

On the importance of low-dimensional structures for data-driven modeling

Jean-Christophe Loiseau

May, 13th 2022

- Maître de Conférences in Fluid Dynamics and Applied Math.
- Machine-learning enthusiast with application to engineering systems.
- Data-efficient models with guarantees of optimality or interpretability.



Given training data $y_i = f(x_i)$ where f is unknown, we want to **learn** a function \hat{f} such that

$$\underset{\hat{f} \in \mathcal{F}}{\text{minimize}} \quad \|f(x) - \hat{f}(x)\|$$

where \mathcal{F} is a particular *class* of models.

A brief overview of SVD

$$\mathbf{A} = \mathbf{U} \ \boldsymbol{\Sigma} \ \mathbf{V}^T$$

Basis for $\text{colspan}(A)$

Basis for $\text{rowspan}(A)$

$$A = \color{red}U\color{black} \Sigma \color{blue}V^T$$

Diagonal matrix

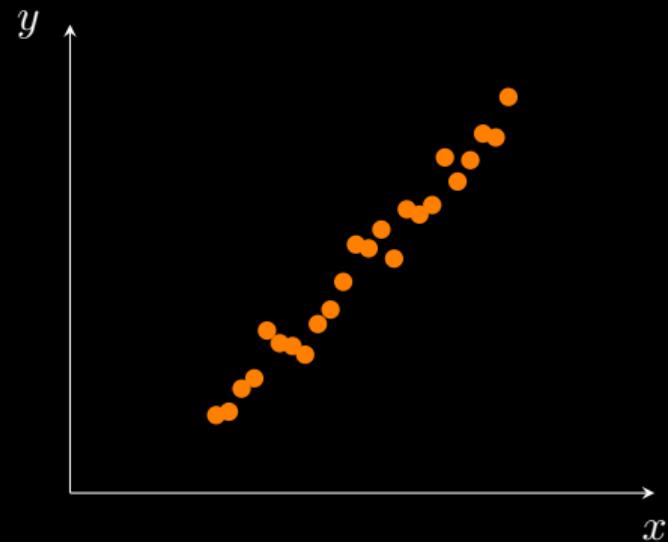
Relation to spectral decomposition

$$\begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix} = \sigma_i \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}$$

Generalization of the *eigenvalue decomposition* to **non-square matrices** by E. Beltrami (1873) and C. Jordan (1874). The first efficient numerical algorithm was developed by G. Golub *et al.* in the late 1960s.

