Estimation de systèmes peu observables en grande dimension par apprentissage statistique

K. KASPER[1], L. MATHELIN[2] & H. Abou-Kandil[1]

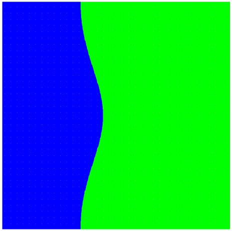
[1]: SATIE - ENS Cachan
[2]: LIMSI, Orsay

Motivation: Inverse problems (IP)

Given set of data S + model F describing state/solution $\boldsymbol{u}$:

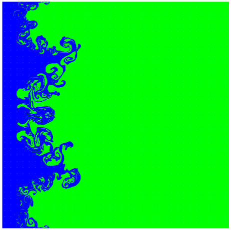
$\longrightarrow$ estimate parameters $\boldsymbol{y}$

parameters: $\boldsymbol{y} \approx D\,\boldsymbol{x}$



space sometimes HD (spatial or temporal fields)

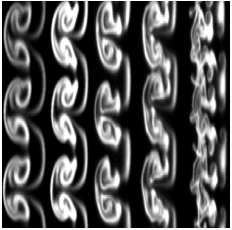
state: $\boldsymbol{u} \approx F\boldsymbol{y}$



F: forward model (expensive, uncertain, non-linear)

state space is HD

data: $S \approx O\boldsymbol{u}$



O: observation operator

data limited, noisy & indirect

Motivation

Goal

Inference of a high-dimensional spatial field from a few sensors.

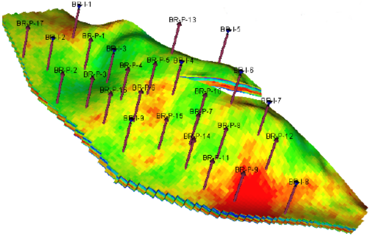


Illustration adapted from Jafarpour & Khaninezhad, 2015.

Motivation

Goal
Inference of a high-dimensional spatial field from a few sensors.

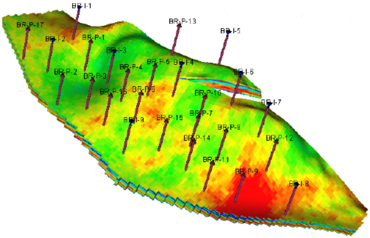


Illustration adapted from Jafarpour & Khaninezhad, 2015.

Context

- High-dimensional spatial field
- Limited number of sensors
- Limited data

➔ Severely ill-posed problem
➔ CPU intensive while need for fast solutions

Standard approach (deterministic framework) — ROM-based state estimation

Derive a reduced-order model: $\boldsymbol{y} \approx \widehat{\boldsymbol{y}} = D_{\mathrm{ROM}}\, \boldsymbol{x}$. $\left\{ D_{\mathrm{ROM}}^{(i)} \right\}_{i=1}^{n_D}$ typically are PCA modes. $\boldsymbol{x}$ is the associated coefficients vector.

From a correlation kernel or a "learning" (unsorted) sequence $Y := \left(\boldsymbol{y}^{(1)} \dots \boldsymbol{y}^{(n_{\mathrm{snap}})} \right) \in \mathbb{R}^{n \times n_{\mathrm{snap}}}$:

$$D_{\mathrm{ROM}}\, \Sigma\, V^* \overset{\text{thin SVD}}{\approx} Y.$$

For a given number n_D of retained modes, leads to the best approximation in the following sense:

$$\widehat{Y} = D_{\mathrm{ROM}}\, \widehat{X} = \underset{\substack{\widetilde{Y} \in \mathbb{R}^{n \times n_{\mathrm{snap}}} \\ \mathrm{rank}\left[\widetilde{Y}\right] \leq n_D}}{\arg\min} \left\| Y - \widetilde{Y} \right\|_F \qquad \text{with } \widehat{X} = \Sigma\, V^*.$$

- $\boldsymbol{y} \in \mathbb{R}^n$ field of interest,
- $D_{\mathrm{ROM}} \in \mathbb{R}^{n \times n_D}$ the approximation basis [Dictionary],
- $\boldsymbol{x} \in \mathbb{R}^{n_D}$ the basis coefficients as estimated from the n_s sensors,
- $\boldsymbol{s} \in \mathbb{R}^{n_s}$, sensor information.

Standard approach (deterministic framework) — ROM-based state estimation

On site, what is measured is $\boldsymbol{s} = O\,F\,\boldsymbol{y} =: G\,\boldsymbol{y} \in \mathbb{R}^{n_s}$ only.

$F : \mathbb{R}^n \to \mathbb{R}^n$, $\boldsymbol{y} \mapsto \boldsymbol{u}$: model operator.
$O : \mathbb{R}^n \to \mathbb{R}^{n_s}$, $\boldsymbol{u} \mapsto \boldsymbol{s}$: observation operator.
$G : \mathbb{R}^n \to \mathbb{R}^{n_s}$, $\boldsymbol{y} \mapsto \boldsymbol{s}$: forward operator.

