Maximum entropy principle for stochastic models in computational mechanics

Christian Soize

Université Paris-Est, France

christian.soize@univ-paris-est.fr

Journée Thématique "Incertitudes" du Groupement MoMaS

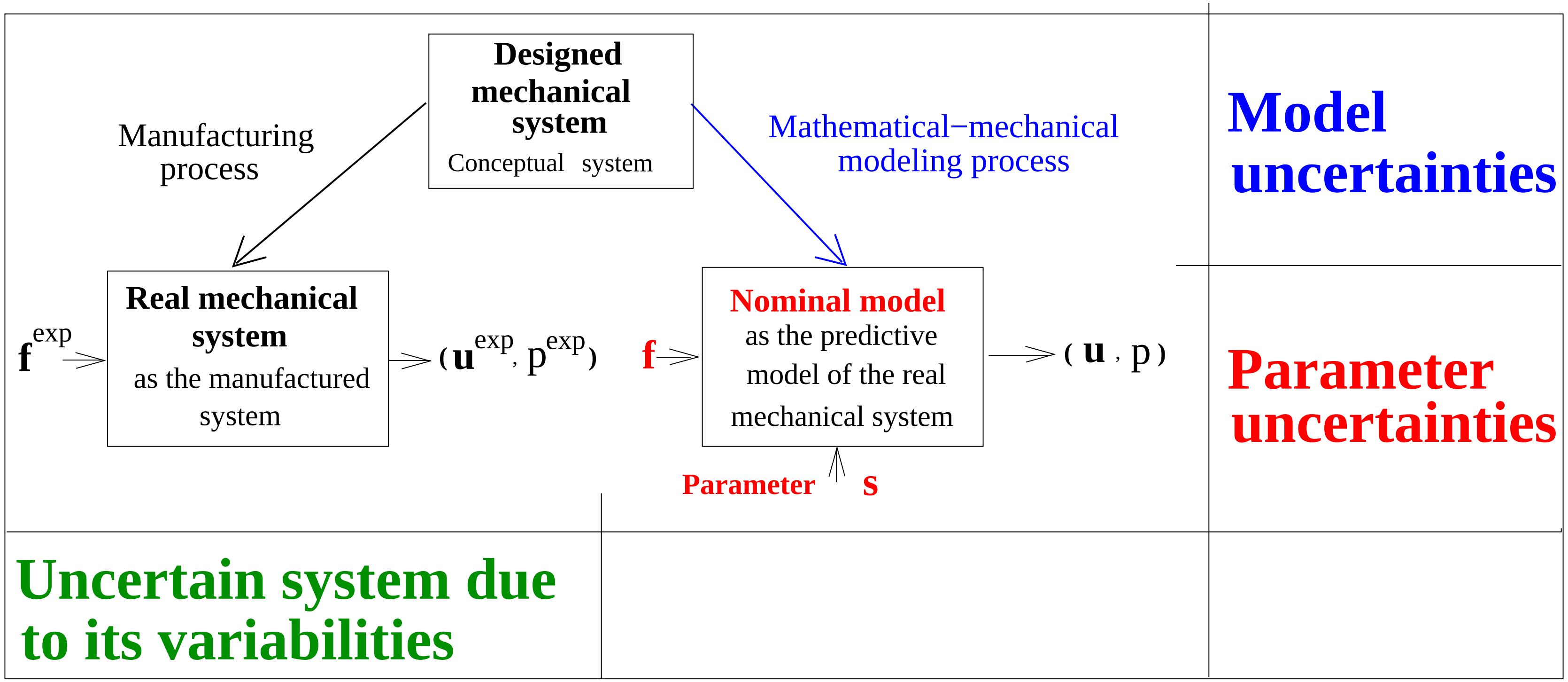
Institut Henri Poincaré

13 Novembre 2008

MOTIVATION

THE USE OF THE PROBABILITY THEORY TO MODEL UNCERTAINTIES IN COMPUTATIONAL MECHANICS

Uncertainties and variabilities in complex mechanical systems



Designed mechanical system

Conceptual system

Manufacturing process

Mathematical−mechanical modeling process

Model uncertainties

$\mathbf{f}^{\text{exp}}$ → **Real mechanical system** as the manufactured system → $(\mathbf{u}^{\text{exp}}, p^{\text{exp}})$

$\mathbf{f}$ → **Nominal model** as the predictive model of the real mechanical system → $(\mathbf{u}, p)$

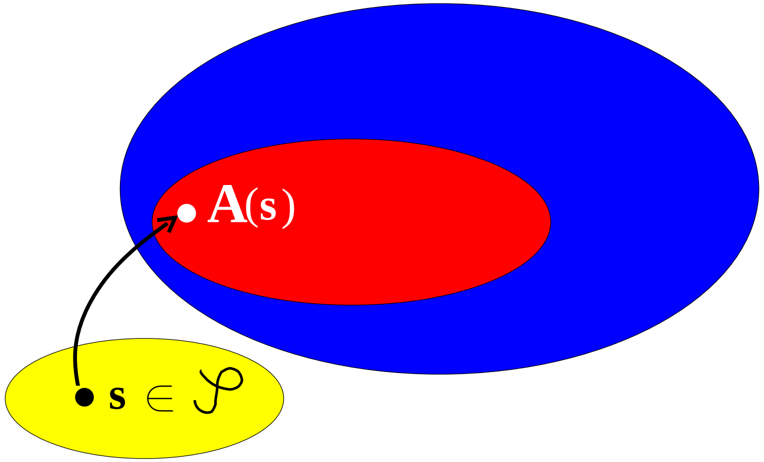
Parameter $\mathbf{s}$

Parameter uncertainties

Uncertain system due to its variabilities

- **Errors** related to the construction of an approximate solution $(\mathbf{u}^n, p^n)$ of $(\mathbf{u}, p)$ have to be reduced and controlled, and **must not be considered as uncertainties**.
- There are two types of uncertainties: **Parameter Uncertainties** and **Model Uncertainties**.
- In general there are **Variabilities** in the real system.

