Metamodels : polynomial chaos expansion and Gaussian process modelling

Chu Mai, Géraud Blatman, EDF R&D

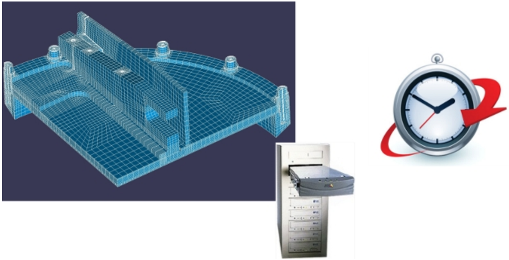
HPC and Uncertainty Treatment
Examples with Open TURNS and URANIE

Airbus Group – CEA – EDF – IMACS - Phimeca

PRACE Advanced Training Center
May 2018

Inaffordable engineering analyses for real world problems

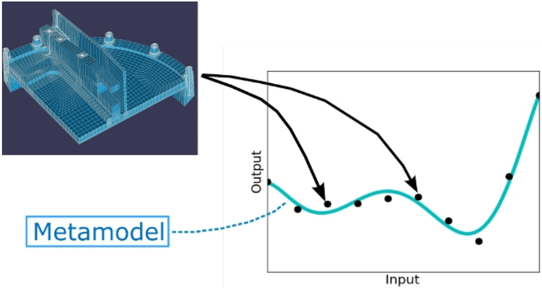
For complex models, a single simulation may take several hours or days to complete.



This makes tasks as optimization, design space exploration, uncertainty propagation and sensitivity analysis impossible in practice.

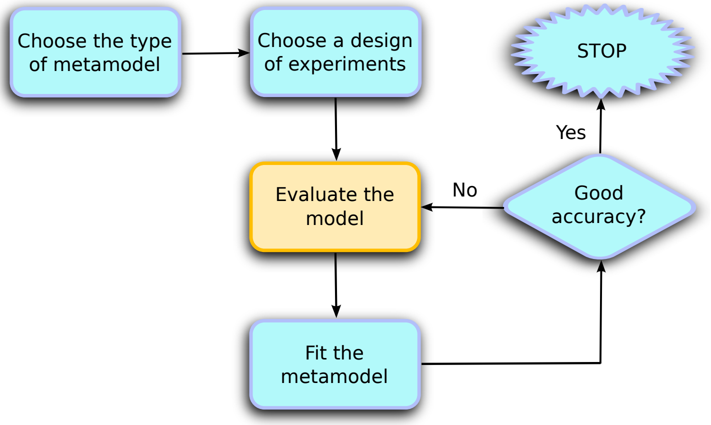
A remedy named metamodel

Construct approximation models known as metamodels that mimic the behaviour of the simulator while being cheaper to evaluate.



Then use the surrogate instead of the simulator in order to perform computationally expensive analyses.

The major steps for constructing a metamodel



Outline

Outline

1. Polynomial chaos expansion
   - Unidimensional case with a uniform variable
   - Unidimensional case with a non uniform variable
   - Multidimensional case
   - Validation and stepwise construction of an optimal chaos approximation
2. A short introduction to Gaussian process modelling
3. Conclusions

Outline

1. Polynomial chaos expansion
   - Unidimensional case with a uniform variable
   - Unidimensional case with a non uniform variable
   - Multidimensional case
   - Validation and stepwise construction of an optimal chaos approximation
2. A short introduction to Gaussian process modelling
3. Conclusions

Instructive example

Model: $Y = f(X) = X \sin X \quad , \quad X \sim \mathcal{U}([0,10])$

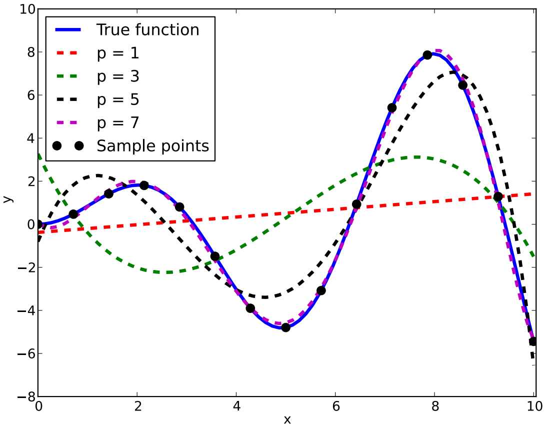
Polynomial metamodel: $\tilde{f}_{\boldsymbol{a}}(X) = \sum_{j=0}^{p} a_j \, X^j$

Coefficients estimation by least squares:

- Design of experiments: n points $x^{(i)}$ evenly distributed in [0,10] ➞ Vecteurs $\mathcal{X} = (x^{(1)}, \ldots, x^{(n)})$ et $\mathcal{Y} = (y^{(1)}, \ldots, y^{(n)})$
- Design matrix: $\boldsymbol{\pi} = \left({x^{(i)}}^j \, , \; i = 1, \ldots, n \; , \; j = 0, \ldots, p \right)$
- Coefficients estimation: $\boxed{\hat{\boldsymbol{a}} = \left(\boldsymbol{\pi}^{\mathsf{T}} \boldsymbol{\pi} \right)^{-1} \boldsymbol{\pi}^{\mathsf{T}} \mathcal{Y}}$

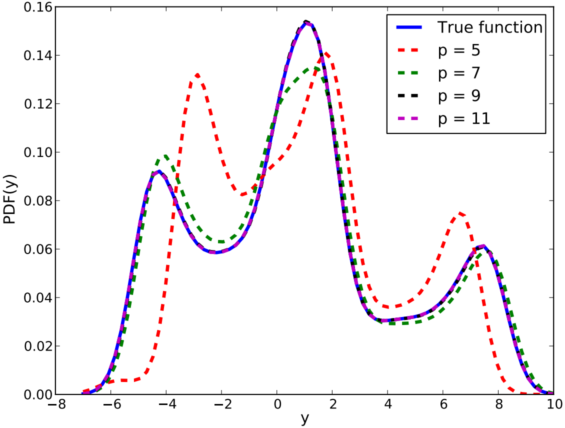
Convergence with respect to the degree p

Design size: $n = 15$



Convergence in terms of probability density function (PDF)

Evaluation of the PDF of each metamodel $\tilde{f}_{\boldsymbol{a}}(X)$ based on $50,000$ realizations of X



Estimation of the moments of Y

Naive strategy: Evaluate $\tilde{f}_{\boldsymbol{a}}(X)$ for a large number of realizations of X and compute the sample moments of the set of results

Drawback: The results are affected by a sampling error

Alternative: Derive analytically the moments by using centered and orthonormal polynomials

Centered and orthonormal polynomials

- Let $(\pi_j)_{j=0,1,2,\dots}$ be a family of polynomials

- The polynomials are said to be centered w.r.t the random variable ξ IFF :

$$\mathbb{E}\left[\pi_i(\xi)\right] \;\equiv\; \int_{\mathcal{D}_\xi} \pi_i(u)\, f_\xi(u)\, du \;=\; 0 \quad \forall i \geq 1$$

where $\mathcal{D}_\xi$ and $f_\xi(u)$ are the support the PDF of ξ

- They are said to be orthonormal w.r.t the random variable ξ IFF :

$$\mathbb{E}\left[\pi_i(\xi)\,\pi_j(\xi)\right] \;\equiv\; \int_{\mathcal{D}_\xi} \pi_i(u)\,\pi_j(u)\, f_\xi(u)\, du \;=\; \begin{cases} 1 \text{ if } i = j \\ 0 \text{ otherwise} \end{cases}$$

Case of a uniform variable – Legendre polynomials

- Let the random variable : $\xi \sim \mathcal{U}([-1,1])$

- The orthonormal polynomials w.r.t ξ are the Legendre polynomials:

$$\pi_0(\xi) = 1 \ , \ \pi_1(\xi) = \sqrt{3}\xi \ , \ \pi_2(\xi) = \frac{\sqrt{5}}{2}(3\xi^2 - 1) \ , \ \ldots$$

$$\boxed{\mathbb{E}\left[\pi_i(\xi)\right] = 0 \ , \ \forall i \geq 1 \qquad \text{et} \qquad \mathbb{E}\left[\pi_i(\xi)\,\pi_j(\xi)\right] = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ otherwise} \end{cases}}$$

Instructive example – Legendre approximation

- Model: $Y \;=\; f(X) \;=\; X \sin X \quad , \quad X \sim \mathcal{U}([0, 10])$
- Change of variable: $\xi \equiv \dfrac{X}{5} - 1 \quad , \quad \xi \sim \mathcal{U}([-1, 1])$
- Metamodel called "polynomial chaos":

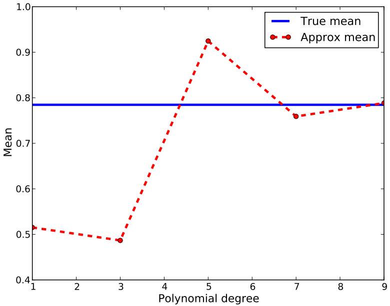
$$\tilde{f}_{\boldsymbol{a}}(X) \;=\; \tilde{h}_{\boldsymbol{a}}(\xi) \;=\; \sum_{j=0}^{p} a_j \, \pi_j(\xi)$$

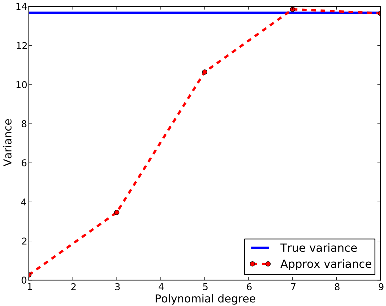
where the π_j's are the Legendre polynomials and the a_j's are the least squares coefficients

