Journal of Computational Physics 260 (2014) 37–53

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Low-rank separated representation surrogates of high-dimensional stochastic functions: Application in Bayesian inference

AbdoulAhad Validi

Mechanical Engineering Department, Michigan State University, East Lansing, MI 48824, USA

ARTICLE INFO

Article history:
Received 25 June 2013
Received in revised form 20 October 2013
Accepted 8 December 2013
Available online 18 December 2013

Keywords:
Separated representation
Inverse problem
Bayesian inference
Uncertainty quantification
High-dimensional PDE/ODE

ABSTRACT

This study introduces a non-intrusive approach in the context of low-rank separated representation to construct a surrogate of high-dimensional stochastic functions, e.g., PDEs/ODEs, in order to decrease the computational cost of Markov Chain Monte Carlo simulations in Bayesian inference. The surrogate model is constructed via a regularized alternative least-square regression with Tikhonov regularization using a roughening matrix computing the gradient of the solution, in conjunction with a perturbation-based error indicator to detect optimal model complexities. The model approximates a vector of a continuous solution at discrete values of a physical variable. The required number of random realizations to achieve a successful approximation linearly depends on the function dimensionality. The computational cost of the model construction is quadratic in the number of random inputs, which potentially tackles the curse of dimensionality in high-dimensional stochastic functions. Furthermore, this vector-valued separated representation-based model, in comparison to the available scalar-valued case, leads to a significant reduction in the cost of approximation by an order of magnitude equal to the vector size. The performance of the method is studied through its application to three numerical examples including a 41-dimensional elliptic PDE and a 21-dimensional cavity flow.

1. Introduction

An inverse problem arises when the inputs/complexities of a model are estimated indirectly from outputs, e.g., noisy observations [58,39,55,12,5,13,57]. In this context, the Bayesian approaches, which have recently attracted much attention [6,19,53], provide applied probability and uncertainty measurements for statistical inference. Indeed, as an extension of conventional statistical methods [34,24,30,58], the solution of the Bayesian inference is a posterior probability distribution over the model input/complexities regarding available/unavailable prior knowledge about them [3]. The computational cost of estimating the posterior distribution is a challenge in practice, and, in response, many asymptotic, deterministic, and sampling based methods have been developed focusing on reductions of or surrogates the forward model [49,43,41,42,33].

Deterministic methods might be reasonable alternatives in low to moderate dimensions, but for high-dimensional and complex problems, the Markov Chain Monte Carlo (MCMC) [45] strategy is a more general and flexible approach [11,56, 22,31]. The MCMC approach requires evaluation of the likelihood function [46], indeed, solving the forward model many times, which might be costly and/or intractable. In the case of intensive computational models, e.g., those described by a system of Ordinary Differential Equations (ODEs) or Partial Differential Equations (PDEs), the cost of such an approach becomes prohibitive. To do so, generalized polynomial chaos (gPC)-based [59], Stochastic Galerkin, and Collocation methods

E-mail address: validiab@msu.edu.

0021-9991/$ – see front matter
http://dx.doi.org/10.1016/j.jcp.2013.12.024

[43,15] have been developed. However, these approaches are attractive alternatives for low (or moderate) dimensions. In the case of high-dimensional problems, *low-rank non-intrusive separated representation approximation* of the model is proposed to construct a surrogate of the forward model [8,9,2,16]. Then, for efficient Bayesian inference, this surrogate model is used in computing the likelihood function, as well as the posterior probability distribution function.

In 1927, Hitchcock [32] introduced separated representation, which is also known as parallel factor analysis or canonical decomposition, to express a Polyadic as a sum of products of rank-one vectors. Subsequently, this approach has been widely used in a variety of areas including chemical kinetics [16], data mining [1,36,29,37], and image processing [18,51]. Here, an approach is proposed to construct a vector-valued separated representation of a continuous stochastic function of a physical random variable ξ and input random variables $\boldsymbol{y}$, i.e., $\boldsymbol{u}(\xi, \boldsymbol{y})$, $\boldsymbol{y} = (y_1, \ldots, y_d)$, $d \in \mathbb{N}$. This function can be approximated with accuracy $\mathcal{O}(\boldsymbol{\epsilon})$ in a separated form as