Observer such that

$$\widehat{\boldsymbol{x}} \in \underset{\widetilde{\boldsymbol{x}} \in \mathbb{R}^{n_D}}{\arg\min} \left\| \boldsymbol{s} - G\,D_{\mathrm{ROM}}\,\widetilde{\boldsymbol{x}} \right\|_2, \qquad \text{[data misfit]}$$

or simply $\widehat{\boldsymbol{x}} = (G\,D_{\mathrm{ROM}})^+\ \boldsymbol{s}$.

- $\boldsymbol{y} \in \mathbb{R}^n$ field of interest,
- $D_{\mathrm{ROM}} \in \mathbb{R}^{n \times n_D}$ the approximation basis [Dictionary],
- $\boldsymbol{x} \in \mathbb{R}^{n_D}$ the basis coefficients as estimated from the n_s sensors,
- $\boldsymbol{s} \in \mathbb{R}^{n_s}$, sensor information.

Standard approach (deterministic framework) — ROM-based state estimation

On site, what is measured is $\boldsymbol{s} = O\,F\,\boldsymbol{y} =: G\,\boldsymbol{y} \in \mathbb{R}^{n_s}$ only.

$F : \mathbb{R}^n \to \mathbb{R}^n$, $\boldsymbol{y} \mapsto \boldsymbol{u}$: model operator.
$O : \mathbb{R}^n \to \mathbb{R}^{n_s}$, $\boldsymbol{u} \mapsto \boldsymbol{s}$: observation operator.
$G : \mathbb{R}^n \to \mathbb{R}^{n_s}$, $\boldsymbol{y} \mapsto \boldsymbol{s}$: forward operator.

Observer such that

$$\widehat{\boldsymbol{x}} \in \underset{\widetilde{\boldsymbol{x}} \in \mathbb{R}^{n_D}}{\arg\min} \left\| \boldsymbol{s} - G\,D_{\mathrm{ROM}}\,\widetilde{\boldsymbol{x}} \right\|_2, \qquad \text{[data misfit]}$$

or simply $\widehat{\boldsymbol{x}} = (G\,D_{\mathrm{ROM}})^+ \; \boldsymbol{s}$.

The reconstructed field is finally:

$$\widehat{\boldsymbol{y}} = D_{\mathrm{ROM}}\,\widehat{\boldsymbol{x}} = D_{\mathrm{ROM}}\;(G\,D_{\mathrm{ROM}})^+ \; \boldsymbol{s}.$$

$L_{\mathrm{ROM}} := D_{\mathrm{ROM}}\;(C\,D_{\mathrm{ROM}})^+$ is the ROM-based lift-up operator.

- $\boldsymbol{y} \in \mathbb{R}^n$ field of interest,
- $D_{\mathrm{ROM}} \in \mathbb{R}^{n \times n_D}$ the approximation basis [Dictionary],
- $\boldsymbol{x} \in \mathbb{R}^{n_D}$ the basis coefficients as estimated from the n_s sensors,
- $\boldsymbol{s} \in \mathbb{R}^{n_s}$, sensor information.

Standard approach (deterministic framework) — ROM-based state estimation

On site, what is measured is $\boldsymbol{s} = O\,F\,\boldsymbol{y} =: G\,\boldsymbol{y} \in \mathbb{R}^{n_s}$ only.

$F : \mathbb{R}^n \to \mathbb{R}^n$, $\boldsymbol{y} \mapsto \boldsymbol{u}$: model operator.
$O : \mathbb{R}^n \to \mathbb{R}^{n_s}$, $\boldsymbol{u} \mapsto \boldsymbol{s}$: observation operator.
$G : \mathbb{R}^n \to \mathbb{R}^{n_s}$, $\boldsymbol{y} \mapsto \boldsymbol{s}$: forward operator.

Observer such that

$$\widehat{\boldsymbol{x}} \in \underset{\widetilde{\boldsymbol{x}} \in \mathbb{R}^{n_D}}{\arg\min} \left\| \boldsymbol{s} - G\,D_{\text{ROM}}\,\widetilde{\boldsymbol{x}} \right\|_2, \qquad \text{[data misfit]}$$

or simply $\widehat{\boldsymbol{x}} = (G\,D_{\text{ROM}})^+ \; \boldsymbol{s}$. $\longleftarrow$ **requires $n_s \geq n_D$** (if no additional hyp.)

The reconstructed field is finally:

$$\widehat{\boldsymbol{y}} = D_{\text{ROM}}\,\widehat{\boldsymbol{x}} = D_{\text{ROM}}\,(G\,D_{\text{ROM}})^+ \; \boldsymbol{s}.$$

$L_{\text{ROM}} := D_{\text{ROM}}\,(C\,D_{\text{ROM}})^+$ is the ROM-based lift-up operator.

- $\boldsymbol{y} \in \mathbb{R}^n$ field of interest,
- $D_{\text{ROM}} \in \mathbb{R}^{n \times n_D}$ the approximation basis [Dictionary],
- $\boldsymbol{x} \in \mathbb{R}^{n_D}$ the basis coefficients as estimated from the n_s sensors,
- $\boldsymbol{s} \in \mathbb{R}^{n_s}$, sensor information.

