Data-Driven estimation of a turbulent flow from wall sensors

Lionel MATHELIN

LIMSI–CNRS, Orsay, France
Dpt. Applied Math., Univ. Washington, USA

Joint work with
S. MURALIDHAR and B. PODVIN

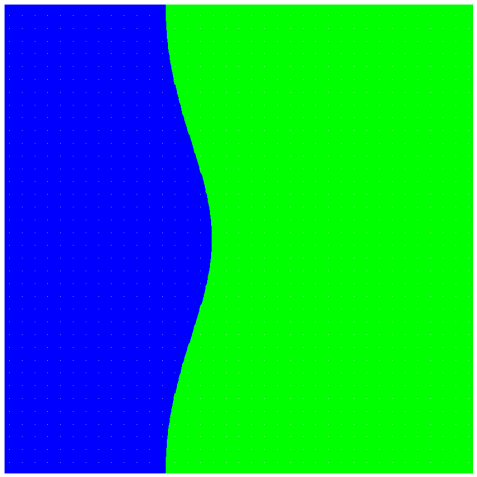
US-Japan workshop – March. 2018

Motivation: Inverse problems (IP)

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

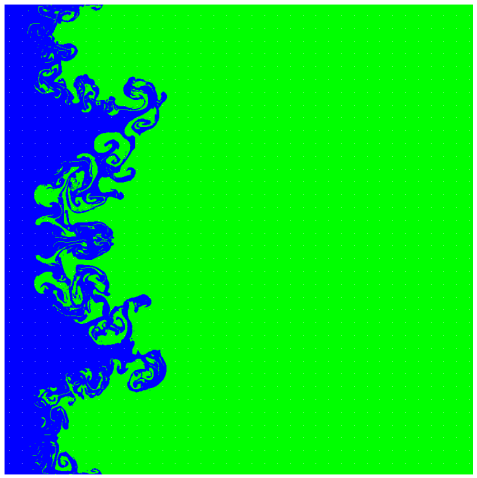
$\longrightarrow$ estimate parameters $\boldsymbol{y}$

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



space sometimes HD (spatial or temporal fields)

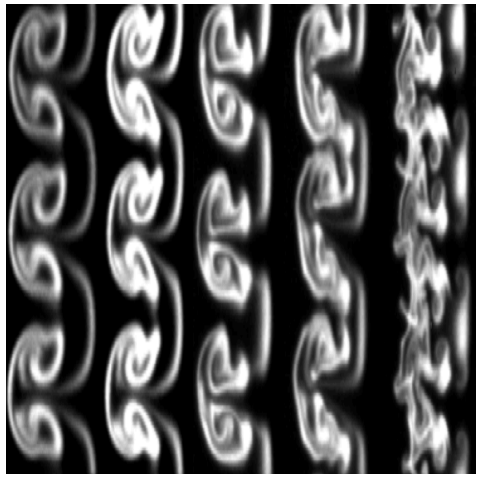
state: $\boldsymbol{u} \approx F\boldsymbol{y}$



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

state space is HD

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



O: observation operator

data limited, noisy & indirect

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

Derive a reduced-order model: $\boldsymbol{y} \approx \widehat{\boldsymbol{y}} = D\,\boldsymbol{x}$. D typically are PCA modes.

From a correlation kernel or a "training" (unsorted) sequence $Y := \left(\boldsymbol{y}^{(1)} \dots \boldsymbol{y}^{(n_{\text{snap}})}\right)$:

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

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

$$\widehat{Y} = D\,\widehat{X} \in \underset{\text{rank}\left[\widetilde{Y}\right] \le n_D}{\arg\min} \left\| Y - \widetilde{Y} \right\|_F \qquad \text{with} \quad \widehat{X} = \Sigma\,V^*.$$

- $\boldsymbol{y} \in \mathbb{R}^n$ field of interest,
- $D \in \mathbb{R}^{n \times n_D}$ 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 $\boldsymbol{s} = G\,\boldsymbol{y}$ only.
$G : \boldsymbol{y} \mapsto \boldsymbol{s}$: forward operator.

