PC-Kriging: the best of polynomial chaos expansions and Gaussian process modelling

Author(s):
Sudret, Bruno; Schöbi, Roland

Publication Date:
2014

Permanent Link:
https://doi.org/10.3929/ethz-a-010346614

Rights / License:
In Copyright - Non-Commercial Use Permitted
PC-Kriging: the best of polynomial chaos expansions and Gaussian process modelling

B. Sudret & Roland Schöbi

Chair of Risk, Safety & Uncertainty Quantification

SIAM UQ’2014 – April 1st, 2014
Consider a physical system whose behaviour is modelled by a black-box function $M$ resulting from the discretization/solving of governing equations:

$$x \in \mathcal{D}_X \subset \mathbb{R}^M \mapsto M(x)$$

Due to uncertainties, the input parameters are modelled by a random vector $X$ with probabilistic joint probability density function $f_X$.

Of interest is the characterization of the statistical properties of the quantities of interest $Y = M(X)$.
Uncertainty Quantification in engineering

- Uncertainty quantification arrives on top of well-defined simulation procedures (legacy codes)

Non intrusive techniques

- The computational models are complex and time-consuming (transient, non linear including e.g. plasticity, large strains, buckling, etc.)

Parsimony

- Input variables are numerous (typically 10-50) and non Gaussian

Curse of dimensionality

Needs for surrogate models in uncertainty quantification
Surrogate models for uncertainty quantification

A surrogate model $\tilde{M}$ is an approximation of the original computational model:

- It is built from a **limited** set of runs of the original model $M$ called the **experimental design** $\mathcal{X} = \{x^{(i)}, i = 1, \ldots, n\}$
- It assumes some regularity of the model $M$ and some general functional shape

<table>
<thead>
<tr>
<th>Name</th>
<th>Shape</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynomial chaos expansions</td>
<td>$\tilde{M}(x) = \sum_{\alpha \in A} y_\alpha \Psi_\alpha(x)$</td>
<td>$y_\alpha$</td>
</tr>
<tr>
<td>Kriging</td>
<td>$\tilde{M}(x) = \beta^T \cdot f(x) + Z(x, \omega)$</td>
<td>$\beta, \sigma_Z^2, \theta$</td>
</tr>
<tr>
<td>Support vector machines</td>
<td>$\tilde{M}(x) = \sum_{i=1}^m y_i K(x_i, x) + b$</td>
<td>$y, b$</td>
</tr>
</tbody>
</table>
A surrogate model $\tilde{M}$ is an approximation of the original computational model:

- It is built from a limited set of runs of the original model $M$ called the experimental design $\mathcal{X} = \{x^{(i)}, i = 1, \ldots, n\}$
- It assumes some regularity of the model $M$ and some general functional shape

<table>
<thead>
<tr>
<th>Name</th>
<th>Shape</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynomial chaos expansions</td>
<td>$\tilde{M}(x) = \sum_{\alpha \in A} y_\alpha \Psi_\alpha(x)$</td>
<td>$y_\alpha$</td>
</tr>
<tr>
<td>Kriging</td>
<td>$\tilde{M}(x) = \beta^T \cdot f(x) + Z(x, \omega)$</td>
<td>$\beta, \sigma^2_Z, \theta$</td>
</tr>
<tr>
<td>Support vector machines</td>
<td>$\tilde{M}(x) = \sum_{i=1} y_i K(x_i, x) + b$</td>
<td>$y, b$</td>
</tr>
</tbody>
</table>
Polynomial chaos expansions

- Have appeared in the early 90’s in mechanical and civil engineering applications, later in CFD
  
  Ghanem & Spanos (1991), Le Maître, Knio et al., Xiu, Karniadakis et al., (2002- )

- Since the early 2000’s, big focus from the applied mathematics community (stochastic partial differential equations)

  Babuška et al., Nobile et al., Cohen, Schwab et al.

- Built-in method for UQ, mean-square convergence (error estimators)
Polynomial chaos expansions vs. Kriging

Kriging

- Historical roots in geostatistics, pioneering papers in computer experiments
  
  Sacks et al. (1989, 1992), O’Hagan et al., Rasmussen et al.

- Many recent developments for optimization, sensitivity analysis, robust design, etc.
  
  Iooss et al., Garnier et al., Roustant, Ginsbourger et al.

- Built-in local error estimation (Kriging variance) allowing for adaptive algorithms

Best of PC expansions and Kriging combined into PC-Kriging
Outline

1 Kriging & Polynomial chaos expansions
   - Kriging
   - Polynomial chaos expansions

2 PC-Kriging

3 Application examples
Kriging (Gaussian process modelling)

Definition

Kriging assumes that $\mathcal{M}(x)$ is a trajectory of an underlying Gaussian process

$$\mathcal{M}(x) \approx \mathcal{M}^{(K)} = \beta^T f(x) + \sigma^2 Z(x, \omega)$$

where:
- $\beta^T f(x)$: trend
- $Z(x, \omega)$: zero mean, unit variance Gaussian process
- $\sigma^2$: variance

Calibration

Parameters $\{\theta, \beta, \sigma_y^2\}$ are computed by maximum likelihood estimation, cross validation, Bayesian calibration, etc.

