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

Fei Lu[1,2], Matthias Morzfeld[1,2], Xuemin Tu[3] and Alexandre J. Chorin[1,2]

[1]Lawrence Berkeley National Laboratory;

[2]Department of Mathematics
University of California, Berkeley;

[3]Department of Mathematics
University of Kansas.

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 require a high order, which makes 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 the appendix.

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

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

where the coefficients a_i are given by

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

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])

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

where C is a constant depending only on m and k, and $\|h\|_{k,2}$ is the weighted Sobolev norm defined by $\|h\|_{k,2}^2 = \sum_{|\alpha|\leq k}\|D^\alpha h\|_{L^2(\mathbb{R}^m, p_0)}^2$ 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 with N.

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

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

where $\gamma_N(d) = \int p_0(\theta)p_\eta(d - h_N(\theta))d\theta$. This surrogate posterior converges to the posterior at the same rate as $\|h - h_N\|_{L^2(\mathbb{R}^m, p_0)}$ converges to zero as $N \to \infty$ (see equation (3)) in the Kullback-Leibler divergence (KLD) [19] and in the Hellinger metric [25].

In practice, PCEs of small to moderate order are used because otherwise PCEs become expensive (see e.g. [15,31] and section 4). This truncation introduces error unless the problem is linear or well represented by a low-order polynomial. If the truncation error is large, then the surrogate posterior may be very different from the posterior. The samples one draws with PCE-based sampling methods approximate the surrogate posterior, which implies that the applicability of PCE-based sampling for inverse problems depends on how well the surrogate posterior approximates the posterior, i.e. on the accuracy of the surrogate posterior.

3 Accuracy of the surrogate posterior

We wish to study the effects of inaccuracies of a truncated PCE on PCE-based sampling methods for inverse problems. Inaccuracies in the PCE are caused by interaction of two mechanisms:

1. The error due to truncation is large if the physical model $h(\theta)$ in (1) is poorly represented by a low-order polynomial.

2. The surrogate posterior must be constructed based on the prior (see above). However, the posterior can put significant probability mass in parameter regions which are unlikely with respect to the prior. Thus, if the model is nonlinear, the PCE may be a poor approximation in the region of large posterior probability. In this case, the surrogate posterior is a poor approximation of the posterior in the region where significant posterior probability mass is located.

We assume throughout that $h(\theta)$ is nonlinear with high-order polynomials in its PCE, so that a truncated PCE (of moderate order) is only locally accurate. What we want to find are the conditions under which the lack of global accuracy of the PCE causes PCE-based sampling in inverse problems to be inaccurate.

Our analysis tool is the KLD of the posterior (2) and the surrogate posterior (4):

$$D_{KL}(p_N||p) := \int p_N(\theta|d) \ln \frac{p_N(\theta|d)}{p(\theta|d)} d\theta.$$

Since the posterior and the surrogate posterior depend on the data, $D_{KL}(p_N||p)$ also depends on the data. We fix the data, as well as the order N of the truncation (assuming non-zero truncation error), and show that the surrogate posterior is inaccurate if the data are informative, i.e. if the likelihood moves the posterior away from the prior. We consider the small noise regime introduced in [28, 29], where the variance of the prior and the likelihood are small. The small noise regime is important in data assimilation because it corresponds to a case where the posterior probability mass localizes in a "small" region of the parameter space. Moreover, the small noise regime allows for a rigorous analysis and the results can be indicative of more general situations. Specifically, we pick a prior $p_{0,\epsilon}$ with mean zero and variance ϵ and set $\eta \sim \mathcal{N}(0, \sigma^2 \epsilon I)$. These choices give the small noise posterior

$$p_\epsilon = \frac{1}{\gamma_\epsilon(d)} \exp\left(-\frac{1}{\epsilon} F(\theta)\right),$$

where

$$F(\theta) = \frac{||d - h(\theta)||^2}{2\sigma^2} + \frac{||\theta||^2}{2},$$

and where $\gamma_\epsilon(d) = \int \exp\left(-F(\theta)/\epsilon\right) d\theta$ is the normalizing constant; here, and for the remainder of this paper $|| \cdot ||$ is the Euclidean norm. We now state our two main results.

Claim 1. The KLD of the posterior and the prior grows as ϵ becomes smaller and smaller. More precisely, we have

$$D_{KL}(p_{0,\epsilon}||p_\epsilon) = \frac{1}{\epsilon}(F(0) - \min_\theta F(\theta)) + o(\frac{1}{\epsilon}), \tag{5}$$

which grows to infinity as $\epsilon \to 0$ when $F(0) > \min_\theta F(\theta)$; here we use the standard notation $f(\epsilon) = o(1/\epsilon)$ for $\limsup_{\epsilon \to 0} \epsilon f(\epsilon) = 0$.

