Reduced Modeling for Inverse Problems

Olga Mula (Paris Dauphine University & Inria Paris)

Journée MaNu
organized online

15/10/2020

General Ideas

Inverse problems: given a set of observations, we look for the casual factors that produced them.

Data Assimilation: time dependent problems, forecasting.

Observations can be **noisy** and of very **different nature**.

This talk:

- Applications involving PDE models.
- We explore whether our algorithms are optimal in some sense.

Neutronics

Observations: Neutron flux

PDEs: Neutron Diffusion/Transport

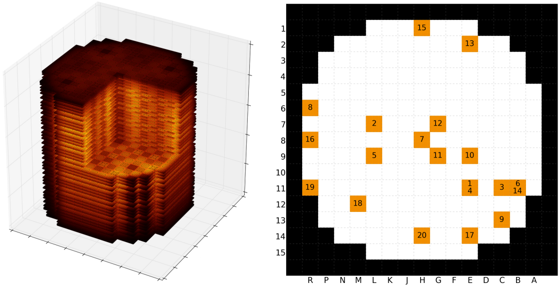


Figure: Sensor placement on a PWR for neutron flux reconstruction.

Structure of the talk

1. Mathematical framework for inverse problems involving PDE models
2. Optimal reconstruction benchmarks
3. Practical algorithms using model reduction and machine learning techniques

Collaborators

Theory: A. Cohen, J. Nichols, W. Dahmen, P. Binev, R. DeVore
Applications: F. Galarce, D. Lombardi, J.F. Gerbeau, J. Aghili, R. Chakir

Part I

Mathematical framework

Ref: [BCD$^+$17] **Data Assimilation in Reduced Modelling.** (SIAM UQ, 2017)

Mathematical setting

Ambient space V:

- Hilbert space over a domain $\Omega \subset \mathbb{R}^k$.
- Potentially very high or infinite dimension.

Parametrized PDE to model complex physical system:

$$\mathcal{B}(y)u = f(y)$$

where

$$y = (y_1, \dots, y_d) \in \mathrm{Y} \subset \mathbb{R}^d$$

is a vector of parameters ranging in some domain $\mathrm{Y} \subset \mathbb{R}^d$.

Parameter to solution map:

$$y \mapsto u(y) \in V$$

Solution manifold:

$$\mathcal{M} := \{u(y) \, : \, y \in \mathrm{Y}\} \subset V$$

is the set of all admissible solutions.

Mathematical setting

Forward problem: Given $y \in \mathrm{Y}$, compute $u(y)$ quickly.

Inverse problem: We observe a vector of linear measurements

$$z = (z_1, \dots, z_m) \in \mathbb{R}^m$$

where

$$z_i = \ell_i(u) = \langle w_i, u \rangle, \quad i = 1, \dots, m.$$

and ℓ_i are independent linear functionals (w_i are the Riesz representers).

Mathematical setting

Types of inverse problems: We have the forward mappings

$$y \in \mathrm{Y} \subset \mathbb{R}^d \quad \mapsto \quad u(y) \in \mathcal{M} \quad \mapsto \quad z \in \mathbb{R}^m$$

with $z_i = \ell_i(u)$.

We seek to approximate the inverse mappings:

- **State Estimation:**

$$z \mapsto u^*(z) \approx u$$

- **Parameter Estimation:**

$$z \mapsto y^*(z) \approx y$$

  when $z = \ell(u(y))$.

- In time-dependent problems: find initial condition, forecast of u...

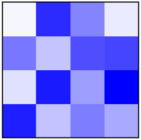
Severely ill-posed problems when $d > m$.

