Quantification of Uncertainty from high-dimensional experimental data

Lionel Mathelin$^{1,2}$

$^1$LIMSI-CNRS, France
$^2$currently at AeroAstro Dpt., MIT
Let $\xi \in \Xi \subseteq \mathbb{R}^d$ belonging to a probability space $(\Xi, B, \mu_\xi)$.

\[
y(\xi) \in \mathbb{R} \quad \rightarrow \quad \mathcal{M}(y(\xi); \xi) = 0, \quad \mu_\xi \text{ a.e.} \quad \rightarrow \quad \text{QoI}(\xi)
\]
A growing issue in numerical simulations

Propagating uncertainty in a model and precisely assessing its output requires accurate description of input uncertainty $y(x, \xi)$.
Propagating uncertainty in a model and precisely assessing its output requires accurate description of input uncertainty $y(x, \xi)$,

But...

- tendency to account for more and more phenomena (multi-physics, etc.) $\rightarrow$ higher number of sources of uncertainty,
- more and more sophisticated models (high fidelity) $\rightarrow$ input description has become the bottleneck of the simulation chain accuracy,
- input data are difficult and/or expensive to acquire (e.g., in situ measurements),
- usually impossible to set-up an experimental design: samples are random and do not obey a sampling strategy (say, like quadrature).
A growing issue in numerical simulations

Propagating uncertainty in a model and precisely assessing its output requires accurate description of input uncertainty $y(x, \xi)$,

But...

- tendency to account for more and more phenomena (multi-physics, etc.) $\rightarrow$ higher number of sources of uncertainty,
- more and more sophisticated models (high fidelity) $\rightarrow$ input description has become the bottleneck of the simulation chain accuracy,
- input data are difficult and/or expensive to acquire (e.g., in situ measurements),
- usually impossible to set-up an experimental design: samples are random and do not obey a sampling strategy (say, like quadrature).

$\Rightarrow$ it is critical to infer the most out of the scarce available data.

As an example, what can reasonably be inferred from a mere 1000 samples of a 100-dimensional vector-valued random variable?
Two reasons for hope:

- “blessing of dimensionality”. Approximation in many bases is **sparse**, very likely that the underlying quantity is **anisotropic**.
Two reasons for hope:

- “blessing of dimensionality”. Approximation in many bases is sparse,
- very likely that the underlying quantity is anisotropic.

⇒ A route to tractability:

- exploit the intrinsic difference between, say, physical and stochastic dimensions
  → separated representation whenever possible,
- take advantage of low correlation orders between dimensions in most physical phenomena,
- efficient evaluation of the basis terms,
- subset selection technique to further reduce the cardinality.
Let’s try anyway...

Two reasons for hope:
- “blessing of dimensionality”. Approximation in many bases is sparse,
- very likely that the underlying quantity is anisotropic.

⇒ A route to tractability:
- exploit the intrinsic difference between, say, physical and stochastic dimensions
  → separated representation whenever possible,
- take advantage of low correlation orders between dimensions in most physical phenomena,
- efficient evaluation of the basis terms,
- subset selection technique to further reduce the cardinality.

\[
\{y^{(1)}, y^{(2)}, \ldots y^{(M)}\} \rightarrow y(x, t, \xi, \ldots)
\]

Lionel Mathelin
UQ from high-dimensional experimental data
Format of the approximation

Information on the QoI: \( \left\{ x^{(m)}, \xi^{(m)}, y^{(m)} \right\}_{m=1}^{M}, x^{(m)} \in \mathbb{R}^{1,2,3,...}, \xi^{(m)} \in \mathbb{R}^{d}, y^{(m)} \in \mathbb{R}. \)
Information on the QoI: \( \{ x^{(m)}, \xi^{(m)}, y^{(m)} \}_{m=1}^{M}, x^{(m)} \in \mathbb{R}^{1,2,3,...}, \xi^{(m)} \in \mathbb{R}^d, y^{(m)} \in \mathbb{R} \).

