Sparse polynomial chaos expansions and application to sensitivity analysis
BOQUSE’2013 – December 18th – INRIA

Author(s):
Sudret, Bruno

Publication Date:
2013

Permanent Link:
https://doi.org/10.3929/ethz-a-010060789

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
Sparse polynomial chaos expansions and application to sensitivity analysis

B. Sudret

Chair of Risk, Safety & Uncertainty Quantification

BOQUSE’2013 – December 18th – INRIA
Chair of Risk, Safety and Uncertainty quantification

The Chair carries out research projects in the field of uncertainty quantification for engineering problems with applications in structural reliability, sensitivity analysis, model calibration and reliability-based design optimization.
The Chair carries out research projects in the field of uncertainty quantification for engineering problems with applications in structural reliability, sensitivity analysis, model calibration and reliability-based design optimization.

Research topics

- Structural reliability analysis
- Polynomial chaos expansions and stochastic finite element methods
- Advanced meta-models (kriging, support vector machines)
- Bayesian model calibration and stochastic inverse problems
- Global sensitivity analysis
- Reliability-based design optimization

http://www.rsuq.ethz.ch
The Chair carries out research projects in the field of uncertainty quantification for engineering problems with applications in structural reliability, sensitivity analysis, model calibration and reliability-based design optimization.

Research topics

- Structural reliability analysis
- **Polynomial chaos expansions** and stochastic finite element methods
- Advanced meta-models (kriging, support vector machines)
- Bayesian model calibration and stochastic inverse problems
- **Global sensitivity analysis**
- Reliability-based design optimization

http://www.rsuq.ethz.ch
Uncertainty quantification in industrial applications

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

- The computational models are complex: coupled problems (thermo-mechanics), plasticity, large strains, contact, buckling, etc.

- A single simulation is already costly (e.g. several hours)

- Engineers focus on so-called quantities of interest, e.g. maximum displacement, average stress, etc.
Uncertainty quantification in industrial applications

- The input variables modelling aleatory uncertainty are often non Gaussian.
- The size of the input random vector is typically 10-100.
- UQ procedures shall be sufficiently general to be applied with little adaptation to a variety of problems.

Need for non intrusive and parsimonious methods for uncertainty quantification.
Global framework for uncertainty quantification

Step A
Model(s) of the system
Assessment criteria

Step B
Quantification of sources of uncertainty

Random variables

Computational model

Step C
Uncertainty propagation

Distribution
Mean, std. deviation
Probability of failure

Step C'
Sensitivity analysis

Global framework for uncertainty quantification

**Step A**
Model(s) of the system
Assessment criteria

**Step B**
Quantification of sources of uncertainty

**Step C**
Uncertainty propagation

1. Random variables
2. Computational model
3. Distribution
   - Mean, std. deviation
   - Probability of failure

**Step C’**
Sensitivity analysis

Global framework for uncertainty quantification

**Step A**
Model(s) of the system
Assessment criteria

**Step B**
Quantification of sources of uncertainty

**Step C**
Uncertainty propagation

Random variables → Computational model → Distribution
Mean, std. deviation
Probability of failure

**Step C’**
Sensitivity analysis

Global framework for uncertainty quantification

**Step A**
Model(s) of the system
Assessment criteria

**Step B**
Quantification of sources of uncertainty

**Step C**
Uncertainty propagation

**Step C’**
Sensitivity analysis

Random variables

Computational model

Distribution
Mean, std. deviation
Probability of failure

Outline

1. Sparse polynomial chaos scheme
   - Sparse truncation schemes
   - Computation of the coefficients
   - Error estimation and validation

2. Adaptive algorithms for sparse expansions

3. Sensitivity analysis
   - Sobol’ indices
   - Case of dependent inputs: ANCOVA

4. Application examples in sensitivity analysis
   - Ishigami function
   - Morris function
   - Bending beam
Spectral approach

- The input parameters are modelled by a random vector $X$ over a probabilistic space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(dx) = f_X(x) \, dx$

- The response random vector $Y = M(X)$ is considered as an element of $L^2(\Omega, \mathcal{F}, \mathbb{P})$

- A basis of multivariate orthogonal polynomials is built up with respect to the input PDF (assuming independent components)

- The response random vector $Y$ is completely determined by its coordinates in this basis
The input parameters are modelled by a random vector $X$ over a probabilistic space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(dx) = f_X(x) \, dx$

The response random vector $Y = \mathcal{M}(X)$ is considered as an element of $L^2(\Omega, \mathcal{F}, \mathbb{P})$

A basis of multivariate orthogonal polynomials is built up with respect to the input PDF (assuming independent components)

The response random vector $Y$ is completely determined by its coordinates in this basis
Spectral approach

- The input parameters are modelled by a random vector $\mathbf{X}$ over a probabilistic space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(d\mathbf{x}) = f_X(x) \, dx$

- The response random vector $\mathbf{Y} = \mathcal{M}(\mathbf{X})$ is considered as an element of $L^2(\Omega, \mathcal{F}, \mathbb{P})$

- A basis of multivariate orthogonal polynomials is built up with respect to the input PDF (assuming independent components)

- The response random vector $\mathbf{Y}$ is completely determined by its coordinates in this basis

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

where:

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

Univariate orthogonal polynomials

For each marginal distribution $f_{X_i}(x_i)$ one can define a functional inner product:

$$\langle \phi_1, \phi_2 \rangle_i = \int_{D_i} \phi_1(x) \phi_2(x) f_{X_i}(x_i) \, dx_i$$

and a family of orthogonal polynomials $\{P^{(i)}_k, k \in \mathbb{N}\}$ such that:

$$\langle P^{(i)}_j, P^{(i)}_k \rangle = \int P^{(i)}_j(x) P^{(i)}_k(x) f_{X_i}(x) \, dx = a^i_{jk} \delta_{jk}$$

