13th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP13
Seoul, South Korea, May 26-30, 2019

Sequential piecewise PCE approximation of likelihood functions in Bayesian inference

Paul-Remo Wagner
PhD Student, Dept. of Civil Engineering, ETH Zurich, Zurich, Switzerland

Stefano Marelli
Senior Scientist, Dept. of Civil Engineering, ETH Zurich, Zurich, Switzerland

Christos Lataniotis
PhD Student, Dept. of Civil Engineering, ETH Zurich, Zurich, Switzerland

Bruno Sudret
Professor, Dept. of Civil Engineering, ETH Zurich, Zurich, Switzerland

ABSTRACT: A method for sequentially constructing polynomial chaos expansions, with the aim of approximating likelihood functions that occur in Bayesian inference problems is presented. The proposed approach is called piecewise polynomial chaos expansion (P-PCE) and is based on sequentially constructing PCEs in refined domains on the residuals of previously constructed PCEs. The obtained local spectral representation allows the computation of posterior expectations by post-processing the PCE coefficients based on the recently developed concept of spectral likelihood expansion (SLE). This paper presents a summary of the proposed theory and showcases the solution of two Bayesian inference problems using the presented approach.

1. INTRODUCTION

In the context of *inverse problems*, the goal is to determine which set of input parameters of a given computational model has produced an observed set of experimental data. One way of addressing such problems is the Bayesian inference framework. It requires a computational model $\tilde{Y} = \mathscr{M}(\boldsymbol{X})$ with uncertain input parameters $\boldsymbol{X} = (X_1, \ldots, X_M)$, $X_i \in \mathscr{D}_{x_i}$, that is to be related with a set of observations $\boldsymbol{y} = (y^{(1)}, \ldots, y^{(N)})$. The link between $\tilde{Y}$ and $\boldsymbol{y}$ is made through a so-called *likelihood function* $\mathscr{L}(\boldsymbol{x}|\boldsymbol{y})$. Let $\pi(\boldsymbol{x})$ be the *prior* distribution of the parameter vector $\boldsymbol{X}$, the Bayesian inference problem is then posed by the following equation:

$$\pi(\boldsymbol{x}|\boldsymbol{y}) = \frac{\mathscr{L}(\boldsymbol{x}|\boldsymbol{y})\pi(\boldsymbol{x})}{\pi(\boldsymbol{y})}, \tag{1}$$

where $\pi(\boldsymbol{x}|\boldsymbol{y})$ corresponds to the *posterior distribution* of $\boldsymbol{X}$ conditioned on $\boldsymbol{y}$. The denominator $\pi(\boldsymbol{y})$ is called *evidence* or *marginal likelihood* and is a mere normalizing constant defined by

$$\pi(\boldsymbol{y}) = \int_{\mathscr{D}_{\boldsymbol{x}}} \mathscr{L}(\boldsymbol{x}|\boldsymbol{y})\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \tag{2}$$

where $\mathscr{D}_{\boldsymbol{x}} = \mathscr{D}_{x_1} \times \ldots \times \mathscr{D}_{x_M}$. The posterior probability density function (PDF), or more frequently its statistics, constitute the solution of the inverse problem. The posterior distribution is often used in subsequent steps to compute expectation values of so-called *quantities of interest* $h(\boldsymbol{X})$, where $\boldsymbol{X} \sim \pi(\boldsymbol{x}|\boldsymbol{y})$. This corresponds to evaluating the following integral:

$$\mathbb{E}\left[h(\boldsymbol{X})|\boldsymbol{y}\right] = \int_{\mathscr{D}_{\boldsymbol{x}}} h(\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}. \tag{3}$$

In real-world scenarios, it is often neither possible to find a closed form expression for the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ nor to directly compute $\mathbb{E}\left[h(\boldsymbol{X})|\boldsymbol{y}\right]$, mainly due to two obstacles:

- the computation of $\pi(\boldsymbol{y})$ requires the evaluation of the possibly high dimensional integral in Eq. (2)
- the repeated evaluation of the function $\mathscr{L}(\boldsymbol{x}|\boldsymbol{y})$ involves the calculation of the forward model $\mathscr{M}(\boldsymbol{x})$, which is often computationally expensive.

The most common methods to compute the posterior and its characteristics are based on Monte-Carlo sampling, more specifically the popular family of Markov chain Monte Carlo (MCMC) algorithms (Robert and Casella, 2004). The first algorithm of this family was the classical Metropolis algorithm from Metropolis et al. (1953) which has been significantly enhanced since then (Hastings, 1970; Haario et al., 2001; Goodman and Weare, 2010). The strength of these algorithms is that they are relatively immune to the curse of dimensionality and avoid evaluating the evidence $\pi(\boldsymbol{y})$. Their weakness, however, is that they require a large amount of potentially expensive forward model runs that may hinder their usage in real-world applications.

Recent attempts to apply MCMC algorithms together with computationally expensive forward models have focused primarily on replacing the forward model with a computationally cheaper surrogate $\tilde{\mathscr{M}}(\boldsymbol{X}) \approx \mathscr{M}(\boldsymbol{X})$ (Marzouk and Xiu, 2009; Higdon et al., 2015). Typically, the construction of such a surrogate relies on the prior parameter distribution $\boldsymbol{X} \sim \pi(\boldsymbol{x})$, which guarantees the surrogate's accuracy over the whole prior domain. This is, however, not always efficient nor accurate as the support of the posterior distribution is typically a small subset of the prior's. This has motivated recent approaches that adaptively refine the surrogate model in regions of high posterior density (see *e.g.* Li and Marzouk (2014)).