\( \rightarrow \) crucial to use a minimal cardinality basis for a given approximation accuracy

\[ y(x, \xi) \approx \sum_{r} w_r(x) \lambda_r(\xi) \quad \text{low-rank approximation}. \]

solved by Galerkin projection:

\[
\left\langle y(x, \xi) - \sum_{r} w_r(x) \lambda_r(\xi), w_R(x) \lambda_R(\xi) \right\rangle = 0, \quad \forall w_R(x) \lambda_R(\xi) \in \mathcal{V} \otimes \mathcal{S}.
\]
Information on the QoI: \( \{ x^{(m)}, \xi^{(m)}, y^{(m)} \}_{m=1}^M, \ x^{(m)} \in \mathbb{R}^{1,2,3,\ldots}, \ \xi^{(m)} \in \mathbb{R}^d, \ y^{(m)} \in \mathbb{R}. \)

\( \rightarrow \) crucial to use a minimal cardinality basis for a given approximation accuracy

\[ y(x, \xi) \approx \sum_r w_r(x) \lambda_r(\xi) \quad \text{low-rank approximation.} \]

solved by Galerkin projection:

\[ \left\langle y(x, \xi) - \sum_{r=1}^R w_r(x) \lambda_r(\xi), w_R(x) \lambda_R(\xi) \right\rangle = 0, \quad \forall w_R(x) \lambda_R(\xi) \in V \otimes S. \]

Alternate projection: Letting \( z(x, \xi) := y(x, \xi) - \sum_{r=1}^{R-1} w_r(x) \lambda_r(\xi), \) a pair \((w_R(x), \lambda_R(\xi))\) is chosen to satisfy

\[ \begin{align*}
\left\langle \lambda_R \Phi c_{w,R}, \lambda_R \phi_I \right\rangle_{M_{\text{coef}}} &= \langle z, \lambda_R \phi_I \rangle_{M_{\text{coef}}}, \quad \forall \phi_I \in \mathcal{V}, \\
\left\langle w_R \Psi c_{\gamma,R}, w_R \psi_k \right\rangle_{M_{\text{coef}}} &= \langle z, w_R \psi_k \rangle_{M_{\text{coef}}}, \quad \forall \psi_k \in \mathcal{S},
\end{align*} \]

with \( < u(x, \xi), v(x, \xi) >_{M_{\text{coef}}} := \sum_{m=1}^{M_{\text{coef}}} u(x^{(m)}, \xi^{(m)}) v(x^{(m)}, \xi^{(m)}) \) the “experimental” inner product in the approximation space.
The vast majority of physics-based random signals exhibit a comparatively low interaction order between input variables so that:

\[ \| P_\gamma \lambda \| \gg \| P_{\gamma',|\gamma'|} \lambda - P_\gamma \lambda \|, \quad P_\gamma \text{ projects on a } |\gamma|-D \text{ canonical hyperplane}, \]

\[ \lambda(\xi) = f_\emptyset + d \sum_{i=1} f_i(\xi_i) + d \sum_{i<j} f_{ij}(\xi_i,\xi_j) + \ldots + f_{12 \ldots d}(\xi_1,\xi_2,\ldots,\xi_d), \approx |\gamma| = N \sum_{\gamma \subseteq \{1,\ldots,d\}} f_\gamma(\xi_\gamma) = f(\xi). \]
The vast majority of physics-based random signals exhibit a comparatively low interaction order between input variables so that:

$$\| P_\gamma \lambda \| \gg \| P_{\gamma', |\gamma'| > |\gamma|} \lambda - P_\gamma \lambda \|,$$

which projects on a $|\gamma|$-D canonical hyperplane, best suits the *High-Dimensional Model Representation* (HDMR), RABITZ & ALIȘ (1999).

$$\lambda(\xi) = f_0 + \sum_{i=1}^{d} f_i(\xi_i) + \sum_{i<j=1}^{d} f_{ij}(\xi_i, \xi_j) + \ldots + f_{12\ldots d}(\xi_1, \xi_2, \ldots, \xi_d),$$

$$\approx \sum_{|\gamma| = N_l < d} f_\gamma(\xi_\gamma) = f(\xi).$$
Functional basis for \( \{ f_\gamma \} \)

\[
\lambda (\xi) \approx \sum_{\gamma \subseteq \{1, \ldots, d\}, |\gamma|=N_l} f_\gamma (\xi_\gamma)
\]