Classical families Xiu & Karniadakis (2002)

<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>$[-1,1]$</td>
<td>Legendre $P_k(x)$</td>
<td>$P_k(x)/\sqrt{\frac{1}{2k+1}}$</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$</td>
<td>Hermite $H_{\ell_k}(x)$</td>
<td>$H_{\ell_k}(x)/\sqrt{k!}$</td>
</tr>
<tr>
<td>Gamma</td>
<td>$x^a e^{-x} \mathbb{1}_{\mathbb{R}^+}(x)$</td>
<td>Laguerre $L^a_k(x)$</td>
<td>$L^a_k(x)/\sqrt{\frac{\Gamma(k+a+1)}{k!}}$</td>
</tr>
<tr>
<td>Beta</td>
<td>$[0,1]$</td>
<td>Jacobi $J^a,b_{\ell_k}(x)$</td>
<td>$J^a,b_{\ell_k}(x)/\sqrt{\frac{\Gamma(k+a+1)\Gamma(k+b+1)}{\Gamma(k+a+b+1)\Gamma(k+1)}}$</td>
</tr>
</tbody>
</table>
Polynomial chaos basis

Univariate orthogonal polynomials

For each marginal distribution $f_{X_i}(x_i)$ one can define a functional inner product:

$$\langle \phi_1, \phi_2 \rangle_i = \int_{D_i} \phi_1(x) \phi_2(x) f_{X_i}(x_i) \, dx_i$$

and a family of orthogonal polynomials $\{P^{(i)}_k, k \in \mathbb{N}\}$ such that:

$$\langle P^{(i)}_j, P^{(i)}_k \rangle = \int P^{(i)}_j(x) P^{(i)}_k(x) f_{X_i}(x) \, dx = a^i_{j} \delta_{jk}$$

Classical families

<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>$[1,1] - 1, 1 \cdot (x)/2$</td>
<td>Legendre $P_k(x)$</td>
<td>$P_k(x) / \sqrt{\frac{1}{2k+1}}$</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$</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>$L_k^a(x) / \sqrt{\frac{\Gamma(k+a+1)}{k!}}$</td>
</tr>
<tr>
<td>Beta</td>
<td>$[1,1] - 1, 1 \cdot \frac{(1-x)^a (1+x)^b}{B(a) B(b)}$</td>
<td>Jacobi $J_k^{a,b}(x)$</td>
<td>$J_k^{a,b}(x) / \sqrt{\frac{\Gamma(k+a+1) \Gamma(k+b+1)}{\Gamma(k+a+b+1) \Gamma(k+1)}}$</td>
</tr>
</tbody>
</table>
Multivariate polynomials

Let us define the multi-indices (tuples) \( \alpha = \{\alpha_1, \ldots, \alpha_M\} \), of degree

\[ |\alpha| = \sum_{i=1}^{M} \alpha_i. \]

The associated multivariate polynomial reads:

\[ \Psi_\alpha(x) = \prod_{i=1}^{M} P_{\alpha_i}^{(i)}(x_i) \]

The set of multivariate polynomials \( \{\Psi_\alpha, \alpha \in \mathbb{N}^M\} \) forms a basis of the space of second order variables:

\[ Y = \sum_{\alpha \in \mathbb{N}^M} y_\alpha \Psi_\alpha(X) \]
Curse of dimensionality

Truncated series

- A truncation scheme is selected and the associated finite set of multi-indices is generated.
- The common truncation scheme considers all polynomials up to a given total degree, e.g.:

\[ \mathcal{A}^{M,p} = \{ \alpha \in \mathbb{N}^M : |\alpha| \leq p \} \]

Drawback: The number of unknown coefficients card \( \mathcal{A}^{M,p} \) grows polynomially both in \( M \) and \( p \):

\[ \text{card } \mathcal{A}^{M,p} = \binom{M + p}{p} \]

"Full" expansions are not tractable when \( M \geq 10 \)

Solutions:

- Sparse truncation schemes, based on the sparsity-of-effect principle
- Adaptive algorithms for the construction of the PC expansion
Why are sparse representations relevant?

Blatman & S., (2008); Cohen, Schwab et al. (2010-13); Doostan & Owhadi (2011)

**Sparsity-of-effects principle:** in usual problems, only low-order interactions between the input variables are relevant. One shall select PC approximations using *low-rank* monomials

**Degree** of a multi-index $\alpha$: total degree of polynomial $\Psi_\alpha$

$$|\alpha| \equiv ||\alpha||_1 = \sum_{i=1}^{M} \alpha_i$$

**Rank** of a multi-index $\alpha$: number of active variables of $\Psi_\alpha$ (non zero terms of multi-index $\alpha$)

$$||\alpha||_0 = \sum_{i=1}^{M} 1_{\{\alpha_i > 0\}}$$
Two selection techniques

- Low-rank index sets:
  \[ \mathcal{A}^{M,p,j} = \{ \alpha \in \mathbb{N}^M : |\alpha| \leq p, ||\alpha||_0 \leq j \} \]

- Hyperbolic sets:
  \[ \mathcal{A}_q^{M,p} = \{ \alpha \in \mathbb{N}^M : ||\alpha||_q \leq p \} \]
  
  where \( ||\alpha||_q \equiv \left( \sum_{i=1}^{M} \alpha_i^q \right)^{1/q} \), \( 0 < q < 1 \)

Limit cases

- \( q = 1 \) : common truncation scheme (all polynomials of maximal total degree \( p \))
- \( q \to 0 \) : additive model (no interaction)
Hyperbolic truncation schemes (cont’)

The hyperbolic norm primarily selects the high-degree polynomials in one single variable and then the polynomials involving few interaction.