Inversion as statistical inference (SI)

Deterministic inversion often ill-posed (existence, uniqueness, stability)

Inversion as statistical inference (SI)

Deterministic inversion often ill-posed (existence, uniqueness, stability)

$\longrightarrow$ sparse basis learning approach

Inversion as statistical inference (SI)

Deterministic inversion often ill-posed (existence, uniqueness, stability)

$\longrightarrow$ sparse basis learning approach

$\longrightarrow$ Statistical inference restates IP as a well-posed extension in a larger space of probability distributions

A dictionary learning algorithm

Without additional hypotheses, impossible to estimate $n_D > n_s$ modes

$\longrightarrow$ derive an over-complete dictionary D for sparse representation of the field $\boldsymbol{y} \in \mathcal{M}_{\boldsymbol{y}}$ to be infered

$$\boldsymbol{y} \underset{\epsilon}{\approx} D\,\boldsymbol{x}, \qquad \boldsymbol{x} \in \mathbb{R}^{n_D}$$

A dictionary learning algorithm

Without additional hypotheses, impossible to estimate $n_D > n_s$ modes

$\longrightarrow$ derive an over-complete dictionary D for sparse representation of the field $\boldsymbol{y} \in \mathcal{M}_{\boldsymbol{y}}$ to be infered

$$\boldsymbol{y} \underset{\epsilon}{\approx} D\,\boldsymbol{x}, \qquad \boldsymbol{x} \in \mathbb{R}^{n_D}$$

$\longrightarrow$ n_s-sparse approximation

$$\boldsymbol{y} \underset{\epsilon}{\approx} D\,\boldsymbol{x} \quad \text{with} \quad \left\|\boldsymbol{x}^{(i)}\right\|_0 \leq n_s, \quad \forall \boldsymbol{y} \in \mathcal{M}_{\boldsymbol{y}}$$

A dictionary learning algorithm

Without additional hypotheses, impossible to estimate $n_D > n_s$ modes

$\longrightarrow$ derive an over-complete dictionary D for sparse representation of the field $\boldsymbol{y} \in \mathcal{M}_{\boldsymbol{y}}$ to be infered

$$\boldsymbol{y} \underset{\epsilon}{\approx} D\,\boldsymbol{x}, \qquad \boldsymbol{x} \in \mathbb{R}^{n_D}$$

$\longrightarrow$ n_s-sparse approximation

$$\boldsymbol{y} \underset{\epsilon}{\approx} D\,\boldsymbol{x} \quad \text{with} \quad \left\|\boldsymbol{x}^{(i)}\right\|_0 \le n_s, \quad \forall \boldsymbol{y} \in \mathcal{M}_{\boldsymbol{y}}$$

$$\text{Find } \{D, X\} \in \underset{\widetilde{D}, \widetilde{X}}{\arg\min} \left\| Y - \widetilde{D}\,\widetilde{X} \right\|_F \quad \text{s.t.} \quad \left\|\widetilde{\boldsymbol{x}}^{(i)}\right\|_0 \le n_s, \quad \forall i, \quad X := \left(\boldsymbol{x}^{(1)} \ldots \boldsymbol{x}^{(n_{\text{snap}})}\right).$$

A dictionary learning algorithm

$\longrightarrow$ use K-SVD algorithm

Repeat

- **Sparse Coding** : $X \in \arg\min_{\widetilde{X}} \left\| Y - D\,\widetilde{X} \right\|_F \quad \text{s.t.} \quad \left\| \widetilde{\boldsymbol{x}}^{(i)} \right\|_0 \leq n_s, \forall i.$
- **CodeBook Update** : Update D and X in order to lower $\|Y - D\,X\|_F$ while maintaining the support of $\left\{ \boldsymbol{x}^{(i)} \right\}_i$.

A dictionary learning algorithm

$\longrightarrow$ use K-SVD algorithm

Repeat

- **Sparse Coding** : $X \in \arg\min_{\widetilde{X}} \left\| Y - D\,\widetilde{X} \right\|_F$ s.t. $\left\| \widetilde{\boldsymbol{x}}^{(i)} \right\|_0 \leq n_s, \forall i.$
- **CodeBook Update** : Update D and X in order to lower $\| Y - D\,X \|_F$ while maintaining the support of $\left\{ \boldsymbol{x}^{(i)} \right\}_i$.

But typically $\boldsymbol{x}^{(i)} \in \mathbb{R}^{n_D}$ **cannot** be estimated from measurements: Observability issue

$\longrightarrow$ determine D given the SI framework for estimating $\boldsymbol{x}^{(i)}$ from $\boldsymbol{s}^{(i)}$ instead of $\boldsymbol{y}^{(i)}$.
$\longrightarrow$ dictionary D both accurate and observable: $D\,(Y, S)$.

$\longrightarrow$ Observability-oriented Dictionary Learning

On-site data: $\boldsymbol{s} = G(\boldsymbol{y})$.