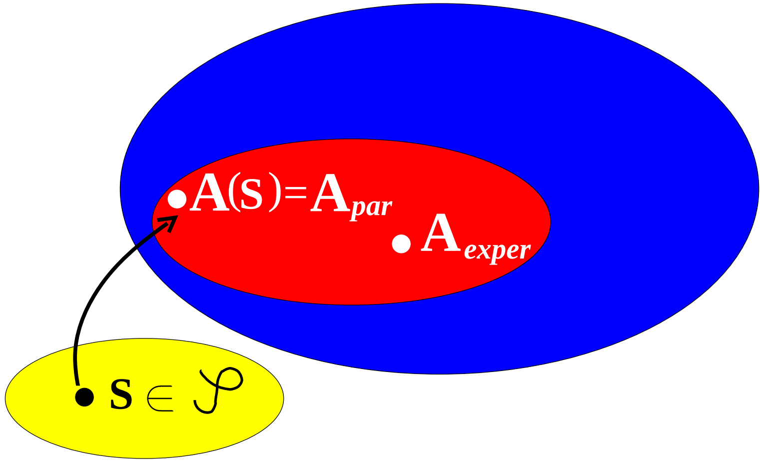
What kind of probabilistic approaches can be used to take into account "parameter uncertainties" and "model uncertainties"?



- Vector-valued parameter $\mathbf{s}$ of the computational model belongs to the admissible **yellow set**.
- $\mathbf{A}(\mathbf{s})$: linear operator in the **red set** $\subset$ **blue set**, depending on the vector-valued parameter $\mathbf{s}$.
- Example for a linear computational dynamic model:

  $\mathbf{A}(\mathbf{s})$ = stiffness matrix of the computational model,
  blue set = all the symmetric positive-definite matrices,
  yellow set = all the admissible values of the parameters (geometry, elasticity tensor,...)

A - Parametric probabilistic approach of "parameter uncertainties"

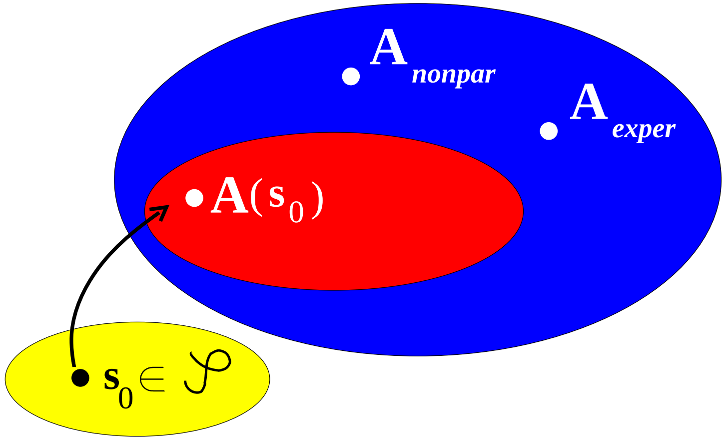


If **vector-valued parameter is uncertain**, then parameter $\mathbf{s}$ is modeled by a random variable $\mathbf{S}$.

Then $\mathbf{A}_{par} = \mathbf{A}(\mathbf{S})$ is a random matrix with values in the **red set**, which is deduced from the random variable $\mathbf{S}$ by the deterministic nonlinear mapping $\mathbf{s} \mapsto \mathbf{A}(\mathbf{s})$.

B - Nonparametric probabilistic approach of "model uncertainties"

"Model uncertainties" cannot be taken into account with the "parametric probabilistic approach" of parameter uncertainties. Another approach is necessary.



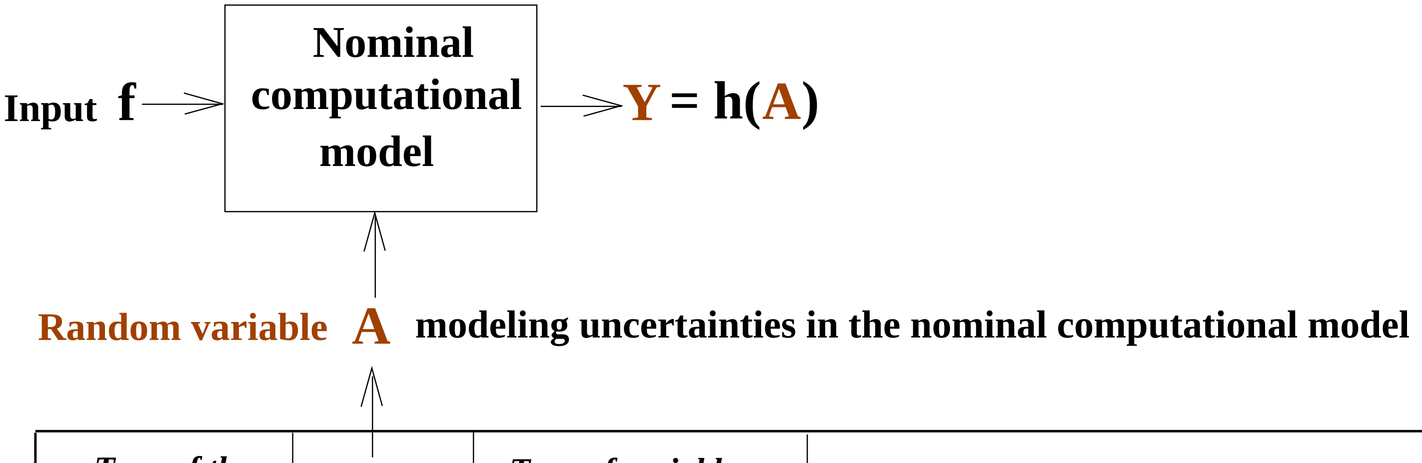
Direct construction of the probability distribution of the random matrix $\mathbf{A}_{nonpar}$ on the **blue set** using the Maximum Entropy Principle with the available information $E\{\mathbf{A}_{nonpar}\} = \mathbf{A}(\mathbf{s}_0)$ in which $\mathbf{s}_0$ is a nominal value of parameter $\mathbf{s}$.