Claim 2. The KLD of the surrogate posterior to the posterior

$$D_{KL}(p_{N,\epsilon}||p_\epsilon) = \frac{1}{\epsilon}(F(\theta_N^*) - \min_\theta F(\theta)) + o(\frac{1}{\epsilon}), \tag{6}$$

grows to infinity as $\epsilon \to 0$ if $F(\theta_N^*) > \min_\theta F(\theta)$; here, $\theta_N^* = \arg\min_\theta F_N(\theta)$ with

$$F_N(\theta) = \frac{||d - h_N(\theta)||^2}{2\sigma^2} + \frac{||\theta||^2}{2}$$

and we assume that the minimizer θ_N^* is unique.

Derivations of these two results are provided in the appendix. The interpretation of the above results is as follows. Claim 1 shows that, under our assumptions of small noise, the data become more informative as $\epsilon \to 0$, because the data shifts the probability mass away from the prior. Claim 2 thus shows that the surrogate posterior diverges from the posterior as the data become more informative. The two claims combined show that the surrogate posterior may not be useful when the data are informative. We will study the effects of these inaccuracies on the solution of Bayesian inverse problems with numerical examples in the next section.

Our results can also be interpreted geometrically. As ϵ is getting smaller, the posterior is getting more sharply peaked around its mode, since, from (3) and (13) one obtains

$$p_\epsilon(\theta|d) = \exp\left(-\frac{1}{\epsilon}(F(\theta) - \min_\theta F) + o(\frac{1}{\epsilon})\right). \tag{7}$$

Similarly, one can show that the surrogate posterior is also sharply peaked around its maximum. Thus, when the maxima of the posterior and the surrogate posterior are different, the surrogate posterior is (almost) singular with respect to the posterior.

4 Efficiency of the PCE for inverse problems

We have shown that, for a fixed N, the KLD of the surrogate posterior from the posterior can be large, i.e. the PCE is not very accurate. To make it accurate one can increase the order N of the truncated PCE (in fact, for $N \to \infty$, the PCE is exact everywhere). What we must find out is the rate at which N must be increased to ensure sufficient accuracy. If this rate is too rapid, the PCE becomes increasingly expensive. For example, a stochastic collocation routinely requires (at least) $N+1$ quadrature points per parameter to compute the coefficients of the PCE, so that the cost of constructing a PCE of order N is about $(N+1)^m$. If accuracy requires large N, then PCE may no longer be a cost effective approach to the inverse problem.

We can estimate the rate at which N must increase from (6), which indicates that the minimizer θ_N^* of F_N must be close to the minimizer θ^* of F, or else the KLD, and, hence, the errors are large. Thus, point-wise accuracy of the truncated PCE is needed at least up to θ^*, i.e. making $\sup_{x \in B} |h(x) - h_N(x)|$ small for a ball $B \in \mathbb{R}^m$ centered at zero and containing θ^*. The estimate

$$\sup_{x \in \mathbb{R}^m} \exp\left(-\frac{||x||^2}{4}\right) ||h(x) - h_N(x)|| \le CN^{\frac{m}{4} - \frac{k}{2}} ||h||_{k,2}.$$

from [32] indicates that this point-wise accuracy can require that

$$N > C \exp\left(\frac{||\theta^*||^2}{(2k - m)}\right). \tag{8}$$

Recall that h is smooth, so that $2k > m$ (i.e. the exponent is positive). Moreover, since the mean of the prior is zero, a large θ^* (in Euclidean norm) is far from where the prior probability mass is. Thus, large θ^* indicates that the data are informative. Equation (8) thus shows that the order required to maintain accuracy in the PCE grows quickly as the data become increasingly informative.

We now investigate the effects of the inaccuracy of the surrogate posterior on the efficiency of PCE-based sampling for inverse problems, using numerical experiments with an elliptic inverse problem. We choose an elliptic inverse problem because it is a popular tool for testing algorithms for parameter estimation and it is also theoretically well understood [11,17–19,25,27]. The example is not very realistic, however it is sufficient to help us make our points.

