Journal of Computational Physics 282 (2015) 138–147

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Short note

Limitations of polynomial chaos expansions in the Bayesian solution of inverse problems

Fei Lu [a,b], Matthias Morzfeld [a,b,*], Xuemin Tu [c], Alexandre J. Chorin [a,b]

[a] Lawrence Berkeley National Laboratory, United States
[b] Department of Mathematics, University of California, Berkeley, United States
[c] Department of Mathematics, University of Kansas, United States

ARTICLE INFO

Article history:
Received 28 April 2014
Received in revised form 28 October 2014
Accepted 3 November 2014
Available online 18 November 2014

Keywords:
Polynomial chaos expansion
Bayesian inverse problem
Monte Carlo sampling

ABSTRACT

Polynomial chaos expansions are used to reduce the computational cost in the Bayesian solutions of inverse problems by creating a surrogate posterior that can be evaluated inexpensively. We show, by analysis and example, that when the data contain significant information beyond what is assumed in the prior, the surrogate posterior can be very different from the posterior, and the resulting estimates become inaccurate. One can improve the accuracy by adaptively increasing the order of the polynomial chaos, but the cost may increase too fast for this to be cost effective compared to Monte Carlo sampling without a surrogate posterior.

1. Introduction

There are many situations in science and engineering where one needs to estimate parameters in a model, for example, the permeability of a porous medium in studies of subsurface flow, on the basis of noisy and/or incomplete data, e.g. pressure measurements. In the Bayesian approach, prior information and a likelihood function for the data are combined to yield a posterior probability density function (pdf) for the parameters. This posterior can be approximated by Monte Carlo sampling and in principle yields all the information one needs, in particular the posterior mean (see e.g. [14,25,26]). However, the sampling may require the evaluation of the posterior for many values of the parameters, which in turn requires repeated solution of the forward problem. This can be expensive, especially in complex high-dimensional problems.

Polynomial chaos expansions (PCE) and generalized PCE provide an approximate representation of the solution of the forward problem (see e.g. [12,15,21,31]) which can be used to reduce the cost of Bayesian inverse problems [2,17–19,23]. The PCE leads to an approximate representation of the posterior, called a "surrogate posterior", which can generate a large number of samples at low cost. However, the resulting samples approximate the surrogate posterior, not the posterior, so that the accuracy of estimates based on these samples depends on how well the surrogate posterior approximates the posterior.

We study how the accuracy of the surrogate posterior depends on the data, and show that when the data are informative (in the sense that the posterior differs significantly from the prior), then the surrogate posterior can be very different from the posterior and PCE-based sampling is either inaccurate or prohibitively expensive. Specifically, we examine the behavior of PCE-based sampling in the small noise regime [28,29], and report results from numerical experiments on an elliptic inverse problem for subsurface flow. In the example, a sufficiently accurate PCE requires a high order, which makes

* Corresponding author.
E-mail address: mmo@math.lbl.gov (M. Morzfeld).

http://dx.doi.org/10.1016/j.jcp.2014.11.010
0021-9991/

PCE-based sampling expensive compared to sampling the posterior directly, without a PCE. Other limitations of PCE have been reported and discussed in other settings as well, e.g. in uncertainty quantification [3,5,13], and statistical hydrodynamics [6,10].

The paper is organized as follows. In Section 2 we explain the use of PCE in the Bayesian solution of inverse problems. In Section 3 we analyze the accuracy of the surrogate posterior in the small noise regime. In Section 4 we study the efficiency of PCE-based sampling with numerical examples. Section 5 provides a summary. Proofs and derivations can be found in Appendix A.