Elliptic PDE with piecewise constant diffusion field

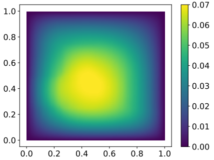
$$-\operatorname{div}(a\nabla u) = 1 \text{ on } \Omega = [0,1]^2, \text{ (well posed in } V = H_0^1(\Omega))$$

$$a = a(x,y) = 1 + 0.9 \sum_j y_j \chi_{D_j}(x), \quad y = (y_j) \in [-1,1]^{16}$$

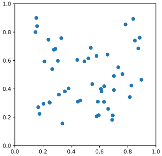
$$\ell_i(u) = \langle w_i, u \rangle = \int_\Omega e^{-\frac{||x-x_i||^2}{\sigma^2}} u(x)\mathrm{d}x$$



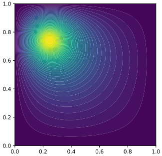
$x \mapsto a(x,y)$



$u(y)$



Pos. Sensors



w_i

Part II

Optimal reconstruction benchmarks

Ref: [CDMN20] **Nonlinear reduced models for state and parameter estimation** (arxiv, 2020)

State Estimation

Running Assumptions: No noise, no model error.

Goal: From the unknown $u \in \mathcal{M}$, we are given

$$\ell_i(u) = \langle \omega_i, u \rangle \,, \quad i = 1, \dots, m,$$

Defining the *sampling space*

$$W := \mathrm{span}\{\omega_1, \dots, \omega_m\}$$

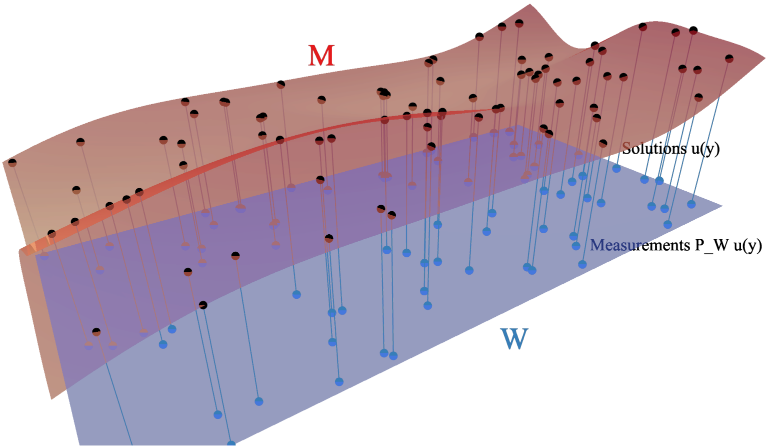
we have the equivalence

$$\ell_i(u), i = 1, \dots, m \quad \Leftrightarrow \quad \omega = P_W u.$$

Our task is to find a reconstruction algorithm

$$A : W \to V$$

such that $A(P_W u)$ approximates the state u.



We look for $A : W \to V$ such that $A(P_W u)$ approximates the state u.

Optimal reconstruction algorithms

The reconstruction performance of an algorithm $A : W \to V$ is

$$E(A, \mathcal{M}) = \max_{u \in \mathcal{M}} ||u - A(P_W u)||$$

and the optimal performance among all algorithms is

$$E^*(\mathcal{M}) = \min_{A: W \to V} E(A, \mathcal{M}).$$

There is a simple mathematical description of an optimal map A^*.

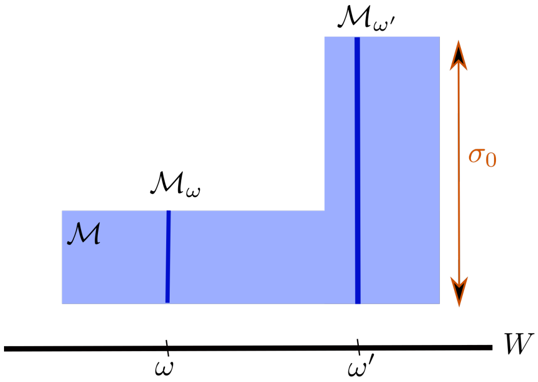
Optimal reconstruction algorithms

Manifold slices: For any $\omega \in W$, we define

$$\mathcal{M}_\omega := \{u \in \mathcal{M} \ : \ P_W u = \omega\}$$

The Chebyshev ball of $\mathcal{M}_\omega$ is the closed ball of minimal radius that contains $\mathcal{M}_\omega$.

