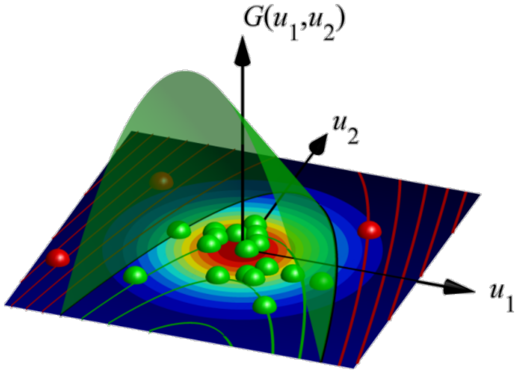
Strengths and limits of reliability assessment methods – Illustration in the field of EMC

Jean-Marc Bourinet

IFMA & Institut Pascal (CNRS UMR 6602)
Clermont-Ferrand, France

Email : bourinet@ifma.fr



Reliability assessment – Rare events

We consider:

✓ a random "object" $\mathbf{X}$ with realizations in a set $\mathcal{X}$

- a random vector (with discrete or/and continuous components)
- a random process (Gaussian, Markov, …)

✓ a scalar real-valued function g, whose entry $\mathbf{x}$ is a realization of $\mathbf{X}$

Note: This function may represent a time and/or space variant problem, *e.g.* expressed in the form: $g(\mathbf{x}) = \max_{t \in [0,T]} g(\mathbf{x}, t)$

We wish to evaluate the probability of the **rare event** corresponding to the following domain of $\mathcal{X}$ ($\equiv$ failure criterion):

$$\mathcal{F}_{\mathbf{x}} = \{\mathbf{x} \in \mathcal{X} : g(\mathbf{x}) \leq 0\}$$

This probability reads: $\mathbb{P}[g(\mathbf{X}) \leq 0] \ll 1$

Some fields of application

- ✓ Risk management of **infrastructures in response to extreme events**: earthquakes, floods, typhoons, hurricanes
- ✓ Prevention of **power system blackouts**
- ✓ **Radioactive waste storage**: How likely is radioactive waste to escape from a repository over the next 10,000 years, depending on unknown aspects of rocks and atmospheric conditions?
- ✓ Packet-switched **telecommunications networks**: Use of buffer of limited size in carrying real-time video. Quantification of the small probabilities of packets loss if the buffers overflow (queuing systems).
- ✓ **Protection of digital contents** (watermarking): Techniques for embedding/hiding information in a digital file (typically audio or video), such that the change is not perceptible and very hard to remove.
- ✓ **Large losses in mathematical finance**: Managers of portfolios of loans need to maintain reserves to protect against rare events involving large losses due to multiple loan defaults.

Time-invariant reliability (a.k.a. static simulation problem)

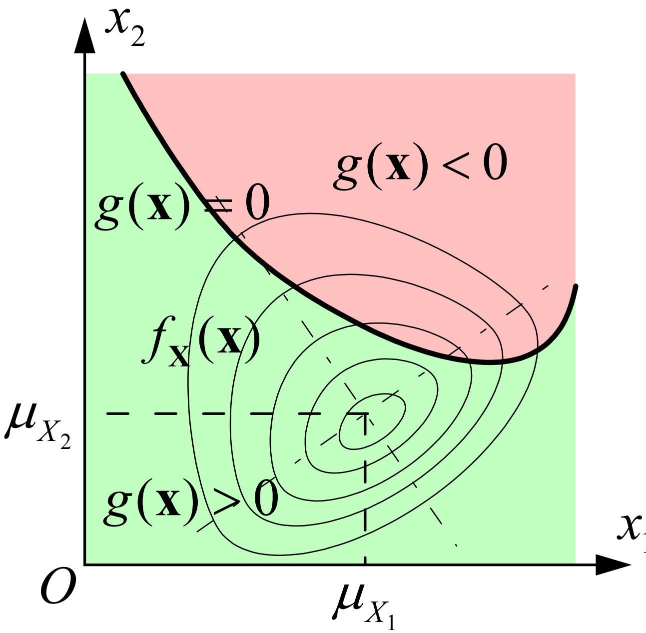
- ✓ $\mathbf{X}$: **continuous random vector** with support $\mathcal{X} \subseteq \mathbb{R}^n$ and joint pdf $f_{\mathbf{X}}$
- ✓ g : **limit-state function** (LSF)
- ✓ $\mathcal{F}_{\mathbf{x}} = \{\mathbf{x} \in \mathcal{X} : g(\mathbf{x}) \le 0\}$: failure domain

  $\overline{\mathcal{F}_{\mathbf{x}}} = \{\mathbf{x} \in \mathcal{X} : g(\mathbf{x}) > 0\}$: safe domain

  $\mathcal{F}_{\mathbf{x}}^0 = \{\mathbf{x} \in \mathcal{X} : g(\mathbf{x}) = 0\}$: limit-state surface (LSS)
- ✓ p_{f} : **failure probability**

$$p_{\mathrm{f}} = \mathbb{P}\left(g(\mathbf{X}) \le 0\right) = \int_{\mathcal{F}_{\mathbf{x}}} f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\,\mathbf{x} = \int_{\mathcal{X}} \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\,\mathbf{x}$$

where: $\mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in \mathcal{F}_{\mathbf{x}} \\ 0 & \text{if } \mathbf{x} \notin \mathcal{F}_{\mathbf{x}} \end{cases}$



Joint probability density function $f_{\mathbf{X}}(\mathbf{x})$

The joint probability density function $f_{\mathbf{X}}$ is rarely accessible.

Statistical information:

✓ **marginal pdfs** f_{X_i} (or cdfs F_{X_i}) of X_i components, for $i = 1, \ldots, n$

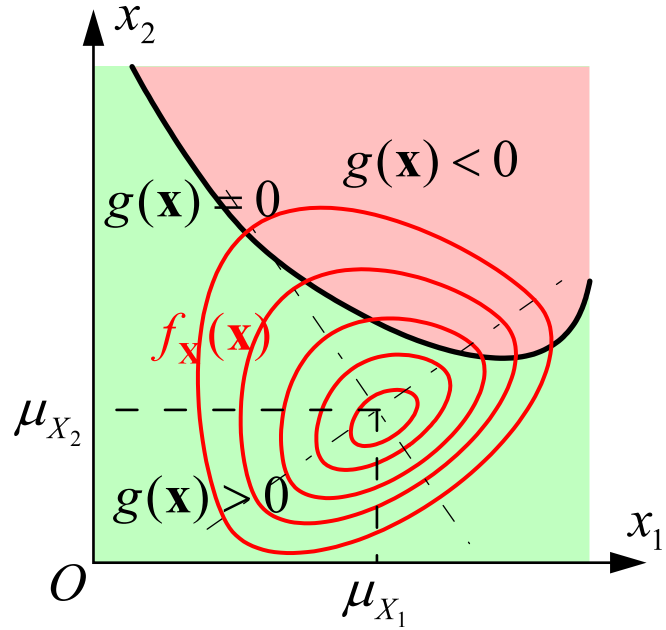
✓ correlation structure between X_i components

- **linear correlation**: $\mathbf{R} = \left[\rho_{ij}\right]_{n \times n}$
- or (much better) **a copula** C such that:

$$F_{\mathbf{X}}(\mathbf{x}) = C\left(F_{X_1}(x_1), \ldots, F_{X_n}(x_n)\right)$$

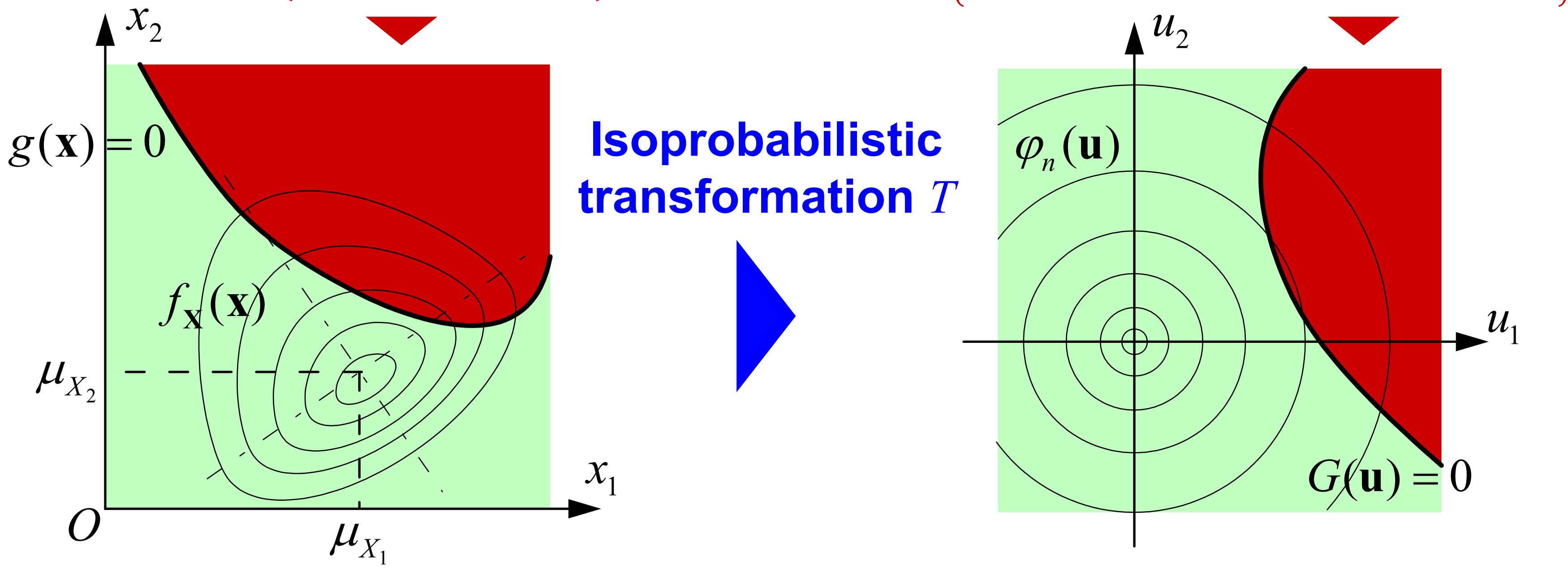
Copula in short ≡ What remains of a joint cdf once the effect of the marginal distributions has been removed.

For continuous marginal distributions, C is unique.



