, one can choose a small learning rate or even a decreasing sequence and run the optimization algorithm long enough to guarantee convergence to an optimal policy. Doing so, however, will likely increase the computational cost manifested in the number of evaluations of  $\mathcal{J}$ . This problem is widely known as the *exploration-exploitation trade-off* in the reinforcement learning literature. Finding an analytically optimal learning rate is beyond the scope of this paper.

**5. Discussion and concluding remarks.** In this work, we presented a new approach for binary optimization that is suitable for optimal design of experiments in Bayesian inverse problems constrained by expensive mathematical models, such as partial differential equations. The regularized utility function is cast into a stochastic objective defined over the parameters of multivariate Bernoulli distribution. A policy gradient algorithm is used to optimize the new objective function and thus yields an approximately optimal probability distribution, that is, policy from which an approximately optimal design is sampled. The proposed approach does not require differentiability of the design utility function nor the penalty function generally employed to enforce sparsity or regularity conditions on the design. Hence, the computational cost of the proposed methods, in terms of the number of forward model solves, is much less than the cost required by traditional gradient-based approach for optimal experimental design. The decrease in computational cost is due mostly to the fact that the proposed method does not require evaluation of the simulation model for each entry of the gradient. Sparsity-enforcing penalty functions such as  $\ell_0$  can be used directly, without the need to utilize a continuation procedure or apply a rounding technique.

The main open issue pertains to the optimal selection of the learning rate parameter. While a decreasing sequence satisfying the Robbins–Monro conditions guarantees convergence of the proposed algorithm almost surely, such a choice may require many iterations before convergence to a degenerate optimal policy. The proposed stochastic formulation can also be solved by other sample-based stochastic optimization methods, including sample average approximation [37, 43, 48]. The learning rate and performance of sample-based methods will be considered in future works.

In sensor placement problems, soft constraints are used to enforce sparsity or a budget on the design, and both classes of constraints have been used in this work. Accommodating hard-constraints, for example, to enforce a maximum number of sensors, has proven to be challenging because one has to map the constraint space to the

space of the policy parameters (i.e., parameters of the Bernoulli distribution). The multivariate Bernoulli distribution was an intuitive choice to define the probability of sensor activation. Other distributions can be employed to define an observational policy. For example, one can consider a mixture distribution to define optimal experimental design in a continuous spatial domain. This, however, requires different formulation of the Bayesian OED problem, the underlying observation error models, and the associated probability measures. These extensions will be considered in future works.

We note that utilizing traditional cost-reduction methods, including randomized matrix methods [13, 45], and other reduced-order modeling approaches (see, e.g., [12, 18, 23, 49]), to reduce the cost of the OED criterion  $\mathcal{J}$  apply to both the relaxed approach and our proposed approach equally. This shows that the proposed algorithms introduce massive computational savings to the OED solution process, compared with the traditional relaxation approach.

**Appendix A. Multivariate Bernoulli distribution.** The probabilities of a Bernoulli random variable  $\zeta \in \{0, 1\}$  are described by

$$(A.1) \quad \mathbb{P}(\zeta|\mathbf{p}) = \mathbf{p}^\zeta (1 - \mathbf{p})^{(1-\zeta)} \equiv \mathbf{p}\zeta + (1 - \mathbf{p})(1 - \zeta); \quad \zeta \in \{0, 1\}, \quad \mathbf{p} \in [0, 1],$$

where  $\mathbf{p}$  can be thought of as the probability of success in a one-trial experiment. Moreover, the following identity holds:

$$(A.2) \quad \frac{\partial \mathbb{P}(\zeta|\mathbf{p})}{\partial \mathbf{p}} = (-1)^{1-\zeta}.$$

Assuming  $\zeta_i, i = 1, 2, \dots, n_s$  are mutually independent Bernoulli random variables with respective success probabilities  $\mathbf{p}_i, i = 1, 2, \dots, n_s$ , then the joint probability mass function (PMF) of the random variable  $\zeta = (\zeta_1, \zeta_2, \dots, \zeta_{n_s})^\top$ , parameterized by  $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{n_s})^\top$ , takes the form

$$(A.3) \quad \mathbb{P}(\zeta|\mathbf{p}) = \prod_{i=1}^{n_s} \mathbf{p}_i^{\zeta_i} (1 - \mathbf{p}_i)^{1-\zeta_i} \equiv \prod_{i=1}^{n_s} (\mathbf{p}_i \zeta_i + (1 - \mathbf{p}_i)(1 - \zeta_i)), \quad \zeta_i \in \{0, 1\}.$$

By using (A.2), the first-order derivative of (A.3) with respect to the parameters  $\mathbf{p}_i$  is described by

$$(A.4) \quad \frac{\partial \mathbb{P}(\zeta|\mathbf{p})}{\partial \mathbf{p}_j} = \frac{\partial}{\partial \mathbf{p}_j} \left( \mathbf{p}_j^{\zeta_j} (1 - \mathbf{p}_j)^{1-\zeta_j} \prod_{\substack{i=1 \\ i \neq j}}^{n_s} \mathbf{p}_i^{\zeta_i} (1 - \mathbf{p}_i)^{1-\zeta_i} \right) = (-1)^{1-\zeta_j} \prod_{\substack{i=1 \\ i \neq j}}^{n_s} \mathbf{p}_i^{\zeta_i} (1 - \mathbf{p}_i)^{1-\zeta_i}.$$