You probably know already!
Kriging (Gaussian process modelling)

Definition

Kriging assumes that $\mathcal{M}(x)$ is a trajectory of an underlying Gaussian process

$$\mathcal{M}(x) \approx \mathcal{M}^{(K)} = \beta^T f(x) + \sigma^2 Z(x, \omega)$$

where:

- $\beta^T f(x)$: trend
- $Z(x, \omega)$: zero mean, unit variance Gaussian process
- $\sigma^2$: variance

Calibration

- Parameters $\{\theta, \beta, \sigma_y^2\}$ are computed by maximum likelihood estimation, cross validation, Bayesian calibration, etc.

You probably know already!
Polynomial chaos expansion in a nutshell

- Consider the input random vector $X$ (dim $X = M$) and a probabilistic space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(dx) = f_X(x) \, dx$

- Assuming that the random output $Y = \mathcal{M}(X)$, it can be cast as the following polynomial chaos expansion:

$$Y = \sum_{\alpha \in \mathbb{N}^M} y_{\alpha} \Psi_\alpha(X)$$

where:

- $y_{\alpha}$: coefficients to be computed (coordinates)
- $\Psi_\alpha(X)$: basis

The basis $\{\Psi_\alpha(X), \alpha \in \mathbb{N}^M\}$ is made of multivariate orthonormal polynomials. Given a tuple $\alpha = \{\alpha_1, \ldots, \alpha_M\}$:

$$\Psi_\alpha(x) \overset{\text{def}}{=} \prod_{i=1}^{M} \Psi_{\alpha_i}^i(x_i) \quad \mathbb{E}[\Psi_\alpha(X)\Psi_\beta(X)] = \delta_{\alpha\beta} \quad \text{(Kronecker symbol)}$$

B. Sudret (Chair of Risk & Safety)
 Classical orthogonal polynomials

<table>
<thead>
<tr>
<th>Type of variable</th>
<th>Weight function</th>
<th>Orthogonal polynomials</th>
<th>Hilbertian basis $\psi_k(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>$[1,1] - 1,1[x]/2$</td>
<td>Legendre $P_k(x)$</td>
<td>$P_k(x)/\sqrt{2k+1}$</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$1/\sqrt{2\pi} e^{-x^2}/2$</td>
<td>Hermite $H_k(x)$</td>
<td>$H_k(x)/\sqrt{k!}$</td>
</tr>
<tr>
<td>Gamma</td>
<td>$x^a e^{-x} 1_{\mathbb{R}^+}(x)$</td>
<td>Laguerre $L_k^a(x)$</td>
<td>$L_k^a(x)/\sqrt{\Gamma(k+a+1)/k!}$</td>
</tr>
<tr>
<td>Beta</td>
<td>$[1,1] - 1,1[x] (1-x)^a(1+x)^b / B(a)B(b)$</td>
<td>Jacobi $J_k^{a,b}(x)$</td>
<td>$J_k^{a,b}(x)/\Gamma(a+k+1)\Gamma(b+k+1)$</td>
</tr>
</tbody>
</table>

Practical implementation

- **A truncation scheme is selected**, e.g.: 

$$\mathcal{A}^{M,p} = \{\alpha \in \mathbb{N}^M : |\alpha| \leq p\} \quad \text{card} \ \mathcal{A}^{M,p} = P = \binom{M+p}{p}$$

- **Expansions coefficients** $\{y_\alpha, \alpha \in \mathcal{A}\}$ are computed:
  - Stochastic Galerkin approach
  - Spectral projection, stochastic collocation
  - Least-square minimization
Least-square minimization

The exact (infinite) series expansion is considered as the sum of a truncated series and a residual:

$$\mathcal{M}(X) = \sum_{\alpha \in \mathcal{A}} y_{\alpha} \Psi_{\alpha}(X) + \varepsilon_P \equiv Y^T \Psi(X) + \varepsilon_P$$

where: \( Y = \{ y_{\alpha}, \alpha \in \mathcal{A} \} \) and \( \Psi(x) = \{ \Psi_{\alpha}(x), \alpha \in \mathcal{A} \} \)

Solution

$$\hat{Y}^{L.S} = \arg \min \frac{1}{n} \sum_{i=1}^{n} \left( Y^T \Psi(x^{(i)}) - \mathcal{M}(x^{(i)}) \right)^2$$

$$= (A^T A)^{-1} A^T \mathcal{M}$$

where:

$$\mathcal{M} = \{ \mathcal{M}(x^{(1)}), \ldots, \mathcal{M}(x^{(n)}) \}^T$$

$$A_{ij} = \Psi_j(x^{(i)}) \quad i = 1, \ldots, n \ ; \ j = 0, \ldots, |\mathcal{A}|$$
Sparse PC expansions

