SIAM J. SCI. COMPUT.
Vol. 38, No. 2, pp. B247–B271

FUNCTIONAL REGRESSION FOR STATE PREDICTION USING LINEAR PDE MODELS AND OBSERVATIONS*

N. C. NGUYEN†, H. MEN†, R. M. FREUND‡, AND J. PERAIRE†

Abstract. Partial differential equations (PDEs) are commonly used to model a wide variety of physical phenomena. A PDE model of a physical problem is typically described by conservation laws, constitutive laws, material properties, boundary conditions, boundary data, and geometry. In most practical applications, however, the PDE model is only an approximation to the real physical problem due to both (i) the deliberate mathematical simplification of the model to keep it tractable and (ii) the inherent uncertainty of the physical parameters. In such cases, the PDE model may not produce a good prediction of the true state of the underlying physical problem. In this paper, we introduce a functional regression method that incorporates observations into a deterministic linear PDE model to improve its prediction of the true state. Our method is devised as follows. First, we augment the PDE model with a random Gaussian functional which serves to represent various sources of uncertainty in the model. We next derive a linear regression model for the Gaussian functional by utilizing observations and adjoint states. This allows us to determine the posterior distribution of the Gaussian functional and the posterior distribution for our estimate of the true state. Furthermore, we consider the problem of experimental design in this setting, wherein we develop an algorithm for designing experiments to efficiently reduce the variance of our state estimate. We provide several examples from the heat conduction, the convection-diffusion equation, and the reduced wave equation, all of which demonstrate the performance of the proposed methodology.

Key words. Gaussian functional, state prediction, linear PDEs, experimental design, data assimilation, nonparametric regression

AMS subject classifications. 60G15, 62G08, 65N75

DOI. 10.1137/14100275X

1. Introduction.

Partial differential equations (PDEs) are widely used to model a wide variety of physical phenomena in the real world and predict associated physical states such as temperature, displacement, velocity, pressure, or density. The predictability of the PDE model depends on how well it captures the real physics. The model captures the physics and yields a good prediction if all aspects of the model, such as constitutive laws, material properties, boundary conditions, boundary data, and geometry can be accurately prescribed. In practice, the PDE model is imperfect due to (i) the deliberate mathematical simplification of the model to keep it tractable (by ignoring certain physics that pose computational difficulties) and (ii) the uncertainty of the model data (by using geometry, material properties, and boundary data that are different from those of the physical problem.). In such cases, the PDE model should be validated and the uncertainty should be quantified.

There are a number of different approaches for dealing with model discrepancy. One approach is to represent physical inputs as random parameters or random fields, thereby resulting in *stochastic PDEs*. There exist several numerical methods for solving stochastic PDEs such as Monte Carlo (MC) methods [12, 16], intrusive poly-

*Submitted to the journal's Computational Methods in Science and Engineering section January 5, 2015; accepted for publication (in revised form) November 30, 2015; published electronically March 23, 2016. This work was supported by the Air Force Office of Scientific Research under the AFOSR grant FA9550-12-1-0357 and AFOSR grant FA9550-15-1-0276.

http://www.siam.org/journals/sisc/38-2/100275.html

†MIT Department of Aeronautics and Astronautics, Cambridge, MA 02139 (cuongng@mit.edu, abbymen@mit.edu, peraire@mit.edu).

‡MIT Sloan School of Management, Cambridge, MA 02139 (rfreund@mit.edu).

nomial chaos [13, 38], nonintrusive polynomial chaos [37, 10, 18], stochastic collocation [37, 3, 28], response surface [5, 14], and Kriging [6, 23, 20, 39, 9].

Data assimilation is another approach that combines the PDE model with observations to estimate the physical state. There are several different data assimilation methods. In *stochastic data assimilation* [19, 24, 34], the state estimate is represented as a stochastic process and is determined by minimizing its variance. In *variational data assimilation* [21, 22, 40], the state estimate is defined as an optimal solution of a least-squares minimization principle. In *parameter estimation* [7, 1, 21, 36], a number of uncertain parameters in the PDE model are determined by matching the model outputs to the observations. In *reduced order modeling* [4, 11, 25, 26, 29, 35], the state is reconstructed by fitting the observations to the snapshots which are computed by solving a parametrized or time-varying PDE model.

In this paper, we focus on a data assimilation method recently introduced in [27]. The method is devised as follows. First, we augment the PDE model with a random Gaussian functional which serves to represent various sources of uncertainty in the model. This gives rise to a stochastic PDE model whose solution is characterized by the Gaussian functional. We next derive a linear regression model for the Gaussian functional by utilizing observations and adjoint states. This functional regression model allows us to compute the posterior distribution for our estimate of the physical state, thereby quantifying the prediction error. A crucial ingredient in our method is the *covariance operator* representing the *prior* distribution of the Gaussian functional. The bilinear covariance operators considered incorporate a number of *hyperparameters* that are determined upon the observations by maximizing the likelihood of the Gaussian functional with respect to the hyperparameters. Furthermore, we consider the problem of experimental design in this setting, wherein we develop an algorithm for designing experiments to efficiently reduce the variance of our state estimate.

We can relate our approach to existing data assimilation methods. Our approach seeks to characterize model uncertainties by introducing a Gaussian functional into the existing numerical model, while stochastic data assimilation directly represents the state estimate as a stochastic process and infers its posterior distribution from the observations, the model outputs, and the priors about the background state. We show in Appendix B that our method and the Kalman method [19] yield exactly the same posterior distribution for a judicious choice of the priors. Our approach can also be shown to yield a posterior mean that satisfies the least-square minimization principle of three-dimensional (3D) variational data assimilation [8, 22].

The proposed method shares the following features with stochastic data assimilation methods. First, it incorporates both the PDE model and the observations into the regression procedure. Second, it can handle the observations given in the form of linear functionals of the field variable. Third, it yields not only the mean estimate but also the posterior variance that quantifies the prediction error. And fourth, it provides a natural mechanism for designing experiments to reduce the posterior variance. Furthermore, the method has a distinctive feature in that it provides a systematic and rational way to construct the model structure (including the prior covariance and experimental noise) based on the observations.

There are a number of new contributions relative to our previous work [27]. Herein we propose an efficient greedy algorithm to rationally select good observations from a large number of possible experiments. We show how our method can be related to Bayesian inference, 3D variational data assimilation, and the Kalman method. Finally, we apply our approach to a variety of PDEs including convection-diffusion equations and Helmholtz equations to demonstrate its performance.

The paper is organized as follows. In section 2, we formulate the problem of interest. In section 3, we present the functional regression method for state estimation using linear PDE models and observations. In section 4, we describe a greedy algorithm for selecting the observations. In section 5, we demonstrate our approach on several examples from heat conduction, convection-diffusion equation, and reduced wave equation. In section 6, we conclude the paper with some remarks on future research. Finally, in the appendix, we show the connection our method to 3D variational data assimilation and the Kalman method.

2. Problem formulation. Let Ω denote a bounded open domain with Lipschitz boundary. Let V be an appropriate finite element approximation space, which is defined on a triangulation of the domain Ω. The Galerkin finite element formulation of a general linear PDE model can be stated as follows. We seek a solution $u^{\mathrm{o}} \in V$ and an output vector $\boldsymbol{s}^{\mathrm{o}} \in \mathbb{C}^M$ such that

$$a(u^{\mathrm{o}}, v) = \ell(v) \quad \forall\, v \in V, \tag{1a}$$

$$s_i^{\mathrm{o}} = c_i(u^{\mathrm{o}}), \quad i = 1, \ldots, M. \tag{1b}$$

Here $a : V \times V \to \mathbb{C}$ is a continuous bilinear form, $\ell : V \to \mathbb{C}$ is a continuous and bounded linear functional, and $c_i : V \to \mathbb{C}, i = 1, \ldots, M$ are continuous and bounded linear functionals. The finite element approximation space V is of N dimensions. We shall take N to be so large that the numerical solution u is indistinguishable from the exact solution of the PDE model at any accuracy level of interest.

We assume that the underlying finite element (FE) model (1) is used to predict the true state of a deterministic and time-independent physical problem. We denote the true state by u^{true} and the associated true output vector by $\boldsymbol{s}^{\mathrm{true}} \in \mathbb{C}^M$. Both u^{true} and $\boldsymbol{s}^{\mathrm{true}}$ are not known. However, we assume that we are given a vector of M observations $\boldsymbol{d} \in \mathbb{C}^M$, which are the measurements of the true output vector $\boldsymbol{s}^{\mathrm{true}}$. We further assume that the observation vector differs from the true output vector $\boldsymbol{s}^{\mathrm{true}}$ by additive Gaussian noise, namely,

$$\boldsymbol{d} = \boldsymbol{s}^{\mathrm{true}} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_M) \in \mathbb{C}^M$ are independent, identically distributed Gaussian random variables with zero mean and variance σ^2.

As mentioned in the introduction, the solution u^{o} of the model (1) may not be a good approximation to the true state u^{true} because of various sources of uncertainty arising from the imprecise knowledge of boundary conditions, geometry, physical parameters, and governing equations. The uncertainty of our PDE model in predicting the true physical state prompts basic research questions such as the following:

- How do we combine the model (1) and the observations (2) to yield a more accurate prediction of the true state?
- How do we quantify the error in the prediction?
- How do we design experiments to improve the accuracy of the prediction?

In the remainder of this paper, we concentrate on addressing these questions. We address the first two questions by using the *functional regression* method described in the next section. We address the third question by devising an experimental design procedure in section 4.

3. Functional regression.

3.1. Gaussian functional. We suppose that we are given a bounded linear functional $g : V \to \mathbb{C}$, and we consider the following model: determine $u \in V$ and $\boldsymbol{s} \in \mathbb{C}^M$ such that

$$a(u, v) + g(v) = \ell(v) \qquad \forall\, v \in V, \tag{3a}$$

$$s_i = c_i(u), \quad i = 1, \ldots, M. \tag{3b}$$

Notice that the new model (3) differs from the original model (1) by the addition of the functional g. Here we can exactly determine u and $\boldsymbol{s}$ only if g is known. Indeed, together with $a(u, v)$ and $\ell(v)$, the functional g determines the solution u and the output vector $\boldsymbol{s}$ of the model (3). Note that the introduction of the functional g is able to capture the uncertainties in the functional ℓ and the bilinear form a, as demonstrated by numerical examples in section 5. Indeed, if we choose $g(v) = \ell(v) - a(u^{\text{true}}, v)$, then $u = u^{\text{true}}$ is the solution of (3). Unfortunately, this particular choice of g requires foreknowledge of the true state u^{true}, which we presumably do not know and which indeed we seek to accurately predict.

In order to capture various sources of uncertainty in the original model (1), we hypothesize the linear functional g as a *Gaussian process* [31]. Specifically, the Gaussian functional g is assumed to have zero mean and covariance operator k, namely,

$$g(v) \sim \mathcal{GP}(0, k(w, v)) \qquad \forall\, w, v \in V\ . \tag{4}$$

We require that the covariance operator $k : V \times V \to \mathbb{C}$ is symmetric positive-definite, namely,

$$k(w, v) = k(v, w), \quad k(v, v) > 0 \quad \forall v \neq 0, \quad \text{and} \quad k(0, 0) = 0. \tag{5}$$

As the covariance operator k characterizes the space of all possible functionals prior to taking into account the observations, it plays an important role in our method. The selection of a covariance operator will be discussed later.

3.2. Functional regression model. Here we develop a linear regression model for the Gaussian functional. To this end, we first obtain the adjoint solutions $\phi_i \in V, i = 1, \ldots, M$ by solving the adjoint problems:

$$a(v, \phi_i) = -c_i(v), \quad \forall v \in V. \tag{6}$$

It follows from (1), (3), and (6) that

$$g(\phi_i) = \ell(\phi_i) - a(u, \phi_i) = a(u^{\text{o}}, \phi_i) + c_i(u) = c_i(u) - c_i(u^{\text{o}}) = s_i - s_i^{\text{o}} \tag{7}$$

for $i = 1, \ldots, M$. Moreover, we assume that the observation d_i differs from the output estimate s_i by Gaussian noise $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, namely,

$$s_i = d_i - \varepsilon_i, \qquad i = 1, \ldots, M. \tag{8}$$

This equation is analogous to (2), which relates the observed data $\boldsymbol{d}$ to the true outputs $\boldsymbol{s}^{\text{true}}$. We substitute (8) into (7) to obtain

$$d_i - s_i^{\text{o}} = g(\phi_i) + \varepsilon_i, \qquad i = 1, \ldots, M. \tag{9}$$

This equation can be viewed as a standard regression model, namely, a linear model with additive Gaussian noise, and we shall now show that the model (9) allows us

to determine the posterior distribution of $g(v)$ for any given test function $v \in V$ as follows.

