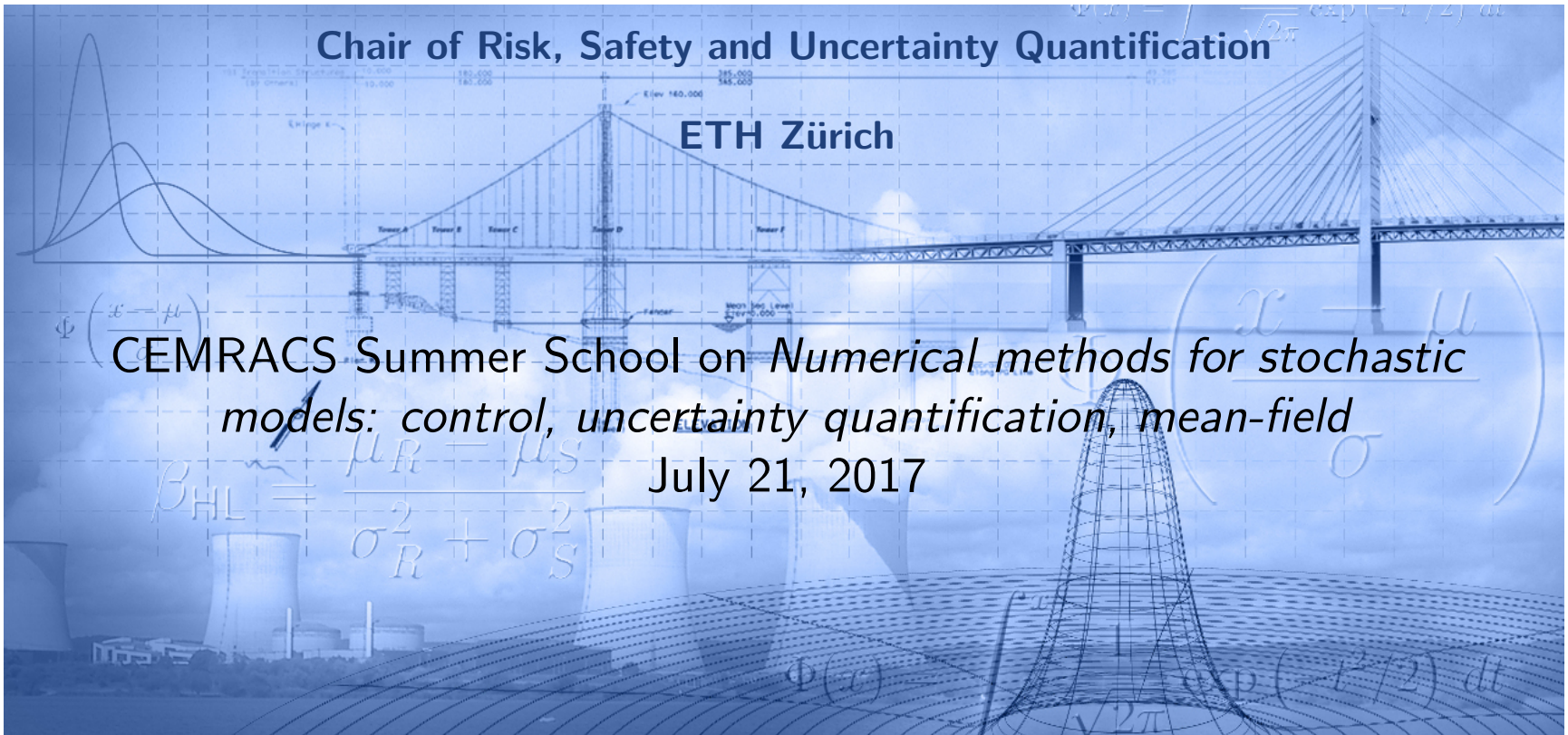
Metamodels in Uncertainty Quantification and Reliability Analysis

S. Marelli and B. Sudret



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

Chair Leader: Prof. Bruno Sudret

Research topics

- Uncertainty modelling for engineering systems
- Structural reliability analysis
- Metamodels (polynomial chaos expansions, Kriging, support vector machines)
- Bayesian model calibration and stochastic inverse problems
- Global sensitivity analysis
- Reliability-based design optimization

`http://www.rsuq.ethz.ch`

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

Chair Leader: Prof. Bruno Sudret

Research topics

- Uncertainty modelling for engineering systems
- **Structural reliability analysis**
- **Metamodels (aka surrogate models)**
  (polynomial chaos expansions, Kriging,
  low-rank tensor approximations, support vector machines)
- Bayesian model calibration and stochastic inverse problems
- Global sensitivity analysis
- Reliability-based design optimization

`http://www.rsuq.ethz.ch`

Credits & acknowledgements

This lecture is largely based on the contents of the following Master- and PhD-level courses offered by the Chair of Risk, Safety and Uncertainty Quantification:

- **Uncertainty Quantification in Engineering**
  Master Course at ETH Zürich
  (B. Sudret and S. Marelli) www.rsuq.ethz.ch/teaching/uncertainty-quantification.html
- **Structural Reliability and Risk Analysis**
  Master Course at ETH Zürich
  (B. Sudret and S. Marelli) www.rsuq.ethz.ch/teaching/structural-reliability.html
- **Uncertainty Quantification and Data Analysis in Applied Sciences**
  PhD Block Course at Computational Science Zürich (first block: Uncertainty Quantification and Reliability Analysis)
  (B. Sudret and S. Marelli) www.zhcs.ch/education/block-course-1/

Outline

Outline

Computational models in engineering

Complex engineering systems are designed and assessed using computational models, a.k.a simulators

A computational model combines:

- A mathematical description of the physical phenomena (governing equations), *e.g.* mechanics, electromagnetism, fluid dynamics, etc.

$$\nabla \cdot \mathbf{D} = \rho$$
$$\nabla \cdot \mathbf{B} = 0$$
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$
$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$$

Computational models in engineering

Complex engineering systems are designed and assessed using computational models, a.k.a simulators

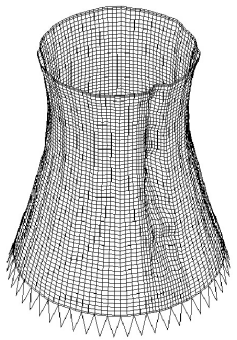
A computational model combines:

- A mathematical description of the physical phenomena (governing equations), *e.g.* mechanics, electromagnetism, fluid dynamics, etc.

$$\nabla \cdot \mathbf{D} = \rho$$
$$\nabla \cdot \mathbf{B} = 0$$
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$
$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$$

- Discretization techniques which transform continuous equations into linear algebra problems

- Algorithms to solve the discretized equations



Computational models in engineering

Computational models are used:

- Together with experimental data for calibration purposes
- To explore the design space ("virtual prototypes")
- To optimize the system (*e.g.* minimize the mass) under performance constraints
- To assess its robustness and its reliability *w.r.t.* uncertainty

Remarks:

- Engineering models are usually very expensive: $\mathcal{O}(1 - 20 \text{ hrs/run})$ even with HPC facilities
- They are often proprietary codes/workflows, hence **black-boxes**

Computational models in engineering

Computational models are used:

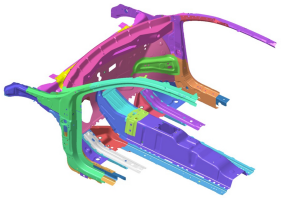
- Together with experimental data for calibration purposes
- To explore the design space ("virtual prototypes")
- To optimize the system (*e.g.* minimize the mass) under performance constraints
- To assess its robustness and its reliability *w.r.t.* uncertainty

Remarks:

- Engineering models are usually very expensive: $\mathcal{O}(1 - 20 \text{ hrs/run})$ even with HPC facilities
- They are often proprietary codes/workflows, hence **black-boxes**

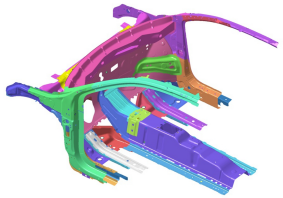
Real world is uncertain

- Differences between the designed and the real system:
  - Dimensions (tolerances in manufacturing)
  - Material properties (*e.g.* variability of the stiffness or resistance)

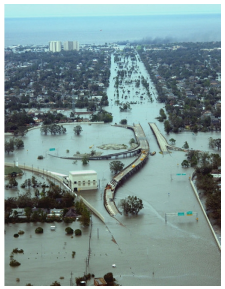


Real world is uncertain

- Differences between the designed and the real system:
  - Dimensions (tolerances in manufacturing)
  - Material properties (*e.g.* variability of the stiffness or resistance)



- Unforecast exposures: exceptional service loads, natural hazards (earthquakes, floods, landslides), climate loads (hurricanes, snow storms, etc.), accidental/malevolent human actions (explosions, fire, etc.)



Global framework for managing uncertainties

Probabilistic Input Model

Quantification of sources of uncertainty

Physical Model

Model(s) of the system

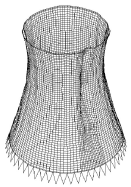
Assessment criteria

Uncertainty Analysis

Uncertainty propagation

Random variables

Computational model



Moments

Probability of failure

Response PDF

Iteration

Sensitivity analysis

Bayesian inversion

Sudret, B. (2007). Uncertainty propagation and sensitivity analysis in mechanical models - Contributions to structural reliability and stochastic spectral methods. Habilitation à diriger des recherches, Université Blaise Pascal, Clermont-Ferrand

Global framework for managing uncertainties

Probabilistic Input Model

Quantification of sources of uncertainty

Physical Model

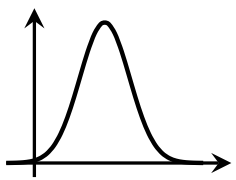
Model(s) of the system

Assessment criteria

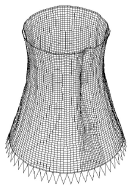
Uncertainty Analysis

Uncertainty propagation

Random variables



Computational model



Moments

Probability of failure

Response PDF

Iteration

Sensitivity analysis

Bayesian inversion

Sudret, B. (2007). Uncertainty propagation and sensitivity analysis in mechanical models - Contributions to structural reliability and stochastic spectral methods. Habilitation à diriger des recherches, Université Blaise Pascal, Clermont-Ferrand

Global framework for managing uncertainties

Probabilistic Input Model

Quantification of sources of uncertainty

Physical Model

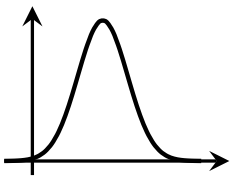
Model(s) of the system

Assessment criteria

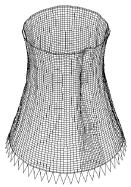
Uncertainty Analysis

Uncertainty propagation

Random variables



Computational model



Moments

Probability of failure

Response PDF

Iteration

Sensitivity analysis

Bayesian inversion

Sudret, B. (2007). Uncertainty propagation and sensitivity analysis in mechanical models - Contributions to structural reliability and stochastic spectral methods. Habilitation à diriger des recherches, Université Blaise Pascal, Clermont-Ferrand

Global framework for managing uncertainties

Probabilistic Input Model

Quantification of sources of uncertainty

Physical Model

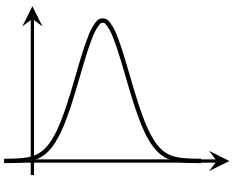
Model(s) of the system

Assessment criteria

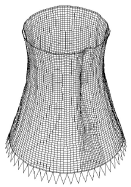
Uncertainty Analysis

Uncertainty propagation

Random variables



Computational model



Moments

Probability of failure

Response PDF

Iteration

Sensitivity analysis

Bayesian inversion

Sudret, B. (2007). Uncertainty propagation and sensitivity analysis in mechanical models - Contributions to structural reliability and stochastic spectral methods. Habilitation à diriger des recherches, Université Blaise Pascal, Clermont-Ferrand

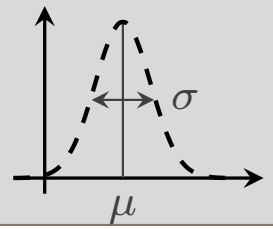
Uncertainty propagation

Goal: given an input random vector $\boldsymbol{X} \sim f_{\boldsymbol{X}}$, estimate the uncertainty/variability of the quantities of interest (QoI) $Y = \mathcal{M}(\boldsymbol{X})$ due to the input uncertainty $f_{\boldsymbol{X}}$

- Output statistics, *i.e.* mean, standard deviation, etc.