Mapping to a standard normal space

$$\mathcal{F}_{\mathbf{x}} = \left\{ \mathbf{x} \in \mathcal{X} : g(\mathbf{x}) \le 0 \right\} \qquad \mathcal{F}_{\mathbf{u}} = \left\{ \mathbf{u} \in \mathbb{R}^n : (g \circ T^{-1})(\mathbf{u}) = G(\mathbf{u}) \le 0 \right\}$$



$$p_{\mathrm{f}} = \mathbb{P}\big(g(\mathbf{X}) \le 0\big) = \mathbb{P}\big(G(\mathbf{U}) \le 0\big)$$

$$= \int_{\mathcal{F}_{\mathbf{x}}} f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{\mathcal{F}_{\mathbf{u}}} \varphi_n(\mathbf{u})\,\mathrm{d}\mathbf{u}$$

where: $\varphi_n(\mathbf{u}) = \dfrac{1}{(2\pi)^{n/2}} \exp\left(-\dfrac{1}{2} \|\mathbf{u}\|_2^2 \right)$

Some usual transformations:

- Nataf (1962,1986)
- Rosenblatt (1952,1981)

Approximation methods

First-Order Reliability Method (FORM)

Problem statement

Find the most probable failure point (MPFP) P^* solution of:

$$\mathbf{u}^* = \arg\min_{\mathbf{u}\in\mathbb{R}^n} -\frac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{u} \quad \text{subject to} \quad G(\mathbf{u}) = 0$$

$\boldsymbol{\alpha} = -\nabla G(\mathbf{u}^*) / \left\|\nabla G(\mathbf{u}^*)\right\|$: unit vector

$\beta = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{u}^*$: Hasofer-Lind reliability index

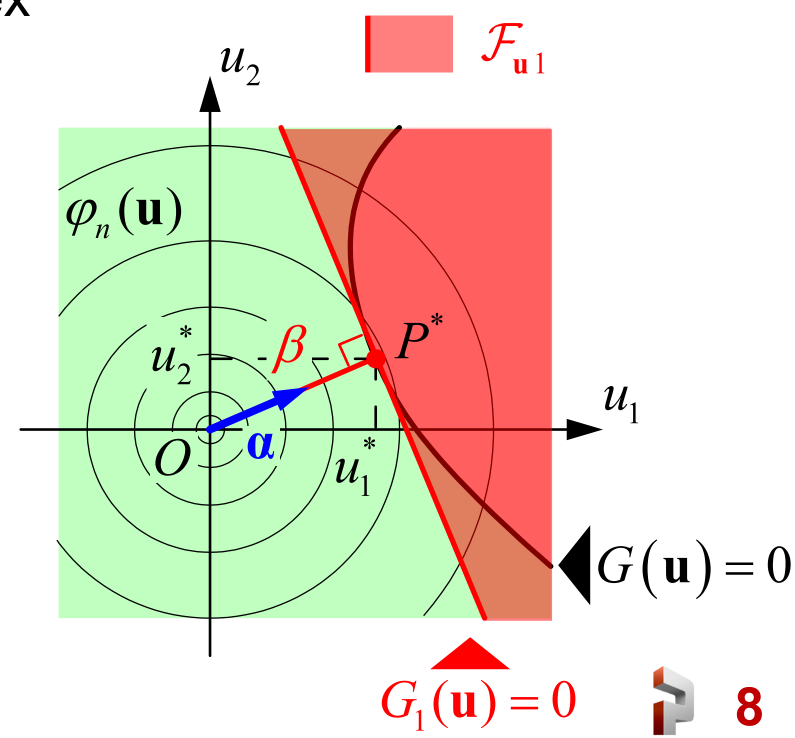
FORM approximation

First-order polynomial of G at P^* :

$$G_1(\mathbf{u}) = G(\mathbf{u}^*) + \nabla G(\mathbf{u}^*)^{\mathrm{T}}(\mathbf{u} - \mathbf{u}^*)$$

where ∇ is the gradient operator.

$$\begin{aligned} p_{\mathrm{f\,FORM}} &= \mathbb{P}\left(G_1(\mathbf{U}) \le 0\right) \\ &= \int_{\mathbb{R}^n} \mathbb{1}_{\mathcal{F}_{\mathbf{u}1}}(\mathbf{u})\, \varphi_n(\mathbf{u})\, \mathrm{d}\mathbf{u} \\ &= \Phi(-\beta) \end{aligned}$$



FORM: strengths and limitations

Strengths

- ✓ Works in several engineering problems of practical interest
- ✓ Computational cost $\propto n$ and rather independent of the failure probability to calculate
- ✓ FORM approximation can be improved by other techniques when the LSS is curved at MPFP (SORM, IS)

Limitations

- ✓ Convergence of the optimization problem tedious or impossible. Main causes: numerical estimation of the gradient of the LSF, intricate shape of the LSS
- ✓ Uncontrolled bias on the failure probability due to nonlinearities of the LSS (can be corrected with SORM) or multiple MPFPs (hard to tackle with FORM)

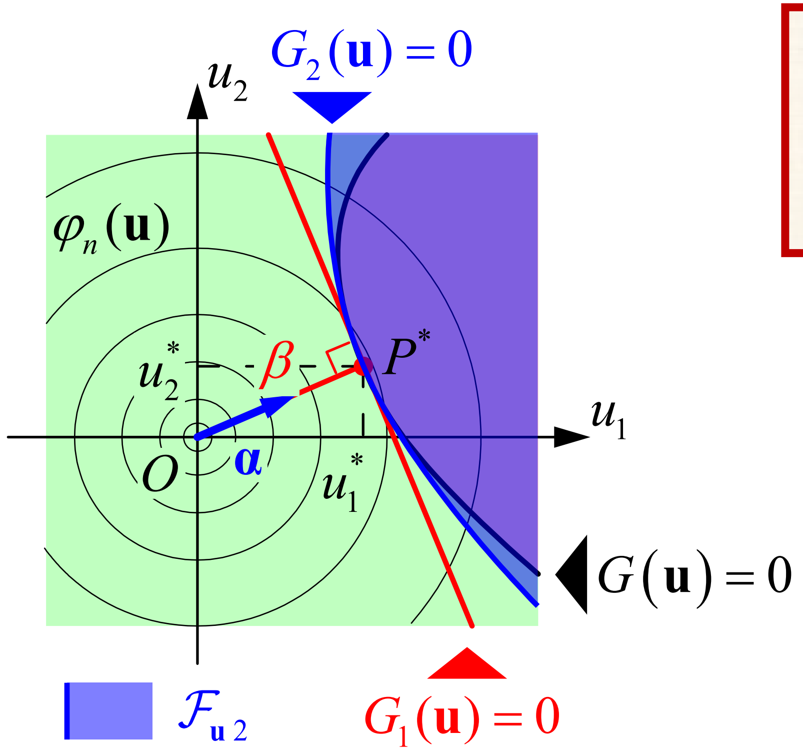
Second-Order Reliability Method (SORM)

SORM approximation

Second-order polynomial of G at P^* (obtained by FORM):

$$G_2(\mathbf{u}) = G(\mathbf{u}^*) + \nabla G(\mathbf{u}^*)^{\mathrm{T}}(\mathbf{u}-\mathbf{u}^*) + \frac{1}{2}(\mathbf{u}-\mathbf{u}^*)^{\mathrm{T}} \nabla^2 G(\mathbf{u}^*)(\mathbf{u}-\mathbf{u}^*)$$

where ∇^2 is the Hessian operator



$$p_{\mathrm{f\,SORM}} = \mathbb{P}\left(G_2(\mathbf{U}) \le 0\right) = \int_{\mathbb{R}^n} \mathbb{1}_{\mathcal{F}_{\mathbf{u}\,2}}(\mathbf{u})\, \varphi_n(\mathbf{u})\, \mathrm{d}\mathbf{u}$$

No closed form solution.

Hohenbichler & Rackwitz approximation:

$$p_{\mathrm{f\,SORM}} \approx \Phi(-\beta) \prod_{i=1}^{n-1} \frac{1}{\sqrt{1+\psi(\beta)\kappa_i}}$$

where κ_i are principal curvatures at P^*

and $\psi(\beta) = \dfrac{\varphi(\beta)}{\Phi(-\beta)}$

SORM: strengths and limitations

Strengths

- ✓ Correction of FORM approximation in case of a single MPFP and a curved LSS

Limitations

- ✓ Computational cost $\propto n^2$ for curvature-fitting SORM.
  An alternative technique known as point-fitting SORM more suitable for high dimensional problems (large n) and slightly noisy LSS
- ✓ Numerical estimation of the Hessian of the LSF by finite differences is often difficult

Sampling methods

Crude Monte Carlo method

Problem statement

$$p_{\mathrm{f}} = \mathbb{P}\left(g(\mathbf{X}) \le 0\right) = \int_{\mathcal{F}_{\mathbf{x}}} f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{\mathcal{X}} \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \mathbb{E}_{f_{\mathbf{X}}}\left[\mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{X})\right]$$

Mean ≡ Mathematical expectation

Crude Monte Carlo estimator

$$p_{\mathrm{f}} = \mathbb{E}_{f_{\mathbf{X}}}\left[\mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{X})\right] \blacktriangleright \widehat{p}_{\mathrm{f\,MC}} = \frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{X}^{(j)})$$

Mean ≡ Sample mean

where $\mathbf{X}^{(j)}, j = 1, \ldots, N$, are a set of N independent and identically distributed samples simulated according to joint pdf $f_{\mathbf{X}}$

Properties of the crude Monte Carlo estimator

- ✓ Unbiased estimator: $\mathbb{E}\left[\widehat{p}_{\mathrm{f\,MC}}\right] = p_{\mathrm{f}}$
- ✓ Coefficient of variation: $\delta_{\widehat{p}_{\mathrm{f\,MC}}} = \frac{\sqrt{\mathrm{Var}\left[\widehat{p}_{\mathrm{f\,MC}}\right]}}{\mathbb{E}\left[\widehat{p}_{\mathrm{f\,MC}}\right]}$ ▶ $\delta_{\widehat{p}_{\mathrm{f\,MC}}} = \sqrt{\frac{1-p_{\mathrm{f}}}{N p_{\mathrm{f}}}}$

Crude MC: strengths and limitations

Strengths

- ✓ No assumption about the LSF
- ✓ Computational cost independent of the dimensionality n

Limitations

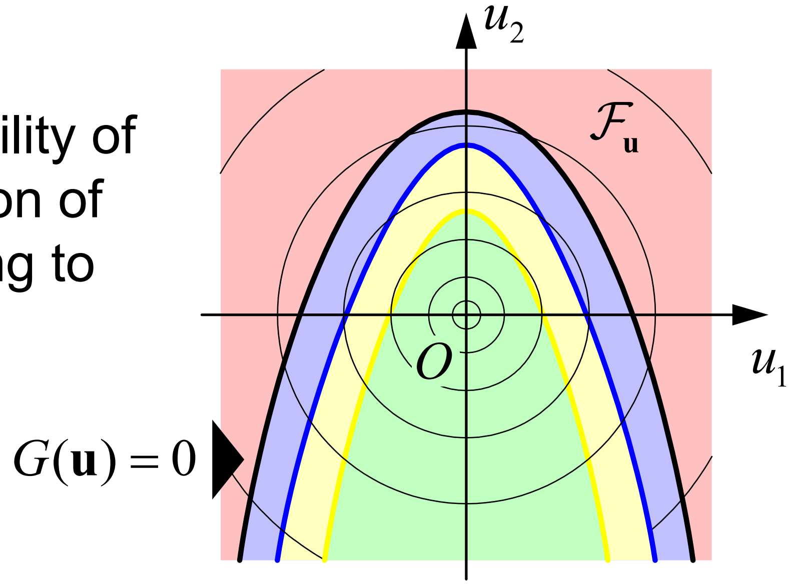
- ✓ Inefficiency for small failure probabilities