4.1 Numerical experiments

Consider the random elliptic equation

$$\begin{cases} -\nabla.(e^{Y(x)}\nabla u) = f(x), & x \in D; \\ u(x) = 0, & x \in \partial D, \end{cases} \tag{9}$$

where $x = (x_1, x_2)$, $D = [0,1] \times [0,1]$, $f(x) = \pi^2 \sin(\pi x_1) \sin(\pi x_2)$, and where Y (often called the log-permeability) is a square integrable stochastic field we wish to estimate from data (e.g. noisy measurements of u at some locations in D). Equation (9) is a simplified model for flow in porous media [11, 27]. We consider a Gaussian log-permeability with mean zero and squared exponential covariance function (see, e.g. [22])

$$R(x_1, x_2, y_1, y_2) = \sigma_1^2 \sigma_2^2 \exp\left(-\frac{(x_1 - y_1)^2}{l_1} - \frac{(x_2 - y_2)^2}{l_2} \right). \tag{10}$$

In our numerical experiments we set $\sigma_1 = \sigma_2 = 1$ and the correlation lengths $l_1 = l_2 = 1$.

We discretize (9) by a standard finite element method on a uniform $(M+1) \times (M+1)$ mesh of triangular 2-D elements [4], with $M = 15$. Solving the PDE is thus equivalent to solving the linear system

$$A(\theta)\hat{u}(\theta) = \hat{f},$$

where $A(\theta)$ is a $M^2 \times M^2$ symmetric positive definite matrix, $\hat{u}(\theta)$ is a M^2-dimensional vector approximating the continuous solution u, and where $\hat{f}$ is a discrete version of f. We evaluate integrals by Gaussian quadrature. We solve these equations with a preconditioned conjugate gradient method. The infinite dimensional random field Y is discretized by a Karhunen–Loève expansion (see, e.g. [12, 15])

$$\hat{Y} = U\theta,$$

where $\theta = (\theta_1, \ldots, \theta_m)$ and where $U = (U_1(x), \ldots, U_m(x))$ is a matrix whose columns are the first m eigenvectors of the squared exponential covariance function, multiplied by their corresponding eigenvalue. Since the squared exponential covariance function has a rapidly decaying spectrum, we can capture 96.66% of the variance with $m = 3$, and this is the choice we make for the following numerical experiments.

This set-up implies a Gaussian prior for θ, with mean zero and variance I. We expand the solution $\hat{u}(\theta)$ of (4.1) in a PCE and approximate it by the truncated PCE, $\hat{u}_N(\theta)$, of order N. We use stochastic collocation with $N+2$ (N even) Gaussian–Hermite quadrature points per parameter to compute the coefficients of the PCE. Thus, $(N+2)^m$ PDE solutions are required to construct the PCE. This approach is only efficient if N or m are small. Here, $m = 3$ is small, and we study the effects of the order N of the truncation, i.e. we will vary N in our numerical experiments.

The example therefore corresponds to a situation in which PCE could be useful, because the number of dimensions is small (it is equal to 3). If we decrease the correlation length, which is perhaps more realistic, we would need to increase m to capture the variance and PCE becomes impractical. However, our goal is to show how inaccuracies in the surrogate posterior can be caused by a PCE which is locally a good approximation, but which lacks global accuracy.

Figure 1 shows our finite element solutions and their PCE approximations of order $N = 4$ and $N = 8$, for two different values of the parameter θ. The first parameter, $\theta = (-2, -1, 1)$, is close to the mean of the prior and we observe that both PCE approximations are accurate (see the top row of Figure 1). As we move the parameter further away from the mean of the prior, e.g. choose $\theta = (-8, -1, 1)$, we observe that the accuracy of the PCE requires that $N \geq 8$ (see the bottom row of Figure 1).