Let $\mathcal{M}_{\boldsymbol{y}} = \left\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_{\mathrm{snap}}}\right\}$ examples of plausible fields. One looks for a recovery procedure minimizing the *Bayes risk*

$$\begin{aligned}\mathbb{E}_{\boldsymbol{s},\boldsymbol{y}}\left[\|\boldsymbol{y} - \widehat{\boldsymbol{y}}(\boldsymbol{s})\|_2^2\right] &= \iint \|\boldsymbol{y} - \widehat{\boldsymbol{y}}(\boldsymbol{s})\|_2^2 \, p(\boldsymbol{s}|\boldsymbol{y})\, p(\boldsymbol{y})\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{y}, \\ &\approx \propto \left\|Y - \widehat{Y}(S)\right\|_F^2. \qquad [\textit{iid} \text{ samples}].\end{aligned}$$

with

$$\begin{aligned}\boldsymbol{y} &\approx \widehat{\boldsymbol{y}} = D\,\boldsymbol{x}, \\ \boldsymbol{s} &\approx \widehat{\boldsymbol{s}} = P(\boldsymbol{x}) \qquad [P\text{: Sparsifying measurement predictor}].\end{aligned}$$

$$\longrightarrow \quad \{D, X, P\} \in \operatorname*{arg\,min}_{\widetilde{D},\widetilde{X},\widetilde{P}} \left\|Y - \widetilde{D}\,\widetilde{X}\left(S; \widetilde{P}\right)\right\|_F.$$

$\longrightarrow$ Observability-oriented Dictionary Learning

On-site data: $\boldsymbol{s} = G(\boldsymbol{y})$.

Let $\mathcal{M}_{\boldsymbol{y}} = \left\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_{\text{snap}}}\right\}$ examples of plausible fields. One looks for a recovery procedure minimizing the *Bayes risk*

$$\begin{aligned} \mathbb{E}_{\boldsymbol{s},\boldsymbol{y}}\left[\left\|\boldsymbol{y} - \widehat{\boldsymbol{y}}(\boldsymbol{s})\right\|_2^2\right] &= \iint \left\|\boldsymbol{y} - \widehat{\boldsymbol{y}}(\boldsymbol{s})\right\|_2^2 p(\boldsymbol{s}|\boldsymbol{y})\, p(\boldsymbol{y})\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{y}, \\ &\approx \propto \left\|Y - \widehat{Y}(S)\right\|_F^2. \qquad [\textit{iid}\ \text{samples}]. \end{aligned}$$

with

$$\begin{aligned} \boldsymbol{y} &\approx \widehat{\boldsymbol{y}} = D\,\boldsymbol{x}, \\ \boldsymbol{s} &\approx \widehat{\boldsymbol{s}} = P(\boldsymbol{x}) \qquad [P\text{: Sparsifying measurement predictor}]. \end{aligned}$$

$$\longrightarrow \quad \{D, X, P\} \in \underset{\widetilde{D},\widetilde{X},\widetilde{P}}{\arg\min} \left\|Y - \widetilde{D}\,\widetilde{X}\left(S; \widetilde{P}\right)\right\|_F.$$

Alternatively:

$$\begin{aligned} \boldsymbol{y} &\approx \widehat{\boldsymbol{y}} = L(\boldsymbol{x}) \qquad [L\text{: Nonlinear sparsifying lift-up operator}], \\ \boldsymbol{s} &\approx \widehat{\boldsymbol{s}} = D_{\text{feat}}\,\boldsymbol{x}. \end{aligned}$$

$$\longrightarrow \quad \{D_{\text{feat}}, X, L\} \in \underset{\widetilde{D}_{\text{feat}},\widetilde{X},\widetilde{L}}{\arg\min} \left\|Y - \widetilde{L}\left(\widetilde{X}\left(S; \widetilde{D}_{\text{feat}}\right)\right)\right\|_F.$$

$\longrightarrow$ Observability-oriented Dictionary Learning

$$\left\{\boldsymbol{x}^{(i)}, D_{\text{feat}}\right\} \in \underset{\widetilde{\boldsymbol{x}}, \widetilde{D}_{\text{feat}}}{\arg\min} \left\|\boldsymbol{s}^{(i)} - \widetilde{D}_{\text{feat}}\, \widetilde{\boldsymbol{x}}\right\|_2^2 \qquad \text{s.t.} \qquad \left\|\widetilde{\boldsymbol{x}}\right\|_0 \leq n_s \quad \forall i, \quad \left\|\widetilde{\boldsymbol{d}}_{\text{feat},l}\right\|_2 = 1 \quad \forall l,$$

$$L \in \underset{\widetilde{L}}{\arg\min} \left\| Y - \widetilde{L}\left(\widetilde{X}\right)\right\|_F .$$

$\longrightarrow$ Observability-oriented Dictionary Learning

$$\left\{ \boldsymbol{x}^{(i)}, D_{\mathrm{feat}} \right\} \in \underset{\widetilde{\boldsymbol{x}}, \widetilde{D}_{\mathrm{feat}}}{\arg\min} \left\| \boldsymbol{s}^{(i)} - \widetilde{D}_{\mathrm{feat}} \, \widetilde{\boldsymbol{x}} \right\|_2^2 \qquad \text{s.t.} \qquad \left\| \widetilde{\boldsymbol{x}} \right\|_0 \leq n_s \quad \forall i, \qquad \left\| \widetilde{\boldsymbol{d}}_{\mathrm{feat},l} \right\|_2 = 1 \quad \forall l,$$

$$L \in \underset{\widetilde{L}}{\arg\min} \left\| Y - \widetilde{L} \left(\widetilde{X} \right) \right\|_F .$$

Let $\Psi^{(i)} := \left\| \boldsymbol{s}^{(i)} - \widetilde{D}_{\mathrm{feat}} \, \widetilde{\boldsymbol{x}}^{(i)} \right\|_2^2 + \mathrm{regul} \left(\widetilde{\boldsymbol{x}}^{(i)} \right)$