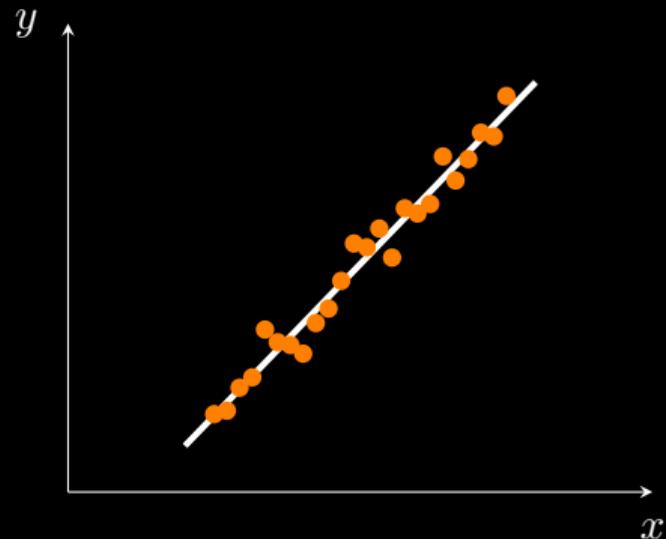
Ordinary least-squares

$$y = ax + b$$



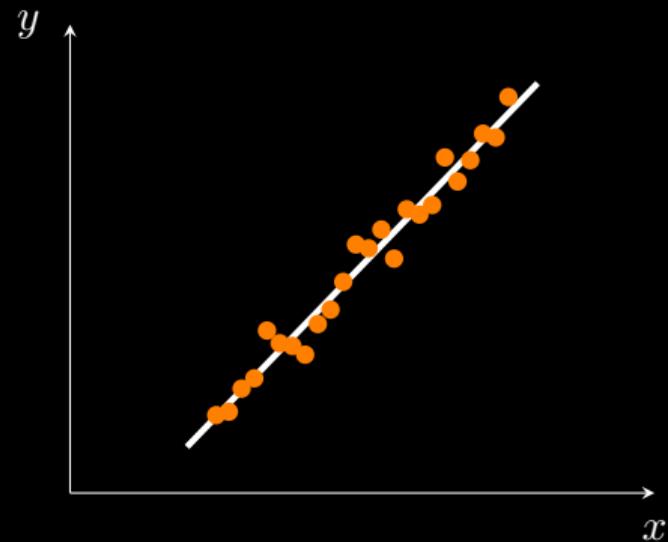
Ordinary least-squares

$$\underset{a,b}{\text{minimize}} \sum_{i=1}^N (y_i - ax_i - b)^2$$



Ordinary least-squares

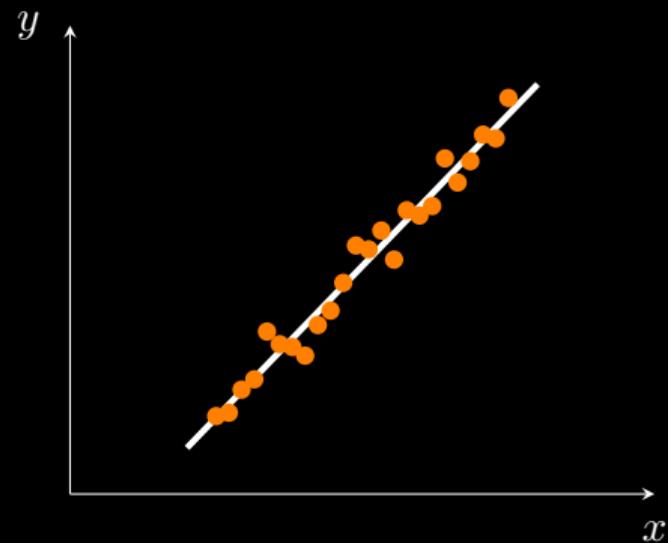
$$\underset{x}{\text{minimize}} \quad \|Ax - b\|_2^2$$



Ordinary least-squares

$$x = (A^T A)^{-1} A^T b$$

Moore-Penrose
pseudoinverse



$$\mathbf{A}^\dagger = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$$

$$A^\dagger = (V\Sigma U^T U\Sigma V^T)^{-1} V\Sigma U^T$$

$$\mathbf{A}^\dagger = \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^T$$