Observer such that

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

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

The reconstructed field is finally:

$$\widehat{\boldsymbol{y}} = D\,\widehat{\boldsymbol{x}} = D\ (G\,D)^+\ \boldsymbol{s}.$$

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

On site, what is measured is $\boldsymbol{s} = G\,\boldsymbol{y}$ only.
$G : \boldsymbol{y} \mapsto \boldsymbol{s}$: forward operator.

Observer such that

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

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

The reconstructed field is finally:

$$\widehat{\boldsymbol{y}} = D\,\widehat{\boldsymbol{x}} = D\ (G\,D)^+\ \boldsymbol{s}.$$

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

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

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

A dictionary learning algorithm

$\longrightarrow$ use K-SVD algorithm

Repeat

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

A dictionary learning algorithm

$\longrightarrow$ use K-SVD algorithm

Repeat

- **Sparse Coding** : $X \in \underset{\widetilde{X}}{\arg\min} \left\| Y - D\widetilde{X} \right\|_F \quad \text{s.t.} \quad \left\| \widetilde{\boldsymbol{x}}^{(i)} \right\|_0 \leq n_s, \forall i.$
- **CodeBook Update** : Update D and X in order to lower $\|Y - DX\|_F$ 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 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 $\mathcal{M}_{\boldsymbol{y}} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_{\text{snap}}}\}$ 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

$$\begin{aligned} \boldsymbol{y} &\approx \widehat{\boldsymbol{y}} = D(\boldsymbol{x}), \\ \boldsymbol{s} &\approx \widehat{\boldsymbol{s}} = D_{\boldsymbol{s}}\, \boldsymbol{x}. \end{aligned}$$

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

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

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

Using block-coordinate descent, alternate solve for

- **Observability-oriented Sparse Coding**

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

- **Estimation CodeBook Update**

$$D \in \underset{\widetilde{D}}{\arg\min} \left\| Y - \widetilde{D}\, X \right\|_F.$$

- **Feature CodeBook Update**

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

$\longrightarrow$ consistent and as realistic as possible

▸ Skip Sparse Bayesian Learning

Bayesian Compressive Sensing

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

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

Bayesian Compressive Sensing

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

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

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

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

with

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

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

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

Bayesian Compressive Sensing (cnt'd)

Sparsity regularization penalty τ unknown
Laplace prior not conjugate to the likelihood model
$\Big\}$ $\longrightarrow$ hierarchical formulation

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

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

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

Bayesian Compressive Sensing (cnt'd)

Sparsity regularization penalty τ unknown
Laplace prior not conjugate to the likelihood model
$\Big\}$ $\longrightarrow$ hierarchical formulation

$$
\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{\gamma}) &= \prod_{i=1}^{n_D} \mathcal{N}(x_i|0,\gamma_i), \qquad \boldsymbol{\gamma} = \left(\gamma_1 \dots \gamma_{n_D}\right), && (1)\\
p(\gamma_i|\lambda) &= \frac{\lambda}{2} \exp\left(-\frac{\lambda\,\gamma_i}{2}\right), && (2)\\
p(\lambda|\nu) &= \Gamma(\lambda|\nu/2, \nu/2).
\end{aligned}
$$