$$\boldsymbol{u}\big(\xi, \boldsymbol{y}(w)\big) = \sum_{l=1}^{r} s_l \boldsymbol{u}_0^l(\xi) \prod_{i=1}^{d} u_i^l\big(y_i(w)\big) + \mathcal{O}(\boldsymbol{\epsilon}), \tag{1}$$

where $\boldsymbol{u}_0^l(\xi)$, which is a vector-valued univariate function of a physical variable ξ; $\{u_i^l(y_i(w))\}_{i=1}^d$, $l = 1, \ldots, r$, which are univariate functions of random variables; and s_l, which are normalization constants; are unknown and must be computed. Because the separation rank, r, as one of the model complexities, is independent of problem dimensionality [7,9,8,16], d, the computational complexity is a weakly-linear function in d, which remarkably reduces the curse of dimensionality, a bottleneck for uncertainty quantification of high-dimensional functions. Furthermore, the model has low-rank separated representation approximation structure if a small separation rank can be found for it.

This study is organized as follows. In Section 2, the Bayesian inference is discussed in more detail. In Section 3, the general problem setup described for either system of ODEs or PDEs. Thereafter, in Section 4, the vector case of the separated representation is introduced, and in Sections 4.1 and 4.2, a regularization approach and an error indicator are proposed for stabilizing the method and finding the optimum construction of the separated model. In Section 5, the results are presented for three different examples: a manufactured function, an elliptic equation, and a cavity flow problem.

2. Bayesian inference

The goal of an inverse problem is to recover anterior information from available data [12,58,5]. The quantity of interest, $\boldsymbol{u}$, in the forward problem context is computed given a mathematical model, $\mathcal{A}$, and parameters, $\boldsymbol{y}$; however, in the inverse problem either the parameters or the mathematical model is computed given the other two quantities.

Considering a general system of equations $\mathcal{A}(\boldsymbol{y}) \approx \boldsymbol{u}$, there are two main approaches for parameter estimation: main classical least squares[1] and Bayesian strategies. In the Bayesian approaches the model is treated as a random variable and the solutions are probability distributions for those model parameters that are sought [3]. Detailed statistical knowledge (e.g., mode, mean, standard deviation, correlation, smoothness, etc.) of the parameters can be revealed using the probability distributions, while in the classical methods the solutions are point quantities and the parameter statistics are not available. In Bayesian approaches, the prior information, which comes from other sources (e.g., physical and experimental observations), is called the "*prior distribution*" of parameters $\boldsymbol{y}$, and is denoted by $p(\boldsymbol{y})$. The "*posterior distribution*", $q(\boldsymbol{y}|\boldsymbol{u})$, can be formulated by incorporating the given data along with the prior distribution in Bayes' theorem [4,17] as follows:

$$q(\boldsymbol{y}|\boldsymbol{u}) = \frac{p(\boldsymbol{u}|\boldsymbol{y})p(\boldsymbol{y})}{\int p(\boldsymbol{u}|\boldsymbol{y})p(\boldsymbol{y})\,d\boldsymbol{y}}. \tag{2}$$

The data are incorporated in the formulation through the likelihood function $p(\boldsymbol{u})$, which can be presented as $L(\boldsymbol{y}) \equiv p(\boldsymbol{u}|\boldsymbol{y})$.

Remark 1. In reality, prior and posterior distributions show the strength of perception about feasible values for the inputs, $\boldsymbol{y}$, before and after experiencing the outputs, $\boldsymbol{u}$. More prior information on the model parameters, e.g., a range of possible values, leads to a more suitable prior distribution. If there is no available prior information about the parameters, then based on "the principle of indifference" an "uninformative" prior distribution is chosen; in which all the model parameter values are assumed to have the same likelihood.

In general, determining the posterior distribution is computationally expensive and problematic due to the integral in (2), which is usually a high-dimensional integral. A typical simplified model is assumed when the value of the integral is not really needed. In these situations two different model posterior distributions are compared by computing the likelihood functions; therefore Eq. (2) can be written as

$$q(\boldsymbol{y}|\boldsymbol{u}) \propto p(\boldsymbol{u}|\boldsymbol{y})p(\boldsymbol{y}). \tag{3}$$

[1] The classical least square approaches are linear and non-linear regression, and data free inference [58,35].