INSTITUT PASCAL

Subset simulation (1/3)

Conceptual idea

Replace the estimation of the probability of a rare event by a sequential estimation of conditional probabilities corresponding to less rare intermediate failure events



$$p_f = \mathbb{P}(E)$$
$$= \mathbb{P}(E_m \mid E_{m-1})\mathbb{P}(E_{m-1})$$
$$= \ldots$$

$$p_f = \mathbb{P}(E_m \mid E_{m-1})\mathbb{P}(E_{m-1} \mid E_{m-2})\ldots\mathbb{P}(E_2 \mid E_1)\mathbb{P}(E_1)$$

where $E = \{G(\mathbf{U}) \leq 0\} = E_m$ is the failure event,

and $E_s, s = 1,\ldots,m,$ are events such that: $E = E_m \subset E_{m-1} \subset \ldots \subset E_2 \subset E_1$

Practically, we take $E_s = \{G(\mathbf{U}) \leq y_s\}$ for $s = 1,\ldots,m,$

where $y_1 > y_2 > \ldots > y_m = 0$

Subset simulation (2/3)

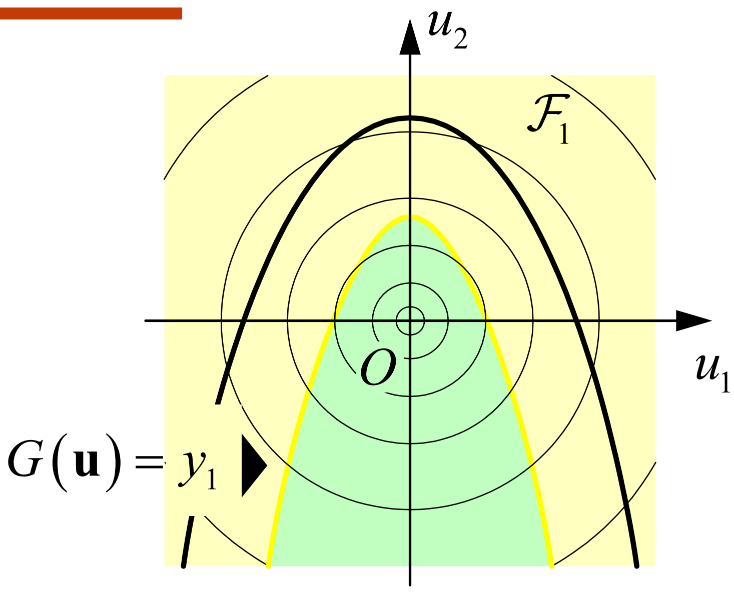
First intermediate level y_1

For $j=1,\dots,N,\ \mathbf{U}^{(j)} \sim \varphi_n$

1st level y_1 : p_0-quantile of $\left(G(\mathbf{u}^{(j)})\right)_{1\le j\le N}$

$$p_1 = \mathbb{P}\left(E_1\right) = \mathbb{E}_{\varphi_n}\left[\mathbb{1}_{\mathcal{F}_1}(\mathbf{U})\right]$$

$$\widehat{p}_1 = \frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\mathcal{F}_1}(\mathbf{u}^{(j)}) = p_0$$



Intermediate levels y_s for $s>1$

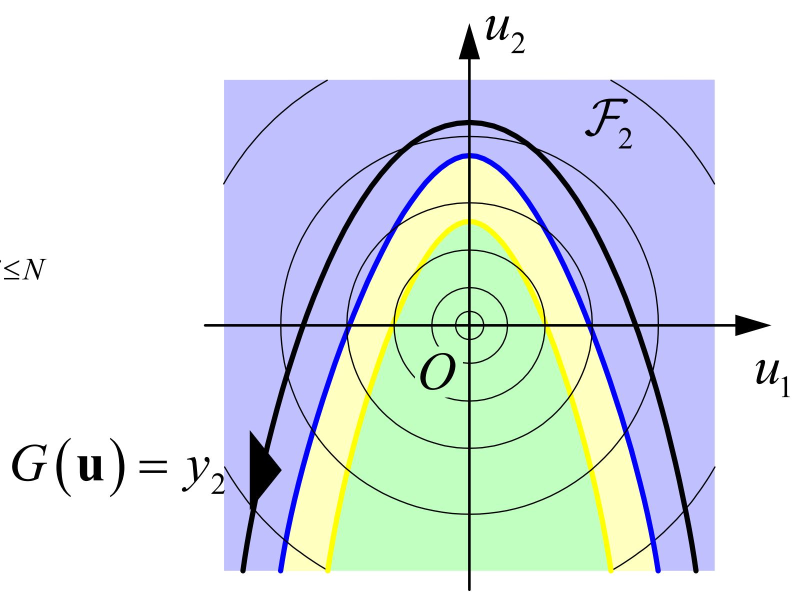
For $j=1,\dots,N,\ \mathbf{U}^{(j)} \sim \varphi_n(\cdot \mid E_{s-1})$

s th level y_s : p_0-quantile of $\left(G(\mathbf{u}^{(j)})\right)_{1\le j\le N}$

If $y_s<0$, then $m=s$ and set $y_s=0$

$$p_s = \mathbb{P}\left(E_s \mid E_{s-1}\right) = \mathbb{E}_{\varphi_n(\cdot\mid E_{s-1})}\left[\mathbb{1}_{\mathcal{F}_s}(\mathbf{U})\right]$$

$$\widehat{p}_s = \frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\mathcal{F}_s}(\mathbf{u}^{(j)}) = p_0 \text{ if } y_s>0$$
$$> p_0 \text{ if } y_s=0$$



Subset simulation (3/3)

Important remarks about subset simulation

- ✓ Conditional MCMC samples generated with the Metropolis-Hastings algorithm modified by Au & Beck (Au & Beck, 2001)
  Main consequence: Generated samples are not independent due to the choice of a proposal pdf
- ✓ Initial states of MCMC follow the target distribution $\varphi_n(\cdot \mid E_{s-1})$
  ▶ No burn in period in the chains
- ✓ Optimal tuning of subset simulation: $p_0 \in [0.1, 0.3]$
  ▶ $\hat{p}_{\mathrm{f\,SS}} = p_0^{m-1} p_m$
- ✓ Subset simulation estimator asymptotically unbiased
  ▶ Practically no significant bias for $N \geq 1000$
- ✓ The coefficient of variation $\delta_{\hat{p}_{\mathrm{f\,SS}}}$ of $\hat{p}_{\mathrm{f\,SS}}$ can be estimated

Subset simulation: strengths and limitations

Strengths

- ✓ High efficiency for small failure probabilities
- ✓ Computational cost independent of the dimensionality n
- ✓ Works well in almost all engineering problems of practical interest

Limitations

- ✓ Still requires several thousands calls to the LSF for a coefficient of variation in the order of a few ten percents

Sensitivity measures in reliability assessment

Sensitivity measures with FORM

Importance factors at MPFP

- ✓ Case of independent inputs

$$\mathrm{Var}\left[G_1(\mathbf{U})/\left\|\nabla G(\mathbf{u}^*)\right\|\right]=\alpha_1^2+\ldots+\boxed{\alpha_i^2}+\ldots+\alpha_n^2=1$$

Contribution from the i th random input

- ✓ Case of linearly correlated inputs

  Importance factors γ_i (Der Kiureghian 2009)

Sensitivities to distribution parameters

$$p_\mathrm{f}=\int_{\mathcal{X}}\mathbb{1}_{\mathcal{F}_\mathbf{x}}(\mathbf{x})\,f_\mathbf{X}(\mathbf{x},\boldsymbol{\theta}_f)\,\mathrm{d}\,\mathbf{x}$$

$\boldsymbol{\theta}_f$: distribution parameters $(\mu, \sigma, p_1, p_2, \ldots)$

$$\nabla_{\boldsymbol{\theta}_f}\beta=\boldsymbol{\alpha}^\mathrm{T}\mathbf{J}_{T,\boldsymbol{\theta}_f}(\mathbf{x}^*,\boldsymbol{\theta}_f)$$

$$\nabla_{\boldsymbol{\theta}_f}p_{\mathrm{f\,FORM}}=-\varphi(\beta)\,\nabla_{\boldsymbol{\theta}_f}\beta$$

Example: sensitivities to means

$$\frac{\partial p_{\mathrm{f\,FORM}}}{\partial \mu_{X_i}} > 0$$ ▶ Sollicitation-type input

$$< 0$$ ▶ Resistance-type input

Sensitivity measures with MC and subset simulation

Sensitivities to distribution parameters of MC and SS based on the **score function**: $\nabla_{\boldsymbol{\theta}_f} \ln f_{\mathbf{X}}(\mathbf{x}, \boldsymbol{\theta}_f)$

Monte Carlo

$$\frac{\partial p_{\mathrm{f}}}{\partial \theta_{fk}} = \int_{\mathcal{X}} \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{x}) \frac{\partial \ln f_{\mathbf{X}}(\mathbf{x}, \theta_{fk})}{\partial \theta_{fk}} f_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x}$$

if $\mathcal{X}$ does not depend on θ_{fk}

$$= \mathbb{E}_{f_{\mathbf{X}}}\left[\mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{X}) \frac{\partial \ln f_{\mathbf{X}}(\mathbf{X}, \theta_{fk})}{\partial \theta_{fk}} \right]$$

$$\widehat{\frac{\partial p_{\mathrm{f}}}{\partial \theta_{fk}}} = \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{X}^{(j)}) \frac{\partial \ln f_{\mathbf{X}}(\mathbf{X}^{(j)}, \theta_{fk})}{\partial \theta_{fk}}$$

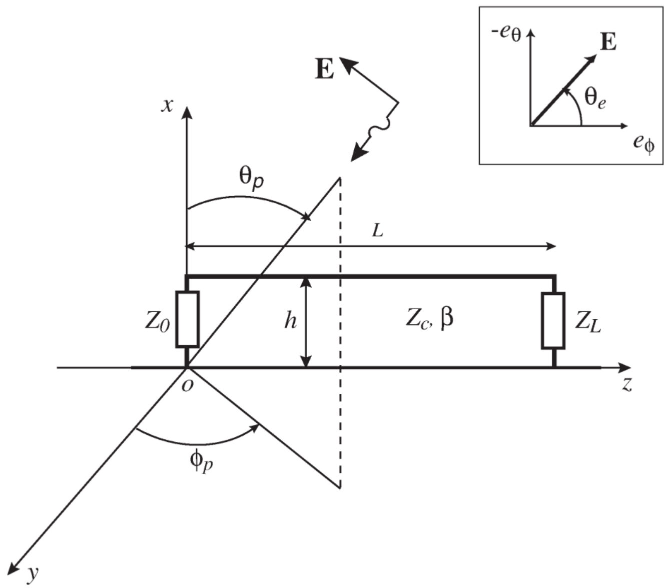
A straightforward postprocessing of a MC analysis