**Approximation of the modes**

Each HDMR mode is somehow naturally approximated within a \( p \)-th total order polynomial expansion format (PC):

\[
f_\gamma \approx \hat{f}_\gamma \equiv \sum_{\alpha, |\alpha|=|\gamma|} c_{\alpha, \gamma} \psi_\alpha (\xi_\gamma).
\]

But for high PC order \( p \) and/or high HDMR order \( N_l \), the PC format of the modes \( \{ f_\gamma \} \) requires too many DOFs. → substitute a low-rank approximation:

\[
f_\gamma \approx \tilde{f}_\gamma \equiv \sum_{r, |\gamma|} \prod_{l=1}^{p} \sum_{\alpha, |\alpha|=|\gamma|} c_{\alpha, \gamma} \psi_\alpha (\xi_\gamma).\]
Functional basis for \( \{ f_\gamma \} \)

\[
\lambda (\xi) \approx \sum_{\gamma \subseteq \{1, \ldots, d\}} f_\gamma (\xi_\gamma)
\]

**Approximation of the modes**

Each HDMR mode is somehow naturally approximated within a \( p \)-th total order polynomial expansion format (PC):

\[
f_\gamma \approx \hat{f}_\gamma \equiv \sum_{\alpha, |\alpha| = |\gamma|} c_{\alpha, \gamma} \psi_\alpha (\xi_\gamma).
\]

**But** for high PC order \( p \) and/or high HDMR order \( N_l \), the PC format of the modes \( \{ f_\gamma \} \) requires too many DOFs.

\[\rightarrow\] substitute a low-rank approximation:

\[
f_\gamma \approx \tilde{f}_\gamma \equiv \sum_r \prod_{l=1}^{p} \sum_{\alpha=1}^{L_l} c_{\alpha, \gamma, l, r} \psi_\alpha (\xi_{\gamma(l)}).
\]
How to exploit \textit{a priori} the likely sparsity of the approximation in $\{f_{\gamma}\}$?

NP-hard problem.
How to exploit *a priori* the likely sparsity of the approximation in \( \{ f_\gamma \} \)?

NP-hard problem.

\[
\mathbf{c} = \arg\min_{\tilde{\mathbf{c}} \in \mathbb{R}^{|\mathcal{J}|}} \| \mathbf{\lambda} - \Psi \tilde{\mathbf{c}} \|^2_2 + \tau \sum_{\gamma \in \{1, \ldots, d\}} \| \tilde{\mathbf{c}} \|^2_{K_{\gamma}},
\]

\( \tau > 0 \) and \( K_{\gamma} \) a positive definite matrix. The *whole* set of predictors associated with a mode \( f_\gamma \) is treated together for subset selection:

- speed-up subset selection step,
- makes the subset selection more robust w.r.t. measurement noise.
How to exploit *a priori* the likely sparsity of the approximation in \( \{ f_\gamma \} \)?

NP-hard problem.

\[
\mathbf{c} = \arg \min_{\tilde{\mathbf{c}} \in \mathbb{R}^{|\mathcal{J}|}} \| \lambda - \Psi \tilde{\mathbf{c}} \|_2^2 + \tau \sum_{\gamma \in \{1, \ldots, d\}} \| \tilde{\mathbf{c}} \|_{K_\gamma},
\]

\( \tau > 0 \) and \( K_\gamma \) a positive definite matrix. The *whole* set of predictors associated with a mode \( f_\gamma \) is treated together for subset selection:

- speed-up subset selection step,
- makes the subset selection more robust w.r.t. measurement noise.

*Compressed Sensing* is suitable but intractable in high-dim framework (top-to-bottom approach).

\[\rightarrow\] modified Least Angle Regression Selection (LARS).

LARS determines a sequence of approximation bases of growing cardinality (explores the Pareto front).