Let $\Phi = [\phi_1, \ldots, \phi_M]$ be a collection of M adjoint states as determined by (6). Let $\Psi = [\psi_1 \in V, \ldots, \psi_N \in V]$ be a collection of N test functions, where $\psi_j, 1 \le j \le N$, are basis functions of the space V. We would like to compute the posterior distribution of $\boldsymbol{g} \in \mathbb{C}^N$ with $g_i = g(\psi_i), i = 1, \ldots, N$. According to the prior (4) and the regression model (9), the joint distribution of $(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})$ and $\boldsymbol{g}$ is given by

$$\begin{bmatrix} \boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}} \\ \boldsymbol{g} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{K}(\Phi,\Phi) + \sigma^2 \boldsymbol{I} & \boldsymbol{K}(\Phi,\Psi) \\ \boldsymbol{K}(\Psi,\Phi) & \boldsymbol{K}(\Psi,\Psi) \end{bmatrix} \right), \tag{10}$$

where $\boldsymbol{K}(\Phi,\Phi) \in \mathbb{C}^{M\times M}$, $\boldsymbol{K}(\Phi,\Psi) \in \mathbb{C}^{M\times N}$, $\boldsymbol{K}(\Psi,\Phi) \in \mathbb{C}^{N\times M}$, and $\boldsymbol{K}(\Psi,\Psi) \in \mathbb{C}^{N\times N}$ have entries

$$\begin{aligned} K_{ij}(\Phi,\Phi) &= k(\phi_i,\phi_j), \qquad i = 1,\ldots,M, j = 1,\ldots,M, \\ K_{ij}(\Phi,\Psi) &= k(\phi_i,\psi_j), \qquad i = 1,\ldots,M, j = 1,\ldots,N, \\ K_{ij}(\Psi,\Phi) &= k(\psi_i,\phi_j), \qquad i = 1,\ldots,N, j = 1,\ldots,M, \\ K_{ij}(\Psi,\Psi) &= k(\psi_i,\psi_j), \qquad i = 1,\ldots,N, j = 1,\ldots,N, \end{aligned} \tag{11}$$

respectively. It thus follows from (10) and the conditional normal distribution formula (see Appendix A.2 in Rasmussen and Williams [31]) that the posterior distribution of $\boldsymbol{g}$ is

$$p\left(\boldsymbol{g} | (\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}}), \Phi, \Psi\right) \sim \mathcal{N}(\bar{\boldsymbol{g}}, \boldsymbol{G}), \tag{12}$$

where the mean vector $\bar{\boldsymbol{g}} \in \mathbb{C}^N$ and the covariance matrix $\boldsymbol{G} \in \mathbb{C}^{N\times N}$ are given by

$$\bar{\boldsymbol{g}} = \boldsymbol{K}(\Psi,\Phi)\boldsymbol{D}^{-1}(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}}), \qquad \boldsymbol{G} = \boldsymbol{K}(\Psi,\Psi) - \boldsymbol{K}(\Psi,\Phi)\boldsymbol{D}^{-1}\boldsymbol{K}(\Phi,\Psi) \tag{13}$$

with $\boldsymbol{D} = \boldsymbol{K}(\Phi,\Phi) + \sigma^2\boldsymbol{I}$. The posterior mean and covariance (13) require the adjoint states (6), the inner products (11), and the inverse of $\boldsymbol{D}$. Since N is typically much larger than M, the computational cost is dominated by the cost of computing the adjoint states.

We note that the posterior mean $\bar{\boldsymbol{g}}$ depends linearly on $(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})$, which is the difference between the observation vector and the output vector of the original model (1). Another way to look at the posterior mean is to view it as a linear combination of M inner products, each one associated with one of the adjoint states, by writing

$$\bar{g}_i = \sum_{j=1}^{M} \beta_j k(\psi_i, \phi_j), \qquad i = 1,\ldots,N, \tag{14}$$

where $\boldsymbol{\beta} = \boldsymbol{D}^{-1}(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})$. We see that the discrepancy between the observed data and the original model (1) has an important effect on the posterior mean.

Also note that the posterior covariance $\boldsymbol{G}$ in (13) does not explicitly depend on the values of the observed outputs but only on the adjoint states Φ and the covariance operator $k(\cdot,\cdot)$. This is based on the premise that the covariance operator $k(\cdot,\cdot)$ is somehow known and given. In reality, of course, we create/invent $k(\cdot,\cdot)$ based on notions both of general principles as well as computational expedients. As will be seen in section 3.3, we will construct/model a suitable covariance operator $k(\cdot,\cdot)$ based on data observations and maximum likelihood estimation. In this way the

posterior covariance $\boldsymbol{G}$ in (13) will implicitly depend on the observed outputs and our modeling assumptions. Furthermore, notice that the covariance matrix $\boldsymbol{G}$ in (13) is the difference of two terms: the first term $\boldsymbol{K}(\Psi, \Psi)$ is simply the prior covariance, while the second term represents a reduction in the covariance due to the observations.

Finally, we express the state estimate as $u(\boldsymbol{x}) = \sum_{n=1}^{N} u_n \psi_n(\boldsymbol{x})$, where $\boldsymbol{u} = (u_1, \ldots, u_N) \in \mathbb{C}^N$ denotes the state vector. Substituting this expression into the model (3) and choosing $v = \psi_i$ for $i = 1, \ldots, N$, we obtain the following linear system:

$$\boldsymbol{A}\boldsymbol{u} = \boldsymbol{l} - \boldsymbol{g}, \tag{15}$$

where $\boldsymbol{A}$, $\boldsymbol{l}$, and $\boldsymbol{g}$ have entries $A_{ij} = a(\psi_i, \psi_j)$, $l_i = \ell(\psi_i)$, and $g_i = g(\psi_i)$ for $i, j = 1, \ldots, N$, respectively. It follows from (12) and (15) that the state vector $\boldsymbol{u}$ obeys a normal distribution

$$\boldsymbol{u}|(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}}), \Phi, \Psi \sim \mathcal{N}(\bar{\boldsymbol{u}}, \boldsymbol{U}), \tag{16}$$

where the posterior mean vector $\bar{\boldsymbol{u}} \in \mathbb{C}^N$ and covariance matrix $\boldsymbol{U} \in \mathbb{C}^{N \times N}$ are given by

$$\bar{\boldsymbol{u}} = \boldsymbol{A}^{-1}(\boldsymbol{l} - \bar{\boldsymbol{g}}), \qquad \boldsymbol{U} = \boldsymbol{A}^{-1}\boldsymbol{G}\boldsymbol{A}^{-\mathrm{H}}. \tag{17}$$

Here $\bar{\boldsymbol{g}}$ and $\boldsymbol{G}$ are given by (13). The superscript H denotes the complex conjugate transpose.

Let us use the notation $\boldsymbol{\psi}(\boldsymbol{x}) = (\psi_1(\boldsymbol{x}), \ldots, \psi_N(\boldsymbol{x})) \in \mathbb{C}^N$ to denote a vector of values of the N basis functions at $\boldsymbol{x}$. It follows from (16) that

$$u(\boldsymbol{x}) \sim \mathcal{N}(\bar{u}(\boldsymbol{x}), \eta(\boldsymbol{x})), \tag{18}$$

where the mean value function $\bar{u}(\boldsymbol{x})$ and the variance function $\eta(\boldsymbol{x})$ are given by

$$\bar{u}(\boldsymbol{x}) = (\boldsymbol{\psi}(\boldsymbol{x}))^{\mathrm{H}}\bar{\boldsymbol{u}}, \qquad \eta(\boldsymbol{x}) = (\boldsymbol{\psi}(\boldsymbol{x}))^{\mathrm{H}}\boldsymbol{U}\boldsymbol{\psi}(\boldsymbol{x}). \tag{19}$$

The mean value function represents the prediction of the true state $u^{\mathrm{true}}(\boldsymbol{x})$, while the variance function quantifies the prediction error.

3.3. Covariance operator. In order to determine the posterior distribution of the state estimate, the covariance operator $k(\cdot, \cdot)$ needs to be specified. Indeed, this is a crucial aspect of our approach because it affects the estimate in (18). We propose a class of bilinear covariance operators of the form

$$k(w, v; \boldsymbol{\theta}) = \sum_{j=1}^{J} \theta_j k_j(w, v), \tag{20}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_J)$ is a vector of *hyperparameters* and $k_j(\cdot, \cdot), 1 \leq j \leq J$ are some given symmetric bilinear forms. The particular definition of the bilinear forms k_j will depend on the type of the PDE model as well as the numerical method used to solve the PDE. We will elaborate on this point later when we discuss our numerical results in section 5.

We now note from (10) that the posterior distribution of $(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})$, conditional on $\boldsymbol{\theta}$, obeys the normal distribution:

$$(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})|(\boldsymbol{\theta}, \sigma) \ \sim \ \mathcal{N}(0, \boldsymbol{D}(\boldsymbol{\theta}, \sigma)), \tag{21}$$

where the matrix $\boldsymbol{D}(\boldsymbol{\theta}, \sigma)$ has entries

$$D_{mq}(\boldsymbol{\theta}, \sigma) = \sum_{j=1}^{J} \theta_j k_j(\phi_m, \phi_q) + \sigma^2 \Delta_{mq}, \qquad m, q = 1, \ldots, M, \tag{22}$$

where $\Delta_{mq} = 1$ if $m = q$ and $\Delta_{mq} = 0$ if $m \neq q$. Let $p((\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})|(\boldsymbol{\theta}, \sigma))$ denote the probability density function of (21), and let us choose the hyperparameter vector $\boldsymbol{\theta}$ guided by the maximum likelihood principle, which by monotonicity is equivalent to log-maximum likelihood. The log likelihood function is given by

$$\begin{aligned} \log p((\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})|(\boldsymbol{\theta}, \sigma)) = &-\tfrac{1}{2}(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})^{\mathrm{H}} \boldsymbol{D}^{-1}(\boldsymbol{\theta}, \sigma)(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}}) \\ &- \tfrac{1}{2} \log(\det(\boldsymbol{D}(\boldsymbol{\theta}, \sigma))) - \tfrac{M}{2} \log(2\pi). \end{aligned} \tag{23}$$

We therefore propose to select $\boldsymbol{\theta}$ by seeking an approximate optimal solution to the log-likelihood optimization problem, yielding

$$(\boldsymbol{\theta}, \sigma) \approx \arg \max_{(\boldsymbol{\theta}', \sigma') \in \Theta} \log p((\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})|(\boldsymbol{\theta}', \sigma')). \tag{24}$$

Here the hyperparameter space Θ is a subset of $\mathbb{R}^{J+1}$ such that the resulting covariance operator k in (20) is positive-definite. The optimization problem (24) does not have a closed form solution, and it is generically nonconvex. We therefore seek approximate solutions of (24) in cases of interest. For example, when the hyperparameter space Θ is low-dimensional, simple grid generation/enumeration and function evaluation will be sufficient for our purposes.

3.4. Bayesian interpretation of the functional regression model. In this section, we present an alternative derivation of the functional regression model via Bayesian analysis. The matrix form of the Galerkin FE formulation (1) seeks $(\boldsymbol{u}^{\mathrm{o}}, \boldsymbol{s}^{\mathrm{o}}) \in \mathbb{C}^N \times \mathbb{C}^M$ such that

$$\boldsymbol{A}\boldsymbol{u}^{\mathrm{o}} = \boldsymbol{l}, \qquad \boldsymbol{s}^{\mathrm{o}} = \boldsymbol{C}^{\mathrm{H}} \boldsymbol{u}^{\mathrm{o}}, \tag{25}$$

where $\boldsymbol{u}^{\mathrm{o}} \in \mathbb{C}^N$ is the vector of degrees of freedom of the FE solution u^{o} and $C_{im} = c_m(\psi_i)$ for $i = 1, \ldots, N$ and $m = 1, \ldots M$. Similarly, the matrix form of the stochastic weak formulation (3) seeks $(\boldsymbol{u}, \boldsymbol{s}) \in \mathbb{C}^N \times \mathbb{C}^M$ such that

$$\boldsymbol{A}\boldsymbol{u} + \boldsymbol{g} = \boldsymbol{l}, \qquad \boldsymbol{s} = \boldsymbol{C}^{\mathrm{H}} \boldsymbol{u}, \tag{26}$$

where $\boldsymbol{g}$ is a *Gaussian random vector* with the prior distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{K})$. Note that the prior covariance matrix $\boldsymbol{K}$ has entries $K_{ij} = k(\psi_i, \psi_j)$.

To define the likelihood function, we compute the adjoint state vectors stored in the matrix $\boldsymbol{\Phi} \in \mathbb{C}^{N \times M}$ by solving the adjoint problem:

$$\boldsymbol{A}^{\mathrm{H}} \boldsymbol{\Phi} = -\boldsymbol{C}. \tag{27}$$

It follows from (25), (26), and (27) that

$$\boldsymbol{\Phi}^{\mathrm{H}} \boldsymbol{g} = \boldsymbol{\Phi}^{\mathrm{H}}(\boldsymbol{l} - \boldsymbol{A}\boldsymbol{u}) = \boldsymbol{\Phi}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{u}^{\mathrm{o}} - \boldsymbol{A}\boldsymbol{u}) = -\boldsymbol{C}^{\mathrm{H}}(\boldsymbol{u}^{\mathrm{o}} - \boldsymbol{u}) = \boldsymbol{s} - \boldsymbol{s}^{\mathrm{o}}. \tag{28}$$

We now substitute $\boldsymbol{s} = \boldsymbol{d} - \boldsymbol{\varepsilon}$ into (28) to arrive at the standard linear regression model:

$$\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}} = \boldsymbol{\Phi}^{\mathrm{H}} \boldsymbol{g} + \boldsymbol{\varepsilon}\ . \tag{29}$$

Since $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$, $(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})|\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{g}, \sigma^2 \boldsymbol{I})$. Therefore, the likelihood function, i.e., the probability density function of $\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}}$ given the vector $\boldsymbol{g}$, is

$$p((\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})|\boldsymbol{g}, \boldsymbol{\Phi}) = \frac{1}{(2\pi\sigma^2)^{M/2}} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}} - \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{g}\|^2\right), \tag{30}$$

where $\|\cdot\|$ denotes the Euclidean norm.

