# ETHzürich



# Dimensionality reduction and surrogate modelling for high-dimensional UQ problems

C. Lataniotis, S. Marelli, B. Sudret Chair of Risk, Safety & Uncertainty Quantification – ETH Zürich



### 1 Introduction

**2** Proposed scheme

### Ingredients

- 4 Validation
- **5** Summary & Outlook

### Problem description

#### UQ framework



#### Standard setup

- Prescribed probabilistic model (marginal PDFs and copula)
- Uncertainty propagation by surrogate models (Kriging, PCE)

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

### Problem description

#### UQ framework



#### New setup

- · High input dimensionality, with possible redundancy in the parameters
- Data sets available instead of prescribed PDFs

Goal of the PhD: optimally combine dimensionality reduction with surrogate modelling

### Applications of Interest

· Earthquake engineering





- Structural health monitoring





### Outline

### Introduction

### **2** Proposed scheme

### **3** Ingredients

### 4 Validation

### **5** Summary & Outlook

### Key ingredients

#### Surrogate models (SM)

A surrogate model  $\tilde{\mathcal{M}}$  is a function s.t.:

 $\mathcal{M}(\mathbf{x}) = \tilde{\mathcal{M}}(\mathbf{x}; \mathbf{\Omega}) + \epsilon$ 

- Inexpensive to evaluate
- Depends on configuration parameters Ω
- Non-intrusive: parameters are inferred from observations X, y

#### Examples of surrogate models:

- Kriging
- Polynomial chaos expansions
- Low-rank approximations

Dimensionality Reduction (DR) A transform  $\mathcal{X} \in \mathbb{R}^M \mapsto \mathcal{Z} \in \mathbb{R}^m$  of the form

$$\mathbf{z} = g(\mathbf{x}; \mathbf{w})$$

- Preserves some properties of x (e.g. information content)
- Parameters w are inferred from observations X

#### Examples of DR:

- PCA
- Kernel PCA
- Autoencoders

### Methodology

Given observations

$$\mathcal{X} = \left\{ \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}; \mathbf{x}^{(i)} \in \mathbb{R}^M \right\}$$

and system responses

$$\mathbf{y} = \left\{ y_1 = \mathcal{M}(\mathbf{x}^{(1)}), \ldots, y_N = \mathcal{M}(\mathbf{x}^{(N)}) \right\}$$

- Instead of computing surrogate  $\tilde{\mathcal{M}}(\mathbf{x})$  calculate  $\tilde{\mathcal{M}}(\mathbf{z})$  were

 $\mathbf{z} = g(\mathbf{x}, \mathbf{w}), \, \mathbf{z} \in \mathbb{R}^m, \, m < M$ 

 Tune DR parameters w having the surrogate model performance as objective

### Methodology

Given observations

$$\mathcal{X} = \left\{ \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}; \mathbf{x}^{(i)} \in \mathbb{R}^M \right\}$$

and system responses

$$\mathbf{y} = \left\{ y_1 = \mathcal{M}(\mathbf{x}^{(1)}), \ldots, y_N = \mathcal{M}(\mathbf{x}^{(N)}) \right\}$$

- Instead of computing surrogate  $\tilde{\mathcal{M}}(\mathbf{x})$  calculate  $\tilde{\mathcal{M}}(\mathbf{z})$  were

 $\mathbf{z} = g(\mathbf{x}, \mathbf{w}), \, \mathbf{z} \in \mathbb{R}^m, \, m < M$ 

 Tune DR parameters w having the surrogate model performance as objective

Treats dim. reduction (DR) and surrogate modelling (SM) as **black boxes** 

### Methodology

Given observations

$$\mathcal{X} = \left\{ \mathbf{x}^{(1)}, \, \dots, \mathbf{x}^{(N)}; \mathbf{x}^{(i)} \in \mathbb{R}^M 
ight\}$$

and system responses

$$\mathbf{y} = \left\{ y_1 = \mathcal{M}(\mathbf{x}^{(1)}), \ldots, y_N = \mathcal{M}(\mathbf{x}^{(N)}) \right\}$$

