Quantification of Uncertainty from high-dimensional experimental data

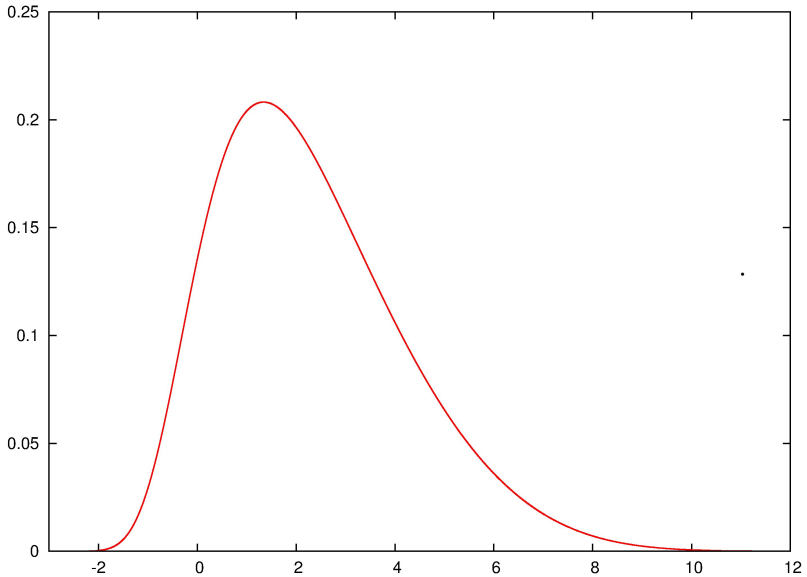
Lionel Mathelin[1,2]

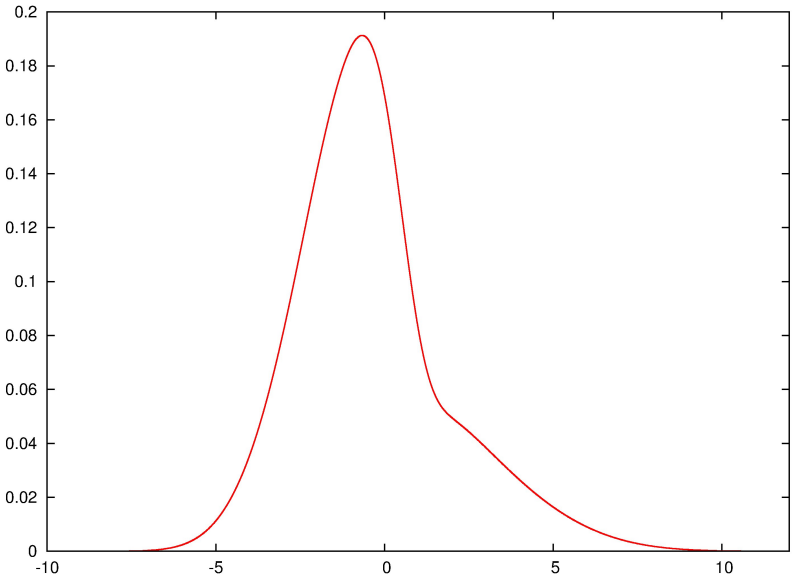
[1]LIMSI-CNRS, France
[2]currently at AeroAstro Dpt., MIT

Propagating parametric uncertainty in a nutshell

Let $\boldsymbol{\xi} \in \Xi \subseteq \mathbb{R}^d$ belonging to a probability space $\left(\Xi, \mathcal{B}, \mu_{\boldsymbol{\xi}}\right)$.

$$y\left(\boldsymbol{\xi}\right) \in \mathbb{R} \quad \longrightarrow \quad \mathcal{M}\left(y(\boldsymbol{\xi}); \boldsymbol{\xi}\right) = 0, \quad \mu_{\boldsymbol{\xi}} - \text{a.e.} \quad \longrightarrow \quad \text{QoI}\left(\boldsymbol{\xi}\right)$$





A growing issue in numerical simulations

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

A growing issue in numerical simulations

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

But...

- tendency to account for more and more phenomena (multi-physics, etc.) $\longrightarrow$ higher number of sources of uncertainty,
- more and more sophisticated models (high fidelity) $\longrightarrow$ 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 $y(\boldsymbol{x}, \boldsymbol{\xi})$,

But...

- tendency to account for more and more phenomena (multi-physics, etc.) $\longrightarrow$ higher number of sources of uncertainty,
- more and more sophisticated models (high fidelity) $\longrightarrow$ 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).

$\Longrightarrow$ 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?

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.

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.

$\Longrightarrow$ A route to tractability:

- exploit the intrinsic difference between, say, physical and stochastic dimensions $\longrightarrow$ 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.

$\Longrightarrow$ A route to tractability:

- exploit the intrinsic difference between, say, physical and stochastic dimensions $\longrightarrow$ 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.

$$\left\{y^{(1)}, y^{(2)}, \ldots y^{(M)}\right\} \longrightarrow y\left(\boldsymbol{x}, t, \boldsymbol{\xi}, \ldots\right)$$

Format of the approximation

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

Format of the approximation

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

$\Longrightarrow$ crucial to use a minimal cardinality basis for a given approximation accuracy

$$\longrightarrow y\left(\boldsymbol{x}, \boldsymbol{\xi}\right) \approx \sum_r w_r\left(\boldsymbol{x}\right) \lambda_r\left(\boldsymbol{\xi}\right) \qquad \textbf{low-rank approximation}.$$

solved by Galerkin projection:

$$\left\langle y\left(\boldsymbol{x}, \boldsymbol{\xi}\right) - \sum_{r}^{R} w_r\left(\boldsymbol{x}\right) \lambda_r\left(\boldsymbol{\xi}\right), w_R\left(\boldsymbol{x}\right) \lambda_R\left(\boldsymbol{\xi}\right) \right\rangle = 0, \qquad \forall\, w_R\left(\boldsymbol{x}\right) \lambda_R\left(\boldsymbol{\xi}\right) \in \mathcal{V} \otimes \mathcal{S}.$$

Format of the approximation

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

$\Longrightarrow$ crucial to use a minimal cardinality basis for a given approximation accuracy

$$\longrightarrow y\left(\boldsymbol{x}, \boldsymbol{\xi}\right) \approx \sum_r w_r\left(\boldsymbol{x}\right) \lambda_r\left(\boldsymbol{\xi}\right) \qquad \textbf{low-rank approximation}.$$

solved by Galerkin projection:

$$\left\langle y\left(\boldsymbol{x}, \boldsymbol{\xi}\right) - \sum_{r}^{R} w_r\left(\boldsymbol{x}\right) \lambda_r\left(\boldsymbol{\xi}\right), w_R\left(\boldsymbol{x}\right) \lambda_R\left(\boldsymbol{\xi}\right) \right\rangle = 0, \qquad \forall\, w_R\left(\boldsymbol{x}\right) \lambda_R\left(\boldsymbol{\xi}\right) \in \mathcal{V} \otimes \mathcal{S}.$$

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

$$\begin{cases} \left\langle \lambda_R\, \Phi\, c_{w,R}, \lambda_R\, \phi_l \right\rangle_{M_{\text{coef}}} &= \left\langle z, \lambda_R\, \phi_l \right\rangle_{M_{\text{coef}}}, & \forall\, \phi_l \in \mathcal{V}, \\ \left\langle w_R\, \Psi\, c_{\gamma,R}, w_R\, \psi_k \right\rangle_{M_{\text{coef}}} &= \left\langle z, w_R\, \psi_k \right\rangle_{M_{\text{coef}}}, & \forall\, \psi_k \in \mathcal{S}, \end{cases}$$

with $< u\left(\boldsymbol{x}, \boldsymbol{\xi}\right), v\left(\boldsymbol{x}, \boldsymbol{\xi}\right) >_{M_{\text{coef}}} := \sum_{m=1}^{M_{\text{coef}}} u\left(\boldsymbol{x}^{(m)}, \boldsymbol{\xi}^{(m)}\right) v\left(\boldsymbol{x}^{(m)}, \boldsymbol{\xi}^{(m)}\right)$ the "experimental" inner product in the approximation space.