\[
\begin{align*}
\|x\|_1 & \leq 3 \\
\|x\|_{0.75} & \leq 3 \\
\|x\|_0.5 & \leq 3 \\
\|x\|_1 & \leq 4 \\
\|x\|_{0.75} & \leq 4 \\
\|x\|_0.5 & \leq 4 \\
\|x\|_1 & \leq 5 \\
\|x\|_{0.75} & \leq 5 \\
\|x\|_0.5 & \leq 5 \\
\|x\|_1 & \leq 6 \\
\|x\|_{0.75} & \leq 6 \\
\|x\|_0.5 & \leq 6
\end{align*}
\]
Index of sparsity

Common truncation \((\mathcal{A}^{M,p})\)  
Hyperbolic truncation \((\mathcal{A}_{q}^{M,p})\)

- Effect of the hyperbolic truncation scheme

\[ IS1 = \frac{\text{card } \mathcal{A}_{q}^{M,p}}{\text{card } \mathcal{A}^{M,p}} \]
Index of sparsity

- Effect of the hyperbolic truncation scheme

\[ IS1 = \frac{\text{card } A_{M,p}^q}{\text{card } A_{M,p}^M} \]
Outline

1. Sparse polynomial chaos scheme
   - Sparse truncation schemes
   - Computation of the coefficients
   - Error estimation and validation

2. Adaptive algorithms for sparse expansions

3. Sensitivity analysis

4. Application examples in sensitivity analysis
Various methods for computing the coefficients

Intrusive approaches

- Historical approaches: projection of the equations residuals in the Galerkin sense
  Ghanem et al.; Le Maître et al.; Babuska, Tempone et al.; Karniadakis et al., etc.
- Proper generalized decompositions
  Nouy et al., 2007-10

Non intrusive approaches

- Non intrusive methods consider the computational model $\mathcal{M}$ as a black box
- They rely upon a design of numerical experiments, i.e. a $n$-sample $\mathcal{X} = \{x^{(i)} \in \mathcal{D}_X, i = 1, \ldots, n\}$ of the input parameters
- Different classes of methods are available:
  - projection: by simulation or quadrature
    Matthies & Keese, 2005; Le Maître et al.
  - stochastic collocation
    Xiu, 2007-09; Nobile, Tempone et al., 2008; Ma & Zabaras, 2009
  - least-square minimization
    Berveiller et al., 2006; Blatman & S., 2008-11
Statistical approach: least-square minimization

Principle

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

\[ Y = M(X) = \sum_{j=0}^{P-1} y_j \Psi_j(X) + \varepsilon_P \equiv Y^T \Psi(X) + \varepsilon_P \]

where:

\[ Y = \{y_0, \ldots, y_{P-1}\} \]

\[ \Psi(x) = \{\Psi_0(x), \ldots, \Psi_{P-1}(x)\} \]
Least-Square Minimization: continuous solution

Least-square minimization

The coefficients are gathered into a vector $\hat{Y}$, and computed by minimizing the mean square error:

$$
\hat{Y} = \arg \min_{\mathcal{Y}} E \left[ (Y^T \Psi(X) - \mathcal{M}(X))^2 \right]
$$

Analytical solution (continuous case)

The least-square minimization problem may be solved analytically:

$$
\hat{Y} = \mathbb{E} \left[ \mathcal{M}(X) \Psi(X) \right]
$$

The solution is identical to the projection solution due to the orthogonality properties of the regressors.
Least-Square Minimization: discretized solution

Resolution

An estimate of the mean square error (sample average) is minimized:

\[
\hat{Y}_{LS} = \arg \min \hat{E} \left[ (Y^T \Psi(X) - M(X))^2 \right]
\]

\[
= \arg \min \frac{1}{n} \sum_{i=1}^{n} \left( Y^T \Psi(x^{(i)}) - M(x^{(i)}) \right)^2
\]
Least-Square Minimization: discretized solution

- **Select an experimental design**
  \[ X = \{ x^{(1)}, \ldots, x^{(n)} \}^T \]
  that covers at best the domain of variation of the parameters

- **Evaluate the model response for each sample** (exactly as in Monte Carlo simulation)
  \[ \mathcal{M} = \{ M(x^{(1)}), \ldots, M(x^{(n)}) \}^T \]

- **Compute the experimental matrix**
  \[ A_{ij} = \Psi_j(x^{(i)}) \quad i = 1, \ldots, n \; ; \; j = 0, \ldots, P - 1 \]

- **Solve the least-square minimization problem**
  \[ \hat{Y} = (A^T A)^{-1} A^T \mathcal{M} \]
Choice of the experimental design

Random designs

- Monte Carlo samples obtained by standard random generators
- Latin Hypercube designs that are both random and “space-filling”
- Quasi-random sequences (e.g. Sobol’ sequence)
Outline

1. Sparse polynomial chaos scheme
   - Sparse truncation schemes
   - Computation of the coefficients
   - Error estimation and validation

2. Adaptive algorithms for sparse expansions

3. Sensitivity analysis

4. Application examples in sensitivity analysis
Validation of the surrogate model

- The truncated series expansions are convergent in the mean square sense. However one does not know in advance where to truncate (problem-dependent).

- Usually the series is truncated according to the total maximal degree of the polynomials, say \( p = 2, 3, 4 \) etc.