- Instead of computing surrogate  $\tilde{\mathcal{M}}(\mathbf{x})$  calculate  $\hat{\mathcal{M}}(\mathbf{z})$  were

 $\mathbf{z} = g(\mathbf{x}, \mathbf{w})$  ,  $\mathbf{z} \in \mathbb{R}^m$  , m < M

 Tune DR parameters w having the surrogate model performance as objective

Treats dim. reduction (DR) and surrogate modelling (SM) as black boxes



Hinton 2007 (neural network with autoencoder) Calandra et al. 2018 (GP regression with autoenc.)



DRSM

LErton, G. E. and Salakhutsinov, R. S. (2006). Reducing the dimensionality of data with neural networks. Science 313:8786, pp. 504-507. Calandro, R. and Paters J. and Raamasen C. E. and Delaemath M.P. (2016). Vanifold Gaussian processes for regression. In Neural Networks (LICNN), 2016. International Joint Conference, pp. 3388-3395. IEEE.

C. Lataniot's, S. Marchi, B. Sudret. (ETH Zürich).

### Motivation



Accuracy of surrogate model not correlated with accuracy of dimensionality reduction

Lataniotis, C. and Marolli, S. and Sudret, B. (2016). Combining frature mapping and Gaussian process modelling in the context of UQ. SIAM Cont. on Uncertainty Quantification (SIAVUQ 2016). Switzerland,

C. Lataniot's, S. Marchi, B. Sudret (ETH Zürich)

Goal: Find optimal w w.r.t. the surrogate modelling performance of  $\tilde{\mathcal{M}}(g(\mathbf{x}, \mathbf{w}))$ 



Given quantities: X, y

- Compute w\* using objective J(w):
  - Compute  $\mathcal{Z} = g(\mathcal{X}, \mathbf{w})$
  - Split Z into training  $(X_t)$  and test set  $(X_{cv})$  with  $Z_t \cup Z_{cv} = Z$  (resp. y to  $y_t, y_{cv}$ )
  - Compute surrogate  $\mathcal{M}(\mathbf{z})$  using  $\mathcal{Z}_t, \mathbf{y}_t$
  - Evaluate the surrogate response  $\hat{\mathcal{M}}(\mathcal{Z}_{cv})$
  - Compute objective (e.g. RMSE):

$$J(\mathsf{w}) = \left(rac{1}{N_{et} \operatorname{var}(y_{ev})} \sum_{k=1}^{N_{et}} \left(y_{ev}^{(i)} - \tilde{\mathcal{M}}\left(\mathbf{z}_{ev}^{(i)}
ight)
ight)^2
ight)^{1/2}$$



- Given quantities: X, y
- Compute w\* using objective J(w):
  - Compute  $\mathcal{Z} = g(\mathcal{X}, \mathbf{w})$
  - Split  $\mathcal{Z}$  into training  $(\mathcal{X}_t)$  and test set  $(\mathcal{X}_{cu})$  with  $\mathcal{Z}_t \cup \mathcal{Z}_{cu} = \mathcal{Z}$  (resp. y to  $\mathbf{y}_t, \mathbf{y}_{cu}$ )
  - Compute surrogate *M*(z) using *Z<sub>l</sub>*, y<sub>t</sub>
  - Evaluate the surrogate response  $\tilde{\mathcal{M}}(\mathcal{Z}_{cv})$
  - Compute objective (e.g. RMSE):

$$J(\mathbf{w}) = \left(\frac{1}{N_{ce}\operatorname{var}(y_{ce})}\sum_{i=1}^{N_{ce}}\left(y_{ce}^{(i)} - \tilde{\mathcal{M}}\left(\mathbf{z}_{ce}^{(i)}\right)\right)^2\right)^{1/2}$$