A new approach for Bayesian inference called spectral likelihood expansion (SLE) was introduced in Nagel and Sudret (2016). There, the likelihood function is represented in a polynomial basis orthogonal to the prior distribution. This spectral representation allows the computation of posterior statistics by mere post-processing of the basis coefficients. While this approach offers appealing analytical results, it is not practical in real-world applications as the polynomial degree required to accurately represent the likelihood function is typically prohibitive.

In this paper the shortcomings of SLE are alleviated by representing the likelihood function as a sum of so-called *partial likelihood functions* $\tilde{\mathscr{L}}^k$ and a residual $\mathscr{R}$:

$$\mathscr{L}(\boldsymbol{x}|\boldsymbol{y}) = \sum_{k=1}^{K} \tilde{\mathscr{L}}^k(\boldsymbol{x}|\boldsymbol{y}) + \mathscr{R}^{K+1}(\boldsymbol{x}). \tag{4}$$

The partial likelihood functions have bounded support $\mathscr{D}_{\boldsymbol{x}}^k \subseteq \mathscr{D}_{\boldsymbol{x}}$ and can be thought of as patches that make up the total likelihood approximation. They are sequentially computed as approximations of the residual:

$$\tilde{\mathscr{L}}^i(\boldsymbol{x}|\boldsymbol{y}) \approx \mathscr{R}^i(\boldsymbol{x}) = \mathscr{L}(\boldsymbol{x}|\boldsymbol{y}) - \sum_{k=1}^{i-1} \tilde{\mathscr{L}}^k(\boldsymbol{x}|\boldsymbol{y}). \tag{5}$$

where, in the initial step, the residual equals the original likelihood function $\mathscr{R}^1(\boldsymbol{x}) = \mathscr{L}(\boldsymbol{x}|\boldsymbol{y})$ on the whole domain $\mathscr{D}_{\boldsymbol{x}}^1 = \mathscr{D}_{\boldsymbol{x}}$.

In the presented piecewise-PCE (P-PCE) approach, the approximation from Eq. (5) is carried out with PCEs, orthogonal to the prior distribution $\pi(\boldsymbol{x})$ in subdomains $\mathscr{D}_{\boldsymbol{x}}^i$. This preserves the appealing analytical properties of SLE. This approach can also be understood as *recursive partitioning regression* (Friedman, 1991) with locally orthogonal basis functions.

After a brief review of PCE in the next section, Section 3 presents the theoretical background of P-PCE. In Section 4 an algorithm is introduced to construct P-PCEs which is used to solve a test problem in Section 5.

2. Polynomial chaos expansions

Polynomial chaos expansion (PCE) is a surrogate modelling technique that has received widespread attention (Xiu and Karniadakis, 2002; Soize and Ghanem, 2004). A brief introduction to the method is presented next, but providing only the bare essentials required for the proposed approach presented in Section 3.

Assume a random vector $\boldsymbol{X} = (X_1, \ldots, X_M)$ with mutually independent components $X_i \sim \pi_i(x_i)$. Its

joint PDF is then given by:

$$\pi(\boldsymbol{x}) = \prod_{i=1}^{M} \pi_i(x_i). \tag{6}$$

Any scalar function $f(\boldsymbol{X}) : \mathbb{R}^M \mapsto \mathbb{R}$ with finite variance can then be approximated by the truncated polynomial chaos expansion:

$$f(\boldsymbol{X}) \approx \tilde{f}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha} \in \mathscr{A}} a_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X}), \tag{7}$$

where $\Psi_{\boldsymbol{\alpha}} \overset{\text{def}}{=} \prod_{i=1}^{M} \psi^i_{\alpha_i}(x_i)$ are polynomials orthogonal with respect to $\pi(\boldsymbol{x})$, $\boldsymbol{\alpha} = (\alpha_i, \ldots, \alpha_M) \in \mathscr{A} \subset \mathbb{N}^M$ is an M-tuple specifying the degree of the polynomial basis functions and $a_{\boldsymbol{\alpha}}$ are the corresponding coefficients.

The univariate polynomials $\psi^i_{\alpha_i}$ are chosen to satisfy:

$$\left\langle \psi^i_k, \psi^i_l \right\rangle_\pi \overset{\text{def}}{=} \int_{\mathscr{D}_{x_i}} \psi^i_k(x_i)\psi^i_l(x_i)\pi_i(x_i)\,\mathrm{d}x_i = \delta_{kl}, \tag{8}$$

where $\delta_{kl} = 1$ if $k = l$ and 0 otherwise. It follows that the multivariate polynomials $\Psi_{\boldsymbol{\alpha}}$ are orthonormal w.r.t. the joint distribution $\pi(\boldsymbol{x})$.

There exist well-known families of polynomial functions that fulfil the fundamental condition of Eq. (8) w.r.t. standard parametric distributions (Askey and Wilson, 1985). However, analytical expressions for families of polynomials that are orthonormal w.r.t. $\pi(\boldsymbol{x})$ on bounded supports $\mathscr{D}^k_{\boldsymbol{x}} \subseteq \mathscr{D}_{\boldsymbol{x}}$, do not exist. In this case, a general recurrence scheme that allows the construction of polynomials $\{\psi^{i,k}_n, n \in \mathbb{N}\}$ orthonormal with respect to arbitrary weight functions $\pi_i(x_i)$ on $\mathscr{D}^k_{x_i}$ has to be used (see Appendix A).