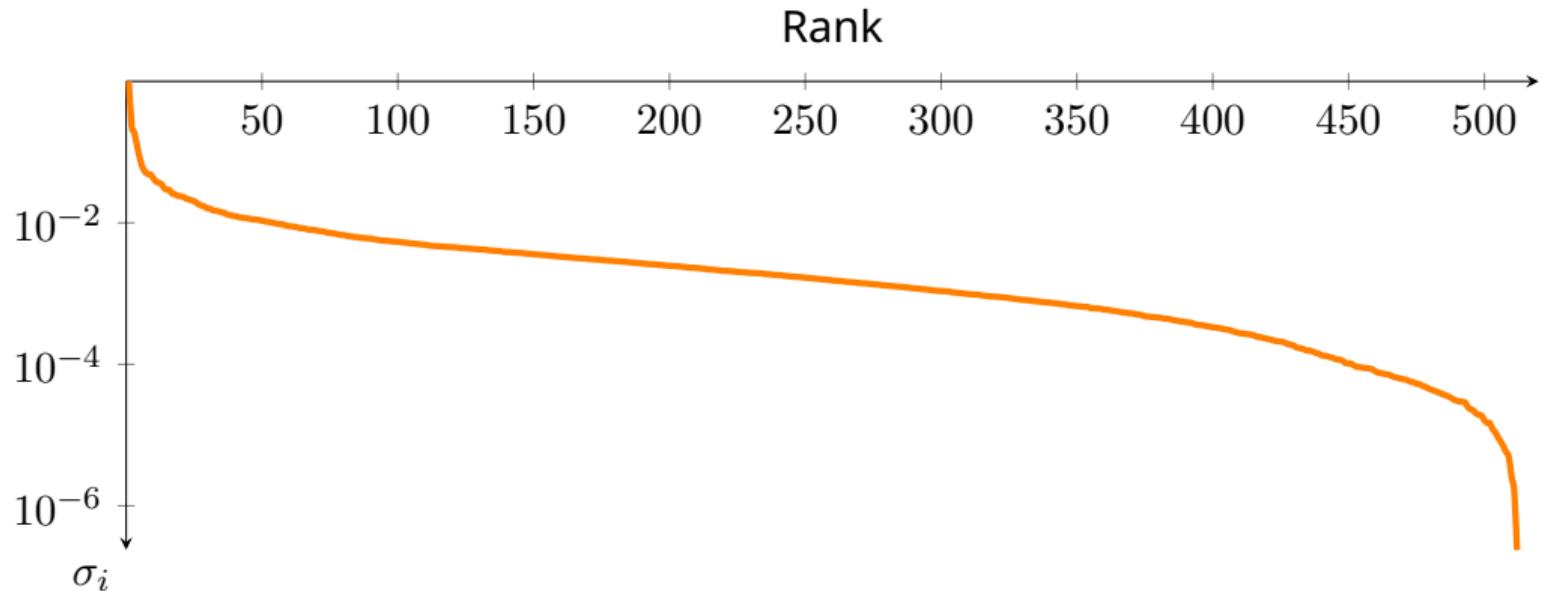
Low-rank approximation

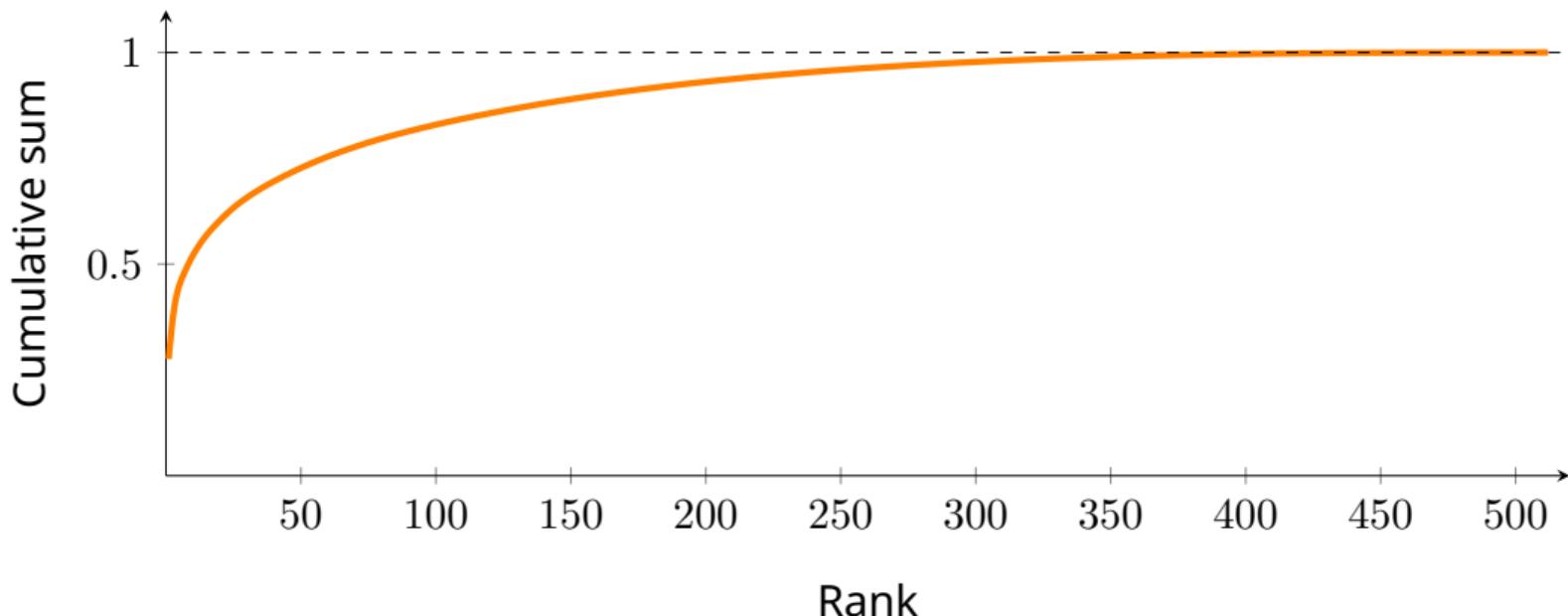