Subset simulation

A similar estimator is available for subset simulation, derived from probabilities $\widehat{p}_s$, $s = 1, \ldots, m$ (Song et al. 2009), see *e.g.* (Dubourg 2011)

Application example
Overcurrent in a transmission line

Model description



Random variable X_i	Mean μ_{X_i}	Coefficient of variation δ_{X_i}	Distribution / Support
$X_1 = L$ (m)	4.2	10%	Lognormal / $\mathbb{R}_+^*$
$X_2 = h$ (m)	0.02	10%	Lognormal / $\mathbb{R}_+^*$
$X_3 = d$ (m)	0.001	5%	Lognormal / $\mathbb{R}_+^*$
$X_4 = Z_L$ (Ω)	1,000	20%	Lognormal / $\mathbb{R}_+^*$
$X_5 = Z_0$ (Ω)	50	5%	Lognormal / $\mathbb{R}_+^*$
$X_6 = a_e$ (V/m)	1	20%	Lognormal / $\mathbb{R}_+^*$
$X_7 = \theta_e$ (rad)	$\pi/4$	57.7%	Uniform / $[0\,;\pi/2]$
$X_8 = \theta_p$ (rad)	$\pi/4$	57.7%	Uniform / $[0\,;\pi/2]$
$X_9 = \phi_p$ (rad)	π	57.7%	Uniform / $[0\,;2\pi[$
$X_{10} = f$ (MHz)	30	9.6%	Uniform / $[25\,;35]$
$X_{11} = \alpha$ (-)	0.0010	28.9%	Uniform / $[0.0005\,;0.0015]$

Limit-state function

$$g(\mathbf{x}) = y_{\text{th}} - r(\mathbf{x})$$

where $r(\mathbf{x})$ is given in closed form (Rannou et al. 2002)

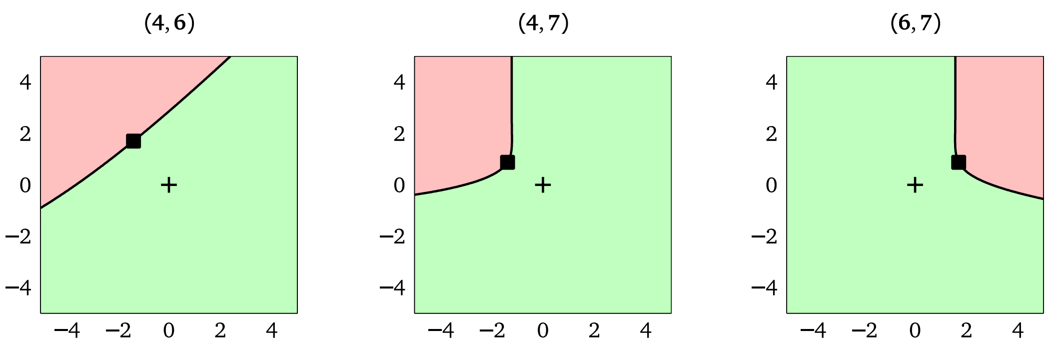
and $y_{\text{th}} = 1.5 \times 10^{-4}\,\text{A}$

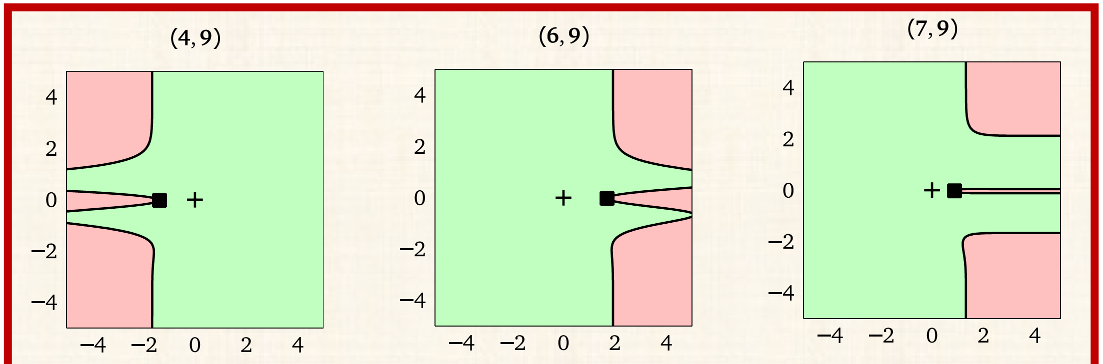
Results (1/4)

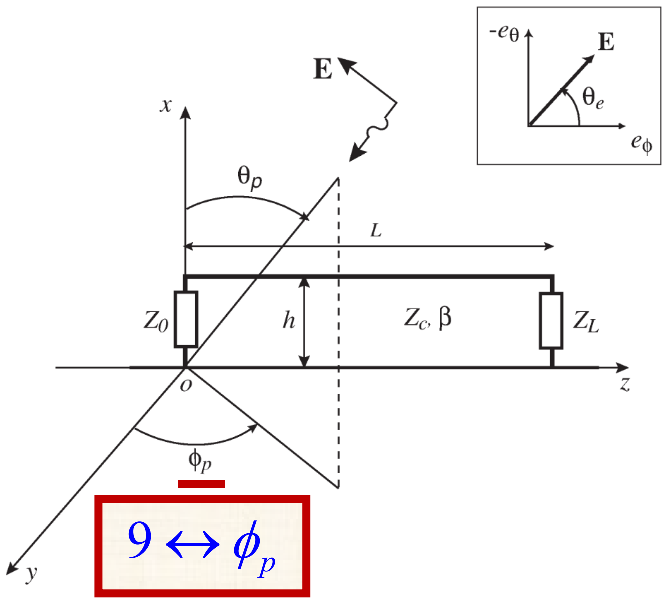
Method	Failure probability (# of calls to LSF)
SS (reference values)	2.27×10^{-4} ($\approx 4 \times 10^6$)
Coefficient of variation of $\widehat{p}_f^{SS}$	1.0%
FORM	31.83×10^{-4} ($9,144$)
SORM^{cf}	1.10×10^{-4} (77)

Results (2/4)

Method	Failure probability (# of calls to LSF)
SS (reference values)	2.27×10^{-4} ($\approx 4 \times 10^{6}$)
Coefficient of variation of $\widehat{p}_{\mathrm{f}}^{\mathrm{SS}}$	1.0%
FORM	31.83×10^{-4} (9,144)
SORM$^{\mathrm{cf}}$	1.10×10^{-4} (77)







Results (3/4)

Alternative method

Solve $p_{\mathrm{f}} = \int_{\mathcal{X}_9} \mathbb{P}\left(E \mid X_9 = x_9\right) f_{X_9}(x_9)\,\mathrm{d}\,x_9$

where $E \mid X_9 = x_9 = \{g(X_1,\ldots,X_8,x_9,X_{10},X_{11}) \leq 0\}$

and $X_9 = \phi_p \sim \mathcal{U}(0,2\pi)$

$$p_{\mathrm{f}} \approx \frac{1}{M}\sum_{j=1}^{M} \mathbb{P}\left(E \mid X_9 = x_{9,j}\right)$$

where $(x_{9,j})_{1\leq j\leq M}$ is a set of M equally-spaced values over $[0,2\pi[$

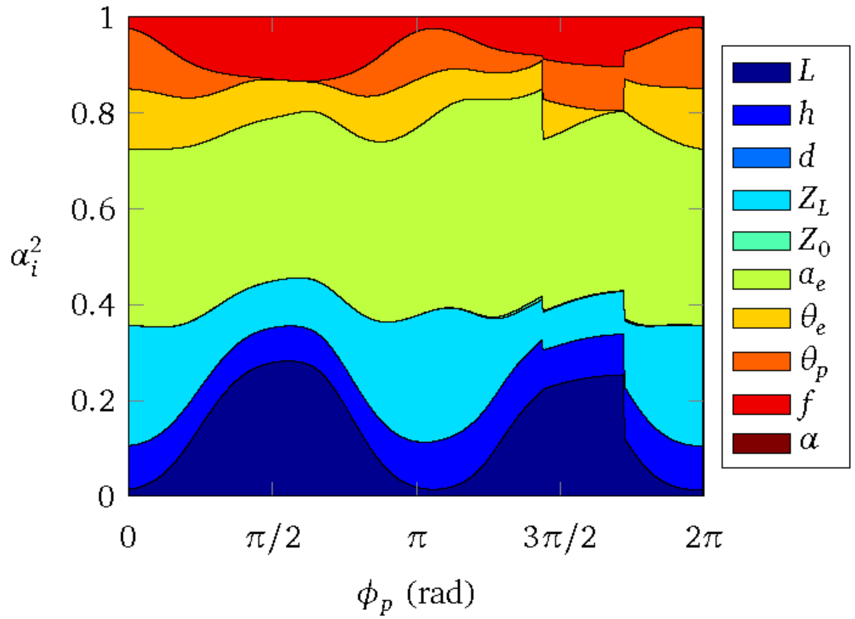
and $\mathbb{P}\left(E \mid X_9 = x_{9,j}\right)$ is determined by FORM and SORM-cf

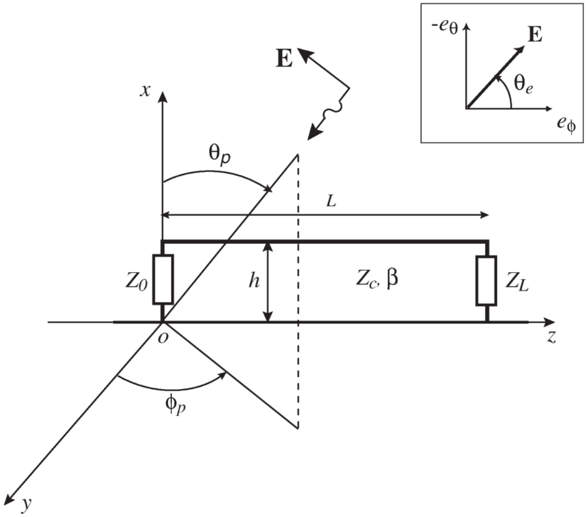
Results (5/4)

Method	Failure probability (# of calls to LSF)
SS (reference values)	2.27×10^{-4} ($\approx 4 \times 10^6$)
Coefficient of variation of $\widehat{p}_f^{SS}$	1.0%
FORM	31.83×10^{-4} (9,144)
SORM^{cf}	1.10×10^{-4} (77)
FORM averaged over ϕ_p	7.42×10^{-4} ($\approx 7.1 \times 10^4$)
SORM^{cf} averaged over ϕ_p	1.95×10^{-4} (6,500)

with $M = 100$

Importance factors





Surrogate models

Learning framework

Input parameter space: $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$ space of data (or patterns)

Output parameter space: $\mathbf{y} \in \mathcal{Y}$ space of responses (or labels)

Scope of the presentation:

- ✓ **Scalar** output y
- ✓ Discrete set $\mathcal{Y}$, *e.g.* $\mathcal{Y} = \{-1; 1\}$
  ▶ **Classification** (or pattern recognition)
- ✓ Continuous set $\mathcal{Y}$: $\mathcal{Y} = \mathbb{R}$
  ▶ **Regression**

Input data

N data pairs: $\mathcal{D} = (\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N) \in \mathcal{X} \times \mathcal{Y}$ ▶ **Training set**

Objective

Construct a model $f(\mathbf{x})$ which predicts the output $\hat{y}$ at any **new** $\mathbf{x} \in \mathcal{X}$

$f: \quad \mathcal{X} \to \mathcal{Y} \qquad \mathbf{x} \mapsto \hat{y} = f(\mathbf{x})$ **(regression)**

$\mathbf{x} \mapsto \hat{y} = \mathrm{sign}\, f(\mathbf{x})$ **(binary classification)**

Optimal model

Construction of an "optimal" model

The model is expected to predict well not only over the training data but also (and more importantly) **over unseen data**.

