Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

**Noname manuscript No.**
(will be inserted by the editor)

# A distributed active subspace method for scalable surrogate modeling of function valued outputs

**Hayley Guy · Alen Alexanderian · Meilin Yu**

arXiv:1908.02694v2 [physics.comp-ph] 8 Aug 2019

the date of receipt and acceptance should be inserted later

**Abstract** We present a distributed active subspace method for training surrogate models of complex physical processes with high-dimensional inputs and function valued outputs. Specifically, we represent the model output with a truncated Karhunen–Loève (KL) expansion, screen the structure of the input space with respect to each KL mode via the active subspace method, and finally form an overall surrogate model of the output by combining surrogates of individual output KL modes. To ensure scalable computation of the gradients of the output KL modes, needed in active subspace discovery, we rely on adjoint-based gradient computation. The proposed method combines benefits of active subspace methods for input dimension reduction and KL expansions used for spectral representation of the output field. We provide a mathematical framework for the proposed method and conduct an error analysis of the mixed KL active subspace approach. Specifically, we provide an error estimate that quantifies errors due to active subspace projection and truncated KL expansion of the output. We demonstrate the numerical performance of the surrogate modeling approach with an application example from biotransport.

**Keywords**

Distributed active subspace; Karhunen–Loève expansion; Dimension reduction; Function valued outputs; Porous medium flow; Biotransport.

## 1 Introduction

Models with uncertain input parameters are common in modeling of complex systems. Computational studies such as forward uncertainty propagation, optimization, or parameter estimation require repeated evaluation of the model. These tasks become challenging for expensive-to-evaluate complex models. To address this challenge, surrogate models are often used. By approximating the mapping from the uncertain input parameters to output quantities of interest (QoIs), using a surrogate model, one can replace expensive model evaluations by inexpensive

Hayley Guy
Department of Mathematics, North Carolina State University
E-mail: hguy@ncsu.edu

Alen Alexanderian
Department of Mathematics, North Carolina State University
E-mail: alexanderian@ncsu.edu

Meilin Yu
Department of Mechanical Engineering, The University of Maryland, Baltimore County
E-mail: mlyu@umbc.edu

surrogate model evaluations. Examples of surrogate modeling tools include polynomial chaos expansion [19,28], multivariate adaptive regression splines [16], and Gaussian processes [36].

In the present work, we consider surrogate construction for models of the form

$$y = f(\boldsymbol{x}, \boldsymbol{\xi}), \tag{1}$$

where $\boldsymbol{x}$ belongs to a spatial domain and $\boldsymbol{\xi} \in \mathbb{R}^{N_p}$ is a vector of uncertain parameters. In our target applications $f$ is defined in terms of the solution of a partial differential equation (PDE) that is parameterized by a high-dimensional input parameter. Albeit, the proposed framework can be adapted to more general settings.

A simple approach is to construct a surrogate model pointwise in $\boldsymbol{x}$. Specifically, discretizing the spatial domain by grid points $\{\boldsymbol{x}_i\}_{i=1}^{N_x}$, we may consider approximating $f(\boldsymbol{x}_i, \boldsymbol{\xi})$, by constructing surrogate models $\hat{f}_i(\boldsymbol{\xi}) \approx f(\boldsymbol{x}_i, \cdot)$, $i = 1, \ldots, N_x$. This straightforward approach can be useful in some cases, however, the power of many of the surrogate modeling approaches can be fully realized if one optimizes them for each $\boldsymbol{x}_i$ in the computational grid. This is a computationally expensive task and might become prohibitive if the parameter dimension $N_p$ is large (e.g., $N_p \geq 100$), the model exhibits large variations in its response to parametric uncertainties over the spatial domain, or the grid point number $N_x$ is large, as can happen in two or three-dimensional geometries.

We seek an efficient surrogate modeling approach, for models of the form (1), whose complexity in terms of the number of model evaluations does not scale with the dimension $N_p$ of the input parameters. To enable this, we need a method that is *structure exploiting*. In particular, we seek to discover and utilize informative low-dimensional subspaces in the input space and low-dimensional spectral representations of the model output. The former uses ideas from active subspace methods [8] and the latter uses Karhunen–Loève (KL) expansions [30,19]. This proposed approach decouples the spatial (i.e. $\boldsymbol{x}$) dimensions and those of the random variable $\boldsymbol{\xi}$ and exposes important structures that can be used for building efficient surrogate models.

**Related work**. KL expansions have been used in many works for representing random field parameters in physics models; see e.g., [18,29,43,27,4,14,38,34,20,15]. KL expansions can also be used to represent random field outputs of physics models, as done in the present work. In models governed by PDEs the output field often exhibits favorable regularity properties and can be represented by a KL expansion with a relatively small number of KL terms. Examples of this appear for instance in our recent works [7,2] for models governed by elliptic PDEs.

The active subspace method [37,11,8] has become a popular approach in recent years for input parameter dimension reduction and surrogate modeling. This method seeks to identify important linear combinations of the input parameters. A brief summary of the active subspace method, for approximating scalar valued models, has been provided in Section 2. The active subspace method has been successfully used for uncertainty analysis in models with scalar valued responses, in a host of engineering applications; examples include scramjet analysis [12], wing shape optimization [31], hydrologic modeling [24], battery modeling [10], and kinetic model uncertainty analysis [25]. Recently, there have been efforts in extending the active subspace method to vector valued functions. In [44] the authors find an upper bound for the error of a ridge function approximation of the original (vector valued) function, and construct an approximating function by minimzing that upper bound. In [25] an interesting approach is introduced for simultaneously approximating multiple outputs using a single low-dimensional shared subspace. The shared subspace is identified by solving a least-squares system to compute an appropriate combination of single-output active subspaces.

We also mention another related and popular class of methods for parameter dimension reduction: variance based [40,41] and derivative based [42,26] global sensitivity analysis (GSA). These methods provide means of identifying unimportant model inputs, hence reducing the dimension of the input parameter vector. Derivative based methods are especially attractive, as they can be used to efficiently screen for unimportant input parameters, after which a surrogate model can be computed as a function of a reduced set of parameters. While GSA approaches have traditionally been applied to models with scalar outputs, extensions of GSA methods to vectorial and function valued outputs appear in several recent works [17,1,7].

In practice, the active subspace method tends to be very effective in reducing the input dimension as important *directions* in the input parameter space are identified. This is, in contrast to seeking reduced *parameter subsets*, as identified by GSA approaches. We note however that utility of GSA goes beyond input parameter dimension reduction—GSA provides valuable insight regarding a model by identifying key contributors to model variability. Also, we mention an interesting link between derivative based GSA and active subspaces through the idea of activity scores as detailed in [9]; computing active subspaces provides, as a byproduct, approximations to derivative based global sensitivity measures, for scalar valued models.

**Our approach and contributions**. In our proposed approach, we combine active subspaces and KL expansions to enable efficient surrogate modeling for function valued QoIs. Specifically, we consider a suitably truncated KL decomposition

$$f(\boldsymbol{x},\boldsymbol{\xi}) \approx \bar{f}(\boldsymbol{x}) + \sum_{k=1}^{N} \sqrt{\lambda_k} f_k(\boldsymbol{\xi})\phi_k(\boldsymbol{x}),$$

where $\bar{f}$ is the mean of the process $f(\boldsymbol{x},\boldsymbol{\xi})$ and $(\lambda_k, \phi_k)$ are the eigenpairs of the covariance operator of the process. This way, the model uncertainty is encoded in the KL modes, $f_k$, $k = 1, \ldots, N$. In many cases a small $N$ can be very effective in approximating $f$. These KL modes can then be approximated efficiently using the active subspace approach. The active subspace approach requires the gradients of $f_k$, which are scalar functions of the random vector $\boldsymbol{\xi}$. For such functions, adjoint state approaches provide efficient means of computing the required gradients. Specifically, gradients can be computed as a cost, in terms of model evaluations, that does not scale with the dimension of the input parameter $\boldsymbol{\xi}$. Realizing the proposed approach requires (i) a rigorous functional framework upon which efficient computational algorithms can be built; and (ii) a systematic computational procedure that guides efficient implementations. These have been detailed in Section 3.

We deploy our proposed framework in the context of flows in biological tissues. Specifically, in Section 4, we tackle biotransport in porous tumors with high-dimensional random inputs (i.e., the permeability field) and random field outputs (i.e., the pressure distribution in tumors). Biological tissues usually have highly heterogeneous, even uncertain, material properties [33]. The biotransport process such as drug delivery in tissues thus can exhibit "unpredictable" behaviors due to the uncertainties in tissue material properties [13]. Since the unpredictability can adversely affect the effectiveness of therapy, quantifying variability in biotransport due to uncertain material properties has a high impact on clinical trial protocol design. However, conducting useful uncertainty quantification studies on biotransport is challenging due to the high dimension of the input parameter space, e.g. permeability and porosity [3]. We find that efficient-to-evaluate and accurate surrogate models can be computed at a modest computational cost, with our proposed framework. The presented computational experiments also indicate that the present method can be used successfully in more general porous medium flow problems, where Darcy flow is a reasonable model. Further investigations of the proposed surrogate modeling method in more complex flow models is subject of our future work.

The contributions of this article are summarized as follows:

- We present a mathematical framework and computational method for surrogate modeling for models with high-dimensional inputs and function valued outputs that uses KL expansions for output dimension reduction, active subspaces for input dimension reduction, and adjoint based gradient computation as needed in the active subspace method.
- We analyze the errors due to active subspace projection and output KL truncation. The presented analysis provides insight on the interplay between these two important sources of errors.
- We present comprehensive computational results in the context of a biotransport application problem that test various aspects of the proposed surrogate modeling method. We mention that the computational studies, to our knowledge, are the first of a kind in the area of biotransport modeling in tissues with uncertain heterogeneous material properties.

## 2 Background on active subspace

One common approach to input dimension reduction is to identify a subset of input model parameters that are the most important to model variability. This is done typically using a local or global sensitivity analysis approach [26, 23, 35]. The active subspace approach is different; it identifies a set of important directions in the input parameter space rather than giving importance to one input over another. We can rotate the coordinates to align with the directions of strongest variation. Each direction can be considered as a set of weights that define a linear combination of all of the inputs. Directions where the inputs do not vary much are ignored.

Consider a model

$$y = g(\boldsymbol{\xi}), \quad \boldsymbol{\xi} \in \mathbb{R}^{N_p}.$$

We assume the uncertain parameters $\boldsymbol{\xi}$ have an associated probability density function $\pi(\boldsymbol{\xi})$ that is supported on $\Omega \subseteq \mathbb{R}^{N_p}$. Assume $g$ is square integrable and has continuous partial derivatives with respect to $\boldsymbol{\xi}$. The following matrix plays a key role in active subspace construction:

$$\mathbf{S} = \int_{\Omega} \nabla g(\boldsymbol{\xi}) \nabla g(\boldsymbol{\xi})^T \pi(\boldsymbol{\xi}) d\boldsymbol{\xi}, \tag{2}$$

where $\nabla g$ is the gradient of $g$. The matrix $\mathbf{S}$ is symmetric and positive semi-definite with spectral decomposition

$$\mathbf{S} = \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}^T, \quad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_{N_p}).$$

The eigenvalues $\lambda_i$'s are sorted in descending order $\lambda_1 \geq \cdots \geq \lambda_{N_p} \geq 0$ and $\mathbf{W}$ contains the orthonormal eigenvectors of $\mathbf{S}$.

The active subspace is determined by the dominant eigenvectors of $\mathbf{S}$. Specifically, we partition the eigenvalues and eigenvectors according to

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_1 & \\ & \boldsymbol{\Lambda}_2 \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_2 \end{bmatrix}, \tag{3}$$

where $\boldsymbol{\Lambda}_1 \in \mathbb{R}^{r \times r}$ is a diagonal matrix with the dominant eigenvalues on its diagonal; and $\mathbf{W}_1$ contains the corresponding eigenvectors $\boldsymbol{w}_1, \ldots \boldsymbol{w}_r$ as its columns. These columns span the dominant eigenspace of $\mathbf{S}$—*the active subspace*.