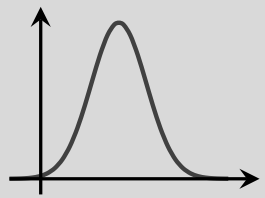
$$\mu_Y = \mathbb{E}_{\boldsymbol{X}}\left[\mathcal{M}(\boldsymbol{X})\right]$$
$$\sigma_Y^2 = \mathbb{E}_{\boldsymbol{X}}\left[\left(\mathcal{M}(\boldsymbol{X}) - \mu_Y\right)^2\right]$$

Mean/std. deviation



- Distribution of the QoI

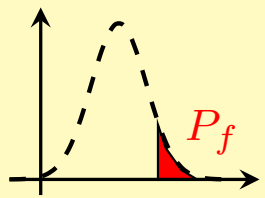
Response PDF



- Probability of exceeding an admissible threshold y_{adm}

$$P_f = \mathbb{P}\left(Y \geq y_{adm}\right)$$

Probability of failure



Monte Carlo simulation

Methodology

- The input random vector $\boldsymbol{X}$ is sampled according to its prescribed joint PDF $f_{\boldsymbol{X}}(\boldsymbol{x})$
- For each sample point $\boldsymbol{x}^{(i)}$, the model response is evaluated, say $y^{(i)} = \mathcal{M}(\boldsymbol{x}^{(i)})$
- The sample set of response quantities $\mathcal{Y} = \{\mathcal{M}(\boldsymbol{x}^{(i)}), i = 1, \ldots, N\}$ is processed, *e.g.*:
  - Moments analysis
  - PDF estimation with kernel smoothing
  - Descriptive statistics

Main drawback: Monte Carlo simulation requires a large number of samples N to achieve proper convergence (*i.e.* typically $N_{MC} \sim 10^{4-6}$)

Monte Carlo simulation

Methodology

- The input random vector $\boldsymbol{X}$ is sampled according to its prescribed joint PDF $f_{\boldsymbol{X}}(\boldsymbol{x})$
- For each sample point $\boldsymbol{x}^{(i)}$, the model response is evaluated, say $y^{(i)} = \mathcal{M}(\boldsymbol{x}^{(i)})$
- The sample set of response quantities $\mathcal{Y} = \{\mathcal{M}(\boldsymbol{x}^{(i)}), i = 1, \ldots, N\}$ is processed, *e.g.*:
  - Moments analysis
  - PDF estimation with kernel smoothing
  - Descriptive statistics

Main drawback: Monte Carlo simulation requires a large number of samples N to achieve proper convergence (*i.e.* typically $N_{MC} \sim 10^{4-6}$)

Meta models

Definition

- A metamodel is an **inexpensive to evaluate analytical function** that accurately approximates a computational model
- It is built from a small sample of point-wise model evaluations (black-box), the *experimental design (ED)*:

$$\mathcal{X} = \left\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(N_{ED})}\right\}, \quad \mathcal{Y} = \left\{\mathcal{M}(\boldsymbol{x}^{(1)}), ..., \mathcal{M}(\boldsymbol{x}^{(N_{ED})})\right\}$$

Selected metamodelling techniques

Polynomial chaos expansions (PCE):

$$\mathcal{M}^{PC}(\boldsymbol{X}) = \sum_{j=0}^{\infty} a_j \boldsymbol{\Psi}_j(\boldsymbol{X})$$

Gaussian process modelling (Kriging):

$$\mathcal{M}^{GP}(\boldsymbol{X}) = \beta^{\mathsf{T}} \boldsymbol{F}(\boldsymbol{X}) + \sigma^2 Z(\boldsymbol{X}, \omega)$$

Meta models

Definition

- A metamodel is an **inexpensive to evaluate analytical function** that accurately approximates a computational model
- It is built from a small sample of point-wise model evaluations (black-box), the *experimental design (ED)*:

$$\mathcal{X} = \left\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(N_{ED})}\right\}, \quad \mathcal{Y} = \left\{\mathcal{M}(\boldsymbol{x}^{(1)}), ..., \mathcal{M}(\boldsymbol{x}^{(N_{ED})})\right\}$$

Selected metamodelling techniques

Polynomial chaos expansions (PCE):

$$\mathcal{M}^{PC}(\boldsymbol{X}) = \sum_{j=0}^{\infty} a_j \boldsymbol{\Psi}_j(\boldsymbol{X})$$

Gaussian process modelling (Kriging):

$$\mathcal{M}^{GP}(\boldsymbol{X}) = \beta^{\mathsf{T}} \boldsymbol{F}(\boldsymbol{X}) + \sigma^2 Z(\boldsymbol{X}, \omega)$$

Metamodels for Uncertainty Propagation

Metamodels as substitutes (surrogates)

- **Sample** an experimental design in the input domain $\Omega_{\boldsymbol{X}}$:

$$\mathcal{X} = \left\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(N_{ED})}\right\}, \quad \mathcal{Y} = \left\{\mathcal{M}(\boldsymbol{x}^{(1)}), ..., \mathcal{M}(\boldsymbol{x}^{(N_{ED})})\right\}$$

- **Calibrate** a metamodel such that $\tilde{\mathcal{M}}(\boldsymbol{X}) \approx \mathcal{M}(\boldsymbol{X})$
- **Substitute** the model $\mathcal{M}(\boldsymbol{X})$ with its surrogate $\tilde{\mathcal{M}}(\boldsymbol{X})$ and perform the MCS analysis

The principle

- MCS with a metamodel is inexpensive ($\sim 10^6$ runs $\cdot\, \mathrm{s}^{-1}$ per core)
- The computational cost of MCS is **traded** for the cost of training the surrogate: $N_{ED} \ll N_{MC}$

Metamodels for Uncertainty Propagation

Metamodels as substitutes (surrogates)

- **Sample** an experimental design in the input domain $\Omega_{\boldsymbol{X}}$:

$$\mathcal{X} = \left\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(N_{ED})}\right\}, \quad \mathcal{Y} = \left\{\mathcal{M}(\boldsymbol{x}^{(1)}), ..., \mathcal{M}(\boldsymbol{x}^{(N_{ED})})\right\}$$

- **Calibrate** a metamodel such that $\tilde{\mathcal{M}}(\boldsymbol{X}) \approx \mathcal{M}(\boldsymbol{X})$
- **Substitute** the model $\mathcal{M}(\boldsymbol{X})$ with its surrogate $\tilde{\mathcal{M}}(\boldsymbol{X})$ and perform the MCS analysis

The principle

- MCS with a metamodel is inexpensive ($\sim 10^6$ runs $\cdot\, \mathrm{s}^{-1}$ per core)
- The computational cost of MCS is **traded** for the cost of training the surrogate: $N_{ED} \ll N_{MC}$

Outline

[Very] Short introduction to Gaussian processes

Gaussian processes in a nutshell

Consider a probability space $(\Omega_Z, \mathcal{F}_Z, \mathbb{P}_Z)$ and $\boldsymbol{x} \in \mathbb{R}^M$. A stochastic process $Z(\boldsymbol{x})$ is Gaussian *i.i.f.* for any finite set $C \in \mathbb{R}^M$ the collection of random variables $Z(C)$ has a Gaussian joint distribution

Notes on Gaussian processes

- A Gaussian process is entirely defined by its mean and covariance functions:

$$\mu(\boldsymbol{x}) = \mathbb{E}\left[Z(\boldsymbol{x})\right]$$
$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}\left[Z(\boldsymbol{x}), Z(\boldsymbol{x}')\right]$$

- The covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$ a positive definite kernel, usually stationary: $k(\boldsymbol{x}, \boldsymbol{x}') = f(|\boldsymbol{x} - \boldsymbol{x}'|)$
- $k(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 R(\boldsymbol{x}, \boldsymbol{x}')$, where $R(\boldsymbol{x}, \boldsymbol{x}')$ is the *auto-correlation function* and σ^2 is the *process variance*

[Very] Short introduction to Gaussian processes

Gaussian processes in a nutshell

Consider a probability space $(\Omega_Z, \mathcal{F}_Z, \mathbb{P}_Z)$ and $\boldsymbol{x} \in \mathbb{R}^M$. A stochastic process $Z(\boldsymbol{x})$ is Gaussian *i.i.f.* for any finite set $C \in \mathbb{R}^M$ the collection of random variables $Z(C)$ has a Gaussian joint distribution

Notes on Gaussian processes

- A Gaussian process is entirely defined by its mean and covariance functions:

$$\mu(\boldsymbol{x}) = \mathbb{E}\left[Z(\boldsymbol{x})\right]$$
$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}\left[Z(\boldsymbol{x}), Z(\boldsymbol{x}')\right]$$

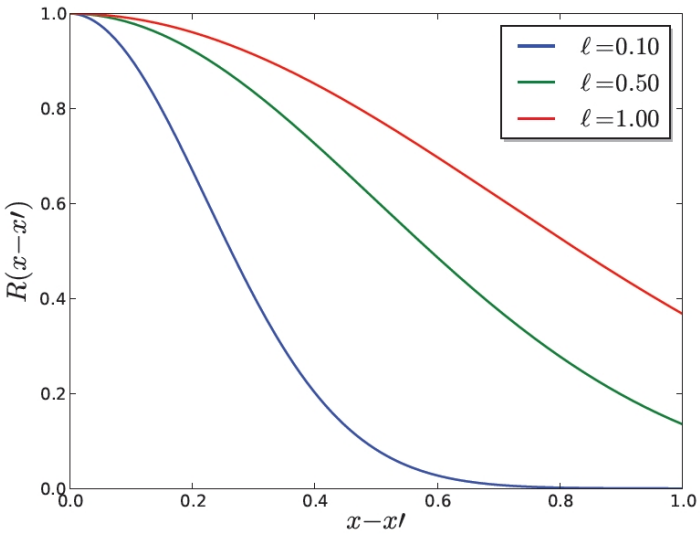
- The covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$ a positive definite kernel, usually stationary: $k(\boldsymbol{x}, \boldsymbol{x}') = f(|\boldsymbol{x} - \boldsymbol{x}'|)$
- $k(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 R(\boldsymbol{x}, \boldsymbol{x}')$, where $R(\boldsymbol{x}, \boldsymbol{x}')$ is the *auto-correlation function* and σ^2 is the *process variance*

Role of the covariance kernel

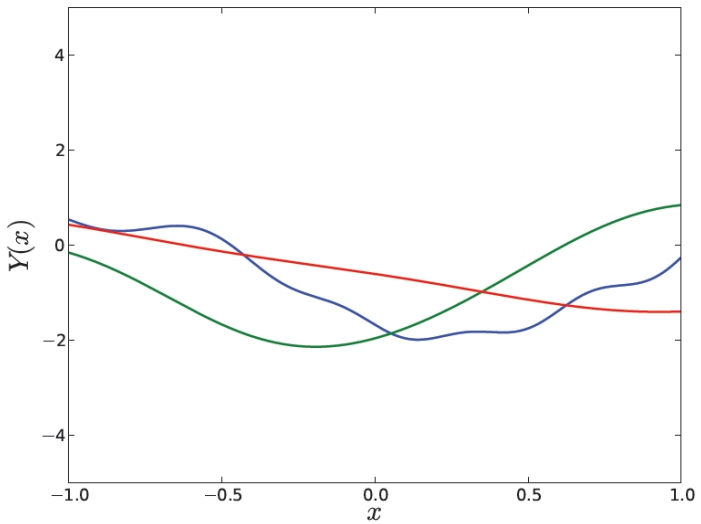
Consider the following parametric Gaussian covariance kernel

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \exp\left(-\sum_{i=1}^{M}\left(\frac{x_i - x_i'}{\theta_i}\right)^2\right)$$

where $\{\theta_i, i = 1, \ldots, d\}$ are scale parameters and σ^2 is the process variance



Covariance kernel



Random process trajectories

Illustrations taken from Dubourg (2011)

Gaussian process modelling (Kriging)

Gaussian process modelling (*a.k.a.* Kriging) assumes that the map $y = \mathcal{M}(\boldsymbol{x})$ is a realization of a Gaussian process:

$$Y(\boldsymbol{x}, \omega) = \sum_{j=1}^{p} \beta_j \, f_j(\boldsymbol{x}) + \sigma \, Z(\boldsymbol{x}, \omega)$$

where:

- $\boldsymbol{f} = \{f_j, j = 1, \ldots, p\}^{\mathsf{T}}$ are predefined (*e.g.* polynomial) functions which form the trend or regression part
- $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_p\}^{\mathsf{T}}$ are the regression coefficients
- σ^2 is the variance of $Y(\boldsymbol{x}, \omega)$
- $Z(\boldsymbol{x}, \omega)$ is a stationary, zero-mean, unit-variance Gaussian process

$$\mathbb{E}\left[Z(\boldsymbol{x}, \omega)\right] = 0 \qquad \mathrm{Var}\left[Z(\boldsymbol{x}, \omega)\right] = 1 \qquad \forall \, \boldsymbol{x} \in \mathbb{X}$$

The Gaussian measure artificially introduced is different from the aleatory uncertainty on the model parameters $\boldsymbol{X}$

Assumptions on the trend and the zero-mean process

Prior assumptions are made based on the existing knowledge on the model to surrogate (linearity, smoothness, etc.)

Trend

- Simple Kriging: $p = 1,\ f_1 = 1$ known constant β_1
- Ordinary Kriging: $p = 1,\ f_1 = 1$, unknown constant β_1
- Universal Kriging: f_j are ha set of arbitratry functions, *e.g.* $\left\{f_j(x) = x^{j-1},\ j = 1, \ldots, p\right\}$ in 1D

Type of auto-correlation function of $Z(\boldsymbol{x})$

A family of auto-correlation function $R(\cdot;\, \boldsymbol{\theta})$ is selected:

$$\mathrm{Cov}\left[Z(\boldsymbol{x}), Z(\boldsymbol{x}')\right] = \sigma^2\, R(\boldsymbol{x}, \boldsymbol{x}';\, \boldsymbol{\theta})$$

e.g. square exponential, generalized exponential, Matérn, etc.