**Issue**: closed-form solution for approximations *linear* in the coefficients while \( \{ \tilde{f}_\gamma \} \) are *nonlinear*.

\[\rightarrow\] Use the PCE-HDMR format for the subset selection *only* and *rk1-HDMR* for approximating \( \lambda (\xi) \).
Solution process for $y(x, \xi) \approx \sum_r w_r(x) \lambda_r(\xi)$

1. Choose $p, N_l, r_{max}, N_x$. Initialize $\{z_m = y_m\}_{m=1}^M$ and set $r \leftarrow 0$,
2. solve a deterministic problem for $\{c_{w,r}\}$: $\langle \lambda_r \Phi c_{w,r}, \lambda_r \phi_l \rangle_{M_{coef}} = \langle z, \lambda_r \phi_l \rangle_{M_{coef}}$
   and normalize $w(x)$,
3. solve a stochastic problem for $\{c_{\gamma,r}\}$: $\langle w_r \Psi c_{\gamma,r}, w_r \psi_k \rangle_{M_{coef}} = \langle z, w_r \psi_k \rangle_{M_{coef}}$
   1. Initialize the stochastic approximation basis $\{f_{\gamma}\} = \emptyset$, $\Gamma = \emptyset$,
   2. solve the (g)LASSO optimization problem for $\lambda_r$ with the (g)LARS algorithm $\rightarrow$ sequence of approximation bases with ordered indices $\{\gamma(n)\}$,
   3. Set $n \leftarrow 0$. Solve the approximation problem: for the next index $\gamma^{(n+1)}$ in the sequence, activate the mode $f_{\gamma^{(n+1)}}$,
      $\Gamma^{(n+1)} = \{\Gamma(n), \gamma^{(n+1)}\}$,
   4. solve for the approximation coefficients $\{c_{\gamma,r}\}$ by (nested) Alternate Least-Squares over the predictors,
      \[
      \begin{cases}
        c_{\gamma,r} = \arg\min_{\tilde{c}_{\gamma} \in \mathbb{R}^{|\gamma|}} \|z - w_r \Psi \{\Gamma^{(n+1)}\}_{\backslash \gamma} c_{\{\Gamma^{(n+1)}\}_{\backslash \gamma}} - w_r \Psi \gamma \tilde{c}_{\gamma}\|_{M_{coef}}^2, & \forall \gamma \in \Gamma^{(n+1)}
      \end{cases}
      \]
   6. estimate the relative approximation error $\epsilon$ by cross-validation. If $\epsilon$ decreases, $n \leftarrow n + 1$ and go back to step 4. Otherwise, exit inner loop.
4. If $\|\lambda_r\|_{M_{coef}}$ converges, set $z \leftarrow z - w_r \lambda_r$, and $r \leftarrow r + 1$. Iterate in step 2.
If one is given $M$ samples...

- Stochastic convection-diffusion problem

Error estimation when the amount of information varies. $d = 8$.

Impact of the dimensionality $d$ onto the recovery performance.

- The anisotropy is successfully exploited.
- If $d = 100 \rightarrow \text{card}\{f_\gamma\} = 3 \times 10^6$ (PC: $2 \times 10^{11}$).
Subset selection. Example for $d = 40$

$$\lambda(\xi) \approx f_0 + \sum_{i=1}^{d} f_i(\xi_i) + \sum_{i<j=1}^{d} f_{ij}(\xi_i, \xi_j) + \sum_{i<j<k=1}^{d} f_{ijk}(\xi_i, \xi_j, \xi_k) + \ldots$$

Connectivity of the 2nd order

Connectivity of the 3rd order

→ essentially activates only a few dimensions.
Connectivity of the 2nd order

Lionel Mathelin

UQ from high-dimensional experimental data
Complexity for approximating $\lambda(\xi)$

$$\mathcal{J}_{\text{subsel}} \sim O\left(\frac{M^4}{p^2}\right) + O\left(M^2 \sum_{l=1}^{N_i} \frac{d!}{(d-l)!l!}\right) \approx O\left(M^2 d^{N_i} p N_i^{-1}\right),$$

(repeated LS pb)

search for most correlated inactive mode

$$\mathcal{J}_{\text{coefs}} \approx O\left(M^2 d^{N_i} p N_i^{-1}\right).$$

Scaling with # samples $M$ ($N_i = 3$).

Scaling with the dimension $d$. UQ from high-dimensional experimental data
Influence of the identification uncertainty and measurement noise – $d = 40$

→ Weighted Total Least Square formulation
First two spatial modes. $d = 6$ (7-dimensional problem), # unknowns: 70,304.