Representing $\lambda(\xi)$

The vast majority of physics-based random signals exhibit a comparatively low interaction order between input variables so that:

$$\|\mathcal{P}_{\boldsymbol{\gamma}}\,\lambda\| \gg \left\|\mathcal{P}_{\boldsymbol{\gamma}',|\boldsymbol{\gamma}'|>|\boldsymbol{\gamma}|}\,\lambda - \mathcal{P}_{\boldsymbol{\gamma}}\,\lambda\right\|, \quad \mathcal{P}_{\boldsymbol{\gamma}} \text{ projects on a } |\boldsymbol{\gamma}|-\text{D canonical hyperplane},$$

Representing $\lambda(\boldsymbol{\xi})$

The vast majority of physics-based random signals exhibit a comparatively low interaction order between input variables so that:

$$\|\mathcal{P}_{\boldsymbol{\gamma}}\,\lambda\| \gg \left\|\mathcal{P}_{\boldsymbol{\gamma}',|\boldsymbol{\gamma}'|>|\boldsymbol{\gamma}|}\,\lambda - \mathcal{P}_{\boldsymbol{\gamma}}\,\lambda\right\|, \quad \mathcal{P}_{\boldsymbol{\gamma}} \text{ projects on a } |\boldsymbol{\gamma}|-\text{D canonical hyperplane},$$

$\Longrightarrow$ best suits the *High-Dimensional Model Representation* (HDMR), Rabitz & Alış (1999).

$$\begin{aligned}\lambda(\boldsymbol{\xi}) &= f_{\emptyset} + \sum_{i=1}^{d} f_i(\xi_i) + \sum_{i<j=1}^{d} f_{ij}(\xi_i,\xi_j) + \ldots + f_{12\ldots d}(\xi_1,\xi_2,\ldots,\xi_d),\\ &\approx \sum_{\boldsymbol{\gamma}\subseteq\{1,\ldots,d\}}^{|\boldsymbol{\gamma}|=N_l<d} f_{\boldsymbol{\gamma}}(\boldsymbol{\xi}_{\boldsymbol{\gamma}}) = f(\boldsymbol{\xi}).\end{aligned}$$

Functional basis for $\{f_\gamma\}$

$$\lambda(\boldsymbol{\xi}) \approx \sum_{\gamma \subseteq \{1,\ldots,d\}}^{|\gamma|=N_l} f_\gamma\left(\boldsymbol{\xi}_\gamma\right)$$

Approximation of the modes

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

$$f_\gamma \approx \hat{f}_\gamma \equiv \sum_{\boldsymbol{\alpha},|\boldsymbol{\alpha}|=|\gamma|} c_{\boldsymbol{\alpha},\gamma}\, \psi_{\boldsymbol{\alpha}}\left(\boldsymbol{\xi}_\gamma\right).$$

Functional basis for $\{f_\gamma\}$

$$\lambda(\boldsymbol{\xi}) \approx \sum_{\gamma \subseteq \{1,\ldots,d\}}^{|\gamma|=N_I} f_\gamma\left(\boldsymbol{\xi}_\gamma\right)$$

Approximation of the modes

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

$$f_\gamma \approx \hat{f}_\gamma \equiv \sum_{\boldsymbol{\alpha},|\boldsymbol{\alpha}|=|\gamma|} c_{\boldsymbol{\alpha},\gamma}\, \psi_{\boldsymbol{\alpha}}\left(\boldsymbol{\xi}_\gamma\right).$$

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

$\longrightarrow$ substitute a low-rank approximation:

$$f_\gamma \approx \tilde{f}_\gamma \equiv \sum_r \prod_{l=1}^{|\gamma|} \sum_{\alpha=1}^{p} c_{\alpha,\gamma,l,r}\, \psi_\alpha\left(\xi_{\gamma(l)}\right).$$

Functional basis for $\{f_\gamma\}$ (cont'd)

How to exploit *a priori* the likely sparsity of the approximation in $\{f_\gamma\}$?

NP-hard problem.

Functional basis for $\{f_\gamma\}$ (cont'd)

How to exploit *a priori* the likely sparsity of the approximation in $\{f_\gamma\}$?

NP-hard problem.

$$\boldsymbol{c} = \underset{\tilde{\boldsymbol{c}} \in \mathbb{R}^{|\mathcal{J}|}}{\arg\min} \|\boldsymbol{\lambda} - \Psi\, \tilde{\boldsymbol{c}}\|_2^2 + \tau \sum_{\gamma \in \{1,\ldots,d\}} \|\tilde{\boldsymbol{c}}\|_{K_\gamma},$$

$\tau > 0$ and K_γ a positive definite matrix. The *whole* set of predictors associated with a mode f_γ is treated together for subset selection:

$\longrightarrow$ speed-up subset selection step,

$\longrightarrow$ makes the subset selection more robust w.r.t. measurement noise.

Functional basis for $\{f_\gamma\}$ (cont'd)

How to exploit *a priori* the likely sparsity of the approximation in $\{f_\gamma\}$?

NP-hard problem.

$$\boldsymbol{c} = \underset{\tilde{\boldsymbol{c}} \in \mathbb{R}^{|\mathcal{J}|}}{\arg\min} \|\boldsymbol{\lambda} - \boldsymbol{\Psi}\, \tilde{\boldsymbol{c}}\|_2^2 + \tau \sum_{\gamma \in \{1,\ldots,d\}} \|\tilde{\boldsymbol{c}}\|_{K_\gamma},$$

$\tau > 0$ and K_γ a positive definite matrix. The *whole* set of predictors associated with a mode f_γ is treated together for subset selection:

$\longrightarrow$ speed-up subset selection step,

$\longrightarrow$ 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).

$\longrightarrow$ **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 $\left\{\tilde{f}_\gamma\right\}$ are *nonlinear*.

$\Longrightarrow$ Use the PCE-HDMR format for the subset selection *only* and *rk1-HDMR* for approximating $\lambda(\boldsymbol{\xi})$.