- Given quantities: X, y
- Compute w\* using objective J(w):
  - Compute  $\mathcal{Z} = g(\mathcal{X}, \mathbf{w})$
  - Split Z into training  $(X_t)$  and test set  $(X_{cv})$  with  $Z_t \cup Z_{cv} = Z$  (resp. y to  $\mathbf{y}_t, \mathbf{y}_{cv}$ )
  - Compute surrogate M(z) using Z<sub>l</sub>, y<sub>t</sub>
  - Evaluate the surrogate response \$\tilde{\mathcal{M}}(\mathcal{Z}\_{cv})\$
  - Compute objective (e.g. RMSE):

$$J(\mathbf{w}) = \left(\frac{1}{N_{\mathrm{ct}} \operatorname{var}(y_{\mathrm{ct}})} \sum_{i=1}^{N_{\mathrm{ct}}} \left(y_{\mathrm{ct}}^{(i)} - \tilde{\mathcal{M}}\left(\mathbf{z}_{\mathrm{ct}}^{(i)}\right)\right)^2\right)^{1/2}$$



- Given quantities: X, y
- Compute w\* using objective J(w):
  - Compute  $\mathcal{Z} = g(\mathcal{X}, \mathbf{w})$
  - Split Z into training  $(X_t)$  and test set  $(X_{cv})$  with  $Z_t \cup Z_{cv} = Z$  (resp. y to  $\mathbf{y}_t, \mathbf{y}_{cv}$ )
  - Compute surrogate  $\tilde{\mathcal{M}}(\mathbf{z})$  using  $\mathcal{Z}_t, \mathbf{y}_t$
  - Evaluate the surrogate response \$\bar{\mathcal{M}}\$ (\$\mathcal{Z}\_{cv}\$)
  - Compute objective (e.g. RMSE):

$$J(\mathbf{w}) = \left(\frac{1}{N_{\mathrm{ce}} \operatorname{var}(y_{\mathrm{ce}})} \sum_{i=1}^{N_{\mathrm{ce}}} \left(y_{\mathrm{ce}}^{(i)} - \tilde{\mathcal{M}}\left(\mathbf{z}_{\mathrm{ce}}^{(i)}\right)\right)^2\right)^{1/2}$$



- Given quantities: X, y
- Compute w\* using objective J(w):
  - Compute  $\mathcal{Z} = g(\mathcal{X}, \mathbf{w})$
  - Split  $\mathcal{Z}$  into training  $(\mathcal{X}_t)$  and test set  $(\mathcal{X}_{cv})$  with  $\mathcal{Z}_t \cup \mathcal{Z}_{cv} = \mathcal{Z}$  (resp. y to  $\mathbf{y}_t, \mathbf{y}_{cv}$ )
  - Compute surrogate  $\tilde{\mathcal{M}}(\mathbf{z})$  using  $\mathcal{Z}_t, \mathbf{y}_t$
  - Evaluate the surrogate response  $\tilde{\mathcal{M}}(\mathcal{Z}_{ev})$
  - Compute objective (e.g. RMSE):

$$J(\mathbf{w}) = \left(\frac{1}{N_{\mathrm{cv}} \operatorname{var}(y_{\mathrm{cv}})} \sum_{i=1}^{N_{\mathrm{cv}}} \left(y_{\mathrm{cv}}^{(i)} - \tilde{\mathcal{M}}\left(\mathbf{z}_{\mathrm{cv}}^{(i)}\right)\right)^2\right)^{1/2}$$