Because $\pi_i(x_i)$ on $\mathscr{D}^k_{x_i} \subseteq \mathscr{D}_{x_i}$ does not necessarily integrate to one, the constant polynomials $\psi^{i,k}_0 \neq 1$ (see Figure 1).

The PCE coefficients $a_{\boldsymbol{\alpha}}$ are then computed by least-square analysis, as detailed in Berveiller et al. (2006); Blatman and Sudret (2011).

To judge the quality of the PCE approximation, it is common practice to use the leave-one-out error as an approximation of the surrogate's generalisation error (Hastie et al., 2001; Arlot and Celisse, 2010).

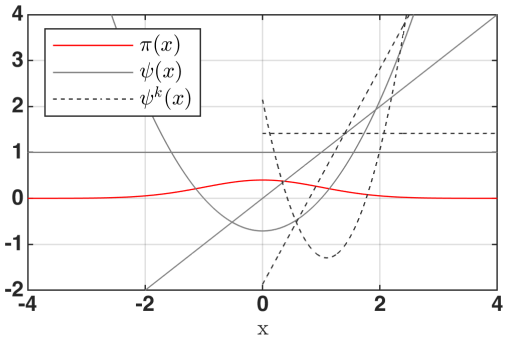


Figure 1: Marginal PDF $\pi(x)$ as a standard normal distribution and associated orthonormal polynomials with degrees $p = 0$ to $p = 2$ on the domains $\mathscr{D}_x = [-\infty, \infty]$ and $\mathscr{D}^k_x = [0, 4]$. Note: the polynomials $\psi(x)$ are the classical Hermite polynomials.

Given an experimental design $\mathfrak{X} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)})$, it is defined by:

$$\varepsilon_{\text{LOO}} = \frac{1}{N} \sum_{i=1}^{N} \left(f(\boldsymbol{x}^{(i)}) - \tilde{f}_{\setminus i}(\boldsymbol{x}^{(i)}) \right)^2, \tag{9}$$

where $\tilde{f}_{\setminus i}$ is the surrogate of f constructed without using the i-th component of the experimental design. It can be estimated efficiently for PCEs as detailed in Blatman and Sudret (2011).

3. Piecewise PCE for likelihood approximation

In this section it is assumed, that a P-PCE has been constructed using the algorithm presented in Section 5. This way, the likelihood function in Bayes' theorem (see Eq. (1)) is expressed as a sum of partial likelihood functions (see Eq. (4)) with negligible residual $\mathscr{R}$. Each $\tilde{\mathscr{L}}^k(\boldsymbol{x}|\boldsymbol{y})$ is given by a PCE as:

$$\tilde{\mathscr{L}}^k(\boldsymbol{x}|\boldsymbol{y}) = \sum_{\boldsymbol{\alpha} \in \mathscr{A}^k} b^k_{\boldsymbol{\alpha}} \Psi^k_{\boldsymbol{\alpha}}(\boldsymbol{x}) \mathbf{1}_{\mathscr{D}^k_{\boldsymbol{x}}}(\boldsymbol{x}), \tag{10}$$

where $\mathscr{A}^k$ is the index set for the k-th PCE. The indicator function $\mathbf{1}_{\mathscr{D}^k_{\boldsymbol{x}}}$ is given by

$$\mathbf{1}_{\mathscr{D}^k_{\boldsymbol{x}}}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{x} \in \mathscr{D}^k_{\boldsymbol{x}}, \\ 0 & \text{if } \boldsymbol{x} \notin \mathscr{D}^k_{\boldsymbol{x}}, \end{cases} \tag{11}$$

and ensures that the PCEs have bounded support in $\mathscr{D}^k_{\boldsymbol{x}}$.

Equation (10) constitutes a local spectral representation of the partial likelihood function. This representation can be used to derive analytical expressions for posterior quantities of interest.

From Eq. (2) and (10) it follows that the evidence $\pi(\boldsymbol{y})$ can be expressed as:

$$\pi(\boldsymbol{y}) = \sum_{k=1}^{K} \sum_{\boldsymbol{\alpha}\in\mathscr{A}^k} b_{\boldsymbol{\alpha}}^k \int_{\mathscr{D}_{\boldsymbol{x}}^k} \Psi_{\boldsymbol{\alpha}}^k(\boldsymbol{x})\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \tag{12}$$

$$= \sum_{k=1}^{K} \frac{b_{\boldsymbol{0}}^k}{\Psi_{\boldsymbol{0}}^k}. \tag{13}$$

That is, $\pi(\boldsymbol{y})$ can be obtained by simple post-processing of the PCEs. Using this result, the expectation of general quantities of interest $h(\boldsymbol{X})$, where $\boldsymbol{X} \sim \pi(\boldsymbol{x}|\boldsymbol{y})$, can also be expressed as:

$$\mathbb{E}\left[h(\boldsymbol{X})|\boldsymbol{y}\right] = \int_{\mathscr{D}_{\boldsymbol{x}}} h(\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x} \tag{14}$$

$$= \frac{1}{\pi(\boldsymbol{y})} \sum_{k=1}^{K} \sum_{\boldsymbol{\alpha}\in\mathscr{A}^k} b_{\boldsymbol{\alpha}}^k \int_{\mathscr{D}_{\boldsymbol{x}}^k} h(\boldsymbol{x})\Psi_{\boldsymbol{\alpha}}^k(\boldsymbol{x})\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \tag{15}$$