- The recent research deals with the development of error estimates:
  - adaptive integration in the projection approach
  - cross validation in the least-squares approach
The least-squares technique is based on the minimization of the mean square error. The generalization error is defined as:

$$E_{gen} = \mathbb{E} \left[ (\mathcal{M}(X) - \mathcal{M}^{PC}(X))^2 \right]$$

It may be estimated by the empirical error using the already computed response quantities:

$$E_{emp} = \frac{1}{n} \sum_{i=1}^{n} \left( \mathcal{M}(x^{(i)}) - \mathcal{M}^{PC}(x^{(i)}) \right)^2$$

The coefficient of determination $R^2$ is often used as an error estimator:

$$R^2 = 1 - \frac{E_{emp}}{\hat{\mathbb{V}}[Y]}$$

$$\hat{\mathbb{V}}[Y] = \frac{1}{n} (\mathcal{M}(x^{(i)}) - \bar{Y})^2$$

This error estimator leads to overfitting
Overfitting – Illustration of the Runge effect

- When using a polynomial regression model that uses the same number of points as the degree of the polynomial, one gets an **interpolating** approximation.

- The empirical error is zero whereas the approximation gets worse and worse.
Error estimators

Leave-one-out cross validation

**Principle**

- In statistical learning theory, cross validation consists in splitting the experimental design $\mathcal{Y}$ in two parts, namely a *training set* (which is used to build the model) and a *validation set*.

- The leave-one-out technique consists in using each point of the experimental design as a single validation point for the meta-model built from the remaining $n - 1$ points.

- $n$ different meta-models are built and the error made on the remaining point is computed, then mean-square averaged.
Cross validation
Implementation

- For each $\mathbf{x}^{(i)}$, a polynomial chaos expansion is built using the following experimental design: $\mathcal{X} \setminus \mathbf{x}^{(i)} = \{ \mathbf{x}^{(j)}, j = 1, \ldots, n, j \neq i \}$, denoted by $\mathcal{M}^{PC \setminus i}(\cdot)$.

- The predicted residual is computed in point $\mathbf{x}^{(i)}$:

  $$\Delta_i = \mathcal{M}(\mathbf{x}^{(i)}) - \mathcal{M}^{PC \setminus i}(\mathbf{x}^{(i)})$$

- The PRESS coefficient (predicted residual sum of squares) is evaluated:

  $$PRESS = \sum_{i=1}^{n} \Delta_i^2$$

- The leave-one-out error and related $Q^2$ error estimator are computed:

  $$E_{LOO} = \frac{1}{n} \sum_{i=1}^{n} \Delta_i^2 \quad Q^2 = 1 - \frac{E_{LOO}}{\hat{\mathbb{V}}[Y]}$$
In practice one does **not need** to explicitly derive the \( n \) different models \( \mathcal{M}^{PC\setminus i}(. \)  

In contrast, a **single** least-square analysis is carried out. The **predicted residual** reads:

\[
\Delta_i = \mathcal{M}(x^{(i)}) - \mathcal{M}^{PC\setminus i}(x^{(i)}) = \frac{\mathcal{M}(x^{(i)}) - \mathcal{M}^{PC}(x^{(i)})}{1 - h_i}
\]

where \( h_i \) is the \( i \)-th diagonal term of matrix \( A(A^TA)^{-1}A^T \), where:

\[
A_{ij} = \Psi_j(x^{(i)})
\]

Thus:

\[
E_{LOO} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\mathcal{M}(x^{(i)}) - \mathcal{M}^{PC}(x^{(i)})}{1 - h_i} \right)^2
\]
Outline

1. Sparse polynomial chaos scheme
2. Adaptive algorithms for sparse expansions
3. Sensitivity analysis
4. Application examples in sensitivity analysis
Index of sparsity

- Effect of the hyperbolic truncation scheme

\[ IS1 = \frac{\text{card } \mathcal{A}_q^{M,p}}{\text{card } \mathcal{A}_M^{M,p}} \]
Index of sparsity

- Effect of the hyperbolic truncation scheme
  \[ IS_1 = \frac{\text{card } \mathcal{A}^M_p}{\text{card } \mathcal{A}^M_p} \]

- Effect of the adaptive selection algorithm
  \[ IS_2 = \frac{\text{card } \mathcal{A}}{\text{card } \mathcal{A}^M_p} \]
How to get sparse expansions?

- Finding the significant coefficients in the PC expansion is a **variable selection problem**.
- It can be addressed by penalized regression techniques: ridge regression ($L_2$ penalty term), **LASSO** ($L_1$ penalty term)
- The **Least Angle Regression** (LAR) algorithm is an efficient approach:
  - A set of candidate basis functions $\mathcal{A}$ is pre-selected, *e.g.* using the hyperbolic truncation scheme
  - A **family** of sparse models are built, containing resp. 1, 2, \ldots, $|\mathcal{A}|$ terms
  - Among these models the best one is retained by applying cross validation techniques
Least angle regression
Implementation

Consider a 3-dimensional vector
The algorithm is initialized with $\hat{f}^0 = 0$. The residual is $R = y$.

The most correlated regressor is $X_1$. 
A move in the direction $X_1$ is carried out so that the residual $y - \gamma_1 X_1$ becomes equicorrelated with $X_1$ and $X_2$

The 1-term sparse approximation of $y$ is $\hat{f}^1$
Least angle regression
Implementation

- A move is jointly made in the directions $X_1, X_2$ until the residual becomes equicorrelated with a new regressor.
- This gives the 2-term sparse approximation.
Least angle regression
Path of solutions

- A path of solutions is obtained containing $1, 2, \ldots, \min(n, \text{Card } A)$ terms
- A $Q^2$-estimator of the accuracy of each solution is evaluated by cross validation and the best result is kept
Basis-and-design adaptive LAR

**Initialization**
- Choose a norm $q$, $0 < q \leq 1$
- Select an initial design $\mathcal{X}$
- Store the model evaluations in $\mathcal{Y}$

**Selection of an optimal PC basis $\mathcal{A}^*$**
For $p = 1, \ldots, p_{max}$:
- Apply LAR to the candidate basis $\mathcal{A}$ which contains all those terms with $q$-norm $\leq p$
- Let $\mathcal{A}^{(p)}$ be the optimal basis obtained by LAR and $\varepsilon_{LOO}^{*}$ the corresponding error estimate (corrected leave-one-out estimate)
- Store $\varepsilon_{LOO, min}^{*} \equiv \min(\varepsilon_{LOO}^{*})$ and the associated basis $\mathcal{A}_{min}$

**STOP** if $\varepsilon_{LOO, min}^{*}$ is less than a target error $\varepsilon_{tgt}$

**Enrich the design $\mathcal{X}$** if $\varepsilon_{LOO}^{*}$ increases twice in a row (overfitting). Restart the procedure from degree $p_{min}$ of basis $\mathcal{A}_{min}$.