- Given quantities: X, y
- Compute w\* using objective J(w):
  - Compute  $\mathcal{Z} = g(\mathcal{X}, \mathbf{w})$
  - Split Z into training  $(X_t)$  and test set  $(X_{cv})$  with  $Z_t \cup Z_{cv} = Z$  (resp. y to  $\mathbf{y}_t, \mathbf{y}_{cv}$ )
  - Compute surrogate  $\tilde{\mathcal{M}}(\mathbf{z})$  using  $\mathcal{Z}_t, \mathbf{y}_t$
  - Evaluate the surrogate response  $\tilde{\mathcal{M}}(\mathcal{Z}_{ev})$
  - Compute objective (e.g. RMSE):

$$J(\mathbf{w}) = \left(\frac{1}{N_{ct}\operatorname{var}(y_{cv})}\sum_{i=1}^{N_{ct}}\left(y_{cv}^{(i)} - \tilde{\mathcal{M}}\left(\mathbf{z}_{cv}^{(i)}\right)\right)^2\right)^{1/2}$$

### Summary

#### DRSM ingredients

- Dimensionality reduction method
- Surrogate modelling method
- Objective function type
- Optimisation method

#### Remarks

- Only the forward DR step is required
- Simplified SMs favour computational efficiency
- Multiple CV-folds are preferred. Extreme case: Leave-One-Out error
- Global optimisation is expected to perform better

### Summary

#### DRSM ingredients

- Dimensionality reduction method : Kernel PCA
- Surrogate modelling method : Kriging
- Objective function type : Leave-One-Out RMSE
- Optimisation method : Genetic algorithm with BFGS refinement

#### Remarks

- Only the forward DR step is required
- Simplified SMs favour computational efficiency
- Multiple CV-folds are preferred. Extreme case: Leave-One-Out error
- Global optimisation is expected to perform better

One particular DRSM setup is explored next

### Outline

- **1** Introduction
- Proposed scheme

### Ingredients

- 4 Validation
- **5** Summary & Outlook

Recall PCA:



- Calculate the covariance matrix C = cov(X)
- Compute the eigen-decomposition of C:

$$Cv = \lambda v$$

- Keep V, an M imes m collection of m eigenvectors v with largest eigenvalues  $\lambda$
- Obtain Z by projecting X onto V:

$$\mathcal{Z} = \mathcal{X} V$$

#### C. Lataniot's, S. Marchi, B. Sudret. (ETH Zürich).

#### Kernel PCA (KPCA) is a non-linear extension of PCA



First maps the input to a high dimensional space (*feature space*):

$$\mathbf{x} \in \mathcal{R}^M \mapsto \Phi(\mathbf{x}) \in \mathcal{H}$$

then PCA is performed in this space

- Infinite dimensional feature spaces are a common choice
- Calculates Z as projections onto the first m eigenvectors of C<sub>H</sub> = cov [Φ(X)]

Problem: In general,  $C_{\mathcal{H}}$  and its eigenvectors cannot be computed

How to solve the eigen-decomposition problem in  ${\mathcal H}$ 

■ Consider Φ that:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$$

• Although  $v_i \rightarrow$  infinite dimensional,  $N \rightarrow$  finite, thus:

$$oldsymbol{v}_i = \sum_{j=1}^N lpha_j^{(i)} \Phi(\mathbf{x}_j) \,, \, i = 1, \, \ldots \,, m$$

where  $\pmb{lpha}^{(i)}$ : the expansion coefficients of  $\pmb{v}_i$ 

The eigen-decomposition problem:

$$\lambda v = C_{\mathcal{H}} v$$

can be cast as:

$$\lambda oldsymbol{lpha} = rac{1}{N} oldsymbol{K} oldsymbol{lpha}$$

where  $\boldsymbol{K}$  is the  $N \times N$  kernel matrix with elements  $K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$ 

By solving this eigen-decomposition the coefficients  $\alpha$  are obtained (not v!)