Kriging in a nutshell

Data

- Given an experimental design $\mathcal{X} = \left\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N_{ED})}\right\}$ and $\boldsymbol{y} = \left\{y^{(1)} = \mathcal{M}(\boldsymbol{x}^{(1)}), \ldots, y^{(N_{ED})} = \mathcal{M}(\boldsymbol{x}^{(N_{ED})})\right\}$

Assumption

- We assume that $\mathcal{M}(\boldsymbol{x})$ is a **realization** of the Gaussian process $Y(\boldsymbol{x}, \omega)$ such that the values $y^{(i)} = \mathcal{M}(\boldsymbol{x}^{(i)})$ are known on $\left\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N_{ED})}\right\}$

Goal

- Of interest is the prediction at a new point $\boldsymbol{x}_0 \notin \mathcal{X}$, denoted by $\hat{Y}_0 \equiv \hat{Y}(\boldsymbol{x}_0, \omega)$, which will be used as a surrogate of $\mathcal{M}(\boldsymbol{x}_0)$

Joint distribution of the observations

- For each point $\boldsymbol{x}^{(i)} \in \mathcal{X}$, $Y^{(i)} \equiv Y(\boldsymbol{x}^{(i)})$ is a Gaussian variable:

$$Y^{(i)} = \sum_{j=1}^{p} \beta_j \, f_j(\boldsymbol{x}^{(i)}) + \sigma Z_i = \boldsymbol{f}_i^{\mathsf{T}} \cdot \boldsymbol{\beta} + \sigma \; Z_i \qquad Z_i \sim \mathcal{N}(0,1)$$

- The joint distribution of $\boldsymbol{Y}$ is Gaussian:

$$Y^{(i)} \sim \mathcal{N}(\boldsymbol{f}_i^{\mathsf{T}} \boldsymbol{\beta}, \sigma^2) \qquad \mathrm{Cov}\left[Y^{(i)}, Y^{(j)}\right] = \sigma^2 \, R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}; \boldsymbol{\theta})$$

that is:

$$\boldsymbol{Y} = \mathcal{N}_{N_{ED}}(\mathbf{F}\boldsymbol{\beta} \, , \, \sigma^2 \, \mathbf{R}(\boldsymbol{\theta}))$$

- Regression matrix $\mathbf{F}$ of size $(N_{ED} \times p)$

$$\mathbf{F}_{ij} = f_j(\boldsymbol{x}^{(i)})$$
$$i = 1, \ldots, N_{ED}, \; j = 1, \ldots, p$$

- Correlation matrix $\mathbf{R}(\boldsymbol{\theta})$ of size $(N_{ED} \times N_{ED})$

$$\mathbf{R}_{ij}(\boldsymbol{\theta}) = R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}; \, \boldsymbol{\theta})$$

Joint distribution of the predictor / observations

- The joint distribution of $\left\{Y_0, Y^{(1)}, \ldots, Y^{(N_{ED})}\right\}^{\mathsf{T}}$ is Gaussian:

$$\left\{ \begin{array}{c} Y_0 \\ \boldsymbol{Y} \end{array} \right\} \sim \mathcal{N}_{1+N_{ED}} \left(\left\{ \begin{array}{c} \boldsymbol{f}_0^{\mathsf{T}}\,\boldsymbol{\beta} \\ \mathbf{F}\,\boldsymbol{\beta} \end{array} \right\}, \sigma^2 \left[\begin{array}{cc} 1 & \boldsymbol{r}_0^{\mathsf{T}} \\ \boldsymbol{r}_0 & \mathbf{R} \end{array} \right] \right)$$

- Regression matrix $\mathbf{F}$ of size $(N_{ED} \times p)$

$$\mathbf{F}_{ij} = f_j(\boldsymbol{x}^{(i)})$$
$$i = 1, \ldots, N_{ED},\ j = 1, \ldots, p$$

- Vector of regressors $\boldsymbol{f}_0$ of size p

$$\boldsymbol{f}_0 = \{f_1(\boldsymbol{x}_0), \ldots, f_p(\boldsymbol{x}_0)\}$$

- Correlation matrix $\mathbf{R}$ of size $(N_{ED} \times N_{ED})$

$$\mathbf{R}_{ij} = R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}; \boldsymbol{\theta})$$

- Cross-correlation vector $\boldsymbol{r}_0$ of size N_{ED}

$$\boldsymbol{r}_{0i} = R(\boldsymbol{x}^{(i)}, \boldsymbol{x}_0; \boldsymbol{\theta})$$

Kriging predictor as the Gaussian process mean

Metamodel: mean predictor

$$\mu_{\widehat{Y}_0} = \boldsymbol{f}_0^{\mathsf{T}}\,\boldsymbol{\beta} + \boldsymbol{r}_0^{\mathsf{T}}\mathbf{R}^{-1}\,(\boldsymbol{y} - \mathbf{F}\,\boldsymbol{\beta})$$

Kriging variance:

$$\sigma^2_{\widehat{Y}_0} = \mathbb{E}\left[(\widehat{Y}_0 - Y_0)^2\right] = \sigma^2\,\left(1 - \boldsymbol{r}_0^{\mathsf{T}}\,\mathbf{R}^{-1}\,\boldsymbol{r}_0\right)$$

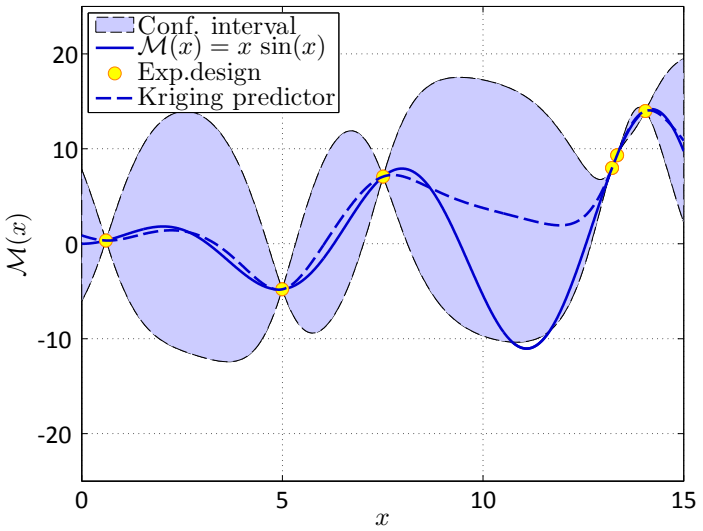
Properties

- The mean predictor has a regression part $\boldsymbol{f}_0^{\mathsf{T}}\,\boldsymbol{\beta} = \sum_{j=1}^{p}\beta_j\,f_j(\boldsymbol{x}_0)$ and a local correction
- It **interpolates** the experimental design:

$$\left.\begin{aligned}\mu_{\widehat{Y}_i} \equiv \mu_{\widehat{Y}(\boldsymbol{x}^{(i)})} &= y^{(i)}\\ \sigma^2_{\widehat{Y}_i} \equiv \sigma^2_{\widehat{Y}(\boldsymbol{x}^{(i)})} &= 0\end{aligned}\right\} \quad \forall\,\boldsymbol{x}^{(i)} \in \mathcal{X}$$

Confidence intervals

- Due to the Gaussianity of the predictor $\widehat{Y}_0 \sim \mathcal{N}(\mu_{\widehat{Y}_0}, \sigma^2_{\widehat{Y}_0})$, one can derive confidence intervals on the prediction



- With confidence level $(1-\alpha)$, *e.g.* 95%, one gets:

$$\mu_{\widehat{Y}_0} - 1.96\,\sigma_{\widehat{Y}_0} \leq \mathcal{M}(\boldsymbol{x}_0) \leq \mu_{\widehat{Y}_0} + 1.96\,\sigma_{\widehat{Y}_0}$$

- The Kriging predictor is asymptotically consistent:

$$\lim_{N\to\infty} \mathbb{E}\left[\left(\widehat{Y}_0 - Y_0\right)^2\right] = 0$$

when the size of the experimental design N tends to infinity

Outline

Kriging inference

So far:

- Kriging predictor assumes that the autocovariance function $\sigma^2 R(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta})$ and the trend coefficients $\boldsymbol{\beta}$ are known

In practice:

- A choice is made for the family of autocorrelation function used, *e.g.* Gaussian, exponential, Matérn-ν, etc.
- The parameters of the covariance function and of the trend, $\left(\vec{\beta}, \sigma^2, \boldsymbol{\theta}\right)$, must be estimated from the data, *i.e.* the experimental design:

$$\mathcal{X} = \left\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N_{ED})}\right\} \qquad \boldsymbol{y} = \left\{y^{(1)} = \mathcal{M}(\boldsymbol{x}^{(1)}), \ldots, y^{(N)} = \mathcal{M}(\boldsymbol{x}^{(N_{ED})})\right\}$$

Maximum likelihood estimation

Maximum likelihood estimation in Kriging

- Assuming that data follows a joint Gaussian distribution $\boldsymbol{Y} \sim \mathcal{N}_N(\mathbf{F}\boldsymbol{\beta}\,, \mathbf{R}(\boldsymbol{\theta}))$ the negative log-likelihood reads:

$$-\log \mathsf{L}\left(\boldsymbol{\beta},\, \sigma^2,\, \boldsymbol{\theta} \mid \boldsymbol{y}\right) = \frac{1}{2\,\sigma^2}\left(\boldsymbol{y} - \mathbf{F}\,\boldsymbol{\beta}\right)^{\mathsf{T}} \mathbf{R}(\boldsymbol{\theta})^{-1}\left(\boldsymbol{y} - \mathbf{F}\,\boldsymbol{\beta}\right) + \frac{N}{2}\,\log\left(2\,\pi\right) + \frac{N}{2}\,\log\left(\sigma^2\right) + \frac{1}{2}\,\log\left(\det \mathbf{R}(\boldsymbol{\theta})\right)$$

- The solution $\left(\widehat{\boldsymbol{\beta}}, \widehat{\sigma^2}\right)$ is obtained by solving:

$$\frac{\partial(-\log \mathsf{L})}{\partial \boldsymbol{\beta}} = 0 \quad ; \quad \frac{\partial(-\log \mathsf{L})}{\partial \sigma^2} = 0$$

Note on the variance: the estimation of the β coefficients adds extra term to the predictor variance:

$$\sigma^2_{\widehat{Y}_0} = \mathbb{E}\left[\left(\widehat{Y}_0 - Y_0\right)^2\right] = \sigma^2\left(1 - \boldsymbol{r}_0^{\mathsf{T}}\,\mathbf{R}^{-1}\,\boldsymbol{r}_0 + \boldsymbol{u}_0^{\mathsf{T}}\left(\mathbf{F}^{\mathsf{T}}\,\mathbf{R}^{-1}\,\mathbf{F}\right)^{-1}\boldsymbol{u}_0\right)$$

with $\boldsymbol{u}_0 = \mathbf{F}^{\mathsf{T}}\,\mathbf{R}^{-1}\,\boldsymbol{r}_0 - \boldsymbol{f}_0$

Maximum likelihood estimation in Kriging

- Assuming that data follows a joint Gaussian distribution $\boldsymbol{Y} \sim \mathcal{N}_N(\mathbf{F}\boldsymbol{\beta}\,, \mathbf{R}(\boldsymbol{\theta}))$ the negative log-likelihood reads:

$$-\log \mathsf{L}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta} \mid \boldsymbol{y}\right) = \frac{1}{2\,\sigma^2}\left(\boldsymbol{y} - \mathbf{F}\,\boldsymbol{\beta}\right)^{\mathsf{T}} \mathbf{R}(\boldsymbol{\theta})^{-1}\left(\boldsymbol{y} - \mathbf{F}\,\boldsymbol{\beta}\right) + \frac{N}{2}\log\left(2\,\pi\right) + \frac{N}{2}\log\left(\sigma^2\right) + \frac{1}{2}\log\left(\det \mathbf{R}(\boldsymbol{\theta})\right)$$

- The solution $\left(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2\right)$ is obtained by solving:

$$\frac{\partial(-\log \mathsf{L})}{\partial \boldsymbol{\beta}} = 0 \quad ; \quad \frac{\partial(-\log \mathsf{L})}{\partial \sigma^2} = 0$$

Note on the variance: the estimation of the β coefficients adds extra term to the predictor variance:

$$\sigma^2_{\widehat{Y}_0} = \mathbb{E}\left[\left(\widehat{Y}_0 - Y_0\right)^2\right] = \sigma^2\left(1 - r_0^{\mathsf{T}}\,\mathbf{R}^{-1}\,r_0 + \boldsymbol{u}_0^{\mathsf{T}}\left(\mathbf{F}^{\mathsf{T}}\,\mathbf{R}^{-1}\,\mathbf{F}\right)^{-1}\boldsymbol{u}_0\right)$$

with $\boldsymbol{u}_0 = \mathbf{F}^{\mathsf{T}}\,\mathbf{R}^{-1}\,r_0 - \boldsymbol{f}_0$

Maximum likelihood estimation

Computation of $\widehat{\beta}$ and $\widehat{\sigma}^2$

- The log-likelihood is quadratic in β

$$\frac{\partial(-\log \mathsf{L})}{\partial \boldsymbol{\beta}} = \boldsymbol{F}^{\mathsf{T}} \mathbf{R}^{-1}(\boldsymbol{\theta})(\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\beta}) = 0$$

that is:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \left(\mathbf{F}^{\mathsf{T}}\, \mathbf{R}(\boldsymbol{\theta})^{-1}\, \mathbf{F}\right)^{-1} \mathbf{F}^{\mathsf{T}}\, \mathbf{R}(\boldsymbol{\theta})^{-1}\, \boldsymbol{y}$$

- Then:

$$\widehat{\sigma^2}(\boldsymbol{\theta}) = \frac{1}{N} \left(\boldsymbol{y} - \mathbf{F} \cdot \widehat{\boldsymbol{\beta}}\right)^{\mathsf{T}} \mathbf{R}(\boldsymbol{\theta})^{-1} \cdot \left(\boldsymbol{y} - \mathbf{F}\, \widehat{\boldsymbol{\beta}}\right)$$

Correlation hyperparameters

- Minimizing $(-\log \mathsf{L})$ is equivalent to minimizing the reduced likelihood function

$$\psi(\boldsymbol{\theta}) = \widehat{\sigma^2}(\boldsymbol{\theta})\, \det \mathbf{R}(\boldsymbol{\theta})^{1/N}$$

- This problem is solved numerically using standard optimization algorithms, *e.g.* gradient-based or global