- Fundamentals of the nonparametric approach for random uncertainties and random matrix theory: C. Soize **PEM 15(3) 277-294 (2000)**.
- Algebraic closure, convergence as dimension goes to infinity and uncertain boundary value problem: C. Soize **JASA 109(5) 1979-1996 (2001)**.
- New ensembles of random matrices for model uncertainties in coupled dynamical systems: C. Soize **CMAME 194(12-16) 1333-1366 (2005)**
- Identification from experiments and inverse problems : C. Soize *et al* **CMAME 198(1), 150-163, (2008)**.

FUNDAMENTAL PROBLEM

HOW TO TAKE INTO ACCOUNT UNCERTAINTIES
IN COMPUTATIONAL MODELS USING THE PROBABILITY THEORY?

Probabilistic modeling of uncertainties in the nominal computational model

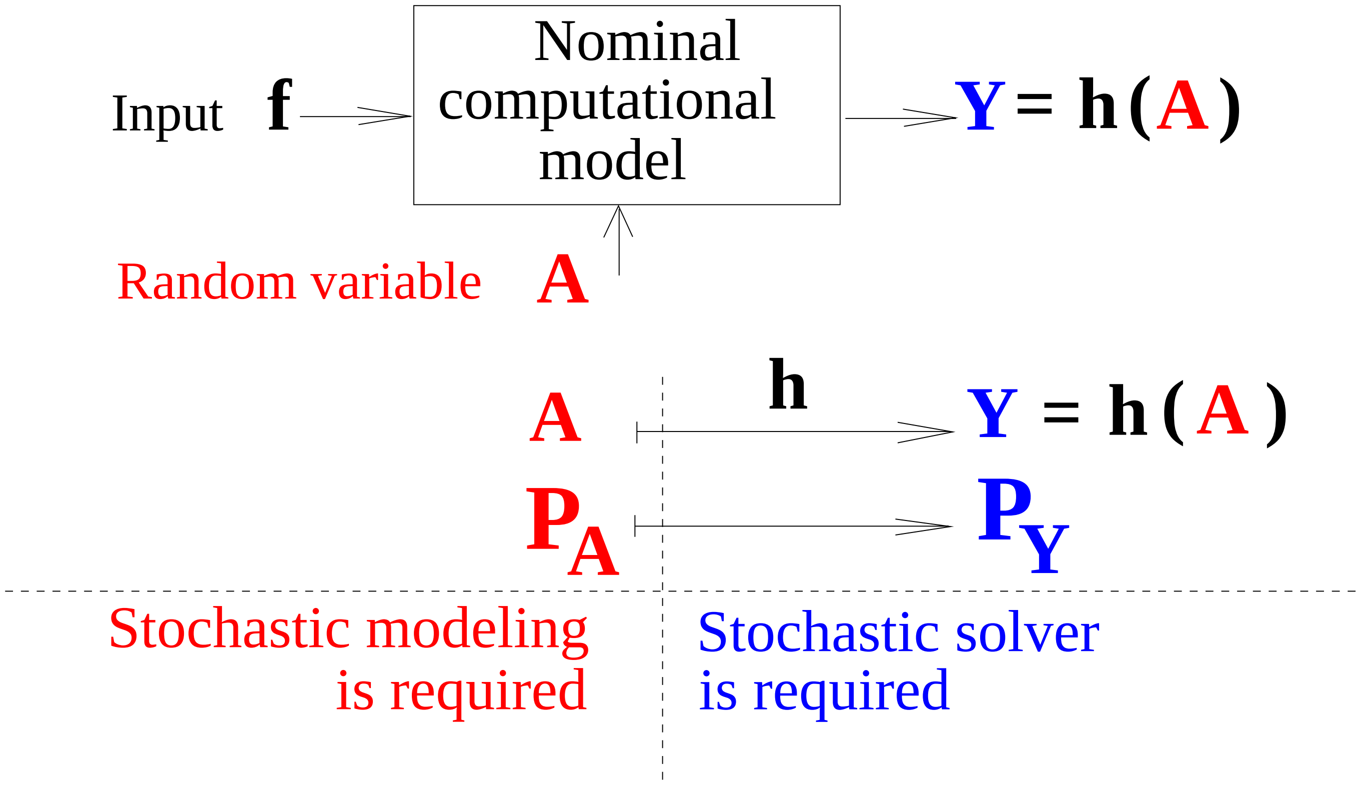


Input $\mathbf{f}$ → **Nominal computational model** → $\mathbf{Y} = \mathbf{h}(\mathbf{A})$

Random variable $\mathbf{A}$ **modeling uncertainties in the nominal computational model**

Type of the probabilistic approach	*Random variable*	*Type of variable of the computational model*	*Mathematical nature of the random variable*
Parametric	$\mathbf{S}$	**Physical parameters of the model**	**Random vector** **Random matrix (ex: (6x6) effective elasticity matrix of a heterogeneous material)** **Finite representation of a vector−valued random field** **Finite representation of a matrix−valued random field**
Nonparametric	$\mathbf{A}_{\text{nonpar}}$	**Matrices of the operators of the model**	**Random matrix in high dimension**

Stochastic computational model and propagation of uncertainties



Input $\mathbf{f}$

Nominal computational model

$\mathbf{Y} = \mathbf{h}(\mathbf{A})$

Random variable $\mathbf{A}$

$$\mathbf{A} \xmapsto{\mathbf{h}} \mathbf{Y} = \mathbf{h}(\mathbf{A})$$

$$P_{\mathbf{A}} \longmapsto P_{\mathbf{Y}}$$

Stochastic modeling is required

Stochastic solver is required

What is the fundamental problem?

- Note that **stochastic modeling** must not be confused with **stochastic solver** .

  For instance, the popular Monte Carlo numerical method :
  - is a **stochastic solver**.
  - does not correspond to a **stochastic modeling**.
  - cannot be used if the **stochastic modeling** has not been carried out.