According to Bayes's theorem (see (2.3) on page 17 in [33]), the posterior distribution of $\boldsymbol{g}$ is given by

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}}, \qquad p(\boldsymbol{g}|(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}}), \boldsymbol{\Phi}) = \frac{p((\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})|\boldsymbol{\Phi}, \boldsymbol{g})p(\boldsymbol{g})}{p((\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})|\boldsymbol{\Phi})}, \tag{31}$$

where the marginal likelihood $p((\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})|\boldsymbol{\Phi})$is given by

$$p((\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})|\boldsymbol{\Phi}) = \int p((\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}})|\boldsymbol{\Phi}, \boldsymbol{g})p(\boldsymbol{g})d\boldsymbol{g}. \tag{32}$$

Writing only the terms from the likelihood and the prior, we obtain

$$\begin{aligned} p(\boldsymbol{g}|(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}}), \boldsymbol{\Phi}) &\propto \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}} - \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{g}\|^2\right) \exp\left(-\frac{1}{2}\boldsymbol{g}^{\mathrm{H}}\boldsymbol{K}^{-1}\boldsymbol{g}\right), \\ &\propto \exp\left(-\frac{1}{2}(\boldsymbol{g} - \bar{\boldsymbol{g}})^{\mathrm{H}}\boldsymbol{G}^{-1}(\boldsymbol{g} - \bar{\boldsymbol{g}})\right), \end{aligned} \tag{33}$$

where $\bar{\boldsymbol{g}}$ and $\boldsymbol{G}$ are given by

$$\bar{\boldsymbol{g}} = \frac{1}{\sigma^2}\boldsymbol{G}\boldsymbol{\Phi}(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}}), \qquad \boldsymbol{G} = \left(\frac{1}{\sigma^2}\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}} + \boldsymbol{K}^{-1}\right)^{-1}. \tag{34}$$

However, since this formula requires the inverse of a matrix of size $N \times N$, we will derive an equivalent expression which is more efficient to compute than (34).

We recall the Sherman–Morrison–Woodbury (SMW) formula for the matrix inversion

$$(\boldsymbol{Z} + \boldsymbol{U}\boldsymbol{W}\boldsymbol{V}^{\mathrm{H}})^{-1} = \boldsymbol{Z}^{-1} - \boldsymbol{Z}^{-1}\boldsymbol{U}(\boldsymbol{W}^{-1} + \boldsymbol{V}^{\mathrm{H}}\boldsymbol{Z}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{Z}^{-1}. \tag{35}$$

Using the SMW formula with $\boldsymbol{Z} = \boldsymbol{K}^{-1}, \boldsymbol{W} = \sigma^{-2}\boldsymbol{I}, \boldsymbol{U} = \boldsymbol{V} = \boldsymbol{\Phi}$ we obtain

$$\boldsymbol{G} = \boldsymbol{K} - \boldsymbol{K}\boldsymbol{\Phi}(\sigma^2\boldsymbol{I} + \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{K}. \tag{36}$$

It thus follows that

$$\begin{aligned} \boldsymbol{G}\boldsymbol{\Phi} &= \boldsymbol{K}\boldsymbol{\Phi} - \boldsymbol{K}\boldsymbol{\Phi}(\sigma^2\boldsymbol{I} + \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{\Phi}, \\ &= \boldsymbol{K}\boldsymbol{\Phi}(\sigma^2\boldsymbol{I} + \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{\Phi})^{-1}\left((\sigma^2\boldsymbol{I} + \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{\Phi}) - \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{\Phi}\right), \\ &= \sigma^2\boldsymbol{K}\boldsymbol{\Phi}(\sigma^2\boldsymbol{I} + \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{\Phi})^{-1}. \end{aligned} \tag{37}$$

Hence, we obtain from (34)–(37) that

$$\bar{\boldsymbol{g}} = \boldsymbol{K}\boldsymbol{\Phi}\boldsymbol{D}^{-1}(\boldsymbol{d} - \boldsymbol{s}^{\mathrm{o}}), \qquad \boldsymbol{G} = \boldsymbol{K} - \boldsymbol{K}\boldsymbol{\Phi}\boldsymbol{D}^{-1}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{K}, \tag{38}$$

where $\boldsymbol{D} = \sigma^2\boldsymbol{I} + \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{\Phi}$. In this way, we need to invert the matrix $\boldsymbol{D}$ of size $M \times M$. Since M is typically much smaller than N, the formula (38) is much more efficient to compute than the original one (34).

We note from the stochastic linear system (26) and the Gaussian property of $\boldsymbol{g}$ that the posterior distribution of $\boldsymbol{u}$ is a multivariate normal distribution

$$\boldsymbol{u}|(\boldsymbol{d}-\boldsymbol{s}^{\mathrm{o}}),\boldsymbol{\Phi}\sim\mathcal{N}(\bar{\boldsymbol{u}},\boldsymbol{U}), \tag{39}$$

where the posterior mean and the posterior covariance matrix are given by (17), namely,

$$\bar{\boldsymbol{u}}=\boldsymbol{A}^{-1}(\boldsymbol{l}-\bar{\boldsymbol{g}}),\qquad \boldsymbol{U}=\boldsymbol{A}^{-1}\boldsymbol{G}\boldsymbol{A}^{-\mathrm{H}}. \tag{40}$$

Substituting (38) into (40) yields

$$\begin{aligned}\bar{\boldsymbol{u}}&=\boldsymbol{A}^{-1}\left(\boldsymbol{l}-\boldsymbol{K}\boldsymbol{\Phi}(\sigma^2\boldsymbol{I}+\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{\Phi})^{-1}(\boldsymbol{d}-\boldsymbol{s}^{\mathrm{o}})\right),\\ \boldsymbol{U}&=\boldsymbol{A}^{-1}\left(\boldsymbol{K}-\boldsymbol{K}\boldsymbol{\Phi}(\sigma^2\boldsymbol{I}+\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{K}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{K}\right)\boldsymbol{A}^{-\mathrm{H}}.\end{aligned} \tag{41}$$

We observe that this result is exactly the same as the one presented in subsection 3.2.

4. Experimental design. The selection of inputs to obtain the observations is an *experimental design problem*. This problem can be posed as choosing the inputs to minimize the posterior covariance with some appropriate design criteria [32], such as A-optimality, D-optimality , E-optimality, and G-optimality [2, 30]. In this section, we develop an experimental design algorithm based on the criterion of minimizing a monotone function of the nonzero eigenvalues.

4.1. Optimizing experimental design. We assume that we are given a large set of L potential observation functionals $b_i(\cdot), 1\le i\le L$, each of which corresponds to a potential experiment to be carried out. Before actually performing any of these potential experiments, we face the following decision questions. Which experiment among the L potential experiments should be implemented first? Given the results of any previous experiments, which experiment should be chosen next? How many experiments are enough to provide an accurate prediction? These decision questions can be summed up as follows: how do we optimally choose a sequence of the actualized observation functionals $c_m(\cdot), 1\le m\le M$, among the potential observation functionals $b_i(\cdot), 1\le i\le L$?

It will be more convenient to state our experimental design problem in the matrix form instead of in the weak form. The expressions involving $\boldsymbol{K}(\Psi,\Psi), \boldsymbol{K}(\Psi,\Phi)$, $\boldsymbol{K}(\Phi,\Psi)$, etc., can be rather unwieldy, so we introduce a compact form of the notation by setting $\boldsymbol{K}=\boldsymbol{K}(\Psi,\Psi)$, $\boldsymbol{K}(\Psi,\Phi)=-\boldsymbol{K}\boldsymbol{A}^{-\mathrm{H}}\boldsymbol{C}$, $\boldsymbol{K}(\Phi,\Psi)=-\boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{K}$, and $\boldsymbol{K}(\Phi,\Phi)=\boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{K}\boldsymbol{A}^{-\mathrm{H}}\boldsymbol{C}$. It follows that the posterior covariance matrix given in (17) can be written as

$$\boldsymbol{U}=\boldsymbol{A}^{-1}\left(\boldsymbol{K}-\boldsymbol{K}\boldsymbol{A}^{-\mathrm{H}}\boldsymbol{C}(\sigma^2\boldsymbol{I}+\boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{K}\boldsymbol{A}^{-\mathrm{H}}\boldsymbol{C})^{-1}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{K}\right)\boldsymbol{A}^{-\mathrm{H}}. \tag{42}$$

We can also rewrite the posterior covariance matrix $\boldsymbol{U}$ as

$$\boldsymbol{U}(\boldsymbol{C})=\boldsymbol{U}_0-\boldsymbol{U}_0\boldsymbol{C}(\sigma^2\boldsymbol{I}+\boldsymbol{C}^{\mathrm{H}}\boldsymbol{U}_0\boldsymbol{C})^{-1}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{U}_0, \tag{43}$$

where $\boldsymbol{U}_0=\boldsymbol{A}^{-1}\boldsymbol{K}\boldsymbol{A}^{-\mathrm{H}}$ is the prior covariance matrix.

We now let $\boldsymbol{B}\in\mathbb{C}^{N\times L}$ be a matrix with entries $B_{nl}=b_l(\psi_n)$ for $n=1,\dots,N$ and $l=1,\dots,L$. Let M denote the number of experiments that we will implement, and consider the following set of matrices induced by the matrix $\boldsymbol{B}$:

$$\mathcal{M}_M(\boldsymbol{B}):=\{\boldsymbol{E}\in\mathbb{C}^{N\times M}\ :\ \text{each column of } \boldsymbol{E} \text{ is chosen among the columns of } \boldsymbol{B}\}. \tag{44}$$

In the experimental design problem, we seek an experiment matrix $\boldsymbol{C} \in \mathcal{M}_M(\boldsymbol{B})$ that minimizes the "size" of $\boldsymbol{U}(\boldsymbol{C})$, where size is measured in various ways in various design formulations such as the product of the eigenvalues of $\boldsymbol{U}(\boldsymbol{C})$, the largest eigenvalue of $\boldsymbol{U}(\boldsymbol{C})$, the sum of the eigenvalues of $\boldsymbol{U}(\boldsymbol{C})$, etc. In the interest of generality, let $h(\cdot) : \mathbb{R}^N \to \mathbb{R}$ be a given monotone function of the ordered eigenvalues $\lambda(\boldsymbol{U}) := (\lambda_1(\boldsymbol{U}), \ldots, \lambda_N(\boldsymbol{U}))$ of $\boldsymbol{U}$, and consider the following optimization problem:

$$\boldsymbol{C}^* = \arg \min_{\boldsymbol{E} \in \mathcal{M}_M(\boldsymbol{B})} h(\lambda(\boldsymbol{U}(\boldsymbol{E}))) \ , \tag{45}$$

where $\boldsymbol{U}(\boldsymbol{E})$ is the posterior covariance matrix as given in (43) for $\boldsymbol{E} = \boldsymbol{C}$. In other words, the optimal experiment design $\boldsymbol{C}^*$ is the matrix that minimizes a particular given monotone function of the eigenvalues of the posterior covariance matrix.

Note that the optimization problem (45) has a nonlinear (and typically nonconvex) objective function and that the set of feasible solutions is an exponentially large discrete set. Furthermore, the optimal observation matrix $\boldsymbol{C}^*$ is not hierarchical in the sense that increasing M might completely change the previously selected columns of $\boldsymbol{C}^*$. Because the optimal set of experiments might be very different for different values of M, one cannot compute the exact optimum by inductively amending the previous optimal set of experiments. For all of the above reasons it therefore is likely to be too computationally expensive to compute an exact optimum of (45). Below we develop a greedy iterative procedure for choosing the experiments based on the eigenvector of the largest eigenvalue of $\boldsymbol{U}(\boldsymbol{C})$.

4.2. A greedy algorithm for choosing the experiments. Here we develop a greedy algorithm for choosing the experiments. We first consider the case where there is no noise in the observation vector $\boldsymbol{d}$, i.e., $\sigma = 0$ for the Gaussian noise. With $\sigma = 0$ it follows that $\boldsymbol{u}$ must satisfy $\boldsymbol{C}^{\mathrm{H}}\boldsymbol{u} = \boldsymbol{d}$. Let us define $\mathcal{T} := \{\boldsymbol{u} : \boldsymbol{C}^{\mathrm{H}}\boldsymbol{u} = \boldsymbol{d}\}$ as the affine set containing the possible values of $\boldsymbol{u}$. Let us assume without loss of generality that the columns of $\boldsymbol{C}$ are linearly independent, so $\dim(\mathcal{T}) = N - M$. Recalling $\bar{\boldsymbol{u}} = \boldsymbol{A}^{-1}(\boldsymbol{l} - \bar{\boldsymbol{g}})$ from (17), it holds that $\bar{\boldsymbol{u}} \in \mathcal{T}$. The posterior covariance matrix (43) can be written as

$$\boldsymbol{U} = \boldsymbol{U}_0 - \boldsymbol{U}_0\boldsymbol{C}(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{U}_0\boldsymbol{C})^{-1}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{U}_0. \tag{46}$$