Thus, the gradient can be written as

$$(A.5) \quad \nabla_{\mathbf{p}} \mathbb{P}(\zeta|\mathbf{p}) = \sum_{j=1}^{n_s} \frac{\partial \mathbb{P}(\zeta|\mathbf{p})}{\partial \mathbf{p}_j} = \sum_{j=1}^{n_s} (-1)^{1-\zeta_j} \prod_{\substack{i=1 \\ i \neq j}}^{n_s} \mathbf{p}_i^{\zeta_i} (1 - \mathbf{p}_i)^{1-\zeta_i} \mathbf{e}_j.$$

Note that the derivative given by (A.4) is the (signed) conditional probability of  $\zeta$  conditioned by  $\zeta_j$  and  $\mathbf{p}_j$ , respectively. The second-order derivatives follow as

$$(A.6) \quad \frac{\partial^2 \mathbb{P}(\zeta|\mathbf{p})}{\partial \mathbf{p}_k \partial \mathbf{p}_j} = (1 - \delta_{kj}) (-1)^{2-\zeta_j-\zeta_k} \prod_{\substack{i=1 \\ i \notin \{j,k\}}}^{n_s} \mathbf{p}_i^{\zeta_i} (1 - \mathbf{p}_i)^{1-\zeta_i},$$

where  $\delta_{kj}$  is the standard Kronecker delta function. The gradient of the log-probabilities, that is, the score function of the multivariate Bernoulli PMF (A.3), is given by

$$\begin{aligned}
 \text{(A.7)} \quad \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p}) &= \nabla_{\mathbf{p}} \log \prod_{i=1}^{n_s} \mathbf{p}_i^{\zeta_i} (1 - \mathbf{p}_i)^{1-\zeta_i} = \sum_{i=1}^{n_s} \nabla_{\mathbf{p}} \log \mathbf{p}_i^{\zeta_i} + \sum_{i=1}^{n_s} \nabla_{\mathbf{p}} \log (1 - \mathbf{p}_i)^{1-\zeta_i} \\
 &= \sum_{i=1}^{n_s} \zeta_i \nabla_{\mathbf{p}} \log \mathbf{p}_i + \sum_{i=1}^{n_s} (1 - \zeta_i) \nabla_{\mathbf{p}} \log (1 - \mathbf{p}_i) = \sum_{i=1}^{n_s} \left( \frac{\zeta_i}{\mathbf{p}_i} + \frac{\zeta_i - 1}{1 - \mathbf{p}_i} \right) \mathbf{e}_i.
 \end{aligned}$$

It follows immediately from (A.7) that

$$\begin{aligned}
 \text{(A.8)} \quad \nabla_{\mathbf{p}} \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p}) &= \sum_{i=1}^{n_s} \left( \frac{-\zeta_i}{\mathbf{p}_i^2} - \frac{1 - \zeta_i}{(1 - \mathbf{p}_i)^2} \right) \mathbf{e}_i \mathbf{e}_i^{\top}, \\
 \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p}) (\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p}))^{\top} &= \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} \left( \frac{\zeta_i}{\mathbf{p}_i} - \frac{1 - \zeta_i}{1 - \mathbf{p}_i} \right) \left( \frac{\zeta_j}{\mathbf{p}_j} - \frac{1 - \zeta_j}{1 - \mathbf{p}_j} \right) \mathbf{e}_i \mathbf{e}_j^{\top}.
 \end{aligned}$$

In the rest of this appendix, we prove some identities essential for convergence analysis of the algorithms proposed in this work. We start with the following basic relations. First note that  $\mathbb{E}[\zeta_i \zeta_j] = \text{cov}(\zeta_i, \zeta_j) + \mathbb{E}[\zeta_i] \mathbb{E}[\zeta_j] = \delta_{ij} \mathbf{p}_i (1 - \mathbf{p}_i) + \mathbf{p}_i \mathbf{p}_j$ . This means that  $\mathbb{E}[\zeta_i^2] = \mathbb{E}[\zeta_i] = \mathbf{p}_i$  and  $\mathbb{E}[\zeta_i \zeta_j] = \mathbf{p}_i \mathbf{p}_j \forall i \neq j$ . Similarly,  $\mathbb{E}[\zeta_i (\zeta_j - 1)] = \mathbb{E}[\zeta_i \zeta_j] - \mathbb{E}[\zeta_i] = \delta_{ij} \mathbf{p}_i (1 - \mathbf{p}_i) + \mathbf{p}_i \mathbf{p}_j - \mathbf{p}_i$ , and  $\mathbb{E}[(\zeta_i - 1)(\zeta_j - 1)] = \mathbb{E}[\zeta_i \zeta_j] - \mathbf{p}_i - \mathbf{p}_j + 1 = \delta_{ij} \mathbf{p}_i (1 - \mathbf{p}_i) + \mathbf{p}_i \mathbf{p}_j - \mathbf{p}_i - \mathbf{p}_j + 1$ .