Solution process for $y(\boldsymbol{x}, \xi) \approx \sum_r w_r(x)\ \lambda_r(\xi)$

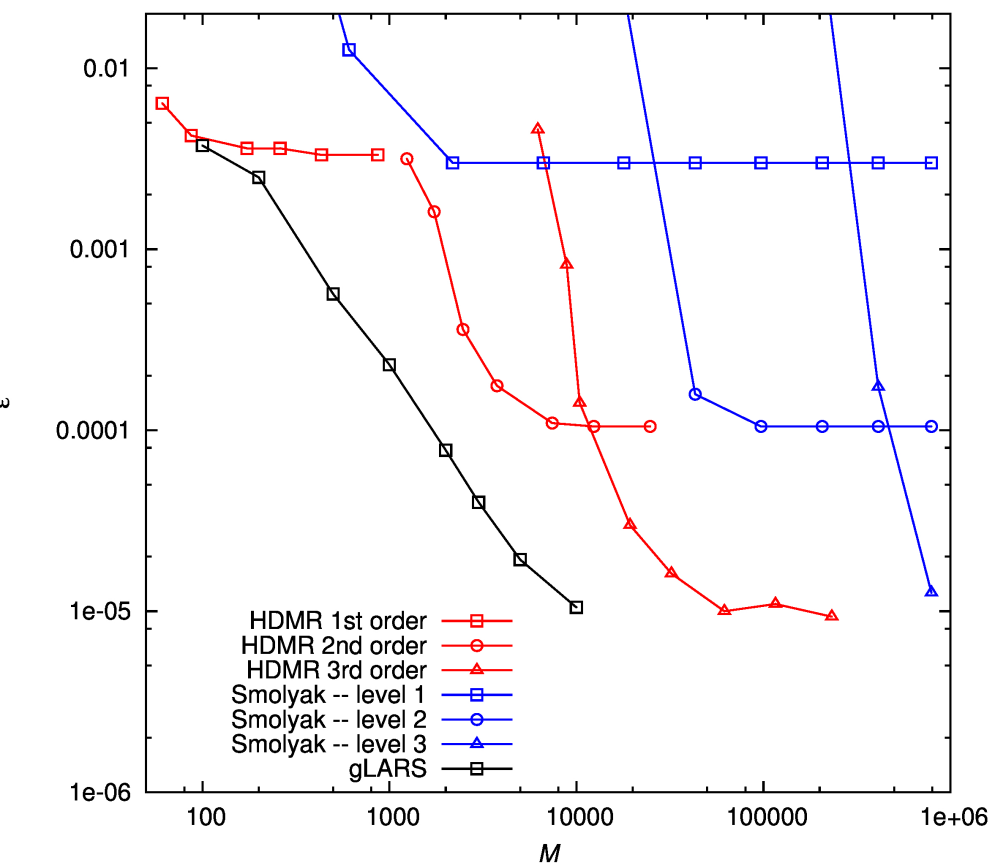
1. Choose p, N_l, r_{max}, N_x. Initialize $\{z_m = y_m\}_{m=1}^{M}$ and set $r \leftarrow 0$,
2. solve a deterministic problem for $\{c_{w,r}\}$: $\langle \lambda_r\, \Phi\, c_{w,r}, \lambda_r\, \phi_l \rangle_{M_{\text{coef}}} = \langle z, \lambda_r\, \phi_l \rangle_{M_{\text{coef}}}$ and normalize $w(\boldsymbol{x})$,
3. solve a stochastic problem for $\{c_{\gamma,r}\}$: $\langle w_r\, \Psi\, c_{\gamma,r}, w_r\, \psi_k \rangle_{M_{\text{coef}}} = \langle z, w_r\, \psi_k \rangle_{M_{\text{coef}}}$
   1. Initialize the stochastic approximation basis $\{f_\gamma\} = \emptyset$, $\boldsymbol{\Gamma} = \emptyset$,
   2. solve the (g)LASSO optimization problem for λ_r with the (g)LARS algorithm $\longrightarrow$ sequence of approximation bases with ordered indices $\{\gamma^{(n)}\}$,
   3. Set $n \leftarrow 0$. Solve the approximation problem:
   4. for the next index $\gamma^{(n+1)}$ in the sequence, activate the mode $f_{\gamma^{(n+1)}}$, $\boldsymbol{\Gamma}^{(n+1)} = \left\{\boldsymbol{\Gamma}^{(n)}, \gamma^{(n+1)}\right\}$,
   5. solve for the approximation coefficients $\{c_{\gamma,r}\}$ by (nested) Alternate Least-Squares over the predictors,

$$\left\{ \boldsymbol{c}_{\gamma,r} = \underset{\tilde{\boldsymbol{c}}_\gamma \in \mathbb{R}^{|\gamma|}}{\arg\min} \left\| \boldsymbol{z} - w_r\, \Psi_{\{\boldsymbol{\Gamma}^{(n+1)}\} \setminus \gamma}\, \boldsymbol{c}_{\{\boldsymbol{\Gamma}^{(n+1)}\} \setminus \gamma} - w_r\, \Psi_\gamma\, \tilde{\boldsymbol{c}}_\gamma \right\|_{M_{\text{coef}}}^2, \quad \forall \gamma \in \boldsymbol{\Gamma}^{(n+1}$$

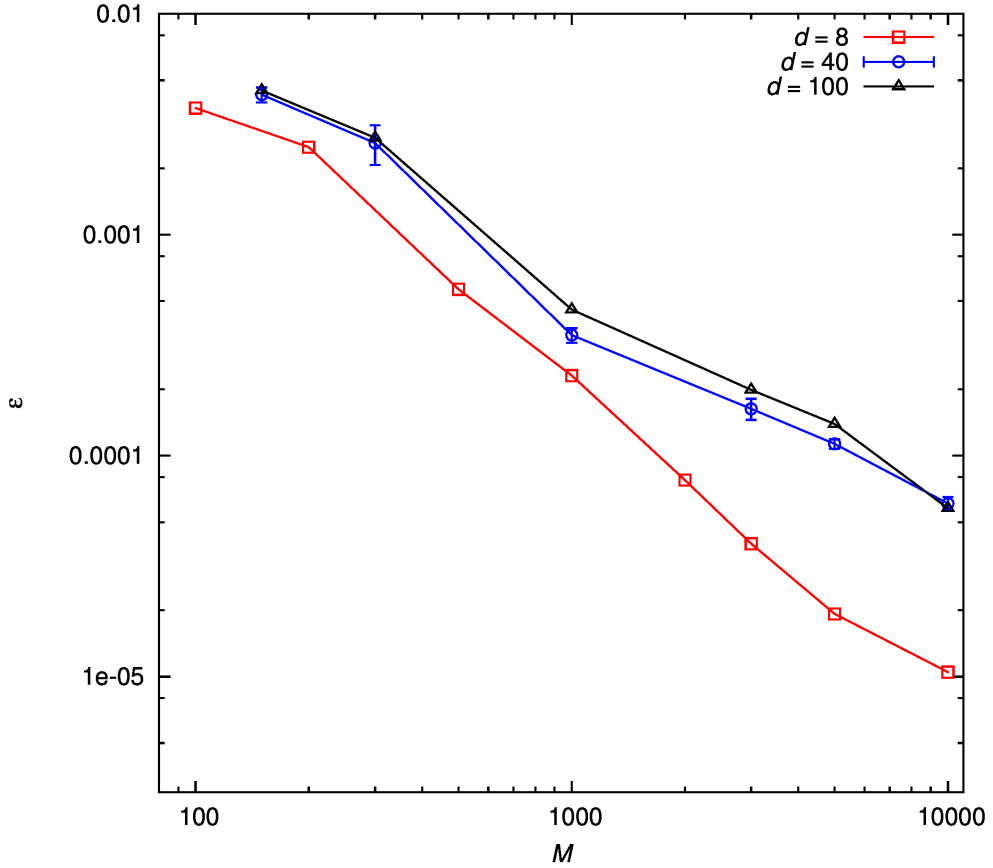
   6. estimate the relative approximation error ϵ by cross-validation. If ϵ decreases, $n \leftarrow n+1$ and go back to step 4. Otherwise, exit inner loop.
4. If $\|\lambda_r\|_{M_{\text{coef}}}$ converges, set $\boldsymbol{z} \leftarrow \boldsymbol{z} - w_r\, \lambda_r$, and $r \leftarrow r+1$. Iterate in step 2.