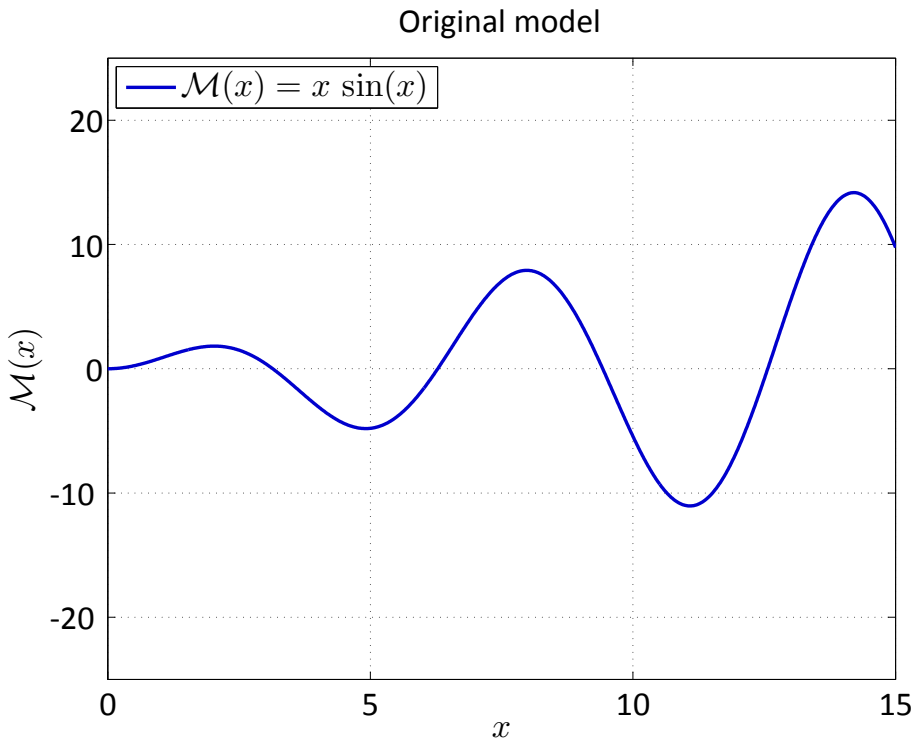
One-dimensional example

Computational model

$$x \mapsto x \sin x \qquad \text{for} \quad x \in [0,\, 15]$$

Experimental design

Six points selected in the range $[0,\, 15]$ using Monte Carlo simulation:



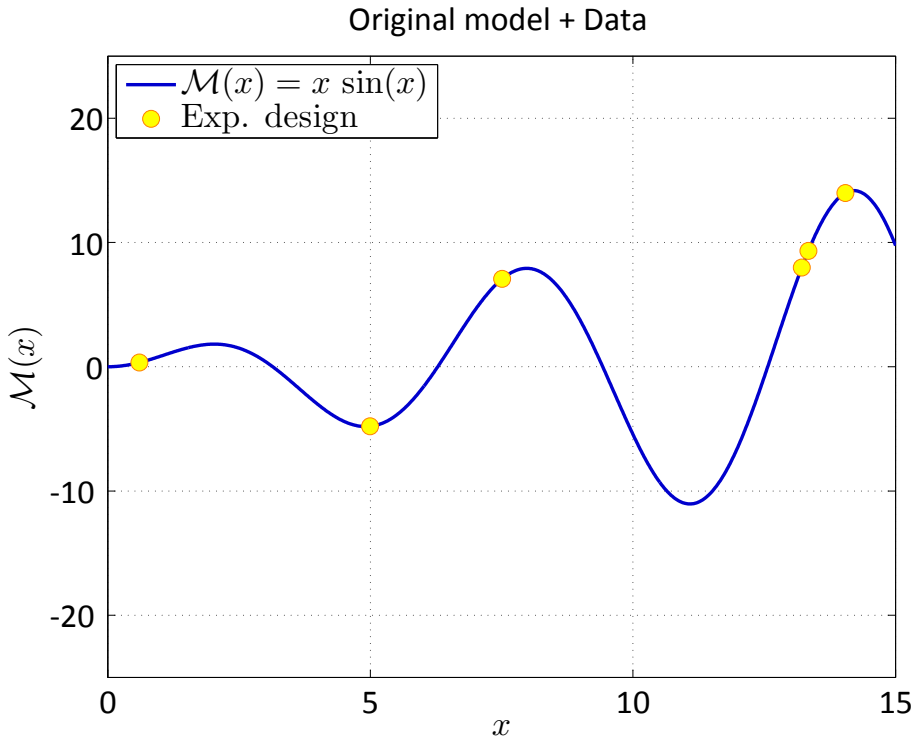
One-dimensional example

Computational model

$$x \mapsto x \sin x \qquad \text{for} \quad x \in [0,\, 15]$$

Experimental design

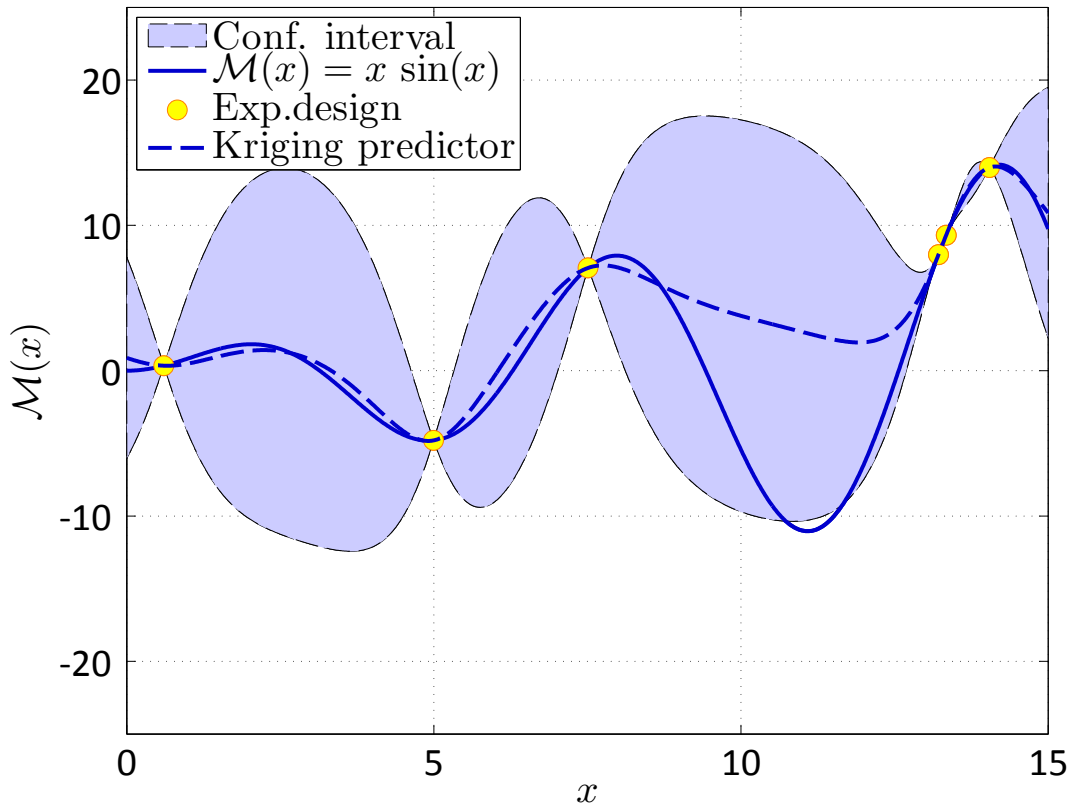
Six points selected in the range $[0,\, 15]$ using Monte Carlo simulation:



$$\mathcal{X} = \{0.6042 \quad 4.9958 \quad 7.5107 \quad 13.2154 \quad 13.3407 \quad 14.0439\}$$

Kriging predictor

- Trend: ordinary
- Covariance kernel: Gaussian
- Optimization method: BFGS (gradient based)



Outline

Beyond surrogates: active learning

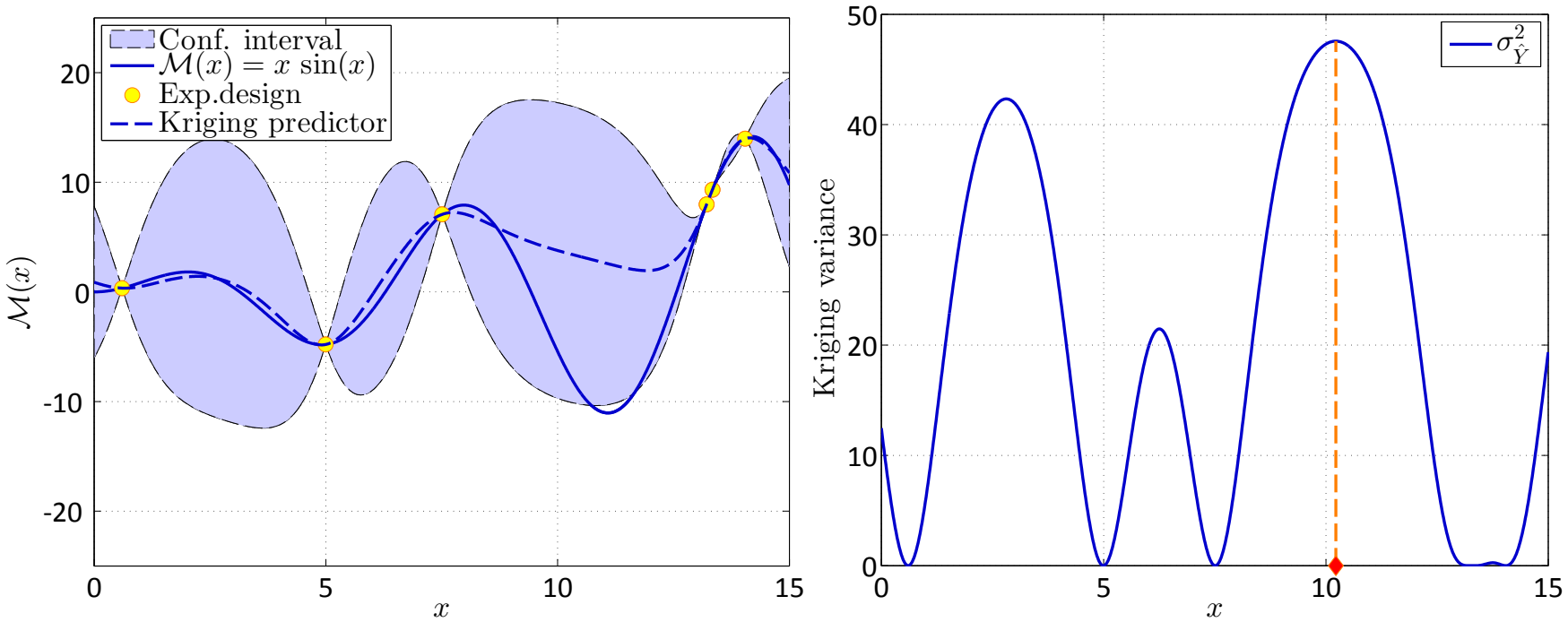
Heuristics

- Adaptively enrich the ED in regions of interest
- Capitalize on the Kriging variance information (**meta**-modelling)
- Naive approach: choose points where the Kriging variance is maximum

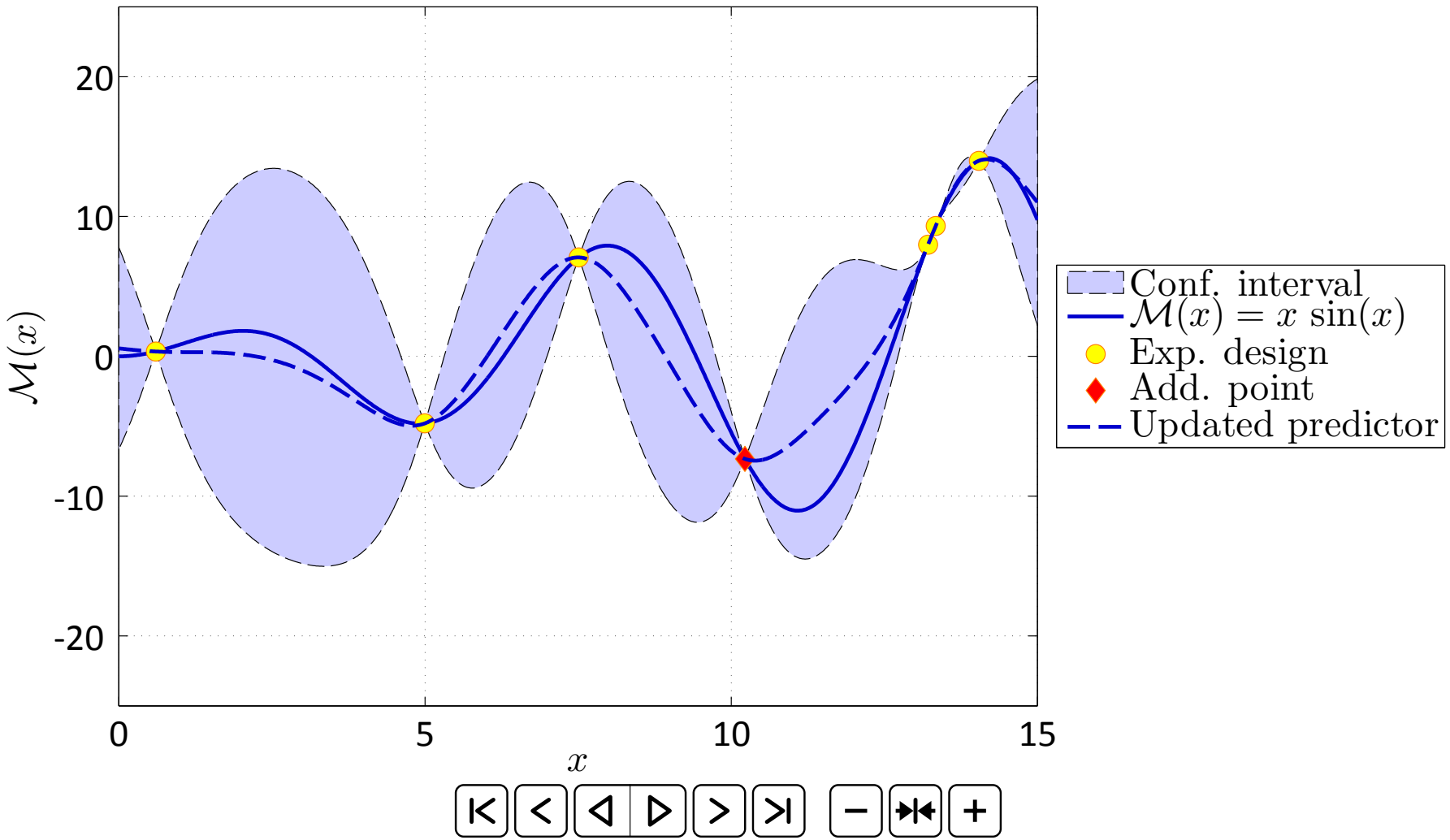
Beyond surrogates: active learning

Heuristics

- Adaptively enrich the ED in regions of interest
- Capitalize on the Kriging variance information (**meta**-modelling)
- Naive approach: choose points where the Kriging variance is maximum



Sequential updating



Outline

Typical engineering questions w.r.t risk and reliability

- What is the scattering of a quantity of interest Y?