Let us also write the eigendecomposition of $\boldsymbol{U}$ as $\boldsymbol{U} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{H}}$ for a suitable orthonormal matrix of eigenvectors $\boldsymbol{Q}$ and diagonal matrix $\boldsymbol{\Lambda}$ of corresponding eigenvalues. Since $\boldsymbol{U}$ is positive semi-definite, i.e., $\boldsymbol{u}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{u} \geq 0$ for all $\boldsymbol{u}$, all eigenvalues of $\boldsymbol{U}$ are nonnegative, and we can partition the eigenvalues into those that are zero and those that are positive, and likewise partition the corresponding eigenvectors, and write

$$\boldsymbol{U} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{H}} = \begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{Q}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}^+ \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_1^{\mathrm{H}} \\ \boldsymbol{Q}_2^{\mathrm{H}} \end{bmatrix} = \boldsymbol{Q}_2\boldsymbol{\Lambda}^+\boldsymbol{Q}_2^{\mathrm{H}} \ , \tag{47}$$

where the columns of $\boldsymbol{Q}_1$ are eigenvectors corresponding to eigenvalues with value 0 and the columns of $\boldsymbol{Q}_2$ are eigenvectors corresponding to positive eigenvalues, and the diagonal matrix $\boldsymbol{\Lambda}^+$ has positive diagonal entries. In a slight change of notation, let us denote these entries as $\Lambda_i^+ > 0$ for $i = 1, \ldots, N - M$, and there is no loss of generality in assuming that $0 < \Lambda_1^+ \leq \Lambda_2^+ \leq \cdots \leq \Lambda_{N-M}^+$. It is straightforward to demonstrate that

$$\boldsymbol{u}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{u} = 0 \iff \boldsymbol{U}\boldsymbol{u} = 0 \iff \boldsymbol{u} \in \mathrm{range}(\boldsymbol{C}) \iff \boldsymbol{u} \in \mathrm{null}(\boldsymbol{Q}_2^{\mathrm{H}}) \ . \tag{48}$$

For the given observation vector $\boldsymbol{d}$, $\boldsymbol{u}$ is restricted to lie in the affine set $\mathcal{T}$, and we write $\boldsymbol{u}|\boldsymbol{d} \sim \mathcal{N}_{\mathcal{T}}(\bar{\boldsymbol{u}}, \boldsymbol{U})$ to designate that the posterior distribution of $\boldsymbol{u}$ is normally distributed on $\mathcal{T}$ with mean $\bar{\boldsymbol{u}}$ and covariance matrix $\boldsymbol{U}$. Even though $\boldsymbol{U}$ has no inverse, the probability density function of $\boldsymbol{u}$ on $\mathcal{T}$, which we denote as $p_{\mathcal{T}}(\boldsymbol{u}|\boldsymbol{d})$, is given by

$$p_{\mathcal{T}}(\boldsymbol{u}|\boldsymbol{d}) = \frac{1}{\pi^{(N-M)/2}\sqrt{\Pi_{i=1}^{N-M}\Lambda_i^+}} e^{-\left[\frac{1}{2}(\boldsymbol{u}-\bar{\boldsymbol{u}})^{\mathrm{H}}\boldsymbol{Q}_2(\boldsymbol{\Lambda}^+)^{-1}\boldsymbol{Q}_2^{\mathrm{H}}(\boldsymbol{u}-\bar{\boldsymbol{u}})\right]} . \tag{49}$$

Notice the precise way that the positive eigenvalues appear in (49).

Now let us suppose that we have already chosen M experiments from among the L experiments $b_i(\cdot)$, $i = 1, \ldots, L$, and let $\boldsymbol{C}$ denote the experiment matrix, and let us consider adding an additional experiment, which would be the $(M+1)$st experiment. Let this experiment be denoted as $\tilde{b}(\cdot)$, where $\tilde{b}(\cdot)$ is one of the L experiments $b_i(\cdot)$, $i = 1, \ldots, L$. In the matrix form, let $\tilde{\boldsymbol{b}}$ be the vector with entries $\tilde{b}(\psi_n)$ for $n = 1, \ldots, N$. We will assume that $\tilde{\boldsymbol{b}} \notin \text{range}(\boldsymbol{C})$, so that the new/updated experiment matrix $\tilde{\boldsymbol{C}} := [\boldsymbol{C}\ \tilde{\boldsymbol{b}}]$ has linearly independent columns. Then letting $\tilde{\boldsymbol{U}} := \boldsymbol{U}(\tilde{\boldsymbol{C}})$, we have from (46) that

$$\tilde{\boldsymbol{U}} = \boldsymbol{U}_0 - \boldsymbol{U}_0\tilde{\boldsymbol{C}}(\tilde{\boldsymbol{C}}^{\mathrm{H}}\boldsymbol{U}_0\tilde{\boldsymbol{C}})^{-1}\tilde{\boldsymbol{C}}^{\mathrm{H}}\boldsymbol{U}_0 . \tag{50}$$

It turns out that $\tilde{\boldsymbol{U}}$ simplifies to

$$\tilde{\boldsymbol{U}} = \boldsymbol{U} - \left(\frac{1}{\tilde{\boldsymbol{b}}^{\mathrm{H}}\boldsymbol{U}\tilde{\boldsymbol{b}}}\right)\boldsymbol{U}\tilde{\boldsymbol{b}}\tilde{\boldsymbol{b}}^{\mathrm{H}}\boldsymbol{U} . \tag{51}$$

(Equation (51) is derived by straightforward, but tedious, substitution using

$$\begin{aligned}
&(\tilde{\boldsymbol{C}}^{\mathrm{H}}\boldsymbol{U}_0\tilde{\boldsymbol{C}})^{-1} \\
&\quad = \begin{bmatrix} \boldsymbol{C}^{\mathrm{H}}\boldsymbol{U}_0\boldsymbol{C} & \boldsymbol{C}^{\mathrm{H}}\boldsymbol{U}_0\tilde{\boldsymbol{b}} \\ \tilde{\boldsymbol{b}}^{\mathrm{H}}\boldsymbol{U}_0\boldsymbol{C} & \tilde{\boldsymbol{b}}^{\mathrm{H}}\boldsymbol{U}_0\tilde{\boldsymbol{b}} \end{bmatrix}^{-1} \\
&\quad = \begin{bmatrix} (\boldsymbol{C}^{\mathrm{H}}\boldsymbol{U}_0\boldsymbol{C})^{-1} - \dfrac{(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{U}_0\boldsymbol{C})^{-1}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{U}_0\boldsymbol{b}\boldsymbol{b}^{\mathrm{H}}\boldsymbol{U}_0\boldsymbol{C}(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{U}_0\boldsymbol{C})^{-1}}{\alpha} & \dfrac{(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{U}_0\boldsymbol{C})^{-1}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{U}_0\boldsymbol{b}}{\alpha} \\ \dfrac{\boldsymbol{b}^{\mathrm{H}}\boldsymbol{U}_0\boldsymbol{C}(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{U}_0\boldsymbol{C})^{-1}}{\alpha} & \dfrac{1}{\alpha} \end{bmatrix},
\end{aligned}$$

where $\alpha := \tilde{\boldsymbol{b}}^{\mathrm{H}}\boldsymbol{U}\tilde{\boldsymbol{b}}$, and using the block form of the inverse above in (50) and simplifying terms. Note that the assumption that $\tilde{\boldsymbol{b}} \notin \text{range}(\boldsymbol{C})$ implies via (48) that $\alpha = \tilde{\boldsymbol{b}}^{\mathrm{H}}\boldsymbol{U}\tilde{\boldsymbol{b}} \neq 0$ so the above objects are all well-defined.)

Let us now compare the ordered eigenvalues of $\boldsymbol{U}$ and $\tilde{\boldsymbol{U}}$, which we denote, respectively, as the arrays

$$\begin{aligned}
\boldsymbol{\lambda} &:= (\ \lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_N\ ) , \\
\tilde{\boldsymbol{\lambda}} &:= (\ \tilde{\lambda}_1 \quad \tilde{\lambda}_2 \quad \cdots \quad \tilde{\lambda}_N\ ) .
\end{aligned}$$

From the linear independence assumption regarding the columns of $\boldsymbol{C}$ and $\tilde{\boldsymbol{C}}$ it follows using (48) that $\lambda_1 = \cdots = \lambda_M = 0$ and $\lambda_{M+1} > 0$ and also that $\tilde{\lambda}_1 = \cdots = \tilde{\lambda}_{M+1} = 0$ and $\tilde{\lambda}_{M+2} > 0$. Similar to (47), we will write

$$\tilde{\boldsymbol{U}} = \begin{bmatrix} \tilde{\boldsymbol{Q}}_1 & \tilde{\boldsymbol{Q}}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \tilde{\boldsymbol{\Lambda}}^+ \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{Q}}_1^{\mathrm{H}} \\ \tilde{\boldsymbol{Q}}_2^{\mathrm{H}} \end{bmatrix} = \tilde{\boldsymbol{Q}}_2\tilde{\boldsymbol{\Lambda}}^+\tilde{\boldsymbol{Q}}_2^{\mathrm{H}} , \tag{52}$$

where the columns of $\tilde{\boldsymbol{Q}}_1$ are eigenvectors corresponding to eigenvalues with value 0 and the columns of $\tilde{\boldsymbol{Q}}_2$ are eigenvectors corresponding to positive eigenvalues, and the diagonal matrix $\tilde{\boldsymbol{\Lambda}}^+$ has positive diagonal entries. As with (52), we denote these entries as $\tilde{\Lambda}_i^+ > 0$ for $i = 1, \ldots, N-M-1$, and there is no loss of generality in assuming that $0 < \tilde{\Lambda}_1^+ \leq \tilde{\Lambda}_2^+ \leq \cdots \leq \tilde{\Lambda}_{N-M-1}^+$. Therefore we have

$$\begin{aligned}
\boldsymbol{\lambda} &= (\ \ 0 \quad 0 \quad \cdots \quad 0 \quad \Lambda_1^+ \quad \Lambda_2^+ \quad \Lambda_3^+ \quad \cdots \quad \Lambda_{N-M}^+ \ \ ) \ , \\
\tilde{\boldsymbol{\lambda}} &= (\ \ 0 \quad 0 \quad \cdots \quad 0 \quad 0 \quad\ \ \tilde{\Lambda}_1^+ \quad \tilde{\Lambda}_2^+ \quad \cdots \quad \tilde{\Lambda}_{N-M-1}^+ \ \ ) \ ,
\end{aligned}$$

where there are M leading zeros in $\boldsymbol{\lambda}$ and $M+1$ leading zeros in $\tilde{\boldsymbol{\lambda}}$ above.

Observe from (51) that $\tilde{\boldsymbol{U}}$ is a rank-1 modification of $\boldsymbol{U}$. As a consequence, it follows from the interlacing eigenvalue theorem (see Theorem 8.1.8 of [15] and discussion therein for example) that the eigenvalues of $\boldsymbol{U}$ and $\tilde{\boldsymbol{U}}$ are *interlaced*, namely,

$$\tilde{\lambda}_1 \leq \lambda_1 \leq \tilde{\lambda}_2 \leq \lambda_2 \leq \cdots \leq \lambda_{N-1} \leq \tilde{\lambda}_N \leq \lambda_N \ . \tag{53}$$

It therefore follows that

$$\Lambda_1^+ \leq \tilde{\Lambda}_1^+ \leq \Lambda_2^+ \leq \tilde{\Lambda}_2^+ \leq \cdots \leq \Lambda_{N-M-1}^+ \leq \tilde{\Lambda}_{N-M-1}^+ \leq \Lambda_{N-M}^+ \ .$$

This chain of inequalities is significant in that it provides lower bounds on the values of the updated eigenvalues $\tilde{\Lambda}_i^+$ for $i = 1, \ldots, N-M-1$:

$$\tilde{\Lambda}_i^+ \geq \Lambda_i^+ \ , \text{for } i = 1, \ldots, N-M-1 \ . \tag{54}$$

Note that this result is true for *any* newly added experiment $\tilde{\boldsymbol{b}}$ satisfying the condition that $\tilde{\boldsymbol{C}} := [\boldsymbol{C} \ \tilde{\boldsymbol{b}}]$ has linearly independent columns.

Let us now consider if there might be an experiment $\tilde{\boldsymbol{b}}$ for which (54) holds at equality for all $i = 1, \ldots, N-M-1$, thus *simultaneously* providing the lowest value of the new eigenvalues $\tilde{\Lambda}_i^+$ for $i = 1, \ldots, N-M-1$. Let $\boldsymbol{q}_N$ be an eigenvector of $\boldsymbol{U}$ corresponding to the largest eigenvalue of $\boldsymbol{U}$, namely, $\lambda_N = \Lambda_{N-M}^+$. Suppose that we are able to set the new experiment $\tilde{\boldsymbol{b}}$ so that $\tilde{\boldsymbol{b}} = \boldsymbol{q}_N$. In this case it follows from (51) and (47) that

$$\begin{aligned}
\tilde{\boldsymbol{U}} &= \boldsymbol{U} - \left(\frac{1}{\boldsymbol{q}_N^{\mathrm{H}} \boldsymbol{U} \boldsymbol{q}_N} \right) \boldsymbol{U} \boldsymbol{q}_N \boldsymbol{q}_N^{\mathrm{H}} \boldsymbol{U} \\
&= \boldsymbol{Q} \boldsymbol{\Lambda} \boldsymbol{Q}^{\mathrm{H}} - \left(\frac{1}{\boldsymbol{q}_N^{\mathrm{H}} \boldsymbol{Q} \boldsymbol{\Lambda} \boldsymbol{Q}^{\mathrm{H}} \boldsymbol{q}_N} \right) \boldsymbol{Q} \boldsymbol{\Lambda} \boldsymbol{Q}^{\mathrm{H}} \boldsymbol{q}_N^{\mathrm{H}} \boldsymbol{q}_N \boldsymbol{Q} \boldsymbol{\Lambda} \boldsymbol{Q}^{\mathrm{H}} \\
&= \boldsymbol{Q} \boldsymbol{\Lambda} \boldsymbol{Q}^{\mathrm{H}} - \lambda_N \boldsymbol{q}_N \boldsymbol{q}_N^{\mathrm{H}} \ ,
\end{aligned}$$

and it follows that the columns of $\boldsymbol{Q}$ comprise a set of orthonormal eigenvectors of $\tilde{\boldsymbol{U}}$ and that the nonzero eigenvalues of $\tilde{\boldsymbol{U}}$ are precisely Λ_i^+ for $i = 1, \ldots, N-M-1$. Therefore the nonzero eigenvalues of $\tilde{\boldsymbol{U}}$ are in fact $\tilde{\Lambda}_i^+ = \Lambda_i^+$ for $i = 1, \ldots, N-M-1$. We therefore see that when $\tilde{\boldsymbol{b}} = \boldsymbol{q}_N$, (54) is satisfied at equality for all $i = 1, \ldots, N-M-1$. In light of the lower bounds given in (54) it holds that $\tilde{\boldsymbol{b}} = \boldsymbol{q}_N$ is the ideal next experiment as it simultaneously minimizes *any* (and hence *every*) monotone function $h(\cdot)$ of the nonzero eigenvalues of $\tilde{\boldsymbol{U}}$.