2. Polynomial chaos expansion for Bayesian inverse problems

Consider the problem of estimating model parameters $\theta \in \mathbb{R}^m$ from noisy data $d \in \mathbb{R}^n$ such that:

$$d = h(\theta) + \eta, \tag{1}$$

where $h : \mathbb{R}^m \to \mathbb{R}^n$ is a smooth nonlinear function describing how the parameters affect the data, and where $\eta \sim p_\eta(\cdot)$ is a random variable with known pdf that represents uncertainty in the measurements. Here, h is the model and often involves a partial differential equation (PDE), or a discretization of a PDE, in which case the evaluation of h can be computationally expensive. Following the Bayesian approach, we assume that prior information about the parameters is available in form of a pdf $p_0(\theta)$. This prior and the likelihood $p(d|\theta) = p_\eta(d - h(\theta))$, defined by (1), are combined in Bayes' rule to give the posterior pdf

$$p(\theta|d) = \frac{1}{\gamma(d)} p_0(\theta) p(d|\theta), \tag{2}$$

where $\gamma(d) = \int p_0(\theta) p(d|\theta) d\theta$ is a normalizing constant (the marginal probability of the data). For simplicity, we assume throughout this paper that $\eta \sim \mathcal{N}(0, \sigma^2 I_n)$ is Gaussian with mean zero and variance $\sigma^2 I_n$, and that the prior is $p_0(\theta) = \mathcal{N}(0, I_m)$ (here, I_k is the identity matrix of order k). These assumptions may be relaxed, however we can make our points in this simplified setting. In this context, it is important to point out that we make no assumptions about the underlying (numerical) model which, in most cases, is nonlinear.

In practice, Monte Carlo (MC) methods such as importance sampling or Markov chain Monte Carlo (MCMC) are used to represent the posterior numerically (see e.g. [7,16]). Most MC sampling methods require repeated evaluation of the posterior for many instances of θ. Since each posterior evaluation involves a likelihood evaluation, many evaluations of the model are needed, which can be computationally expensive.

To reduce the computational cost of MC sampling one can approximate the model by a truncation of its PCE, because the evaluation of the truncated PCE is often less expensive than the evaluation of the model (e.g. solving a PDE). It is natural to construct the PCE before the data are available, i.e. one expands h using the prior. With a Gaussian prior one uses (multivariate) Hermite polynomials, which form a complete orthonormal basis in $L^2(\mathbb{R}^m, p_0)$. Let $i = (i_1, \ldots, i_m) \in \mathbb{N}^m$ be a multi-index and let $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$ be the parameter we wish to estimate. The multivariate Hermite polynomials $\{\Phi_i(\theta) : |i| = i_1 + \cdots + i_m < \infty\}$ are defined by

$$\Phi_i(\theta) = H_{i_1}(\theta_1) \cdots H_{i_m}(\theta_m),$$

where $H_k(x)$ is the normalized kth-order Hermite polynomial (see e.g. [24,31]). Assuming that $h \in L^2(\mathbb{R}^m, p_0)$ we define the Nth-order PCE of h by

$$h_N(\theta) = \sum_{|i| \le N} a_i \Phi_i(\theta),$$

where the coefficients a_i are given by

$$a_i = E\big[h(\theta)\Phi_i(\theta)\big] = \int h(\theta)\Phi_i(\theta) p_0(\theta) d\theta.$$

As $N \to \infty$, h_N converges to h in $L^2(\mathbb{R}^m, p_0)$. The rate of convergence depends on the regularity of h and is estimated by (see e.g. [32])

$$\|h - h_N\|_{L^2(\mathbb{R}^m, p_0)} \le C N^{-\frac{k}{2}} \|h\|_{k,2}, \tag{3}$$

where C is a constant depending only on m and k, and $\|h\|_{k,2}$ is the weighted Sobolev norm defined by $\|h\|_{k,2}^2 = \sum_{|\alpha| \le k} \|D^\alpha h\|_{L^2(\mathbb{R}^m, p_0)}^2$ with $D^\alpha h = \frac{\partial^{|\alpha|}}{\partial_{x_1}^{\alpha_1} \cdots \partial_{x_m}^{\alpha_m}} h$. For the remainder of this paper we assume enough regularity of h, so that $\|h - h_N\|_{L^2(\mathbb{R}^m, p_0)}$ converges quickly as N increases.

In PCE-based sampling for Bayesian inverse problems, one replaces the model h in (1) by its truncated PCE h_N, and obtains the surrogate posterior

$$p_N(\theta|d) = \frac{1}{\gamma_N(d)} p_0(\theta) p_\eta\big(d - h_N(\theta)\big), \tag{4}$$