  PDF f_Y
  $\hat{\mu}_Y, \hat{\sigma}_Y$

- What are the parameters that drive the uncertainty on the QoI ?

  Sensitivity indices

- What is the probability of failure (resp. non performance) of the system ?

  $p_f = \mathbb{P}(Y \geq y_{adm})$

- What is the optimal design (e.g. minimal cost) that guarantees some performance

  $\boldsymbol{d}^* = \arg\min \mathfrak{c}(\boldsymbol{d})$ s.t.
  $\mathbb{P}(g(\boldsymbol{X}(\boldsymbol{d}), \boldsymbol{Z}) \leq 0) \leq p_{f,adm}$

- What are the best-fit model parameters that allow one to reproduce experimental data

  Bayesian inversion

Typical engineering questions w.r.t risk and reliability

- What is the scattering of a quantity of interest Y? — PDF f_Y, $\hat{\mu}_Y, \hat{\sigma}_Y$
- What are the parameters that drive the uncertainty on the QoI ? — Sensitivity indices
- **What is the probability of failure (resp. non performance) of the system ?** — $p_f = \mathbb{P}(Y \geq y_{adm})$
- What is the optimal design (e.g. minimal cost) that guarantees some performance — $\boldsymbol{d}^* = \arg\min \mathfrak{c}(\boldsymbol{d})$ s.t. $\mathbb{P}(g(\boldsymbol{X}(\boldsymbol{d}), \boldsymbol{Z}) \leq 0) \leq p_{f,adm}$
- What are the best-fit model parameters that allow one to reproduce experimental data — Bayesian inversion

Limit state function

- For the assessment of the system's performance, failure criteria are defined, *e.g.* :

$$\text{Failure} \quad \Leftrightarrow \quad QoI = \mathcal{M}(\boldsymbol{x}) \geq q_{adm}$$

Examples:

+ admissible stress / displacements in civil engineering
+ max. temperature in heat transfer problems
+ crack propagation criterion in fracture mechanics

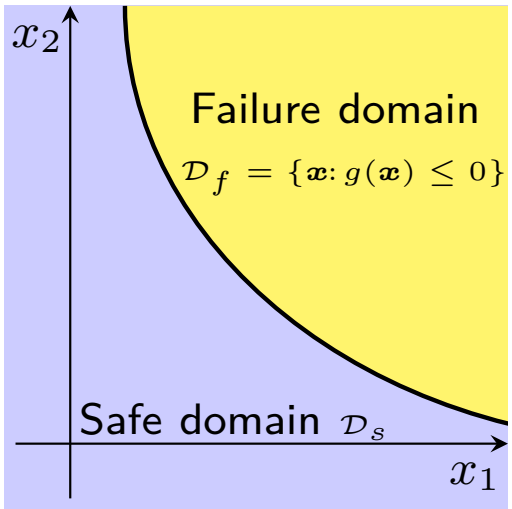
- The failure criterion is cast as a limit state function (performance function) $g:\ \boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}} \mapsto \mathbb{R}$ such that:

$$g\left(\boldsymbol{x}, \mathcal{M}(\boldsymbol{x})\right) \leq 0 \quad \text{Failure domain } \mathcal{D}_f$$

$$g\left(\boldsymbol{x}, \mathcal{M}(\boldsymbol{x})\right) > 0 \quad \text{Safety domain } \mathcal{D}_s$$

$$g\left(\boldsymbol{x}, \mathcal{M}(\boldsymbol{x})\right) = 0 \quad \text{Limit state surface}$$

e.g. $g(\boldsymbol{x}) = q_{adm} - \mathcal{M}(\boldsymbol{x})$

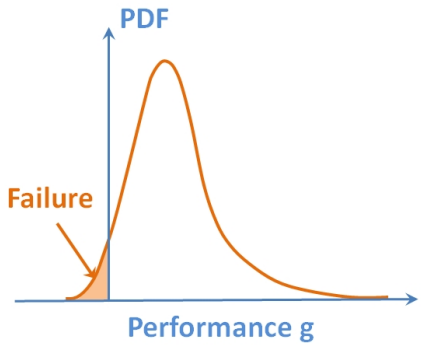


Probability of failure

Definition

$$P_f = \mathbb{P}\left(\{\boldsymbol{X} \in D_f\}\right) = \mathbb{P}\left(g\left(\boldsymbol{X}, \mathcal{M}(\boldsymbol{X})\right) \leq 0\right)$$

$$P_f = \int_{\mathcal{D}_f = \{\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}} : g(\boldsymbol{x}, \mathcal{M}(\boldsymbol{x})) \leq 0\}} f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x}$$



Features

- Multidimensional integral, whose dimension is equal to the number of basic input variables $M = \dim \boldsymbol{X}$
- Implicit domain of integration defined by a condition related to the sign of the limit state function:

$$\mathcal{D}_f = \{\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}} : \quad g(\boldsymbol{x}, \mathcal{M}(\boldsymbol{x})) \leq 0\}$$

- Failures are (usually) rare events: sought probability in the range 10^{-2} to 10^{-8}

Monte Carlo simulation

Reformulation

- Indicator function of the failure domain

$$\mathbf{1}_{\mathcal{D}_f}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } g\left(\boldsymbol{x}, \mathcal{M}\left(\boldsymbol{x}\right)\right) \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

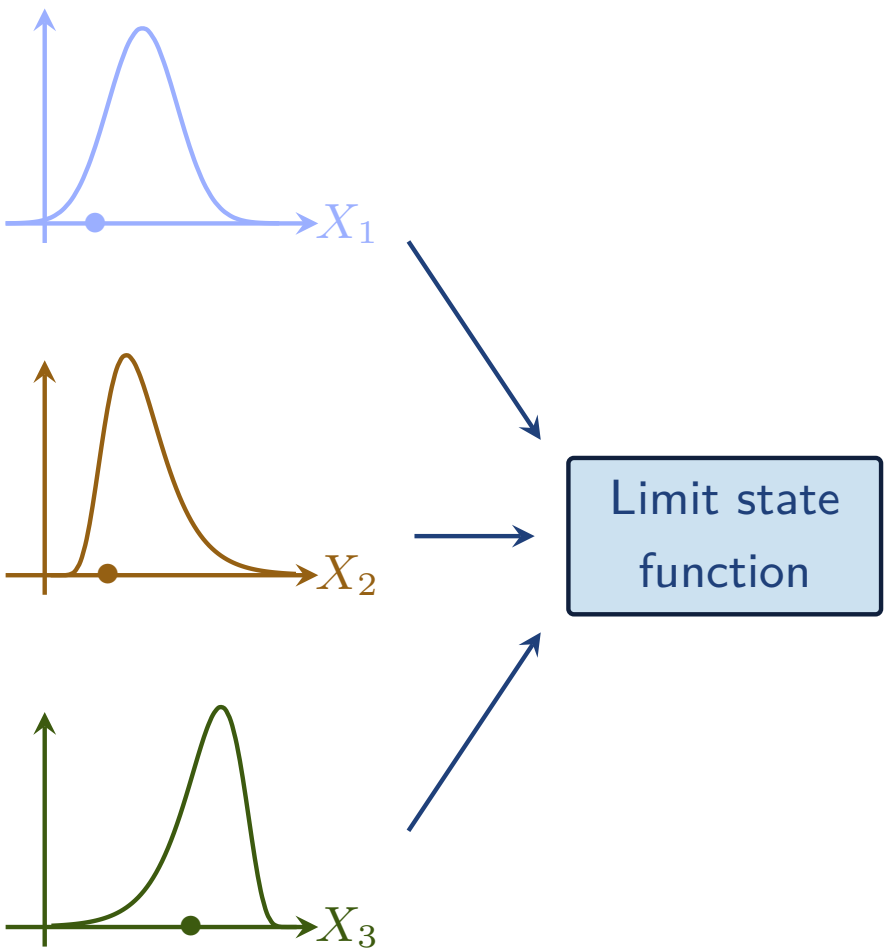
- Probability of failure:

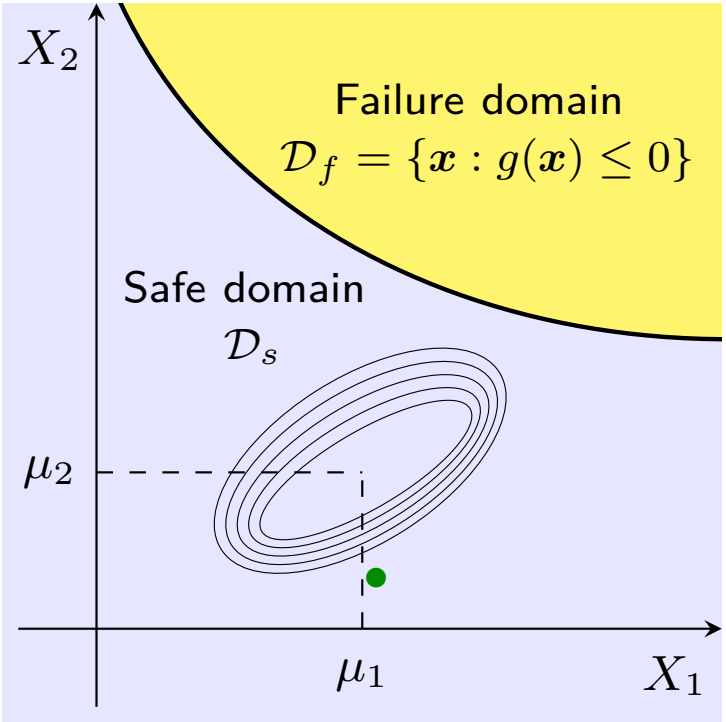
$$P_f = \int_{D_f} f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\mathcal{D}_{\boldsymbol{X}}} \mathbf{1}_{\mathcal{D}_f}(\boldsymbol{x})\, f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} = \mathbb{E}\left[\mathbf{1}_{\mathcal{D}_f}(\boldsymbol{X})\right]$$

Crude Monte Carlo estimator

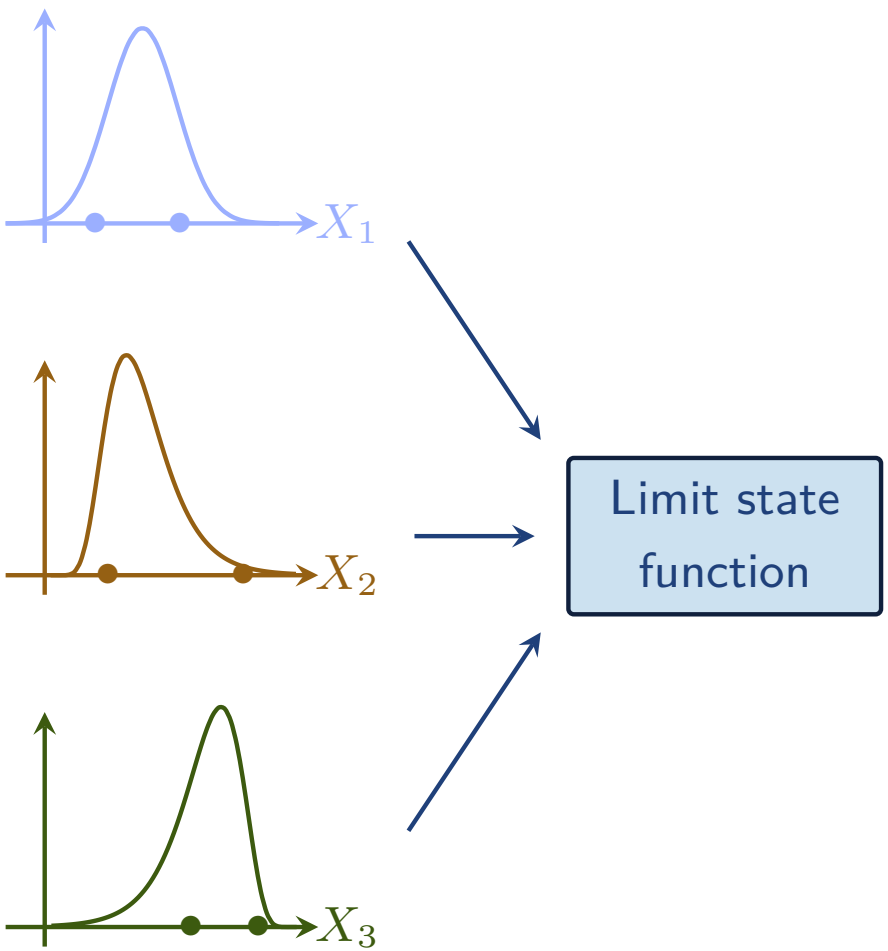
$$\hat{P}_f = \frac{1}{N} \sum_{i=1}^{N} 1_{\mathcal{D}_f}\left(\boldsymbol{X}_i\right) \qquad \boldsymbol{X}_i : \text{ i.i.d copies of } \boldsymbol{X}$$