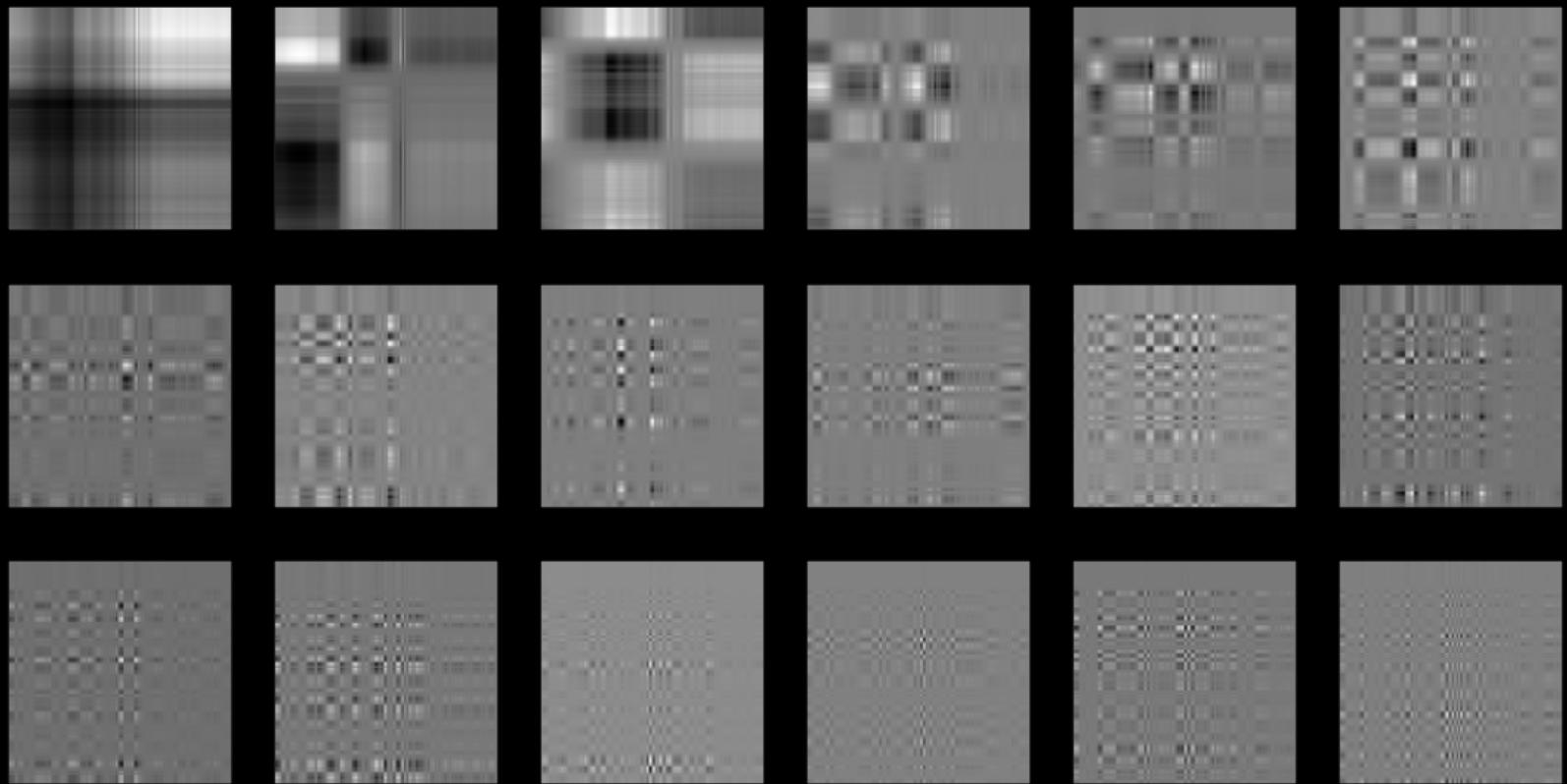
How to compress this image ?