If one is given M samples...

- Stochastic convection-diffusion problem



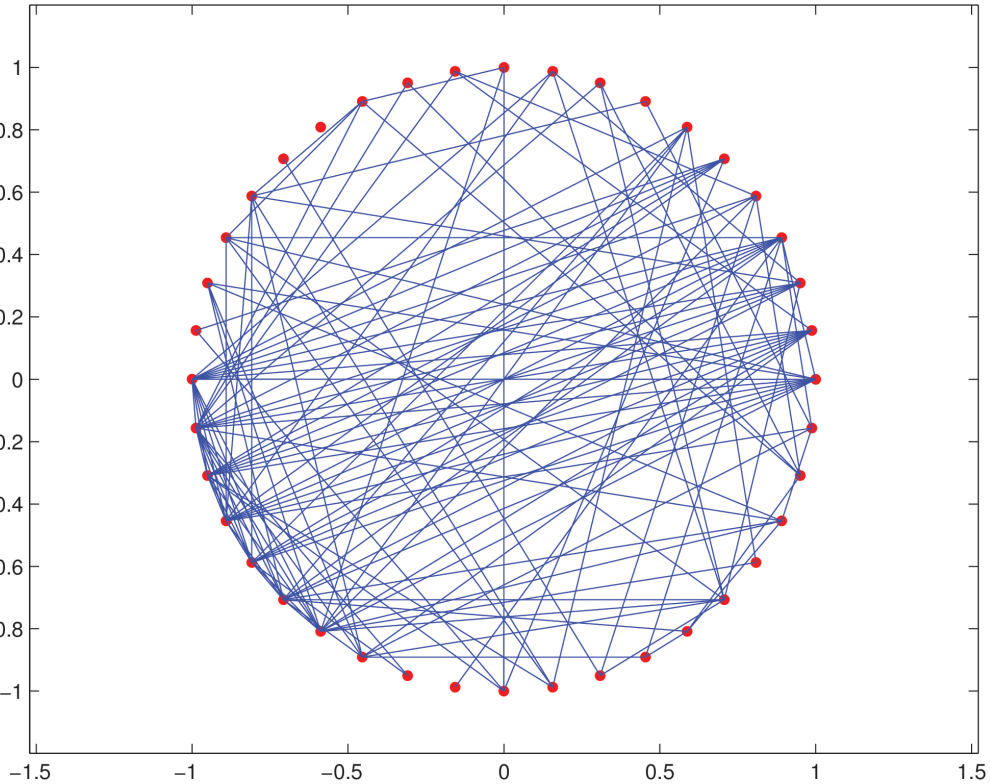
Error estimation when the amount of information varies. $d = 8$.

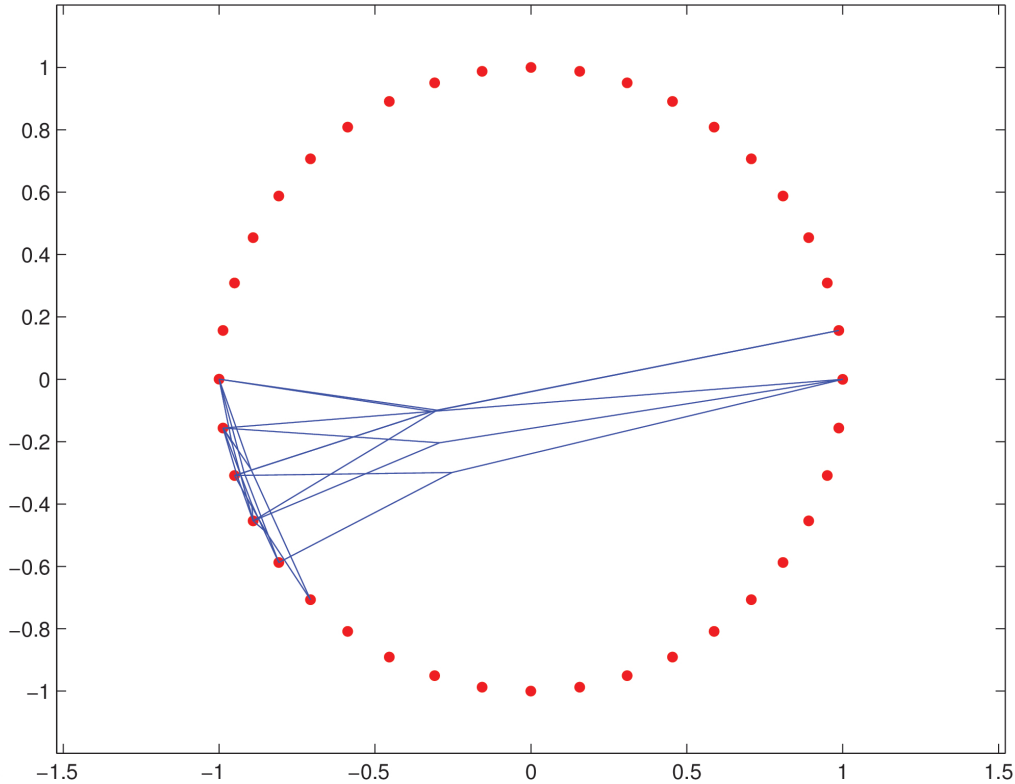


Impact of the dimensionality d onto the recovery performance
$\longrightarrow$ the anisotropy is successfully exploited.
If $d = 100 \longrightarrow \text{card}\{f_{\boldsymbol{\gamma}}\} = 3\,10^6$ (PC: $2\,10^{11}$).