Example: $V = \mathbb{R}^2$, $W = \text{span}\{e_x\}$ and $\mathcal{M}$ is L-shaped.



Optimal algorithms are not feasible in practice

Lemma: An optimal reconstruction map is given by

$$A^*_{\mathrm{wc}}(\omega) = \mathrm{cen}(\mathcal{M}_\omega)$$

where $\mathrm{cen}(\mathcal{M}_\omega)$ is the center of the Chebyshev ball of $\mathcal{M}_\omega$.

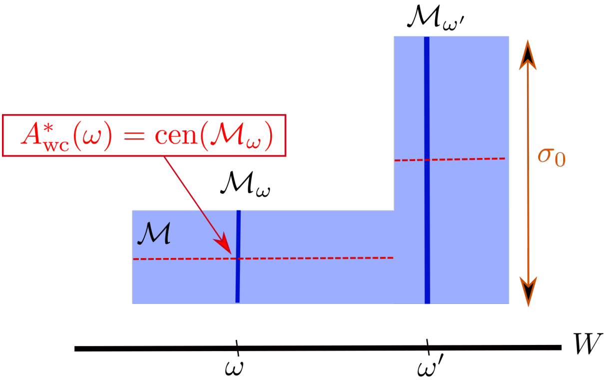
Reconstruction width: We define the diameter of $\mathcal{M}$ from W by

$$\sigma_0 := \sigma(\mathcal{M}, W) = \max\{\mathrm{diam}(\mathcal{M}_\omega) \, : \, \forall \omega \in W\}.$$

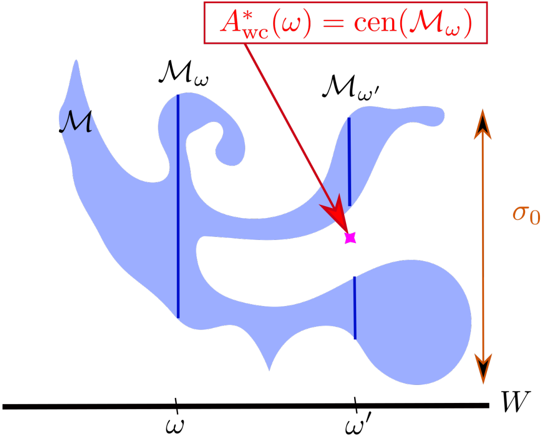
Any algorithm A cannot deliver a performance better than $\sigma_0/2$,

$$E^*(\mathcal{M}) = \sigma_0/2$$

Example: $V = \mathbb{R}^2$, $W = \text{span}\{e_x\}$ and $\mathcal{M}$ is L-shaped.



Example: $V = \mathbb{R}^2$, $W = \text{span}\{e_x\}$ and $\mathcal{M}$ is any shape.



Practical issue: A^*_{wc} is not easily computable since $\mathcal{M}$ may have a complicated geometry which is in general not given explicitly.

Part III

Practical algorithms using model reduction and machine learning techniques

▷ **Linear/Affine mappings** $A : W \to V$

- **Nonlinear mappings** $A : W \to V$
  - **Piecewise affine algorithms**
  - **Beyond piecewise affine:** OT, Neural Networks (ongoing works).

Ref: [CDD$^+$20] **Optimal reduced model algorithms for data-based state estimation** (SINUM, 2020)

Characterisation: For any given affine map $A : W \to V$ there exists an affine space $V_n^{\text{aff}} = \bar{u} + V_n$ of dimension $1 \leq n \leq m$ such that

$$A(\omega) = \underset{v \in \omega + W^\perp}{\arg\min} \; \text{dist}(v, \bar{u} + V_n), \quad \forall \omega \in W$$

where

$$\text{dist}(v, \bar{u} + V_n) = \|(v - \bar{u}) - P_{V_n}(v - \bar{u})\|.$$

Conversely: For any given affine space $V_n^{\text{aff}} = \bar{u} + V_n$, the above formula for $A(\omega)$ yields an affine reconstruction algorithm.