By projecting the function $h(\boldsymbol{x})$ on all K bases individually, *i.e.* $h(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}\in\mathscr{A}^k} c_{\boldsymbol{\alpha}}^k \Psi_{\boldsymbol{\alpha}}^k$ for $k = 1,\ldots,K$, Eq. (14) can be further simplified into

$$\mathbb{E}\left[h(\boldsymbol{X})|\boldsymbol{y}\right] = \frac{1}{\pi(\boldsymbol{y})} \sum_{k=1}^{K} \left\langle \sum_{\boldsymbol{\alpha}\in\mathscr{A}^k} c_{\boldsymbol{\alpha}}^k \Psi_{\boldsymbol{\alpha}}^k, \sum_{\boldsymbol{\alpha}\in\mathscr{A}^k} b_{\boldsymbol{\alpha}}^k \Psi_{\boldsymbol{\alpha}}^k \right\rangle_{\pi,k} \tag{16}$$

$$= \frac{1}{\pi(\boldsymbol{y})} \sum_{k=1}^{K} \sum_{\boldsymbol{\alpha}\in\mathscr{A}^k} c_{\boldsymbol{\alpha}}^k b_{\boldsymbol{\alpha}}^k. \tag{17}$$

This result can be used to compute general expectations according to Eq. (3) and simple posterior moments. To calculate the first moment of the i-th posterior marginal distribution $\pi(x_i|\boldsymbol{y})$, one chooses $h(x_i) = x_i$ in Eq. (14), while the posterior covariance between the i-th and j-th variable can be computed by setting $h(x_i,x_j) = (x_i - \mathbb{E}\left[X_i\right])(x_j - \mathbb{E}\left[X_j\right])$.

The P-PCE representation of the likelihood function also makes it easy to find analytical expressions for the marginal posterior distributions. After gathering the parameters that are not to be marginalized in $\boldsymbol{x}_*$ and the rest in $\boldsymbol{x}_{\sim *}$, the marginal posterior distribution reads:

$$\pi(\boldsymbol{x}_*|\boldsymbol{y}) = \int_{\mathscr{D}_{\boldsymbol{x}_{\sim *}}} \pi(\boldsymbol{x}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}_{\sim *} \tag{18}$$

$$= \frac{1}{\pi(\boldsymbol{y})} \sum_{k=1}^{K} \sum_{\boldsymbol{\alpha}\in\mathscr{A}^k} b_{\boldsymbol{\alpha}}^k \int_{\mathscr{D}_{\boldsymbol{x}_{\sim *}}^k} \Psi_{\boldsymbol{\alpha}}^k(\boldsymbol{x})\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}_{\sim *}. \tag{19}$$

Because of the orthogonality in the local bases, the integral expression on the right-hand side of Eq. (18) equals zero for all $\boldsymbol{\alpha}$ that are non-constant in at least one marginalizing parameter gathered in $\boldsymbol{x}_{\sim *}$. For all $\boldsymbol{\alpha}$ that are constant in all marginalizing parameters, the integral expression evaluates to one over the product of the marginalized constant polynomials ($1/\Psi_{\boldsymbol{0}_{\sim *}}^k$).

New sets of multi-indices $\mathscr{A}_*^k \subset \mathscr{A}^k$ are then defined for each PCE that contains only terms that are constant in all marginalizing parameters $\boldsymbol{x}_{\sim *}$. Using these sets, the marginal posterior distribution in Eq. (18) can be written as

$$\pi(\boldsymbol{x}_*|\boldsymbol{y}) = \frac{1}{\pi(\boldsymbol{y})} \sum_{k=1}^{K} \sum_{\boldsymbol{\alpha}\in\mathscr{A}_*^k} \frac{b_{\boldsymbol{\alpha}}^k}{\Psi_{\boldsymbol{0}_{\sim *}}^k} \Psi_{\boldsymbol{\alpha}}^k(\boldsymbol{x}_*)\pi(\boldsymbol{x}_*), \tag{20}$$

where $\Psi_{\boldsymbol{\alpha}}^k(\boldsymbol{x}_*)$ is the product of the univariate polynomials in the non marginalized dimensions with corresponding prior $\pi(\boldsymbol{x}_*)$. These equations reduce to the ones given in Nagel and Sudret (2016) for the case of $K = 1$ and $\mathscr{D}_{\boldsymbol{x}}^1 = \mathscr{D}_{\boldsymbol{x}}$.

4. Computing P-PCE

An algorithm to construct a P-PCE is outlined next. It consists of sequentially refining PCEs with increasingly smaller support and summing up their contributions to approximate the total likelihood function $\mathscr{L}(\boldsymbol{x}|\boldsymbol{y})$. The procedure can be visualized by the graph shown in Figure 2. The algorithm at each refinement level r constructs N_r PCEs using the residuals of the previous refinement level. In the r-th refinement level there are $N_r = 2^{Mr}$ subdomains $\mathscr{D}_{\boldsymbol{x}}^{r,i}$. In this section $\tilde{\mathscr{L}}^{r,i}$ and $\mathscr{R}^{r,i}$ is used to refer to the i-th partial likelihood and residual of the r-th refinement level respectively.

