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.
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

\[
y = (y_1, \ldots, y_d) \in Y \subset \mathbb{R}^d
\]

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

**Parameter to solution map:**

\[
y \mapsto u(y) \in V
\]

**Solution manifold:**

\[
\mathcal{M} := \{ u(y) : y \in Y \} \subset V
\]

is the set of all admissible solutions.
Mathematical setting

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

Inverse problem: We observe a vector of linear measurements

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

where

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

and $\ell_i$ are independent linear functionals ($w_i$ are the Riesz representers).
Types of inverse problems: We have the forward mappings

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

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

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

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

\[\ell_i(u) = \langle w_i, u \rangle = \int_\Omega e^{- \frac{||x - x_i||^2}{\sigma^2}} u(x) dx\]

\[x \mapsto a(x, y) \quad u(y) \quad \text{Pos. Sensors} \quad w_i\]
Part II

Optimal reconstruction benchmarks

Ref: [CDMN20] Nonlinear reduced models for state and parameter estimation (arxiv, 2020)
Running Assumptions: No noise, no model error.

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

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

Defining the sampling space

$$\mathcal{W} := \text{span}\{\omega_1, \ldots, \omega_m\}$$

we have the equivalence

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

Our task is to find a reconstruction algorithm

$$A : \mathcal{W} \rightarrow V$$

such that $A(P_W u)$ approximates the state $u$. 
We look for $A : W \rightarrow V$ such that $A(P_W u)$ approximates the state $u$. 
The reconstruction performance of an algorithm $A : W \rightarrow V$ is

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

and the optimal performance among all algorithms is

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

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.
Lemma: An optimal reconstruction map is given by

$$A_{wc}^*(\omega) = \text{cen}(\mathcal{M}_\omega)$$

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

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

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

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

$$A_{wc}^*(\omega) = \text{cen}(\mathcal{M}_\omega)$$

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 \( A : W \rightarrow V \)
- Nonlinear mappings \( A : W \rightarrow V \)
  - 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 $A : W \rightarrow V$ there exists an affine space $V_{n}^{\text{aff}} = \bar{u} + V_{n}$ of dimension $1 \leq n \leq m$ such that

$$A(\omega) = \arg \min_{v \in \omega + W} \text{dist}(v, \bar{u} + V_{n}), \quad \forall \omega \in W$$

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 $V_{n}^{\text{aff}}$ are polynomials, reduced models...
**Practical computation:** For a given $V_n^{\text{aff}} = \bar{u} + V_n$, computing

$$A(\omega) = \arg \min_{v \in \omega + W} \text{dist}(v, \bar{u} + V_n), \quad \forall \omega \in W$$

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

\[
A(\omega) = \arg\min_{u \in \omega + W^\perp} \text{dist}(u, \bar{u} + V_n), \quad \forall \omega \in W
\]

Error:

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

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

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

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

with $\dim(E) \leq m + 1$ in the sense that

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

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

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

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

Ref: [CDMN20] Nonlinear reduced models for state and parameter estimation (arxiv, 2020)
Consider a partition of the parameter domain

\[ Y = Y_1 \cup \cdots \cup Y_K \quad \Rightarrow \quad \mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_K. \]

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

\[ V_k, \quad \dim(V_k) = n_k, \quad k = 1, \ldots, K \]

such that

\[ \varepsilon_k := \max_{u \in \mathcal{M}_k} \text{dist}(u, V_k) \]

and bounded inverse inf-sup constant

\[ \beta_k := \beta(V_k, W) > 0. \]
Consider a partition of the parameter domain

\[ Y = Y_1 \cup \cdots \cup Y_K \implies \mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_K. \]

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

\[ V_k, \quad \dim(V_k) = n_k, \quad k = 1, \ldots, K \]

such that

\[ \varepsilon_k := \max_{u \in \mathcal{M}_k} \text{dist}(u, V_k) \]

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

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

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

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

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

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

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

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

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

But this estimate is not feasible since it uses the knowledge of $k(u)$. We only know the data $\omega$ and want to use it for selecting a $k^* = k(\omega)$. 
Model selection

We would like to select the reconstruction that is closest to $M$

$$k^* = k(\omega) = \arg\min_{k=1,\ldots,K} \text{dist}(A_k(\omega), M),$$

but

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

is not easily computable.
Model selection

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

$$k^* = k(\omega) = \arg\min_{k=1,\ldots,K} \text{dist}(A_k(\omega), \mathcal{M}),$$

but

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

is not easily computable.