**Compute the coefficients associated with $\mathcal{A}_{min}$ by least-square regression**
Outline

1. Sparse polynomial chaos scheme
2. Adaptive algorithms for sparse expansions
3. Sensitivity analysis
   - Sobol' indices
   - Case of dependent inputs: ANCOVA
4. Application examples in sensitivity analysis
Sensitivity analysis

Sobol’ decomposition

- Sensitivity analysis aims at quantifying what are the input parameters (or combinations thereof) that influence the most the response variability.
- Global sensitivity analysis relies on so-called variance decomposition techniques.

Consider a model \( \mathcal{M} : \mathbf{x} \in [0, 1]^M \rightarrow \mathcal{M}(\mathbf{x}) \in \mathbb{R} \). The Hoeffding-Sobol’ decomposition reads:

\[
\mathcal{M}(\mathbf{x}) = \mathcal{M}_0 + \sum_{i=1}^{M} \mathcal{M}_i(x_i) + \sum_{1 \leq i < j \leq M} \mathcal{M}_{ij}(x_i, x_j) + \cdots + \mathcal{M}_{12\ldots M}(\mathbf{x})
\]

where:
- \( \mathcal{M}_0 \) is the mean value of the function
- \( \mathcal{M}_i(x_i) \) are univariate functions
- \( \mathcal{M}_{ij}(x_i, x_j) \) are bivariate functions
- etc.
Sobol’ decomposition

Properties

- The functional decomposition is **unique** if the orthogonality of the terms (with respect to the uniform measure) is imposed \([0, 1]^M\), i.e. \(\{i_1, \ldots, i_s\} \neq \{j_1, \ldots, j_t\}\):

\[
\int_{[0,1]^M} M_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s})M_{j_1 \ldots j_t}(x_{j_1}, \ldots, x_{j_t}) \, dx = 0
\]

- A construction definition of the terms is obtained by recurrence:

\[
M_i(x_i) = \int_{[0,1]^{M-1}} M(x) \, dx_{\sim i} - M_0
\]

\[
M_{ij}(x_i, x_j) = \int_{[0,1]^{M-2}} M(x) \, dx_{\sim \{ij\}} - M_i(x_i) - M_j(x_j) - M_0
\]

where \(\int_{[0,1]^{M-1}} (\cdot) \, dx_{\sim i}\) denoted the integration over all variables except for the \(i\)-th one.
Variance decomposition

- Assume $X_i \sim U(0, 1)$, $i = 1, \ldots, M$ (possibly after some isoprobabilistic transform)

- Due to the orthogonality of the decomposition:

$$D \equiv \text{Var} [\mathcal{M}(\mathbf{X})] = \mathbb{E} \left[ (\mathcal{M}(\mathbf{X}) - \mathcal{M}_0)^2 \right]$$

$$= \mathbb{E} \left[ \left( \sum_{\{i_1, \ldots, i_s\} \subset \{1, \ldots, M\}} \mathcal{M}_{i_1 \ldots i_s}(X_{i_1}, \ldots, X_{i_s}) \right)^2 \right]$$

$$= \sum_{\{i_1, \ldots, i_s\} \subset \{1, \ldots, M\}} \mathbb{E} \left[ \mathcal{M}_{i_1 \ldots i_s}^2(X_{i_1}, \ldots, X_{i_s}) \right]$$
**Sobol’ indices**

**Partial variance**

- Consider:

\[
D_{i_1...i_s} = \int_{[0,1]^s} \mathcal{M}_{i_1...i_s}^2(x_{i_1}, \ldots, x_{i_s}) \, dx_{i_1} \ldots dx_{i_s}
\]

- Then:

\[
D \equiv \text{Var}[Y] = \sum_{i=1}^{M} D_i + \sum_{1 \leq i < j \leq M} D_{ij} + \ldots + D_{12\ldots M}
\]

- The Sobol’ indices are obtained by normalization:

\[
S_{i_1...i_s} = \frac{D_{i_1...i_s}}{D}
\]

They represent the fraction of the total variance \(\text{Var}[Y]\) that can be attributed to each input variable \(i\) (\(S_i\)) or combinations of variables \(\{i_1 \ldots i_s\}\).
First order and total Sobol’ indices

First order Sobol’ indices

\[ S_i = \frac{D_i}{D}, \quad D_i = \text{Var}_{X_i} [\mathcal{M}_i(X)] = \text{Var}_{X_i} [\mathbb{E} [\mathcal{M}(X)|X_i = x_i]] \]

They quantify the (additive) effect of each input parameter separately, i.e. the reduction of variance to expect if \( X_i \) is set equal to \( x_i \).