- **Stochastic modeling** is then absolutely required and is a fundamental problem which has to be solved.

- This means that the probability measure $P_{\mathbf{A}}$ must objectively be constructed and not arbitrarily chosen.

A method useful to perform such a stochastic modeling

- The **Maximum Entropy Principle** is a method which allows the

prior probability distribution $P_{\mathbf{A}}^{prior}(d\mathbf{a};\mathbf{w})$

to be constructed using only the **available information** and which can depend on a vector-valued parameter **w**.

- Parameter **w** of the **prior probability distribution**:
  - can be estimated using mathematical statistics if measurements are available (such as the maximum likelihood method),
  - can be used as a sensitivity parameter if no measurement is available.

- If measurements are available, a

posterior probability distribution $P_{\mathbf{A}}^{post}$

can be updated from the prior probability distribution using the Bayes's method.

- Here, we are interested in the construction of the **prior probability distribution** using the **Maximum Entropy Principle**.

MAXIMUM ENTROPY PRINCIPLE

CONSTRUCTION OF THE PROBABILITY DISTRIBUTIONS OF VECTOR-VALUED RANDOM VARIABLES

Entropy as a measure of uncertainties for a vector-valued random variable

---

The measure of uncertainties of a vector-valued random variable $\mathbf{A} = (A_1, .., A_N)$ is defined by the entropy $\mathcal{E}(p_{\mathbf{A}})$ of its p.d.f $p_{\mathbf{A}}$ (Information theory, Shannon 1948):

$$\mathcal{E}(p_{\mathbf{A}}) = -E\{\log(p_{\mathbf{A}}(\mathbf{A}))\} = -\int_{\mathbb{R}^N} p_{\mathbf{A}}(\mathbf{a}) \,\log(p_{\mathbf{A}}(\mathbf{a}))\, d\mathbf{a}$$

- If the level of uncertainties decreases, then the entropy decreases.
- The maximum of uncertainty corresponds to the maximum of entropy.

Maximum entropy principle (Information theory, Shannon 48, Jaynes 56)

- Maximum entropy principle corresponds to the **maximization of uncertainties** and allows the p.d.f $p_{\mathbf{A}}$ of the vector-valued random variable **A** to be constructed using **only the available information**.

- *Let* $\mathcal{C}_{ad}$ *be the admissible set of all the p.d.f satisfying the constraints defined by the available information. Then the p.d.f* $p_{\mathbf{A}}(\mathbf{a})$ *is solution of the following optimization problem*:

$$p_{\mathbf{A}} = \arg \max_{p \in \mathcal{C}_{ad}} \mathcal{E}(p)$$

A - Definition of the available information

The available information can mathematically be written as:

Eq. (1) Supp $p_{\mathbf{A}} = \mathcal{A} \subset \mathbb{R}^N$ with $\mathcal{A}$ bounded or not bounded.

Eq. (2) $E\{\mathbf{g}(\mathbf{A})\} = \mathbf{f}$

$\mathbf{f}$ is a given vector

$\mathbf{a} \mapsto \mathbf{g}(\mathbf{a})$ is a given vector-valued function.

B - Definition of the admissible space for p.d.f. $p_{\mathbf{A}}$

$\mathcal{C}_{ad} =$ { set of all the p.d.f with support defined by **Eq. (1)** and satisfying the vector-valued constraint defined by **Eq. (2)** }

C- A formulation of the solution of the optimization problem

- Let $\boldsymbol{\lambda}$ be a vector-valued Lagrange multiplier belonging to an admissible set. **Let $\mathbf{B}_{\boldsymbol{\lambda}}$ be the $\mathbb{R}^N$-valued random variable** whose p.d.f is

$$p(\mathbf{b},\boldsymbol{\lambda}) = \mathbb{1}_{\mathcal{A}}(\mathbf{b})\, c_{\boldsymbol{\lambda}}\, \exp\{-\Phi(\mathbf{b},\boldsymbol{\lambda})\} \quad , \quad \Phi(\mathbf{b},\boldsymbol{\lambda}) = <\boldsymbol{\lambda}, \mathbf{g}(\mathbf{b})>$$

  in which $c_{\boldsymbol{\lambda}}$ is the positive constant defined by the normalization condition.

- Then the random variable $\mathbf{A}$ whose p.d.f $p_{\mathbf{A}}$ is constructed with the Maximum Entropy Principle is such that

$$p_{\mathbf{A}}(\mathbf{a}) = p(\mathbf{a},\boldsymbol{\lambda}^{sol}) \qquad \textit{i.e.} \qquad \mathbf{A} = \mathbf{B}_{\boldsymbol{\lambda}^{sol}}$$

  in which $\boldsymbol{\lambda}^{sol}$ is a solution in $\boldsymbol{\lambda}$ of the equation: $E\{\mathbf{g}(\mathbf{B}_{\boldsymbol{\lambda}})\} = \mathbf{f}$.

D- Problem to be solved and difficulties arising in high dimension

- Equation $E\{\mathbf{g}(\mathbf{B}_{\boldsymbol{\lambda}})\} = \mathbf{f}$ *i.e.* $\int_{\mathbb{R}^N} \mathbf{g}(\mathbf{b})\, p(\mathbf{b}, \boldsymbol{\lambda})\, d\mathbf{b} = \mathbf{f}$

  has to be solved in $\boldsymbol{\lambda}$ with an appropriate algorithm such that:

  - the interior-reflective Newton method
  - the Powell dogleg method
  - or the trust-region dogleg algorithm
  - etc

- Consequently, for each given $\boldsymbol{\lambda}$, we have to calculate $E\{\mathbf{g}(\mathbf{B}_{\boldsymbol{\lambda}})\}$

- When N **is large**, an **integral in high dimension** must be calculated

  - with an efficient algorithm
  - for which convergence can be controlled
  - and for which the numerical cost is low or is reasonable

Effective construction in high dimension

[**C. Soize**], Construction of probability distributions in high dimension using the maximum entropy principle. Applications to stochastic processes, random fields and random matrices,
International Journal for Numerical Methods in Engineering, **76**(10), 1583-1611 (2008).