In uniformly coercive problems, we have that the residual

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

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

$$S(A_k(\omega), \mathcal{M}) := \min_{y \in Y} \mathcal{R}(v, y), \quad \rightarrow \quad \hat{k}(\omega)$$

which is a convex problem if affinely parametrized PDE.
We define the $\delta$-offset of $\mathcal{M}$

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

and its diameter with respect to $\mathcal{W}$

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

**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.*
\[ A_{\text{wc}}^*(\omega) = \text{cen}(\mathcal{M}_{\delta,\omega}) \]
Practical algorithm for model selection

**Goal:** Generate a partition in $Y$ such that

$$\max_{k=1,\ldots,K} \epsilon_k \leq \epsilon \quad \text{and} \quad \max_{k=1,\ldots,K} \beta_k \geq \beta.$$ 

or such that

$$\max_{k=1,\ldots,K} \beta_k^{-1} \epsilon_k \leq \delta$$

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

$$Y = Y_1 \cup \ldots Y_{K_j} \leadsto M = M_1 \cup \ldots \cup M_{K_j}.$$ 

For each $k$, we associate the hierarchy of reduced bases

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

with

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

and

$$\text{dist}(M_k, V_{n,k}) \leq \epsilon_{n,k}, \quad \text{and} \quad \beta_{n,k} := \beta(V_{n,k}, W).$$
Split: Depending on the goal, define the test quantity

\[ \tau_k = \min_{n=0,\ldots,m} \max \left\{ \frac{\epsilon_{n,k}}{\epsilon}; \frac{\beta}{\beta_{n,k}} \right\} \]

or

\[ \tau_k = \min_{n=0,\ldots,m} \frac{\beta_{n,k}^{-1}\epsilon_{n,k}}{\delta}. \]

If

\[ \tau_k > 1 \quad \Rightarrow \quad \text{Split cell } k. \]
Elliptic PDE with piecewise constant diffusion field

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

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

\[c_j = \begin{cases} 
0.9j^{-2} & \text{elliptic ++} \\
0.99j^{-2} & \text{elliptic +} \\
0.9j^{-1} & \text{elliptic -} \\
0.99j^{-1} & \text{elliptic - -} 
\end{cases}\]
Numerical example

Error bounds $\sigma_k$ w/ $\dim(W_m) = 16$, $\dim(Y) = 4$

$\sigma_k = \max_{\ell = 1, \ldots, K} \| e_{\ell, k} \|_U$

Error bounds $\sigma_k$ w/ $\dim(W_m) = 4$, $\dim(Y) = 4$

$\sigma_k = \max_{\ell = 1, \ldots, K} \| e_{\ell, k} \|_U$

Error bounds $\sigma_k$ w/ $\dim(W_m) = 16$, $\dim(Y) = 16$

$\sigma_k = \max_{\ell = 1, \ldots, K} \| e_{\ell, k} \|_U$

Error bounds $\sigma_k$ w/ $\dim(W_m) = 4$, $\dim(Y) = 16$

$\sigma_k = \max_{\ell = 1, \ldots, K} \| e_{\ell, k} \|_U$

$c_1 = 0.99 t^{-3}$
$c_1 = 0.9 t^{-3}$
$c_1 = 0.99 t^{-2}$
$c_1 = 0.9 t^{-2}$

Olga MULA (Paris Dauphine)  ROM for Inverse Problems
Part III
Practical algorithms using model reduction and machine learning techniques

- Linear/Affine mappings $A : W \rightarrow V$
- Nonlinear mappings $A : W \rightarrow V$
  - 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)
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: \( A : W \mapsto \mathbb{R}^2, \quad A(x) = (x, y), \quad A = \mathcal{N}\mathcal{N}_\theta \)

Training: \( \min_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \| u_i - \mathcal{N}\mathcal{N}_\theta(x_i) \|^2 \)
Approximation of Burgers equation with $n = 5$ modes.

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