From a practical point of view, we would not expect there to be an experiment $b_i(\cdot)$ whose corresponding $\tilde{\boldsymbol{b}}_i$ vector satisfies $\tilde{\boldsymbol{b}}_i = \boldsymbol{q}_N$ (or $\tilde{\boldsymbol{b}}_i = -\boldsymbol{q}_N$). Instead, it makes good intuitive sense to seek an experiment $\tilde{\boldsymbol{b}}_i$ that makes the smallest angle with $\pm \boldsymbol{q}_N$.

This leads to the criterion to choose the next experiment by finding the index i^*_{next} such that

$$i^*_{\text{next}} = \arg \max_{i \in \{1,\dots,L\}} \frac{|\boldsymbol{b}_i^{\mathrm{H}} \boldsymbol{q}_N|}{\|\boldsymbol{b}_i\|} .$$

This is formalized in the following algorithm.

0. (Initial values.) $M \leftarrow 0$. Given the matrix $\boldsymbol{K}$, set the initial posterior covariance matrix $\boldsymbol{U}$ to be the prior covariance matrix:
$$\boldsymbol{U} = \boldsymbol{A}^{-1} \boldsymbol{K} \boldsymbol{A}^{-\mathrm{H}}. \tag{55}$$
1. (Compute eigenvector of largest eigenvalue.) Solve the eigenvalue problem $\boldsymbol{U}\boldsymbol{v} = \lambda \boldsymbol{v}$ to obtain the maximum eigenvalue $\lambda_{\max}$ and the corresponding eigenvector $\boldsymbol{v}_{\max}$.
2. (Determine the next experiment to add.) Find an index i^*_{next} such that
$$i^*_{\text{next}} = \arg \max_{i \in \{1,\dots,L\}} \frac{|\boldsymbol{b}_i^{\mathrm{H}} \boldsymbol{v}_{\max}|}{\|\boldsymbol{b}_i\|}, \tag{56}$$
and append $\boldsymbol{b}_{i^*_{\text{next}}}$ to the current matrix $\boldsymbol{C}$ to form the updated matrix $\boldsymbol{C} \leftarrow [\boldsymbol{C} \;\; \boldsymbol{b}_{i^*_{\text{next}}}]$. Perform the new experiment to obtain the experiment outcome value d_{M+1}. Update the number of experiments: $M \leftarrow M+1$. If $M \geq M_{\max}$, stop. (Optional: apply other stopping criterion to determine whether to stop.)
3. (Update posterior covariance matrix.) Given the updated matrix $\boldsymbol{C}$, calculate the updated posterior covariance matrix $\boldsymbol{U}$ as
$$\boldsymbol{U} = \boldsymbol{A}^{-1} \left(\boldsymbol{K} - \boldsymbol{K}\boldsymbol{A}^{-\mathrm{H}} \boldsymbol{C} (\sigma^2 \boldsymbol{I} + \boldsymbol{C}^{\mathrm{H}} \boldsymbol{A}^{-1} \boldsymbol{K} \boldsymbol{A}^{-\mathrm{H}} \boldsymbol{C})^{-1} \boldsymbol{C}^{\mathrm{H}} \boldsymbol{A}^{-1} \boldsymbol{K} \right) \boldsymbol{A}^{-\mathrm{H}}. \tag{57}$$
4. (Optional: update hyperparameters.) Optionally, given the updated experiments and their outcomes, reset the hyperparameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_J)$ of the covariance operator $k(\cdot, \cdot, \boldsymbol{\theta})$ in (20) by approximately solving the likelihood maximization problem (24) as described in section 3.3.
5. Go to step 1.

While the algorithm is best conceptualized for the case when $\sigma = 0$, we have found that in practice it also works well for $\sigma > 0$, using the intuitive notion that for small values of σ the mathematics developed and used in this section remains approximately valid. Indeed, as we will see in section 5, the algorithm works very well for $\sigma > 0$ in the applications of interest herein.

Notice that the computational complexity of this algorithm scales very favorably with the number of possible experiments L. Specifically, the algorithm needs to evaluate L scalar products $\boldsymbol{b}_i^{\mathrm{H}} \boldsymbol{v}_{\max}, i = 1, \dots, L$ only once at each of the M steps. Therefore, the algorithm can accommodate a large list of possible experiments without too much computational effort so long as M is relatively small.

Let us now discuss the optional step 4 in the above algorithm. By our way of choosing the covariance operator discussed in subsection 3.4, the matrix $\boldsymbol{K}$ depends on the observations. Ultimately, the prior covariance matrix $\boldsymbol{K}$ plays an important role in the posterior covariance matrix $\boldsymbol{U}$ as it is given by (57). By updating the hyperparameters when a new experiment is chosen, we hope to further reduce the "size" of the posterior covariance matrix $\boldsymbol{U}$ through a good choice of the prior covariance matrix $\boldsymbol{K}$. In all numerical examples presented in the next section, we carry out the optional step 4 in our greedy algorithm.

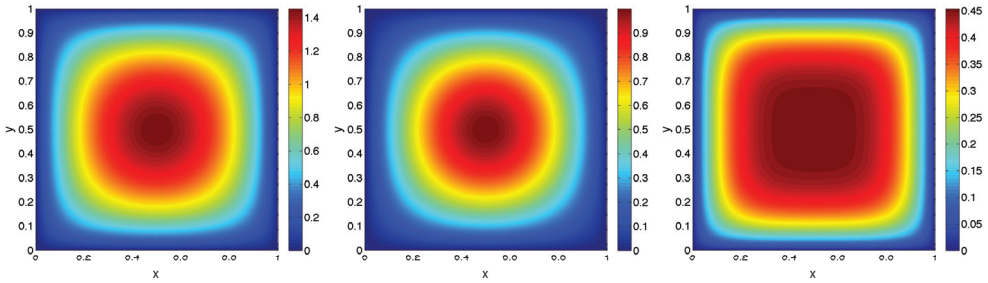


FIG. 1. *Plot of* u^{o} *(left),* u^{true} *(middle), and their difference* $u^{\mathrm{o}} - u^{\mathrm{true}}$ *(right).*

5. Numerical results. We will demonstrate the proposed method on numerical examples from a variety of second-order linear PDEs such as heat equations, convection-diffusion equations, and Helmholtz equations. In order to verify the performance of our method, we will specify the true state and generate the observations by adding Gaussian noise to the true outputs. Since our examples involve second-order PDEs, we will consider the covariance operator of the form

$$k(w, v; \boldsymbol{\theta}) = \theta_1 \int_\Omega wv d\boldsymbol{x} + \theta_2 \int_\Omega \nabla w \cdot \nabla v d\boldsymbol{x} \ , \tag{58}$$

which is related to the $H^1(\Omega)$ inner product. Note that the covariance operator is positive-semidefinite when the hypeparameters are nonnegative.

5.1. Heat conduction example. We consider the following PDE model for heat conduction in a unit square domain:

$$-\Delta u^{\mathrm{o}} = f^{\mathrm{o}} \quad \text{in } \Omega \quad \text{and} \quad u^{\mathrm{o}} = 0 \quad \text{on } \partial\Omega, \tag{59}$$

where $\Omega \equiv (0,1) \times (0,1)$ and $f^{\mathrm{o}} = 2\pi^2$. The Garlerkin FE formulation of this model problem has the bilinear form and the linear functional

$$a(w, v) = \int_\Omega \nabla w \cdot \nabla v dx dy, \qquad \ell(v) = \int_\Omega f^{\mathrm{o}} v dx dy \qquad \forall w, v \in V, \tag{60}$$

respectively. Here V is a FE approximation space of piecewise linear polynomials defined on a finite element mesh of 5,000 elements. Figure 1 shows a plot of the finite element solution of our PDE model (59). The possible observation functionals are specified as

$$b_i(v) = \int_\Omega \exp\left(-\frac{(x - x_i)^2 + (y - y_i)^2}{0.25^2}\right) v dx dy, \quad i = 1, \ldots, L, \tag{61}$$

where $(x_i, y_i), i = 1, \ldots, L = 961$ are the points of a 31×31 uniform grid as shown in Figure 2. These spatial points represent possible measurement locations at which we obtain the observations. We assume that the observations are noise-free.

Furthermore, we assume that the true state is $u^{\mathrm{true}} = \sin(\pi x)\sin(\pi y)$ as depicted in Figure 1. We note that the true state u^{true} is the solution of the PDE model (59) in which the source term $f^{\mathrm{o}} = 2\pi^2$ is replaced with $f^{\mathrm{true}} = 2\pi^2 \sin(\pi x)\sin(\pi y)$. Therefore, the source of uncertainty in our PDE model (59) comes from the source term. Due to this uncertainty, as shown in Figure 1, the solution u^{o} of the PDE model

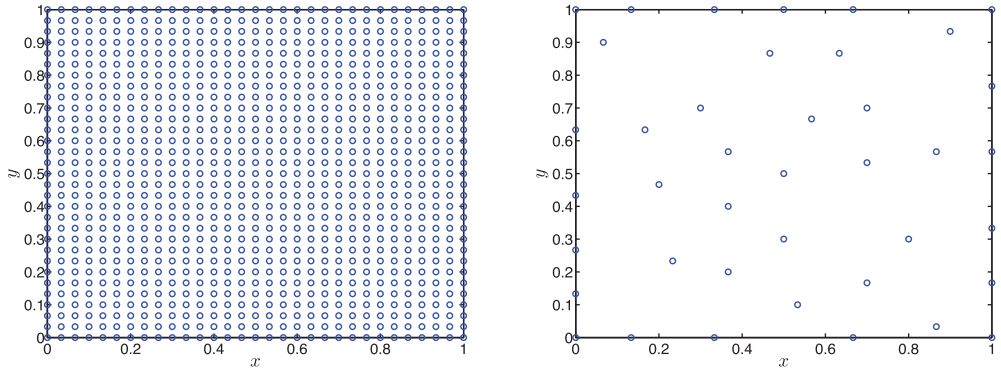


FIG. 2. *Potential measurement points (left) and selected measurement locations (right).*

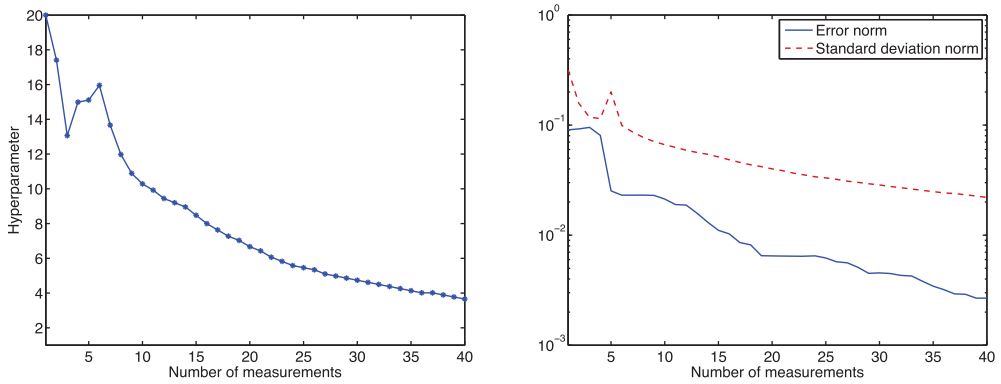


FIG. 3. *The first hyperparameter θ_1 (left) and the error norm and the standard deviation norm (right) as a function of M. Note that the maximum likelihood approach yields $\theta_2 = 0$ and $\sigma = 0$.*

(59) is significantly different from the true state u^{true}. In what follows, we combine the PDE model (59) with observations to improve the prediction of the true state by using the functional regression method.

The hyperparameters $(\theta_1, \theta_2, \sigma)$ are chosen to maximize the log marginal likelihood on a uniform grid $50 \times 40 \times 40$ of the hyperparameter domain $\Theta \equiv [0.1, 25] \times [0, 1] \times [0, 0.1]$. We next pursue the greedy algorithm to determine the measurement locations, which are shown in Figure 2. We observe that the selected measurement locations are distributed over the whole domain. Figure 3 shows the hyperparameter θ_1 and the $L^2(\Omega)$ norm of the standard function $\|\sqrt{\eta}\|_\Omega$ as well as the error norm $\|u^{\text{true}} - \bar{u}\|_\Omega$ as a function of the number of measurements. It is interesting to note that the maximum likelihood approach yields $\theta_2 = 0$ and $\sigma = 0$. This means that our bilinear covariance operator k is collinear with the L^2 inner product and that $\sigma = 0$ coincides with the noise-free observations. Figure 4 shows the posterior variance function and the selected measurement points at each iteration of our greedy algorithm. We observe that the variance reduces as we increase the number of measurements and that the measurement points are typically selected near the maximum peak of the posterior variance function.