Given $\boldsymbol{\xi}$ we define $\boldsymbol{y} = \mathbf{W}_1^T\boldsymbol{\xi} \in \mathbb{R}^r$ and $\boldsymbol{z} = \mathbf{W}_2^T\boldsymbol{\xi} \in \mathbb{R}^{N_p - r}$, and note that

$$\boldsymbol{\xi} = \mathbf{W}_1\mathbf{W}_1^T\boldsymbol{\xi} + \mathbf{W}_2\mathbf{W}_2^T\boldsymbol{\xi} = \mathbf{W}_1\boldsymbol{y} + \mathbf{W}_2\boldsymbol{z}.$$

The elements of $\boldsymbol{y}$ and $\boldsymbol{z}$ are the sets of active and inactive variables, respectively. As discussed in detail in [11], the active variables, i.e., elements of $\boldsymbol{y}$, are responsible for most of the variations of the function $g$.

We can consider approximating $g$ in the active subspace, $g(\boldsymbol{\xi}) \approx g(\mathbf{W}_1\mathbf{W}_1^T\boldsymbol{\xi})$. It is convenient to define the function $G : \mathbb{R}^r \to \mathbb{R}$ by

$$G(\boldsymbol{y}) = g(\mathbf{W}_1\boldsymbol{y}), \quad \boldsymbol{y} \in \mathbb{R}^r,$$

and write the approximation to $g$ as

$$g(\boldsymbol{\xi}) \approx G(\mathbf{W}_1^T\boldsymbol{\xi}).$$

In practice, the function $G$ is typically approximated by a surrogate model. Specifically, we can compute a polynomial regression fit, which we denote by $\tilde{G}(\boldsymbol{y})$, for the function $G(\boldsymbol{y})$. Moreover, Monte Carlo sampling is used to approximate the matrix $\mathbf{S}$ in (2):

$$\mathbf{S} \approx \frac{1}{N_s} \sum_{i=1}^{N_s} \nabla g(\boldsymbol{\xi}_i) \nabla g(\boldsymbol{\xi}_i)^T.$$

Typically, a modest Monte Carlo sample is sufficient for computing reliable approximations to the dominant eigenvalues and eigenvectors of $\mathbf{S}$.

For readers' convenience, we provide the steps for computing an active subspace-based surrogate model for the function $g$ in Algorithm 1. For further details, we refer the readers to [11].

## 3 Active subspace-based surrogate models for function valued QoIs

In this section, we outline our approach for computing an active subspace-based surrogate model for a function valued QoI. Specifically, we consider models of the form

$$y = f(\boldsymbol{x}, \boldsymbol{\xi}), \quad \boldsymbol{x} \in \mathcal{X}, \boldsymbol{\xi} \in \Omega, \tag{4}$$

where $\mathcal{X} \subset \mathbb{R}^n$, with $n = 2$ or 3, is a compact set, and $\Omega \subset \mathbb{R}^{N_p}$ is a sample space. The set $\mathcal{X}$ will be a (sub-) region of a computational domain, in our target applications. Here $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_{N_p})^T$ is a vector of uncertain parameters, with probability density function $\pi(\boldsymbol{\xi})$. We make the following assumptions about $f(\boldsymbol{x}, \boldsymbol{\xi})$.

**Assumption 1** *We assume*

*(a) $f \in L^2(\mathcal{X} \times \Omega)$ and $f$ is a mean square continuous process.*

**Algorithm 1** Computation of an active subspace-based surrogate model for a scalar-valued function $g$.

**Input:** A set of $N_s$ data points $(\boldsymbol{\xi}_i, g(\boldsymbol{\xi}_i)), i = 1, \ldots, N_s$, drawn from the law of $\boldsymbol{\xi}$.

**Output:** Surrogate model $\tilde{G}(\mathbf{W}_1^T \boldsymbol{\xi}) \approx g(\boldsymbol{\xi})$

Compute the gradients $\boldsymbol{D}_i = \nabla g(\xi_i)$, $i = 1, \ldots, N_s$

Compute

$$\mathbf{S} = \frac{1}{N_s} \sum_{i=1}^{N_s} \boldsymbol{D}_i \boldsymbol{D}_i^T.$$

Compute spectral decomposition $\mathbf{S} = \mathbf{W} \boldsymbol{\Lambda} \mathbf{W}^T$.

Based on decay of the eigenvalues, determine the dimension $r$ of the active subspace, and partition $\boldsymbol{\Lambda}$ and $\boldsymbol{W}$ as in (3).

Compute $\boldsymbol{y}_i = \mathbf{W}_1^T \boldsymbol{\xi}_i$, $i = 1, \ldots, N_s$.

Compute a regression fit $\tilde{G}(\boldsymbol{y})$ to $G(\boldsymbol{y})$ using the data points $(\boldsymbol{y}_i, g(\boldsymbol{\xi}_i)), i = 1, \ldots, N_s$.

*(b)* $\frac{\partial f}{\partial \xi_i}(x, \boldsymbol{\xi})$ *exists for all* $x \in \mathcal{X}, \boldsymbol{\xi} \in \Omega, i = 1, \ldots, N_p$.

*(c)* $\frac{\partial f}{\partial \xi_i}(x, \boldsymbol{\xi}) \in L^2(\mathcal{X} \times \Omega), i = 1, \ldots, N_p$.

Letting $\mathbb{E}[\cdot]$ denote expectation with respect to $\boldsymbol{\xi}$, the covariance function, $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, of $f$ is defined as

$$c(\boldsymbol{x}, \boldsymbol{y}) := \mathbb{E}[f(\boldsymbol{x}, \cdot) f(\boldsymbol{y}, \cdot)] - \mathbb{E}[f(\boldsymbol{x}, \cdot)] \mathbb{E}[f(\boldsymbol{y}, \cdot)]$$

and the associated covariance operator $C_f : L^2(\mathcal{X}) \to L^2(\mathcal{X})$ is given by

$$[C_f u](\boldsymbol{x}) := \int_{\mathcal{X}} c(\boldsymbol{x}, \boldsymbol{y}) u(\boldsymbol{y}) d\boldsymbol{y}.$$

Assumption 1(a) ensures that the covariance function $c$ and the process mean

$$\bar{f}(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x}, \xi)]$$

are continuous on $\mathcal{X} \times \mathcal{X}$ and $\mathcal{X}$, respectively; see e.g., [22, Theorem 7.3.2].

### 3.1 Computational method

We reduce the dimension of the output by computing a low-rank approximation using a truncated KL expansion:

$$f(\boldsymbol{x}, \boldsymbol{\xi}) \approx \hat{f}(\boldsymbol{x}, \boldsymbol{\xi}) = \bar{f}(\boldsymbol{x}) + \sum_{k=1}^{N} \sqrt{\lambda_k(C_f)} f_k(\boldsymbol{\xi}) \phi_k(\boldsymbol{x}).$$

Here $\lambda_k(C_f)$ and $\phi_k$ are the eigenvalues and eigenvectors of the covariance operator $C_f$ and $f_k$'s are given by

$$f_k(\boldsymbol{\xi}) = \frac{1}{\sqrt{\lambda_k(C_f)}} \int_{\mathcal{X}} (f(\boldsymbol{x}, \boldsymbol{\xi}) - \bar{f}(\boldsymbol{x})) \phi_k(\boldsymbol{x}) d\boldsymbol{x}.$$

We refer to the coefficients $f_k$ as the *KL modes* of $f$. In the present work, the (generalized) eigenvalue problem,

$$C_f \phi_k = \lambda_k(C_f) \phi_k, \quad \int_{\mathcal{X}} \phi_k(\boldsymbol{x})^2 \, d\boldsymbol{x} = 1, \quad k = 1, 2, \ldots, \tag{5}$$

is solved using Nyström's method. In practice, $N$ can be chosen such that $(\sum_{k=1}^{N} \lambda_k(C_f))/(\sum_{k=1}^{\infty} \lambda_k(C_f)) < \text{tol}$, where tol is a user-specified tolerance. In many applications of interest, where $f$ is defined in terms of the solution of a differential equation, the eigenvalues $\lambda_k(C_f)$ exhibit rapid decay, which enables low-rank representations. In particular, this is observed in the application problem considered in the present work.

The next step is to approximate the KL modes $f_k(\boldsymbol{\xi})$ using active subspaces [11]. For each KL mode $f_k$, $k = 1, \ldots, N$, we compute an active subspace by considering the symmetric positive semidefinite matrix, $\mathbf{S}_k \in \mathbb{R}^{N_p \times N_p}$,

defined by

$$\mathbf{S}_k = \int_{\mathbb{R}^{N_p}} \big(\nabla f_k(\boldsymbol{\xi})\big)\big(\nabla f_k(\boldsymbol{\xi})\big)^T \pi(\boldsymbol{\xi}) d\boldsymbol{\xi}. \tag{6}$$

As before, we compute the spectral decomposition $\mathbf{S}_k = \mathbf{W}_k \boldsymbol{\Lambda}_k \mathbf{W}_k^T$, and partition the eigenpairs according to

$$\boldsymbol{\Lambda}_k = \begin{bmatrix} \boldsymbol{\Lambda}_{k,1} & \\ & \boldsymbol{\Lambda}_{k,2} \end{bmatrix}, \quad \mathbf{W}_k = \begin{bmatrix} \mathbf{W}_{k,1} & \mathbf{W}_{k,2} \end{bmatrix},$$

where $\boldsymbol{\Lambda}_{k,1} \in \mathbb{R}^{r_k \times r_k}$ is a diagonal matrix with the dominant eigenvalues of $\mathbf{S}_k$ on its diagonal; and $\mathbf{W}_k$ contains the corresponding eigenvectors. Defining,

$$G_k(\boldsymbol{y}) = f_k(\mathbf{W}_{k,1}\boldsymbol{y}), \quad \boldsymbol{y} \in \mathbb{R}^{r_k}, k = 1, \ldots, N,$$

the KL modes can be approximated by

$$f_k(\boldsymbol{\xi}) \approx G_k(\mathbf{W}_{k,1}^T \boldsymbol{\xi}), \quad k = 1, \ldots, N.$$

In practice, the active subspace-based surrogates for the KL modes $f_k$ are constructed by following Algorithm 1, with $g$ replaced by $f_k$, $k = 1, \ldots, N$. Thus, we obtain surrogate models

$$f_k(\boldsymbol{\xi}) \approx \tilde{G}_k(\mathbf{W}_{k,1}^T \boldsymbol{\xi}), \quad k = 1, \ldots, N.$$

The overall surrogate model for $f(\boldsymbol{x}, \boldsymbol{\xi})$ is then given by

$$f(\boldsymbol{x}, \boldsymbol{\xi}) \approx \hat{f}(\boldsymbol{x}, \boldsymbol{\xi}) = \bar{f}(\boldsymbol{x}) + \sum_{k=1}^{N} \sqrt{\lambda_k} \tilde{G}_k(\mathbf{W}_{k,1}^T \boldsymbol{\xi}) \phi_k(\boldsymbol{x}). \tag{7}$$

The computational steps for computing $\hat{f}(\boldsymbol{x}, \boldsymbol{\xi})$ can be divided into three main steps:

1. Compute the truncated KL expansion of $f(\boldsymbol{x}, \boldsymbol{\xi})$.
2. Compute the active subspace-based approximation to KL modes $f_k$, $k = 1, \ldots, N$.
3. Form the overall surrogate model as in (7).