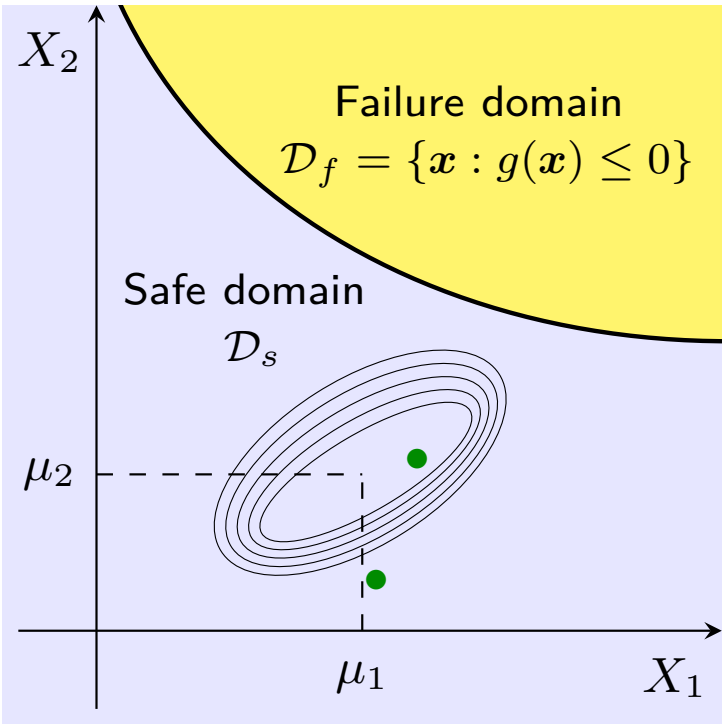
Monte Carlo simulation



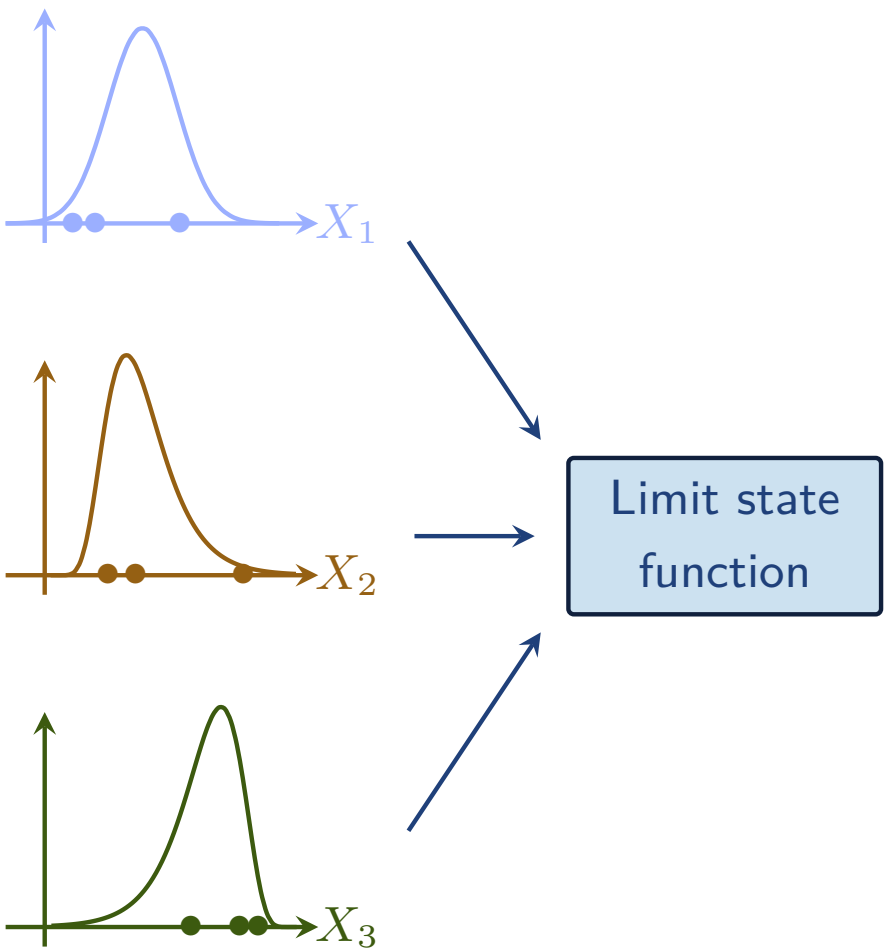


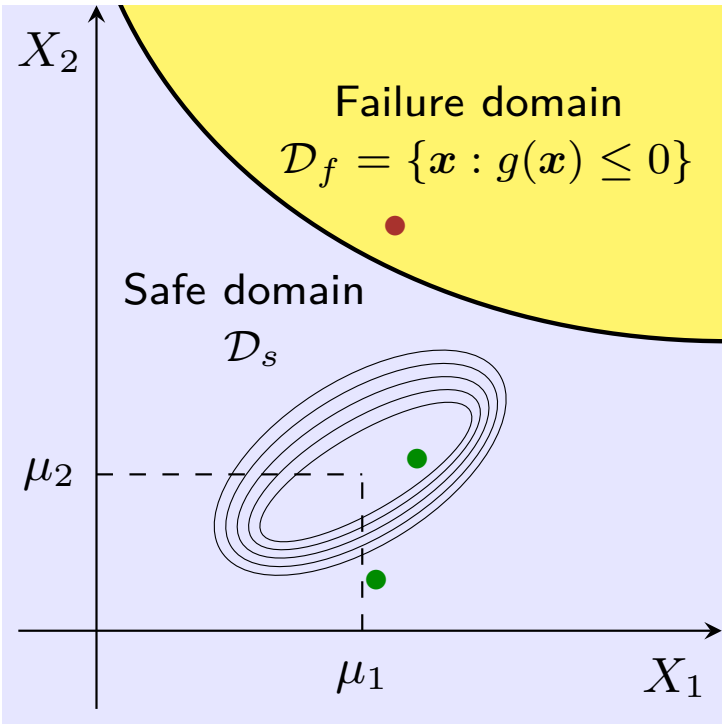
Monte Carlo simulation



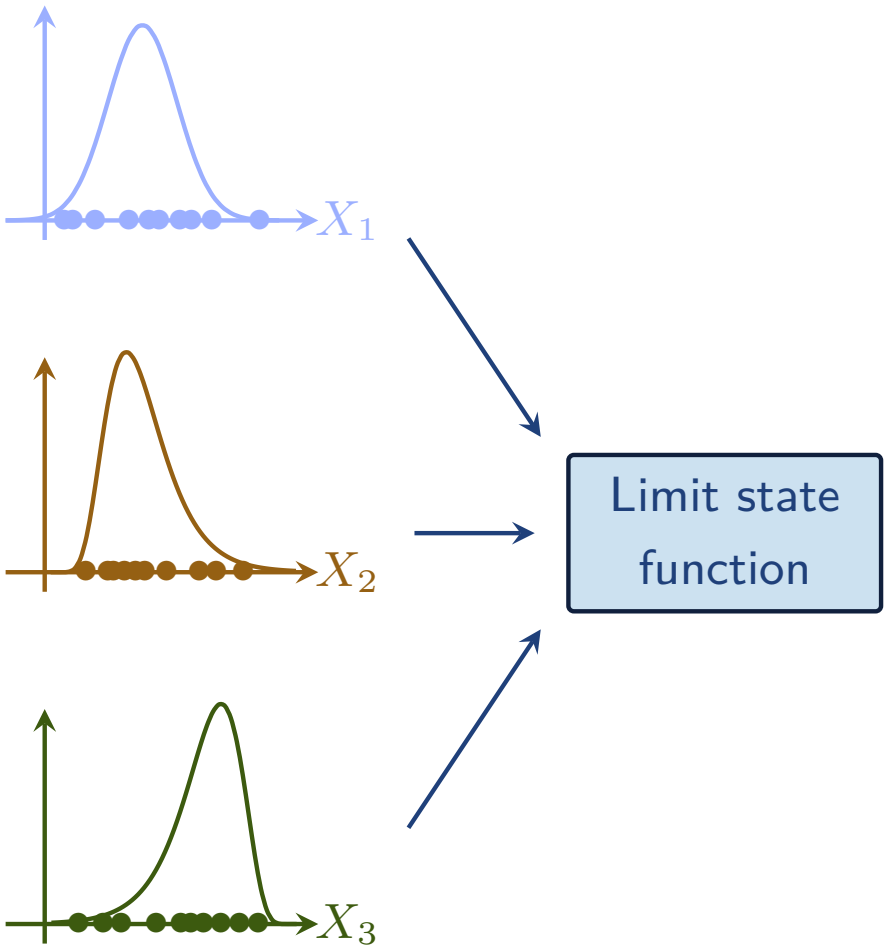


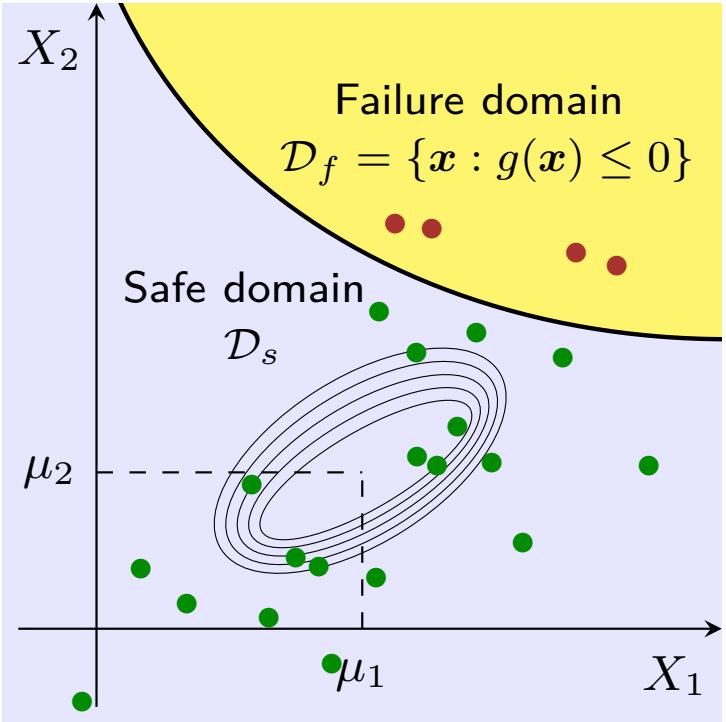
Monte Carlo simulation





Monte Carlo simulation





Estimator of the probability of failure P_f

- The estimator $\hat{P}_f$ is a sum of Bernoulli variables: it has a binomial distribution with:

  Mean value: $\mathbb{E}\left[\hat{P}_f\right] = P_f$ — Unbiasedness

  Variance: $\mathrm{Var}\left[\hat{P}_f\right] = \frac{1}{N} P_f \left(1 - P_f\right)$ — Convergence

- Its coefficient of variation reduces to $CV_{P_f} \approx 1/\sqrt{N\,P_f}$ for rare events.

Convergence rate of Monte Carlo simulation $\propto 1/\sqrt{N}$

Minimal size of the sample set

Suppose the probability of failure under consideration is of magnitude $P_f = 10^{-k}$ and an accuracy of 5% is aimed at.

$$CV_{P_f} = \frac{1}{\sqrt{N\,P_f}}$$

$$CV_{P_f} \leq 5\% \Longrightarrow N \geq 4.10^{k+2}$$

P_f	N_{min}
10^{-2}	40,000
10^{-3}	400,000
10^{-4}	4,000,000
10^{-6}	400,000,000

Estimator of the probability of failure P_f

- The estimator $\hat{P}_f$ is a sum of Bernoulli variables: it has a binomial distribution with:

  Mean value: $\mathbb{E}\left[\hat{P}_f\right] = P_f$ — Unbiasedness

  Variance: $\mathrm{Var}\left[\hat{P}_f\right] = \frac{1}{N} P_f \left(1 - P_f\right)$ — Convergence

- Its coefficient of variation reduces to $CV_{P_f} \approx 1/\sqrt{N\,P_f}$ for rare events.

Convergence rate of Monte Carlo simulation $\propto 1/\sqrt{N}$

Minimal size of the sample set

Suppose the probability of failure under consideration is of magnitude $P_f = 10^{-k}$ and an accuracy of 5% is aimed at.

$$CV_{P_f} = \frac{1}{\sqrt{N\,P_f}}$$

$$CV_{P_f} \leq 5\% \Longrightarrow N \geq 4.10^{k+2}$$

P_f	N_{min}
10^{-2}	40,000
10^{-3}	400,000
10^{-4}	4,000,000
10^{-6}	400,000,000

A note on Reliability analysis

An active research field

Reliability analysis (aka Structural Reliability) is a research field that has been active in the last 40 years, producing a rich literature on advanced methods to estimate low-probability events

Overview of solution strategies

- **Methods based on approximation**: FORM, SORM
- **Methods based on simulation**: MCS, Importance Sampling, Line Sampling, Subset Simulation, Asymptotic sampling, etc.
- **Methods based on metamodels**: Active learning-based methods

A note on Reliability analysis

An active research field

Reliability analysis (aka Structural Reliability) is a research field that has been active in the last 40 years, producing a rich literature on advanced methods to estimate low-probability events

Overview of solution strategies

- **Methods based on approximation**: FORM, SORM
- **Methods based on simulation**: MCS, Importance Sampling, Line Sampling, Subset Simulation, Asymptotic sampling, etc.
- **Methods based on metamodels**: Active learning-based methods

Outline

Use of Kriging for structural reliability analysis

- From a given experimental design $\mathcal{X} = \left\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}\right\}$, Kriging yields a mean predictor $\mu_{\hat{g}}(\boldsymbol{x})$ and the Kriging variance $\sigma_{\hat{g}}(\boldsymbol{x})$ of the limit state function g

- The mean predictor is substituted for the "true" limit state function, defining the surrogate failure domain

$$\mathcal{D}_f{}^0 = \{\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}} \ : \ \mu_{\hat{g}}(\boldsymbol{x}) \leq 0\}$$

- The probability of failure is approximated by: Kaymaz, Struc. Safety (2005)

$$P_f^0 = \mathbb{P}\left[\mu_{\hat{g}}(\boldsymbol{X}) \leq 0\right] = \int_{\mathcal{D}_f^0} f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} = \mathbb{E}\left[\mathbf{1}_{\mathcal{D}_f^0}(\boldsymbol{X})\right]$$