Subset selection. Example for $d = 40$

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

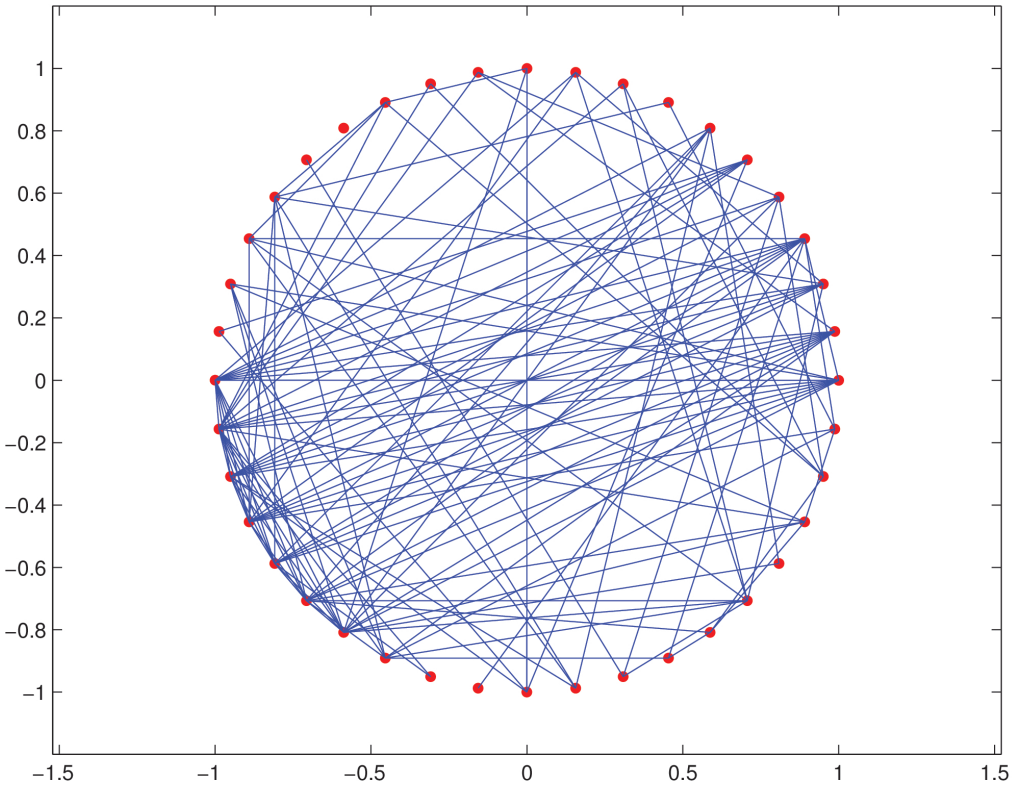




Connectivity of the 2nd order Connectivity of the 3rd order

$\longrightarrow$ essentially activates only a few dimensions.

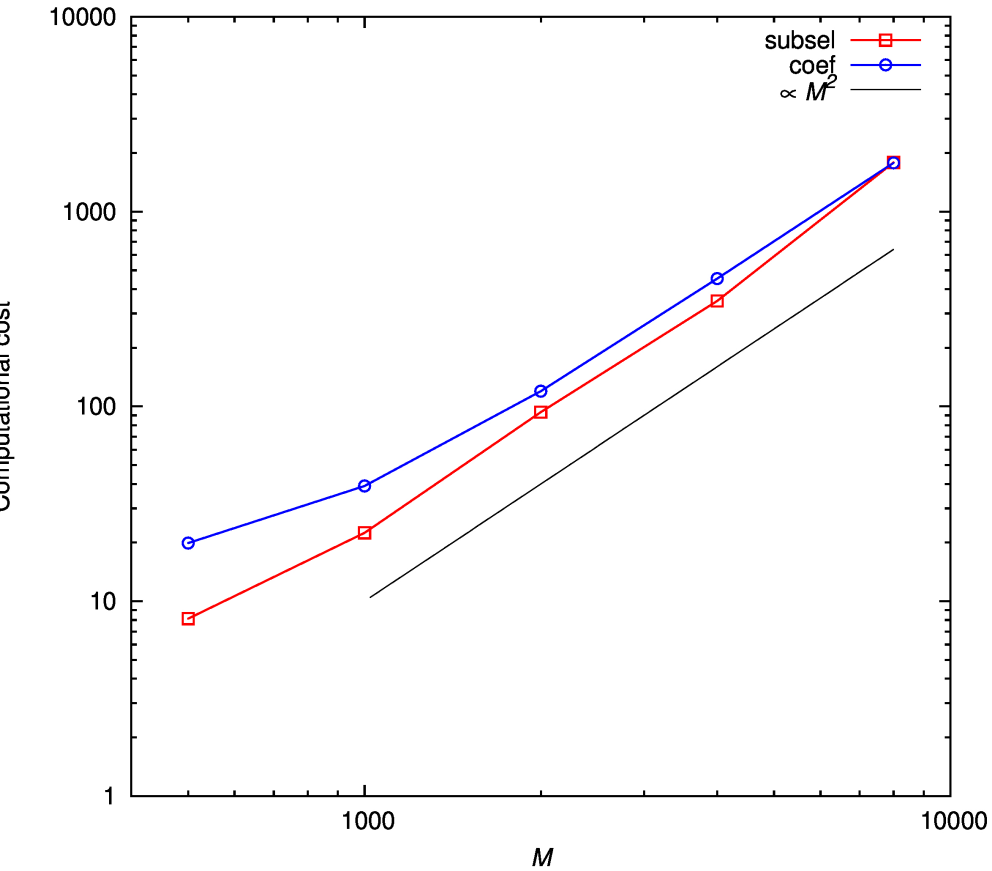
Connectivity of the 2nd order



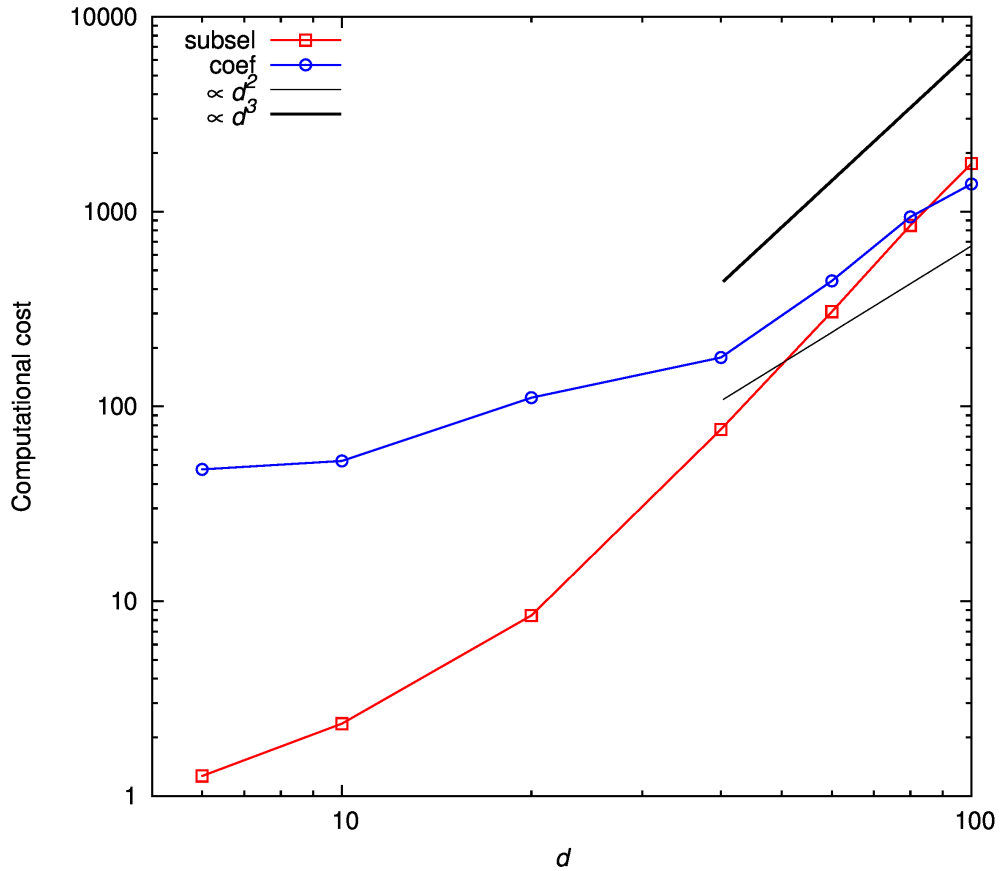
Complexity for approximating $\lambda(\xi)$