The most computationally challenging part of the above process is the first step, in which we require an ensemble of model evaluations $f(\cdot, \boldsymbol{\xi}_i)$, $i = 1, \ldots, N_s$. These model evaluations will be used to compute the KL expansion of the model $f(\boldsymbol{x}, \boldsymbol{\xi})$. For this, we use Algorithm 1 in [2], that uses Nyström's method to compute $\lambda_k(C_f)$ and the corresponding eigenvectors $\phi_k(\cdot)$, $k = 1, \ldots, N$. This process also provides the evaluations of the KL modes, $f_k(\boldsymbol{\xi}_i), k = 1, \ldots, N$, $i = 1, \ldots, N_s$. As shown in our numerical results, often a modest choice of $N_s$ is sufficient for obtaining reliable approximations to (i) the dominant eigenpairs, and (ii) KL modes $f_k$.

Notice that for implementing the proposed method, differentiability of the KL modes is required. Moreover, as seen below, for the purposes of error analysis, Lipschitz continuity of the output KL modes is needed. These requirements can be satisfied through suitable boundedness assumptions on the partial derivatives of $f(\boldsymbol{x}, \boldsymbol{\xi})$, as we now explain. For convenience, and with no loss of generality, we consider the case where $\bar{f}(\boldsymbol{x}) \equiv 0$ and consider

$$F(\boldsymbol{\xi}) = \int_{\mathcal{D}} f(\boldsymbol{x}, \boldsymbol{\xi}) v(\boldsymbol{x}) d\boldsymbol{x}, \tag{8}$$

where we use a generic element $v \in L^2(\mathcal{D})$, with $\|v\|_{L^2(\mathcal{D})} = 1$, in place of the eigenvectors $\phi_k$. This $F$ can be thought of as a generic *unnormalized* KL mode. In addition to the earlier assumptions on $f$, we also require

$$\left| \frac{\partial f}{\partial \xi_j}(\boldsymbol{x}, \boldsymbol{\xi}) \right| \leq b_j(\boldsymbol{x}), \quad \text{for all } \boldsymbol{x} \in \mathcal{D}, \boldsymbol{\xi} \in \Omega,$$

where $b_j$, $j = 1, \ldots, N$, are square integrable.

**Lemma 1** *Suppose the process $f(\boldsymbol{x}, \boldsymbol{\xi})$ satisfies Assumption 1 and (8). Then, $F$ is differentiable and is Lipschitz continuous.*

*Proof* Differentiability of $F$ can be shown using the standard arguments of differentiating under the integral sign; see e.g., [7], where it is shown that under the present set of assumptions, $\frac{\partial F}{\partial \xi_j} = \int_{\mathcal{D}} \frac{\partial f}{\partial \xi_j}(\boldsymbol{x}, \boldsymbol{\xi}) v(\boldsymbol{x})\, d\boldsymbol{x}$. We also have

$$|\partial_j F(\boldsymbol{\xi})| \leq \int_{\mathcal{D}} |\partial_j f(\boldsymbol{x}, \boldsymbol{\xi}) v(\boldsymbol{x})| d\boldsymbol{x} \leq \int_{\mathcal{D}} |b_j(\boldsymbol{x})||v(\boldsymbol{x})|\, d\boldsymbol{x} \leq \|b_j\|_{L^2(\mathcal{D})} \|v\|_{L^2(\mathcal{D})},$$

for every $\boldsymbol{x} \in \mathcal{D}$. Using this, it is straightforward to show, $|F(\boldsymbol{\xi}_1) - F(\boldsymbol{\xi}_2)| \leq L\|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|$, for all $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in \Omega$, with $L = \left(\sum_{j=1}^{N} \|b_j\|^2_{L^2(\mathcal{D})}\right)^{1/2}$. □

### 3.2 Gradient computation

The present active subspace-based surrogate modeling approach requires computing the gradient of the output KL modes. Here, it is more convenient to consider the "unnormalized KL modes", $F_k$, $k = 1, \ldots, N$.

$$F_k(\boldsymbol{\xi}) = \int_{\mathcal{X}} (f(\boldsymbol{x}, \boldsymbol{\xi}) - \bar{f}(\boldsymbol{x}))\phi_k(\boldsymbol{x}) d\boldsymbol{x}. \tag{9}$$

Finite-difference approximations provide a simple approach, but will be prohibitive when the input dimension $N_p$ is large and model evaluations are expensive. For field quantities defined in terms of the solution $u$ of a PDE parameterized by $\boldsymbol{\xi}$, $F_k$ is a functional of $u$. This is an ideal situation for deploying adjoint based gradient computation. To illustrate this, we assume the model $f(\boldsymbol{x}, \boldsymbol{\xi})$ is a function of the solution $u(\boldsymbol{x}, \boldsymbol{\xi})$ of a PDE. For instance $f$ can be the restriction of $u$ to a sub-domain, or $f$ can be flux of $u$ through a boundary. For illustration, we consider the case where the PDE is of the form $\mathcal{A}(\boldsymbol{\xi})u = b$, where $\mathcal{A}$ denotes a differential operator that is parameterized by the uncertain parameter vector $\boldsymbol{\xi}$ and $b$ is a source term, and assume $f(\boldsymbol{x}, \boldsymbol{\xi}) = u(\boldsymbol{x}, \boldsymbol{\xi})$. To simplify the presentation further, we consider the discretized problem, where the discretized KL mode is defined by

$$\boldsymbol{F}_k(\boldsymbol{\xi}) = (\boldsymbol{u} - \bar{\boldsymbol{u}})^T \mathbf{W} \boldsymbol{\phi}_k, \quad \text{where} \quad \mathbf{A}(\boldsymbol{\xi})\boldsymbol{u} = \boldsymbol{b}.$$

Here $\boldsymbol{u}$ is the discretized state variable, $\mathbf{W}$ is a diagonal matrix with quadrature weights on diagonal, $\boldsymbol{\phi}_k$ is the discretized $k$th eigenvector of output covariance, $\mathbf{A}$ is the discretized PDE operator, and $\boldsymbol{b}$ is the discretized source term. Computing the gradient of $\boldsymbol{F}_k$ with respect to $\boldsymbol{\xi}$ can be done using a standard Lagrangian formalism [21]. Namely, we define the Lagrangian

$$\mathcal{L}(\boldsymbol{u}, \boldsymbol{\xi}, \boldsymbol{q}) = (\boldsymbol{u} - \bar{\boldsymbol{u}})^T \mathbf{W} \boldsymbol{\phi}_k + \boldsymbol{q}^T \big(\mathbf{A}(\boldsymbol{\xi})\boldsymbol{u} - \boldsymbol{b}\big),$$

where $\boldsymbol{q}$ is a Lagrange multiplier. Setting $\frac{\partial \mathcal{L}}{\partial \boldsymbol{q}} = 0$ recovers the state equation, and setting $\frac{\partial \mathcal{L}}{\partial \boldsymbol{u}} = 0$ gives the adjoint equation

$$\mathbf{A}^T(\boldsymbol{\xi})\boldsymbol{q} = -\mathbf{W}\boldsymbol{\phi}_k.$$

Then, the gradient of $\boldsymbol{F}_k$ is given by

$$\nabla \boldsymbol{F}_k(\boldsymbol{\xi})^T = \frac{\partial \mathcal{L}}{\partial \boldsymbol{\xi}} = \boldsymbol{q}^T \frac{\partial \mathbf{A}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \boldsymbol{u}.$$

Note that evaluation of $\nabla \boldsymbol{F}_k$ requires one state (forward) equation solve and one adjoint equation solve, independently of the dimension of the uncertain parameter vector $\boldsymbol{\xi}$. (The forward solves can be reused across the output KL modes.) Thus, to compute $\nabla \boldsymbol{F}_k$, $k = 1, \ldots, N$, we need one solution of the state equation, and $N$ adjoint solves. A small $N$, is often sufficient for suitable representations of the output field due to the rapid decay of the eigenvalues of the output covariance operator, observed in many applications.

Computing the active subspaces for the output KL modes can be done with a set $N_s$ of model evaluations, used across all KL modes. Thus, the computational cost of active subspace discovery for the output KL modes is $N_s(1 + N)$, independent of the parameter dimension $N_p$. Typically a modest $N_s$ is sufficient, as seen in our numerical results. Furthermore, the same set of model evaluations can be used for surrogate model construction for the output KL modes. This enables efficient computation of active subspaces for individual KL modes, at a cost that does not scale with the dimension of $\boldsymbol{\xi}$.

Notice that the present illustration uses an equation $\mathbf{A}(\boldsymbol{\xi})\boldsymbol{u} = \boldsymbol{b}$, which is linear in the state variable and nonlinear in the uncertain parameter $\boldsymbol{\xi}$. Such an equation can result from discretizing linear (in state) PDEs that are parameterized by uncertain parameters. The adjoint approach can more generally be applied to nonlinear PDE models; see e.g., [21].

In the present work, we use adjoint based gradient computation for computing the gradient of the output KL modes for models governed by elliptic PDEs with a random coefficient function; see section 4.

### 3.3 Error analysis

The presented computational strategy involves several approximations for computing the active subspace-based surrogate for $f(\boldsymbol{x}, \boldsymbol{\xi})$. In this section, we analyze the errors incurred due to (i) truncation of the output KL expansion and (ii) active subspace approximation of the KL modes. For the purposes of the presented analysis, it is more convenient to work with unnormalized KL modes (9), and consider

$$f_N(\boldsymbol{x}, \boldsymbol{\xi}) = \bar{f}(\boldsymbol{x}) + \sum_{k=1}^{N} F_k(\boldsymbol{\xi})\phi_k(\boldsymbol{x}). \tag{10}$$

The active subspace strategy is flexible regarding the distribution law of the random vector $\boldsymbol{\xi}$; see e.g., [8]. However, in the present work, where we consider models with uncertain coefficient functions that are modeled using log-Gaussian random fields, $\boldsymbol{\xi}$ is a standard Gaussian random vector, i.e., $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

As detailed in [8], while Algorithm 1 provides a practical surrogate modeling framework, it is not directly amenable to theoretical analysis. To enable error analysis, following [8], we define the functions $G_k$, used for active subpace projection of the KL modes in the following way:

$$G_k(\boldsymbol{y}) = \int f_k(\mathbf{W}_{k,1}\boldsymbol{y} + \mathbf{W}_{k,2}\boldsymbol{z})\, \pi_{Y|Z}(\boldsymbol{z}|\boldsymbol{y}) d\boldsymbol{z}, \tag{11}$$

where $\pi_{Y|Z}$ is the conditional density,

$$\pi_{Y|Z}(\boldsymbol{z}|\boldsymbol{y}) := \frac{\pi(\mathbf{W}_{k,1}\boldsymbol{y} + \mathbf{W}_{k,2}\boldsymbol{z})}{\pi_Z(\boldsymbol{z})}, \quad \text{with } \pi_Z(\boldsymbol{z}) = \int \pi(\mathbf{W}_{k,1}\boldsymbol{y} + \mathbf{W}_{k,2}\boldsymbol{z})\, d\boldsymbol{y}.$$

In practice, the *marginalized* $G_k$, which for convenience we can denote by

$$G_k(\boldsymbol{y}) = \mathbb{E}_{\boldsymbol{z}}\{f_k(\mathbf{W}_{k,1}\boldsymbol{y} + \mathbf{W}_{k,2}\boldsymbol{z})\},$$

can be approximated by Monte Carlo sampling,

$$G_k(\boldsymbol{y}) \approx \hat{G}_k(\boldsymbol{y}) := \frac{1}{N_{\text{AS}}} \sum_{i=1}^{N_{\text{AS}}} f_k(\mathbf{W}_{k,1}\boldsymbol{y} + \mathbf{W}_{k,2}\boldsymbol{z}_i). \tag{12}$$

However, as seen below, even a very small Monte Carlo Sample (even with $N_{\text{AS}} = 1$) can be acceptable. This partly justifies and explains the effectiveness of Algorithm 1, which can be seen as a special case of (12), with $N_{\text{AS}} = 1$ and $\boldsymbol{z}_1 = 0$.