We study the accuracy of the PCE approximation further by focusing attention on the grid point $x = (0.5, 0.5625)$, which is the point on our grid where the first eigenvector of the squared exponential covariance function is maximum. We vary θ_1 between $-8, \ldots, 2$ and fix $\theta_2 = -1$ and $\theta_3 = 1$ as before, i.e. one parameter, θ_1, may be far from its value assumed by the prior, while the other two are within two standard deviations of the mean of the prior. We restrict $\theta_1 \leq 2$
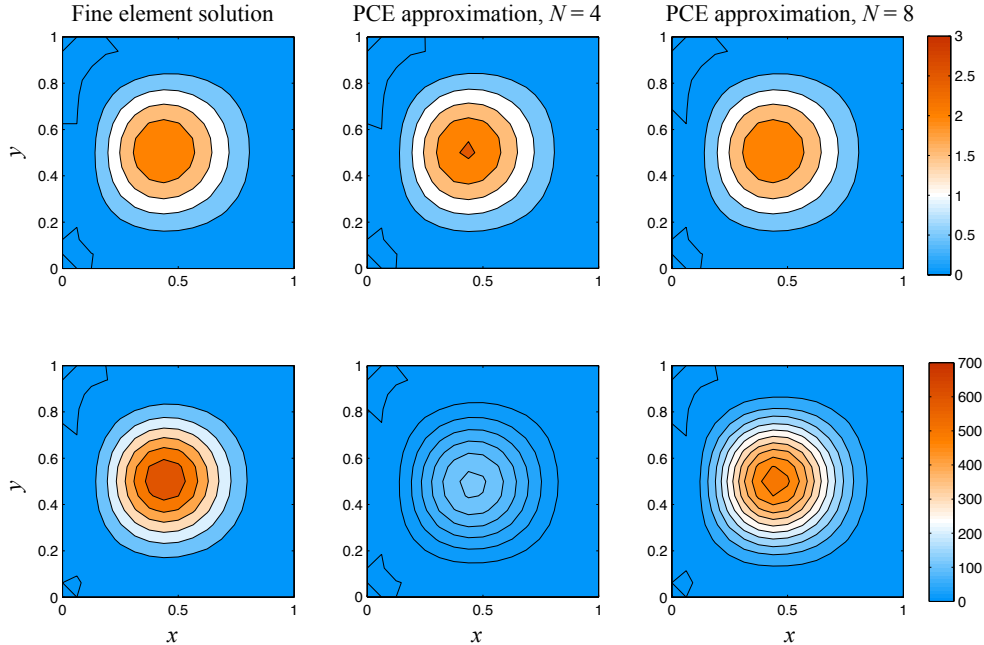
6

Figure 1: The finite element solutions (left column), and their PCE approximations of order $N = 4$ (middle column) and of order $N = 8$ (right column) for two different parameters: $\theta = (-2, -1, 1)$ (top row) and $\theta = (-8, -1, 1)$ (bottom row).

Order	Parameter θ_1										
	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2
$N = 4$	81.1	69.4	53.2	33.4	13.5	0.26	2.56	2.68	1.71	7.72	85.3
$N = 8$	21.2	10.6	3.51	0.39	0.13	0.03	0.02	0.01	0.02	0.04	0.02

Table 1: Relative error in percent as a function of of θ_1 at the grid point $x = (0.5, 0.5625)$.

because the finite element solution is very small otherwise (less than the standard deviation of the noise in the data), which corresponds to a situation in which no inference can be successful since there is almost no information in the signal. The results of our calculations are shown in table 1, where we show the relative error as a function of θ_1 for PCEs with $N = 4$ and $N = 8$. The relative error is defined by the absolute value of the difference of the finite element solution and the PCE approximation, divided by the absolute value of the finite element solution (all evaluated at $x = (0.5, 0.5625)$).

We find that the PCE approximation of order $N = 4$ is accurate (in the sense that the error less than 10%) if θ_1 is within a standard deviation of the mean of the prior (i.e., roughly between -2 and 1). One can extend the region where the PCE is accurate by increasing its order, e.g. a PCE with $N = 8$ is accurate (with errors less than 10%) for $-6 \leq \theta_1 \leq 2$. However, for finite m, the PCE is always inaccurate for large enough θ_1, as is indicated by large errors in Table 1 (global accuracy can only occur if $h(\theta)$ is a low-order polynomial).

With the PCE in place, we solve the inverse problem of estimating the log-permeability θ from noisy measurements of the solution u at several locations in the domain D. The data are "synthetic", i.e. we generate data using our numerical model. This has the advantage that the
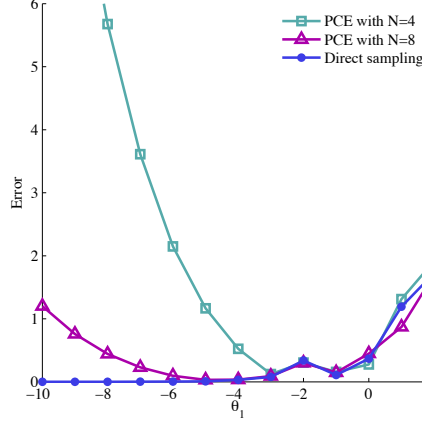
Figure 2: Errors in the solution of the inverse problem: norm of the error as a function of θ_1 for implicit sampling (blue dots), PCE-based implicit sampling with $N = 4$ (turquoise squares), and PCE-based implicit sampling with $N = 8$ (purple triangles).

sampling algorithms operate under ideal conditions (since the model is compatible with the data). We collect data in $n = 16$ locations in the center of the domain, i.e.