$M = 3,300 \rightarrow$ about 3.2 samples per dimension.

$\rightarrow$ they compare rather well with exact separated solution modes (from Karhunen-Loève).
First spatial mode. $d = 8$ (10-dimensional problem), # unknowns: 781,000+.

$M = 19,300 \rightarrow$ about 2.7 samples per dimension.
A solution technique to get an accurate representation of input variables from experimental data to feed numerical models: $y(x, \xi) \approx \sum_r w_r(x) \lambda_r(\xi)$,

few data available $\rightarrow$ key of success is a well suited functional representation,

efficient subset selection technique to derive a stochastic basis even for high-dimensional problems,

approximation somehow robust w.r.t. noise (Weighted Total Least Squares) – still in progress.
Concluding remarks

- A solution technique to get an accurate representation of input variables from experimental data to feed numerical models: \( y(x, \xi) \approx \sum_r w_r(x) \lambda_r(\xi) \),
- few data available \( \rightarrow \) key of success is a well suited functional representation,
- efficient subset selection technique to derive a stochastic basis even for high-dimensional problems,
- approximation somehow robust w.r.t. noise (Weighted Total Least Squares) – still in progress.

Down the road . . .

- Application to realistic problems (oil spill in Gulf of Mexico),
- theoretical analysis supporting the choice of the representation format for \( \lambda(\xi) \):
  CANDECOMP-like? HDMR? Tensor Trains? . . . ?
  \( \rightarrow \) format selection issue.
Concluding remarks

A solution technique to get an accurate representation of input variables from experimental data to feed numerical models: $y(x, \xi) \approx \sum_r w_r(x) \lambda_r(\xi)$,

few data available $\rightarrow$ key of success is a well suited functional representation,

efficient subset selection technique to derive a stochastic basis even for high-dimensional problems,

approximation somehow robust w.r.t. noise (Weighted Total Least Squares) – still in progress.

Down the road...

Application to realistic problems (oil spill in Gulf of Mexico),

theoretical analysis supporting the choice of the representation format for $\lambda(\xi)$: CANDECOMP-like? HDMR? Tensor Trains? ...?
$\rightarrow$ format selection issue.

Accuracy of input is obviously critical for accuracy of output of numerical models...

... and can often be achieved even from scarce experimental datasets
Stochastic diffusion equation on $\Omega \times \Xi$, $\Omega = [x_-, x_+]$ with deterministic Dirichlet boundary conditions:

$$\nabla (\nu(x, \xi) \nabla y(x, \xi)) = F(x, \xi),$$
$$y(x_-, \xi) = y_-, \quad y(x_+, \xi) = y_+.$$

$F$ and $\nu$ defined by

$$\nu(x, \xi') = \nu_0(x) + \sum_k \sqrt{\sigma_{\nu,k}} \omega_{\nu,k}(x) \xi'_k,$$
$$F(x, \xi'') = F_0(x) + \sum_k \sqrt{\sigma_{F,k}} \omega_{F,k}(x) \xi''_k,$$

with $\nu_0 = 1$ and $F_0 = -1$. The spatial modes $\omega_{\nu,k}(x)$ and $\omega_{F,k}(x)$, and their associated amplitude $\sqrt{\sigma_{\nu,k}}$ and $\sqrt{\sigma_{F,k}}$, are the first dominant eigenfunctions of eigenproblems associated with Gaussian correlation kernels:

$$K_\nu(x, x') = \sigma^2_\nu e^{-\frac{(x-x')^2}{L^2_{c,\nu}}}, \quad K_F(x, x') = \sigma^2_F e^{-\frac{(x-x')^2}{L^2_{c,F}}},$$

with $\sigma_\nu = \sqrt{0.5}$, $\sigma_F = \sqrt{0.5}$, $L_{c,\nu} = \sqrt{0.2}$, $L_{F,\nu} = \sqrt{0.2}$.
Low rank or HDMR?

Weak solution of the stochastic diffusion problem. Approximation of $y(x^*, \xi) \equiv \lambda(\xi)$. Higher order HDMR modes activated upon sensitivity Sobol estimates at previous order.

→ HDMR is a sound choice, even with this naive basis adaption scheme.