Recall the approximation of the KL modes $F_k(\boldsymbol{\xi}) \approx G_k(\boldsymbol{W}_{k,1}^T\boldsymbol{\xi})$. As a first step in our error analysis, we quantify the error in this approximation. For each $k \in \{1, \ldots, N\}$, we define

$$\delta_k := \left( \sum_{j=r_k+1}^{N_p} \lambda_j(\mathbf{S}_k) \right)^{1/2}, \tag{13}$$

where $\{\lambda_j(\mathbf{S}_k)\}_{j=1}^{N_p}$ are the eigenvalues of $\mathbf{S}_k$ defined in (6) and $r_k$ is the dimension of the active subspace for the $k$th KL mode $F_k(\boldsymbol{\xi})$. The following result bounds the error in approximating the individual KL modes.

**Lemma 2** *Let $\delta_k$ be as in* (13)*, and let $G_k$ and $\hat{G}_k$ be as in* (11) *and* (12)*. Then,*

*(a)* $\displaystyle\int_\Omega (F_k(\boldsymbol{\xi}) - G_k(\boldsymbol{W}_{k,1}^T\boldsymbol{\xi}))^2 \pi(\boldsymbol{\xi}) d\boldsymbol{\xi} \le \delta_k^2.$

*(b)* $\displaystyle\int_\Omega (F_k(\boldsymbol{\xi}) - \hat{G}_k(\boldsymbol{W}_{k,1}^T\boldsymbol{\xi}))^2 \pi(\boldsymbol{\xi}) d\boldsymbol{\xi} \le (1 + N_{AS}^{-1/2})\delta_k^2.$

*Proof* The KL modes $F_k$ are square integrable, mean zero, and by Lemma 1, they are differentiable and Lipschitz continuous. Thus, the first statement follows from using [8, Theorem 4.3] and the second one follows from [8, Theorem 4.4]. □

Next, we consider the error, due to active subspace projection (for individual KL modes), in approximating the KL expansion:

$$e(\boldsymbol{x}, \boldsymbol{\xi}) = |f_N(\boldsymbol{x}, \boldsymbol{\xi}) - \hat{f}_N(\boldsymbol{x}, \boldsymbol{\xi})|$$

where $f_N$ is as in (10), and

$$\hat{f}_N(\boldsymbol{x},\boldsymbol{\xi}) := \sum_{k=1}^{N} G_k(\boldsymbol{W}_{k,1}^T\boldsymbol{\xi})\phi_k(\boldsymbol{x}) \tag{14}$$

its active subspace-based approximation where $G_k(\boldsymbol{W}_{k,1}^T\boldsymbol{\xi})$ is our approximation of the KL modes $F_k(\boldsymbol{\xi})$, as defined before. Then let

$$\bar{e}(\boldsymbol{\xi}) = \int_{\mathcal{X}} e(\boldsymbol{x},\boldsymbol{\xi})d\boldsymbol{x}.$$

**Theorem 1** $\mathbb{E}\{\bar{e}\} \le |\mathcal{X}|^{\frac{1}{2}} \sum_{k=1}^{N} \delta_k$, *where* $\delta_k$ *is as in* (13).

*Proof* See Appendix A. □

*Remark 1* Note that in view of Lemma 2(b), if we use $\hat{G}_k$ defined in (12), instead of $G_k$ in (14), we can repeat the argument in proof of Theorem 1 to get the following estimate:

$$\mathbb{E}\{\bar{e}\} \le |\mathcal{X}|^{\frac{1}{2}}(1 + N_{\mathrm{AS}}^{-1/2}) \sum_{k=1}^{N} \delta_k.$$

Finally, we consider the overall error of approximating $f(\boldsymbol{x},\boldsymbol{\xi})$, due to KL truncation and active subspace projection:

$$E(\boldsymbol{x},\boldsymbol{\xi}) = |f(\boldsymbol{x},\boldsymbol{\xi}) - \hat{f}_N(\boldsymbol{x},\boldsymbol{\xi})|$$

where $f$ is the original QoI and $\hat{f}_N$ is the active subspace-based approximation as before. We consider,

$$\bar{E}(\boldsymbol{\xi}) := \int_{\mathcal{X}} E(\boldsymbol{x},\boldsymbol{\xi})dx.$$

We have the following result:

**Theorem 2**

$$\mathbb{E}\{\bar{E}(\boldsymbol{\xi})\} \le |\mathcal{X}|^{\frac{1}{2}} \left[ \left( \sum_{k=N+1}^{\infty} \lambda_k(C_f) \right)^{1/2} + \sum_{k=1}^{N} \left( \sum_{j=r_k+1}^{N_p} \lambda_j(\mathbf{S}_k) \right)^{1/2} \right]. \tag{15}$$

*Proof* See Appendix A. □

We note that the first term in (15) indicates error due to truncation of the output KL expansion. Recall that $\sum_{k=1}^{\infty} \lambda_k(C_f) = \int_{\mathcal{X}} c_f(\boldsymbol{x},\boldsymbol{x})\, d\boldsymbol{x} < \infty$, where the equality is due to Mercer's Theorem, and the finiteness of the integral is due to continuity of the covariance operator, which is a consequence of the mean square continuity assumption on the process. Therefore, the first term in (15) can be made arbitrarily small by taking $N$ sufficiently large. However, choosing a large $N$ could entail accumulation of error due to active subspace projection error in the second term in (15); this error, however, can be controlled by increasing $r_k$. In practice, in many applications, a small number of output KL modes (i.e., a small $N$), can be used to obtain an accurate KL representation for the output. Moreover, typically low-dimensional (in many cases one- or two-dimensional) active subspaces can be afforded for approximating the dominant KL modes. We demonstrate these issues numerically in Section 4.

*Remark 2* In view of Remark 1, if we use $\hat{G}_k$ defined in (12), instead of $G_k$ in (14), we can repeat the argument in proof of Theorem 2 to get the following estimate:

$$\mathbb{E}\{\bar{E}(\boldsymbol{\xi})\} \le |\mathcal{X}|^{\frac{1}{2}} \left[ \left( \sum_{k=N+1}^{\infty} \lambda_k(C_f) \right)^{1/2} + (1 + N_{\mathrm{AS}}^{-1/2}) \sum_{k=1}^{N} \left( \sum_{j=r_k+1}^{N_p} \lambda_j(\mathbf{S}_k) \right)^{1/2} \right].$$

**Additional sources of error**. The above error analysis only concerns errors assiciated with active subspace projection and output KL truncation. In practical computations there are a number of other errors. Most of these errors are related to the active subspace approach used for approximating the KL modes $F_k$. These include errors in approximating the eigenvalues and eigenvectors of $\mathbf{S}_k$'s, incurred due to sample average approximation to these

matrices and surrogate modeling errors incurred in approximating $G_k$'s. Errors in approximating $\mathbf{\Lambda}_k$ and $\mathbf{W}_k$ are analyzed for instance in [8]. Errors due to surrogate modeling of $G_k$'s are difficult to quantify in general, as these errors depend on the choice of surrogate modeling framework.

There are also further errors related to KL approximation of the output field. Namely, we need to approximate the mean field $\mathbb{E}\{f(\boldsymbol{x},\cdot)\}$. If simple Monte Carlo sampling is used, approximations of the mean field exhibit the usual Monte Carlo convergence behavior. However, quasi-Monte Carlo approaches can provide efficient means of obtaining more accurate estimates. Finally, there will be errors in computing the eigenvalues $\lambda_k(C_f)$ and the corresponding eigenvectors. These errors are due to (i) sample average approximation to the output covariance function, and (ii) errors due to discretizing the generalized eigenvalue problem (5). The discretization errors of course depend on the numerical method used for solving the eigenvalue problem. For example, if Nystrom's method is used, as done in the present work, the discretization errors can be controlled by the resolution of the computational grid, and the quadrature method used.

We demonstrate numerically that once a suitable output dimension reduction is determined, the active subspace approach can be deployed to obtain approximations to the output KL modes and an overall surrogate model that captures the statistical properties of $f$ reliably. We find that this can be accomplished with an ensemble of function evaluations $\{f(\cdot,\boldsymbol{\xi}_j)\}_{j=1}^{N_s}$, with a modest $N_s$.

## 4 Application to biotransport in tumors

In this section, we present our computational results in the context of a biotransport application problem. We begin by describing the governing model in Section 4.1. Next, we discuss computation of the KL expansion of the output in Section 4.2. This is followed by our results on active subspace discovery and surrogate model construction in Section 4.3. We test the accuracy of the computed surrogate models in Section 4.4.

### 4.1 The governing model

In this section we describe the biotransport problem we seek to investigate using the proposed method. We are interested in understanding the impact of uncertainty in the material properties of cancerous tumors. Specifically we seek to characterize the uncertainties in the pressure field when a single needle injection occurs at the center of a spherical tumor with uncertain heterogeneous structure. We focus on a 2D cross-section, and consider Darcy's law constrained by mass conservation in a domain $\mathcal{D}\subset\mathbb{R}^2$ given by a circle of radius $R_{\text{tumor}}=5$ mm, centered at the origin, with an inner circle of radius $R_{\text{needle}}=0.25$ mm, modeling the injection site, removed. We denote the inner and outer boundaries of the domain by $\Gamma_{\text{N}}$ and $\Gamma_{\text{D}}$, respectively. The following elliptic PDE governs the fluid pressure $p$:

$$\begin{aligned} -\nabla\cdot\left(\frac{\kappa}{\mu}\nabla p\right) &= 0 \quad \text{in } \mathcal{D},\\ p &= 0 \quad \text{on } \Gamma_{\text{D}},\\ \nabla p\cdot\boldsymbol{n} &= \frac{Q\mu}{2\pi R_{\text{needle}}\kappa} \quad \text{on } \Gamma_{\text{N}}. \end{aligned} \tag{16}$$

In this equation, $\kappa$ denotes the absolute permeability field, $\mu$ is the fluid dynamic viscosity, $Q$ is the volume flow rate per unit length, and $\boldsymbol{n}$ is the outward-pointing normal vector. The nominal values for the above parameters are $\kappa=0.5\ md$, $\mu=8.9\times10^{-4}\ Pa\cdot s$, and $Q=1\ mm^2/min$. These values are chosen according to previous investigations of fluid transport in tumors [39,32,6]. As noted in a number of previous works, tumors exhibit complex structures due to their invasive nature. Generally, tumors consist of loosely organized abnormal cells, fibers, vasculature, and lymphatics [5], resulting in disordered tissues with complex heterogeneous structures.

In the present work, we model the permeability field $\kappa$ by a log-Gaussian random field as follows. Let $z(\boldsymbol{x},\omega)$ be a centered Gaussian process; here $\omega\in\Omega$ where $\Omega$ is an appropriate sample space. We assume $z$ has unit pointwise variance and has correlation function

$$c_z(\boldsymbol{x},\boldsymbol{y})=\exp\left\{-\frac{1}{\ell}\|\boldsymbol{x}-\boldsymbol{y}\|_1\right\},\quad \boldsymbol{x},\boldsymbol{y}\in\mathcal{D}, \tag{17}$$

where $\ell > 0$ is the correlation length. In the present study we set the correlation length $\ell = 1\ mm$. Then, we define the log-permeability field $a = \log \kappa$ according to

$$a(\boldsymbol{x}, \omega) = a_0(\boldsymbol{x}) + \sigma_a z(\boldsymbol{x}, \omega), \quad \boldsymbol{x} \in \mathcal{D}, \omega \in \Omega,$$

where $a_0$ and $\sigma_a^2$ represent the pointwise mean and variance, respectively. We can represent $a(\boldsymbol{x}, \omega)$ using a truncated KL expansion:

$$a(\boldsymbol{x}, \omega) \approx \hat{a}(\boldsymbol{x}, \omega) := a_0(\boldsymbol{x}) + \sum_{j=1}^{N_p} \sqrt{\lambda_j(C_a)} \xi_j(\omega) e_j(\boldsymbol{x}), \tag{18}$$

where $a_0$ is the mean field, $(\lambda_j(C_a), e_j)$ are the eigenpairs of the covariance operator $C_a$ of $a(\boldsymbol{x}, \omega)$, $N_p$ is the input parameter dimension, and $\xi_j$ are independent standard normal random variables. With this parameterization, the uncertainty in the (approximate) log permeability field $\hat{a} = \log \kappa$ is completely characterized by the random vector $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_{N_p})^T$. That is, $\hat{a}(\boldsymbol{x}, \omega) = \hat{a}(\boldsymbol{x}, \boldsymbol{\xi}(\omega))$, and thus, we can consider the (approximate) log-permeability field as a random process $\hat{a} : \mathcal{D} \times \Omega \to \mathbb{R}$. The QoI under study here is the pressure field $p(\boldsymbol{x}, \boldsymbol{\xi})$.