$\longrightarrow$ Observability-oriented Dictionary Learning

$$\left\{\boldsymbol{x}^{(i)}, D_{\text{feat}}\right\} \in \underset{\widetilde{\boldsymbol{x}}, \widetilde{D}_{\text{feat}}}{\arg\min} \left\|\boldsymbol{s}^{(i)} - \widetilde{D}_{\text{feat}}\, \widetilde{\boldsymbol{x}}\right\|_2^2 \qquad \text{s.t.} \qquad \left\|\widetilde{\boldsymbol{x}}\right\|_0 \le n_s \quad \forall i, \quad \left\|\widetilde{\boldsymbol{d}}_{\text{feat},l}\right\|_2 = 1 \quad \forall l,$$

$$L \in \underset{\widetilde{L}}{\arg\min} \left\| Y - \widetilde{L}\left(\widetilde{X}\right)\right\|_F .$$

Let $\Psi^{(i)} := \left\|\boldsymbol{s}^{(i)} - \widetilde{D}_{\text{feat}}\, \widetilde{\boldsymbol{x}}^{(i)}\right\|_2^2 + \text{regul}\left(\widetilde{\boldsymbol{x}}^{(i)}\right)$, it leads to

$\longrightarrow$ Double-optimization problem

$$\{D_{\text{feat}}, X, L\} \in \underset{\widetilde{D}\text{feat}, \widetilde{X}, \widetilde{L}}{\arg\min} \left\| Y - \widetilde{L}\left(\widetilde{X}\right)\right\|_F ,$$

$$\text{s.t.} \qquad \boldsymbol{c}^{(i)} := \frac{\partial \Psi^{(i)}}{\partial \widetilde{\boldsymbol{x}}^{(i)}} = \boldsymbol{0}, \qquad \forall\, 1 \le i \le n_{\text{snap}},$$

$$\frac{\partial \Psi}{\partial \widetilde{\boldsymbol{d}}_{\text{feat},l}} = \boldsymbol{0}, \qquad \text{s.t.} \qquad \left\|\widetilde{\boldsymbol{d}}_{\text{feat},l}\right\|_2 = 1, \qquad \forall\, 1 \le l \le n_D.$$

$\longrightarrow$ consistent and as realistic as possible

$\longrightarrow$ Observability-oriented Dictionary Learning

$$\left\{\boldsymbol{x}^{(i)}, D_{\text{feat}}\right\} \in \underset{\widetilde{\boldsymbol{x}}, \widetilde{D}_{\text{feat}}}{\arg\min} \left\|\boldsymbol{s}^{(i)} - \widetilde{D}_{\text{feat}}\,\widetilde{\boldsymbol{x}}\right\|_2^2 \quad \text{s.t.} \quad \left\|\widetilde{\boldsymbol{x}}\right\|_0 \leq n_s \quad \forall i, \quad \left\|\widetilde{\boldsymbol{d}}_{\text{feat},l}\right\|_2 = 1 \quad \forall l,$$

$$L \in \underset{\widetilde{L}}{\arg\min} \left\|Y - \widetilde{L}\left(\widetilde{X}\right)\right\|_F.$$

Let $\Psi^{(i)} := \left\|\boldsymbol{s}^{(i)} - \widetilde{D}_{\text{feat}}\,\widetilde{\boldsymbol{x}}^{(i)}\right\|_2^2 + \text{regul}\left(\widetilde{\boldsymbol{x}}^{(i)}\right)$, it leads to

$\longrightarrow$ Double-optimization problem

$$\{D_{\text{feat}}, X, L\} \in \underset{\widetilde{D}\text{feat}, \widetilde{X}, \widetilde{L}}{\arg\min} \left\|Y - \widetilde{L}\left(\widetilde{X}\right)\right\|_F,$$

$$\text{s.t.} \quad \boldsymbol{c}^{(i)} := \frac{\partial \Psi^{(i)}}{\partial \widetilde{\boldsymbol{x}}^{(i)}} = \mathbf{0}, \qquad \forall\, 1 \leq i \leq n_{\text{snap}},$$

$$\frac{\partial \Psi}{\partial \widetilde{\boldsymbol{d}}_{\text{feat},l}} = \mathbf{0}, \qquad \text{s.t.} \quad \left\|\widetilde{\boldsymbol{d}}_{\text{feat},l}\right\|_2 = 1, \qquad \forall\, 1 \leq l \leq n_D.$$

$\longrightarrow$ consistent and as realistic as possible

- constraints $\boldsymbol{c}^{(i)}$ are independent $\longrightarrow$ *parallel solving*,
- *e.g.*, use *l − BFGS*.

$\longrightarrow$ Observability-oriented Dictionary Learning – Linear framework

On-site data: $\boldsymbol{s} = G\,\boldsymbol{y}$. Linear relationship

$L \equiv D \in \mathbb{R}^{n \times n_D} \in \mathrm{span}\left\{\boldsymbol{y}^{(i)}\right\}$. Let $Y \overset{\mathrm{QR}}{=} Q\,R$, $D = Y\,B$.

$\longrightarrow$ Observability-oriented Dictionary Learning – Linear framework

On-site data: $\boldsymbol{s} = G\,\boldsymbol{y}$. Linear relationship

$L \equiv D \in \mathbb{R}^{n \times n_D} \in \operatorname{span}\left\{\boldsymbol{y}^{(i)}\right\}$. Let $Y \stackrel{\mathrm{QR}}{=} Q\,R$, $D = Y\,B$.