LEMMA A.1. *Let  $\zeta \in \Omega_{\zeta} := \{0, 1\}^{n_s}$  be a random variable following the joint Bernoulli distribution (A.3) and assume that  $\text{var}(\zeta)$  is the total variance operator that evaluates the trace of the variance-covariance matrix of the random variable  $\zeta$ . Then the following identities hold:*

$$\text{(A.9)} \quad \mathbb{E}[\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})] = 0; \quad \text{var}(\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})) = \sum_{i=1}^{n_s} \frac{1}{\mathbf{p}_i - \mathbf{p}_i^2}.$$

*Proof.* The first identity follows as

$$\text{(A.10)} \quad \mathbb{E}[\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})] = \sum_{\zeta} \nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p}) \mathbb{P}(\zeta|\mathbf{p}) = \sum_{\zeta} \nabla_{\mathbf{p}} \mathbb{P}(\zeta|\mathbf{p}) = \nabla_{\mathbf{p}} \sum_{\zeta} \mathbb{P}(\zeta|\mathbf{p}) = 0.$$

By definition of the covariance matrix, and since  $\mathbb{E}[\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})] = 0$ , then

$$\begin{aligned}
 \text{(A.11)} \quad \text{var}(\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})) &= \text{Tr} \left( \mathbb{E} \left[ (\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})) (\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p}))^{\top} \right] \right) \\
 &= \mathbb{E} \left[ \text{Tr} \left( (\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})) (\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p}))^{\top} \right) \right] = \mathbb{E} \left[ (\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p}))^{\top} (\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})) \right],
 \end{aligned}$$

where we utilized the circular property of the trace operator and the fact that the matrix trace is a linear operator. Thus,

(A.12)

$$\begin{aligned}
\text{var}(\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})) &= \mathbb{E}\left[(\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p}))^\top (\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p}))\right] = \mathbb{E}\left[\sum_{i=1}^{n_s} \left(\frac{\partial \log \mathbb{P}(\zeta|\mathbf{p}_i)}{\partial \mathbf{p}_i}\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n_s} \left(\frac{\zeta_i}{\mathbf{p}_i} + \frac{\zeta_i - 1}{1 - \mathbf{p}_i}\right)^2\right] = \sum_{i=1}^{n_s} \left(\frac{\mathbb{E}[\zeta_i^2]}{\mathbf{p}_i^2} + \frac{\mathbb{E}[\zeta_i^2] - 2\mathbb{E}[\zeta_i] + 1}{(1 - \mathbf{p}_i)^2}\right) \\
&= \sum_{i=1}^{n_s} \left(\frac{\mathbf{p}_i}{\mathbf{p}_i^2} + \frac{\mathbf{p}_i - 2\mathbf{p}_i + 1}{(1 - \mathbf{p}_i)^2}\right) = \sum_{i=1}^{n_s} \left(\frac{1}{\mathbf{p}_i} + \frac{1 - \mathbf{p}_i}{(1 - \mathbf{p}_i)^2}\right) = \sum_{i=1}^{n_s} \frac{1}{\mathbf{p}_i - \mathbf{p}_i^2},
\end{aligned}$$

where we used the fact that  $\mathbb{E}[\zeta_i^2] = \mathbf{p}_i$ , as discussed above. This is also obvious since  $\zeta^2 = \zeta$ .  $\square$

LEMMA A.2. *Let  $\zeta \in \Omega_\zeta := \{0, 1\}^{n_s}$  be a random variable following the joint Bernoulli distribution (A.3). Then for any  $\zeta \in \Omega_\zeta$ , the following bounds hold:*

$$(A.13) \quad \|\nabla_{\mathbf{p}} \mathbb{P}(\zeta|\mathbf{p})\| \leq \sqrt{n_s} \max_{j=1, \dots, n_s} \min_{\substack{i=1, \dots, n_s \\ k \neq j}} \mathbb{P}(\zeta_i|\mathbf{p}_i),$$

$$(A.14) \quad \mathbb{E}_\zeta \left[ \|\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})\|^2 \right] \equiv \text{var}(\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})) \leq \frac{n_s}{\min_i \mathbf{p}_i} + \frac{n_s}{1 - \max_i \mathbf{p}_i}.$$

*Proof.*

(A.15)

$$\begin{aligned}
\|\nabla_{\mathbf{p}} \mathbb{P}(\zeta|\mathbf{p})\|^2 &= \left\| \sum_{j=1}^{n_s} (-1)^{1-\zeta_j} \prod_{\substack{i=1 \\ k \neq j}}^{n_s} \mathbf{p}_i^{\zeta_i} (1 - \mathbf{p}_i)^{1-\zeta_i} \mathbf{e}_j \right\|^2 \leq \sum_{j=1}^{n_s} |(-1)^{1-\zeta_j}| \left( \prod_{\substack{i=1 \\ k \neq j}}^{n_s} \mathbf{p}_i^{\zeta_i} (1 - \mathbf{p}_i)^{1-\zeta_i} \right)^2 \\
&\leq \sum_{j=1}^{n_s} \min_{\substack{i=1, \dots, n_s \\ k \neq j}} (\mathbb{P}(\zeta_i|\mathbf{p}_i))^2 \leq n_s \max_{j=1, \dots, n_s} \min_{\substack{i=1, \dots, n_s \\ k \neq j}} (\mathbb{P}(\zeta_i|\mathbf{p}_i))^2,
\end{aligned}$$

which prove the first inequality (A.13). By utilizing (A.7), we have

$$\begin{aligned}
(A.16) \quad \|\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})\|^2 &= \left\| \sum_{i=1}^{n_s} \left(\frac{\zeta_i}{\mathbf{p}_i} + \frac{\zeta_i - 1}{1 - \mathbf{p}_i}\right) \mathbf{e}_i \right\|^2 = \sum_{i=1}^{n_s} \left(\frac{\zeta_i}{\mathbf{p}_i} + \frac{\zeta_i - 1}{1 - \mathbf{p}_i}\right)^2 \\
&= \sum_{i=1}^{n_s} \left(\frac{\zeta_i}{\mathbf{p}_i^2} + \frac{\zeta_i - 1}{(1 - \mathbf{p}_i)^2}\right),
\end{aligned}$$

where the last relation follows given the fact that  $\zeta_i \in \{0, 1\}$ , and hence  $\zeta_i(1 - \zeta_i) = 0$ , and  $\zeta_i^2 = \zeta_i \forall i = 1, 2, \dots, n_s$ . Taking the expectation of both sides, we get