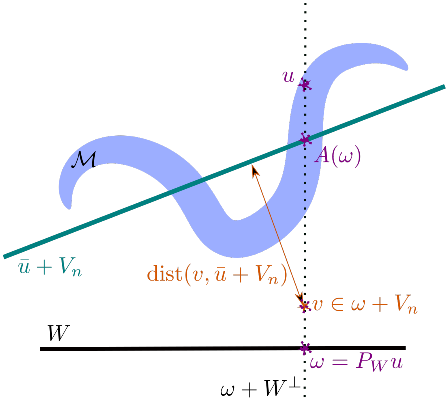
Common choices for V_n^{aff} are polynomials, reduced models...

Affine reconstruction algorithms

Practical computation: For a given $V_n^{\text{aff}} = \bar{u} + V_n$, computing

$$A(\omega) = \underset{v \in \omega + W^\perp}{\arg\min} \; \text{dist}(v, \bar{u} + V_n), \quad \forall \omega \in W$$

is easy (least-squares problem with a correction).



Affine reconstruction algorithms

$$A(\omega) = \underset{u \in \omega + W^\perp}{\arg\min}\ \mathrm{dist}(u, \bar{u} + V_n), \quad \forall \omega \in W$$

Error:

$$E(A, \mathcal{M}) = \max_{u \in \mathcal{M}} \|u - A(\omega)\| \leq \frac{1}{\beta(V_n, W_m)} \max_{u \in \mathcal{M}} \mathrm{dist}(u, \bar{u} + V_n)$$

where

$$\beta(V_n, W_m) := \inf_{v \in V_n} \frac{\|P_{W_m} v\|}{\|v\|} \in (0, 1]$$

plays the role of a stability constant. It can be interpreted as

$$\beta(V_n, W_m) = \cos(\theta_{V_n, W_m}), \quad \theta_{V_n, W_m} \in [0, \pi/2].$$

Limitations of Affine Algorithms

We said that if $A : W \to V$ is an affine mapping, then its image $A(W)$ is contained in a linear space of dimension $\leq m+1$.

So its performance is limited by below by the Kolmogorov $m+1$-width,

$$d_{m+1}(\mathcal{M}) := \min_{\substack{E \subseteq V \\ \dim(E) \leq m+1}} \max_{u \in \mathcal{M}} \operatorname{dist}(u, E)$$

in the sense that

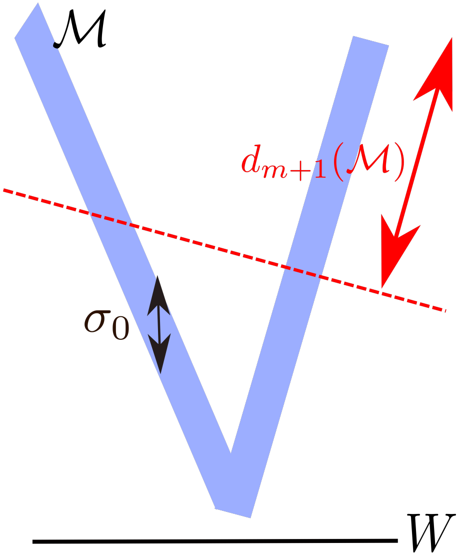
$$d_{m+1}(\mathcal{M}) \leq \min_{\substack{A:W \to V \\ A \text{ linear}}} \max_{u \in \mathcal{M}} ||u - A(P_W u)||.$$

Depending on $\mathcal{M}$ and W, we may have

$$\frac{1}{2}\sigma_0 = \min_{\substack{A:W \to V \\ A \text{ any mapping}}} \max_{u \in \mathcal{M}} ||u - A(P_W u)|| \ll d_{m+1}(\mathcal{M})$$

In order to overcome the limitation of $d_{m+1}(\mathcal{M})$ for the linear mappings, we have to build nonlinear ones.

We can have $\frac{1}{2}\sigma_0 \ll d_{m+1}(\mathcal{M})$



Part III

Practical algorithms using model reduction and machine learning techniques

- **Linear/Affine mappings** $A : W \to V$
- **Nonlinear mappings** $A : W \to V$
  - ▷ **Piecewise affine algorithms**
  - **Beyond piecewise affine:** OT, Neural Networks (ongoing works).