Using block-coordinate descent, alternate solve for

- **Observability-oriented Sparse Coding**

$$\boldsymbol{x}^{(i)} \in \underset{\widetilde{\boldsymbol{x}} \in \mathbb{R}^{n_D}}{\arg\min} \left\| \boldsymbol{s}^{(i)} - D_{\mathrm{feat}}\,\widetilde{\boldsymbol{x}} \right\|_2, \quad \text{s.t.} \quad \|\boldsymbol{x}\|_0 \leq n_s,$$

- **Estimation CodeBook Update**

$$D \in \underset{\widetilde{D} \in \mathbb{R}^{n \times n_D}}{\arg\min} \left\| Y - \widetilde{D}\,X \right\|_F.$$

$\longrightarrow$ solved directly: $D \leftarrow Y\,X^+$ [using `qrginv` (Katsikis *et al.*, 2011)].

$\longrightarrow$ Recovery error: $\left\| R\,\left(I_{n_{\mathrm{snap}}} - B\,X\right)\right\|_F$ [inexpensive to evaluate].

- **Feature CodeBook Update**

$$\boldsymbol{d}_{\mathrm{feat},l} \propto \left(S - D_{\mathrm{feat},\setminus l}\,X_{\setminus l}\right)\,\widehat{\boldsymbol{x}}_l^{\mathsf{T}}, \qquad \left\|\boldsymbol{d}_{\mathrm{feat},l}\right\|_2 = 1, \quad \forall\, 1 \leq l \leq n_D,$$

with $\quad \widehat{R\,\boldsymbol{b}_l}\,\widehat{\boldsymbol{x}}_l \stackrel{\text{rank-1}}{\approx} R\,\left(I_{n_{\mathrm{snap}}} - B_{\setminus l}\,X_{\setminus l}\right).$

$\longrightarrow$ consistent and as realistic as possible

Bayesian Compressive Sensing

Relax $\|\boldsymbol{x}\|_0$ (NP-hard optimization problem) in $\|\cdot\|_1$.

$$\boldsymbol{x} \in \underset{\widetilde{\boldsymbol{x}} \in \mathbb{R}^{n_D}}{\arg\min} \left\| \boldsymbol{s} - D_{\text{feat}}\, \widetilde{\boldsymbol{x}} \right\|_2^2 + \tau \left\| \widetilde{\boldsymbol{x}} \right\|_1 ,$$

Bayesian Compressive Sensing

Relax $\|\boldsymbol{x}\|_0$ (NP-hard optimization problem) in $\|\cdot\|_1$.

$$\boldsymbol{x} \in \underset{\widetilde{\boldsymbol{x}} \in \mathbb{R}^{n_D}}{\arg\min} \left\|\boldsymbol{s} - D_{\text{feat}}\, \widetilde{\boldsymbol{x}}\right\|_2^2 + \tau \left\|\widetilde{\boldsymbol{x}}\right\|_1 ,$$

Full information on the solution $\longrightarrow$ Bayesian modeling
Independent data during training.

$$\boldsymbol{x} \in \underset{\widetilde{\boldsymbol{x}} \in \mathbb{R}^{n_D}}{\arg\max}\ q\left(\widetilde{\boldsymbol{x}} \middle| \boldsymbol{s}\right) \propto L\left(\widetilde{\boldsymbol{x}} \middle| \boldsymbol{s}\right)\, p\left(\widetilde{\boldsymbol{x}}\right), \qquad \widehat{\boldsymbol{s}} = D_{\text{feat}}\, \boldsymbol{x},$$

with

$$L\left(\boldsymbol{x} | \boldsymbol{s}, \beta\right) \sim \mathcal{N}\left(\boldsymbol{s} \middle| D_{\text{feat}}\, \boldsymbol{x}, \beta^{-1}\right),$$

$$p\left(\boldsymbol{x}\right) = \frac{\lambda}{2} \exp\left(-\frac{\lambda}{2} \left\|\boldsymbol{x}\right\|_1\right), \qquad \text{[Laplace prior]},$$

$$\tau = \lambda / \beta \qquad \text{[sparsity penalty]}.$$

Bayesian Compressive Sensing (cnt'd)

$$\left.\begin{array}{l}\text{Laplace prior not conjugate to the likelihood model}\\ \text{Sparsity regularization penalty } \tau \text{ unknown}\end{array}\right\} \longrightarrow \text{hierarchical formulation}$$

$$p(\boldsymbol{x}|\gamma) = \prod_{i=1}^{n_D} \mathcal{N}(x_i|0,\gamma_i), \qquad \boldsymbol{\gamma} = (\gamma_1 \ldots \gamma_{n_D}), \tag{1}$$

$$p(\gamma_i|\lambda) = \frac{\lambda}{2} \exp\left(-\frac{\lambda \gamma_i}{2}\right), \tag{2}$$

$$p(\lambda|\nu) = \Gamma(\lambda|\nu/2, \nu/2).$$

Each of the n_D independent hyperparameters γ_i controls the strength of the prior $\longrightarrow$ introduces sparsity in the model.