- Closed-form formulas of the moments:

$$\boxed{\mathbb{E}\left[\tilde{f}_{\boldsymbol{a}}(X)\right] \;=\; a_0 \quad , \quad \mathbb{V}\mathrm{ar}\left[\tilde{f}_{\boldsymbol{a}}(X)\right] \;=\; \sum_{j=1}^{p} a_j^2}$$

Convergence of the moments





Accurate moment estimates from $p = 7$

Outline

1. Polynomial chaos expansion
   - Unidimensional case with a uniform variable
   - Unidimensional case with a non uniform variable
   - Multidimensional case
   - Validation and stepwise construction of an optimal chaos approximation
2. A short introduction to Gaussian process modelling
3. Conclusions

Instructive example with a Gaussian variable

Model: $Y \;=\; f(X) \;=\; X \sin X \qquad , \qquad X \sim \mathcal{N}(5,1)$

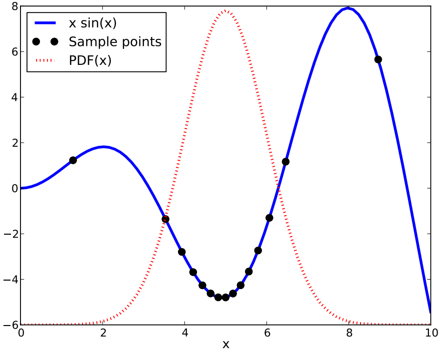
Change of variable : $\xi \equiv X - 5 \quad , \quad \xi \sim \mathcal{N}(0,1)$

Polynomial chaos: $$\tilde{f}_{\boldsymbol{a}}(X) \;=\; \tilde{h}_{\boldsymbol{a}}(\xi) \;=\; \sum_{j=0}^{p} a_j \, \pi_j(\xi)$$

where the π_j's are the Hermite polynomials which are centered and orthonormal w.r.t ξ

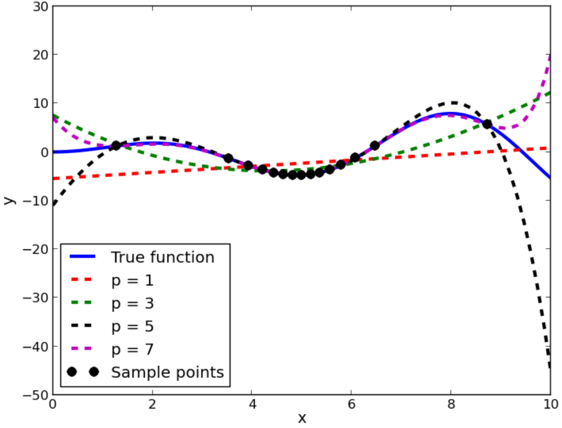
Design of experiments to compute the coefficients

- Sampling of the $x^{(i)}$'s according to the PDF $\mathcal{N}(5, 1)$

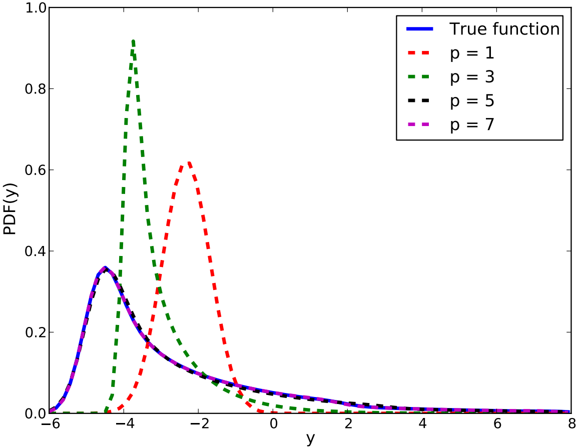


- Change of variable for each realization: $\xi^{(i)} = x^{(i)} - 5$

Approximation of the output



Approximation of the output PDF



Classical families of polynomials

Sudret, HDR(2008)

PDF of ξ	Support	Polynomials
Normal (Gaussian)	$\mathbb{R}$	Hermite
Uniform	$[-1, 1]$	Legendre
Gamma	$(0, \infty)$	Laguerre
Chebyshev	$(-1, 1)$	Chebyshev
Beta	$(-1, 1)$	Jacobi

Change of variable

General formula $$\boxed{\xi \;=\; F_\xi^{-1}\left[F_X(X)\right]}$$

where F_ξ and F_X are the (known) cumulative distribution functions of ξ and X

Guidelines for choosing the couple (ξ, polynomials)

- If the type of the X PDF belongs to the table of classical families, select the ξ with the same PDF (e.g. $\mathcal{N}(210, 18) \rightarrow \mathcal{N}(0, 1)$)
- Otherwise, select a ξ with the same type of support (infinite, semi-bounded or bounded) as X