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

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

Sequential Sparse Bayesian Learning

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

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

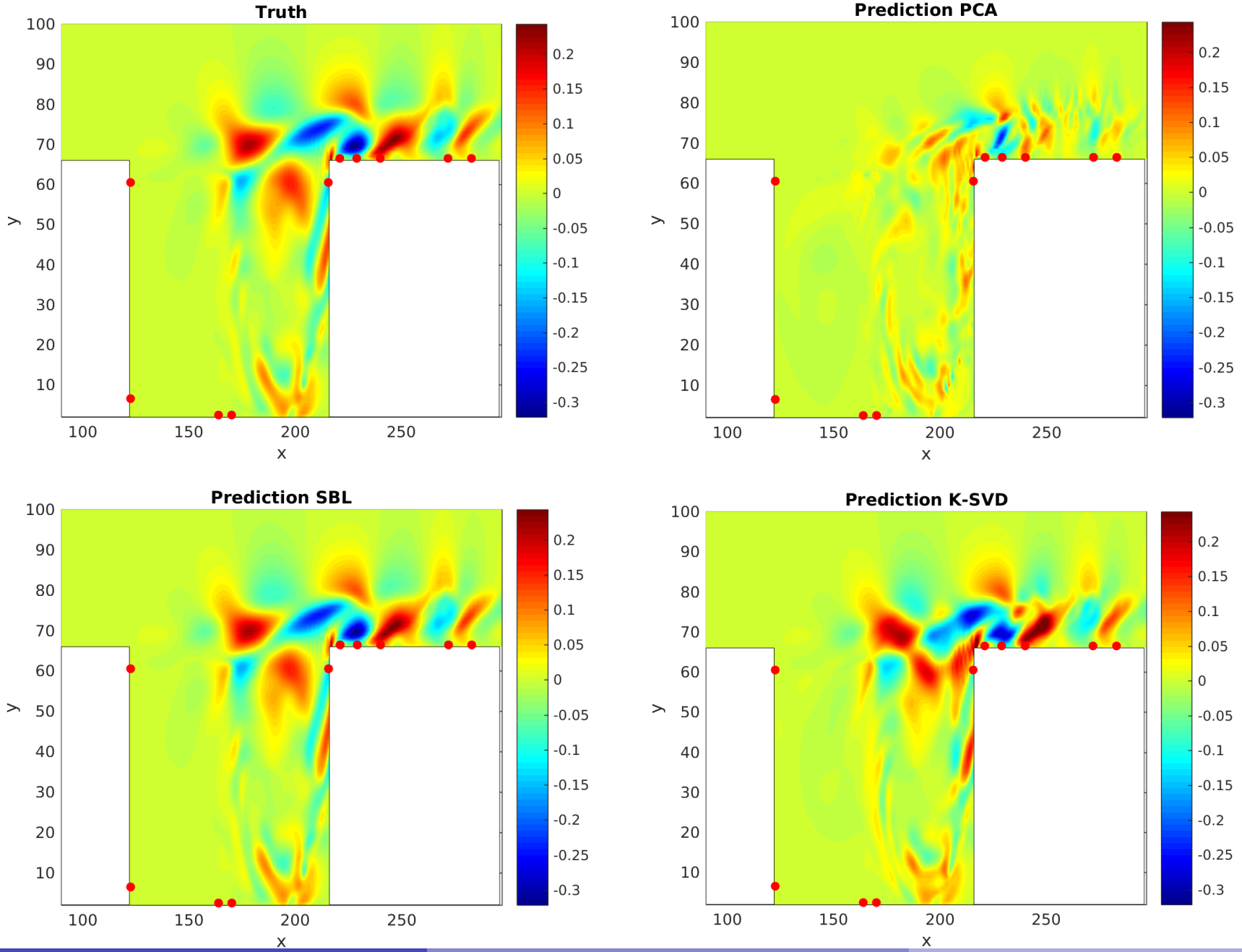
Follows Babacan *et al.* (2010).