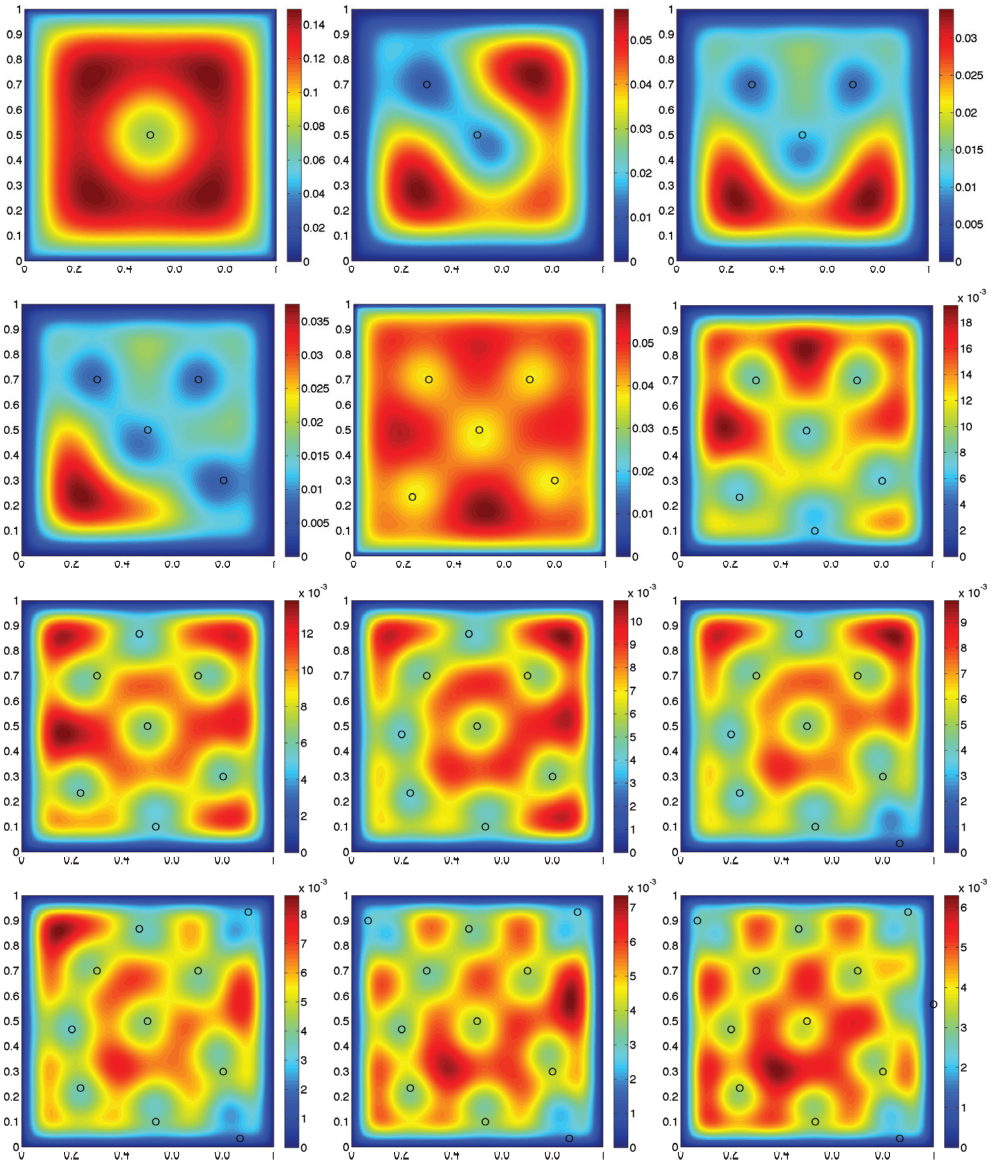


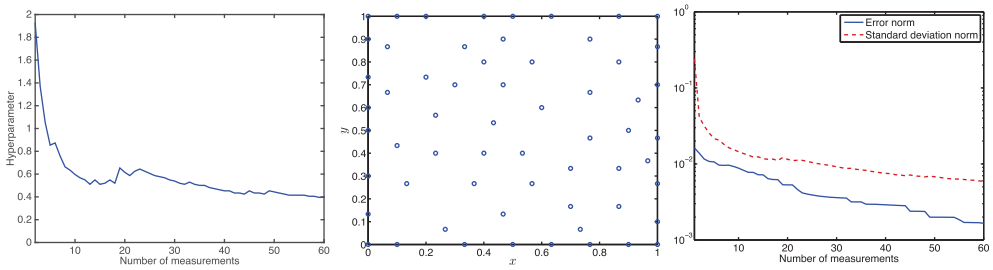
FIG. 4. *The variance function $\eta(\boldsymbol{x})$ and the measurement points for $M = 1, \ldots, 12$, where the order goes from from left to right and top to bottom.*

5.2. Convection-diffusion example. We next consider a convection-diffusion problem on a unit square $\Omega \equiv (0,1) \times (0,1)$:

$$-\Delta u^{\mathrm{o}} + \boldsymbol{c}^{\mathrm{o}} \cdot \nabla u^{\mathrm{o}} = f^{\mathrm{o}} \quad \text{in } \Omega \quad \text{and} \quad u^{\mathrm{o}} = 0 \quad \text{on } \partial\Omega, \tag{62}$$

where $\boldsymbol{c}^{\mathrm{o}} = (10, 10)$ and $f^{\mathrm{o}} = 10$. The Garlerkin FE formulation of the above PDE model is derived with the bilinear form and the linear functional

$$a(w,v) = \int_\Omega \nabla w \cdot \nabla v dx dy + \int_\Omega (\boldsymbol{c}^{\mathrm{o}} \cdot \nabla w) v dx dy, \qquad \ell(v) = \int_\Omega f^{\mathrm{o}} v dx dy \qquad \forall w, v \in V, \tag{63}$$



(a) (b) (c)

FIG. 5. *The hyperparameter θ_1 versus M* (a), *the measurement locations* (b), *and the error and the standard deviation versus M* (c). *Note that $\theta_2 = 0$ and $\sigma = 0$.*

where V is a FE approximation space of piecewise polynomials of degree $p = 3$ and is defined on a mesh of 800 elements. The possible measurement points and observation functionals are the same as in the previous example. We assume that the observations are noise-free.

To assess our method, we assume that the true state u^{true} is the solution of the PDE model (62) in which the convective velocity $\boldsymbol{c}^{\text{o}} = (10, 10)$ is replaced with $\boldsymbol{c}^{\text{true}} = (15, 15)$. Hence, the source of uncertainty in our PDE model (62) comes from the convective velocity. In this example, the uncertainty is present in the bilinear form a, whereas in the previous example the uncertainty is present in the linear functional ℓ. This example serves to demonstrate the effectiveness of our method on cases where uncertainty affects PDE operators.

The covariance operator has the form (58), where (θ_1, θ_2) and σ are computed by maximizing the likelihood on a uniform grid $200 \times 50 \times 50$ of the domain $\Theta \equiv [0.1, 10] \times [0, 1] \times [0, 0.1]$. Note that $\theta_2 = 0$ and $\sigma = 0$, while θ_1 is shown in Figure 5(a). Figure 5 also shows the measurement locations and the norm of the standard function $\|\sqrt{\eta}\|_\Omega$ as well as the error norm $\|u^{\text{true}} - \bar{u}\|_\Omega$ as a function of the number of measurements. As expected, both the error norm and the standard deviation norm decrease as M increases. Hence, increasing the number of measurements reduces the uncertainty. This can be seen more clearly in Figure 6, which shows the absolute error function and the standard deviation function for $M = 20, 40$, and 60. We observe that the error near the top right conner of the domain is much larger than the one near the bottom left conner of the domain, while the standard deviation is spread over the physical domain much more evenly than the error.

5.3. A Helmholtz example. We consider the sound-hard scattering of an incident plane wave $u^{\text{inc}} = \exp(ikx)$ by a circle of radius $a = 1$, where k is the wave number. The scattered field satisfies the exterior Helmholtz problem

$$
\begin{aligned}
\Delta u^{\text{true}} + k^2 u^{\text{true}} &= 0 \quad \text{in } \mathbb{R}^2 \backslash \Omega_{\text{c}}, \\
\nabla u^{\text{true}} \cdot \boldsymbol{n} + \nabla u^{\text{inc}} \cdot \boldsymbol{n} &= 0 \quad \text{on } \Gamma_{\text{c}}, \\
\lim_{r \to \infty} \sqrt{r} \left(\frac{\partial u^{\text{true}}}{\partial r} - iku^{\text{true}} \right) &= 0,
\end{aligned}
\tag{64}
$$

where Ω_{c} is the domain of the unit circle and Γ_{c} is the boundary of the unit circle. The last condition is known as Sommerfeld radiation condition. It is known that the

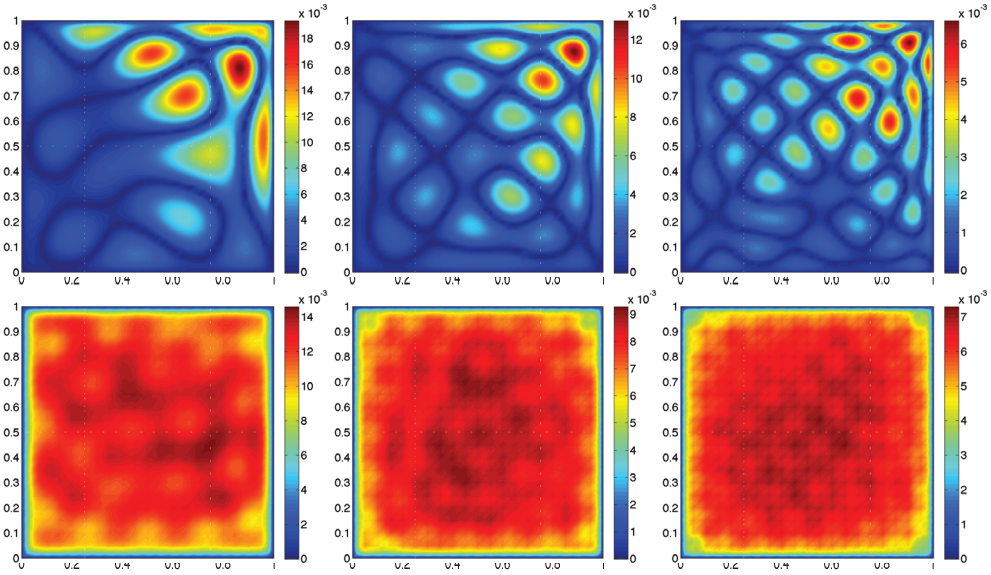


FIG. 6. *The absolute error function* $|\bar{u}(\boldsymbol{x}) - u^{\text{true}}(\boldsymbol{x})|$ *(top row) and the standard deviation function* $\sqrt{\eta(\boldsymbol{x})}$ *(bottom row) for* $M = 20$ *(left),* $M = 40$ *(middle), and* $M = 60$ *(right).*

solution of the above Helmholtz problem has the analytical form

$$u^{\text{true}}(x,y) = -\frac{J_0'(ka)}{H_0^{1\prime}(ka)} H_0^1(kr) - 2\sum_{n=1}^{\infty} i^n \frac{J_n'(ka)}{H_n^{1\prime}(ka)} H_n^1(kr)\cos(n\theta), \tag{65}$$

where (r,θ) are polar coordinates of (x,y), and J_n and H_n^1 are the Bessel function and the Hankel function of the first kind, respectively [17]. Note that the prime denotes the derivative of a function with respect to its argument.

Since the scatter is simple the solution of the exterior Helmholtz problem (64) is known analytically. For more complicated scatterers, however, the problem is usually solved by using a numerical method. In the case of a finite element method, the unbounded domain need to be truncated to a bounded domain and the Sommerfeld radiation condition should be replaced with a suitable absorbing boundary condition at the exterior boundary of the truncated domain. This gives rise to the following PDE model:

$$\begin{aligned} \Delta u^{\text{o}} + k^2 u^{\text{o}} &= 0 \quad \text{in } \Omega, \\ \nabla u^{\text{o}} \cdot \boldsymbol{n} + \nabla u^{\text{inc}} \cdot \boldsymbol{n} &= 0 \quad \text{on } \Gamma_{\text{c}}, \\ \nabla u^{\text{o}} \cdot \boldsymbol{n} - iku^{\text{o}} &= 0 \quad \text{on } \Gamma_{\text{o}}, \end{aligned} \tag{66}$$

where Ω is the truncated domain as shown in Figure 7 and Γ_{o} is the exterior boundary of the truncated domain. Note here that we approximate the Sommerfeld radiation condition by the first-order absorbing boundary condition. The weak formulation of the PDE model (66) is given by

$$a(u^{\text{o}}, v) = \ell(v) \qquad \forall v \in X, \tag{67}$$

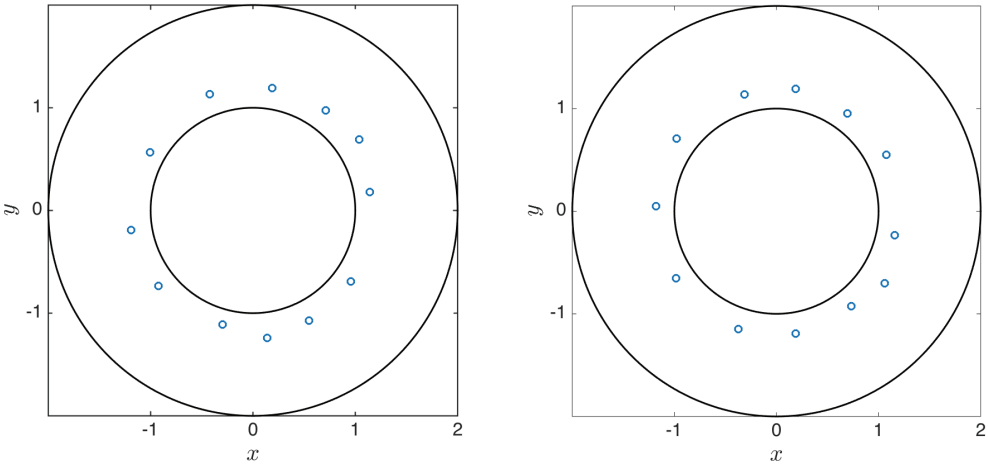


FIG. 7. *The measurement locations obtained using the greedy algorithm for two different noise levels:* $\sigma_{\text{noise}} = 0.005$ *(left) and* $\sigma_{\text{noise}} = 0.01$ *(right).*

where, for all $w, v \in X$,

$$
\begin{aligned}
a(w,v) &= \int_{\Omega} (\nabla w \cdot \nabla v - k^2 wv) dxdy - \int_{\Gamma_o} ikwvdxdy, \\
\ell(v) &= -\int_{\Gamma_c} (\nabla u^{\text{inc}} \cdot \boldsymbol{n}) v dxdy,
\end{aligned}
\tag{68}
$$

and $X \equiv H^1(\Omega)$. In practice, we replace X with a finite-dimensional space V which is a FE approximation space of piecewise polynomials of degree $p = 3$ and is defined on a mesh of 800 elements. This FE discretization is fine enough such that the numerical error is negligible.