- Monte Carlo simulation can be used on the metamodel:

$$\widehat{P_f^0} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{\mathcal{D}_f^0}(\boldsymbol{x}_k)$$

Confidence bounds on the probability of failure

Shifted failure domains

Dubourg *et al.* , Struct. Mult. Opt. (2011)

- Let us define a confidence level $(1-\alpha)$ and $k_{1-\alpha} = \Phi^{-1}(1-\alpha/2)$, *i.e.* 1.96 if $1-\alpha = 95\%$, and:

$$\mathcal{D}_f^- = \{\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}} \ : \ \mu_{\hat{g}}(\boldsymbol{x}) + k_{1-\alpha}\,\sigma_{\hat{g}}(\boldsymbol{x}) \leq 0\}$$

$$\mathcal{D}_f^+ = \{\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}} \ : \ \mu_{\hat{g}}(\boldsymbol{x}) - k_{1-\alpha}\,\sigma_{\hat{g}}(\boldsymbol{x}) \leq 0\}$$

- Interpretation ($1-\alpha = 95\%$):
  - If $\boldsymbol{x} \in \mathcal{D}_f^0$ it belongs to the true failure domain with at worst a 50% chance
  - If $\boldsymbol{x} \in \mathcal{D}_f^+$ it belongs to the true failure domain with at worst 95% chance: conservative estimation

Bounds on the probability of failure

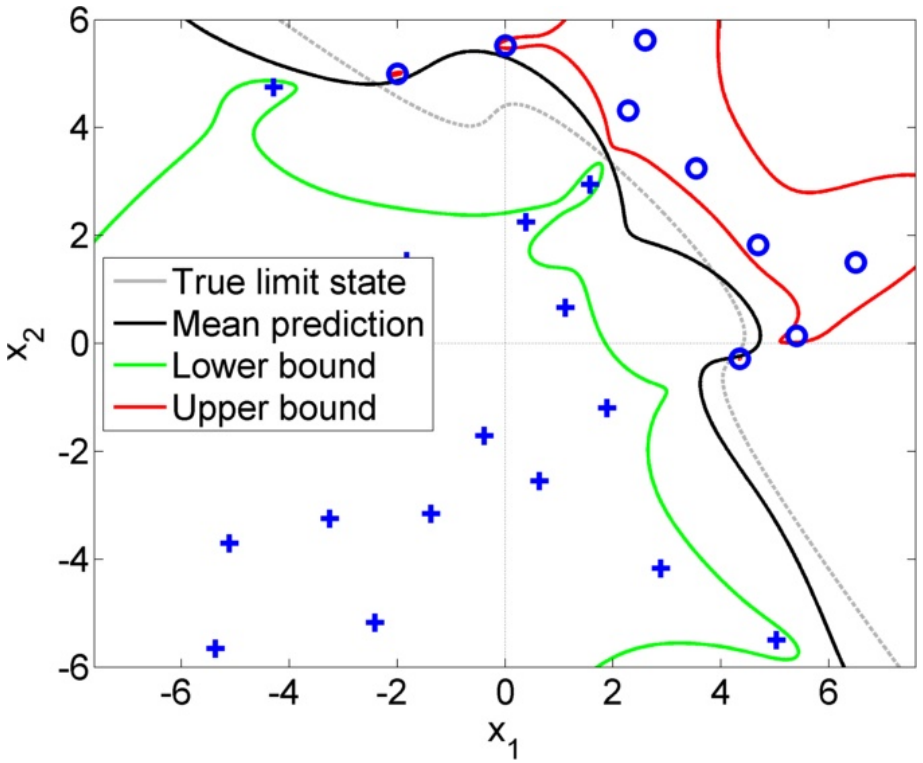
$$\mathcal{D}_f^- \subset \mathcal{D}_f^0 \subset \mathcal{D}_f^+ \qquad \Leftrightarrow \qquad P_f^- \leq P_f^0 \leq P_f^+$$

Example: hat function

Problem statement

$$g(\boldsymbol{x}) = 20 - (x_1 - x_2)^2 - 8\,(x_1 + x_2 - 4)^3$$

where X_1 , $X_2 \sim \mathcal{N}(0, 1)$



- Ref. solution:

$$P_f = 1.07 \cdot 10^{-4}$$

- Kriging surrogate:

$$P_f^- = 7.70 \cdot 10^{-6}$$
$$P_f^0 = 4.43 \cdot 10^{-4}$$
$$P_f^+ = 5.52 \cdot 10^{-2}$$

How to improve the results?

Heuristics

- The Monte Carlo estimate of P_f reads:

$$\widehat{P_f} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{\mathcal{D}_f}(\boldsymbol{x}_k) \approx \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{\mathcal{D}_f^0}(\boldsymbol{x}_k)$$

- The Kriging-based prediction is accurate when:

$$\mathbf{1}_{\mathcal{D}_f^0}(\boldsymbol{x}_k) = \mathbf{1}_{\mathcal{D}_f}(\boldsymbol{x}_k) \quad \text{for almost all } \boldsymbol{x}_k$$

i.e. if $\mu_{\hat{g}}(\boldsymbol{x})$ is of the same sign as $g(\boldsymbol{x})$ for almost all sample points

Ensure that the mean predictor $\mu_{\hat{g}}(\boldsymbol{x})$ classifies properly the MCS samples according to the sign of $g(\boldsymbol{x})$

Adaptive Kriging for structural reliability

Procedure

- Start from an initial experimental design $\mathcal{X}$ and a Kriging surrogate
- At each iteration:
  - Select the next point(s) to be added to $\mathcal{X}$: enrichment criterion
  - Update the Kriging surrogate
  - Compute an estimation of P_f and bounds
  - Check convergence

Different enrichment criteria

Requirements

- It shall be based on the available information: $\left(\mu_{\hat{g}}(\boldsymbol{x}), \sigma_{\hat{g}}(\boldsymbol{x})\right)$
- It shall favor new points in the vicinity of the limit state surface
- If possible, it shall yield the best K points when distributed computing is available

Different enrichment criteria

- Margin indicator function — Ph.D Deheeger (2008); Bourinet *et al.* , Struc. Safety (2011)
- Margin classification function — Ph.D Dubourg (2011); Dubourg *et al.* , PEM (2013)
- Learning function U — Ph.D Échard (2012); Échard & Gayton, RESS (2011)
- Expected feasibility function — Bichon *et al.* , AIAA (2008); RESS (2011)
- Stepwise uncertainty reduction (SUR) — Bect *et al.* , Stat. Comput. (2012)

Learning function $U(\boldsymbol{x})$

Definition

- The learning function U is defined by: Échard *et al.* (2011)

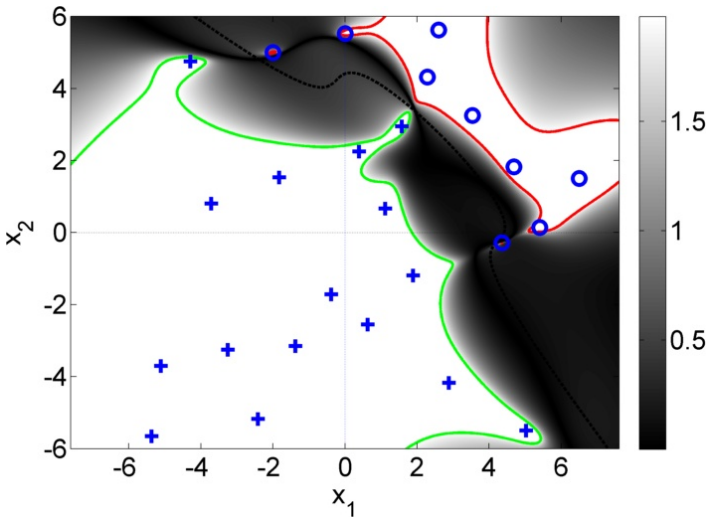
$$U(\boldsymbol{x}) = \frac{|\mu_{\hat{g}}(\boldsymbol{x})|}{\sigma_{\hat{g}}(\boldsymbol{x})}$$

Interpretation

- It describes the distance of the mean predictor $\mu_{\hat{g}}$ to zero in terms of a number of Kriging standard deviations $\sigma_{\hat{g}}$
- A small value of $U(\boldsymbol{x})$ means that:
  - $\mu_{\hat{g}}(\boldsymbol{x}) \approx 0$: $\boldsymbol{x}$ is close to the limit state surface
  - and / or $\sigma_{\hat{g}}(\boldsymbol{x}) >> 0$: the uncertainty in the prediction at point $\boldsymbol{x}$ is large
- The probability of misclassification of a point $\boldsymbol{x}$ is equal to $\Phi(-U(\boldsymbol{x}))$

Bect *et al.* , Stat. Comput. (2012)

Comparison of the enrichment criteria



Learning function U

Optimization of the enrichment criterion

$$\boldsymbol{x}_U^* = \arg \min_{\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}}} U(\boldsymbol{x})$$

Requires the solution of a complex optimization problem in each iteration

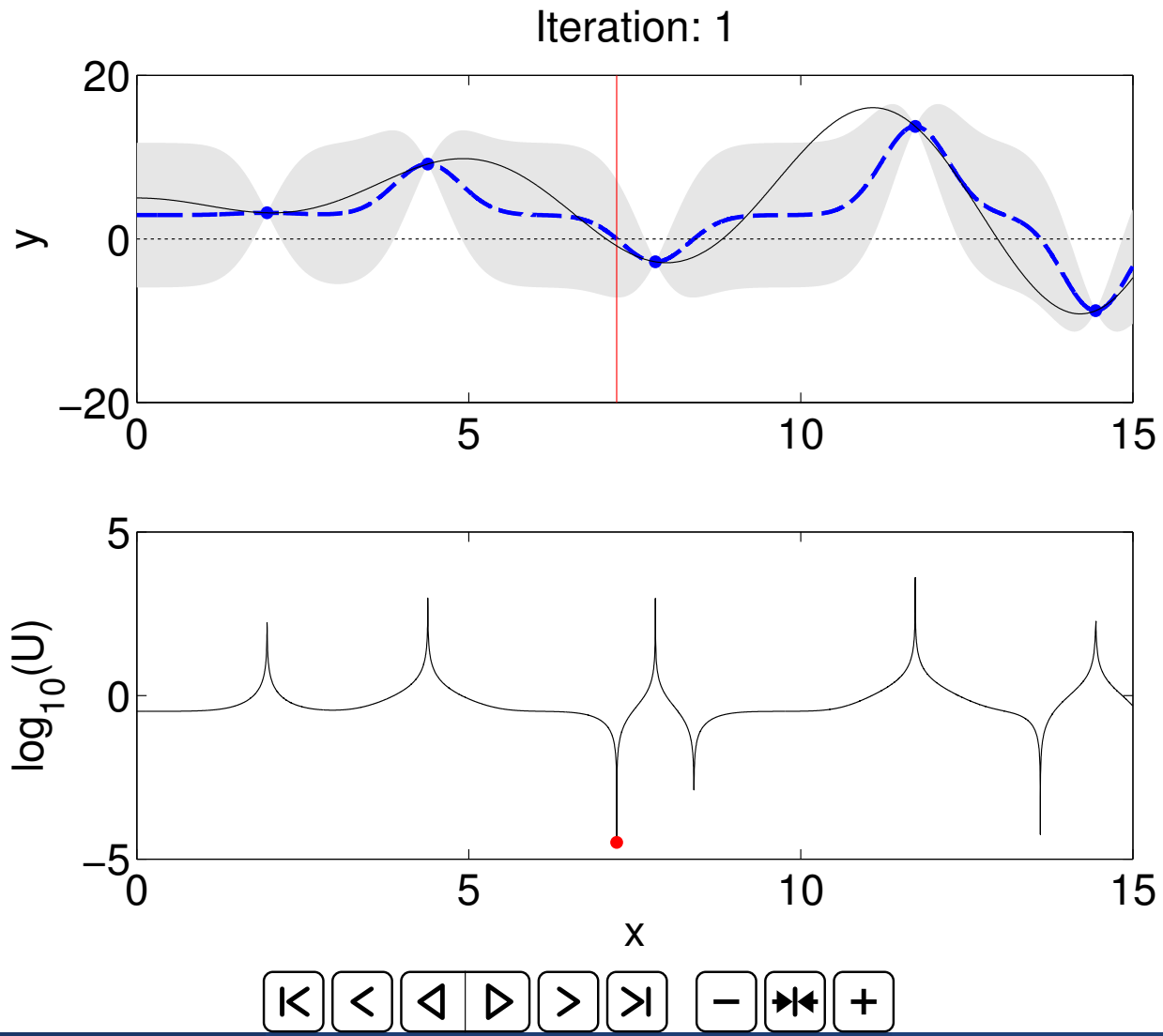
Discrete optimization over a large Monte Carlo sample $\mathfrak{X} = \left\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N_{MC})}\right\}$

$$\boldsymbol{x}_U^* = \arg \min_{i=1,\ldots,n} \left\{U(\boldsymbol{x}^{(1)}), \ldots, U(\boldsymbol{x}^{(N_{MC})})\right\}$$

Echard, B., Gayton, N. & Lemaire, M. AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation, Structural Safety (2011)

1D Application example - U function

Limit state function: $g(x) = 5 - x \sin x$

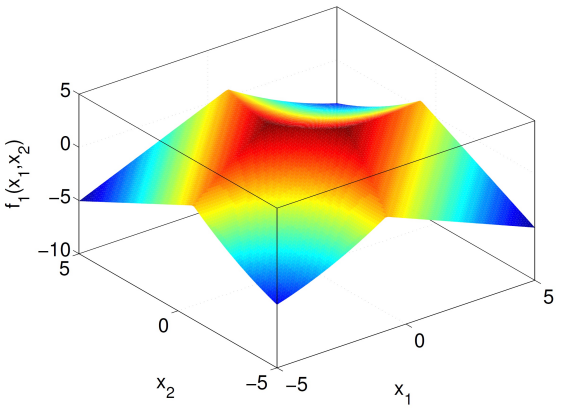


Series system

Schöbi *et al.* , ASCE J. Risk Unc. (2016)