Outline

1. Polynomial chaos expansion
   - Unidimensional case with a uniform variable
   - Unidimensional case with a non uniform variable
   - Multidimensional case
   - Validation and stepwise construction of an optimal chaos approximation
2. A short introduction to Gaussian process modelling
3. Conclusions

Instructive example

Model:

$$Y = f(X_1, X_2) = X_1 \sin X_2$$

$$X_1 \sim \mathcal{U}([0, 10]) \quad , \quad X_2 \sim \mathcal{LN}(5, 1)$$

Changes of variable :

- $\xi_1 = X_1/5 - 1 \quad , \quad \xi_1 \sim \mathcal{U}([-1, 1])$ → Legendre polynomials
- $\xi_2 = \exp(X_2) - 5 \quad , \quad \xi_2 \sim \mathcal{N}(0, 1)$ → Hermite polynomials

Construction of a bivariate orthonormal basis

- Legendre polynomials:

$$\pi_0^{(1)}(\xi_1) = 1 \quad , \quad \pi_1^{(1)}(\xi_1) = \sqrt{3}\xi_1 \quad , \quad \pi_2^{(1)}(x) = \frac{\sqrt{5}}{2}(3\xi_1^2 - 1) \quad , \quad \ldots$$

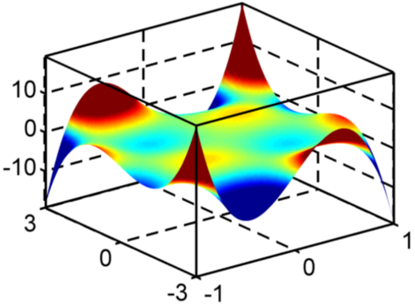
- Hermite polynomials:

$$\pi_0^{(2)}(\xi_2) = 1 \quad , \quad \pi_1^{(2)}(\xi_1) = \xi_2 \quad , \quad \pi_2^{(2)}(\xi_2) = \frac{\sqrt{2}}{2}\left(\xi_2^2 - 1\right) \quad , \quad \ldots$$

- Construction of bivariate polynomials:

$$\begin{aligned}
\psi_{0,0}(\xi_1,\xi_2) &= \pi_0^{(1)}(\xi_1) \times \pi_0^{(2)}(\xi_2) = 1 \\
\psi_{1,0}(\xi_1,\xi_2) &= \pi_1^{(1)}(\xi_1) \times \pi_0^{(2)}(\xi_2) = \sqrt{3}\xi_1 \\
\psi_{1,2}(\xi_1,\xi_2) &= \pi_1^{(1)}(\xi_1) \times \pi_2^{(2)}(\xi_2) = \tfrac{\sqrt{6}}{2}\xi_1(\xi_2^2 - 1)
\end{aligned}$$

Construction of a bivariate orthonormal basis



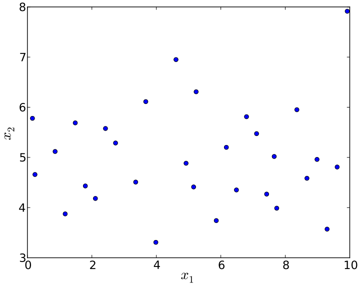
$\psi_{3,3}(\xi_1, \xi_2)$

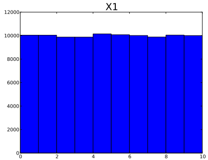
Fitting of a polynomial chaos approximation

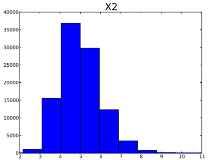
"Hybrid" Legendre-Hermite polynomial chaos:

$$\tilde{f}_{\boldsymbol{a}}(X_1, X_2) \;=\; \tilde{h}_{\boldsymbol{a}}(\xi_1, \xi_2) \;=\; \sum_{i_1+i_2 \leq p} a_{i_1,i_2}\; \psi_{i_1,i_2}(\xi_1, \xi_2)$$

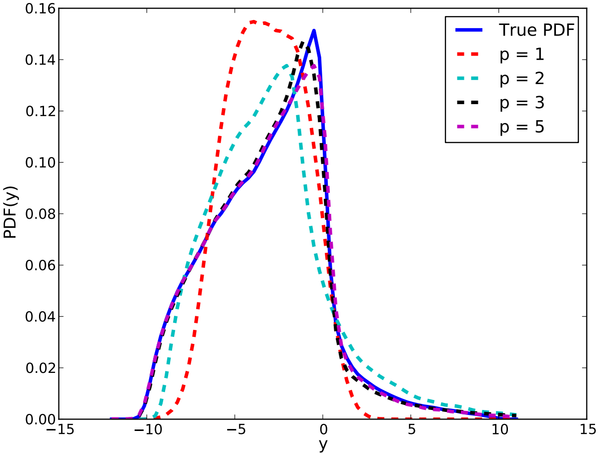
Quasi-random design of experiments ($n = 30$):