$$d = G\hat{u}(\theta) + \eta,$$

where G is an $n \times M^2$ matrix which projects the finite element solution to the n observed components, and where η is an n–dimensional random vector with distribution $\mathcal{N}(0, 0.05^2 I)$. More precisely, data are collected 3 steps away from the top and right boundary and five steps away from the bottom and left boundaries, and two steps away from each other. The prior and likelihood define the posterior

$$p(\theta|d) \propto \exp\left(-F(\theta)\right),$$

where

$$F(\theta) = \frac{1}{2}||\theta||^2 + \frac{1}{2\sigma^2}(d - G\hat{u}(\theta))^T(d - G\hat{u}(\theta)). \tag{11}$$

For PCE-based MC sampling we replace h with its truncated PCE (we consider $N = 4$ and $N = 8$) and compute the surrogate posterior

$$p_N(\theta|d) \propto \exp\left(-F_N(\theta)\right),$$

where

$$F_N(\theta) = \frac{1}{2}||\theta||^2 + \frac{1}{2\sigma^2}(d - G\hat{u}_N(\theta))^T(d - G\hat{u}_N(\theta)).$$

We generate synthetic data sets in which we vary $\theta_1 = -10, -9, \ldots, 2$, while $\theta_2 = -1$, $\theta_3 = 1$ are kept fixed so that the data become increasingly informative as θ_1 increases in magnitude and use implicit sampling [1, 8, 9, 20] for each data set to sample the surrogate posterior. Implicit sampling is an importance sampling method that generates samples close to the mode of the posterior by first locating the mode via numerical optimization, and then solving data dependent algebraic equations to generate samples in the vicinity of the mode. The weights of the samples are then computed from the importance density which is implicitly defined by these algebraic equations. For further details, also about the implementation of implicit sampling, see [1, 8, 9, 20].

Figure 2 summarizes the results of our numerical experiments. We plot the error one makes when solving the inverse problem for three different sampling schemes. The error is defined as the
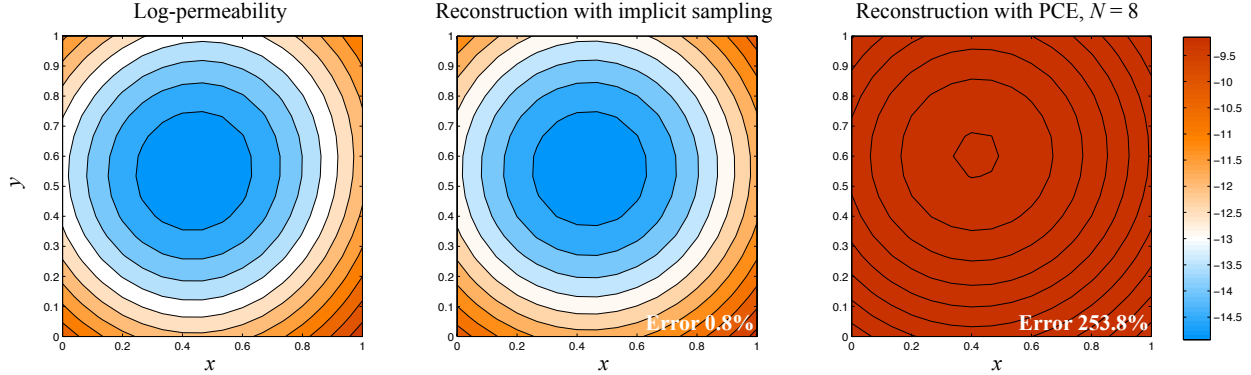
Figure 3: The log-permeability (left), its approximation via implicit sampling (middle) and its approximation via PCE-based MCMC sampling with order $N = 8$.

Euclidean norm of the difference between the parameter θ and its approximation by the conditional mean (the minimum mean square error estimator) obtained via sampling. We use implicit sampling with 20 particles, and vary the accuracy of the PCE from $N = 4$, to $N = 8$, to $N = \infty$, i.e. we sample the posterior directly (without using PCE). The latter is the reference solution.

The figure illustrates that PCE-based sampling with $N = 4$ can only be accurate where the PCE itself is accurate, i.e. only if the parameter is within 2 standard deviations of the mean of the prior. The error increases steeply as the parameter becomes larger (in magnitude). The figure also indicates that the region of applicability of PCE-based sampling can be increased by increasing the order of the PCE, since we obtain much better results with a surrogate posterior when $N = 8$. In this case, the parameter can be far from what is assumed by the prior, however ultimately, one can not guarantee the accuracy of this approach due to the lack of global accuracy of the PCE.

We have obtained the same results with a random walk Metropolis MCMC algorithm (see e.g. [16]), which also shows that the failure of PCE we observe here is independent of the sampling method one uses for sampling the (surrogate) posterior.