$$\mathscr{J}_{\text{subsel}} \sim \underbrace{\mathcal{O}\left(\frac{M^4}{p^2}\right)}_{\text{repeated LS pb}} + \underbrace{\mathcal{O}\left(M^2 \sum_{l=1}^{N_l} \frac{d!}{(d-l)!\,l!}\right)}_{\text{search for most correlated inactive mode}} \lessapprox \mathcal{O}\left(M^2\, d^{N_l}\, p\, {N_l}^{-1}\right),$$

$$\mathscr{J}_{\text{coefs}} \lessapprox \mathcal{O}\left(M^2\, d^{N_l}\, p\, {N_l}^{2}\right).$$

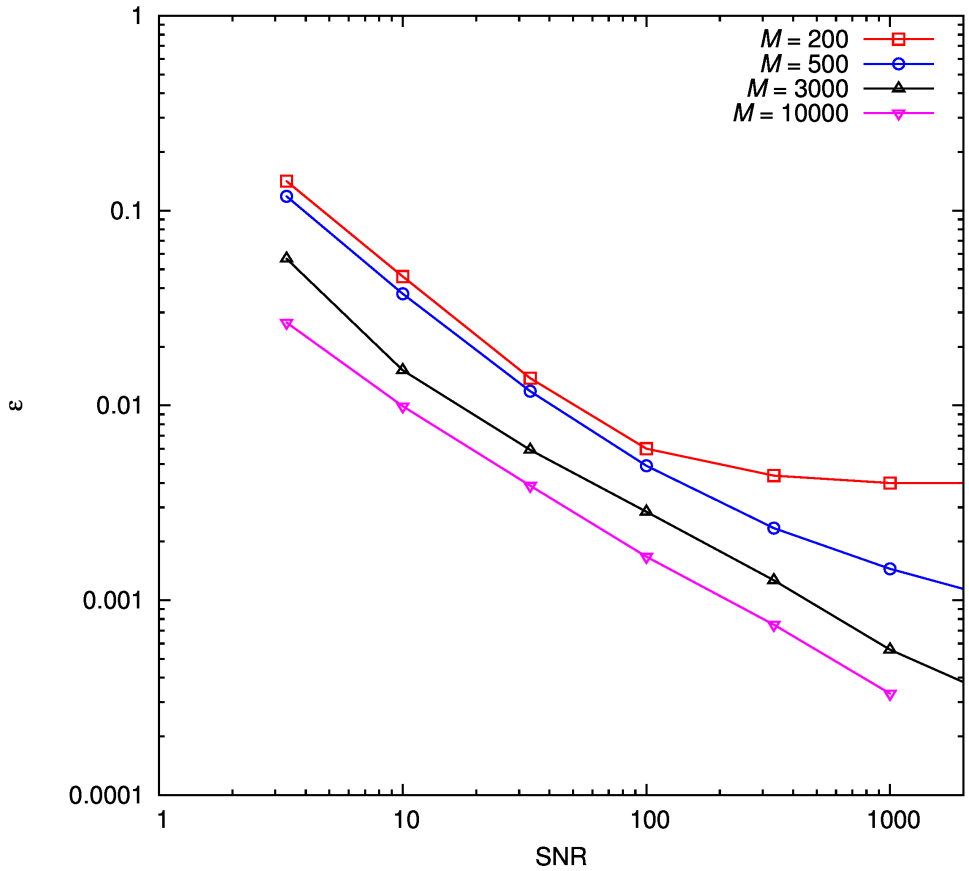


Scaling with # samples M ($N_l = 3$).



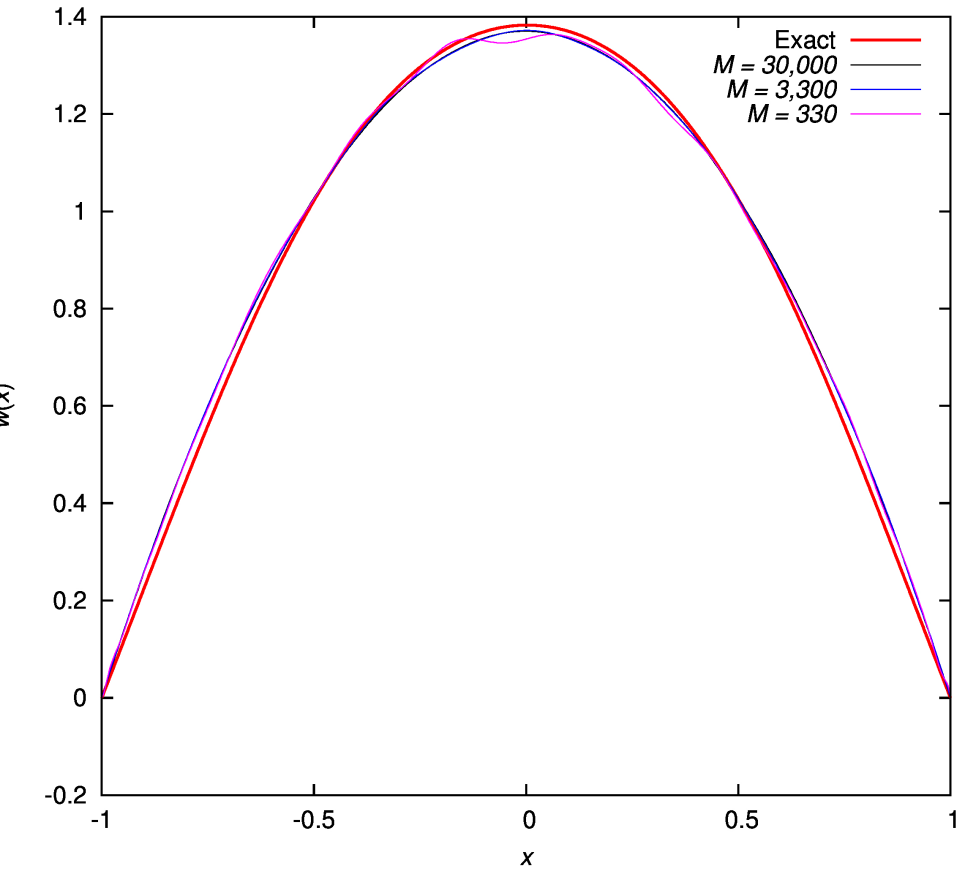
Scaling with the dimension d.