Total Sobol’ indices

\[ S^{T}_{i} \overset{\text{def}}{=} \sum_{i \subseteq \{i_1 \ldots i_s\}} S_{i_1 \ldots i_s} \]

They quantify the total effect of \( X_i \) including the first order effect and the interactions with the other variables.
Consider \( Y = \mathcal{M}(\mathbf{X}) \) where \( \mathbf{X} \sim f_X \) with independent components:

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

- Due to the orthogonality properties of the polynomial chaos basis, one gets:
  \[
  \mathbb{E}[\Psi_\alpha(\mathbf{X})] = 0 \quad \mathbb{E}[\Psi_\alpha(\mathbf{X}) \Psi_\beta(\mathbf{X})] = \delta_{\alpha\beta}
  \]

- Thus the mean and variance:
  \[
  \mathcal{M}_0 = \mathbb{E}[\mathcal{M}(\mathbf{X})] = y_0
  \]
  \[
  D = \text{Var}[\mathcal{M}(\mathbf{X})] = \sum_{\substack{\alpha \in \mathbb{N}^M \\ \alpha \neq 0}} y^2_\alpha
  \]
Interaction sets

Let $A_u$ be the set of multi-indices depending exactly on the subset of variables $u = \{i_1, \ldots, i_s\}$:

$$A_u = \{\alpha \in \mathbb{N}^M : k \in u \iff \alpha_k \neq 0\} \bigcup_{u \subset \{1, \ldots, M\}} A_u = \mathbb{N}^M$$

Sobol’ decomposition

By unicity of the Sobol’ decomposition one gets ($x_u \overset{\text{def}}{=} \{x_{i_1}, \ldots, x_{i_s}\}$):

$$\mathcal{M}(x) = \mathcal{M}_0 + \sum_{u \subset \{1, \ldots, M\}} \mathcal{M}_u(x_u)$$

where:

$$\mathcal{M}_u(x_u) \overset{\text{def}}{=} \sum_{\alpha \in A_u} y_{\alpha} \Psi_{\alpha}(x)$$
Partial variances

The partial variances $D_u \overset{\text{def}}{=} D_{i_1...i_s} = \text{Var}[\mathcal{M}_u(X)]$ are obtained by summing up the square of selected PC coefficients

First order contribution

$$D_i = \sum_{\alpha \in A_i} y^2_{\alpha} \quad A_i = \{ \alpha \in \mathbb{N}^M : \alpha_i > 0, \alpha_j \neq i = 0 \}$$

Higher order contribution

$$D_u = \sum_{\alpha \in A_u} y^2_{\alpha} \quad A_{i_1...i_s} = \{ \alpha \in \mathbb{N}^M : k \in u \iff \alpha_k > 0 \}$$

- The Sobol' indices come after normalization:

$$S_u = \frac{D_u}{D}$$
Sobol’ decomposition from PC expansions

First order indices

\[ S_i = \sum_{\alpha \in A_i} y_\alpha^2 / D \quad A_i = \{ \alpha \in \mathbb{N}^M : \alpha_i > 0, \alpha_j \neq i = 0 \} \]

Higher order indices

\[ S_{i_1, \ldots, i_s} = \sum_{\alpha \in A_{i_1, \ldots, i_s}} y_\alpha^2 / D \quad A_{i_1, \ldots, i_s} = \{ \alpha \in \mathbb{N}^M : k \in \{i_1, \ldots, i_s\} \Leftrightarrow \alpha_j \neq 0 \} \]

Total indices

\[ S_i^T = \sum_{\alpha \in A_i^T} y_\alpha^2 / D \quad A_i^T = \{ \alpha \in \mathbb{N}^M : \alpha_i > 0 \} \]
Example

Computational model

\[ Y = \mathcal{M}(X_1, X_2) \]

Probabilistic model

\[ X_i \sim \mathcal{N}(\mu_i, \sigma_i) \]

Isoprobabilistic transform

\[ X_i = \mu_i + \sigma_i \xi_i \]

Chaos degree

\[ p = 3, \text{ i.e. } P = 10 \text{ terms} \]

### Variance

\[ D = \sum_{j=1}^{9} a_j^2 \]

### Sobol’ indices

\[ S_1 = \left( \frac{a_1^2 + a_3^2 + a_6^2}{D} \right) \]
\[ S_2 = \left( \frac{a_2^2 + a_5^2 + a_9^2}{D} \right) \]
\[ S_{12} = \left( \frac{a_4^2 + a_7^2 + a_8^2}{D} \right) \]
Outline

1. Sparse polynomial chaos scheme
2. Adaptive algorithms for sparse expansions
3. Sensitivity analysis
   - Sobol’ indices
   - Case of dependent inputs: ANCOVA
4. Application examples in sensitivity analysis
Consider $Y = \mathcal{M}(X)$ where $X$ has a joint PDF $f_X$ (i.e. with dependent components).

Assume a functional decomposition exists (e.g. the High Dimensional Model Representation (HDMR), see Rabitz et al.):

$$\mathcal{M}(x) = m_0 + \sum_{i=1}^{M} m_i(x_i) + \sum_{1 \leq i < j \leq M} m_{ij}(x_i, x_j) + \ldots$$

The various terms in the expansion are not necessarily orthogonal anymore with respect to $\mathbb{P}(dx) = f_X(x) \, dx$:

$$\mathbb{E}_X \left[ m_u(X_u) \, m_v(X_v) \right] \neq 0$$

The covariance decomposition is an extension of the Sobol’ decomposition accounting from the cross terms.
Covariance decomposition

\[
\text{Var} [m(X)] = \text{Cov} \left[ m_0 + \sum_{u \subset \{1, \ldots, M\}} m_u(X_u), m(X) \right] = \sum_{u \subset \{1, \ldots, M\}} \text{Cov} [m_u(X_u), m(X)]
\]

ANCOVA sensitivity indices

\[
S_u^{(\text{cov})} = \frac{\text{Cov} [m_u(X_u), m(X)]}{\text{Var} [m(X)]} = \frac{\text{Var} [m_u(X_u)] + \sum_{v \neq u} \text{Cov} [m_u(X_u); m_v(X_v)]}{\text{Var} [m(X)]}
\]