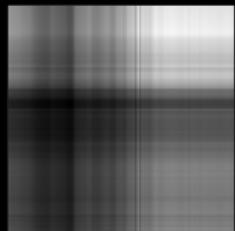
Low-rank approximation

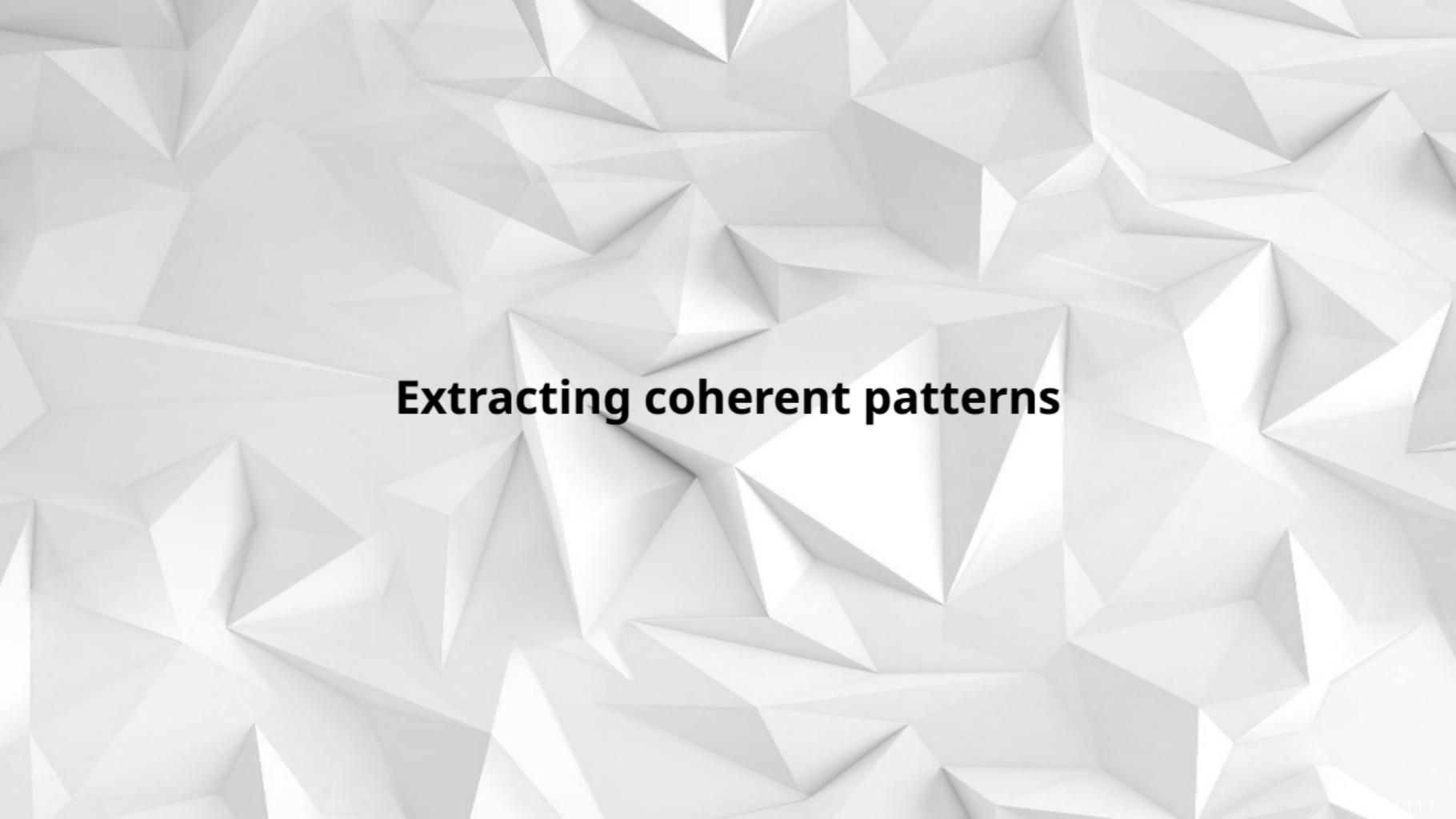
$$\begin{aligned} & \underset{\mathbf{X}}{\text{minimize}} && \|\mathbf{A} - \mathbf{X}\|_F^2 \\ & \text{subject to} && \text{rank } \mathbf{X} = r \end{aligned}$$