Hyperparameters maximize the (log-) marginal likelihood

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

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

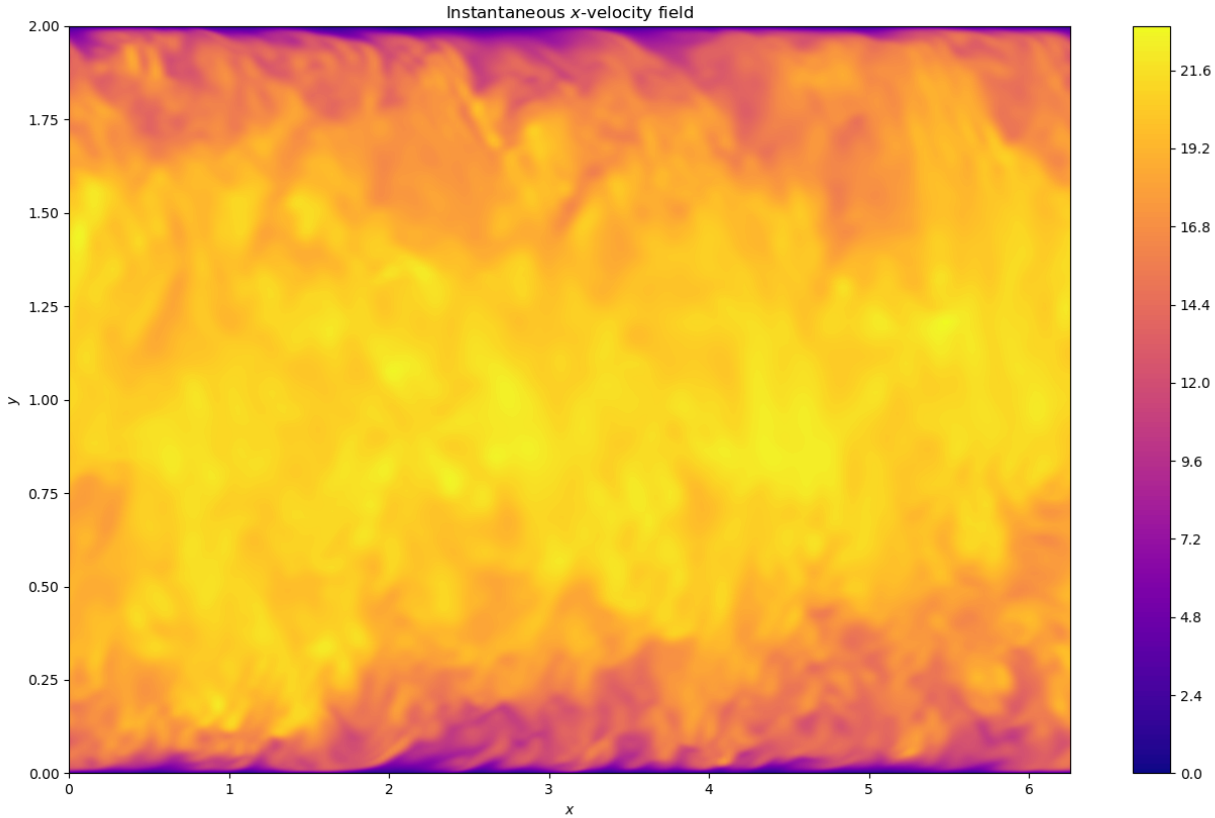
Recovery performance



Some remarks

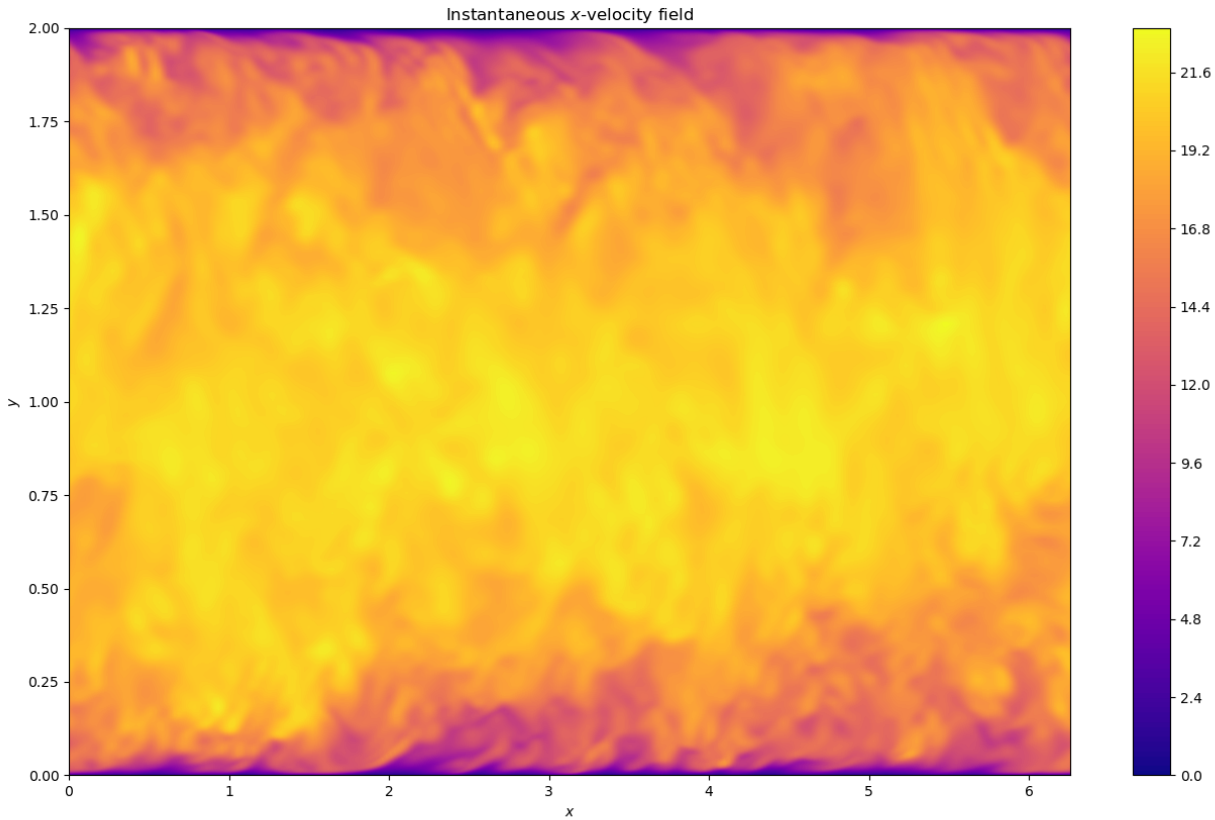
- Offline/Online strategy for inference. First learn about the system at hand, then exploit,
- Basis learning philosophy is a key for a realistic and successful approach,
- Need to balance between representation accuracy and observability when using a dictionary $\longrightarrow$ observability-oriented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

A more challenging situation: estimating a 3-D turbulent channel flow



Estimation of the velocity field: $\boldsymbol{y} \in \mathbb{R}^{2\,10^7}$
30 pressure sensors at the lower wall: $\boldsymbol{s}_{\text{inst}}$.
Dataset: 7,000 snapshots: Y
No model available. Catalog of data from past experiments and/or simulations.

A more challenging situation: estimating a 3-D turbulent channel flow



Estimation of the velocity field: $\boldsymbol{y} \in \mathbb{R}^{2\,10^7}$
30 pressure sensors at the lower wall: $\boldsymbol{s}_{\text{inst}}$.
Dataset: 7,000 snapshots: Y
No model available. Catalog of data from past experiments and/or simulations.

Disclaimer: I will not solve the problem. . .

Preliminary observations

- ☹ The spectrum of the database is very slowly decaying $\longrightarrow$ cannot exploit low-rank structure.
- ☹ To fill a d-D hypercube with ϵ-balls $\mathcal{B}_\epsilon\left(L^2\right)$, one needs

$$n_{\text{snap}} \sim \epsilon^{-d}\, d^{d/2} \quad \text{samples}$$

$\longrightarrow$ NN-based approach is doomed to fail.

- ☹ Not so much data $\longrightarrow$ precludes deep-$\star$ approaches.