Given: The P-PCE algorithm begins with a given prior distribution $\pi(\boldsymbol{x})$, a likelihood function

$\mathscr{L}(\boldsymbol{x}|\boldsymbol{y})$ and an integer $N_{\text{enr.}}$, specifying the number of new samples added in each enrichment step. The initial refinement level is $r=0$ and draw $N_{\text{enr.}}$ samples $\mathfrak{X}=(\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(N_{\text{enr.}})})$ are drawn from a uniform distribution $\tilde{\pi}(\boldsymbol{x})$ that has the same support as $\pi(\boldsymbol{x})$. In the case of prior distributions with unbounded support (*e.g.* normal distribution) the initial domain $\mathscr{D}_{\boldsymbol{x}}^{0,1}$ ($r=0$ and $i=N_r=1$) is restricted to the ε-quantile and the $(1-\varepsilon)$-quantile in each marginal distribution, for a small number ε. These samples are then used to evaluate the likelihood function and the results are stored in $\mathfrak{L}=\mathscr{L}(\mathfrak{X}|\boldsymbol{y})$. An orthonormal basis $\Psi_{\alpha}^{0,1}$ with respect to $\pi(\boldsymbol{x})$ is computed on the domain $\mathscr{D}_{\boldsymbol{x}}^{0,1}$. Then $\mathfrak{X}$ and $\mathfrak{L}$ can be used to approximate the polynomial coefficients:

$$b_{\boldsymbol{\alpha}}^{0,1}=\left\langle \Psi_{\boldsymbol{\alpha}}^{0,1},\mathscr{L}(\boldsymbol{x}|\boldsymbol{y})\right\rangle_{\pi,0,1}. \tag{21}$$

In the present implementation the coefficients are estimated by least squares regression (Berveiller et al., 2006).

$\tilde{\mathscr{L}}^{0,1}(\boldsymbol{x}|\boldsymbol{y})=\sum_{\boldsymbol{\alpha}\in\mathscr{A}^{0,1}} b_{\boldsymbol{\alpha}}^{0,1}\Psi_{\alpha}^{0,1}(\boldsymbol{x})$ is obtained as an initial approximation of the likelihood function.

Until converged: Increase refinement level $r=r+1$

1. **Subdivide domain:** Subdivide each domain $\mathscr{D}_{\boldsymbol{x}}^{r-1,i}$ of the previous refinement level $r-1$ into 2^M equally sized subdomains. This defines the subdomains $\mathscr{D}_{\boldsymbol{x}}^{r,i}$ for $i=1,\ldots,N_r$.

2. **Loop over subdomains** $\mathscr{D}_{\boldsymbol{x}}^{r,i}$: For $i=1,\ldots,N_r$

   (a) **Enrich experimental design:** Add $N_{\text{enr.}}$ new samples to the experimental design $\mathfrak{X}$ by sampling from the uniform distribution $\tilde{\pi}^{r,i}(x)$ defined on the subdomain $\mathscr{D}_{\boldsymbol{x}}^{r,i}$. Evaluate the likelihood at these new samples and update the residuals vector $\mathfrak{R}$ by subtracting all lower level PCEs from the likelihood evaluations:

   $$\mathfrak{R}=\mathscr{L}(\mathfrak{X}|\boldsymbol{y})-\sum_{s=0}^{r-1}\sum_{j=1}^{N_r}\tilde{\mathscr{L}}^{s,j}(\mathfrak{X}|\boldsymbol{y}) \tag{22}$$

   (b) **PCE:** Construct orthonormal basis $\{\Psi_{\boldsymbol{\alpha}}^{r,i}:\alpha\in\mathscr{A}^{r,i}\}$ w.r.t the truncated prior in subdomain $\mathscr{D}_{\boldsymbol{x}}^{r,i}$ and compute the coefficients using the residuals $\mathfrak{R}$:

   $$b_{\boldsymbol{\alpha}}^{r,i}=\langle\Psi_{\boldsymbol{\alpha}}^{r,i},\mathscr{R}^{r,i}(\boldsymbol{x})\rangle_{\pi,r,i}. \tag{23}$$

   $\tilde{\mathscr{L}}^{r,i}(\boldsymbol{x}|\boldsymbol{y})=\sum_{\boldsymbol{\alpha}\in\mathscr{A}^{r,i}} b_{\boldsymbol{\alpha}}^{r,i}\Psi_{\boldsymbol{\alpha}}^{r,i}(\boldsymbol{x})$ is obtained as an approximation to $\mathscr{R}^{r,i}(\boldsymbol{x})$.

   (c) **Compute residuals and error:** Update the residual $\mathfrak{R}=\mathfrak{R}-\mathscr{L}^{r,i}(\mathfrak{X}|\boldsymbol{y})$. Estimate the leave-one-out error of the current PCE using $\mathfrak{R}=(\boldsymbol{r}^{(1)},\ldots,\boldsymbol{r}^{(N)})^{\intercal}$ and $\mathfrak{X}=(\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(N)})^{\intercal}$:

   $$\varepsilon_{\text{LOO}}^{r,i}=\frac{1}{N_{r,i}}\sum_{j=1}^{N}\left(\boldsymbol{r}^{(j)}-\tilde{\mathscr{L}}_{\setminus j}^{r,i}(\boldsymbol{x}^{(j)}|\boldsymbol{y})\right)^2, \tag{24}$$

   where $N_{r,i}$ is the number of samples in $\mathscr{D}_{\boldsymbol{x}}^{r,i}$.