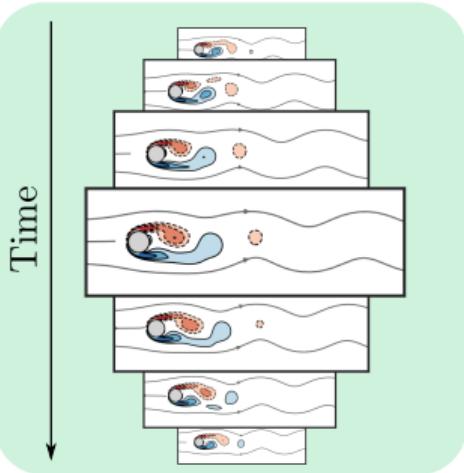
The background of the image is a white surface composed of numerous small, sharp-edged triangles, creating a geometric, crystalline or paper-folded effect.

Extracting coherent patterns

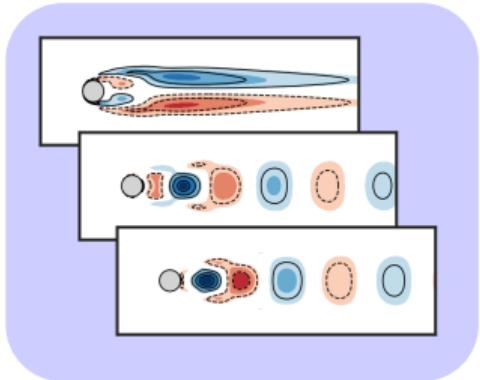


Add video of the cavity

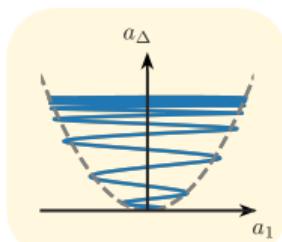
Navier-Stokes simulation

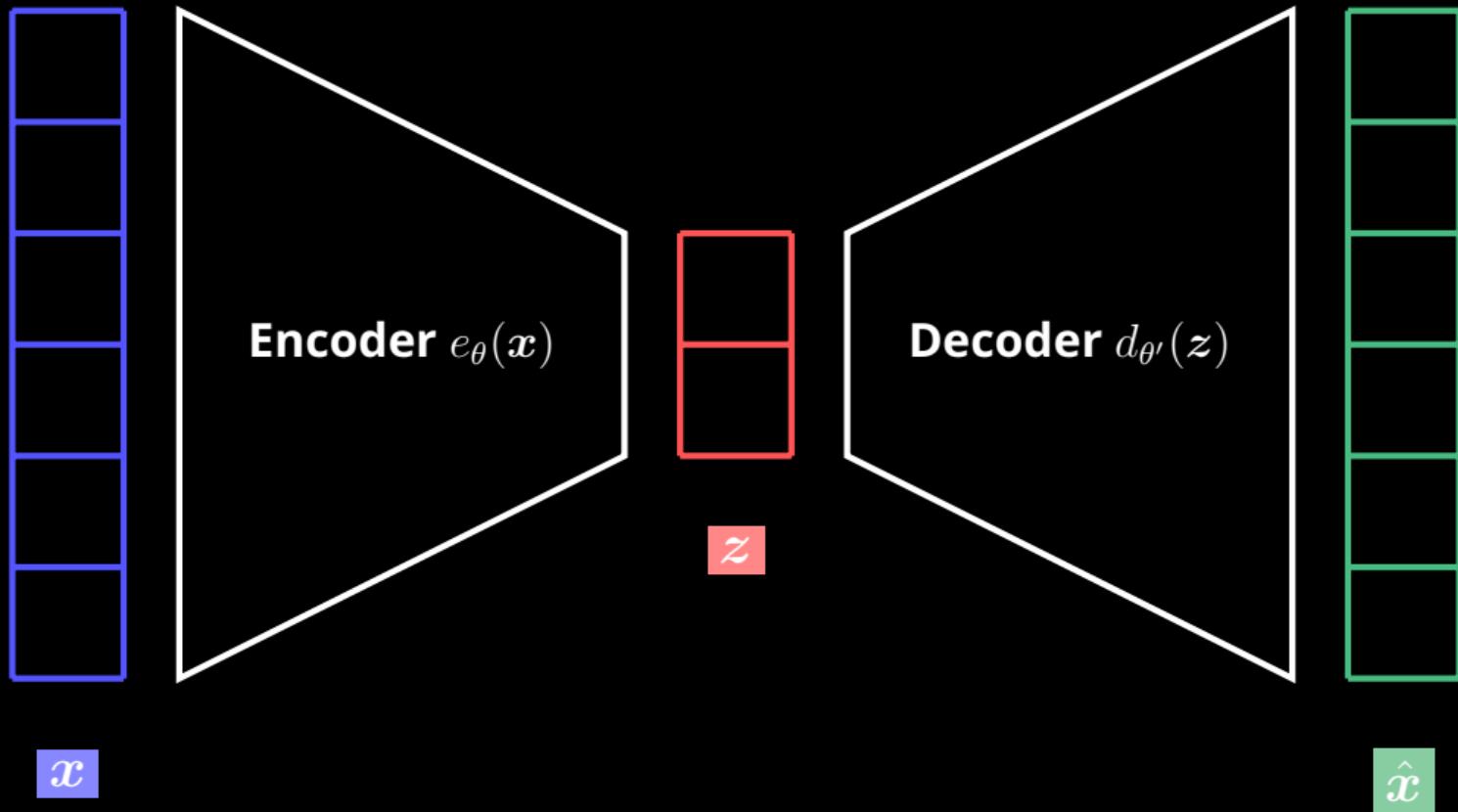


Dimensionality reduction



Simple representation





$$\min_{\theta, \theta'} \sum_{i=1}^N \| \mathbf{x}_i - (d_{\theta'} \circ e_{\theta})(\mathbf{x}_i) \|_2^2$$

Estimate



Ground truth

$$\begin{aligned} & \underset{\boldsymbol{P}, \boldsymbol{Q}}{\text{minimize}} && \sum_{i=1}^N \|\boldsymbol{x}_i - \boldsymbol{P}\boldsymbol{Q}^T\boldsymbol{x}_i\|_2^2 \\ & \text{subject to} && \text{rank } \boldsymbol{P} = \text{rank } \boldsymbol{Q} = r \end{aligned}$$

$$\begin{aligned} & \underset{\boldsymbol{P}}{\text{minimize}} && \sum_{i=1}^N \|\boldsymbol{x}_i - \boldsymbol{P}\boldsymbol{P}^T\boldsymbol{x}_i\|_2^2 \\ & \text{subject to} && \text{rank } \boldsymbol{P} = r \end{aligned}$$

$$\begin{aligned} & \underset{\boldsymbol{P}}{\text{minimize}} && \|\boldsymbol{X} - \boldsymbol{P}\boldsymbol{P}^T\boldsymbol{X}\|_F^2 \\ & \text{subject to} && \boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{I}_r \end{aligned}$$

Proper Orthogonal Decomposition

$$P\Lambda = C_{xx}P$$

P corresponds to the left singular vectors of X . The latent representation is given by $z_i = P^T x_i$. The optimal rank of the model can be inferred from the distribution of the PCA eigenvalues $\Lambda = \Sigma^2$.



The whole dataset can be correctly approximated using only 500 so-called eigen-faces.