Preliminary observations

- ☹ The spectrum of the database is very slowly decaying $\longrightarrow$ cannot exploit low-rank structure.
- ☹ To fill a d-D hypercube with ϵ-balls $\mathcal{B}_\epsilon\left(L^2\right)$, one needs

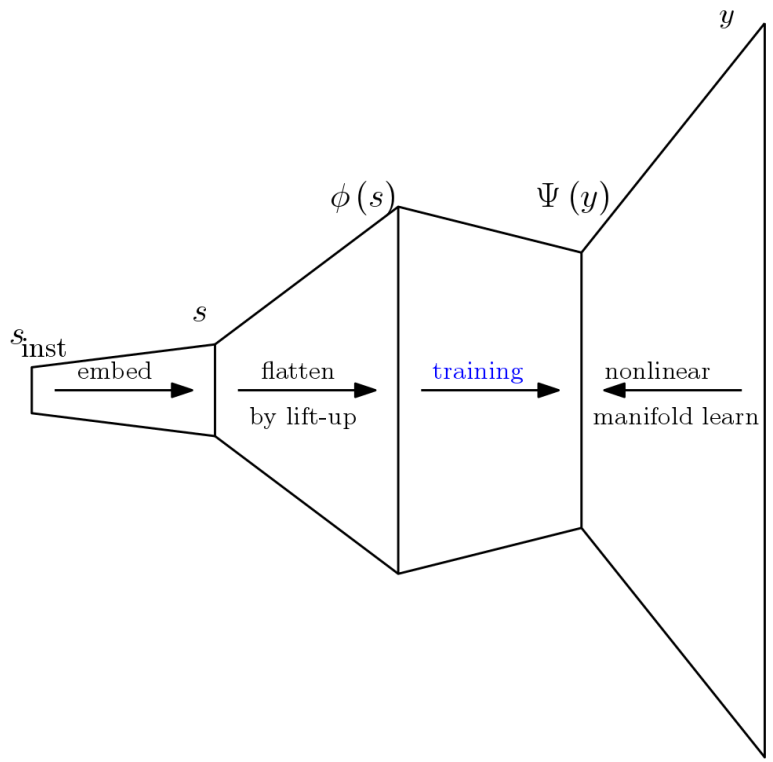
$$n_{\text{snap}} \sim \epsilon^{-d}\, d^{d/2} \quad \text{samples}$$

$\longrightarrow$ NN-based approach is doomed to fail.

- ☹ Not so much data $\longrightarrow$ precludes deep-$\star$ approaches.

What we may do

- Alleviate the loss of Markovianity by embedding the measurements
- Unfold the measurements to disentangle the information.
- Discover the manifold the flow data lie on. Hopefully of reasonable dimension.
- "Flatten" its relationship with the measurement features.



Flow features Ψ

Nonlinear manifold learning: Diffusion Maps

Flow features Ψ

Nonlinear manifold learning: Diffusion Maps

Kernel N_{DM} to quantify the distance between snapshots $\{\boldsymbol{y}_i\}_i$:

$$N_{\mathrm{DM}}\left(\boldsymbol{y}_i, \boldsymbol{y}_j\right) = \exp\left(-\left(\boldsymbol{y}_i - \boldsymbol{y}_j\right)^{\mathsf{T}} \Sigma_{N_{\mathrm{DM}}}^{-1} \left(\boldsymbol{y}_i - \boldsymbol{y}_j\right)\right).$$

With D the degree matrix, the Markov matrix quantifies the transition probabilities:

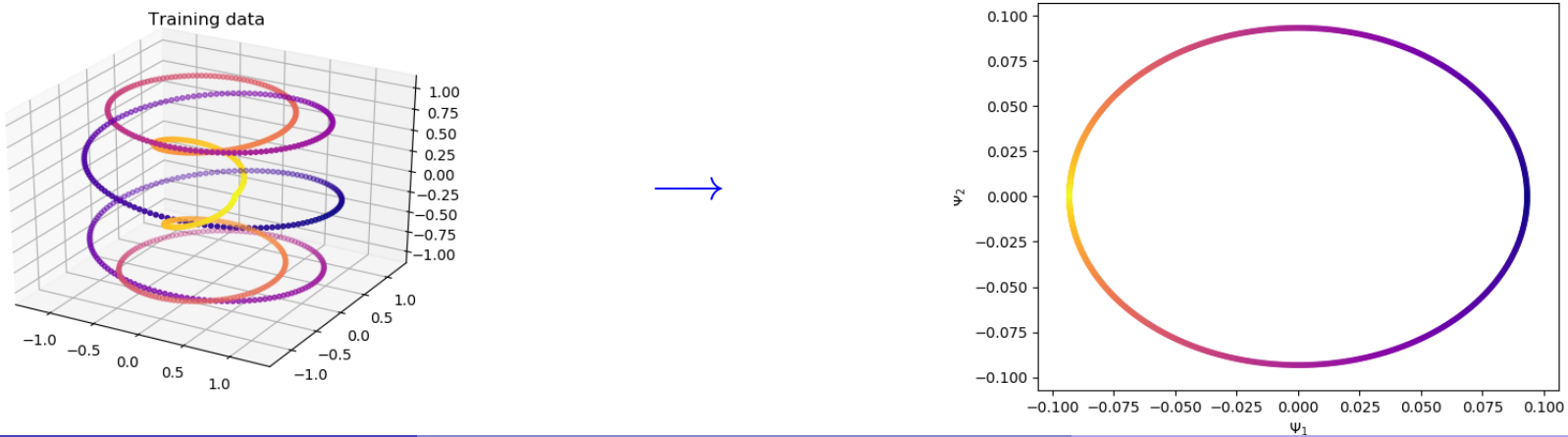
$$P = D^{-1}\, N_{\mathrm{DM}}, \qquad A = D^{-1/2}\, N_{\mathrm{DM}}\, D^{-1/2},$$

t-step transition matrix, with $t \in \mathbb{N}$ the diffusion "time" (acting as a scale):

$$P^t = R\, \Gamma^t\, L^{\mathsf{T}}.$$

Flow features Ψ_Y are modes of the diffusion map:

$$\Psi_Y = \Gamma^t\, R^{\mathsf{T}}.$$



Measurement features Φ

Lagrangian writes

$$\begin{aligned} \mathscr{L} &:= \lambda \, \|W\|_F^2 + \|\Psi_Y - \Psi_Y \, W \, \Phi(S)\|_F^2 \,, \\ &\propto \mathrm{Tr}\left[\lambda \, W \, W^{\mathsf{T}} + \Xi \, \Xi^{\mathsf{T}} + \alpha^{\mathsf{T}} \, (\Psi_Y - \Psi_Y \, W \, \Phi_S - \Xi)\right] , \qquad \text{[Dual version]} \end{aligned}$$

with $\Xi := \Psi_Y - \Psi_Y \, W \, \Phi_S$ and α the matrix of adjoint variables.

Measurement features Φ

Lagrangian writes

$$\begin{aligned} \mathscr{L} &:= \lambda \, \|W\|_F^2 + \|\Psi_Y - \Psi_Y \, W \, \Phi(S)\|_F^2 \,, \\ &\propto \mathrm{Tr}\left[\lambda \, W \, W^{\mathsf{T}} + \Xi \, \Xi^{\mathsf{T}} + \alpha^{\mathsf{T}} \, (\Psi_Y - \Psi_Y \, W \, \Phi_S - \Xi)\right], \qquad \text{[Dual version]} \end{aligned}$$

with $\Xi := \Psi_Y - \Psi_Y \, W \, \Phi_S$ and α the matrix of adjoint variables.

The measurement features Φ are determined using a Multi Kernel Learning technique.
Generalized minimax problem to learn the kernel of measurement features $K_{i,j} := \left\langle \Phi_{\boldsymbol{s}_i}, \Phi_{\boldsymbol{s}_j} \right\rangle$:

$$K(\cdot,\cdot) = \sum_k \mu_k \, K_k(\cdot,\cdot).$$

Measurement features Φ

Lagrangian writes

$$\begin{aligned}\mathscr{L} &:= \lambda \, \|W\|_F^2 + \|\Psi_Y - \Psi_Y \, W \, \Phi(S)\|_F^2 \,, \\ &\propto \mathrm{Tr}\left[\lambda \, W \, W^{\mathsf{T}} + \Xi \, \Xi^{\mathsf{T}} + \alpha^{\mathsf{T}} \, (\Psi_Y - \Psi_Y \, W \, \Phi_S - \Xi)\right], \qquad \text{[Dual version]}\end{aligned}$$

with $\Xi := \Psi_Y - \Psi_Y \, W \, \Phi_S$ and α the matrix of adjoint variables.