For illustration, two sets of realizations of the permeability field and the corresponding pressure field with correlation length $\ell = 1\ mm$ are presented in Figure 1. We observe large fluctuations in the permeability field and relatively mild fluctuations in the pressure field.

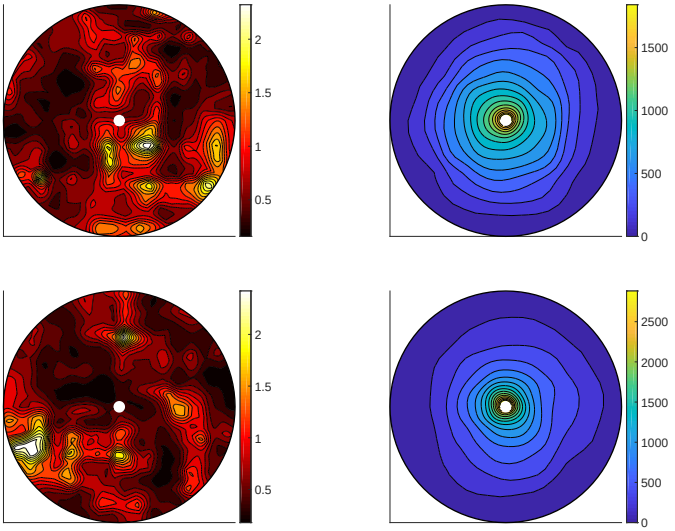


**Fig. 1** Two realizations of the log-permeability field (left) and the corresponding pressure field (right) using correlation length $\ell = 1\ mm$.

## 4.2 Spectral representation of the model output

We represent the QoI, $p$, using a truncated KL expansion

$$p(\boldsymbol{x}, \boldsymbol{\xi}) \approx \bar{p}(\boldsymbol{x}) + \sum_{k=1}^{N} \sqrt{\lambda_k} p_k(\boldsymbol{\xi}) \phi_k(\boldsymbol{x}), \qquad \text{where} \qquad p_k(\boldsymbol{\xi}) = \frac{1}{\sqrt{\lambda_k}} \int_{\mathcal{D}} (p(\boldsymbol{x}, \boldsymbol{\xi}) - \bar{p}(\boldsymbol{x})) \phi_k(\boldsymbol{x}) d\boldsymbol{x}. \tag{19}$$

Here $\lambda_k, \phi_k(\boldsymbol{x})$ are the eigenvalues and corresponding eigenfunctions of the covariance operator $C_f$ of $p(\boldsymbol{x}, \boldsymbol{\xi})$.

In Figure 2 (left) we see that the eigenvalues of the covariance operator $C_f$ show faster decay than those of the log-permeability field $a(x, \boldsymbol{\xi})$. In Figure 2 (middle) we compute the ratio $\rho_k = (\sum_{k=1}^{N} \lambda_k)/\mathrm{Tr}(C_f)$, where $\lambda_k$ are the eigenvalues of the covariance operator $C_f$ and $N$ is the number of KL modes retained. Using just $N = 10$ KL modes gives us $\rho_k \approx 0.9$ indicating that 90% of the variance is captured. If we add five more modes so that

$N = 15$ then we capture nearly 95% of the average variance in the model output $p$. Figure 2 (right) shows the first 40 eigenvalues of the covariance operator $C_f$, for a number of different sample sizes $N_s$ used to approximate the covariance function of $p$. Note that using $N_s = 300$ samples we can approximate the dominant eigenvalues reasonably well. In the current study we set $N_p$ the input parameter dimension to be 200 and to retain $N = 15$ KL modes. From the results that follow we will see that we can achieve significant dimension reduction for the input parameter.

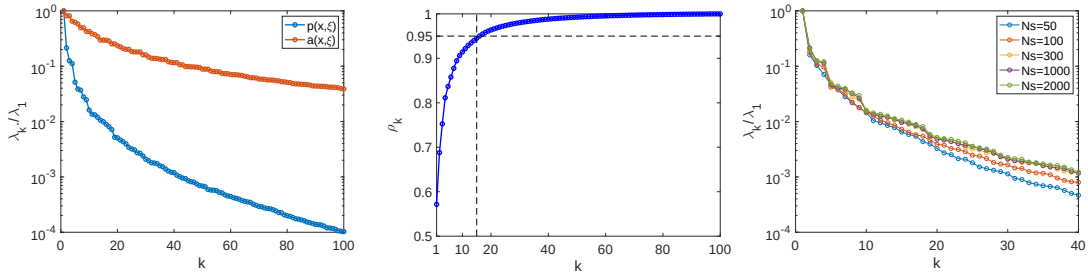


**Fig. 2** Eigenvalue spectrum of $a(\boldsymbol{x}, \boldsymbol{\xi})$ versus $p(\boldsymbol{x}, \boldsymbol{\xi})$ (left). Ratio showing saturation of average variance for $p(\boldsymbol{x}, \boldsymbol{\xi})$ (middle). First 40 eigenvalues of $C_f$ for different sample sizes $N_s$ (right).

### 4.3 Active subspace discovery and surrogate model construction

To construct the active subspace-based surrogate model for $p(\boldsymbol{x}, \boldsymbol{\xi})$, we need the gradients of the output KL modes defined in (19). For this, we use the adjoint method. The adjoint based expression for $p_k(\boldsymbol{\xi})$ can be derived using a formal Lagrange approach; see, e.g., [21]. For a basic derivation of the gradient of the output KL modes for models governed by elliptic PDEs, we refer the reader to [7], where output KL expansions are used within the context of derivative based global sensitivity analysis. The adjoint based expression for the partial derivatives of $p_k$'s are given by,

$$\frac{\partial p_k}{\partial \xi_j} = \sqrt{\lambda_j(C_a)} \int_{\mathcal{D}} e_j(\boldsymbol{x}) e^{\hat{a}(\boldsymbol{x},\boldsymbol{\xi})} \nabla p(\boldsymbol{x}) \cdot \nabla q(\boldsymbol{x})\, d\boldsymbol{x}. \tag{20}$$

where $p$ is the solution of the (forward) PDE (16) and $q$ is the solution of the adjoint equation:

$$\begin{aligned} -\nabla \cdot \left(\frac{\kappa}{\mu} \nabla q\right) &= -\frac{1}{\sqrt{\lambda_k(C_f)}} \phi_k \quad \text{in } \mathcal{D}, \\ q &= 0 \quad \text{on } \Gamma_D, \\ \nabla q \cdot n &= 0 \quad \text{on } \Gamma_N. \end{aligned} \tag{21}$$

Note that in the forward and adjoint equation, we let the permeability field be $\kappa(\boldsymbol{x}, \boldsymbol{\xi}) = e^{\hat{a}(\boldsymbol{x},\boldsymbol{\xi})}$.

Guided by the results in the previous subsection, we focus on the first $N = 15$ output KL modes. For each $k \in \{1, \ldots, N\}$, we generate a sample $\{\nabla p_k(\boldsymbol{\xi}_i)\}_{i=1}^{N_s}$, with $N_s = 300$. For $\{\boldsymbol{\xi}_i\}_{i=1}^{N_s}$, we used the same set of 300 parameter samples used in computing the KL expansion of the output. Using (20),

$$\nabla p_k(\boldsymbol{\xi}_i) = \begin{bmatrix} \sqrt{\lambda_1(C_a)} \displaystyle\int_{\mathcal{D}} e_1(\boldsymbol{x}) e^{\hat{a}(\boldsymbol{x},\boldsymbol{\xi}_i)} \nabla p(\boldsymbol{x},\boldsymbol{\xi}_i) \cdot \nabla q(\boldsymbol{x},\boldsymbol{\xi}_i)\, d\boldsymbol{x} \\ \sqrt{\lambda_2(C_a)} \displaystyle\int_{\mathcal{D}} e_2(\boldsymbol{x}) e^{\hat{a}(\boldsymbol{x},\boldsymbol{\xi}_i)} \nabla p(\boldsymbol{x},\boldsymbol{\xi}_i) \cdot \nabla q(\boldsymbol{x},\boldsymbol{\xi}_i)\, d\boldsymbol{x} \\ \vdots \\ \sqrt{\lambda_{N_p}(C_a)} \displaystyle\int_{\mathcal{D}} e_{N_p}(\boldsymbol{x}) e^{\hat{a}(\boldsymbol{x},\boldsymbol{\xi}_i)} \nabla p(\boldsymbol{x},\boldsymbol{\xi}_i) \cdot \nabla q(\boldsymbol{x},\boldsymbol{\xi}_i)\, d\boldsymbol{x} \end{bmatrix}, \quad i = 1, \ldots, N_s.$$

To compute these, we reuse the model evaluations $\{p(\boldsymbol{x}, \boldsymbol{\xi}_i)\}_{i=1}^{N_s}$, from the computation of the output KL expansion earlier; the adjoint variables $q(\cdot, \boldsymbol{\xi}_i)$ are computed by solving the adjoint equation (21), with $\kappa = e^{\hat{a}(\boldsymbol{x},\boldsymbol{\xi}_i)}$, $i = 1, \ldots, N_s$.

Using the gradient samples, we approximate the matrix $\mathbf{S}_k$ defined in (6) for each KL mode, $k = 1, \ldots, N$:

$$\hat{\mathbf{S}}_k = \frac{1}{N_s} \sum_{i=1}^{N_s} \nabla p_k(\boldsymbol{\xi}_i) \nabla p_k(\boldsymbol{\xi}_i)^T.$$

In each case, we consider the corresponding spectral decomposition $\hat{\mathbf{S}}_k = \hat{\mathbf{W}}_k \hat{\boldsymbol{\Lambda}}_k \hat{\mathbf{W}}_k^T$. To identify the active subspace and where to partition the eigenpairs we examine the spectrum of $\hat{\mathbf{S}}_k$. As an illustration, in Figure 3 (top) we present the spectrum for the first three output KL modes. Visually we observe a gap between the first and second eigenvalue for each of the modes indicating one-dimensional active subspaces. In Figure 3 (bottom), we show the corresponding sufficient summary plots (SSPs) for the corresponding modes. A sufficient summary plot here is a scatter plot of the active variables $y = \mathbf{W}_{k,1}^T \boldsymbol{\xi}$ versus the output KL modes $p_k(\boldsymbol{\xi})$. We observe a strong univariate trend which further indicates the presence of one-dimensional active subspaces. To provide a consistent truncation approach, we can use a threshold $\gamma$ and choose the dimension $r_k$ of the active subspace according to $\lambda_{k,1}/\lambda_{k,r_{k+1}} > \gamma$. In the current study we use $\gamma = 10$. Using this approach, we identified a one-dimensional active subspace for each of the first 15 output modes, expect modes 11 and 12 where two-dimensional active subspaces were identified. To illustrate, we report the spectrum of $\mathbf{S}_{12}$ and the SSP for $p_{12}$ in Figure 4. Based on the determined values of $r_k$, we partition

$$\boldsymbol{\Lambda}_k = \begin{bmatrix} \boldsymbol{\Lambda}_{k,1} & \\ & \boldsymbol{\Lambda}_{k,2} \end{bmatrix}, \quad \mathbf{W}_k = \begin{bmatrix} \mathbf{W}_{k,1} & \mathbf{W}_{k,2} \end{bmatrix}, \quad k = 1, \ldots, N.$$

$\boldsymbol{\Lambda}_{k,1}$ contains the dominant eigenvalues and $\mathbf{W}_{k,1}$ the corresponding eigenvectors.

