Journal of Computational Physics 336 (2017) 36–50

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Multi-fidelity Gaussian process regression for prediction of random fields

L. Parussini [a], D. Venturi [b,*], P. Perdikaris [c], G.E. Karniadakis [d]

[a] Department of Engineering and Architecture, University of Trieste, Italy
[b] Department of Applied Mathematics and Statistics, University of California Santa Cruz, USA
[c] Department of Mechanical Engineering, Massachusetts Institute of Technology, USA
[d] Division of Applied Mathematics, Brown University, USA

ARTICLE INFO

Article history:
Received 17 July 2016
Received in revised form 10 November 2016
Accepted 23 January 2017
Available online xxxx

Keywords:
Gaussian random fields
Multi-fidelity modeling
Recursive co-kriging
Uncertainty quantification

ABSTRACT

We propose a new multi-fidelity Gaussian process regression (GPR) approach for prediction of random fields based on observations of surrogate models or hierarchies of surrogate models. Our method builds upon recent work on recursive Bayesian techniques, in particular *recursive co-kriging*, and extends it to vector-valued fields and various types of covariances, including separable and non-separable ones. The framework we propose is general and can be used to perform uncertainty propagation and quantification in model-based simulations, multi-fidelity data fusion, and surrogate-based optimization. We demonstrate the effectiveness of the proposed recursive GPR techniques through various examples. Specifically, we study the stochastic Burgers equation and the stochastic Oberbeck–Boussinesq equations describing natural convection within a square enclosure. In both cases we find that the standard deviation of the Gaussian predictors as well as the absolute errors relative to benchmark stochastic solutions are very small, suggesting that the proposed multi-fidelity GPR approaches can yield highly accurate results.

1. Introduction

High-fidelity numerical simulations of complex stochastic dynamical systems require substantial computing time and data storage even in modern parallel architectures. This inherently limits the number of system states we can reliably simulate, thereby affecting accuracy when inferring statistical properties of any phase space function such as the performance of a certain engineering design. This basic observation has recently driven an explosive growth of fundamental and practical research at the interface of high-performance scientific computing, probability theory, and applied mathematics. One of the main features of such research is to replace expensive computational models with cheap surrogates or *hierarchies of surrogates*, and then come up with mathematical techniques leveraging on cross-correlations between the output of different surrogates to infer a quantity of interest. This yields a *multi-fidelity approach computational modeling*, which was first proposed by Kennedy and O'Hagan [19,20] in a Bayesian setting, and since then used in many different disciplines (see, e.g., [25,9]). For example, in [3] the authors presented a non-intrusive framework based on treed multi-output Gaussian processes, in which the response statistics are obtained through sampling a properly trained surrogate model of the physical system. The tree is built in a sequential way and its refinement depends on the observations through a global measure of

* Corresponding author.
E-mail address: venturi@ucsc.edu (D. Venturi).

http://dx.doi.org/10.1016/j.jcp.2017.01.047
0021-9991/

the uncertainty in the prediction, the inferred length scales, as well as the input probability distribution. Gaussian process regression has also been studied in the context of hierarchical multi-scale modeling of materials. For example, in [22] an adaptive moving kriging interpolation method was proposed to reduce the number of model evaluations at the fine-scale and inform the coarse-scale model with essential constitutive information. A multi-fidelity approach for minimizing the number of evaluations of expensive high-fidelity models was also proposed in [23], where the statistics of the high-fidelity model are computed based on realizations of a corrected low-fidelity surrogate. The correction function can be additive, multiplicative, or a combination of the two and it may be updated occasionally by high fidelity model evaluations.

In this paper we propose a new multi-fidelity Gaussian process regression (GPR) approach for predicting finite-dimensional random fields based on observations of surrogate models or hierarchies of surrogate models. The key idea relies on representing the map between the space of (random) Fourier or other spectral coefficients associated with any series expansion (relative to a spatial basis) and the physical space in which the random field develops. In this setting, the multi-fidelity inference problem for a random field reduces to an inference problem of a multivariate random vector of Fourier coefficients given data, i.e., vector samples produced by surrogate models at different levels of fidelity. To perform such inference, one can apply the recursive co-kriging technique recently developed by Le Gratiét et al. [15,13] (see also [25]) to each component of the vector of Fourier coefficients. From a Bayesian standpoint, this is equivalent to assuming independent priors for each model output, which may result in loss of information. To overcome this issue, we extend the recursive co-kriging technique to a multivariate setting (statistical models with vector outputs), to capture the cross-correlation structure among different vector components. We consider both separable and non-separable priors and quantify the advantages and trade-offs of each approach. We remark that although the main focus of this paper is revolving around uncertainty propagation and quantification in model-based computations, the proposed framework can be readily applied to more general parametric studies such as multivariate inverse problems and multi-objective surrogate-based optimization.