(A.17)

$$\begin{aligned}
\mathbb{E}_\zeta \left[ \|\nabla_{\mathbf{p}} \log \mathbb{P}(\zeta|\mathbf{p})\|^2 \right] &= \mathbb{E}_\zeta \left[ \sum_{i=1}^{n_s} \left(\frac{\zeta_i}{\mathbf{p}_i^2} + \frac{\zeta_i - 1}{(1 - \mathbf{p}_i)^2}\right) \right] = \sum_{i=1}^{n_s} \left(\frac{\mathbb{E}[\zeta_i]}{\mathbf{p}_i^2} + \frac{\mathbb{E}[\zeta_i] - 1}{(1 - \mathbf{p}_i)^2}\right) \\
&= \sum_{i=1}^{n_s} \left(\frac{\mathbf{p}_i}{\mathbf{p}_i^2} + \frac{\mathbf{p}_i - 1}{(1 - \mathbf{p}_i)^2}\right) = \sum_{i=1}^{n_s} \left(\frac{1}{\mathbf{p}_i} + \frac{1}{(1 - \mathbf{p}_i)}\right) \\
&= \sum_{i=1}^{n_s} \frac{1}{\mathbf{p}_i} + \sum_{i=1}^{n_s} \frac{1}{(1 - \mathbf{p}_i)} \leq \frac{n_s}{\min_i \mathbf{p}_i} + \frac{n_s}{1 - \max_i \mathbf{p}_i},
\end{aligned}$$

which completes the proof of (A.14).  $\square$

**Appendix B. Proofs of theorems and lemmas discussed in subsection 3.4.**

*Proof of Theorem 3.3.* The first bound is obtained as follows:

$$(B.1) \quad \begin{aligned} \|\nabla_{\mathbf{p}} \Upsilon(\mathbf{p})\|^2 &= \left\| \sum_{k=1}^{2^{n_s}} \mathcal{J}(\zeta[k]) \nabla_{\mathbf{p}} \mathbb{P}(\zeta[k]|\mathbf{p}) \right\|^2 \leq \sum_{k=1}^{2^{n_s}} \|\mathcal{J}(\zeta[k]) \nabla_{\mathbf{p}} \mathbb{P}(\zeta[k]|\mathbf{p})\|^2 \\ &= \sum_{k=1}^{2^{n_s}} (\mathcal{J}(\zeta[k]))^2 \|\nabla_{\mathbf{p}} \mathbb{P}(\zeta[k]|\mathbf{p})\|^2 \leq C^2 \sum_{k=1}^{2^{n_s}} \|\nabla_{\mathbf{p}} \mathbb{P}(\zeta[k]|\mathbf{p})\|^2. \end{aligned}$$

Lemma A.2 provides a bound on the gradient of log probability of the multivariate Bernoulli distribution (3.1) as  $\|\nabla_{\mathbf{p}} \mathbb{P}(\zeta|\mathbf{p})\| \leq \sqrt{n_s} \max_{j=1, \dots, n_s} \min_{k=1, \dots, n_s, k \neq j} \mathbb{P}(\zeta_k|\mathbf{p}_k)$ . It then follows from (B.1) that

$$\|\nabla_{\mathbf{p}} \Upsilon(\mathbf{p})\|^2 \leq n_s C^2 \sum_{i=1}^{2^{n_s}} \max_{j=1, \dots, n_s} \min_{k=1, \dots, n_s, k \neq j} \mathbb{P}(\zeta_k[i]|\mathbf{p}_k),$$

and hence

$$(B.2) \quad \|\nabla_{\mathbf{p}} \Upsilon(\mathbf{p})\|^2 \leq n_s 2^{n_s} C^2 \max_{i=1, \dots, 2^{n_s}} \max_{j=1, \dots, n_s} \min_{k=1, \dots, n_s, k \neq j} \mathbb{P}(\zeta_k[i]|\mathbf{p}_k).$$

Because  $\mathbb{P}(\zeta|\mathbf{p}) \leq 1$ , it follows immediately that  $\|\nabla_{\mathbf{p}} \Upsilon(\mathbf{p})\|^2 \leq n_s 2^{n_s} C^2$ , and the bound (3.13a) follows by taking the square root on both sides. The second bound (3.13b) follows by noting that

$$(B.3) \quad \|\nabla_{\mathbf{p}} \Upsilon(\mathbf{p}[1]) - \nabla_{\mathbf{p}} \Upsilon(\mathbf{p}[2])\| \leq \|\nabla_{\mathbf{p}} \Upsilon(\mathbf{p}[1])\| + \|\nabla_{\mathbf{p}} \Upsilon(\mathbf{p}[2])\| \leq 2\sqrt{n_s |\Omega_{\zeta}|} C.$$