Problem: Compute $E\{\mathbf{h}(\mathbf{B}_\lambda)\}$ for N large in which $\mathbf{b} \mapsto \mathbf{h}(\mathbf{b})$ is any vector-valued function defined on $\mathbb{R}^N$.

A- Usual methodologies

- Explicit solution (very particular case).
- Analytical calculation using integration in Gaussian spaces (not efficient)
- Approximate methods based on the use of Markov Chain Monte Carlo methods (MCMM) (efficient):
  - Metropolis-Hastings algorithm
  - Gibbs sampling

  are two efficient algorithms but can also be not efficient if there exists attraction regions which do not correspond to the invariant measure $p(\mathbf{b}, \boldsymbol{\lambda})\, d\mathbf{b}$.

B- An alternative method in the class of the MCMM

- **Case 1**: $\mathcal{A} = \mathbb{R}^N$ and $\mathbf{u} \mapsto \|\boldsymbol{\nabla}_{\mathbf{u}} \Phi(\mathbf{u}, \boldsymbol{\lambda})\|$ is a locally bounded function.

It can be proven that $\mathbf{B}_{\boldsymbol{\lambda}} = \lim_{r \to +\infty} \mathbf{U}(r)$ *(*in probability distribution *) in which* $\{(\mathbf{U}(r), \mathbf{V}(r)), r \geq 0\}$ *is the diffusion stochastic process satisfying, for all* $r > 0$*, the following Itô Stochastic Differential Equation (ISDE)*

$$
\begin{aligned}
d\mathbf{U}(r) &= \mathbf{V}(r)\, dr \\
d\mathbf{V}(r) &= -\boldsymbol{\nabla}_{\mathbf{u}} \Phi(\mathbf{U}(r), \boldsymbol{\lambda})\, dr - \tfrac{1}{2} f_0 \mathbf{V}(r)\, dr + \sqrt{f_0}\, d\mathbf{W}(r)
\end{aligned}
$$

with the initial condition $\mathbf{U}(0) = \mathbf{u}_0$ *and* $\mathbf{V}(0) = \mathbf{v}_0$*,*

$f_0 > 0$ free parameter to kill the transient part and thus to get more rapidly the stationary solution corresponding to the invariant measure $p(\mathbf{b}, \boldsymbol{\lambda})\, d\mathbf{b}$.

- **Case 2**: $\mathcal{A}$ is a bounded part of $\mathbb{R}^N$ (see the conference paper and IJNME 2008)

C- Solver of the ISDE, mathematical expectation, random generator

• Solving the Itô stochastic differential equation

- Giving a sampling $r_1, \ldots, r_M$ of r
- Giving any realization $\mathbf{W}(r_1,\theta), \ldots, \mathbf{W}(r_M,\theta)$ of the Wiener process
- Compute the realization $\mathbf{U}(r_1,\theta), \ldots, \mathbf{U}(r_M,\theta)$ solving the ISDE with an **explicit scheme** or a **semi-implicit scheme**.

• Estimation of the mathematical expectation can be obtained using

- **ergodic method**: $E\{\mathbf{h}(\mathbf{B}_\lambda)\} \simeq \frac{1}{M-M_0+1} \sum_{k=M_0}^{M} \mathbf{h}(\mathbf{U}(r_k,\theta))$

- **Monte Carlo method**: $E\{\mathbf{h}(\mathbf{B}_\lambda)\} \simeq \frac{1}{n_s} \sum_{\ell=1}^{n_s} \mathbf{h}(\mathbf{B}_\lambda(\theta_\ell))$

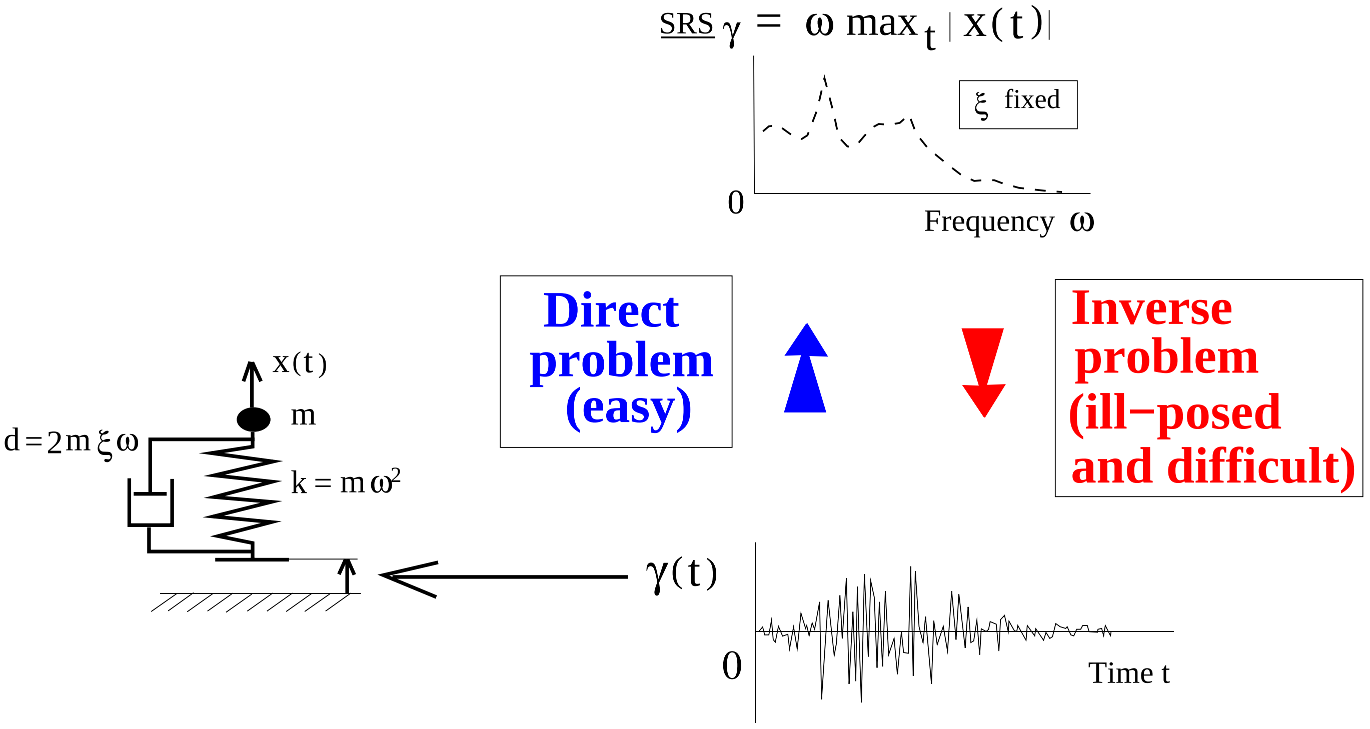
  in which $\mathbf{B}_\lambda(\theta_\ell) \simeq \mathbf{U}(r_M,\theta_\ell)$ (with r_M sufficiently large) are independent realizations of $\mathbf{B}_\lambda$ (**random generator of independent realizations of $\mathbf{B}_\lambda$**)