▶ **Ability of the model to generalize**

f chosen from a known set of candidate functions $\mathcal{F}$ (hypothesis space for SVMs) by minimization of an error criterion defined over the training set (our only source of information)

How (im)perfect is this "optimal" model?

The "optimal" model is chosen from a **set of specified models**
For Kriging: assumption about trend, type of covariance function, …
For SVMs: type of loss function, type of kernel, type of regularization

We do not know if the selected type of model will allow us to obtain an approximation sufficiently close to the real data!

Surrogate of a numerical model

We consider $(y_1, \ldots, y_N) \in \mathcal{Y}^N$ as outputs of a **true model**:

$$f_{\text{true}} : \mathcal{X} \to \mathcal{Y}$$
$$\mathbf{x} \mapsto y = f_{\text{true}}(\mathbf{x})$$

$$f_{\text{true}}(\mathbf{x}) \equiv \text{LSF } G(\mathbf{u})$$

$$p_{\text{f}} = \mathbb{P}(G(\mathbf{U}) \leq 0) = \int_{\mathcal{F}_{\mathbf{u}}} \varphi_n(\mathbf{u})\,\mathrm{d}\mathbf{u} = \mathbb{E}\left[\mathbb{1}_{\mathcal{F}_{\mathbf{u}}}(\mathbf{X})\right]$$

The training set now becomes: $\mathcal{D} = (\mathbf{x}_1, f_{\text{true}}(\mathbf{x}_1)), \ldots, (\mathbf{x}_N, f_{\text{true}}(\mathbf{x}_N))$

The true model is supposed costly to evaluate (*e.g.* finite elements).

► **We want to learn f from a training set $\mathcal{D}$ as small as possible. (small N)**

f is called **surrogate model** (**machine** in statistical learning, **metamodel** or **response surface** in engineering, …),
e.g. polynomial response surfaces, Kriging (a.k.a. GP emulators), PCE, **support vector machines (SVM)**, ANN, …

Support Vector Machines (SVMs)

SVM as a ℓ_2-regularized learning problem (1/2)

$$\min_{h \in \mathcal{H}_k,\, b \in \mathbb{R}} \; C \sum_{i=1}^{N} \ell\left(y_i, h(\mathbf{x}_i) + b\right) + \frac{1}{2} \|h\|_{\mathcal{H}_k}^2$$

where $\ell : \mathcal{Y} \times \mathbb{R} \rightarrow \mathbb{R}^+$ is a **loss function** (convex w.r.t. second parameter but not necessarily differentiable),

and $C > 0$ is a **regularization constant**.

- The **first term** enforces the fit of the SVM model $f(\mathbf{x}) = h(\mathbf{x}) + b$ to the training data for the problem of interest (classification or regression)
- The **second term** enforces a small norm $\|h\|_{\mathcal{H}_k}$ in the RKHS which results in a sufficiently smooth solution (regularization effect)

SVM as a ℓ_2-regularized learning problem (2/2)

$$\min_{h \in \mathcal{H}_k,\, b \in \mathbb{R}} \; C \sum_{i=1}^{N} \ell\big(y_i, h(\mathbf{x}_i) + b\big) + \frac{1}{2} \|h\|_{\mathcal{H}_k}^2$$

Solution h in the form $h(\mathbf{x}) = \sum_{i=1}^{N} c_i k(\mathbf{x}_i, \mathbf{x})$

where $\mathbf{c} = (c_1, \ldots, c_N)$ is a vector of unknown expansion coefficients

and k is a positive definite kernel (reproducing kernel of the RKHS)

- $\|h\|_{\mathcal{H}_k}^2 = \langle h, h \rangle_{\mathcal{H}_k} = \mathbf{c}^{\mathrm{T}} \mathbf{K} \mathbf{c}$

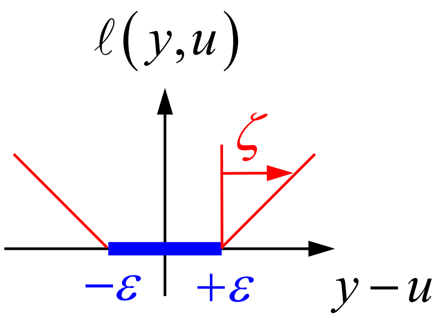
  where $\mathbf{K} = \big(k(\mathbf{x}_i, \mathbf{x}_j)\big)_{i,j=1}^{N}$ is the **Gram (or kernel) matrix**

- $h(\mathbf{x}_i) = \sum_{j=1}^{N} c_j k(\mathbf{x}_j, \mathbf{x}_i) = (\mathbf{K}\mathbf{c})_i$

INSTITUT PASCAL

ε–insensitive loss function and Gaussian RBF kernel

ε–insensitive loss function



No penalization of deviations lower than ε, linear penalization otherwise

$\ell(y,u) = (|y-u| - \varepsilon)_+$ where $(x)_+ = \max(x,0)$

Gaussian RBF Kernel

$$k(\mathbf{x},\mathbf{x}') = \exp\left(-\gamma \|\mathbf{x}-\mathbf{x}'\|^2\right)$$ where $\gamma \in \mathbb{R}_+^*$

Most usual form: $k(\mathbf{x},\mathbf{x}') = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}\right)$ σ : **bandwidth parameter**

Remarks:

- ✓ Stationary, isotropic (or radial)
- ✓ Good pick for smooth decision functions
- ✓ One single parameter to tune
- ✓ Most widely used kernel with SVMs

L1–ε–SVR, dual optimization problem and solution

Dual optimization problem

$$\max_{\boldsymbol{\alpha},\boldsymbol{\alpha}^*} \; -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)\boxed{k(\mathbf{x}_i,\mathbf{x}_j)}+\sum_{i=1}^{N}y_i\left(\alpha_i-\alpha_i^*\right)-\varepsilon\sum_{i=1}^{N}\left(\alpha_i+\alpha_i^*\right)$$

subject to
$$\begin{cases} \sum_{i=1}^{N}\left(\alpha_i-\alpha_i^*\right)=0 \\ 0\le\alpha_i \boxed{\le C} \quad \text{for } i=1,\dots,N \\ 0\le\alpha_i^* \boxed{\le C} \quad \text{for } i=1,\dots,N \end{cases}$$

SVR model

$$f(\mathbf{x})=\left(\sum_{i\in\mathcal{S}\cup\mathcal{S}^*}\left(\alpha_i-\alpha_i^*\right)\boxed{k(\mathbf{x}_i,\mathbf{x})}\right)+b$$

where $\mathcal{S}=\{i:0<\alpha_i<C\}$ and $\mathcal{S}^*=\{i:0<\alpha_i^*<C\}$
($\mathcal{S}\cup\mathcal{S}^*$ represents the set of indices of **support vectors**)

and b is conveniently obtained as a byproduct of any interior point optimization method.

Hyperparameter selection

Optimal selection of hyperparameters

Hyperparameters

Hyperparameters = Kernel parameter(s) + Other model parameters

$$\boldsymbol{\theta}_{\mathrm{k}} = (\gamma) \qquad \boldsymbol{\theta}_{\mathrm{m}} = (C, \varepsilon)$$

Objective

$$\left(\boldsymbol{\theta}_{\mathrm{m}}^{*}, \boldsymbol{\theta}_{\mathrm{k}}^{*}\right) = \underset{\boldsymbol{\theta}_{\mathrm{m}}, \boldsymbol{\theta}_{\mathrm{k}}}{\arg\min}\ \widehat{\mathrm{Err}}_{\mathrm{gen}}\left(\boldsymbol{\theta}_{\mathrm{m}}, \boldsymbol{\theta}_{\mathrm{k}}\right)$$

where $\widehat{\mathrm{Err}}_{\mathrm{gen}}(\boldsymbol{\theta}_{\mathrm{m}}, \boldsymbol{\theta}_{\mathrm{k}})$ is a “sufficiently” accurate approximation of the **generalization error** for given values of parameters $\boldsymbol{\theta}_{\mathrm{m}}$ and $\boldsymbol{\theta}_{\mathrm{k}}$.

Note: The exact error (called true risk in SVMs) is not accessible!

Two key issues

- ✓ What can we take for $\widehat{\mathrm{Err}}_{\mathrm{gen}}(\boldsymbol{\theta}_{\mathrm{m}}, \boldsymbol{\theta}_{\mathrm{k}})$?
- ✓ How to efficiently solve the optimization problem?

Approximation of the generalization error

Objective

$$\left(\boldsymbol{\theta}_{\mathrm{m}}^{*}, \boldsymbol{\theta}_{\mathrm{k}}^{*}\right) = \underset{\boldsymbol{\theta}_{\mathrm{m}}, \boldsymbol{\theta}_{\mathrm{k}}}{\arg\min}\ \widehat{\mathrm{Err}}_{\mathrm{gen}}\left(\boldsymbol{\theta}_{\mathrm{m}}, \boldsymbol{\theta}_{\mathrm{k}}\right)$$

Two key issues

- ✓ **What can we take for $\widehat{\mathrm{Err}}_{\mathrm{gen}}\left(\boldsymbol{\theta}_{m}, \boldsymbol{\theta}_{k}\right)$?**
- ✓ How to efficiently solve the optimization problem?