Ref: [CDMN20] **Nonlinear reduced models for state and parameter estimation** (arxiv, 2020)

Piecewise-affine algorithms

Consider a partition of the parameter domain

$$\mathrm{Y} = \mathrm{Y}_1 \cup \cdots \cup \mathrm{Y}_K \quad \rightsquigarrow \mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_K.$$

For each $\mathcal{M}_k$, we may construct a family of affine spaces

$$V_k, \quad \dim(V_k) = n_k, \quad k = 1, \dots, K$$

such that

$$\varepsilon_k := \max_{u \in \mathcal{M}_k} \operatorname{dist}(u, V_k)$$

and bounded inverse inf-sup constant

$$\beta_k := \beta(V_k, W) > 0.$$

Piecewise-affine algorithms

Consider a partition of the parameter domain

$$\mathrm{Y} = \mathrm{Y}_1 \cup \cdots \cup \mathrm{Y}_K \quad \rightsquigarrow \mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_K.$$

For each $\mathcal{M}_k$, we may construct a family of affine spaces

$$V_k, \quad \dim(V_k) = n_k, \quad k = 1, \dots, K$$

such that

$$\varepsilon_k := \max_{u \in \mathcal{M}_k} \mathrm{dist}(u, V_k)$$

and bounded inverse inf-sup constant

$$\beta_k := \beta(V_k, W) > 0.$$

For any prescribed $\varepsilon > 0$ and $1 \geq \beta > 0$, by taking K large enough, we may impose that

$$\max_{k=1,\dots,K} \varepsilon_k \leq \varepsilon \quad \text{and} \quad \max_{k=1,\dots,K} \beta_k \geq \beta > 0.$$

Model selection

To each V_k corresponds an affine algorithm A_k.

From the given data $\omega = P_W u$, we need to select between the reconstructions

$$u_k = A_k(w), \quad k = 0, \dots, K.$$

Model selection

To each V_k corresponds an affine algorithm A_k.

From the given data $\omega = P_W u$, we need to select between the reconstructions

$$u_k = A_k(w), \quad k = 0, \dots, K.$$

Note that since $u \in \mathcal{M}$ there exist $k = k(u)$ such that $u \in \mathcal{M}_k$. Therefore, for this particular k,

$$\|u - A_k(u)\| \leq \beta_k^{-1} \varepsilon_k \leq \beta^{-1} \varepsilon.$$

Model selection

To each V_k corresponds an affine algorithm A_k.

From the given data $\omega = P_W u$, we need to select between the reconstructions

$$u_k = A_k(w), \quad k = 0, \dots, K.$$

Note that since $u \in \mathcal{M}$ there exist $k = k(u)$ such that $u \in \mathcal{M}_k$. Therefore, for this particular k,

$$\|u - A_k(u)\| \leq \beta_k^{-1} \varepsilon_k \leq \beta^{-1} \varepsilon.$$

But this estimate is not feasible since it uses the knowledge of $k(u)$.

We only know the data ω and want to use it for selecting a $k^* = k(\omega)$.

Model selection

We would like to select the reconstruction that is closest to $\mathcal{M}$

$$k^* = k(\omega) = \mathrm{argmin}_{k=1,\dots,K} \mathrm{dist}(A_k(\omega), \mathcal{M}),$$

but

$$\mathrm{dist}(A_k(\omega), \mathcal{M}) := \min_{y \in \mathrm{Y}} \|u(y) - A_k(\omega)\|.$$

is not easily computable.

Model selection

We would like to select the reconstruction that is closest to $\mathcal{M}$

$$k^* = k(\omega) = \operatorname{argmin}_{k=1,\dots,K} \operatorname{dist}(A_k(\omega), \mathcal{M}),$$

but

$$\operatorname{dist}(A_k(\omega), \mathcal{M}) := \min_{y \in \mathrm{Y}} \|u(y) - A_k(\omega)\|.$$

is not easily computable.