The second-order derivative of the objective  $\Upsilon$  is

$$(B.4) \quad \text{Hess}_{\mathbf{p}} \Upsilon = \nabla_{\mathbf{p}} \nabla_{\mathbf{p}} \Upsilon = \sum_{\zeta} \mathcal{J}(\zeta) \nabla_{\mathbf{p}} \nabla_{\mathbf{p}} \mathbb{P}(\zeta|\mathbf{p}) = \sum_{i=k}^{2^{n_s}} \mathcal{J}(\zeta[k]) \nabla_{\mathbf{p}} \nabla_{\mathbf{p}} \mathbb{P}(\zeta[k]|\mathbf{p}).$$

Equation (B.4) shows that the  $(i, j)$ th entry of the Hessian (B.4) is  $\sum_{k=1}^{2^{n_s}} \mathcal{J}(\zeta[k]) \frac{\partial^2 \mathbb{P}(\zeta|\mathbf{p})}{\partial \mathbf{p}_i \partial \mathbf{p}_j}$ . The second-order derivative of the Bernoulli distribution is discussed in detail in Appendix A, and it follows from (A.6) that  $\frac{\partial^2 \mathbb{P}(\zeta|\mathbf{p})}{\partial \mathbf{p}_i \partial \mathbf{p}_j} \leq 1$ . This means that the entries of the Hessian (B.4) are bounded as follows,

$$(B.5) \quad \left( \frac{\partial^2 \Upsilon}{\partial \mathbf{p}_i \partial \mathbf{p}_j} \right)^2 = \left( \sum_{i=k}^{2^{n_s}} \mathcal{J}(\zeta[k]) \frac{\partial^2 \mathbb{P}(\zeta|\mathbf{p})}{\partial \mathbf{p}_i \partial \mathbf{p}_j} \right)^2 \leq C^2 \sum_{k=1}^{2^{n_s}} \left( \frac{\partial^2 \mathbb{P}(\zeta[k]|\mathbf{p})}{\partial \mathbf{p}_i \partial \mathbf{p}_j} \right)^2 \leq C^2 2^{n_s},$$

and the bound in (3.14) follows by taking the square root of both sides.  $\square$

*Proof of Lemma 3.4.* For any realization of the random design  $\zeta \sim \mathbb{P}(\zeta|\mathbf{p})$ , it holds that

$$\begin{aligned}
\mathbb{E}[\mathcal{J}(\zeta)\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta|\mathbf{p})] &= \sum_{\zeta}\mathcal{J}(\zeta)\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta|\mathbf{p})\mathbb{P}(\zeta|\mathbf{p}) = \sum_{\zeta}\mathcal{J}(\zeta)\nabla_{\mathbf{p}}\mathbb{P}(\zeta|\mathbf{p}) \\
\text{(B.6)} \qquad \qquad \qquad &= \nabla_{\mathbf{p}}\sum_{\zeta}\mathcal{J}(\zeta)\mathbb{P}(\zeta|\mathbf{p}) = \nabla_{\mathbf{p}}\mathbb{E}[\mathcal{J}(\zeta)] = \nabla_{\mathbf{p}}\Upsilon(\mathbf{p}).
\end{aligned}$$

Thus, the unbiasedness of the estimator  $\widehat{\mathbf{g}}$  follows because

$$\text{(B.7)} \quad \mathbb{E}[\widehat{\mathbf{g}}] = \mathbb{E}\left[\frac{1}{N_{\text{ens}}}\sum_{j=1}^{N_{\text{ens}}}\mathcal{J}(\zeta[j])\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta[j]|\mathbf{p})\right] \stackrel{\text{(B.6)}}{=} \frac{1}{N_{\text{ens}}}\sum_{j=1}^{N_{\text{ens}}}\nabla_{\mathbf{p}}\Upsilon(\mathbf{p}) = \nabla_{\mathbf{p}}\Upsilon(\mathbf{p}).$$

The total variance  $\text{var}(\widehat{\mathbf{g}})$  follows by the definition of  $\widehat{\mathbf{g}}$  as

$$\text{(B.8)} \quad \text{var}(\widehat{\mathbf{g}}) = \text{var}\left(\frac{1}{N_{\text{ens}}}\sum_{j=1}^{N_{\text{ens}}}\mathcal{J}(\zeta[j])\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta[j]|\mathbf{p})\right) = \frac{1}{N_{\text{ens}}}\text{var}(\mathcal{J}(\zeta)\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta|\mathbf{p})).$$

To prove (3.17), we rewrite (B.8) as follows,

$$\begin{aligned}
\text{(B.9)} \quad \mathbb{E}[\widehat{\mathbf{g}}^{\text{T}}\widehat{\mathbf{g}}] &= \text{var}(\widehat{\mathbf{g}}) + \mathbb{E}[\widehat{\mathbf{g}}]^{\text{T}}\mathbb{E}[\widehat{\mathbf{g}}] = \frac{1}{N_{\text{ens}}}\text{var}(\mathcal{J}(\zeta)\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta|\mathbf{p})) + \mathbf{g}^{\text{T}}\mathbf{g} \\
&\leq \frac{1}{N_{\text{ens}}}\text{var}(C\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta|\mathbf{p})) + \mathbf{g}^{\text{T}}\mathbf{g} = \frac{C^2}{N_{\text{ens}}}\text{var}(\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta|\mathbf{p})) + \mathbf{g}^{\text{T}}\mathbf{g}. \\
&\leq \frac{C^2 n_s}{N_{\text{ens}}}\left(\frac{1}{\min_i \mathbf{p}_i} + \frac{1}{1 - \max_i \mathbf{p}_i}\right) + \mathbf{g}^{\text{T}}\mathbf{g},
\end{aligned}$$

where the last inequality follows by (A.14), and, as before,  $C = \max_{\zeta \in \Omega_{\zeta}} \{|\mathcal{J}(\zeta)|\}$ . From Theorem 3.3, the gradient is bounded. By setting  $K_1 = \frac{C^2 n_s}{N_{\text{ens}}}\left(\frac{1}{\min_i \mathbf{p}_i} + \frac{1}{1 - \max_i \mathbf{p}_i}\right)$  and  $K_2 = 1$ , one can guarantee that  $\mathbb{E}[\widehat{\mathbf{g}}^{\text{T}}\widehat{\mathbf{g}}] \leq K_1 + K_2 \mathbf{g}^{\text{T}}\mathbf{g}$ .  $\square$