In Figure 3 we indicate what the errors in Figure 2 mean for the physics of this inverse problem. In the left panel we plot the log-permeability, in the middle panel its approximation obtained via implicit sampling of the posterior ($N = \infty$), and in the right panel we plot the estimated log-permeability obtained with PCE-based sampling $N = 8$. The parameter $\theta_1 = -15$, i.e. far from what is assumed in the prior, while $\theta_2 = -1$ and $\theta_3 = 1$ are within the range predicted by the prior. This is a scenario in which the data are informative and move the posterior away from the prior. It is evident from this figure that using the surrogate posterior gives catastrophically large errors. However, it is feasible to solve this inverse problem by sampling the posterior (without constructing a surrogate posterior).

Finally, we compare the computational costs of MC sampling with and without a PCE-based surrogate posterior. The cost of solving the inverse problem is dominated by the cost of the required PDE solutions. Constructing a PCE with stochastic collocation with $N + 2$ Gaussian–Hermite quadrature points for each parameter dimension, requires (at least) $(N+2)^m$ PDE solutions. Since $m = 3$ and $N = 4$ or $N = 8$, between 216 and 1000 PDE solutions are required for constructing the PCE. Implicit sampling of the posterior requires 153 evaluations of the posterior, which amounts to 153 PDE solutions (most of them occur during the optimization to find the mode of the posterior, since 20 samples appear sufficient to accurately describe it). Neglecting the cost of evaluation of

9

the surrogate posterior, this is a larger cost than implicit sampling of the posterior, however the results we obtain are less accurate. Since PCE-based sampling is more costly, but less accurate, we conclude that sampling the posterior (without constructing a surrogate posterior) is a better method for this problem.

4.2 Discussion

In the small noise regime, the approximation of the posterior in (7) indicates that significant information resides in the neighborhood of the maximum of the posterior. Hence a successful sampling method should generate samples around this maximum, otherwise information will be lost. Implicit sampling is such a sampling method and therefore is efficient in the above example which corresponds to a small noise scenario. In other situations, other sampling schemes may show a better performance in terms of balancing computational cost and accuracy.

Our simulations also suggest that a successful sampling scheme could result from an adaptive construction of a PCE so that the surrogate posterior is close to the posterior near the maximizer of the posterior. For instance, one could compute a PCE with respect to $\mathcal{N}(\mu, H^{-1})$, where μ is the maximizer of the posterior and H is the Hessian of F in (11) at the maximum. This construction can gain efficiency and accuracy over the prior-based surrogate, since PCEs with low to moderate order may be locally sufficiently accurate. However, such a scheme adds the cost of the optimization to the cost of the PCE construction (neglecting the cost of using the PCE during sampling). It is not clear to us that this strategy will be more efficient than sampling the posterior directly, since e.g. implicit sampling can generate samples close to the mode of the posterior at a low cost once the mode is located [1, 8, 9, 20].

Last, we wish to point out that one can anticipate similar problems and modes of failure with other reduced order modeling techniques for sampling and inverse problems, since the failures we describe are due to the lack of global accuracy, which is common to all reduced order models.

5 Conclusions

We have suggested possible mechanisms of failure of PCE-based sampling in Bayesian inverse problems. In particular, we showed that if the data contain information beyond what is assumed in the prior, then PCE-based sampling can lead to catastrophically large errors or require excessive computational cost. The reason is that PCEs of finite order are not globally accurate (unless the model itself is a low-order polynomial). We presented a rigorous analysis of the failure in the small noise limit, which is characterized by a prior and a likelihood that have "small" variances. We also investigated the efficiency of PCE-based sampling in numerical experiments with an elliptic inverse problem. We observed that a sufficiently accurate PCE for this problem requires a high order, which makes the approach impractical compared to directly sampling the posterior (without constructing a PCE). Moreover, even at a low accuracy, PCE-based sampling was found to be more costly than sampling the posterior without a PCE.