Convergence of the output PDF



Multivariate polynomial chaos

- Consider the model:

$$Y = f(\boldsymbol{X}) \quad , \quad \mathrm{dim}(\boldsymbol{X}) = M \quad , \quad \mathrm{dim}(Y) = 1$$

- Assumption: independence of the components of $\boldsymbol{X}$

- Approximation by tensorizing the 1D scheme:

$$\tilde{f}_{\boldsymbol{a}}(\boldsymbol{X}) \quad = \quad \sum_{i_1 + \cdots + i_M \leq p} a_{i_1,\ldots,i_M} \, \pi_{i_1}^{(1)}(X_1) \times \cdots \times \pi_{i_M}^{(M)}(X_M)$$

The centering and orthonormality properties still hold

$\rightarrow$ Closed-form formulas of the moments

and the Sobol' sensitivity indices

Sobol' sensitivity indices

Sobol' sensitivity indices: Quantify the part of the variance of Y that is due to the variance of each variable X_i

Definition: $$S[X_i] \;=\; \frac{\mathbb{V}\text{ar}\left[\mathbb{E}\left[Y|X_i\right]\right]}{\mathbb{V}\text{ar}[Y]}$$

Direct estimation from model evaluations can reveal inaffordable!

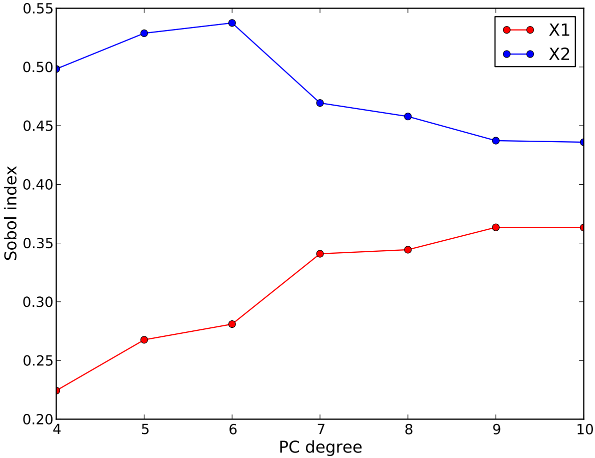
Estimation of the numerator from the chaos proxy :

$$\boxed{\mathbb{V}\text{ar}\left[\mathbb{E}\left[Y|X_i\right]\right] \quad \approx \quad \sum_{\boldsymbol{j}\in\mathcal{I}_i} a_j^2}$$

where $\mathcal{I}_i$ is set of the "multi-indices" $\boldsymbol{j}$ such that $j_k = 0$ if $k \neq i$

Sensitivities to interactions can also be derived!

Instructive example – Convergence of the Sobol' indices

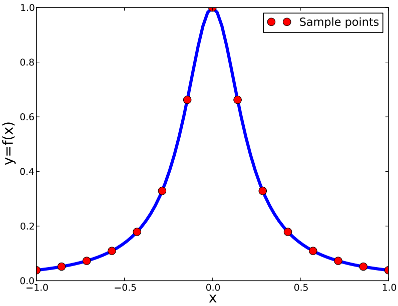


Outline

Runge function

Model: $$Y = f(X) = \frac{1}{1+25X^2} \quad , \quad X \equiv \xi \sim \mathcal{U}([-1,1])$$

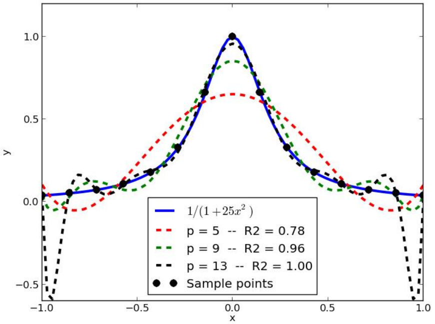
Design of experiments: $n = 15$ points $x^{(i)}$ evenly located over $[-1,1]$



Legendre chaos approximation : $$Y \approx \tilde{f}_{\boldsymbol{a}}(\xi) = \sum_{j=0}^{p} a_j \, \pi_j(\xi)$$

The R^2 accuracy estimator

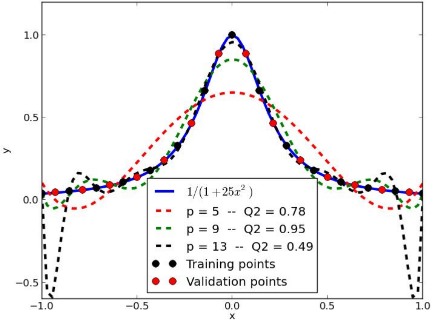
$$R^2 = 1 - Err \quad , \quad Err \propto \sum_{i=1}^{n} \left(f(\xi^{(i)}) - \tilde{f}_{\boldsymbol{a}}(\xi^{(i)}) \right)^2$$