3. **Check convergence:** Average the leave one out errors from all subdomains of level r to obtain an estimate of the surrogate model accuracy at the current refinement level:

   $$\varepsilon_{\text{LOO}}^{r}=\frac{1}{N_r}\sum_{i=1}^{N_r}\varepsilon_{\text{LOO}}^{r,i}. \tag{25}$$

   Terminate the algorithm if a sufficient accuracy is reached.

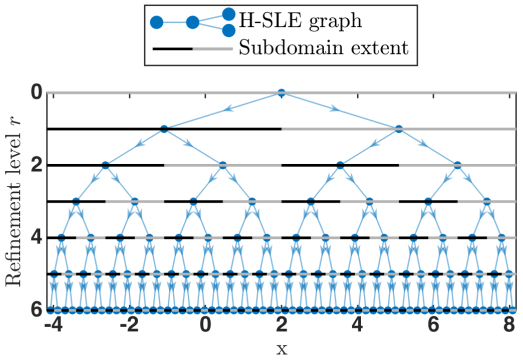


Figure 2: Visualization of the subdomains $\mathscr{D}_{\boldsymbol{x}}^{k}$ in a 1D example created during the P-PCE procedure. Each node represents one PCE of a partial likelihood.

5. RESULTS

Two case studies are presented that demonstrate the concepts of P-PCE. In both cases an inference problem is solved, where both the prior $\pi(\boldsymbol{x})$ and the likelihood function $\mathscr{L}(\boldsymbol{x}|\boldsymbol{y})$ are given by a normal distribution:

$$\pi(\boldsymbol{x}) = \mathscr{N}(\boldsymbol{x}|\boldsymbol{\mu}_{\text{prior}}, \boldsymbol{\Sigma}_{\text{prior}}), \tag{26}$$

$$\mathscr{L}(\boldsymbol{x}|\boldsymbol{y}) = \prod_{i=1}^{N} \mathscr{N}(\boldsymbol{y}^{(i)}|\boldsymbol{x}, \boldsymbol{\Sigma}). \tag{27}$$

The prior parameters $\boldsymbol{\mu}_{\text{prior}}$, $\boldsymbol{\Sigma}_{\text{prior}}$ and the measurement noise $\boldsymbol{\Sigma}$ of the likelihood function are known. Then, a set of N synthetic data measurements $\boldsymbol{y}^{(i)}$ are generated to complete the inference problem. The posterior distribution in this case can be calculated analytically by:

$$\begin{aligned} \pi(\boldsymbol{x}|\boldsymbol{y}) &= \mathscr{N}(\boldsymbol{x}|\boldsymbol{\mu}^{\text{post}}, \boldsymbol{\Sigma}_{\text{post}}) \quad \text{with} \\ \boldsymbol{\Sigma}_{\text{post}} &= \left(\boldsymbol{\Sigma}_{\text{prior}}^{-1} + N\boldsymbol{\Sigma}^{-1}\right)^{-1} \\ \boldsymbol{\mu}_{\text{post}} &= \boldsymbol{\Sigma}_{\text{post}} \left(\boldsymbol{\Sigma}_{\text{prior}}^{-1}\boldsymbol{\mu}_{\text{prior}} + N\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{y}}\right), \end{aligned} \tag{28}$$

where $\bar{\boldsymbol{y}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{y}^{(i)}$ is the data sample mean. The analytically available posterior distribution makes it possible to accurately judge the convergence of the proposed P-PCE method.

In the presented case studies, the maximum polynomial degree at each refinement level is set to $p = 3$ and a standard truncation scheme $\mathscr{A}^k = \left\{\boldsymbol{\alpha} \in \mathbb{N}^M : ||\boldsymbol{\alpha}||_1 \leq p\right\}$ is used, with $||\boldsymbol{\alpha}||_1 = \sum_{i=1}^{M} \alpha_i$ for all K PCEs, where all basis functions up to a total degree p are included. The experimental design is enriched with $N_{\text{enr.}} = 2 \cdot \text{card}(\mathscr{A}^k)$ in each domain. This is in accordance with the recommendations given in Berveiller et al. (2006).

At every refinement step, the experimental design is thus enriched with $N_{\text{enr.}}$ samples in every one of the 2^M new subdomains, that is, N grows according to:

$$N(r) = N_{\text{enr.}} \cdot \sum_{k=0}^{r} \left(2^M\right)^k. \tag{29}$$

5.1. 1D conjugate problem

The first problem aims to show how the P-PCE procedure approximates a given likelihood function. A 1D example with $\mu_{\text{prior}} = 2$ and $\sigma^2_{\text{prior}} = 4$ is used with 10 synthetic measurements $\boldsymbol{y}$ sampled from $\mathscr{N}(1, \sigma^2)$ with $\sigma^2 = 1$.

Following the procedure described in Section 4, $N_{enr.} = 2 \cdot \text{card}(\mathscr{A}) = 8$ samples are drawn in the initial domain $\mathscr{D}_{\boldsymbol{x}} = [-4.18, 8.18]$ that corresponds to the 10^{-3} and $1 - 10^{-3}$ quantiles of the prior distribution.