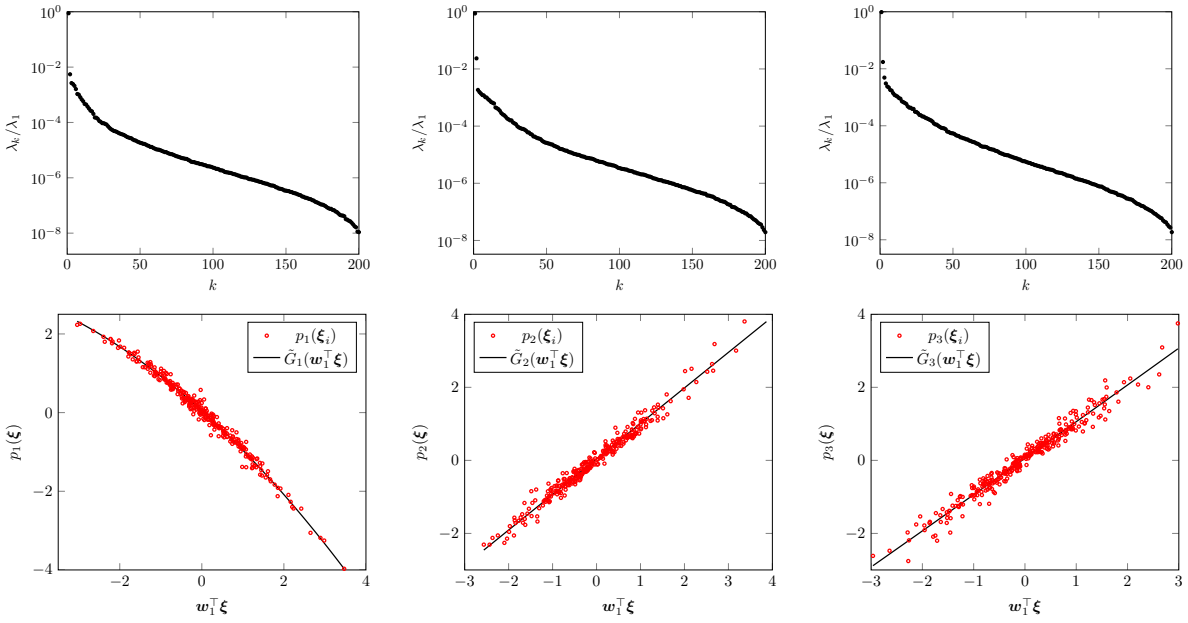


**Fig. 3** Top: eigenvalues of the matrix $\hat{\mathbf{S}}_k$ for $k = 1$ (left) $k = 2$ (middle), and $k = 3$ (right); bottom: SSPs for output KL modes $p_1$ (left), $p_2$ (middle), and $p_3$ (right).

Recall that the active subspace approach essentially seeks "important linear combinations" of the input parameters. To illustrate this, in Figure 5, we present the components of the dominant eigenvector for the first three output KL modes $p_1, p_2$ and $p_3$. The smaller inset plot shows the first 50 components of the dominant eigenvector for $p_1, p_2$ and $p_3$. The magnitude of the components give us a sensitivity measure for each of the input parameters. A large component value indicates that that particular input is important in defining the direction of most variation in our function $p$. We note that the first output KL mode is sensitive to a few components of the input parameter vector. In contrast, the second and especially the third output KL modes show sensitivity to a larger number of components of $\boldsymbol{\xi}$.

Finally, we note that the present results indicate a significant dimension reduction. The dominant output KL modes, each a function of 200 parameters, can be approximated in one or two dimensional active subspaces.

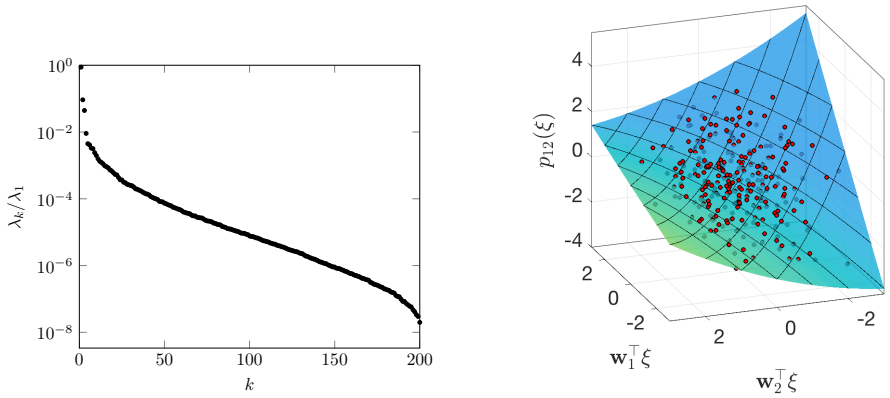


**Fig. 4** The normalized eigenvalues of output mode $p_{12}$ (left) and the 2D sufficient summary plot corresponding to $p_{12}$ (right). Note that the first two eigenvalues are very close to each other.

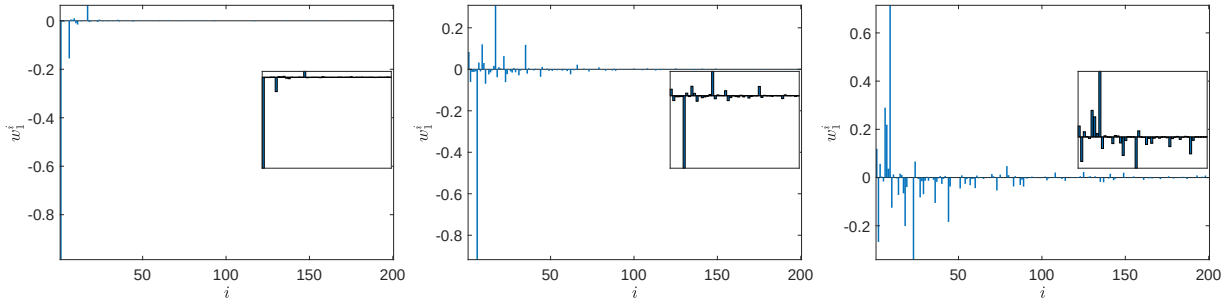


**Fig. 5** Components of the dominant eigenvector $\boldsymbol{w}$ for the output KL modes, $p_1$, $p_2$, and $p_3$ (left, middle, and right images, respectively). The inset plot shows the first 50 components in each case.

Next, we compute the surrogate models for the output KL modes, following the strategy described in Section 3. The surrogate models $\tilde{G}_k(\mathbf{W}_{k,1}^T\boldsymbol{\xi}) \approx p_k(\boldsymbol{\xi})$, $k = 1, \ldots, N$ are constructed by regression fit. Specifically, letting $p_{i,k} = p_k(\boldsymbol{\xi}_i)$ and $y_i = \mathbf{W}_{k,1}^T\boldsymbol{\xi}$, $i = 1, \ldots, N_s$, we compute least squares approximating polynomials $\tilde{G}_k$ that minimize

$$\sum_{i=1}^{N_s}(\tilde{G}_k(y_i) - p_{i,k})^2, \quad k = 1, \ldots, N.$$

In our computations, we found that linear regressions fits were suitable for the dominant output KL modes, except modes 1, 11, and 12, for which we used a quadratic fit. For illustration, we report the computed surrogate models for the first three output modes in Figure 3 (bottom) and for the mode $p_{12}$ in Figure 4 (right). We now have all the pieces to form the surrogate model for the pressure field $p(\boldsymbol{x}, \boldsymbol{\xi})$:

$$p(\boldsymbol{x}, \boldsymbol{\xi}) \approx \hat{p}(\boldsymbol{x}, \boldsymbol{\xi}) = \bar{p}(\boldsymbol{x}) + \sum_{k=1}^{N} \sqrt{\lambda_k}\tilde{G}_k(\mathbf{W}_{k,1}^T\boldsymbol{\xi})\phi_k(\boldsymbol{x}). \tag{22}$$

Below, we examine the accuracy of the computed surrogate model and show its effectiveness in capturing the statistical properties of the pressure field.

### 4.4 The accuracy of the surrogate model

In this section, we provide various tests of accuracy that examine different aspects of the proposed method. To provide a baseline for comparision, we computed 10,000 realizations of the exact pressure field and its surrogate model approximation.

We begin by examining the success of the low-rank KL approximation of the pressure field in capturing the variance of the process. The variance of the pressure field can be obtained from $\mathrm{Var}(p(\boldsymbol{x},\boldsymbol{\xi})) = \mathbb{E}\{p(\boldsymbol{x},\boldsymbol{\xi})^2\} - \mathbb{E}\{p(\boldsymbol{x},\boldsymbol{\xi})\}^2$, which we approximate using the computed samples of $p(\boldsymbol{x},\boldsymbol{\xi})$. The variance field for the truncated KL exansion of $p$ is determined completely by the spectral decomposition of its (approximate) covariance operator:

$$\mathrm{Var}(\hat{p}(\boldsymbol{x},\boldsymbol{\xi})) = \sum_{k=1}^{N} \lambda_k \phi_k(\boldsymbol{x})^2,$$

where $\lambda_k$ and $\phi_k(\boldsymbol{x})$ are the eigenvalues and eigenvectors of the covariance operator $C_p$. Taking the square root we obtain the standard deviation of both the exact pressure field and its approximation; results are shown in Figure 6. We note that for both, the standard deviation is highest at the center of the tumor and decreases to zero as we move away from the center. We also observe that even a low-rank approximation to $p(\boldsymbol{x},\boldsymbol{\xi})$ captures the standard deviation of the pressure field well.

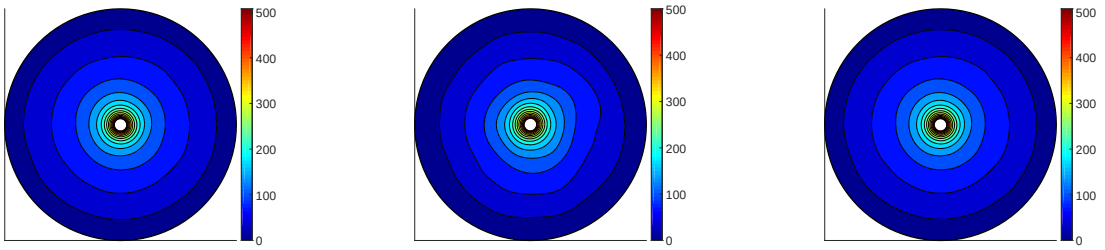


**Fig. 6** (Left) standard deviation of $p(\boldsymbol{x},\boldsymbol{\xi})$ using sampling, (middle) standard deviation of $p(\boldsymbol{x},\boldsymbol{\xi})$ using KLE (15 modes), and (right) standard deviation of $p(\boldsymbol{x},\boldsymbol{\xi})$ using KLE (50 modes).