Sequential Sparse Bayesian Learning

$\longrightarrow$ three-stage hierarchy. Eqs. (1)-(2) result in a Laplace distribution $p(\boldsymbol{x}|\lambda)$,

$\longrightarrow$ $q(\boldsymbol{x}|\boldsymbol{s},\gamma,\beta,\lambda)$ is a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\Sigma)$,

$$\boldsymbol{\mu} = \beta\,\Sigma\,D_{\text{feat}}^{\mathsf{T}}\,\boldsymbol{s}, \qquad \Sigma = \left(\beta\,D_{\text{feat}}^{\mathsf{T}}\,D_{\text{feat}} + \Lambda\right)^{-1}, \quad \Lambda = \operatorname{diag}\left(\gamma_1^{-1},\ldots,\gamma_{n_D}^{-1}\right).$$

Follows Babacan *et al.* (2010).

Hyperparameters maximize the (log-) marginal likelihood

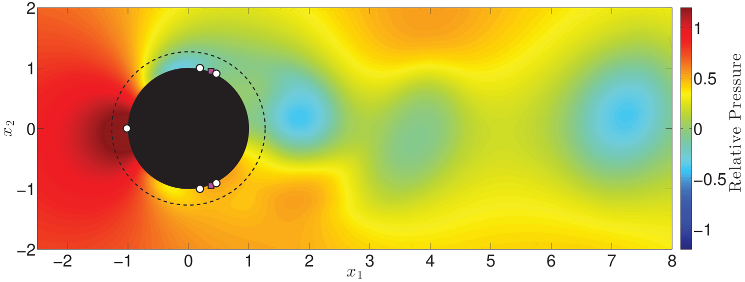
$$\log\,p(\boldsymbol{s},\gamma,\beta,\lambda) = \log\int p(\boldsymbol{s}|\boldsymbol{x},\beta)\;p(\boldsymbol{x}|\gamma)\;p(\gamma|\lambda)\;p(\lambda)\;p(\beta)\;\mathrm{d}\boldsymbol{x}$$

$\longrightarrow$ MAP estimate: $\widehat{\boldsymbol{y}} = D\,\boldsymbol{\mu}$.

In a nutshell

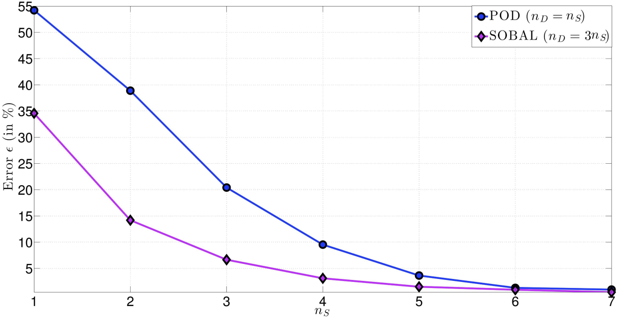
1. Basis learning **[offline]**
   - Form a snapshot matrix Y of expected realizations of the field and corresponding matrix S,
   - Given the sensors, learn representation dictionaries D (and D_{feat}) using Sparse Bayesian Dictionary Learning,
2. Field reconstruction **[online]**
   - Use SBL with the measure $\boldsymbol{s}$ to estimate $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma)$.
   - Reconstruct the MAP total field from $\widehat{\boldsymbol{y}} = D\,\boldsymbol{\mu}$ or use Markov chain Monte Carlo.

Numerical experiment: pressure field



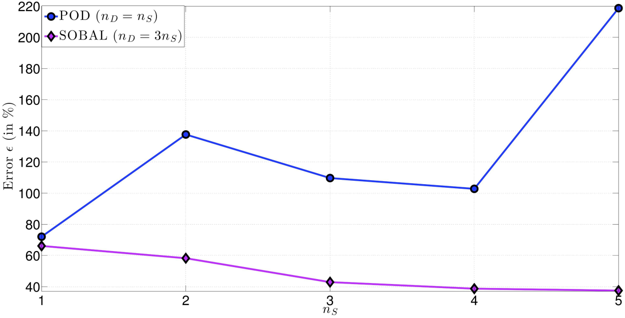
- ***y***: pressure field
- ***s***: measurements at the surface ($n_s \leq 5$)

Observability issue. Full information: $\boldsymbol{s} \equiv \boldsymbol{y}$



$\longrightarrow$ significantly better L^2-reconstruction performance than PCA/KL.

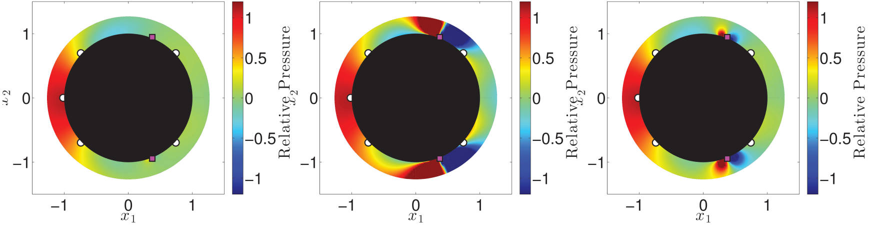
Observability issue. Sensor information only



$\longrightarrow$ significantly better L^2-reconstruction performance than PCA/KL.

$\longrightarrow$ PCA coefficients are not accurately estimated from the sensors.