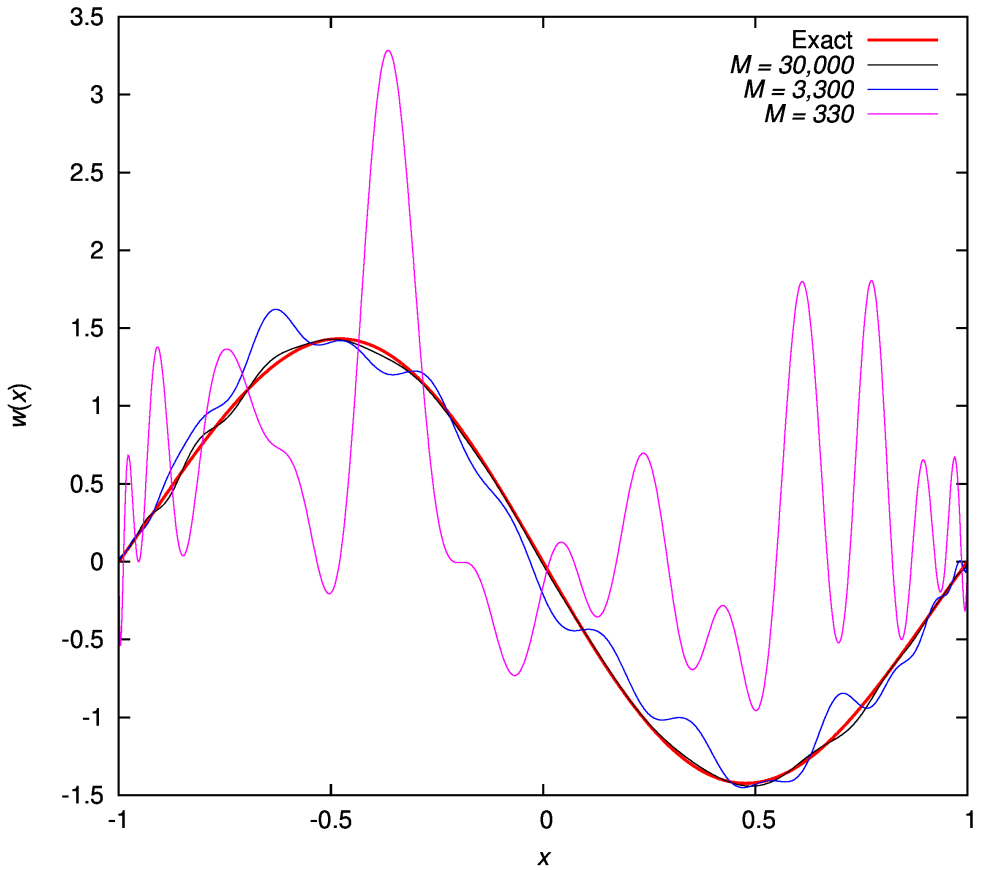
Influence of the identification uncertainty and measurement noise – $d = 40$



$\longrightarrow$ **Weighted Total Least Square formulation**

Spatial modes $\{w_r(\boldsymbol{x})\}$



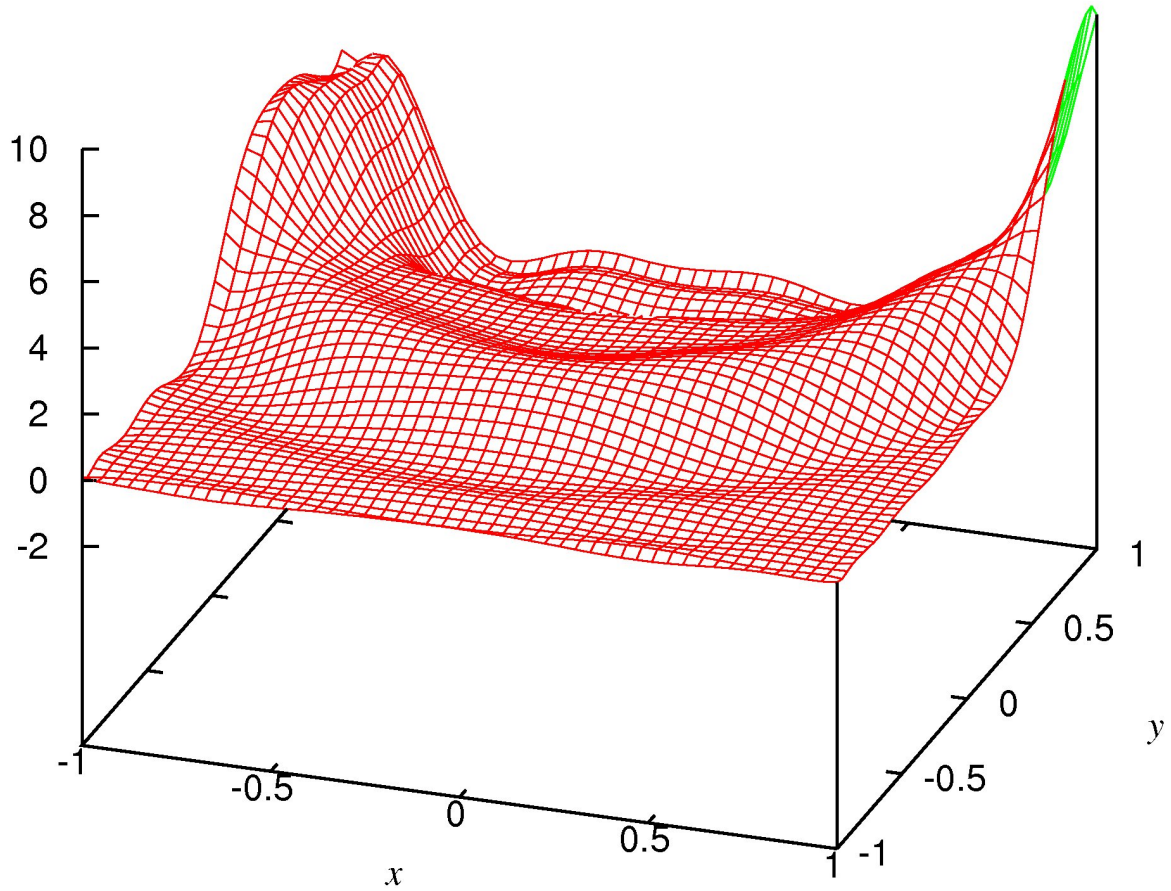


First two spatial modes. $d = 6$ (7-dimensional problem), # unknowns: 70,304.

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

$\longrightarrow$ they compare rather well with exact separated solution modes (from Karhunen-Loève).

Shallow Water Equations



First spatial mode. $d = 8$ (10-dimensional problem), # unknowns: 781,000+.

$M = 19,300 \Longrightarrow$ about 2.7 samples per dimension.

Concluding remarks

- A solution technique to get an accurate representation of input variables from experimental data to feed numerical models: $y(\boldsymbol{x}, \boldsymbol{\xi}) \approx \sum_r w_r(\boldsymbol{x})\ \lambda_r(\boldsymbol{\xi})$,
- few data available $\longrightarrow$ 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: $y(\boldsymbol{x}, \boldsymbol{\xi}) \approx \sum_r w_r(\boldsymbol{x})\ \lambda_r(\boldsymbol{\xi})$,
- few data available $\longrightarrow$ 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 $\lambda(\boldsymbol{\xi})$: CANDECOMP-like? HDMR? Tensor Trains? ...?
  $\longrightarrow$ format selection issue.