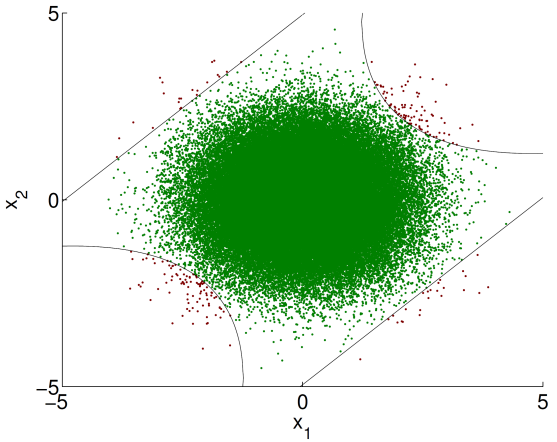
Consider the system reliability analysis defined by:

$$g(\boldsymbol{x}) = \min \begin{pmatrix} 3 + 0.1\,(x_1 - x_2)^2 - \frac{x_1+x_2}{\sqrt{2}} \\ 3 + 0.1\,(x_1 - x_2)^2 + \frac{x_1+x_2}{\sqrt{2}} \\ (x_1 - x_2) + \frac{6}{\sqrt{2}} \\ (x_2 - x_1) + \frac{6}{\sqrt{2}} \end{pmatrix}$$



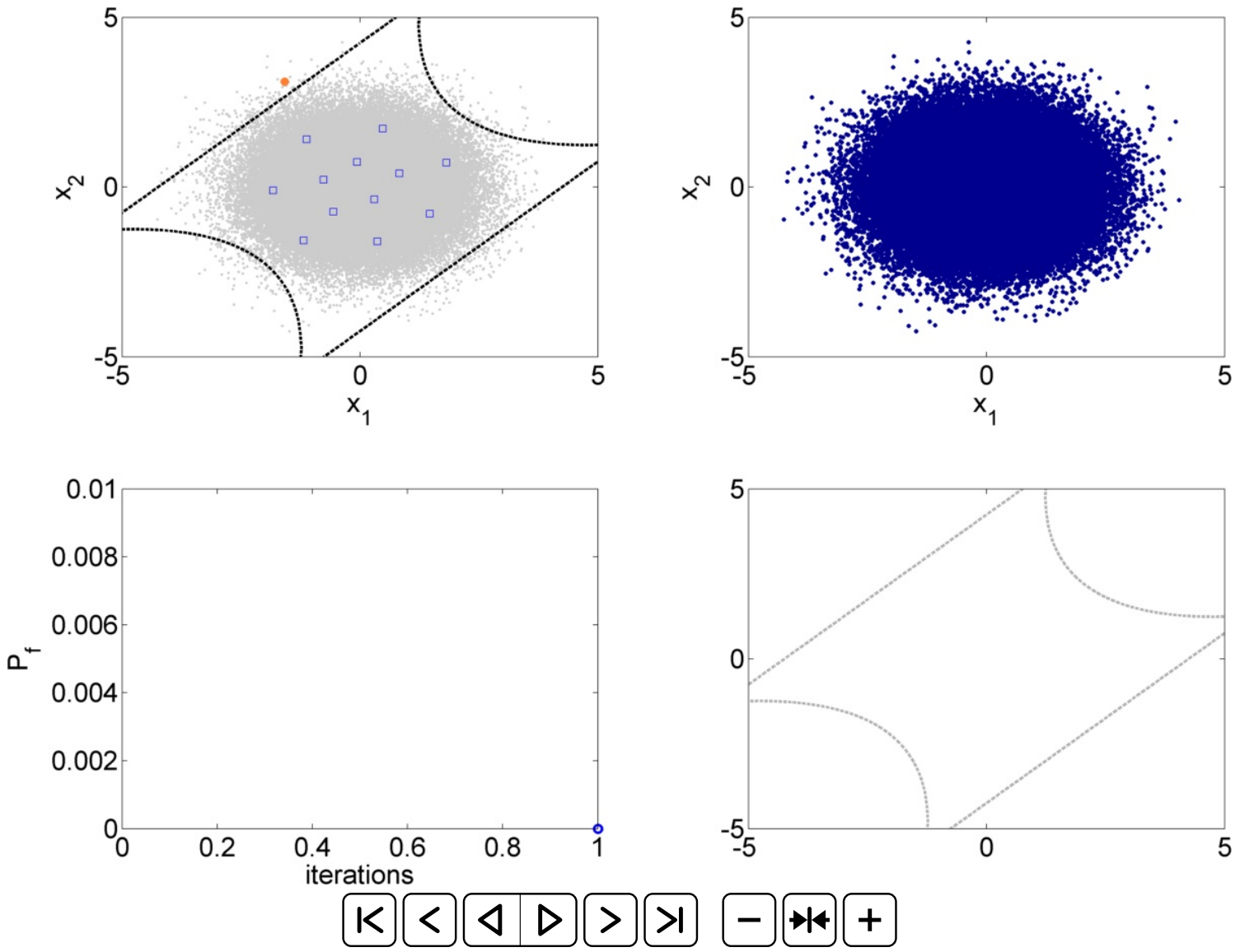
where $X_1, X_2 \sim \mathcal{N}(0,1)$

- Initial design: LHS of size 12 (transformed into the standard normal space)
- In each iteration, one point is added (maximize the probability of missclassification)



- The mean predictor $\mu_{\widehat{\mathcal{M}}}(\boldsymbol{x})$ is used, as well as the bounds $\mu_{\widehat{\mathcal{M}}}(\boldsymbol{x}) \pm 2\sigma_{\widehat{\mathcal{M}}}(\boldsymbol{x})$ so as to get bounds on P_f: $\hat{P}_f^- \leq \hat{P}_f^0 \leq \hat{P}_f^+$

Results with classical Kriging (AK-MCS)



Outline

Conclusions

Conclusions

- Metamodels are ideal tools to deal with uncertainty when the models are black-boxes
- Estimating low probabilities of failure requires more refined algorithms than plain MCS
- Recent research on metamodels and active learning has brought new extremely efficient algorithms
- Accurate estimations of P_f are obtained with $\mathcal{O}(10^2)$ runs independently of their magnitude

Remark

- More advanced techniques combine active learning with recent metamodels (*e.g.* PC-Kriging), as well as proper simulation-based algorithms (*e.g.* subset simulation)

Questions ?

Acknowledgements: R. Schöbi

Chair of Risk, Safety & Uncertainty Quantification

www.rsuq.ethz.ch

The Uncertainty Quantification Laboratory

www.uqlab.com

Thank you very much for your attention!

ETH zürich

Risk, Safety & Uncertainty Quantification

BACKUP SLIDES

Matérn autocorrelation function (1D)

Definition

$$R_1(x,x') = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\sqrt{2\,\nu}\, \frac{|x-x'|}{\theta} \right)^{\nu} \kappa_\nu \left(\sqrt{2\,\nu}\, \frac{|x-x'|}{\theta} \right)$$

where $\nu \geq 1/2$ is the shape parameter, θ is the scale parameter, $\Gamma(\cdot)$ is the Gamma function and $\kappa_\nu(\cdot)$ is the modified Bessel function of the second kind

Properties

The values $\nu = 3/2$ and $\nu = 5/2$ are usually used $\left(h = \frac{|x-x'|}{\theta} \right)$:

$$R_1(h;\, \nu = 3/2) = (1 + \sqrt{3}\, h) \exp(-\sqrt{3}\, h)$$

$$R_1(h;\, \nu = 5/2) = (1 + \sqrt{5}\, h + \frac{5}{3}\, h^2) \exp(-\sqrt{5}\, h)$$

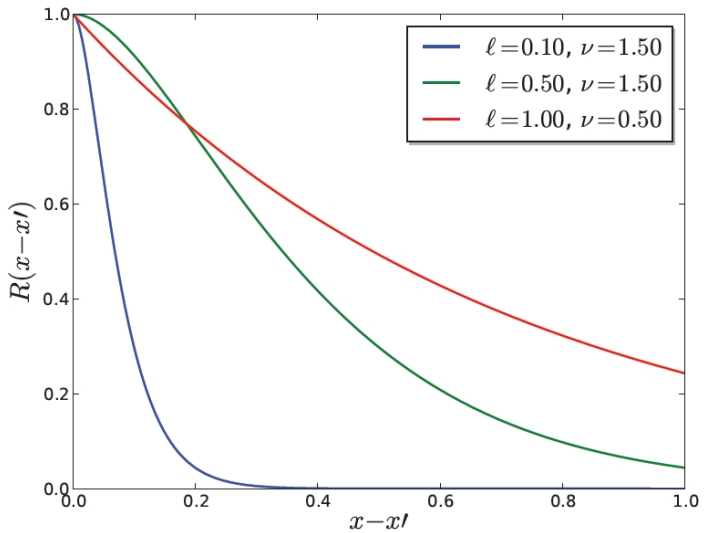
Matérn autocorrelation function

Parameter ν controls the regularity (smoothness) of the trajectories

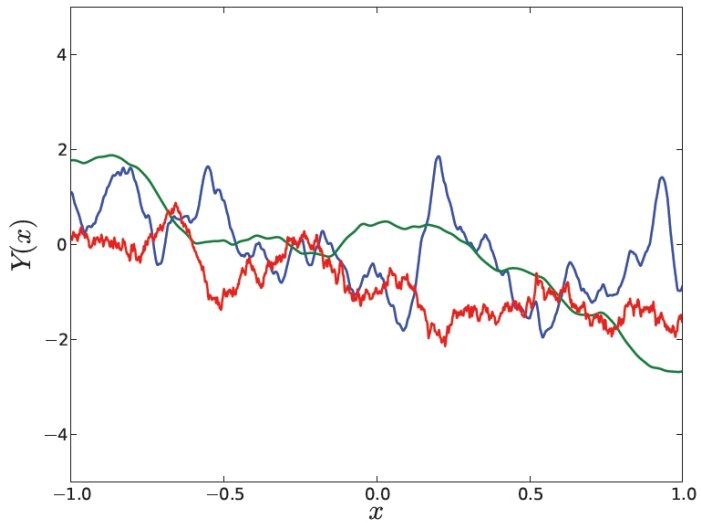
- The trajectories of such a process are $\lfloor \nu \rfloor$ times differentiable:

$$\begin{aligned} \nu = 1/2 \quad &: \quad \mathcal{C}^0 \text{ (continuous, non differentiable)} \\ \nu = 3/2 \quad &: \quad \mathcal{C}^1 \\ \nu = 5/2 \quad &: \quad \mathcal{C}^2 \end{aligned}$$

- When $\nu \to +\infty$, $R_1(h; \nu)$ tends to the square exponential autocorrelation



Autocorrelation function



Trajectories

Kriging variance

- The Kriging variance reads:

$$\sigma^2_{\widehat{Y}_0} = \mathbb{E}\left[(\widehat{Y}_0 - Y_0)^2\right] = \sigma^2 \left(1 - \boldsymbol{r}_0^{\mathsf{T}} \mathbf{R}^{-1} \boldsymbol{r}_0 + \boldsymbol{u}_0^{\mathsf{T}} \left(\mathbf{F}^{\mathsf{T}} \mathbf{R}^{-1} \mathbf{F}\right)^{-1} \boldsymbol{u}_0\right)$$

  with $\boldsymbol{u}_0 = \mathbf{F}^{\mathsf{T}} \mathbf{R}^{-1} \boldsymbol{r}_0 - \boldsymbol{f}_0$

- It is made of two parts:
  - $\sigma^2 \left(1 - \boldsymbol{r}_0^{\mathsf{T}} \mathbf{R}^{-1} \boldsymbol{r}_0\right)$ corresponds to the simple Kriging (when the trend is known)
  - the rest corresponds to the uncertainty due to the estimation of $\boldsymbol{\beta}$ from the data

- The predictor is interpolating the data in the experimental design:

$$\sigma^2_{\widehat{Y}_i} \equiv \sigma^2_{\widehat{Y}(\boldsymbol{x}^{(i)})} = 0 \qquad \forall\, \boldsymbol{x}^{(i)} \in \mathcal{X}$$

PC-Kriging

Schöbi & Sudret, IJUQ (2015); Kersaudy *et al.* , J. Comp. Phys (2015)

Heuristics: Combine polynomial chaos expansions (PCE) and Kriging

- PCE approximates the global behaviour of the computational model
- Kriging allows for local interpolation and provides a local error estimate

Universal Kriging model with a sparse PC expansion as a trend

$$\mathcal{M}(\boldsymbol{x}) \approx \mathcal{M}^{(\mathrm{PCK})}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) + \sigma^2 Z(\boldsymbol{x}, \omega)$$

PC-Kriging calibration

- Sequential PC-Kriging: least-angle regression (LAR) detects a sparse basis, then PCE coefficients are calibrated together with the auto-correlation parameters
- Optimized PC-Kriging: universal Kriging models are calibrated at each step of LAR

PC-Kriging

Schöbi & Sudret, IJUQ (2015); Kersaudy *et al.* , J. Comp. Phys (2015)

Heuristics: Combine polynomial chaos expansions (PCE) and Kriging

- PCE approximates the global behaviour of the computational model
- Kriging allows for local interpolation and provides a local error estimate

Universal Kriging model with a sparse PC expansion as a trend

$$\mathcal{M}(\boldsymbol{x}) \approx \mathcal{M}^{(\mathrm{PCK})}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) + \sigma^2 Z(\boldsymbol{x}, \omega)$$

PC-Kriging calibration

- Sequential PC-Kriging: least-angle regression (LAR) detects a sparse basis, then PCE coefficients are calibrated together with the auto-correlation parameters
- Optimized PC-Kriging: universal Kriging models are calibrated at each step of LAR

Results with PC Kriging