*Proof of Lemma 3.5.* The first identity follows from

$$\text{(B.10)} \quad \mathbb{E}[\mathbf{d}] = \mathbb{E}\left[\frac{1}{N_{\text{ens}}}\sum_{j=1}^{N_{\text{ens}}}\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta[j]|\mathbf{p})\right] = \frac{1}{N_{\text{ens}}}\sum_{j=1}^{N_{\text{ens}}}\mathbb{E}[\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta[j]|\mathbf{p})] = 0,$$

where the last equality follows from (A.9). Because  $\mathbb{E}[\mathbf{d}] = 0$ , and by utilizing Lemma A.1, the second identity follows as

$$\begin{aligned}
\text{(B.11)} \quad \text{var}(\mathbf{d}) &= \mathbb{E}[\mathbf{d}^{\text{T}}\mathbf{d}] - \mathbb{E}[\mathbf{d}]^{\text{T}}\mathbb{E}[\mathbf{d}] = \mathbb{E}[\mathbf{d}^{\text{T}}\mathbf{d}] = \frac{1}{N_{\text{ens}}^2}\sum_{j=1}^{N_{\text{ens}}}\text{var}(\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta[j]|\mathbf{p})) \\
&= \frac{1}{N_{\text{ens}}^2}\sum_{j=1}^{N_{\text{ens}}}\sum_{i=1}^{n_s}\frac{1}{\mathbf{p}_i - \mathbf{p}_i^2} = \frac{1}{N_{\text{ens}}}\sum_{i=1}^{n_s}\frac{1}{\mathbf{p}_i - \mathbf{p}_i^2}. \quad \square
\end{aligned}$$

*Proof of Lemma 3.6.* The estimator is unbiased because  $\nabla_{\mathbf{p}}\Upsilon^{\text{b}}(\mathbf{p}) = \nabla_{\mathbf{p}}\Upsilon(\mathbf{p})$ , and

(B.12)

$$\begin{aligned} \mathbb{E}[\widehat{\mathbf{g}}^b] &= \mathbb{E}\left[\frac{1}{N_{\text{ens}}}\sum_{j=1}^{N_{\text{ens}}}\mathcal{J}(\zeta[j]-b)\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta[j]|\mathbf{p})\right] = \frac{1}{N_{\text{ens}}}\sum_{j=1}^{N_{\text{ens}}}\mathbb{E}[\mathcal{J}(\zeta[j]-b)\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta[j]|\mathbf{p})] \\ &= \frac{1}{N_{\text{ens}}}\sum_{j=1}^{N_{\text{ens}}}\mathbb{E}[\mathcal{J}(\zeta[j])\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta[j]|\mathbf{p})] - \frac{1}{N_{\text{ens}}}\sum_{j=1}^{N_{\text{ens}}}\mathbb{E}[b\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta[j]|\mathbf{p})] \\ &= \nabla_{\mathbf{p}}\Upsilon(\mathbf{p}) - \frac{b}{N_{\text{ens}}}\sum_{j=1}^{N_{\text{ens}}}\mathbb{E}[\nabla_{\mathbf{p}}\log\mathbb{P}(\zeta[j]|\mathbf{p})] = \nabla_{\mathbf{p}}\Upsilon(\mathbf{p}), \end{aligned}$$

where the last step follows by Lemma A.1. The total variance of the estimator  $\widehat{\mathbf{g}}^b$  is given by

$$(B.13) \quad \text{var}(\widehat{\mathbf{g}}^b) = \mathbb{E}\left[(\widehat{\mathbf{g}}^b)^\top \widehat{\mathbf{g}}^b\right] - \mathbb{E}[\widehat{\mathbf{g}}^b]^\top \mathbb{E}[\widehat{\mathbf{g}}^b] = \mathbb{E}\left[(\widehat{\mathbf{g}}^b)^\top \widehat{\mathbf{g}}^b\right] - \mathbf{g}^\top \mathbf{g},$$

where  $\mathbf{g} = \nabla_{\mathbf{p}}\Upsilon^b = \nabla_{\mathbf{p}}\Upsilon$ . The first term follows as