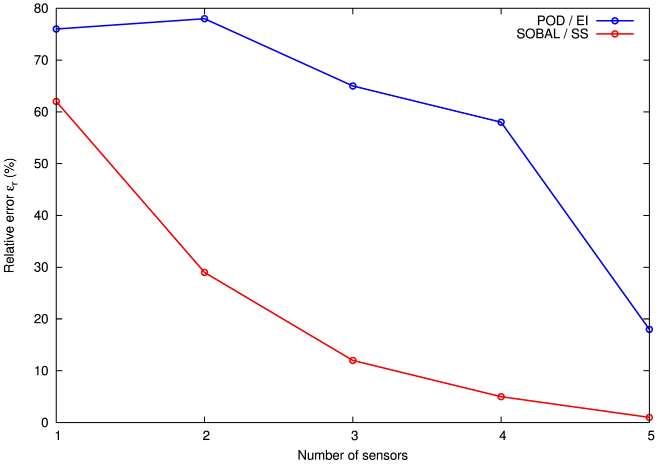
Reconstruction



Relative pressure field. Exact (left), estimated from PCA/KL (middle), estimated from present approach (right).

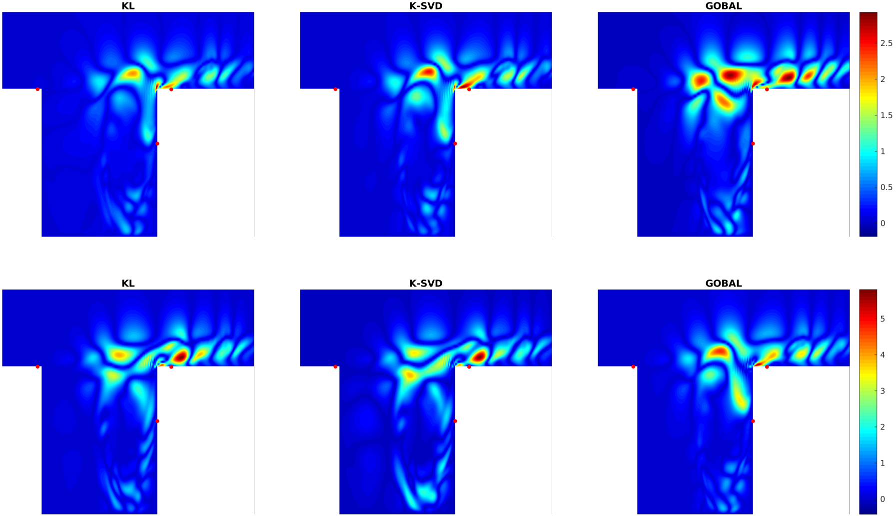
⟶ Much better reconstruction performance.

Combined with SensorSpace...

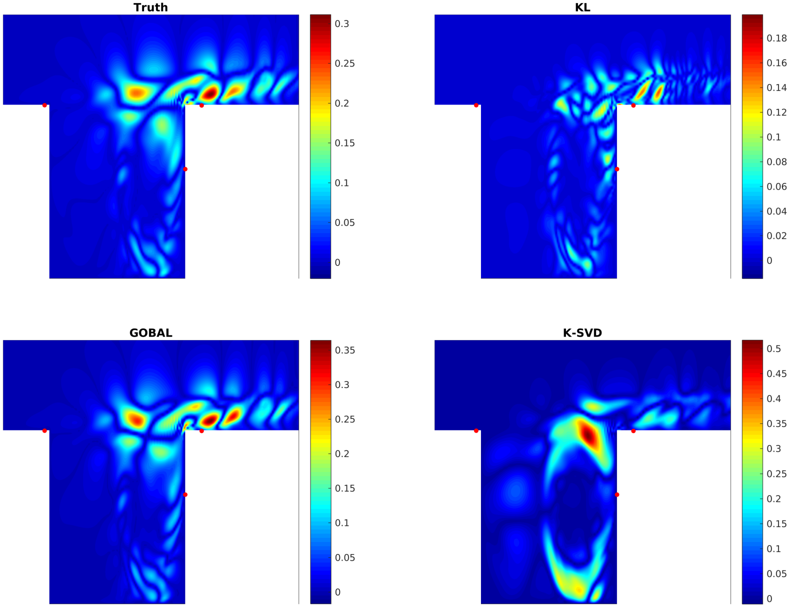


⟶ PCA/Effective Independence compared to SOBAL / SensorSpace.

Cavity flow – Dictionary



Recovery performance – 3 sensors



Closing remarks

- Offline/Online strategy for inference. First learn about the system at hand, then exploit,
- Basis learning philosophy is a key for a *realistic* and successful approach,
- Need to balance between representation accuracy and observabity when using a dictionary $\longrightarrow$ *observability-oriented sparse Bayesian learning*.

Closing remarks

- Offline/Online strategy for inference. First learn about the system at hand, then exploit,
- Basis learning philosophy is a key for a *realistic* and successful approach,
- Need to balance between representation accuracy and observabity when using a dictionary $\longrightarrow$ *observability-oriented sparse Bayesian learning*.

On-going efforts:

- High-dimensional fields, nonlinear observation operator,
- Combined with an assimilation technique, *e.g.*, sparsity exploiting EnKF.