The convergence behaviour of the P-PCE procedure is shown in Figure 3 and the estimated posterior marginals can be seen in Figure 4. After the 5-th refinement step, the relative error of both moments falls below 0.1%, at a total cost of $N = 256$ model evaluations.

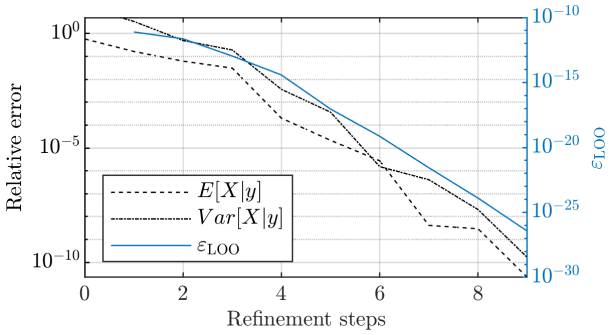


Figure 3: 1D example – Relative error convergence of the mean and variance of $X|\boldsymbol{y}$ and the behaviour of the average leave-one-out error defined in Eq. (25).

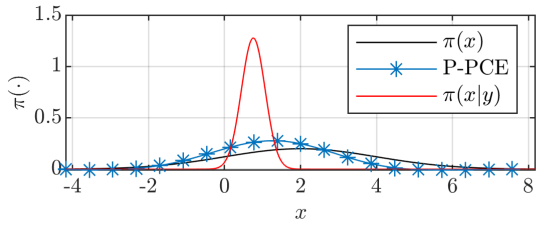
5.2. 2D conjugate problem

In the second problem the shortcomings of the current implementation of P-PCE are presented by setting the dimensionality of the problem to $M = 2$ and investigating an example with $\boldsymbol{\mu}_{\text{prior}} = (2,2)^{\intercal}$ and $\boldsymbol{\Sigma}_{\text{prior}} = \text{diag}(4,4)$. A set of 10 synthetic measurements $\boldsymbol{y}$ is drawn from $\mathscr{N}((1,2)^{\intercal}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \text{diag}(4,1)$.

The same procedure as before is followed, but this time starting with $N_{\text{enr.}} = 2 \cdot \text{card}(\mathscr{A}^k) = 20$ samples in the initial domain. At every refinement step, the experimental design is enriched with 20 samples in every subdomain.

The convergence behaviour is visualized in Figure 5 and the approximated posterior marginals af-

(a) Refinement step 1



(b) Refinement step 5

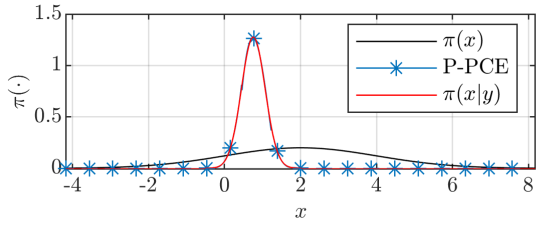


Figure 4: 1D example – Analytical prior and posterior distributions and the P-PCE posterior approximation after 1 and 5 refinement steps.

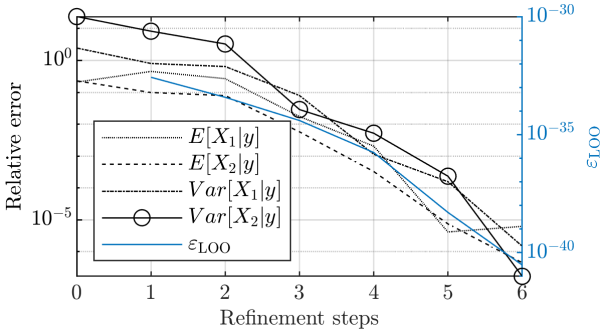


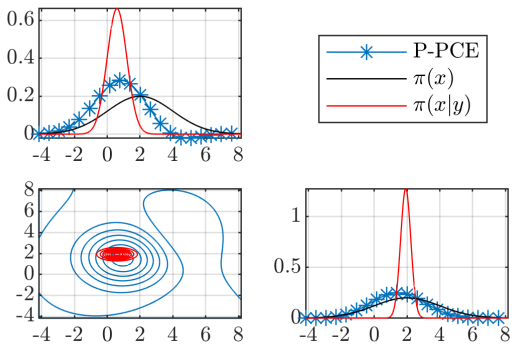
Figure 5: 2D example – Relative error convergence of the mean and variance of $X_i|\mathbf{y}$ and the behaviour of the average leave-one-out error defined in Eq. (25).

ter the 1-st and 5-th refinement step are shown in Figure 6. The number of model evaluations at the 5-th refinement step, where all moment errors fall below 1%, is $N(5) = 6820$.

6. Conclusions

The presented approach for constructing piecewise PCEs to approximate likelihood functions has very promising properties. It was shown that, using the sequential polynomial approximation from Eq. (10), all posterior moments and quantities of interest can be calculated by mere post-processing of the underlying PCE coefficients. Furthermore the approach of summing up individual PCEs allows

(a) Refinement step 1



(b) Refinement step 5

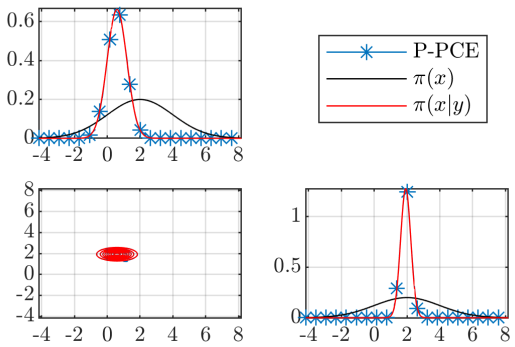


Figure 6: 2D example – Analytical prior and posterior distributions and the P-PCE posterior approximation after 1 and 5 refinement steps.

to drop the requirement of a globally positive PCE from standard SLE (Nagel and Sudret, 2016), because the individual summands can be negative as long as they sum up to a positive value.