Structural (uncorrelated):

\[
S_u^{(U)} = \frac{\text{Var} [m_u(X_u)]}{\text{Var} [m(X)]}
\]

Correlated:

\[
S_u^{(C)} = \frac{\sum_{v \neq u} \text{Cov} [m_u(X_u); m_v(X_v)]}{\text{Var} [m(X)]}
\]
A functional decomposition \( m(x) \overset{\text{def}}{=} M_A(x) \) is obtained by computing a truncated PC expansion of \( M(Z) \) where \( Z \) is made of independent variables of marginal PDF \( f_{X_i} \)

\[
f_Z(z) = \prod_{i=1}^{M} f_{X_i}(z_i)
\]

The moments of \( Y = M(X) \) are computed by Monte Carlo simulation using a \( n \)-sample drawn according to \( f_X \) (i.e. with correlation):

\[
\bar{y}_A = \frac{1}{n} \sum_{i=1}^{n} M_A(x_i)
\]

\[
\text{Var}[Y_A] = \frac{1}{n-1} \sum_{i=1}^{n} (M_A(x_i) - \bar{y}_A)^2
\]

\( \bar{y}_A \neq y_0 \) since \( \mathbb{E}_X[\Psi_\alpha(X)] \neq \mathbb{E}_Z[\Psi_\alpha(Z)] \)
Monte Carlo estimators of ANCOVA indices

**ANCOVA index**

\[ \overline{y_u} = \frac{1}{n} \sum_{i=1}^{n} M_u(x_u, i) \]

\[ S_u^{(cov)} = \frac{\sum_{i=1}^{n} (M_A(x_i) - \overline{y_A}) (M_u(x_u, i) - \overline{y_u})}{\sum_{i=1}^{n} (M_A(x_i) - \overline{y_A})^2} \]

**Structural contribution**

\[ S_u^{(U)} = \frac{\sum_{i=1}^{n} (M_u(x_u, i) - \overline{y_u})^2}{\sum_{i=1}^{n} (M_A(x_i) - \overline{y_A})^2} \]

**Correlated contribution**

\[ S_u^{(C)} = S_u^{(cov)} - S_u^{(U)} \]
Outline

1. Sparse polynomial chaos scheme
2. Adaptive algorithms for sparse expansions
3. Sensitivity analysis
4. Application examples in sensitivity analysis
   - Ishigami function
   - Morris function
   - Bending beam
Ishigami function

Definition

\[ Y = \sin X_1 + a \sin^2 X_2 + b X_3^4 \sin X_1 \quad a = 7, \ b = 0.1 \]

where \( X_i \sim U[-\pi, \pi] \) are independent uniform random variables

Analytical solution

\[ D = \frac{a^2}{8} + \frac{b \pi^4}{5} + \frac{b^2 \pi^8}{18} + \frac{1}{2} \]

\[ D_1 = \frac{b \pi^4}{5} + \frac{b^2 \pi^8}{50} + \frac{1}{2}, \quad D_2 = \frac{a^2}{8}, \quad D_3 = 0 \]

\[ D_{12} = D_{23} = 0, \quad D_{13} = \frac{8 b^2 \pi^8}{225}, \quad D_{123} = 0 \]
First order Sobol’ indices

\[
\begin{array}{cccc}
\text{Index} & \text{Value} \\
S_1 & 0.3138 \\
S_2 & 0.4424 \\
S_3 & 0 \\
S_{12} & 0 \\
S_{13} & 0.2436 \\
S_{23} & 0 \\
\end{array}
\]
First order Sobol’ indices

Sample size $N$

$S_i - S_i^{(ref)}$

$S_1$ (MCS) $S_1$ (PC)
$S_2$ (MCS) $S_2$ (PC)
$S_{13}$ (MCS) $S_{13}$ (PC)
Total Sobol’ indices

Sample size $N$

$S_{T_i}^T$ (MCS)
$S_{T_i}^T$ (PC)
$S_{T_2}^T$ (MCS)
$S_{T_2}^T$ (PC)
$S_{T_3}^T$ (MCS)
$S_{T_3}^T$ (PC)

B. Sudret (Chair of Risk & Safety)
Total Sobol’ indices – small design and replication

- Sparse polynomial chaos scheme
- Adaptive algorithms for sparse expansions
- Sensitivity analysis
- Application examples

- Ishigami function
- Morris function
- Bending beam

- Total Sobol’ indices – small design and replication

---

**Graphs:**

- **$S_1^T$ (MCS) vs. $S_1^T$ (PC)**
- **$S_2^T$ (MCS) vs. $S_2^T$ (PC)**
- **$S_3^T$ (MCS) vs. $S_3^T$ (PC)**

*Sample size $N$*
Morris function

Definition

\[ Y = \beta_0 + \sum_{i=1}^{20} \beta_i w_i + \sum_{i<j}^{20} \beta_{ij} w_i w_j + \sum_{i<j<l}^{20} \beta_{ijl} w_i w_j w_l + \sum_{i<j<l<s}^{20} \beta_{ijls} w_i w_j w_l w_s \]

where:

\[ w_i = \begin{cases} 
2 \left( \frac{1.1 X_i}{X_i + 0.1} - 0.5 \right) & \text{if } i = 3, 5, 7 \\
2(X_i - 0.5) & \text{otherwise} 
\end{cases} \quad X_i \sim \mathcal{U}(0, 1) \]

and:

\[
\begin{align*}
\beta_i &= 20 \quad \text{for } i = 1, \ldots, 10 \quad \text{and} \quad \beta_i = (-1)^i \quad \text{otherwise} \\
\beta_{ij} &= -15 \quad \text{for } i = 1, \ldots, 6 \quad \text{and} \quad \beta_{ij} = (-1)^{i+j} \quad \text{otherwise} \\
\beta_{ijl} &= -10 \quad \text{for } i = 1, \ldots, 5 \quad \text{and} \quad \beta_{ijl} = 0 \quad \text{otherwise} \\
\beta_{ijls} &= 5 \quad \text{for } i = 1, \ldots, 4 \quad \text{and} \quad \beta_{ijls} = 0 \quad \text{otherwise}
\end{align*}
\]
Morris function
Sensitivity results