Approximations used for $\widehat{\mathrm{Err}}_{\mathrm{gen}}\left(\boldsymbol{\theta}_{m}, \boldsymbol{\theta}_{k}\right)$

- ✓ **Span approximation of LOO error for ε-SVR (Chang & Lin 2005)**

$$\mathrm{Err}_{\mathrm{LOO},\,\ell_1} = \frac{1}{N}\sum_{i=1}^{N} \ell_1\left(y_i, f^{(-i)}(\mathbf{x}_i)\right) = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - f^{(-i)}(\mathbf{x}_i)\right|$$

where $\ell_1(y,u) = |y-u|$ (ℓ_1 **loss function**)

$$\widehat{\mathrm{Err}}_{\mathrm{gen}}\left(\boldsymbol{\theta}_{\mathrm{m}}, \boldsymbol{\theta}_{\mathrm{k}}\right) = \mathrm{Err}_{\mathrm{LOO},\,\ell_1} \approx \varepsilon + \frac{1}{N}\sum_{i\in\mathcal{S}_{\mathcal{U}}}\left(\alpha_i + \alpha_i^{*}\right)S_i^2 + \frac{1}{N}\sum_{i\in\mathcal{S}_{\mathcal{U}}}\left(\xi_i + \xi_i^{*}\right)$$

Span approximation for L1–ε–SVR LOO error

- **LOO error used for ε-SVR (Chang & Lin 2005)**

$$\mathrm{Err}_{\mathrm{LOO},\,\ell_1} = \frac{1}{N}\sum_{i=1}^{N} \ell_1\left(y_i, f^{(-i)}(\mathbf{x}_i)\right) = \frac{1}{N}\sum_{i=1}^{N} \left|y_i - f^{(-i)}(\mathbf{x}_i)\right|$$

where $\ell_1(y,u) = |y-u|$ (ℓ_1 **loss function**)

- **Span of SV $\mathbf{x}_i$**

$$S_i^2 = \frac{1}{\left(\widetilde{\mathbf{K}}_{\mathcal{S}_\mathcal{U}}^{-1}\right)_{ii}} \quad \text{where} \quad \widetilde{\mathbf{K}}_{\mathcal{S}_\mathcal{U}} = \begin{bmatrix} \mathbf{K}_{\mathcal{S}_\mathcal{U}} & \mathbf{1}_{|\mathcal{S}_\mathcal{U}|\times 1} \\ \mathbf{1}_{|\mathcal{S}_\mathcal{U}|\times 1}^{\mathrm{T}} & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{K}_{\mathcal{S}_\mathcal{U}} = \left(k(\mathbf{x}_i, \mathbf{x}_j)\right)_{i,j\in\mathcal{S}_\mathcal{U}}$$

where $\mathcal{S}_\mathcal{U} = \left\{ i \;:\; 0 < \alpha_i + \alpha_i^* < C \right\}$ for L1-ε-SVR

- **Span approximation** (L1-ε-SVR)

$$\widehat{\mathrm{Err}_{\mathrm{LOO},\,\ell_1}} = \varepsilon + \frac{1}{N}\sum_{i\in\mathcal{S}_\mathcal{U}} \left(\alpha_i + \alpha_i^*\right) S_i^2 + \frac{1}{N}\sum_{i\in\mathcal{S}_\mathcal{U}} \left(\xi_i + \xi_i^*\right)$$

Assumption:

$$\mathcal{S}_\mathcal{U}^{(-i)} = \mathcal{S}_\mathcal{U}$$

Finding optimal values for hyperparameters

Objective

$$\left(\boldsymbol{\theta}_m^*, \boldsymbol{\theta}_k^*\right) = \underset{\boldsymbol{\theta}_m, \boldsymbol{\theta}_k}{\arg\min}\ \widehat{\mathrm{Err}}_{gen}\left(\boldsymbol{\theta}_m, \boldsymbol{\theta}_k\right)$$

Two key issues

- ✓ What can we take for $\widehat{\mathrm{Err}}_{gen}\left(\boldsymbol{\theta}_m, \boldsymbol{\theta}_k\right)$?
- ✓ **How to efficiently solve the optimization problem?**

Techniques used for solving the optimization problem

- ✓ Deterministic search strategies (grid search most often)
- ✓ **Stochastic optimization methods**
  - Simulated annealing (Pai & Hong 2006, Lin et al. 2008),
  - Genetic algorithms (Pai 2006, Chen 2007),
  - Particle swarm optimization (Lin et al. 2008, Fei et al. 2009),
  - Particle filter (Wei et al. 2013),
  - **Cross-entropy method**.

Reliability assessment based on Support Vector Machines

Adaptive surrogate models for reliability assessment

Problem setup

- ✓ Training points with calls to the LSF: $\left(N_{\mathrm{a},s}\right)_{s=1}^{s_{\max}} = \left(N_{\mathrm{a},1}, N_{\mathrm{a},2}, \ldots N_{\mathrm{a},s_{\max}}\right)$
- ✓ **Sequence of surrogate models:** $\left(f_s\right)_{s=1}^{s_{\max}} = \left(f_1, f_2, \ldots f_{s_{\max}}\right)$
- ✓ Probability estimates: $\left(p_s\right)_{s=1}^{s_{\mathrm{final}}} = \left(p_1, p_2, \ldots p_{s_{\mathrm{final}}} = p_{\mathrm{f\,ASVR}}\right)$
- ✓ Cumulative number of points: $N_{\mathrm{t}} = \sum_{i=1}^{s_{\max}} N_{\mathrm{a},s}$

Prior SVM-based adaptive strategies for RA

- ✓ SVMs are far less common than Kriging for function approximation
- ✓ Earliest references about reliability assessment based on SVMs (Rocco & Moreno 2002, Hurtado & Alvarez 2003)
- ✓ Classification more used than regression
- ✓ Gaussian RBF kernel widely used, polynomial kernel in some works, *e.g.* (Basudhar & Missoum 2008)
- ✓ **Very few details about hyperparameter selection** (which is the key point!)

Adaptive Support Vector Regression (ASVR)

- ✓ Sampling and training **in the standard space**
- ✓ Failure probability estimate: $p_{\mathrm{f\,ASVR}} = \mathbb{E}_{\varphi_n}\left[\mathbb{1}_{\widetilde{\mathcal{F}}_{\mathbf{u}}}\left(\mathbf{U}\right)\right]$

  where $\widetilde{\mathcal{F}}_{\mathbf{u}} = \left\{\mathbf{u} \in \mathbb{R}^n : \widetilde{G}_{s_{\max}}\left(\mathbf{u}\right) \leq 0\right\}$
- ✓ Learning intermediate LSFs with high accuracy is a waste of time

  ► **Fast exploration with MCMC training samples which progressively reach and populate the failure domain**
- ✓ Only the sign of $G\left(\mathbf{u}\right)$ is used in SVC, the exact value is not accounted for (loss of information)

  ► Use of L1-ε-SVR (**regression**) in ASVR method
- ✓ QP solved by an **interior point method**

  ► High accuracy on α_i and α_i^*
- ✓ True k-fold cross validation avoided

  ► Use of the **span approximation of the LOO error**
- ✓ Inaccuracy of grid search for hyperparameter selection

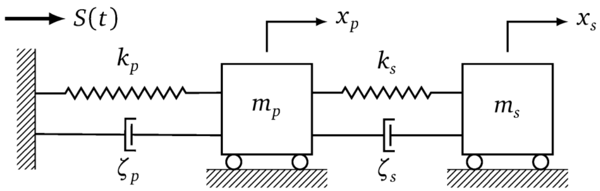
  ► **Stochastic search with the cross entropy-method**

Application examples

Example 1 (from Der Kiureghian & de Stefano 1990)

Limit-state function

$$g(\mathbf{x}) = F_s - 3k_s\sqrt{\frac{\pi S_0}{4\zeta_s\omega_s^3}\left[\frac{\zeta_a\zeta_s}{\zeta_p\zeta_s(4\zeta_a^2+\theta^2)+\gamma\zeta_a^2}\frac{\left(\zeta_p\omega_p^3+\zeta_s\omega_s^3\right)\omega_p}{4\zeta_a\omega_a^4}\right]}$$



where $\omega_p = \sqrt{k_p/m_p}$, $\omega_s = \sqrt{k_s/m_s}$, $\omega_a = (\omega_p + \omega_s)/2$, $\zeta_a = (\zeta_p + \zeta_s)/2$, $\gamma = m_s/m_p$ and $\theta = (\omega_p - \omega_s)/\omega_a$.

Random variables (8)

Variable	m_p	m_s	k_p	k_s	ζ_p	ζ_s	F_s	S_0
Distribution	Lognormal							
Mean	1.5	0.01	1	0.01	0.05	0.02	[15.0–27.5]	100
c.o.v.	0.1	0.1	0.2	0.2	0.4	0.5	0.1	0.1

Previous results (Bourinet et al. 2011)

	FORM[a] $e_1 = 1 \times 10^{-3}$ $e_2 = 5 \times 10^{-3}$	SORM	SS 3×10^5/step $\times 500$	[2]SMART 573/step $\times 50$
μ_{F_s}	P_f $N_{\text{call}}(N_{\text{iter}})$	P_f N_{call}	P_f N_{call} (emp. c.o.v.)	P_f N_{call} (emp. c.o.v.)
27.5	3.91×10^{-6} 2727 (303)[b]	3.70×10^{-7} (2727+)44	$\mathbf{3.78 \times 10^{-7}}$ 21×10^5 (0.040)	3.66×10^{-7} 4011 (0.096)

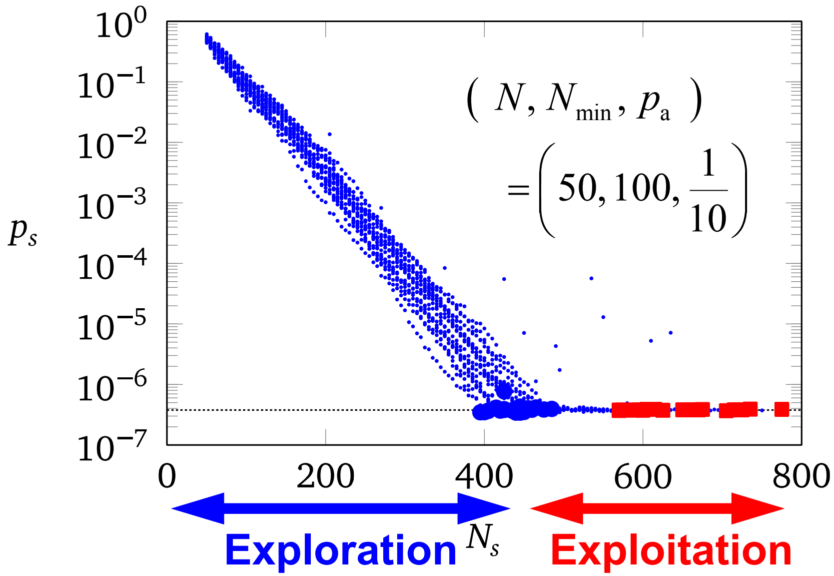
Reference value in bold characters.
[a] Details on e_1 and e_2 convergence criteria in [22].
[b] Stepsize = 0.025 (Armijo's rule otherwise).