This example serves to demonstrate our approach when the uncertainty sources originate from the domain truncation and the boundary condition. We aim to improve the prediction of the scattered field u^{true} by combining the PDE model (66) with observations of the scattered field. Of particular interest is the value of the scattered field on the unit circle, as it ultimately determines the radar cross section. The possible observation functionals are specified as

$$
b_i(v) = \int_{\Omega} \exp\left(-\frac{(x-x_i)^2 + (y-y_i)^2}{0.25^2}\right) v dxdy, \quad i = 1, \ldots, L, \tag{69}
$$

where $(x_i, y_i), i = 1, \ldots, L = 3720$ coincide with the mesh points. We will consider noisy observations which are obtained by adding Gaussian noise of zero mean and standard deviation σ_{noise} to the noise-free observations.

The hyperparameters $(\theta_1, \theta_2, \sigma)$ are computed by maximizing the likelihood on a uniform grid $100 \times 50 \times 50$ of the domain $\Theta \equiv [0.01, 1] \times [0, 1] \times [0, 0.02]$. We present in Table 1 the hyperparameters for $M = 4, 8, 12$. We observe that σ is quite close to σ_{noise} for $M = 12$. Hence, the likelihood maximization approach yields a good estimate of the noise level when the number of measurements is sufficient. We present in Figure 7 the selected measurement locations for two different noise levels in the observations: $\sigma_{\text{noise}} = 0.005$ and $\sigma_{\text{noise}} = 0.01$. It is interesting to see that in all cases

TABLE 1
The hyperparameters $(\theta_1, \theta_2, \sigma)$ for the two different noise levels in the observations.

	$\sigma_{\text{noise}} = 0.005$			$\sigma_{\text{noise}} = 0.01$		
M	θ_1	θ_2	σ	θ_1	θ_2	σ
4	0.0300	0.0000	0.0000	0.0100	0.0000	0.0125
8	0.0282	0.0000	0.0081	0.0400	0.0000	0.0070
12	0.0455	0.0000	0.0049	0.0336	0.0000	0.0086

the measurement locations are quite close to the unit circle and distributed quite uniformly along the unit circle.

We next show the true state, the mean prediction, and the 95% confidence region (shaded area) for $\sigma_{\text{noise}} = 0.005$ in Figure 8, and for $\sigma_{\text{noise}} = 0.01$ in Figure 9. Here the 95% confidence region is an area bounded by the mean prediction plus and minus two times the standard deviation function. We see that increasing M can improve the accuracy of the mean prediction for $\sigma_{\text{noise}} = 0.005$. However, for $\sigma_{\text{noise}} = 0.01$, the mean prediction does not get better as we increase M. This means that when σ_{noise} is equal or greater than 0.01, the observations do not add a positive contribution to the prediction of the true state. For $\sigma_{\text{noise}} = 0.005$ the 95% confidence region shrinks quite rapidly as M increases, whereas for $\sigma_{\text{noise}} = 0.01$ the 95% confidence region does not shrink as M increases. Therefore, when the observations become too noisy, adding more observations does not help improve the prediction. Clearly, the mean prediction and error estimate get worse as the noise level increases.

6. Conclusions. We have presented a statistical method that combines a linear PDE model with observations to predict the state of a physical problem. First, a random functional is introduced into the PDE model to account for various sources of uncertainty in the model. This random functional is posed as a Gaussian process with zero mean and prior covariance operator. Next, a linear regression model for the Gaussian functional is derived by utilizing the adjoint states and the observations. This regression model allows us to compute the posterior distribution of both the Gaussian functional and the state estimate. A key ingredient of our method is the prior covariance operator of the Gaussian functional. We propose a class of affine bilinear forms for the prior covariance operator and determine the associated hyperparameters based on the observations. Furthermore, we devise a greedy algorithm to select observations among a large number of possible measurements.

We would like to extend our approach to nonlinear PDEs. Two new challenges arise in the nonlinear case. First, although the functional g is Gaussian, the state u is no longer a Gaussian random field because of the nonlinearity. Second, the nonlinearity prevents a direct link between the adjoint solutions and the observations, thereby complicating the regression model. Our research will focus on addressing these issues.

Appendix A. Relationship with variational data assimilation. Let $\boldsymbol{\beta} = \boldsymbol{D}^{-1}(\boldsymbol{d} - \boldsymbol{s}^{\text{o}})$ and $\bar{\boldsymbol{q}} = \boldsymbol{\Phi}\boldsymbol{\beta}$. We can easily show that $(\bar{\boldsymbol{u}}, \bar{\boldsymbol{q}}, \boldsymbol{\beta})$ is the optimal solution of the following least squares minimization problem

$$\begin{aligned} \min_{(\boldsymbol{z},\boldsymbol{w},\boldsymbol{\gamma}) \in \mathbb{C}^N \times \mathbb{C}^N \times \mathbb{C}^M} \quad & \tfrac{1}{2}\boldsymbol{w}^{\text{H}}\boldsymbol{K}\boldsymbol{w} + \tfrac{1}{2}\sigma^2\boldsymbol{\gamma}^{\text{H}}\boldsymbol{\gamma} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{z} + \boldsymbol{K}\boldsymbol{w} = \boldsymbol{l}, \\ & \boldsymbol{C}^{\text{H}}\boldsymbol{z} + \sigma^2\boldsymbol{\gamma} = \boldsymbol{d}\,. \end{aligned} \tag{70}$$

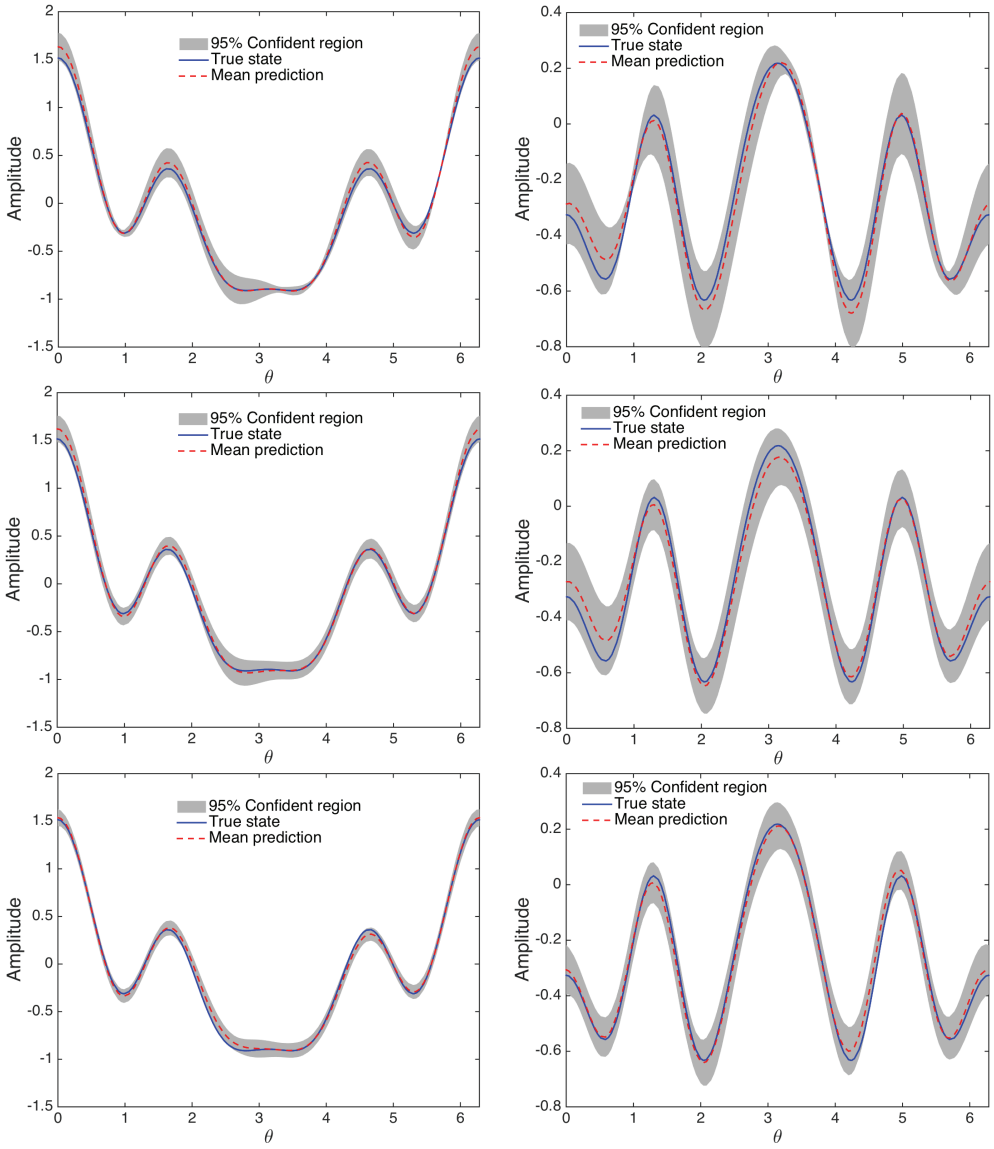


FIG. 8. *Panels show the true state and the mean prediction as a function of θ: the real part (left) and the imaginary part (right). These results are obtained using $M = 4$ (top row), $M = 8$ (middle row), and $M = 12$ (bottom row) observations with $\sigma_{\text{noise}} = 0.005$. In these plots the shaded area represents the mean prediction plus and minus two times the standard deviation function (corresponding to the 95% confidence region).*

By eliminating the constraints in the least-squares problem (70) we obtain the following result:

$$\bar{\boldsymbol{u}} := \arg\min_{\boldsymbol{z}\in\mathbb{C}^N} \frac{1}{2}(\boldsymbol{A}\boldsymbol{z}-\boldsymbol{l})^{\mathrm{H}}\boldsymbol{K}^{-1}(\boldsymbol{A}\boldsymbol{z}-\boldsymbol{l}) + \frac{1}{2}\sigma^{-2}(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{z}-\boldsymbol{d})^{\mathrm{H}}(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{z}-\boldsymbol{d}). \tag{71}$$

We see that the posterior mean is optimal in the sense that it minimizes an error objective function, which is defined as the sum of the model error and the output

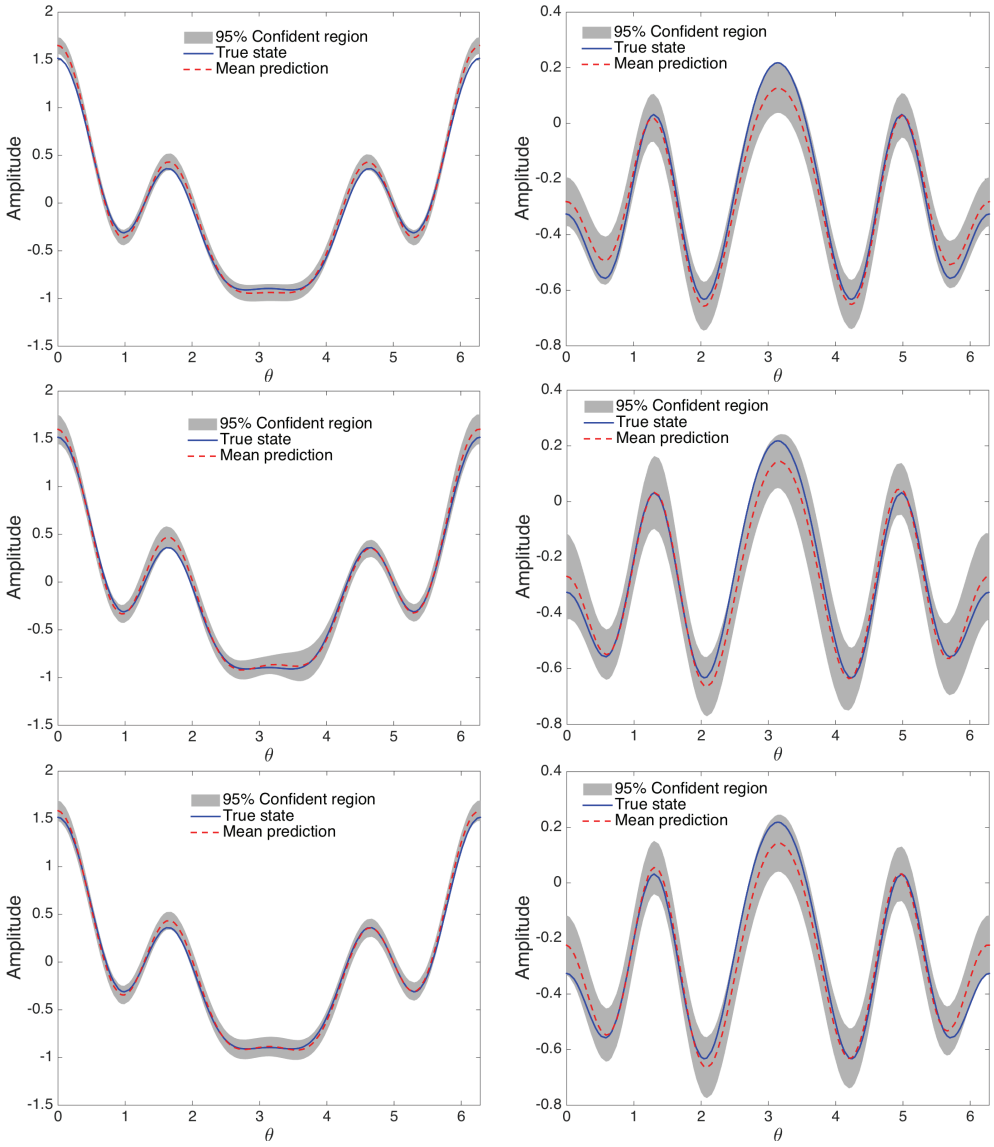


FIG. 9. *Panels show the true state and the mean prediction as a function of θ: the real part (left) and the imaginary part (right). These results are obtained using $M = 4$ (top row), $M = 8$ (middle row), and $M = 12$ (bottom row) observations with $\sigma_{\text{noise}} = 0.01$. In these plots the shaded area represents the mean prediction plus and minus two times the standard deviation function (corresponding to the 95% confidence region).*