- Reference: $N = 440,000$ Monte Carlo simulations + bootstrap
- Adaptive sparse PC: $Q_{tgt}^2 = 0.9$ (resp. 0.99)
Morris function
Sparsity of the model

<table>
<thead>
<tr>
<th>Sensitivity indices</th>
<th>Sparse PCE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Q^2_{0.9}$</td>
</tr>
<tr>
<td>$S_4^T$</td>
<td>0.26</td>
</tr>
<tr>
<td>$S_1^T$</td>
<td>0.25</td>
</tr>
<tr>
<td>$S_2^T$</td>
<td>0.24</td>
</tr>
<tr>
<td>$S_4^T$</td>
<td>0.16</td>
</tr>
<tr>
<td>$S_3^T$</td>
<td>0.10</td>
</tr>
<tr>
<td>$S_2^T$</td>
<td>0.08</td>
</tr>
<tr>
<td>$S_8^T$</td>
<td>0.11</td>
</tr>
<tr>
<td>$S_2^T$</td>
<td>0.10</td>
</tr>
<tr>
<td>$S_6^T$</td>
<td>0.10</td>
</tr>
<tr>
<td>$S_5^T$</td>
<td>0.08</td>
</tr>
<tr>
<td># model evaluations</td>
<td>250</td>
</tr>
<tr>
<td>PC degree</td>
<td>3</td>
</tr>
<tr>
<td>Index of sparsity $IS$</td>
<td>$250/1771 \approx 0.14$</td>
</tr>
</tbody>
</table>

- **Full chaos:**
  \[ \text{card } A^{20,11} = \binom{20+11}{11} = 84,672,315 \]
- **Hyperbolic truncature:**
  \[ \text{card } A^{20,11}_{0.4} = 4,234 \]

LAR: 339 terms (IS2 = 8%)
Outline

1. Sparse polynomial chaos scheme
2. Adaptive algorithms for sparse expansions
3. Sensitivity analysis
4. Application examples in sensitivity analysis
   - Ishigami function
   - Morris function
   - Bending beam
Elastic bending beam

Problem statement

Spatially varying Young’s modulus modelled by a stationary lognormal random field:

\[ \mu_E = 210,000 \text{ MPa}, \ CV_E = 20\% \]

- Input random field:

\[ E(x, \omega) = \exp[\lambda_E + \zeta_E g(x, \omega)] \quad g(x, \omega) \sim \mathcal{N}(0, 1) \]

\[ \text{Cov}[g(x) g(x')] = e^{-|x' - x|/\ell} \quad (\ell = 0.5 \text{ m} = L/6) \]

- Quantity of interest: maximal deflection at midspan
Analytical solution of the Fredholm integral equation

100 terms retained for a discretization accuracy \( \sim 1\% \)
Beam deflection

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>LAR - $\varepsilon_{tgt} = 0.01$</th>
<th>LAR - $\varepsilon_{tgt} = 0.001$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (mm)</td>
<td>2.83</td>
<td>2.81</td>
<td>2.83</td>
</tr>
<tr>
<td>Standard Deviation (mm)</td>
<td>0.37</td>
<td>0.36</td>
<td>0.36</td>
</tr>
<tr>
<td>Skewness</td>
<td>[0.38 ; 0.52] †</td>
<td>0.30</td>
<td>0.41</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>[3.12 ; 3.70] †</td>
<td>3.09</td>
<td>3.30</td>
</tr>
<tr>
<td>Number of FE runs</td>
<td>10,000</td>
<td>200</td>
<td>1,200</td>
</tr>
<tr>
<td>Error estimate</td>
<td>$10^{-3}$</td>
<td>$4 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>PC degree</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Number of PC terms</td>
<td>10</td>
<td>246</td>
<td></td>
</tr>
<tr>
<td>Index of sparsity $IS_1$</td>
<td>$2 \times 10^{-3}$</td>
<td>$3 \times 10^{-6}$</td>
<td></td>
</tr>
<tr>
<td>Index of sparsity $IS_2$</td>
<td>7%</td>
<td>4%</td>
<td></td>
</tr>
</tbody>
</table>

- PC expansion of max. degree $p = 6$ in 100 dimensions:
  $P = 1,705,904,746$ terms
- Hyperbolic truncature: 5,118 terms
- LAR: 246 terms
Only the modes #1 and #3 impact the maximal deflection (modes #2 #4 are antisymmetric).

Another analysis using only 3 terms in the KL expansion (and $N = 40$ finite element runs) provide the same second moments within 1% accuracy ($L_2$-error of $2.10^{-4}$)
Conclusions

- Polynomial chaos expansions suffer from the **curse of dimensionality** when classical Galerkin/ stochastic collocation techniques are used.

- Hyperbolic truncation schemes allow one to select **a priori** approximation spaces of lower dimensionality that are consistent with the **sparsity-of-effect** principle.

- **Sparse expansions** may be obtained by casting the problem of computing the PC coefficients in terms of regression analysis and variable selection (**Least Angle Regression**).

- Although stochastic problems (**e.g.** involving random fields) may appear as large dimensional, the **effective dimension** for quantities of interest is usually much smaller.
Questions?

Thank you very much for your attention!

UQLab...

... The Uncertainty Quantification Toolbox in Matlab

http://www.rsuq.ethz.ch