Concluding remarks

- A solution technique to get an accurate representation of input variables from experimental data to feed numerical models: $y(\boldsymbol{x}, \boldsymbol{\xi}) \approx \sum_r w_r(\boldsymbol{x})\ \lambda_r(\boldsymbol{\xi})$,
- few data available $\longrightarrow$ 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 $\lambda(\boldsymbol{\xi})$: CANDECOMP-like? HDMR? Tensor Trains? ...?
  $\longrightarrow$ 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

Format for $\{\lambda(\xi)\}$, a motivating example

Stochastic diffusion equation on $\Omega \times \Xi$, $\Omega = [x_-, x_+]$ with deterministic Dirichlet boundary conditions:

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

F and ν defined by

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

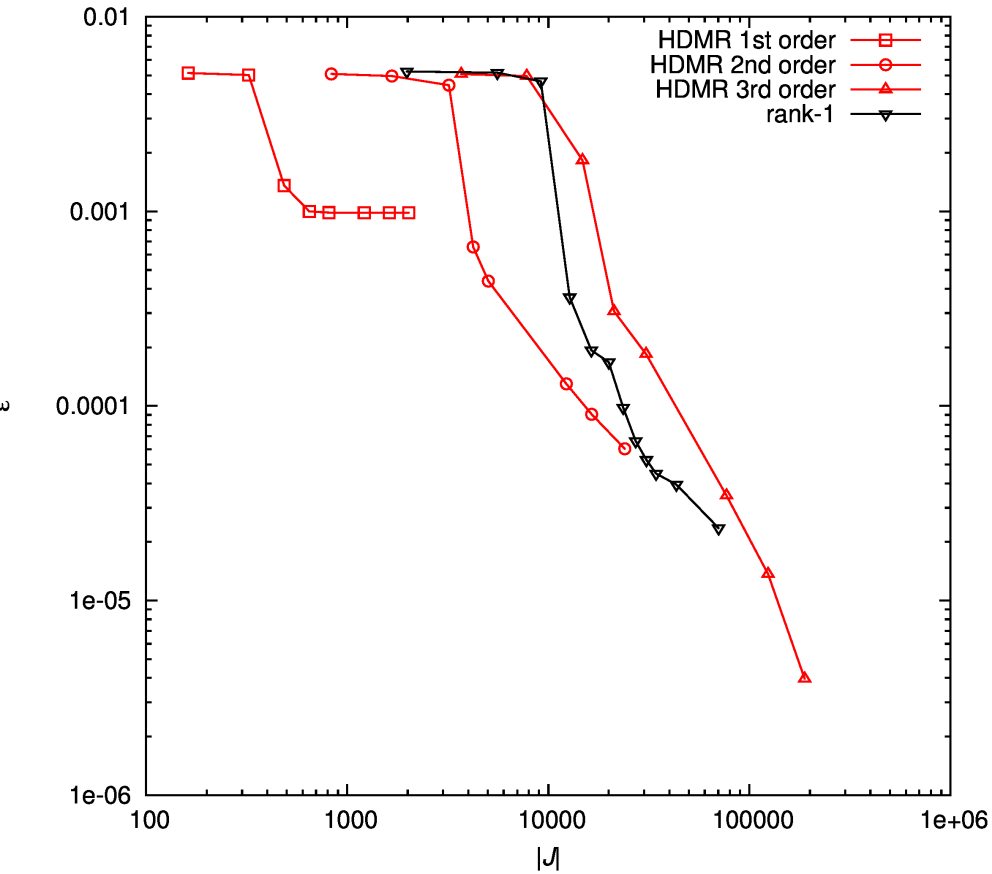
with $\nu_0 = 1$ and $F_0 = -1$. The spatial modes $\omega_{\nu,k}(x)$ and $\omega_{F,k}(x)$, and their associated amplitude $\sqrt{\sigma_{\nu,k}}$ and $\sqrt{\sigma_{F,k}}$, are the first dominant eigenfunctions of eigenproblems associated with Gaussian correlation kernels:

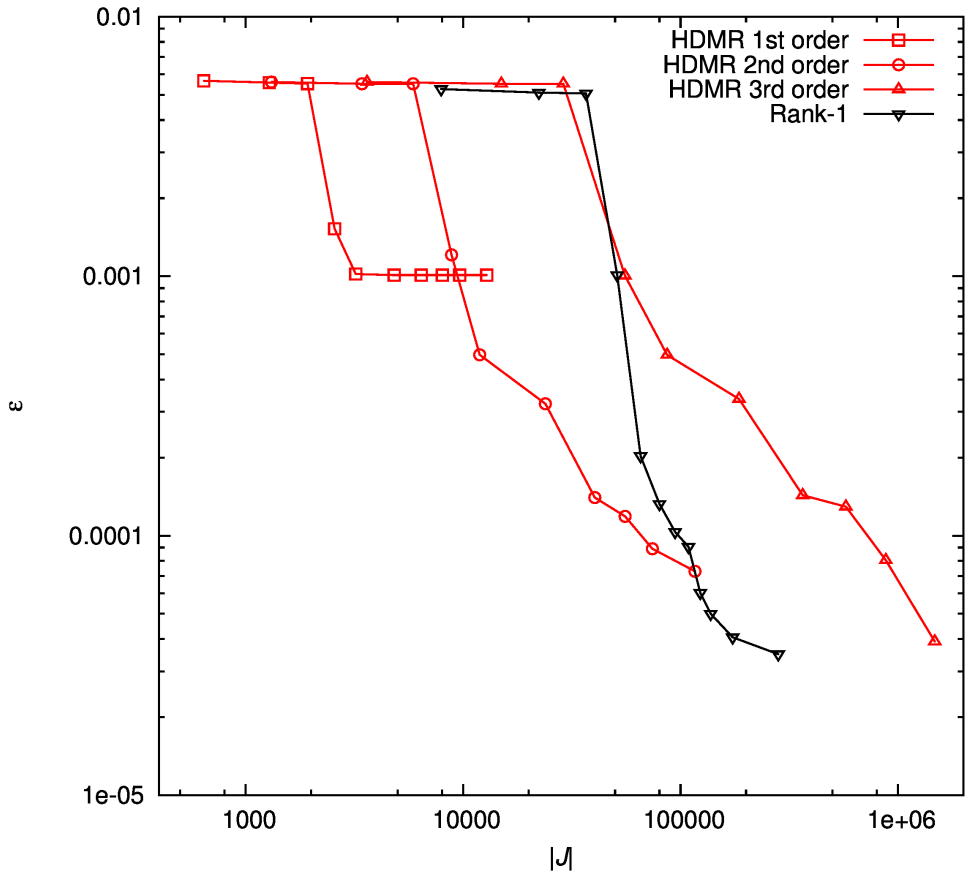
$$K_\nu(x,x') = \sigma_\nu^2\, e^{-\frac{(x-x')^2}{L_{c,\nu}^2}}, \qquad K_F(x,x') = \sigma_F^2\, e^{-\frac{(x-x')^2}{L_{c,F}^2}},$$

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 $y(\boldsymbol{x}^\star, \boldsymbol{\xi}) \equiv \lambda(\boldsymbol{\xi})$. Higher order HDMR modes activated upon sensitivity Sobol estimates at previous order.





$\longrightarrow$ HDMR is a sound choice, even with this naive basis adaption scheme.

Back...