error. Furthermore, we can write (71) as

$$\bar{\boldsymbol{u}} := \arg\min_{\boldsymbol{z}\in\mathbb{C}^N} \frac{1}{2}(\boldsymbol{z}-\boldsymbol{u}^{\text{o}})^{\text{H}}\boldsymbol{U}_0^{-1}(\boldsymbol{z}-\boldsymbol{u}^{\text{o}}) + \frac{1}{2}\sigma^{-2}(\boldsymbol{C}^{\text{H}}\boldsymbol{z}-\boldsymbol{d})^{\text{H}}(\boldsymbol{C}^{\text{H}}\boldsymbol{z}-\boldsymbol{d}), \tag{72}$$

where $\boldsymbol{U}_0 = \boldsymbol{A}^{-1}\boldsymbol{K}\boldsymbol{A}^{-\text{H}}$. The optimization formulation (72) is known as 3D variational data assimilation [8, 22]. It shows the relation between our method and 3D variational data assimilation.

Appendix B. Relationship with the Kalman method. Let us review the Kalman method [19] for state estimation. We begin by assuming that the state vector $\boldsymbol{u}$ is Gaussian with the prior distribution

$$p(\boldsymbol{u}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{u}-\boldsymbol{u}^{\mathrm{b}})^{\mathrm{H}}\boldsymbol{B}^{-1}(\boldsymbol{u}-\boldsymbol{u}^{\mathrm{b}})\right), \tag{73}$$

where $\boldsymbol{u}^{\mathrm{b}}$ is the background state vector and $\boldsymbol{B}$ is the background covariance matrix. Furthermore, the data $\boldsymbol{d}$ is assumed to have a Gaussian probability density function with covariance $\boldsymbol{R}$ and mean $\boldsymbol{C}^{\mathrm{H}}\boldsymbol{u}$. The likelihood function is thus given by

$$p(\boldsymbol{d}|\boldsymbol{u}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{u}-\boldsymbol{d})^{\mathrm{H}}\boldsymbol{R}^{-1}(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{u}-\boldsymbol{d})\right). \tag{74}$$

The posterior distribution of $\boldsymbol{u}$ is given by Bayes' theorem (see equation (2.3) on page 17 in [33]):

$$p(\boldsymbol{u}|\boldsymbol{d}) \propto p(\boldsymbol{d}|\boldsymbol{u})p(\boldsymbol{u}). \tag{75}$$

It can be shown by algebraic manipulations that the posterior distribution is also Gaussian

$$p(\boldsymbol{u}|\boldsymbol{d}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{u}-\hat{\boldsymbol{u}})^{\mathrm{H}}\hat{\boldsymbol{U}}^{-1}(\boldsymbol{u}-\hat{\boldsymbol{u}})\right). \tag{76}$$

Here the posterior mean $\hat{\boldsymbol{u}}$ and covariance $\hat{\boldsymbol{U}}$ are given by the Kalman formulas

$$\hat{\boldsymbol{u}} = \boldsymbol{u}^{\mathrm{b}} + \boldsymbol{H}\left(\boldsymbol{d}-\boldsymbol{C}^{\mathrm{H}}\boldsymbol{u}^{\mathrm{b}}\right), \qquad \hat{\boldsymbol{U}} = \boldsymbol{B}-\boldsymbol{H}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{B}, \tag{77}$$

where

$$\boldsymbol{H} = \boldsymbol{B}\boldsymbol{C}\left(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{C}+\boldsymbol{R}\right)^{-1} \tag{78}$$

is the so-called Kalman gain matrix.

We now show that our method is related to the Kalman method for a particular choice of the background state vector and the background covariance matrix. In particular, let us choose the prior information as

$$\boldsymbol{u}^{\mathrm{b}} = \boldsymbol{u}^{\mathrm{o}}, \qquad \boldsymbol{B} = \boldsymbol{A}^{-1}\boldsymbol{K}\boldsymbol{A}^{-\mathrm{H}}, \qquad \boldsymbol{R} = \sigma^2\boldsymbol{I}. \tag{79}$$

It then follows from (77)–(79) that

$$\begin{aligned}\hat{\boldsymbol{u}} &= \boldsymbol{u}^{\mathrm{o}} + \boldsymbol{A}^{-1}\boldsymbol{K}\boldsymbol{A}^{-\mathrm{H}}\boldsymbol{C}\left(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{K}\boldsymbol{A}^{-\mathrm{H}}\boldsymbol{C}+\sigma^2\boldsymbol{I}\right)^{-1}\left(\boldsymbol{d}-\boldsymbol{s}^{\mathrm{o}}\right),\\ \hat{\boldsymbol{U}} &= \boldsymbol{A}^{-1}\boldsymbol{K}\boldsymbol{A}^{-\mathrm{H}} - \boldsymbol{A}^{-1}\boldsymbol{K}\boldsymbol{A}^{-\mathrm{H}}\boldsymbol{C}\left(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{C}+\sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{K}\boldsymbol{A}^{-\mathrm{H}}.\end{aligned} \tag{80}$$

We see that the Kalman method yields exactly the same posterior distribution as our method when the priors are chosen by (79). Our method provides a systematic way to choose appropriate priors which are very important in the context of the Kalman method.

REFERENCES

[1] A. P. Demster, N. M. Laird, and D. B. Rubin, *Maximum Likelihood from Incomplete Data via the EM Algorithm*. J. Roy. Statist. Soc. Ser. B, 39 (1977), pp. 1–38.

[2] A. Atkinson, A. Donev, and R. Tobias, *Optimum Experimental Designs, with SAS*, Oxford University Press, Oxford, 2007.

[3] I. Babuška, F. Nobile, and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal., 45 (2007), pp. 1005–1034.

[4] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera, *An "empirical interpolation" method: Application to efficient reduced-basis discretization of partial differential equations*, C. R. Math., 339 (2004), pp. 667–672.

[5] H.-S. Chung and J. J. Alonso, *Using Gradients to Construct Response Surface Models for High-Dimensional Design Optimization Problems*, AIAA paper 2001-0922, 2001.

[6] H.-S. Chung and J. J. Alonso, *Using Gradients to Construct Cokriging Approximation Models for High-Dimensional Design Optimization Problems*, AIAA paper 2002-0317, 2002.

[7] A. Cohen, M. A. Davenport, and D. Leviatan, *On the stability and accuracy of least squares approximations*, Found. Comput. Math., 13 (2013), pp. 819–834.

[8] P. Courtier, E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux, *The ECMWF implementation of three-dimensional variational assimilation (3D-Var).* I: *Formulation*, Quart. J. Roy. Meteor. Soc., 124 (1998), pp. 1783–1807.

[9] R. P. Dwight and Z.-H. Han, *Efficient Uncertainty Quantification Using Gradient-Enhanced Kriging*, AIAA paper 2009-2276, 2009.

[10] M. S. Eldred, C. G. Webster, and P. G. Constantine, *Design Under Uncertainty Employing Stochastic Expansion Methods*, AIAA paper 2008-6001, 2008.

[11] R. Everson and L. Sirovich, *Karhunen–Loeve procedure for gappy data*, Opt. Soc. Am. A, 12 (1995), pp. 1657–1664.

[12] G. Fishman, *Monte Carlo: Concepts, Algorithms, and Applications*, Springer, New York, 1996.

[13] R. Ghanem and J. Red-Horse, *Propagation of probabilistic uncertainty in complex physical systems using a stochastic finite element approach*, Phys. D, 133 (1999), pp. 137–144.

[14] A. A. Giunta, M. S. Eldred, and J. P. Castro, *Uncertainty quantification using response surface approximations*, in Proceedings of the 9th ASCE Joint Specialty Conference on Probabilistic Mechanics and Structural Reliability, Albuquerque, NM,, 2004.

[15] G. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 996.

[16] J. M. Hammersley, D. C. Handscomb, and G. Weiss, *Monte Carlo methods*, Phys. Today, 18 (1965), p. 55.

[17] I. Harari and T. J. R. Hughes, *Galerkin/least-squares finite element methods for the reduced wave equation with non-reflecting boundary conditions in unbounded domains*, Comput. Methods Appl. Mech. Engrg., 98 (1992), pp. 411–454.

[18] J. Jakeman, M. Eldred, and D. Xiu, *Numerical approach for quantification of epistemic uncertainty*, J. Comput. Phys., doi:10.1016/j.jcp.2010.03.003.

[19] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Trans. ASME J. Basic Eng., 82 (1960), pp. 35–45.

[20] J. Laurenceau and P. Sagaut, *Building efficient response surfaces of aerodynamic functions with Kriging and cokriging*, AIAA Journal, 46 (2008), pp. 498–507.

[21] J. M. Lewis, S. Lakshmivarahan, and S. Dhall, *Dynamic Data Assimilation: A Least Squares Approach*, Cambridge University Press, Cambridge, 2006.

[22] Z. Li and I. M. Navon, *Optimality of variational data assimilation and its relationship with the Kalman filter and smoother*, Quart. J. Roy. Meteor. Soc., 127 (2001), pp. 661–683.

[23] W. Liu and S. M. Batill, *Gradient-Enhanced Response Surface Approximations Using Kriging Models*, AIAA paper 2002-5456, 2002.

[24] A. C. Lorenc, *A global three-dimensional multivariate statistical interpolation scheme*, Monthly Weather Rev., 109 (1981), pp. 701–721.

[25] Y. Maday, N. C. Nguyen, A. T. Patera, and G. S. H. Pau, *A general multipurpose interpolation procedure: The magic points*, Comm. Pure Appl. Anal., 8 (2008), pp. 383–404.

[26] N. C. Nguyen, A. T. Patera, and J. Peraire, *A "best points" interpolation method for efficient approximation of parametrized functions*, Internat. J. Numer. Methods Engrg., 73 (2008), pp. 521–543.

[27] N. C. Nguyen and J. Peraire, *Gaussian functional regression for linear partial differential equations*, Comput. Methods Appl. Mech. Engrg., 287 (2015), pp. 69–89.

[28] F. Nobile, R. Tempone, and C. G. Webster, *A sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal., 46 (2008), pp. 2309–2345.
[29] A. T. Patera and E. M. Ronquist, *Regression on parametric manifolds: Estimation of spatial fields, functional outputs, and parameters from noisy data*, C. R. Math., 350 (2012), pp. 543–547.
[30] F. Pukelsheim, *Optimal Design of Experiments*, Wiley & Sons, New York, 1993.
[31] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, 2006.
[32] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, *Design and analysis of computer experiments*, Statist. Sci., 4 (1989), pp. 409–423.
[33] S. Särkkä, *Bayesian Filtering and Smoothing*, Cambridge University Press, Cambridge, MA, 2013.
[34] T. Sondergaard and P. F. J. Lermusiaux, *Data assimilation with Gaussian mixture models using the dynamically orthogonal field equations. Part* I: *Theory and scheme*, Monthly Weather Rev., 141 (2013), pp. 1737–1760.
[35] K. Willcox, *Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition*, Comput. & Fluids, 35 (2006), pp. 208–226.
[36] J. Wolberg, *Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments*, Springer, New York, 2005.
[37] D. Xiu and J. A. Hesthaven, *High–order collocation methods for differential equations with random inputs*, SIAM J. Sci. Comput., 27 (2005), pp. 1118–1139.
[38] D. Xiu and G. E. Karniadakis, *Modeling uncertainty in flow simulations via generalized polynomial chaos*, J. Comput. Phys., 187 (2003), pp. 137–167.
[39] W. Yamazaki, M. P. Rumpfkeil, and D. J. Mavriplis, *Design Optimization Utilizing Gradient/Hessian Enhanced Surrogate Model*, AIAA paper 2010-4363, 2010.
[40] M. Yano, J. D. Penn, and A. T. Patera, *A model-data weak formulation for simultaneous estimation of state and model bias*, C. R. Math., 351 (Dec. 2013), pp. 937–941.