Problem : R^2 increases with the error (overfitting) → Not reliable estimator

Alternative: use of an independent validation set

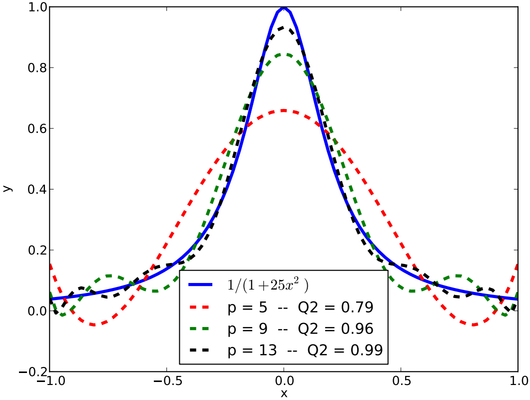
- Sampling of m new points $\{x'^{(1)}, \ldots, x'^{(m)}\}$
- Calculation of the R^2 counterpart from these points $\rightarrow$ Coefficient Q^2



Q^2 "detects" overfitting $\rightarrow$ More reliable estimator

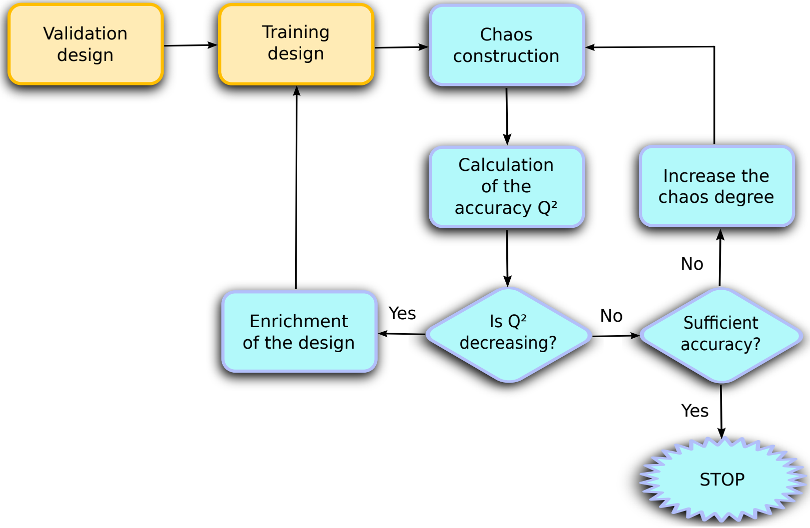
Improving the approximation by adding design points

Addition of 15 new points in the design of experiments



Adaptive construction of an “optimal” polynomial chaos

Blatman 2009



Some remarks

– There exists other methods to estimate the coefficients (numerical integration)

– The design size $n <<$ the number of chaos coefficients: risk of overfitting

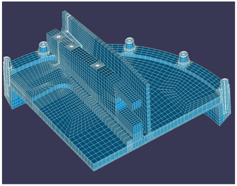
→ Obtaining an accurate approximation in high dimensions (≥ 10 input parameters) is uneasy

Solution:

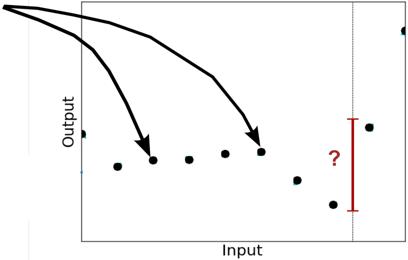
sparse adaptive PCE (Least Angle Regression) Blatman & Sudret 2011

Outline

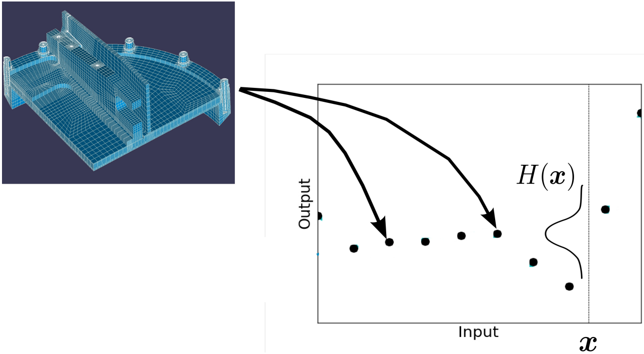
Prediction at a new point (1)



Output value at a new location?