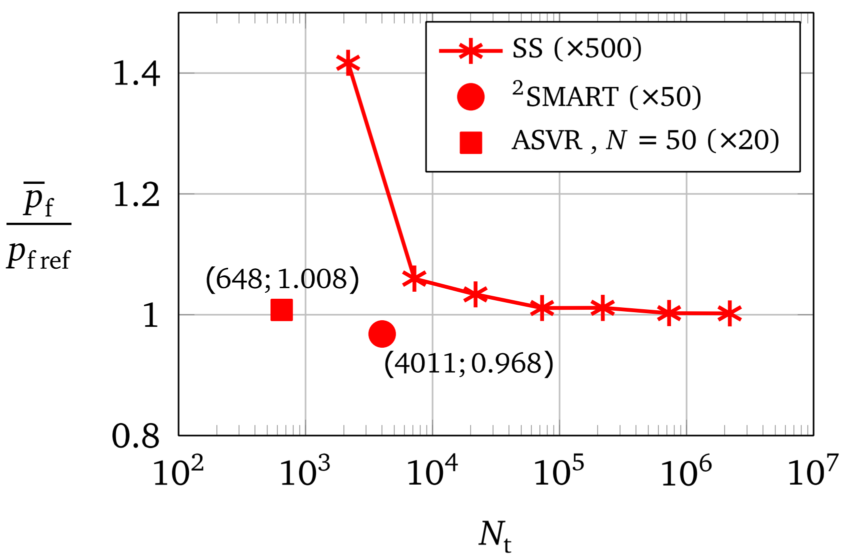
Example 1 (from Der Kiureghian & de Stefano 1990)

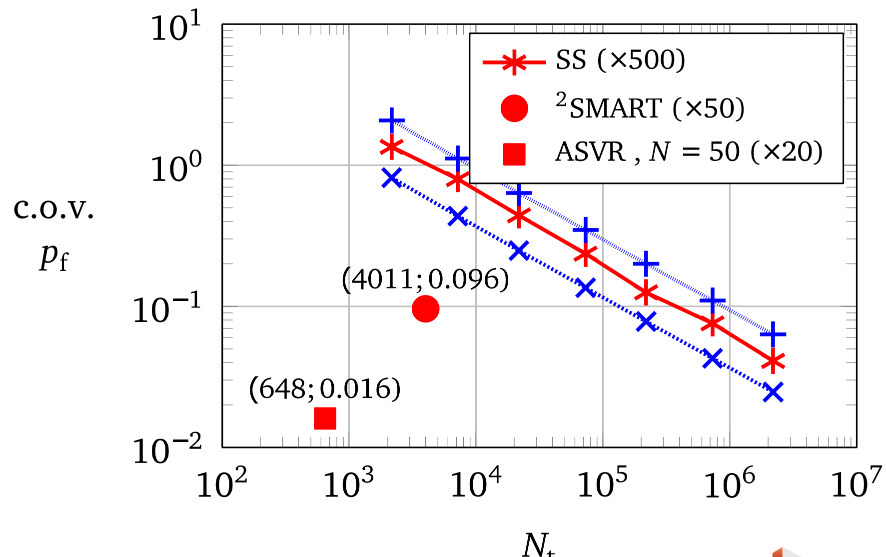
Main characteristics of the reliability assessment problem

- ✓ Rare event: $p_{\mathrm{f\,ref}} = 3.78 \times 10^{-7}$
- ✓ Single MPFP (but very hard to find with FORM!)
- ✓ Highly curved limit-state surface
- ✓ Variable sensitivities at MPFP are different from those at mean

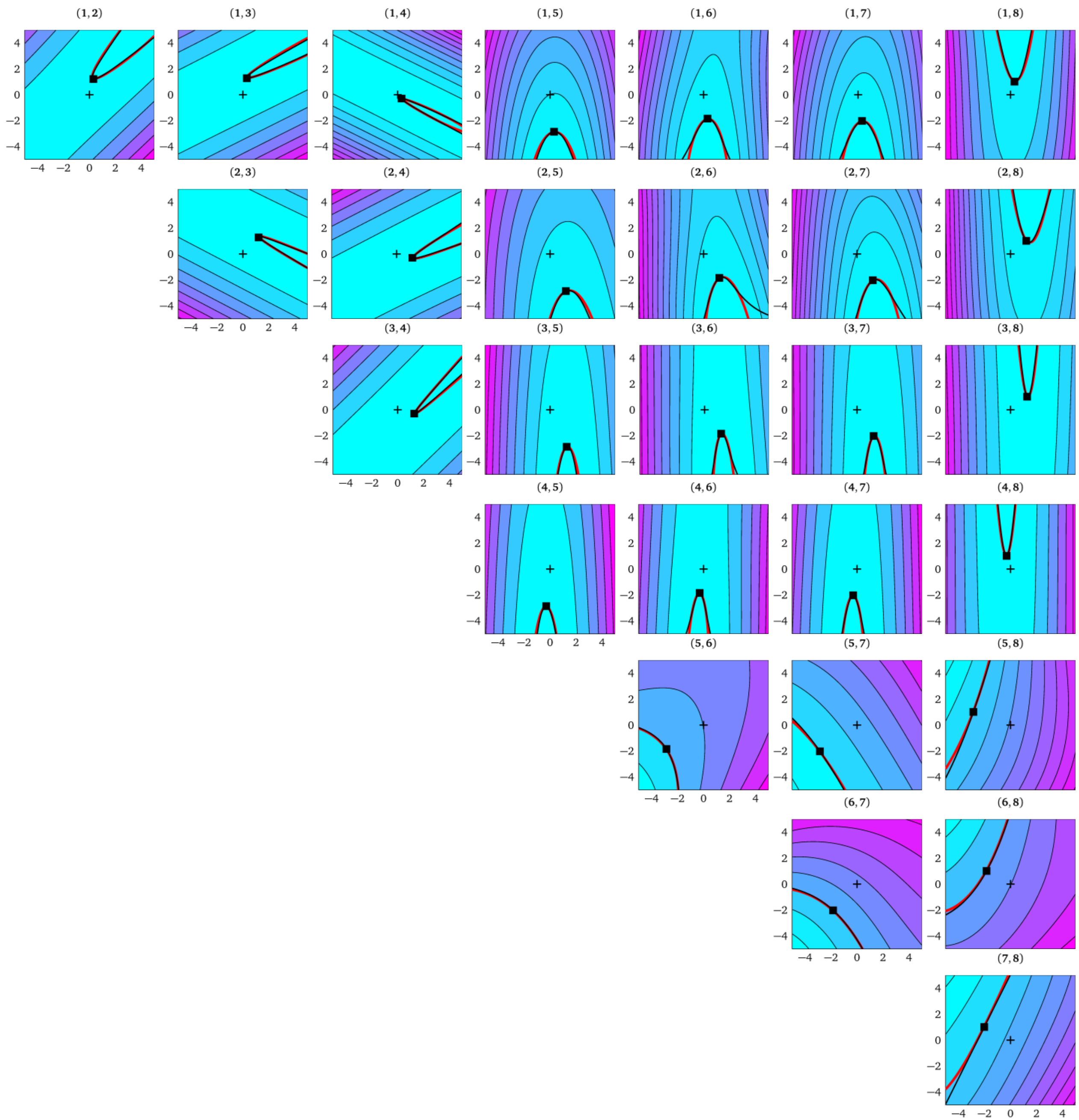
Failure probability estimates







Shape of the limit-state surface



Example 2 (Rackwitz 2001)

- **Limit-state function** $$g(\mathbf{x}) = g(x_1, \ldots, x_n) = (n + a\sigma\sqrt{n}) - \sum_{i=1}^{n} x_i$$
- **Random variables ($n = 100 , 250$)**
  where X_i, $i = 1, \ldots, n$, are i.i.d. lognormal random variables, with unit means and standard deviations equal to $\sigma = 0.2$. We take here $a = 3$.
- **Previous results (Bourinet et al. 2011)**

	FORM $e_1 = 1 \times 10^{-3}$ $e_2 = 1 \times 10^{-3}$	SORM	SS $\times 500$	^{2}SMART $\times 20$
n	P_f $N_{call}(N_{iter})$	P_f N_{call}	P_f N_{call} (emp. c.o.v.)	P_f N_{call} (emp. c.o.v.)
100	4.20×10^{-5} 315 (3)	2.97×10^{-3} (315+)5150	100,000/step $\mathbf{1.73 \times 10^{-3}}$ 300,000 (0.023)	2012/step 1.74×10^{-3} 6036 (0.022)
250	2.82×10^{-6} 765 (3)	5.77×10^{-3} (765+)31,625	100,000/step $\mathbf{1.59 \times 10^{-3}}$ 300,000 (0.023)	3569/step 1.61×10^{-3} 10,707 (0.025)

Reference value in bold characters.

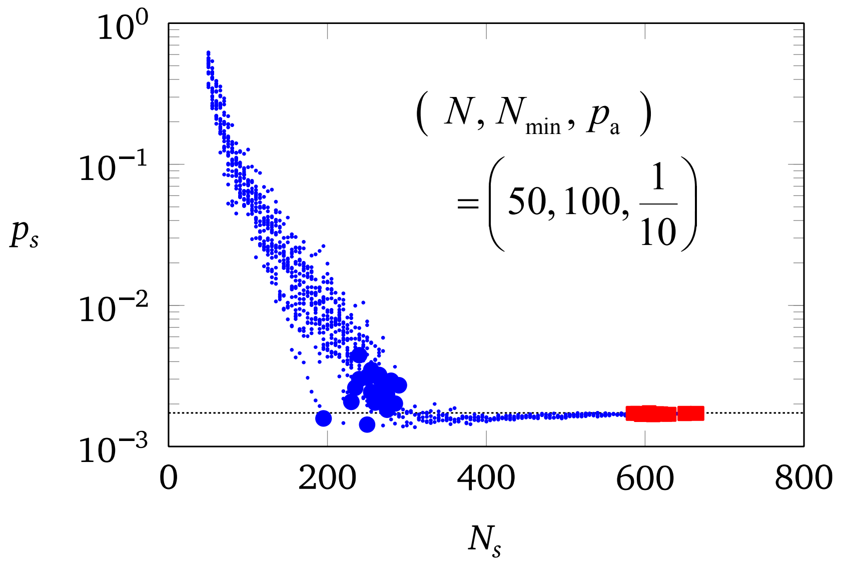
INSTITUT PASCAL

Example 2 (Rackwitz 2001)

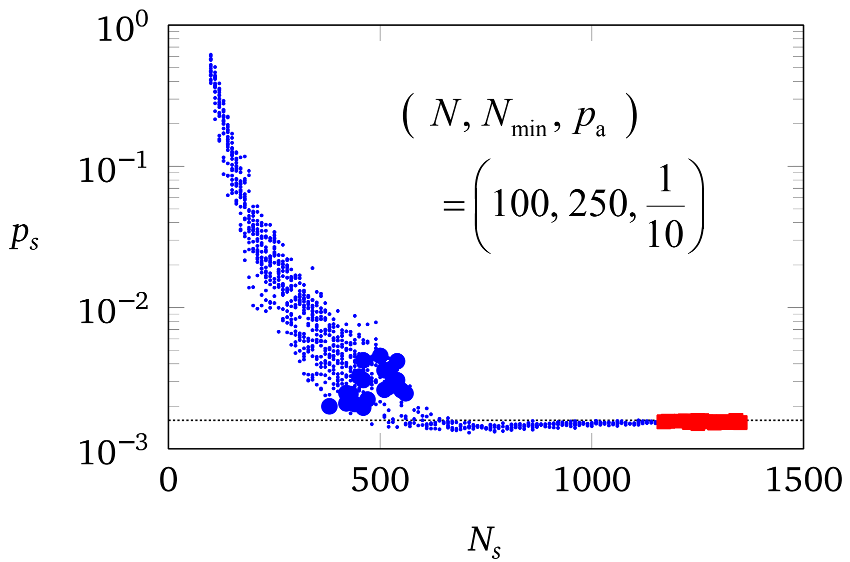
Main characteristics of the reliability assessment problem