#### How to calculate $\ensuremath{\mathcal{Z}}$

The k-th component of the *i*-th sample of Z, denoted by  $z_i^{(k)}$ , is the projection of  $\Phi(\mathbf{x}_i)$  onto  $\mathbf{v}_k$ :

$$z_i^{(k)} = \Phi(\mathbf{x}_i)^T \mathbf{v}_k$$
$$= \sum_{j=1}^N \alpha_j^{(k)} \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$$
$$= \sum_{j=1}^N \alpha_j^{(k)} \kappa(\mathbf{x}_i, \mathbf{x}_j)$$

#### Remarks

- Some details are omitted here (*e.g.* centering the kernel)
- In contrast to PCA going from Z back to X is non-trivial (pre-image problem)
- The ARD kernel will be used:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j; \mathbf{w}) = \exp\left(-\frac{1}{2} \sum_{k=1}^M \frac{1}{w_k^2} \left(x_i^{(k)} - x_j^{(k)}\right)^2\right)$$

Schölkopf, B., Smola, A. J., and Müller, K.-R., (1999). Kernel Principal Component Analysis, Advances in Kernel Methods-Support Vector Learning, MIT Press, Cambridge, MA pp. 327-352.

Ham, J., Lee, D. D., Mika, S., and Schölkopf, B. (2004). A kernel view of the dimensionality reduction of manifolds. In Proceedings of the twenty-first international conference on Machine learning (p. 47), ACM.

#### C. Lataniotis, S. Marelli, B. Sudret (ETH Zürich)

### Kriging reminder

#### Kriging metamodelling in a nutshell

Universal Kriging:

$$\tilde{\mathcal{M}}(\boldsymbol{x}) = \boldsymbol{\beta}^T \boldsymbol{f}(\boldsymbol{x}) + \sigma^2 Z(\boldsymbol{x}, R(\boldsymbol{x}, \boldsymbol{x}_{\mathsf{ED}}; \boldsymbol{\theta}))$$
  
 $\hat{\boldsymbol{\theta}} = \arg\min J(\hat{\boldsymbol{\theta}})$ 

Objective function varies depending on the estimation method (e.g. maximum likelihood, cross-validation, etc. )



#### Using a Kriging metamodel as a surrogate

A Kriging metamodel forms a group of curves characterized by their:

Mean (model output):

$$\mu_{\dot{Y}}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{\beta} + \boldsymbol{r}(\boldsymbol{x})^T \boldsymbol{R}^{-1} \left(\boldsymbol{y} - \boldsymbol{F} \boldsymbol{\beta}\right)$$

Variance (local error estimate):

$$\sigma_{\hat{Y}}^2(x) = \sigma^2 \left(1 - r^T(x) R^{-1} r(x) + u^T(x) (F^T R^{-1} F)^{-1} u(x)
ight)$$

Ingredients Objective function

### DRSM objective: Kriging Leave-One-Out error

#### Recall DRSM algorithm



The Leave-One-Out RMSE error reads:

$$J(\mathbf{w}) = \left(\frac{1}{N \text{Var}\left[\mathbf{y}\right]} \sum_{i=1}^{N} \left(\mathcal{M}(\boldsymbol{z}_{i}) - \mu_{\hat{Y},(-i)}(\boldsymbol{z}_{i})\right)^{2}\right)^{1/2}$$

Ingredients Objective function

### DRSM objective: Kriging Leave-One-Out error

#### Recall DRSM algorithm



The Leave-One-Out RMSE error reads:

$$J(\mathbf{w}) = \left(\frac{1}{N \text{Var}\left[\mathbf{y}\right]} \sum_{i=1}^{N} \left(\mathcal{M}(\boldsymbol{z}_{i}) - \mu_{\hat{Y},(-i)}(\boldsymbol{z}_{i})\right)^{2}\right)^{1/2}$$

For Kriging surrogates  $\mu_{\hat{Y}_i(-i)}(z_i)$  can be computed efficiently

Duorule, O. (1983). Cross validation of kriging in a unique neighborhood. Vathematical Geology 15.6, pp. 697-699.