Prediction at a new point (2)



Assumption: The response is a realization of a Gaussian random variable whose moments depend on the design points

Gaussian process assumption

The model output is a realization of a Gaussian random process of the form :

$$H(\boldsymbol{x}, \omega) \quad = \boxed{\boldsymbol{r}(\boldsymbol{x}) \cdot \boldsymbol{\beta}} + \boxed{Z(\boldsymbol{x}, \omega)}$$

Trend (deterministic)

Linear regression
on a fixed basis

Random fluctuations

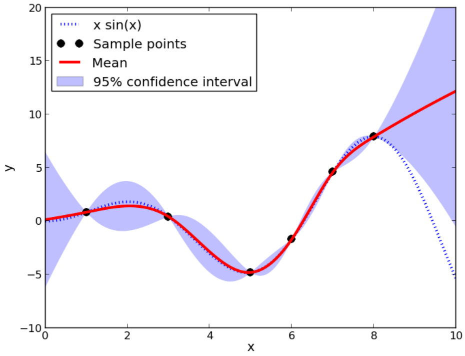
Zero-mean and
stationary
Gaussian process

$$\mathbb{C}\mathrm{ov}_Z(\boldsymbol{x}, \boldsymbol{x'}) \; = \; \sigma^2 \rho(\|\boldsymbol{x} - \boldsymbol{x'}\|)$$

Kriging

Instructive example — $y = f(x) = x\sin(x)$

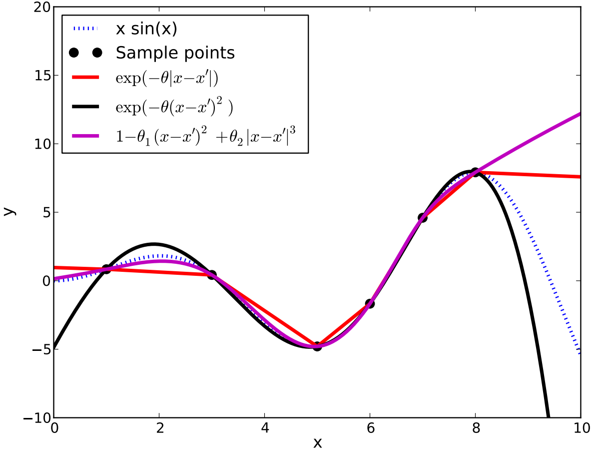
$$H(x,\omega) \;=\; \boldsymbol{r}(x)\cdot\boldsymbol{\beta} \;+\; Z(x,\omega) \qquad , \qquad \mathbb{C}\mathrm{ov}_Z(x,x') \;=\; \sigma^2 e^{-\theta(x-x')^2}$$



– The mean is used as a metamodel (interpolator)

– The prediction bands are used as error indicators

Influence of the autocorrelation function



The autocorrelation function type has an effect on the metamodel

Expressions of the mean and the variance

Notations: $$\boldsymbol{k}(\boldsymbol{x}) \equiv \left\{\rho\left(\boldsymbol{x}, \boldsymbol{x}^{(1)}\right), \ldots, \rho\left(\boldsymbol{x}, \boldsymbol{x}^{(N)}\right)\right\}^{\mathsf{T}}$$

$$\boldsymbol{R} \equiv \left(r_j(\boldsymbol{x}^{(i)})\right)_{1 \leq i,j \leq N} \quad , \quad \boldsymbol{K} \equiv \left(\rho(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})\right)_{1 \leq i,j \leq N}$$

Mean: $$\boxed{\mu(\boldsymbol{x}) = \boldsymbol{r}^{\mathsf{T}}(\boldsymbol{x})\boldsymbol{\beta} + \boldsymbol{k}^{\mathsf{T}}(\boldsymbol{x})\boldsymbol{K}^{-1}(\mathcal{Y} - \boldsymbol{R}\boldsymbol{\beta})}$$

Variance: $$\boxed{\sigma^2(\boldsymbol{x}) = \sigma^2 - \boldsymbol{k}^{\mathsf{T}}(\boldsymbol{x})\, \boldsymbol{K}^{-1}\, \boldsymbol{k}^{\mathsf{T}}(\boldsymbol{x})}$$

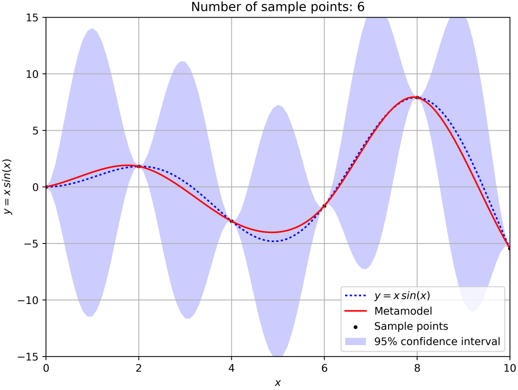
Parameter fitting

- To apply the previous formulas, the parameters $(\boldsymbol{\beta}, \sigma, \boldsymbol{\theta})$ have to be estimated from the design points

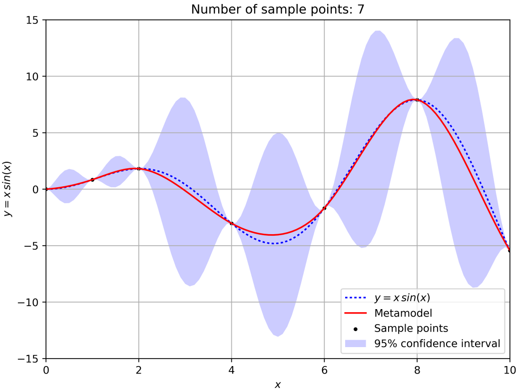
→ Optimal correlation parameter $\hat{\boldsymbol{\theta}}$ estimated by the maximum likelihood estimate (Marrel et al. 2008) or cross validation (Bachoc 2013)