Acknowledgements

Appendix

Derivation of Claim 1

To prove (5), note that

$$
\begin{aligned}
D_{KL}(p_{0,\epsilon}||p_\epsilon) &= \int (2\pi\epsilon)^{-\frac{m}{2}} \exp\left(-\frac{||\theta||^2}{2\epsilon}\right) \ln\left(\frac{\gamma_\epsilon(d)}{(2\pi\epsilon)^{\frac{m}{2}}} \exp\left(\frac{||d - h(\theta)||^2}{2\epsilon\sigma^2}\right)\right) d\theta, \\
&= \ln\frac{\gamma_\epsilon(d)}{(2\pi\epsilon)^{\frac{m}{2}}} + \frac{1}{\epsilon}E\left[\frac{||d - h(X_\epsilon)||^2}{2\sigma^2}\right],
\end{aligned} \tag{12}
$$

where $X_\epsilon \sim \mathcal{N}(0, \epsilon I_m)$. The normalizing constant $\gamma_\epsilon(d)$ can be written as

$$
\gamma_\epsilon(d) = \int \exp\left(-\frac{1}{\epsilon}F(\theta)\right) d\theta = (2\pi\epsilon)^{\frac{m}{2}} E\left[\exp\left(-\frac{1}{\epsilon}\frac{||d - h(X_\epsilon)||^2}{2\sigma^2}\right)\right]. \tag{13}
$$

Laplace's principle (see e.g. [30]) indicates that

$$
\lim_{\epsilon\to 0} \ln E\left[\exp\left(-\frac{1}{\epsilon}\frac{||d - h(X_\epsilon)||^2}{2\sigma^2}\right)\right] = \min_\theta F(\theta),
$$

which can be written as

$$
E\left[\exp\left(-\frac{1}{\epsilon}\frac{||d - h(X_\epsilon)||^2}{2\sigma^2}\right)\right] = \exp\left(-\frac{1}{\epsilon}\min_\theta F(\theta) + o(\frac{1}{\epsilon})\right).
$$

Substituting this equality into (13), we have

$$
\gamma_\epsilon(d) = (2\pi\epsilon)^{\frac{m}{2}} \exp\left(-\frac{1}{\epsilon}\min_\theta F(\theta) + o(\frac{1}{\epsilon})\right). \tag{14}
$$

Since

$$
E\left[\frac{||d - h(X_\epsilon)||^2}{2\sigma^2}\right] \to \frac{||d - h(0)||^2}{2\sigma^2} = F(0)
$$

as $\epsilon \to 0$, we can write the second term in (12) as

$$
\frac{1}{\epsilon}E\left[\frac{||d - h(X_\epsilon)||^2}{2\sigma^2}\right] = \frac{1}{\epsilon}F(0) + o(\frac{1}{\epsilon}).
$$

Then (5) follows by substituting the above equality and (14) into (12).

Derivation of Claim 2

To prove (6), express the surrogate posterior as

$$
p_{N,\epsilon} = \frac{1}{\gamma_{N,\epsilon}(d)} \exp\left(-\frac{1}{\epsilon}F_N(\theta)\right),
$$

where $\gamma_{N,\epsilon}(d) := \int \exp\left(-F_N(\theta)/\epsilon\right) d\theta$. The definition of the KLD then gives

$$
\begin{aligned}
D_{KL}(p_{N,\epsilon}||p_\epsilon) &= \int p_{N,\epsilon}(\theta) \ln\left(\frac{\gamma_\epsilon(d)}{\gamma_{N,\epsilon}(d)} \exp\left(\frac{1}{\epsilon}\left(F(\theta) - F_N(\theta)\right)\right)\right) d\theta, \\
&= \ln\frac{\gamma_\epsilon(d)}{\gamma_{N,\epsilon}(d)} + \frac{1}{\epsilon}\int p_{N,\epsilon}(\theta)\left(F(\theta) - F_N(\theta)\right) d\theta.
\end{aligned} \tag{15}
$$

As before (see (14)), we have that

$$\gamma_{N,\epsilon}(d) = (2\pi\epsilon)^{\frac{m}{2}} \exp\left(-\frac{1}{\epsilon} \min_\theta F_N(\theta) + o(\frac{1}{\epsilon})\right). \qquad (16)$$

Thus,

$$p_{N,\epsilon} = \exp\left(-\frac{1}{\epsilon}(F_N(\theta) - \min F_N) + o(\frac{1}{\epsilon})\right),$$

converges to the delta function $\delta_{\theta_N^*}(\theta)$ as $\epsilon \to 0$. It follows that

$$\lim_{\epsilon \to 0} \int p_{N,\epsilon}(\theta)\,(F(\theta) - F_N(\theta))\,d\theta = F(\theta_N^*) - \min_\theta F_N(\theta),$$

which implies that the second term in (15) can be written as $(F(\theta_N^*) - \min_\theta F_N(\theta))/\epsilon + o(1/\epsilon)$. Then equation (6) follows by substituting (14) and (16) into (15).

References

[1] E. Atkins, M. Morzfeld, and A. J. Chorin. Implicit particle methods and their connection with variational data assimilation. *Mon. Wea. Rev.*, 141(6):1786–1803, 2013.