### Outline

- Introduction
- Proposed scheme
- **3** Ingredients
- 4 Validation
- **5** Summary & Outlook

### Sobol function

#### Model description



$$\mathcal{M}(\mathbf{x}) = \prod_{i=1}^{20} \frac{|4x_i - 2| + \alpha_i}{1 + a_i} , \, x_i \in [0, 1]$$

#### where

$$\boldsymbol{\alpha} = [1, 2, 5, 10, 20, 50, 100, 500, \dots, 500]^T$$

#### **DRSM** parameters

KPCA		
Kernel:	ARD	
Kriging		
Correlation function:	Matérn 5/2 (isotropic)	
Trend type:	linear	
Optimisation:	GA + BFGS	
Optimisation		
Method:	GA+ BFGS	
Population size:	60	
Max. Iterations:	100	
Conv. tolerance:	$10^{-3}$	

**Training**: 400 random samples **Validation**: 1000 random samples

### Sobol function: m vs surrogate model performance



C. Lataniot's, S. Marcili, B. Sudret. (ETH Zürich).

Validation Sobol function

### Sobol function: Optimal KPCA parameters (m = 4)

ARD Kernel reminder:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j; \mathbf{w}) = \exp\left(-\frac{1}{2} \sum_{k=1}^{M} \frac{1}{w_k^2} \left(x_i^{(k)} - x_j^{(k)}\right)^2\right)$$



 $1/w_i$  (ARD kernel)

**Global sensitivity analysis** 



### DRSM vs plain Kriging



### Electrical resistor networks

#### Dataset description



- Courtesy of SANDIA National Labs
- Network comprised of 80 uncertain resistors
- Output of interest is voltage at V
- Effect of resistors to V decays with distance

#### DRSM parameters

KPCA		
Kernel:	ARD	
Kriging		
Correlation function:	Matérn 5/2	
	(isotropic)	
Trend type:	linear	
Optimisation:	GA + BFGS	
Optimisation		
Method:	GA + BFGS	
Population size:	150	
Max. Iterations:	100	
Conv. tolerance:	$10^{-3}$	

**Training**: 500 samples **Validation**: 500 samples

Jakeman, D., Michael S. E., and Khachik S. (2015). Enhancing II-minimization estimates of polynomial chaos expansions using basis selection. Journal of Computational Physics 289, pp. 18-34.

### Resistor networks: m vs surrogate model performance



Selected m = 8

### Resistor networks: Optimal KPCA parameters (m = 8)



Validation Electrical resistor networks

### Resistor networks: DRSM vs plain Kriging



### Outline

- Introduction
- Proposed scheme
- **3** Ingredients
- 4 Validation
- **5** Summary & Outlook

### Summary & Outlook

#### Summary

- A new framework for SM with high-dimensional inputs is proposed
- It is currently at validation stage
- It shows superior performance compared to traditional approaches (disjoint DR and SM)
- An interesting by-product of DRSM with KPCA using ARD kernel: pseudo-sensitivity indices

#### Outlook

- Explore different setups, e.g. PCE instead of Kriging and Autoencoders instead of KPCA
- Test performance of various DRSM set-ups on problems of interest (e.g. time-series inputs)

### Summary & Outlook

#### Summary

- A new framework for SM with high-dimensional inputs is proposed
- It is currently at validation stage
- It shows superior performance compared to traditional approaches (disjoint DR and SM)
- An interesting by-product of DRSM with KPCA using ARD kernel: pseudo-sensitivity indices

#### Outlook

- Explore different setups, e.g. PCE instead of Kriging and Autoencoders instead of KPCA
- Test performance of various DRSM set-ups on problems of interest (e.g. time-series inputs)

## Thank you very much for your attention!



www.uqlab.com

Chair of Risk, Safety & Uncertainty Quantification

http://www.rsuq.ethz.ch

C. Lataniot's, S. Marchi, B. Sudret. (ETH Zürich).