In uniformly coercive problems, we have that the residual

$$\mathcal{R}(v, y) := \|\mathcal{B}(y)v - f(y)\|^2_{V'}, \quad \forall (v, y) \in V \times \mathrm{Y}$$

is uniformly equivalent to the norm

$$r\|v - u(y)\|_V \leq \mathcal{R}(v, y) \leq R\|v - u(y)\|_V, \quad \forall v \in V.$$

We can thus equivalently compute

$$\mathcal{S}(A_k(\omega), \mathcal{M}) := \min_{y \in \mathrm{Y}} \mathcal{R}(v, y), \qquad \xrightarrow[\min_{k=1,\dots,K}]{} \qquad \hat{k}(\omega)$$

which is a convex problem if affinely parametrized PDE.

Model selection

We define the δ-offset of $\mathcal{M}$

$$\mathcal{M}_\delta := \mathcal{M} + B(0, \delta)$$

and its diameter with respect to W

$$\sigma_\delta = \sigma_\delta(\mathcal{M}, W) := \max\{\|u - v\| \ : \ u, v \in \mathcal{M}_\delta, \ u - v \in W^\perp\}$$

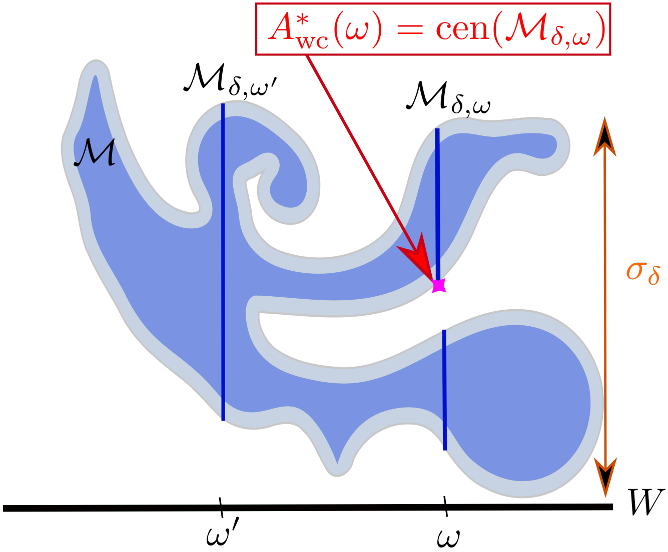
Theorem 1

For the above selection $\hat{k}(\omega)$ with the residual, we have

$$\frac{1}{2}\sigma_0 \leq \max_{u \in \mathcal{M}} \|u - A_{\hat{k}}(\omega)\| \leq \frac{1}{2}\sigma_{\kappa\beta^{-1}\varepsilon},$$

with $\kappa = R/r$.

We can make $\beta^{-1}\varepsilon \to 0$ by increasing K. In the limit, we reach the performance of the optimal algorithm.



$A^*_{\mathrm{wc}}(\omega) = \mathrm{cen}(\mathcal{M}_{\delta,\omega})$
$\mathcal{M}_{\delta,\omega'}$
$\mathcal{M}_{\delta,\omega}$
$\mathcal{M}$
σ_δ
W
ω'
ω

Practical algorithm for model selection

Goal: Generate a partition in Y such that

$$\max_{k=1,\dots,K} \varepsilon_k \leq \varepsilon \quad \text{and} \quad \max_{k=1,\dots,K} \beta_k \geq \beta.$$

or such that

$$\max_{k=1,\dots,K} \beta_k^{-1} \varepsilon_k \leq \delta$$

Dyadic partitioning: Step $j > 0$: We start from

$$Y = Y_1 \cup \dots Y_{K_j} \rightsquigarrow \mathcal{M} = \mathcal{M}_1 \cup \dots \cup \mathcal{M}_{K_j}.$$