The bottleneck of the current procedure is the construction of the P-PCE detailed in Section 4. Notice that 2^{Mr} new PCEs have to be computed at each refinement level. Consequently, in moderate to high dimensions, the required number of samples grows exponentially, rendering the approach computationally infeasible.

Nonetheless, the analytically available posterior quantities of interest and likelihood approximation make the proposed framework appealing. Future research will focus on facilitating the construction of P-PCE by means of adaptive construction of the subdomains, as well as sparse polynomial chaos expansions in each subdomain coupled with adaptive sampling.

A. STIELTJES PROCEDURE

For details refer to Gautschi (2004). To construct a series of univariate orthonormal polynomials with respect to a weight function $\pi(x)$ on a domain $\mathscr{D}$ the following recurrence relation can be used for $p \in \mathbb{N}_0$:

$$\sqrt{\beta_{p+1}}\psi_{p+1}(x) = (x-\alpha_p)\psi_p(x) - \sqrt{\beta_p}\psi_{p-1}(x). \tag{30}$$

The initial polynomials are given by $\psi_{-1} = 0$ and $\psi_0 = 1/\sqrt{\int_{\mathscr{D}} \pi(x)\,\mathrm{d}x}$. The coefficients can be calculated through the Christoffel-Darboux formulae

$$\alpha_p = \frac{\langle x\psi_p, \psi_p\rangle_\pi}{\langle \psi_p, \psi_p\rangle_\pi}, \qquad \beta_p = \frac{\langle \psi_p, \psi_p\rangle_\pi}{\langle \psi_{p-1}, \psi_{p-1}\rangle_\pi}, \tag{31}$$

with the inner product $\langle x_1, x_2\rangle_\pi = \int_{\mathscr{D}} x_1 x_2 \pi(x)\,\mathrm{d}x$.

REFERENCES

Arlot, S. and Celisse, A. (2010). "A survey of cross-validation procedures for model selection." *Stat. Surv.*, 4, 40–79.

Askey, R. and Wilson, J. (1985). "Some basic hypergeometric polynomials that generalize Jacobi polynomials." *Mem. Amer. Math. Soc.*, 319.

Berveiller, M., Sudret, B., and Lemaire, M. (2006). "Stochastic finite elements: a non intrusive approach by regression." *Eur. J. Comput. Mech.*, 15(1-3), 81–92.

Blatman, G. and Sudret, B. (2011). "Adaptive sparse polynomial chaos expansion based on Least Angle Regression." *J. Comput. Phys*, 230, 2345–2367.

Friedman, J. (1991). "Multivariate adaptive regression splines." *Annals Statistics*, 19, 1–1416.

Gautschi, W. (2004). *Orthogonal Polynomials: Computation and Approximation*. Numerical mathematics and scientific computation. Oxford University Press.

Goodman, J. and Weare, J. (2010). "Ensemble samplers with affine invariance." *Comm. Appl. Math. Comput. Sci.*, 5(1), 65–80.

Haario, H., Saksman, E., and Tamminen, J. (2001). "An adaptive Metropolis algorithm." *Bernoulli*, 7(2), 223–242.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The elements of statistical learning: Data mining, inference and prediction*. Springer, New York.

Hastings, W. K. (1970). "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika*, 57(1), 97–109.

Higdon, D., McDonnell, J. D., Schunck, N., Sarich, J., and Wild, S. M. (2015). "A Bayesian approach for parameter estimation and prediction using a computationally intensive model." *J. Phys. G: Nuclear and Particle Physics*, 42(3), 1–18.

Li, J. and Marzouk, Y. M. (2014). "Adaptive construction of surrogates for the Bayesian solution of inverse problems." *Meth. Alg. Sci. Comput.*, 36(3), A1163–A1186.

Marzouk, Y. and Xiu, D. (2009). "A stochastic collocation approach to Bayesian inference in inverse problems." *Comm. Comput. Phys.*, 6(4), 826–847.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). "Equation of state calculations by fast computing machines." *J. Chem. Phys.*, 21(6), 1087–1092.

Nagel, J. and Sudret, B. (2016). "Spectral likelihood expansions for Bayesian inference." *J. Comput. Phys.*, 309, 267–294.

Robert, C. and Casella, G. (2004). *Monte Carlo statistical methods (2^{nd} Ed.)*. Springer Series in Statistics. Springer Verlag.

Soize, C. and Ghanem, R. (2004). "Physical systems with random uncertainties: chaos representations with arbitrary probability measure." *SIAM J. Sci. Comput.*, 26(2), 395–410.

Xiu, D. and Karniadakis, G. (2002). "The Wiener-Askey polynomial chaos for stochastic differential equations." *SIAM J. Sci. Comput.*, 24(2), 619–644.