The paper is organized as follows. In Section 2 we introduce the general framework along with some theoretical background. In Section 3 we give a brief overview of Gaussian process regression (GPR) methods for vector-valued random fields and introduce our multi-fidelity recursive GPR technique. In Section 4 we evaluate the accuracy of the proposed multi-fidelity GPR approach by applying it to the stochastic Burgers equation and a stochastic thermal convection problem. Finally, in Section 5 we summarize our main findings.

2. Methodology

Consider a scalar random field $u(\boldsymbol{x}, \boldsymbol{\xi})$ depending on a set of coordinates (or design variables) $\boldsymbol{x} \in \mathbb{R}^n$, as well as on a set of random parameters $\boldsymbol{\xi} \in \mathbb{R}^d$. The field u could be, e.g., the solution to a partial differential equation in which the boundary conditions are set to be random and represented in terms of $\boldsymbol{\xi}$. Suppose that $u(\boldsymbol{x}, \boldsymbol{\xi})$ is in a separable Hilbert space. This allows us to write the series expansion

$$u(\boldsymbol{x}, \boldsymbol{\xi}) \cong \sum_{i=1}^{k} a_i(\boldsymbol{\xi}) L_i(\boldsymbol{x}), \tag{1}$$

where $L_i(\boldsymbol{x})$ are basis functions depending on the coordinates (or design variables) $\boldsymbol{x}$ while $a_i(\boldsymbol{\xi})$ are functions of random variables $\boldsymbol{\xi}$. If $u(\boldsymbol{x}, \boldsymbol{\xi})$ is the solution to a stochastic PDE model, then $L_i(\boldsymbol{x})$ are usually set a priori (spatial basis functions), while the functions $a_i(\boldsymbol{\xi})$ are determined by the PDE, e.g., by computing its solution at specific values of $\boldsymbol{\xi}$ through the probabilistic collocation method [1,8,4]. At this point we pose the following question: can we determine a model for the random vector field

$$\boldsymbol{a}(\boldsymbol{\xi}) = [a_1(\boldsymbol{\xi}) \cdots a_k(\boldsymbol{\xi})] \tag{2}$$

based on data collected at a specific nodes in the $\boldsymbol{\xi}$-space? This question is obviously not new and researchers have been working on it for decades. For instance, one can use polynomial interpolation of each $a_i(\boldsymbol{\xi})$ at Chebyshev sparse grids [2]. However, this implicitly assumes that $a_i(\boldsymbol{\xi})$ is a multivariate polynomial (which we do not know for sure), and also that we can predict the value of $a_i(\boldsymbol{\xi})$ at a non-observed location with probability 1, i.e., with no uncertainty. This is obviously not correct from a statistical standpoint. In a more robust setting, $a_i(\boldsymbol{\xi})$ should be considered as a random field with known values at observed points. Following the classical literature [18,19,6,29,12], we shall assume that the distribution of the vector $\boldsymbol{a}(\boldsymbol{\xi})$, conditional to the realization $\{\boldsymbol{a}_1 = \boldsymbol{a}(\boldsymbol{\xi}_1), ..., \boldsymbol{a}_n = \boldsymbol{a}(\boldsymbol{\xi}_n)\}$, is Gaussian with mean $\boldsymbol{m}(\boldsymbol{\xi})$ and (matrix-valued) covariance function $\boldsymbol{C}(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)$, i.e.,

$$\boldsymbol{a}(\boldsymbol{\xi}) | \boldsymbol{a}_1, ..., \boldsymbol{a}_n \sim \mathcal{GP}\left(\boldsymbol{m}(\boldsymbol{\xi}), \boldsymbol{C}(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)\right). \tag{3}$$

As we shall see in the subsequent sections, this setting allows us to build a multi-fidelity Gaussian process regression framework, in which observations of $\boldsymbol{a}(\boldsymbol{\xi})$ obtained from models with different levels of fidelity are combined in a seamless way to yield a highly accurate Gaussian predictor of $\boldsymbol{a}(\boldsymbol{\xi})$.