$$\begin{aligned} \mathbb{E}\left[(\widehat{\mathbf{g}}^b)^\top \widehat{\mathbf{g}}^b\right] &= \mathbb{E}\left[(\widehat{\mathbf{g}} - b\mathbf{d})^\top (\widehat{\mathbf{g}} - b\mathbf{d})\right] = \mathbb{E}[\widehat{\mathbf{g}}^\top \widehat{\mathbf{g}}] - 2\mathbb{E}[\widehat{\mathbf{g}}^\top b\mathbf{d}] + \mathbb{E}[b\mathbf{d}^\top b\mathbf{d}] \\ &= \text{var}(\widehat{\mathbf{g}}) + \mathbb{E}[\widehat{\mathbf{g}}]^\top \mathbb{E}[\widehat{\mathbf{g}}] - 2b\mathbb{E}[\widehat{\mathbf{g}}^\top \mathbf{d}] + b^2\text{var}(\mathbf{d}) + b^2\mathbb{E}[\mathbf{d}]^\top \mathbb{E}[\mathbf{d}] \\ (B.14) \quad &= \text{var}(\widehat{\mathbf{g}}) + \mathbf{g}^\top \mathbf{g} - 2b\mathbb{E}[\widehat{\mathbf{g}}^\top \mathbf{d}] + b^2\text{var}(\mathbf{d}) \\ &= \text{var}(\widehat{\mathbf{g}}) + \mathbf{g}^\top \mathbf{g} - 2b\mathbb{E}[\widehat{\mathbf{g}}^\top \mathbf{d}] + \frac{b^2}{N_{\text{ens}}}\sum_{i=1}^{n_s}\frac{1}{\mathbf{p}_i - \mathbf{p}_i^2}. \end{aligned}$$

From (B.13), (B.14), the total variance follows as

$$(B.15) \quad \text{var}(\widehat{\mathbf{g}}^b) = \text{var}(\widehat{\mathbf{g}}) - 2b\mathbb{E}[\widehat{\mathbf{g}}^\top \mathbf{d}] + \frac{b^2}{N_{\text{ens}}}\sum_{i=1}^{n_s}\frac{1}{\mathbf{p}_i - \mathbf{p}_i^2}. \quad \square$$

**Acknowledgment.** We thank three anonymous referees and the associate editor for their detailed and insightful comments that helped us improve our manuscript.

REFERENCES

- [1] A. ALEXANDERIAN, *Optimal experimental design for infinite-dimensional Bayesian inverse problems governed by PDEs: A review*, Inverse Problems, 2021.
- [2] 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.
- [3] A. ALEXANDERIAN, N. PETRA, G. STADLER, AND O. GHATTAS, *A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized  $\ell_0$ -sparsification*, SIAM J. Sci. Comput., 36 (2014), pp. A2122–A2148.
- [4] 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.
- [5] A. ALEXANDERIAN, N. PETRA, G. STADLER, AND I. SUNSERI, *Optimal design of large-scale Bayesian linear inverse problems under reducible model uncertainty: Good to know what you don't know*, SIAM/ASA J. Uncertain. Quantif., 9 (2021), pp. 163–184.
- [6] 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.
- [7] B. AROUNA, *Adaptive Monte Carlo method, a variance reduction technique*, Monte Carlo Methods Appl., 10 (2004), pp. 1–24.

- [8] A. ATTIA, *DOERL: Design of Experiments Using Reinforcement Learning*, 2021, <https://gitlab.com/ahmedattia/doerl>.
- [9] 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).
- [10] A. ATTIA AND E. CONSTANTINESCU, *Optimal Experimental Design for Inverse Problems in the Presence of Observation Correlations*, preprint, <https://arxiv.org/abs/2007.14476>, 2020.
- [11] A. ATTIA AND A. SANDU, *A Hybrid Monte Carlo sampling filter for non-Gaussian data assimilation*, *AIMS Geosci.*, 1 (2015), pp. 41–78.
- [12] A. ATTIA, R. STEFANESCU, AND A. SANDU, *The reduced-order hybrid Monte Carlo sampling smoother*, *Internat. J. Numer. Methods Fluids*, 83 (2017), pp. 28–51.
- [13] H. AVRON AND S. TOLEDO, *Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix*, *J. ACM*, 58 (2011), pp. 1–34.
- [14] R. BANNISTER, *A review of operational methods of variational and ensemble-variational data assimilation*, *Quart. J. Roy. Meteor. Soc.*, 143 (2017), pp. 607–633.
- [15] D. P. BERTSEKAS AND J. TSITSIKLIS, *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA, 1996.
- [16] S. P. BRADLEY, A. C. HAX, AND T. L. MAGNANTI, *Applied Mathematical Programming*, Addison-Wesley, Reading, MA, 1977.
- [17] 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.
- [18] T. CUI, Y. MARZOUK, AND K. WILLCOX, *Scalable posterior approximations for large-scale Bayesian inverse problems via likelihood-informed parameter and state reduction*, *J. Comput. Phys.*, 315 (2016), pp. 363–387.
- [19] R. DALEY, *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, 1991.
- [20] A. DEFazio, F. BACH, AND S. LACOSTE-JULIEN, *SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives*, in *Proceedings of Advances in Neural Information Processing Systems*, 2014, pp. 1646–1654.
- [21] J. DUPACOVÁ AND R. WETS, *Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems*, *Ann. Statist.*, 16 (1988), pp. 1517–1549.
- [22] V. FEDOROV AND J. LEE, *Design of experiments in statistics*, in *Handbook of Semidefinite Programming*, R. S. H. Wolkowicz and L. Vandenberghe, eds., *Internat. Ser. Oper. Res. Management Sci.* 27, Kluwer Academic Publishers, Boston, 2000, pp. 511–532.
- [23] 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.
- [24] E. HABER, L. HORESH, AND L. TENORIO, *Numerical methods for experimental design of large-scale linear ill-posed inverse problems*, *Inverse Problems*, 24 (2008), pp. 125–137.
- [25] E. HABER, L. HORESH, AND L. TENORIO, *Numerical methods for the design of large-scale nonlinear discrete ill-posed inverse problems*, *Inverse Problems*, 26 (2010), 025002.
- [26] E. HABER, Z. MAGNANT, C. LUCERO, AND L. TENORIO, *Numerical methods for A-optimal designs with a sparsity constraint for ill-posed inverse problems*, *Comput. Optim. Appl.*, 52 (2012), pp. 293–314.
- [27] X. HUAN AND Y. MARZOUK, *Gradient-based stochastic optimization methods in Bayesian experimental design*, *Int. J. Uncertain. Quantif.*, 4 (2014).
- [28] X. HUAN AND Y. M. MARZOUK, *Simulation-based optimal Bayesian experimental design for nonlinear systems*, *J. Comput. Phys.*, 232 (2013), pp. 288–317.
- [29] 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.
- [30] J. KAIPIO AND E. SOMERSALO, *Statistical and Computational Inverse Problems*, *Appl. Math. Sci.* 160, Springer, New York, 2006.
- [31] A. J. KING AND R. T. ROCKAFELLAR, *Asymptotic theory for solutions in statistical estimation and stochastic programming*, *Math. Oper. Res.*, 18 (1993), pp. 148–162.
- [32] A. J. KLEYWEGT, A. SHAPIRO, AND T. HOMEM-DE MELLO, *The sample average approximation method for stochastic discrete optimization*, *SIAM J. Optim.*, 12 (2002), pp. 479–502.
- [33] K. KOVAL, A. ALEXANDERIAN, AND G. STADLER, *Optimal experimental design under irreducible uncertainty for linear inverse problems governed by PDEs*, *Inverse Problems*, 36 (2020).
- [34] P. L’ECUYER, *Efficiency improvement and variance reduction*, in *Proceedings of the Winter Simulation Conference*, IEEE, 1994, pp. 122–132.