→ Parameters $(\hat{\boldsymbol{\beta}}, \hat{\sigma})$ estimated by empirical best linear unbiased estimator (BLUE) (Santner et al. 2003)

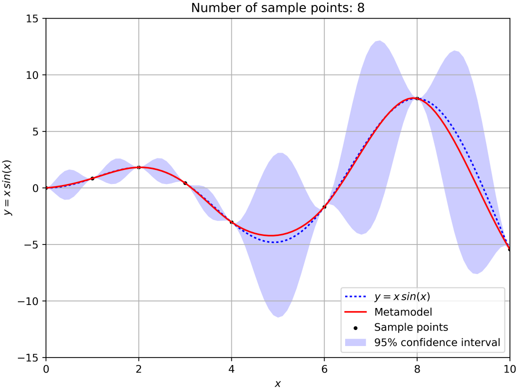
Approximation of $f(x) = x\sin(x)$ – Convergence



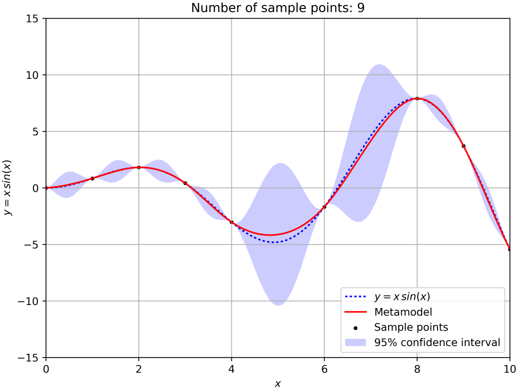
Approximation of $f(x) = x\sin(x)$ – Convergence



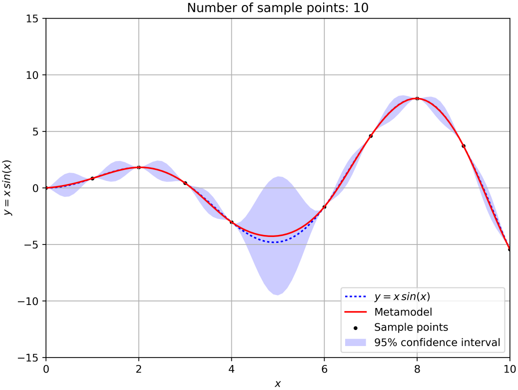
Approximation of $f(x) = x \sin(x)$ – Convergence



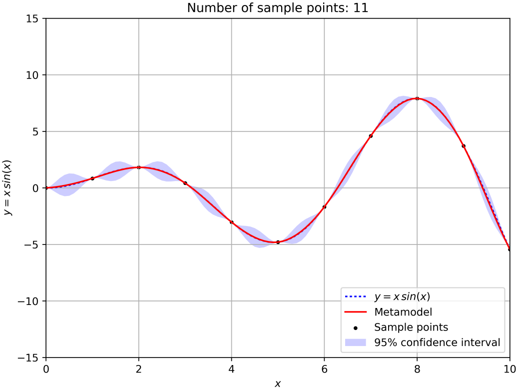
Approximation of $f(x) = x\sin(x)$ – Convergence



Approximation of $f(x) = x\sin(x)$ – Convergence



Approximation of $f(x) = x\sin(x)$ – Convergence



Outline

1. Polynomial chaos expansion
2. A short introduction to Gaussian process modelling
3. Conclusions

Polynomial chaos or Gaussian process?

	Polynomial chaos	**Gaussian process**
Pros	Closed-form formulas of the moments and the sensitivities	Error indicators (prediction intervals)
	Algorithmic simplicity	Possibility of refining optimally the design of experiments
Cons	No "local" error indicator	Uneasy parameter estimation for dimensions ≥ 15

More references:

- Couple Polynomial chaos and Kriging: Schöbi & Sudret 2014
- Confront Polynomial chaos and Kriging: Le Gratiet et al. 2016

Some computer tools (non exhaustive list!)

Polynomial chaos	Gaussian process
OpenTURNS (Python)	
UQLab (Matlab)	
URANIE	
chaospy (Python)	Scikits (Python)
UQToolbox (Python)	DACE (Matlab)
NISP (Scilab)	STK (Matlab)
	DiceKriging (R)

A few messages...

- Polynomial chaos applied to several industrial studies (civil engineering, non destructive testing, etc.)
- For a given problem, ideally test several types of metamodels
- Check that the designs are consistent with the input parameter PDFs!
- Assess the metamodel accuracy (validation) prior to computing quantities of interest!
- Metamodelling may not work for highly nonlinear models

Thank you for your attention!