Next, we study the accuracy of the computed surrogate model by focusing on the relative error indicator

$$E_{rel}(\boldsymbol{\xi}) = \frac{\int_{\mathcal{D}} |p(\boldsymbol{x},\boldsymbol{\xi}) - \hat{p}(\boldsymbol{x},\boldsymbol{\xi})|\, d\boldsymbol{x}}{\int_{\mathcal{D}} |p(\boldsymbol{x},\boldsymbol{\xi})\, d\boldsymbol{x}},$$

where $\hat{p}(\boldsymbol{x},\boldsymbol{\xi})$ is computed according to (22). In Figure 7 (left), we plot the expected value of $E_{rel}$ as the number of output KL modes increases; the dashed lines indicate the fifth and ninty fifth percentiles. We note that with $N = 15$ output KL modes, the relative error is about 5% on average. To better understand the behavior of the relative error, we report its distribution in Figure 7 (right), where we used $N = 15$ output KL modes in computing $\hat{p}$. We also show the distribution of the relative error computed over a subdomin $\mathcal{D}' = \{\boldsymbol{x} \in \mathcal{D} : \|\boldsymbol{x}\| \leq 2\}$ in Figure 7 (right). This is done to quantify the approximation errors near the injection site. We observe that the distribution of the error is shifted to the left, indicating smaller approximation errors in $\mathcal{D}'$ (with high probability). To see more clearly how the distribution of the relative error evolves as the number of KL modes increase, we show the probability density function of $E_{rel}$ corresponding to different number of output KL modes in Figure 8. Note that not only does the mode of the distribution get smaller, the spread of the error also decreases, which can be inferred from Figure 7 (left) as well.

In Figure 9, we report the distribution of the average pressure along concentric circles of various radii. This shows that the computed surrogate captures statistical properties of the pressure well in different parts of the domain.

## 5 Concluding remarks

We have presented a distributed active subspace method for scalable surrogate modeling of PDE-governed physical processes with high-dimensional inputs and function valued outputs. To save the modeling efforts spent on function valued outputs, we employ the truncated KL expansion to decouple the spatial dimensions with those of the random variables. As a result, the randomness in outputs is fully represented by scalar valued KL modes. For elliptic PDEs, as observed in the work, a low-rank KL representation of the model output is usually sufficient for an accurate representation. To reduce the dimension of inputs, we construct active subspaces for each of the dominant output

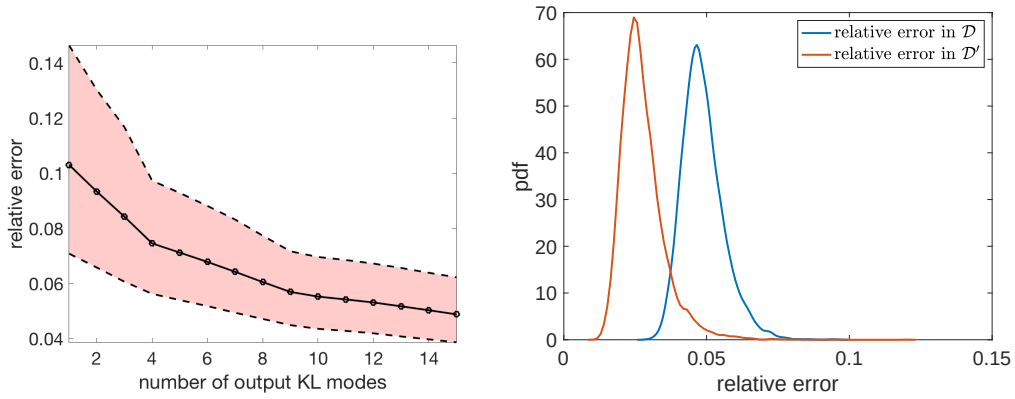


**Fig. 7** Left: The expected value of the error $E_{rel}$ (solid black line) along with the fifth and the ninty fifth percentiles (dashed lines) of $E_{rel}$. Right: Distribution of $E_{rel}$ computed over the entire domain $\mathcal{D}$ (blue) and over the subdomian $\mathcal{D}'$ (red).

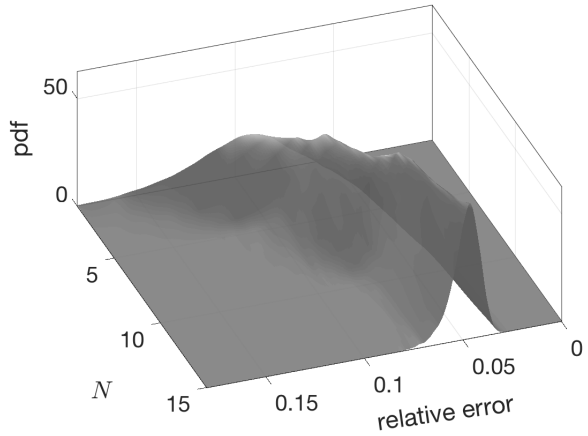


**Fig. 8** Distribution of the error $E_{rel}$ as the number of output KL modes increase from $N = 1$ to $N = 15$.

KL modes. Since output gradients with respect to the random variables need to be calculated when constructing active subspaces, we develop an adjoint-based framework to ensure that the computational cost does not scale with the input dimension. The method development is complemented by a rigorous mathematical formulation as well as theoretical analysis of errors due to active subspace projection and output KL representation.

We then deploy the distributed active subspace method to conduct surrogate modeling of the pressure field in a biotransport model in tumors with an uncertain permeability field. We demonstrate that a low-rank representation of the pressure field (i.e., output) can be achieved with the truncated KL expansion. The input (i.e., random variables used to represent the uncertain log-permeability field) dimension can be very high (e.g., several hundreds) when the correlation length of the log-permeability field is small. However, we observe that dominant output KL modes admit low (one or two) dimensional active subspaces. We observe that the average relative error between the surrogate model and the PDE solution can be controlled under 5% even when a very low-rank representation of the outputs is employed. We also show that the surrogate models can capture statistical properties of the pressure well in different parts of the domain.

In future work, we will investigate extensions of the distributed active subspace method to surrogate modeling of the time-dependent diffusion and convection-diffusion processes in biological and geological flows. We envision that the truncated KL expansion can be used to extract spatiotemporal coherent structures from the physical process, and the low-dimensional active subspaces can then be constructed for the scalar valued KL modes associated with these structures. This surrogate modeling approach will contribute to cost-effective forward uncertainty quantification, and facilitate solution of inverse problems under uncertainty, such as Bayesian inversion of tissue material properties from medical images.

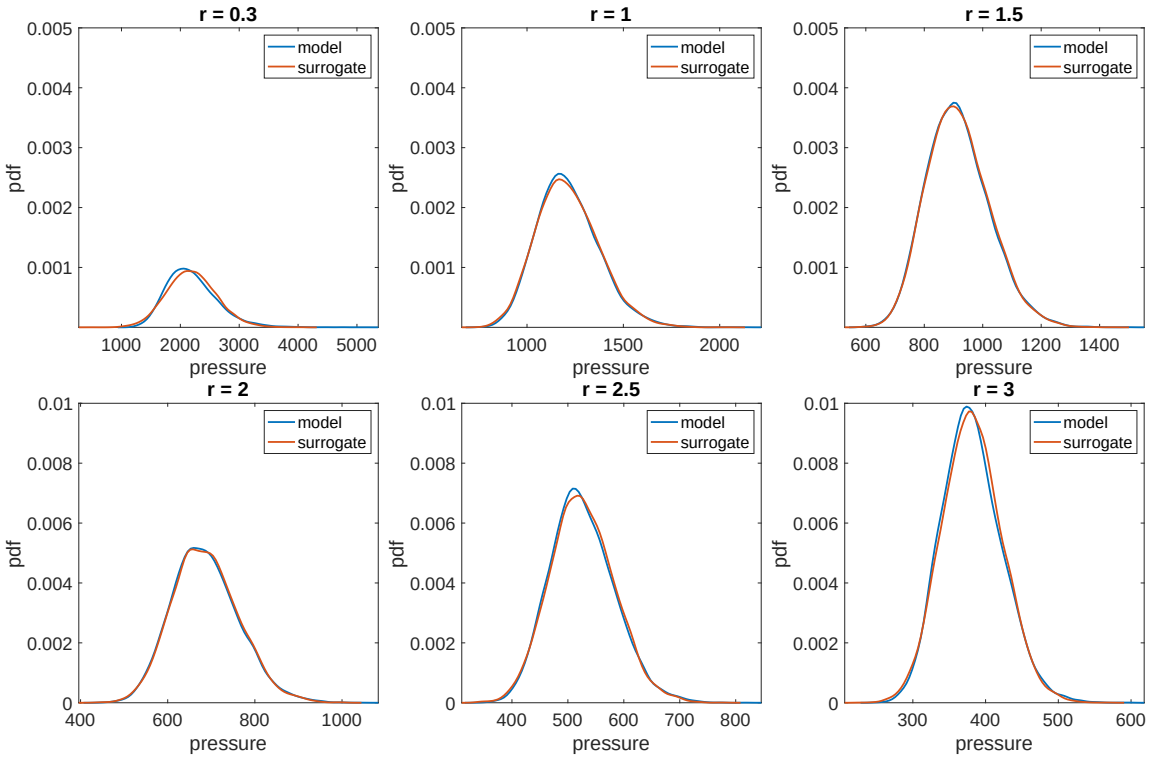


**Fig. 9** Distribution of the pressure averaged over concentric circles of various radii.

## References


1. Alexanderian, A., Gremaud, P., Smith, R.: Variance-based sensitivity analysis for time-dependent processes. In review (arXiv: `https://arxiv.org/abs/1711.08030`) (2019)
2. Alexanderian, A., Reese, W., Smith, R.C., Yu, M.: Model input and output dimension reduction using Karhunen–Loève expansions with application to biotransport. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems Part B: Mechanical Engineering **Accepted** (2019). `https://ui.adsabs.harvard.edu/#abs/2019arXiv190306314A`
3. Alexanderian, A., Zhu, L., Salloum, M., Ma, R., Yu, M.: Investigation of biotransport in a tumor with uncertain material properties using a non-intrusive spectral uncertainty quantification method. J. Biomech. Eng. pp. 091006–1–091006–11 (2017)
4. Babuška, I., Nobile, F., Tempone, R.: A stochastic collocation method for elliptic partial differential equations with random input data. SIAM Journal on Numerical Analysis **45**(3), 1005–1034 (2007)
5. Clark, W.H.: Tumour progression and the nature of cancer. Br J Cancer **64**, 631–44 (1991)
6. Clark, W.H.: Biphasic finite element model of solute transport for direct infusion into nervous tissue. Annals of Biomedical Engineering **35**, 2145âĂŞ–2158 (2007)
7. Cleaves, H.L., Alexanderian, A., Guy, H., Smith, R.C., Yu, M.: Derivative-based global sensitivity analysis for models with high-dimensional inputs and functional outputs. arXiv e-prints arXiv:1902.04630 (2019)
8. Constantine, P.: Active Subspaces: Emerging Ideas in Dimension Reduction for Parameter Studies. SIAM, Philadelphia (2015)
9. Constantine, P.G., Diaz, P.: Global sensitivity metrics from active subspaces. Reliability Engineering & System Safety **162**, 1–13 (2017)
10. Constantine, P.G., Doostan, A.: TimeâĂŘdependent global sensitivity analysis with active subspaces for a lithium ion battery model. Statistical Analysis and Data Mining: The ASA Data Science Journal **10**, 243–262 (2017)
11. Constantine, P.G., Dow, E., Wang, Q.: Active subspace methods in theory and practice: applications to kriging surfaces. SIAM Journal on Scientific Computing **36**(4), A1500–A1524 (2014)
12. Constantine, P.G., Emory, M., Larsson, J., Iaccarino, G.: Exploiting active subspaces to quantify uncertainty in the numerical simulation of the hyshot ii scramjet. Journal of Computational Physics **302**, 1–20 (2015)
13. Debbage, P.: Targeted drugs and nanomedicine: present and future. Current Pharmaceutical Design **15**, 153–72 (2009)
14. Doostan, A., Ghanem, R.G., Red-Horse, J.: Stochastic model reduction for chaos representations. Computer Methods in Applied Mechanics and Engineering **196**(37-40), 3951–3966 (2007)
15. Elman, H.: Solution algorithms for stochastic galerkin discretizations of differential equations with random data. Handbook of Uncertainty Quantification pp. 1–16 (2017)
16. Friedman, J.: Fast MARS. Tech. Rep. 110, Laboratory for Computational Statistics, Department of Statistics, Stanford University (1993)
17. Gamboa, F., Janon, A., Klein, T., Lagnoux, A., et al.: Sensitivity analysis for multidimensional and functional outputs. Electronic Journal of Statistics **8**(1), 575–603 (2014)