- [35] W.-K. MAK, D. P. MORTON, AND R. K. WOOD, *Monte Carlo bounding techniques for determining solution quality in stochastic programs*, Oper. Res. Lett., 24 (1999), pp. 47–56.
- [36] I. M. NAVON, *Data assimilation for numerical weather prediction: A review*, in Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications, Springer, New York, 2009, pp. 21–65.
- [37] A. NEMIROVSKI, A. JUDITSKY, G. LAN, AND A. SHAPIRO, *Robust stochastic approximation approach to stochastic programming*, SIAM J. Optim., 19 (2009), pp. 1574–1609.
- [38] J. NOCEDAL AND S. WRIGHT, *Numerical Optimization*, Springer, New York, 2006.
- [39] A. PÁZMAN, *Foundations of Optimum Experimental Design*, D. Reidel, Boston, 1986.
- [40] N. PETRA AND G. STADLER, *Model Variational Inverse Problems Governed by Partial Differential Equations*, Technical report 11-05, The Institute for Computational Engineering and Sciences, The University of Texas at Austin, 2011.
- [41] F. PUKELSHEIM, *Optimal Design of Experiments*, John Wiley & Sons, New York, 1993.
- [42] S. J. REDDI, S. SRA, B. PÓCZOS, AND A. SMOLA, *Fast Incremental Method for Nonconvex Optimization*, preprint, <https://arxiv.org/abs/1603.06159>, 2016.
- [43] H. ROBBINS AND S. MONRO, *A stochastic approximation method*, Ann. Math. Statist., 22 (1951), pp. 400–407.
- [44] S. SAGER, H. G. BOCK, AND M. DIEHL, *The integer approximation error in mixed-integer optimal control*, Math. Program., 133 (2012), pp. 1–23.
- [45] A. K. SAIBABA, A. ALEXANDERIAN, AND I. C. IPSEN, *Randomized matrix-free trace and log-determinant estimators*, Numer. Math., 137 (2017), pp. 353–395.
- [46] M. SCHMIDT, N. LE ROUX, AND F. BACH, *Minimizing finite sums with the stochastic average gradient*, Math. Program., 162 (2017), pp. 83–112.
- [47] A. SHAPIRO, *Asymptotic analysis of stochastic programs*, Ann. Oper. Res., 30 (1991), pp. 169–186.
- [48] J. C. SPALL, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, Wiley Ser. Discrete Math. Optim. 65, John Wiley & Sons, New York, 2005.
- [49] A. SPANTINI, A. SOLONEN, T. CUI, J. MARTIN, L. TENORIO, AND Y. MARZOUK, *Optimal low-rank approximations of Bayesian linear inverse problems*, SIAM J. Sci. Comput., 37 (2015), pp. A2451–A2487.
- [50] R. S. SUTTON, D. A. MCALLESTER, S. P. SINGH, AND Y. MANSOUR, *Policy gradient methods for reinforcement learning with function approximation*, in Proceedings of Advances in Neural Information Processing Systems, 2000, pp. 1057–1063.
- [51] A. TARANTOLA, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, PA, 2005.
- [52] D. UCINSKI, *Optimal sensor location for parameter estimation of distributed processes*, Internat. J. Control, 73 (2000), pp. 1235–1248.
- [53] 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), pp. 1–34.
- [54] R. J. WILLIAMS, *Simple statistical gradient-following algorithms for connectionist reinforcement learning*, Mach. Learn., 8 (1992), pp. 229–256.
- [55] L. A. WOLSEY AND G. L. NEMHAUSER, *Integer and Combinatorial Optimization*, Wiley Ser. Discrete Math. Optim. 55, John Wiley & Sons, New York, 1999.
- [56] 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.