[2] S. Balakrishnan, A. Roy, M.G. Ierapetritou, G.P. Flach, and P.G. Georgopoulos. Uncertainty reduction and characterization for complex environmental fate and transport models: an empirical Bayesian framework incorporating the stochastic response surface method. *Water Resour. Res.*, 39(12):1350–1362, 2003.

[3] I. Bilionis and N. Zabaras. Solution of inverse problems with limited forward solver evaluations: a Bayesian perspective. *Inverse Problems*, 30(1):015004 (32pp), 2014.

[4] D. Braess. *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, 1997.

[5] M. Branicki and A.J. Majda. Fundamental limitations of polynomial chaos for uncertainty quantification in systems with intermittent instabilities. *Comm. Math. Sci.*, 11(1):55–103, 2013.

[6] A.J. Chorin. Gaussian fields and random flow. *J. Fluid Mech.*, 63:21–32, 1974.

[7] A.J. Chorin and O.H. Hald. *Stochastic Tools in Mathematics and Science*. Springer, 3rd edition, 2013.

[8] A.J. Chorin, M. Morzfeld, and X. Tu. Implicit particle filters for data assimilation. *Commun. Appl. Math. Comput. Sci.*, 5(2):221–240, 2010.

[9] A.J. Chorin and X. Tu. Implicit sampling for particle filters. *Proc. Nat. Acad. Sc. USA*, 106:17249–17254, 2009.

[10] S.C. Crow and G.H. Canavan. Relationship between a Wiener-Hermite expansion and an energy cascade. *J. Fluid Mech.*, 41:387–403, 1970.

[11] M. Dashti and A.M. Stuart. Uncertainty quantification and weak approximation of an elliptic inverse problem. *SIAM J. Numer. Anal.*, 49(6):2524–2542, 2011.

[12] R. Ghanem and P. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Dover Publications, 2003.

[13] T. Hou, W. Luo, B. Rozovskii, and H.M. Zhou. Wiener chaos expansions and numerical solutions of randomly forced equations of fluid mechanics. *J. Comput. Phys.*, 217:687–706, 2006.

[14] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer, 2005.

[15] O.P. LeMaître and O.M. Knio. *Spectral Methods for Uncertainty Quantification: with Applications to Computational Fluid Dynamics*. Springer, 2010.

[16] J. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, 2001.

[17] Y.M. Marzouk and H.N. Najm. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *J. Comput. Phys.*, 228(6):1862–1902, 2009.

[18] Y.M. Marzouk, H.N. Najm, and L.A. Rahn. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *J. Comput. Phys.*, 224(2):560–586, 2007.

[19] Y.M. Marzouk and D.B. Xiu. A stochastic collocation approach to Bayesian inference in inverse problems. *Commun. Comput. Phys.*, 6:826–847, 2009.

[20] M. Morzfeld, X. Tu, E. Atkins, and A.J. Chorin. A random map implementation of implicit filters. *J. Comput. Phys.*, 231(4):2049–2066, 2012.

[21] H.N. Najm. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Ann. Rev. Fluid Mech.*, 41:35–52, 2009.

[22] C.E. Rasmussen and C.K.I. Wiliams. *Gaussian processes for machine learning*. MIT Press, 2006.

[23] F. Rizzi, H.N. Najm, B.J. Debusschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, and O.M. Knio. Uncertainty quantification in MD simulations. Part II: Bayesian inference of force-field parameters. *Multiscale Model. Simul.*, 10(4):1460–1492, 2012.

[24] W. Schoutens. *Stochastic Processes and Orthogonal Polynomials*. Springer, 2000.

[25] A.M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.

[26] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.

[27] X. Tu, M. Morzfeld, J. Wilkening, and A. J. Chorin. Implicit sampling for an elliptic inverse problem in underground hydrodynamics. arXiv:1308.4640, 2013.

[28] E. Vanden-Eijnden and J. Weare. Rare event simulation and small noise diffusions. *Commun. Pure Appl. Math.*, 65:1770– 1803, 2012.

[29] E. Vanden-Eijnden and J. Weare. Data assimilation in the low noise regime with application to the Kuroshio. *Mon. Wea. Rev.*, 141:1822–1841, 2013.

[30] S. R. S. Varadhan. *Large Deviations and Applications*. CBMS-NSF Regional Conference Series in Applied Mathematics, 46. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1984.

[31] D.B. Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach.* Princeton University Press, 2010.

[32] C. Xu and B. Guo. Hermite spectral and pseudo-spectral methods for nonlinear partial differential equations in multiple dimensions. *Comput. Appl. Math.*, 22(2):167–193, 2003.