APPLICATION 1

INVERSE PROBLEM RELATIVE TO THE CONSTRUCTION OF A RANDOM GENERATOR OF ACCELEROGRAMS FROM A GIVEN VELOCITY RESPONSE SPECTRUM

What is a Velocity Response Spectrum (VRS) and what is the inverse problem associated with?

Direct problem: Given a deterministic transient signal $\{\gamma(t), t \in [0\,, T]\}$ (acceleration in a given direction), calculate the $\underline{\mathrm{VRS}}_\gamma$ (which consists in evaluating the maximum of the dynamical response $x(t)$ of a family of SDOF linear damped oscillators excited at their bases by γ).



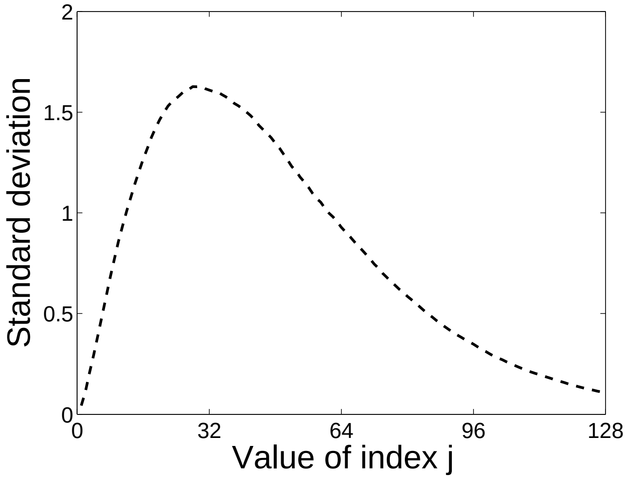
$\underline{\mathrm{SRS}}_\gamma = \omega \max_t |x(t)|$

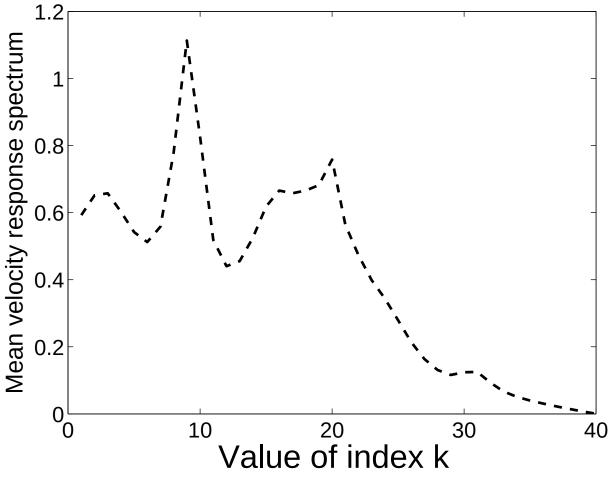
Direct problem (easy)

Inverse problem (ill−posed and difficult)

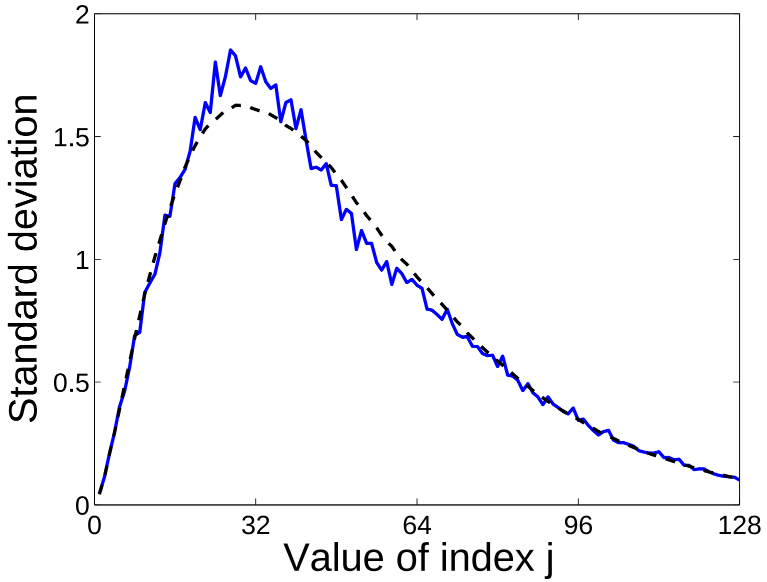
Formulation of the inverse problem using the Maximum Entropy Principle