The measurement features Φ are determined using a Multi Kernel Learning technique.
Generalized minimax problem to learn the kernel of measurement features $K_{i,j} := \left\langle \Phi_{\boldsymbol{s}_i}, \Phi_{\boldsymbol{s}_j} \right\rangle$:

$$K(\cdot,\cdot) = \sum_k \mu_k \, K_k(\cdot,\cdot).$$

The adjoint variables $\widetilde{\alpha}$ satisfy a Sylvester equation:

$$\lambda \left(\Psi_Y \Psi_Y^{\mathsf{T}}\right)^{-1} \widetilde{\alpha} + \widetilde{\alpha} \, K_{SS} = \Psi_Y.$$

Wrapping-up...

Finally

$$\boldsymbol{s} \longrightarrow \Psi_{\boldsymbol{y}} \approx \widetilde{\alpha}\, K_S\left(\boldsymbol{s}\right).$$

Wrapping-up...

Finally

$$\boldsymbol{s} \longrightarrow \Psi_{\boldsymbol{y}} \approx \widetilde{\alpha}\, K_S\left(\boldsymbol{s}\right).$$

How to recover $\boldsymbol{y}$ from $\Psi_{\boldsymbol{y}}$? $\longrightarrow$ pre-image kernel problem.

If Gaussian kernel and isotropic $\longrightarrow$ Nyström extension technique to estimate the transition probabilities $p\left(\boldsymbol{y}, \boldsymbol{y}_i\right)$ and distances $D\left(\Psi_{\boldsymbol{y}_i}, \Psi_{\boldsymbol{y}}\right)$.

$\longrightarrow$ easy to relate to distances $d\left(\boldsymbol{y}_i, \boldsymbol{y}\right)$.

$\longrightarrow$ can estimate $\widehat{\boldsymbol{y}} \approx \boldsymbol{y}$ from knowning its distance to training samples $\{\boldsymbol{y}_i\}_i$.

Wrapping-up...

Finally

$$\boldsymbol{s} \longrightarrow \Psi_{\boldsymbol{y}} \approx \widetilde{\alpha}\, K_S\left(\boldsymbol{s}\right).$$

How to recover $\boldsymbol{y}$ from $\Psi_{\boldsymbol{y}}$? $\longrightarrow$ pre-image kernel problem.

If Gaussian kernel and isotropic $\longrightarrow$ Nyström extension technique to estimate the transition probabilities $p\left(\boldsymbol{y}, \boldsymbol{y}_i\right)$ and distances $D\left(\Psi_{\boldsymbol{y}_i}, \Psi_{\boldsymbol{y}}\right)$.

$\longrightarrow$ easy to relate to distances $d\left(\boldsymbol{y}_i, \boldsymbol{y}\right)$.

$\longrightarrow$ can estimate $\widehat{\boldsymbol{y}} \approx \boldsymbol{y}$ from knowning its distance to training samples $\{\boldsymbol{y}_i\}_i$.

At the end of the day:

$$\boldsymbol{s}_{\text{inst}} \longrightarrow \boldsymbol{s} \longrightarrow \Phi\left(\boldsymbol{s}\right) \longrightarrow \widehat{\Psi}\left(\widehat{\boldsymbol{y}}\right) \longrightarrow \widehat{\boldsymbol{y}}$$

Very much work in progress. . .

- Basis learning philosophy is a key enabler,
- Need to balance between representation accuracy and observability.

On-going efforts:

- More robust determination of the preimage,
- Anisotropic features kernel ($\rightarrow$ Laplace-Beltrami operator),
- Observability of the flow features yet to be enforced,
- More robust definition of the Bayes risk, *e.g.*, Wasserstein distance.