For each k, we associate the hierarchy of reduced bases

$$V_{n,k} = \bar{u}_k + V_{n,k}, \quad n = 0, \dots, m,$$

with

$$V_{0,k} \subset \dots \subset V_{n,k} \subset \dots \subset V_{m,k}, \quad \dim(V_{n,k}) = n,$$

and

$$\text{dist}(\mathcal{M}_k, V_{n,k}) \leq \varepsilon_{n,k}, \quad \text{and} \quad \beta_{n,k} := \beta(V_{n,k}, W).$$

Practical algorithm for model selection

Split: Depending on the goal, define the test quantity

$$\tau_k = \min_{n=0,\dots,m} \max \left\{ \frac{\varepsilon_{n,k}}{\varepsilon}; \frac{\beta}{\beta_{n,k}} \right\}$$

or

$$\tau_k = \min_{n=0,\dots,m} \frac{\beta_{n,k}^{-1} \varepsilon_{n,k}}{\delta}.$$

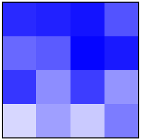
If

$$\tau_k > 1 \quad \Rightarrow \quad \text{Split cell } k.$$

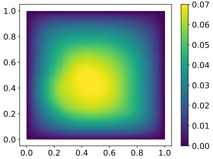
Elliptic PDE with piecewise constant diffusion field

$$-\operatorname{div}(a\nabla u) = 1 \text{ on } \Omega = [0,1]^2, \text{ (well posed in } V = H^1_0(\Omega))$$

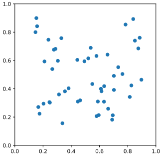
$$a = a(y) = 1 + \sum_j c_j y_j \chi_{D_j}, \quad y = (y_j) \in [-1,1]^{16}, \quad \ell_i(u) = \int_\Omega e^{-\frac{||x-x_i||^2}{\sigma^2}} u(x)\mathrm{d}x$$



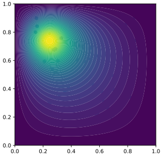
y



$u(y)$



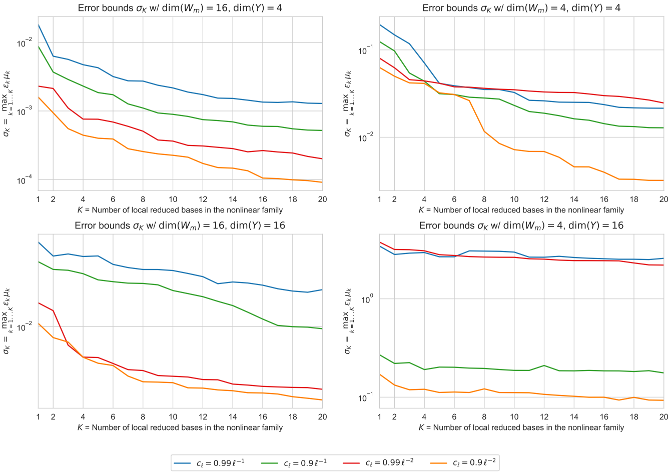
Pos. Sensors



w_i

$$c_j = \begin{cases} 0.9 j^{-2} & \text{elliptic ++} \\ 0.99 j^{-2} & \text{elliptic +} \\ 0.9 j^{-1} & \text{elliptic -} \\ 0.99 j^{-1} & \text{elliptic - -} \end{cases}$$

Numerical example



Part III

Practical algorithms using model reduction and machine learning techniques

- **Linear/Affine mappings** $A : W \to V$
- **Nonlinear mappings** $A : W \to V$
  - **Piecewise affine algorithms**
  - ▷ **Beyond piecewise affine:** OT, Neural Networks (ongoing works).

Ref: [ELMV20] **Nonlinear model reduction on metric spaces. Application to one-dimensional conservative PDEs in Wasserstein spaces** (M2AN, 2020)

Reconstruction with Machine Learning techniques

Reasons to go beyond piecewise affine:

- Partition may be suboptimal
- Can we discover in one step the best partition?
- In transport dominated problems, linear reduced spaces do not give much accuracy.

Ongoing research on:

- Can approximation classes such as Neural Networks help for forward and inverse reduced modelling?
- What can popular Machine Learning metrics such as Optimal Transport distances bring?