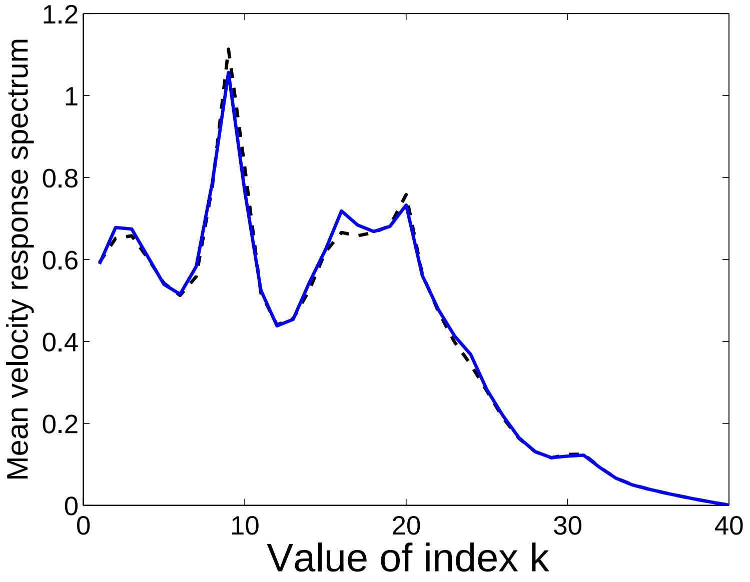
- Construction of the probability distribution of the **time series**: $\boldsymbol{\Gamma} = \{\Gamma_1, \ldots, \Gamma_{128}\}$.
- Construction of the **random** VRS: $\mathbf{VRS}_{\Gamma} = \{\mathcal{S}_1(\boldsymbol{\Gamma}), \ldots, \mathcal{S}_{40}(\boldsymbol{\Gamma})\}$ (40 frequency sampling points) where $\boldsymbol{\gamma} \mapsto \{\mathcal{S}_1(\boldsymbol{\gamma}), \ldots, \mathcal{S}_{40}(\boldsymbol{\gamma})\}$ is a **given nonlinear mapping**.
- **Available information**:
  - Centered time series
  - Given standard deviation $\boldsymbol{\sigma}_{\boldsymbol{\Gamma}} = \underline{\boldsymbol{\sigma}}_{\gamma}$
  - Given mean value $E\{\mathbf{VRS}_{\boldsymbol{\Gamma}}\} = \underline{\mathrm{VRS}}_{\gamma}$.





Estimation of the constraints (available information) using the random generator of the time series $\boldsymbol{\Gamma}$





$\boldsymbol{\lambda}^{sol}$ is computed using the interior-reflective Newton method and the ergodic method.

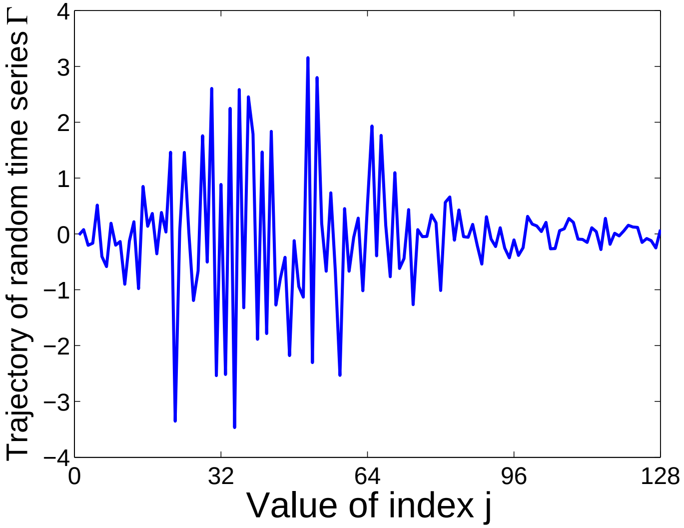
Black dashed line: Given values

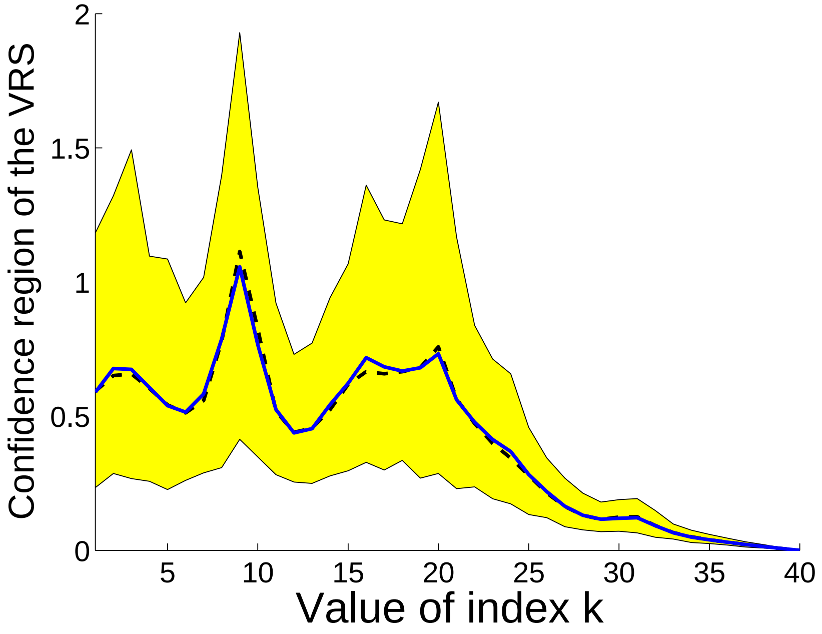
Blue solid line: Estimation with the random generator

Left figure: Predicted $\boldsymbol{\sigma}_{\boldsymbol{\Gamma}}$ compares to the given standard deviation $\underline{\boldsymbol{\sigma}}_{\gamma}$

Right figure: Predicted $E\{\mathbf{VRS}_{\boldsymbol{\Gamma}}\}$ compares to the given $\underline{\mathrm{VRS}}_{\gamma}$

Properties constructed with the random generator of the time series Γ





Left figure: Example of a realization of the **time series Γ** given by the generator

Right figure:

Black dashed line: Given $\underline{\text{VRS}}_\gamma$.
Blue solid line: $E\{\mathbf{VRS}_\Gamma\}$ estimated with the random generator.
Yellow region: Confidence region of the random $\mathbf{VRS}_\Gamma$ estimated
with the random generator for $P_c = 0.98$.