- ✓ $p_{\mathrm{f\,ref}} = 1.73 \times 10^{-3}$ ($n = 100$) , $p_{\mathrm{f\,ref}} = 1.59 \times 10^{-3}$ ($n = 250$)
- ✓ Single MPFP
- ✓ High dimension ($n = 100$ and $n = 250$)
- ✓ Smooth limit-state surface
- ✓ All variables of equal importance

Failure probability estimates

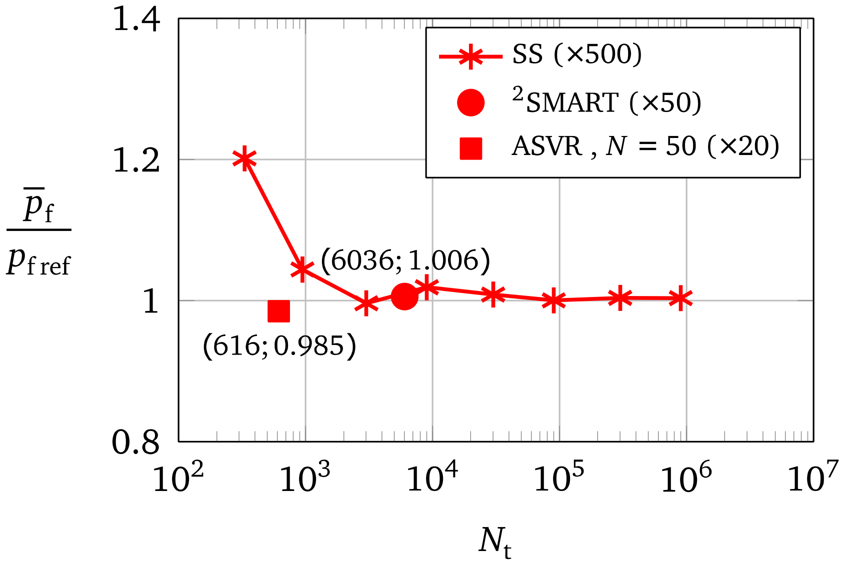


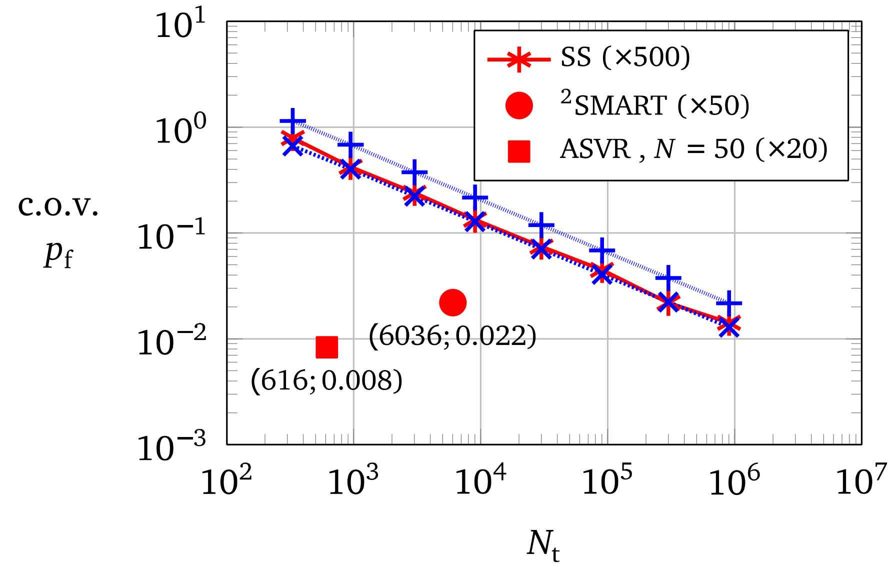
$n = 100$

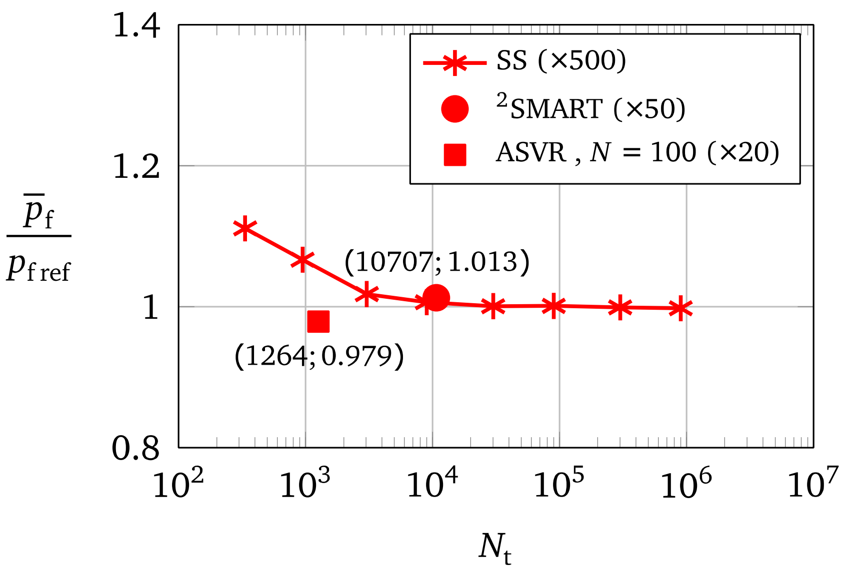


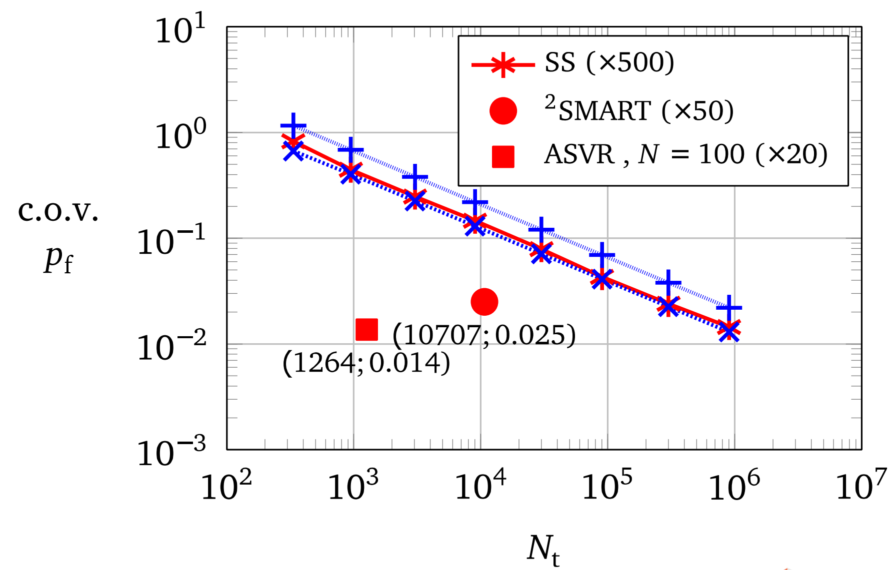
$n = 250$

Failure probability estimates





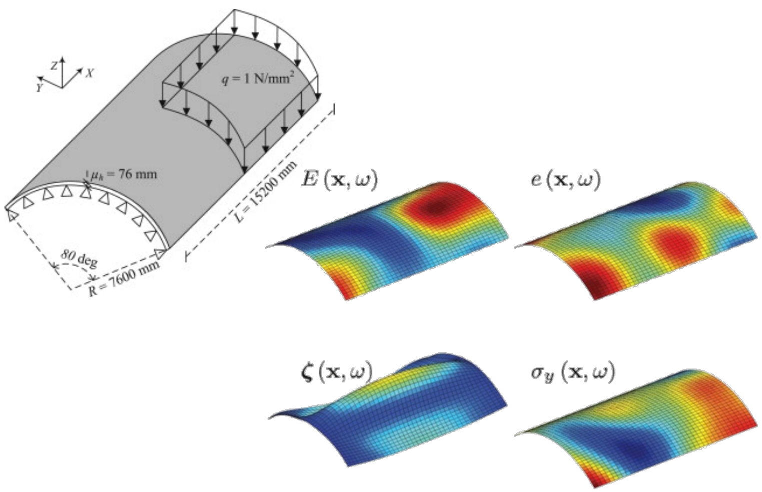


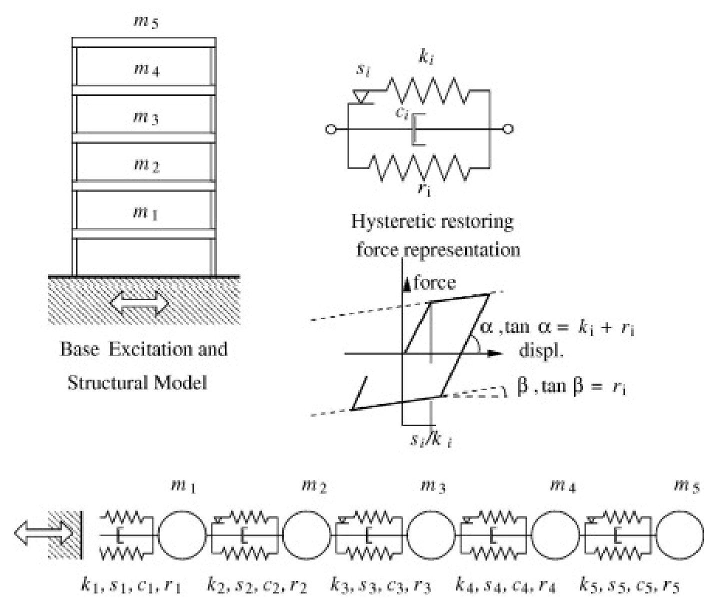


Future works in relation with the ASVR method

◢ Ongoing work

- ✓ Multiple MPFPs
- ✓ Moderately high dimensional problems (up to about 100-250 r.v.s) with random fields / random processes





◢ In the long run

- ✓ Bias correction with sampling
- ✓ Improve some computational aspects (distributed training tasks, efficiency of QP solvers)

Thank you for your attention

Further details in:

- Bourinet J.-M. Probability of rare events based on adaptive support vector machine regression. Reliability Engineering & System Safety (submitted).
- Kouassi A., Bourinet J.-M., Lalléchère S., Bonnet P., Fogli M. Reliability and sensitivity analysis of transmission lines in a probabilistic EMC context. IEEE Transactions on Electromagnetic Compatibility (submitted).

Jean-Marc Bourinet

IFMA & Institut Pascal (CNRS UMR 6602)
Clermont-Ferrand, France

Email : bourinet@ifma.fr
http://www.ifma.fr/FERUM