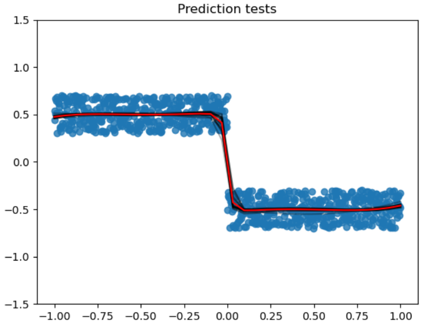
Neural Networks can discover the right partition

State: $u = (x, y) \in \mathcal{M}$ (step)

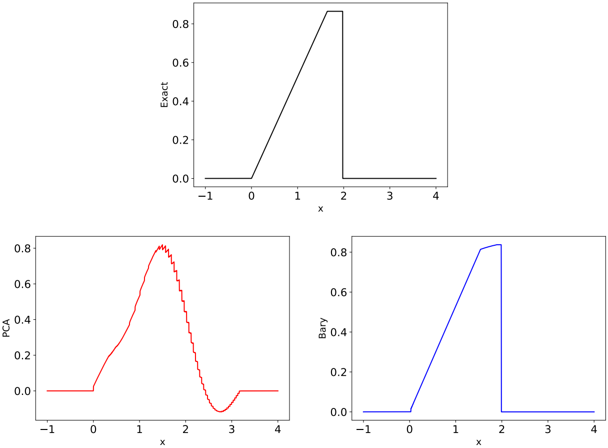
Observation $P_W u = x, \quad W = \text{span}\{e_x\}.$

Algorithm: $A : W \mapsto \mathbb{R}^2, \quad A(x) = (x, y), \quad A = \mathcal{NN}_\theta$

Training: $\min_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \| u_i - \mathcal{NN}_\theta(x_i) \|^2$



Approximation of Burgers equation with $n = 5$ modes.



$V = L_2$ $\qquad$ $V = W_2$

ROM using O.T. distances can reconstruct shocks with few modes.

CEMRACS 2021

Data Assimilation and Model Reduction in high-dimensional problems

CIRM, Luminy, Marseille.
July 17 — August 27, 2021

Organising Committee:

Virginie Ehrlacher
Damiano Lombardi
Olga Mula
Fabio Nobile
Tommaso Taddei

Scientific Committee:

Albert Cohen
Yvon Maday
Gianluigi Rozza
Karen Veroy

B. Adcock, A. C. Hansen, and C. Poon, *Beyond consistent reconstructions: optimality and sharp bounds for generalized sampling, and application to the uniform resampling problem*, SIAM Journal on Mathematical Analysis **45** (2013), no. 5, 3132–3167.

P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Data assimilation in reduced modeling*, SIAM/ASA Journal on Uncertainty Quantification **5** (2017), no. 1, 1–29.

P. Binev, A. Cohen, O. Mula, and J. Nichols, *Greedy algorithms for optimal measurements selection in state estimation using reduced models*, SIAM/ASA Journal on Uncertainty Quantification **6** (2018), no. 3, 1101–1126.

A. Cohen, W. Dahmen, R. DeVore, J. Fadili, O. Mula, and J. Nichols, *Optimal reduced model algorithms for data-based state estimation*, SIAM Journal on Numerical Analysis (2020).

A. Cohen, W. Dahmen, O. Mula, and J. Nichols, *Nonlinear reduced models for state and parameter estimation*, arXiv:2009.02687 (2020).

References II

V. Ehrlacher, D. Lombardi, O. Mula, and F.-X. Vialard, *Nonlinear model reduction on metric spaces. Application to one-dimensional conservative PDEs in Wasserstein spaces*, ESAIM, Math. Model. Numer. Anal. (2020).

F. Galarce, D. Lombardi, and O. Mula, *Reconstructing haemodynamics quantities of interest from doppler ultrasound imaging*, Submitted (2020).

Y. Maday, A. T. Patera, J. D. Penn, and M. Yano, *A parameterized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics*, International Journal for Numerical Methods in Engineering **102** (2015), no. 5, 933–965.