APPLICATION 2

CONSTRUCTION OF THE PROBABILITY MODEL OF THE KARHUNEN-LOEVE REPRESENTATION OF A NON-GAUSSIAN POSITIVE-VALUED RANDOM FIELD

Just giving the formulation as an example of the use of the maximum entropy principle.

Context

- Computational model including the discretization of an **elliptic operator**.
- Let $\mathbf{Z} = (Z_1, \ldots, Z_n)$ be the spatial sampling of the **positive-valued random field** defining the elliptic operator and with mean value $\underline{\mathbf{z}} = E\{\mathbf{Z}\}$.
- **Normalized** random vector $\mathbf{G} = (G_1, \ldots, G_n)$ such that $Z_j = \underline{z}_j \, G_j$.
- **Karhunen-Loeve** expansion $\mathbf{G}^N$ of the $\mathbf{G}$ yields

$$\mathbf{G}^N = \underline{\mathbf{G}} + \sum_{\alpha=1}^{N} \sqrt{v_\alpha} \, A_\alpha \, \boldsymbol{\varphi}^\alpha$$

A fundamental problem is the construction of the probability distribution of the vector-valued random variable A $= (A_1, \ldots, A_N)$ **which can be performed using the maximum entropy principle**.

Available information used to construct the probability model

- **The support** $\mathcal{A}$ of $p_{\mathbf{A}}$ has to be such that

$$\mathcal{A} = \{\mathbf{a} = (a_1, \ldots, a_N) \quad \text{such that for all } j \quad , \quad \underline{G}_j + \sum_{\alpha=1}^{N} \sqrt{v_\alpha}\, a_\alpha\, \varphi_j^\alpha > 0\}$$

- **Centered** vector-valued random variable **A**
- **Covariance matrix** of **A** is the unity matrix
- **Ellipticity condition** requires that $E\{s(\mathbf{A})\} = \kappa < +\infty$ in which $s(\mathbf{a}) = \sum_{j=1}^{n} (\underline{G}_j + \sum_{\alpha=1}^{N} \sqrt{v_\alpha}\, a_\alpha\, \varphi_j^\alpha)^{-2}$

Then the probability distribution of A and the corresponding random generator of independent realizations of A can be carried out with the method presented.

See the developments and the results in the conference paper and in the publication:

[**C. Soize**], Construction of probability distributions in high dimension using the maximum entropy principle. Applications to stochastic processes, random fields and random matrices,
International Journal for Numerical Methods in Engineering, **76**(10), 1583-1611 (2008).

APPLICATION 3

CONSTRUCTION OF THE PROBABILITY MODEL OF RANDOM MATRICES

A FIRST EXAMPLE to demonstrate the capability of the method proposed: Probability model for positive-definite band random matrices

---

- Let $[\,\mathbf{G}\,]$ be the band random matrix with values in $\mathbb{M}_n^+(\mathbb{R})$ with $n = 4$, for which the band structure is such that

$$[\,\mathbf{G}\,] = \begin{bmatrix} G_{11} & G_{12} & 0 & 0 \\ G_{12} & G_{22} & G_{23} & 0 \\ 0 & G_{23} & G_{33} & G_{34} \\ 0 & 0 & G_{34} & G_{44} \end{bmatrix}$$

- **The problem is the construction of the probability distribution on $\mathbb{M}_n^+(\mathbb{R})$ of random matrix $[\,\mathrm{G}\,]$ using the maximum entropy principle under the constraints defined by the available information**.

• Choosing an algebraic representation of the random matrix

Since $[\,\mathbf{G}\,]$ is positive definite a.s, random matrix $[\,\mathbf{G}\,]$ can be written as

$$[\,\mathbf{G}\,] = [\,\mathcal{G}(\mathbf{A})] = [\,\mathcal{L}(\mathbf{A})]^T\,[\,\mathcal{L}(\mathbf{A})] \quad , \quad [\,\mathcal{L}(\mathbf{A})] = \begin{bmatrix} A_1^2 & A_2 & 0 & 0 \\ 0 & A_3^2 & A_4 & 0 \\ 0 & 0 & A_5^2 & A_6 \\ 0 & 0 & 0 & A_7^2 \end{bmatrix}$$

which **defines the nonlinear deterministic mapping** $\mathbf{A} \mapsto [\,\mathcal{G}(\mathbf{A})\,]$.

• Available information for the vector-valued random variable $\mathbf{A} = (A_1, .., \; A_N)$

- Mean value: $E\{[\,\mathcal{G}(\mathbf{A})\,]\} = [\,I_n]$

- Second-order random matrix: $E\{\|[\,\mathcal{G}(\mathbf{A})\,]\|_F^2\} = n(\delta^2 + 1)$

- Invertibility condition: $E\{\|[\,\mathcal{G}(\mathbf{A})\,]^{-1}\|_F^2\} = \alpha < +\infty$

• Results

- The probability distribution of the vector-valued random variable **A**
- and the corresponding random generator of independent realizations of **A**

can be carried out with the method presented.

See the developments and the results in the conference paper and in the publication:

[**C. Soize**], Construction of probability distributions in high dimension using the maximum entropy principle. Applications to stochastic processes, random fields and random matrices,
International Journal for Numerical Methods in Engineering, **76**(10), 1583-1611 (2008)**.**

CONCLUSIONS

- The probability theory is a powerful mathematical theory to model uncertainties in computational mechanics and more generally in computational sciences.

- The maximum entropy principle is a very effective method to construct the prior probability distribution of random quantities modeling uncertainties in a computational model.

- This principle can be used for complicate mathematical objects such as vector-valued random variables or random matrices in high dimension, vector-valued stochastic processes and tensor-valued random fields.

- For high dimension problems, the Markov Chain Monte Carlo methods such as Metropolis-Hastings algorithm or Gibbs sampling can be used and we have proposed an alternative method based on stochastic analysis.

- Experimental validations have been presented for complex uncertain dynamical systems.