Premise

- Standard least-square analysis requires to choose the truncation scheme $\mathcal{A}$ a priori, and to have an experimental of size $n > |\mathcal{A}|$, i.e. typically $n = 3 \cdot |\mathcal{A}|$

Sparse expansions

- Finding the relevant sparse basis among a large candidate basis $\mathcal{A}$ is a variable selection problem: $\ell_1$ penalization

- The Least Angle Regression (LAR) has proven efficient:
  - A sufficient large candidate truncated basis $\mathcal{A}$ is selected
  - LAR provides a sequence of less and less sparse models
  - The best one is finally selected from leave-one-out cross-validation

$$
\epsilon_{LOO} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathcal{M}(\mathbf{x}^{(i)}) - \mathcal{M}^{PC\setminus i}(\mathbf{x}^{(i)}) \right\}^2
$$
Heuristics

- Combine polynomial chaos expansions (PCE) and Kriging:
  - PCE approximates the global behaviour of the computational model.
  - Kriging allows for local interpolation.

Universal Kriging model with a sparse PC expansion as a trend

\[ M(x) \approx M^{(PCK)}(x) = \sum_{\alpha \in A} a_{\alpha} \psi_{\alpha}(x) + \sigma^2 Z(x, \omega) \]

where \( \{\Psi_{\alpha}(x), \alpha \in A\} \) is a set of orthonormal polynomials found by LARS.

Two approaches for calibration

- **Sequential** PC-Kriging
- **Optimal** PC-Kriging
Sequential PC-Kriging

Input Distributions \rightarrow LARS \rightarrow y = M^{(K)}(P* = P) \rightarrow M^{(FPCK)} \rightarrow Full-PC-Kriging \rightarrow Prediction

Experimental Design \rightarrow Autocorr. Function
Optimal PC-Kriging

Input Distributions → Experimental Design → Autocorr. Function

LARS

\[ y = M^{(K)}(P^*=1) \]
\[ y = M^{(K)}(P^*=2) \]
\[ \cdots \]
\[ y = M^{(K)}(P^*=P) \]

LOO

\[ M^{(OPCK)} \]

Optimal-PC-Kriging

Prediction
Ishigami function \((M = 3)\)

**Definition**

\[
f_1(X) = \sin X_1 + 7 \sin^2 X_2 + 0.1 X_3^4 \sin X_1
\]

where:

\[
X_i = \mathcal{U}(-\pi, \pi), \quad i = 1, 2, 3
\]

**Error estimation**

- Each metamodel is calibrated from a Latin Hypercube Sampling (LHS) experimental design of size \(n\)
- For each metamodel \(\hat{M}\) the relative generalization error is computed from a large validation set:

\[
\epsilon_{gen} \equiv \frac{\sum_{i=1}^{n_{val}} \left( M(x_i) - \hat{M}(x_i) \right)^2}{\sum_{i=1}^{n_{val}} (M(x_i) - \mu_y)^2}
\]

\(n_{val} = 10^5\)

- The analysis is replicated \(B = 50\) times and box plots are obtained
Ishigami function – small experimental design

(a) Ordinary Kriging

(b) PCE

(c) Sequential PC-Kriging

(d) Optimal PC-Kriging
Ishigami function – large experimental design

- **(e) Ordinary Kriging**
- **(f) PCE**
- **(g) Sequential PC-Kriging**
- **(h) Optimal PC-Kriging**
Morris function \((M = 20)\)

**Definition**

\[
f_2(X) = \sum_{i=1}^{20} \beta_i w_i + \sum_{i<j}^{20} \beta_{ij} w_i w_j + \sum_{i<j<l}^{20} \beta_{ijl} w_i w_j w_l + 5 w_1 w_2 w_3 w_4
\]

where:

- \(X_i = \mathcal{U}(-1, 1), \ i = 1, \ldots, 20\)
- \(w_i = 2(X_i - 1/2)\) for all \(i\) except for \(i = 3, 5, 7\) where
  \[
  w_i = 2 \left( \frac{1.1 X_i}{X_i + 0.1} - 1/2 \right)
  \]
- The coefficients are defined as: \(\beta_i = 20, \ i = 1, \ldots, 10;\)
  \(\beta_{ij} = -15, \ i, j = 1, \ldots, 6;\)
  \(\beta_{ijl} = -10, \ i, j, l = 1, \ldots, 5\)
Morris function – small experimental design

(i) Ordinary Kriging

(j) PCE

(k) Sequential PC-Kriging

(l) Optimal PC-Kriging
Conclusions

- Polynomial chaos expansions and Kriging are two powerful meta-modelling techniques (historically) developed by different communities.

- **PC-Kriging** combines the two by considering a sparse PC expansion as the trend of a universal Kriging model.

- PCE allows for a global approximation while the Kriging part interpolates the local variability.

- The coupled approach is always as good as the best of pure PC / pure Kriging.

- There is a significant improvement for small experimental designs.
Questions ?

Thank you very much for your attention!

Chair of Risk, Safety & Uncertainty Quantification

http://www.rsuq.ethz.ch

UQLab ...

... The Uncertainty Quantification Toolbox in Matlab