18. Ghanem, R.: Probabilistic characterization of transport in heterogeneous media. Computer Methods in Applied Mechanics and Engineering **158**(3), 199 – 220 (1998). DOI https://doi.org/10.1016/S0045-7825(97)00250-8. URL http://www.sciencedirect.com/science/article/pii/S0045782597002508
19. Ghanem, R.G., Spanos, P.D.: Stochastic finite elements: a spectral approach. Springer-Verlag, New York (1991). DOI 10.1007/978-1-4612-3094-6. URL http://dx.doi.org/10.1007/978-1-4612-3094-6
20. Graham, I.G., Kuo, F.Y., Nichols, J.A., Scheichl, R., Schwab, C., Sloan, I.H.: Quasi-monte carlo finite element methods for elliptic pdes with lognormal random coefficients. Numerische Mathematik **131**(2), 329–368 (2015)
21. Gunzburger, M.: Perspectives in flow control and optimization, vol. 5. Siam (2003)
22. Hsing, T., Eubank, R.: Theoretical foundations of functional data analysis, with an introduction to linear operators. John Wiley & Sons (2015)
23. Iooss, B., Saltelli, A.: Introduction to sensitivity analysis. In: R. Ghanem, D. Higdon, H. Owhadi (eds.) Handbook of uncertainty quantification, pp. 1103–1122. Springer (2017)
24. Jefferson, J., Gilbert, J., Constantine, P., Maxwell, R.: Active subspaces for sensitivity analysis and dimension reduction of an integrated hydrologic model. Computers & Geosciences **83**, 127–138 (2015)
25. Ji, W., Wang, J., Zahm, O., Marzouk, Y.M., Yang, B., Ren, Z., Law, C.K.: Shared low-dimensional subspaces for propagating kinetic uncertainty to multiple outputs. Combustion and Flame **190**, 146–157 (2018)
26. Kucherenko, S., Iooss, B.: Derivative-based global sensitivity measures. In: R. Ghanem, D. Higdon, H. Owhadi (eds.) Handbook of Uncertainty Quantification. Springer (2017)
27. Le Maıtre, O., Knio, O., Najm, H., Ghanem, R.: Uncertainty propagation using wiener–haar expansions. Journal of computational Physics **197**(1), 28–57 (2004)
28. Le Maître, O.P., Knio, O.M.: Spectral methods for uncertainty quantification. Scientific Computation. Springer, New York (2010). DOI 10.1007/978-90-481-3520-2. URL http://dx.doi.org/10.1007/978-90-481-3520-2. With applications to computational fluid dynamics
29. Le Maître, O.P., Reagan, M.T., Najm, H.N., Ghanem, R.G., Knio, O.M.: A stochastic projection method for fluid flow: Ii. random process. Journal of computational Physics **181**(1), 9–44 (2002)
30. Loève, M.: Probability theory. I, fourth edn. Springer-Verlag, New York-Heidelberg (1977). Graduate Texts in Mathematics, Vol. 45
31. Lukaczyk, T.W., Palacios, F., Alonso, J.J., Constantine, P.: Active subspaces for shape optimization. In: the 10th AIAA Multidisciplinary Design Optimization Conference. National Harbor, Maryland (2014). AIAA-2014-1171
32. Ma, R., Su, D., Zhu, L.: Multiscale simulation of nanopartical transport in deformable tissue during an infusion process in hyperthermia treatments of cancers. In: W.J. Minkowycz, E. Sparrow, J.P. Abraham (eds.) Nanoparticle Heat Transfer and Fluid Flow, Computational & Physical Processes in Mechanics & Thermal Science Series, vol. 4. CRC Press, Taylor & Francis Group (2012)
33. Mangado, N., Piella, G., Noailly, J., Pons-Prats, J., ÃĄngel GonzÃąlez Ballester, M.: Analysis of uncertainty and variability in finite element computational models for biomedical engineering: Characterization and propagation. Front Bioeng Biotechnol. **4**, 85 (2016)
34. Matthies, H.G., Keese, A.: Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. Computer methods in applied mechanics and engineering **194**(12-16), 1295–1331 (2005)
35. Prieur, C., Tarantola, S.: Variance-based sensitivity analysis: Theory and estimation algorithms. In: R. Ghanem, D. Higdon, H. Owhadi (eds.) Handbook of Uncertainty Quantification, pp. 1217–1239. Springer (2017)
36. Rasmussen, C.E., Williams, C.: Gaussian processes for machine learning cambridge (2006)
37. Russi, T.M.: Uncertainty quantification with experimental data and complex system models. Ph.D. thesis, University of California, Berkeley (2010)
38. Saad, G., Ghanem, R.: Characterization of reservoir simulation models using a polynomial chaos-based ensemble kalman filter. Water Resources Research **45**(4) (2009)
39. Salloum, M., Ma, R., Weeks, D., Zhu, L.: Controlling nanoparticle delivery in magnetic nanoparticle hyperthermia for cancer treatment: experimental study in agarose gel. Int. J. Hyperthermia **24**, 337–345 (2008)
40. Sobol, I.: Estimation of the sensitivity of nonlinear mathematical models. Matematicheskoe Modelirovanie **2**(1), 112–118 (1990)
41. Sobol, I.: Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. Mathematics and Computers in Simulation **55**(1-3), 271 – 280 (2001)
42. Sobol', I., Kucherenko, S.: Derivative based global sensitivity measures and their link with global sensitivity indices. Mathematics and Computers in Simulation **79**(10), 3009–3017 (2009)
43. Xiu, D., Karniadakis, G.E.: Modeling uncertainty in flow simulations via generalized polynomial chaos. Journal of Computational Physics **187**(1), 137 – 167 (2003). DOI https://doi.org/10.1016/S0021-9991(03)00092-5. URL http://www.sciencedirect.com/science/article/pii/S0021999103000925
44. Zahm, O., Constantine, P., Prieur, C., Marzouk, Y.: Gradient-based dimension reduction of multivariate vector-valued functions. arXiv preprint arXiv:1801.07922 (2018)

## A Proofs

*Proof of Theorem 1:* Note that,

$$e(\boldsymbol{x},\boldsymbol{\xi}) = \left| \sum_{k=1}^{N} (F_k(\boldsymbol{\xi}) - G_k(\boldsymbol{W}_{k,1}^T\boldsymbol{\xi}))\phi_k(\boldsymbol{x}) \right| \leq \sum_{k=1}^{N} |F_k(\boldsymbol{\xi}) - G_k(\boldsymbol{W}_{k,1}^T\boldsymbol{\xi})||\phi_k(\boldsymbol{x})|.$$

Then,

$$\begin{aligned}\mathbb{E}\{\bar{e}\} &= \mathbb{E}\Big\{ \int_{\mathcal{X}} \sum_{k=1}^{N} |F_k(\boldsymbol{\xi}) - G_k(\boldsymbol{W}_{k,1}^T\boldsymbol{\xi})||\phi_k(\boldsymbol{x})|dx \Big\} \\ &= \sum_{k=1}^{N} \Big( \int_{\mathcal{X}} |\phi_k(\boldsymbol{x})|dx \Big) \mathbb{E}\{|F_k(\boldsymbol{\xi}) - G_k(\boldsymbol{W}_{k,1}^T\boldsymbol{\xi})|\} \\ &\leq \sum_{k=1}^{N} \left[ \int_{\mathcal{X}} |\phi_k(\boldsymbol{x})|^2 dx \right]^{\frac{1}{2}} \left[ \int_{\mathcal{X}} 1^2 dx \right]^{\frac{1}{2}} \mathbb{E}\{|F_k(\boldsymbol{\xi}) - G_k(\boldsymbol{W}_{k,1}^T\boldsymbol{\xi})^2|\}^{1/2} \leq |\mathcal{X}|^{\frac{1}{2}} \sum_{k=1}^{N} \delta_k\end{aligned}$$

where we have used Cauchy–Schwarz inequlity and Lemma 2(a). □

*Proof of Theorem 2:* We have

$$\begin{aligned}\bar{E}(\boldsymbol{\xi}) = \int_{\mathcal{X}} E(\boldsymbol{x},\boldsymbol{\xi})dx &= \int_{\mathcal{X}} |f(\boldsymbol{x},\boldsymbol{\xi}) - f_N(\boldsymbol{x},\boldsymbol{\xi}) + f_N(\boldsymbol{x},\boldsymbol{\xi}) - \hat{f}_N(\boldsymbol{x},\boldsymbol{\xi})| \\ &\leq \int_{\mathcal{X}} |f(\boldsymbol{x},\boldsymbol{\xi}) - f_N(\boldsymbol{x},\boldsymbol{\xi})|dx + \int_{\mathcal{X}} |f_N(\boldsymbol{x},\boldsymbol{\xi}) - \hat{f}_N(\boldsymbol{x},\boldsymbol{\xi})|dx \\ &\leq \Big\{ \int_{\mathcal{X}} |f(\boldsymbol{x},\boldsymbol{\xi}) - f_N(\boldsymbol{x},\boldsymbol{\xi})|^2 dx \Big\}^{1/2} |\mathcal{X}|^{\frac{1}{2}} + \bar{e}(\boldsymbol{\xi}) \\ &= \Big\{ \sum_{k=N+1}^{\infty} F_k(\boldsymbol{\xi})^2 \Big\}^{1/2} |\mathcal{X}|^{\frac{1}{2}} + \bar{e}(\boldsymbol{\xi}).\end{aligned}$$

Now we consider

$$\begin{aligned}\mathbb{E}\{\bar{E}(\boldsymbol{\xi})\} &\leq \mathbb{E}\left( \Big\{ \sum_{k=N+1}^{\infty} F_k(\boldsymbol{\xi})^2 \Big\}^{1/2} \right) |\mathcal{X}|^{\frac{1}{2}} + \mathbb{E}\{\bar{e}(\boldsymbol{\xi})\} \\ &\leq |\mathcal{X}|^{\frac{1}{2}} \left[ \mathbb{E}\left( \Big\{ \sum_{k=N+1}^{\infty} F_k(\boldsymbol{\xi})^2 \Big\}^{1/2} \right) + \sum_{k=1}^{N} \delta_k \right] \leq |\mathcal{X}|^{\frac{1}{2}} \left[ \mathbb{E}\left( \sum_{k=N+1}^{\infty} F_k(\boldsymbol{\xi})^2 \right)^{1/2} + \sum_{k=1}^{N} \delta_k \right].\end{aligned}$$

Note that $\mathbb{E}\left( \sum_{k=N+1}^{\infty} F_k(\boldsymbol{\xi})^2 \right) = \sum_{k=N+1}^{\infty} \mathbb{E}\big(F_k(\boldsymbol{\xi})^2\big) = \sum_{k=N+1}^{\infty} \lambda_k(C_f)$. Therefore, we have our desired result

$$\begin{aligned}\mathbb{E}\{\bar{E}(\boldsymbol{\xi})\} &\leq |\mathcal{X}|^{\frac{1}{2}} \left[ \left( \sum_{k=N+1}^{\infty} \lambda_k(C_f) \right)^{1/2} + \sum_{k=1}^{N} \delta_k \right] \\ &= |\mathcal{X}|^{\frac{1}{2}} \left[ \left( \sum_{k=N+1}^{\infty} \lambda_k(C_f) \right)^{1/2} + \sum_{k=1}^{N} \left( \sum_{j=r_k+1}^{N_p} \lambda_j(\mathbf{S}_k) \right)^{1/2} \right].\end{aligned}$$

□