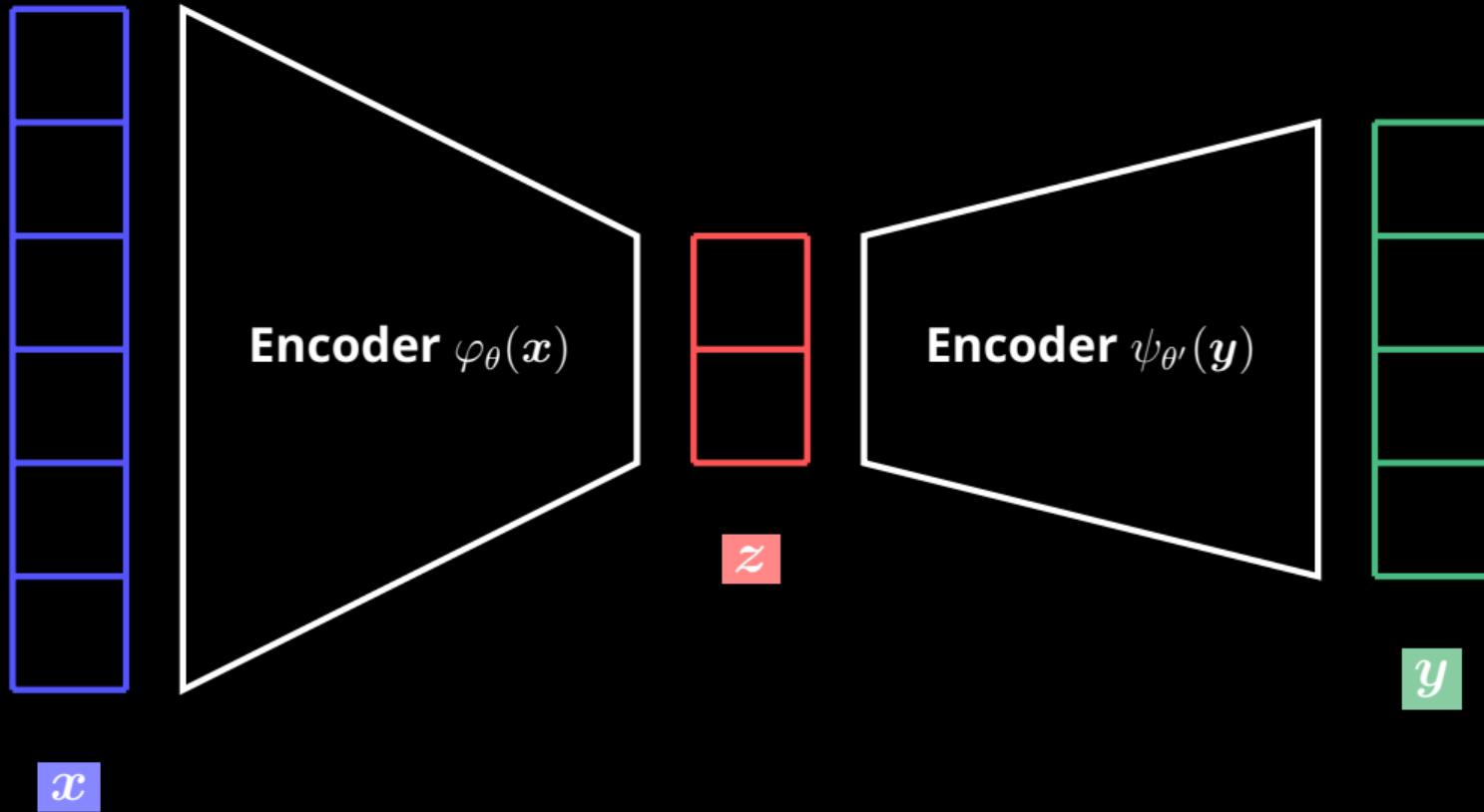
Shear-driven cavity POD modes

Add POD modes

Shear-driven cavity POD modes

Add phase portraits

Add cylinder flow and pressure coefficient



$$\min_{\theta, \theta'} \quad \sum_{i=1}^N \|\varphi_\theta(\mathbf{x}_i) - \psi_{\theta'}(\mathbf{y}_i)\|_2^2$$

$$\begin{aligned} & \underset{\boldsymbol{P}, \boldsymbol{Q}}{\text{minimize}} && \sum_{i=1}^N \|\boldsymbol{P}^T \boldsymbol{y}_i - \boldsymbol{Q}^T \boldsymbol{x}_i\|_2^2 \\ & \text{subject to} && \text{rank } \boldsymbol{P} = \text{rank } \boldsymbol{Q} = r \end{aligned}$$

$$\begin{aligned} & \underset{\boldsymbol{P}, \boldsymbol{Q}}{\text{minimize}} && \|\boldsymbol{P}^T \boldsymbol{Y} - \boldsymbol{Q}^T \boldsymbol{X}\|_F^2 \\ & \text{subject to} && \boldsymbol{P}^T \boldsymbol{C}_{yy} \boldsymbol{P} = \boldsymbol{Q}^T \boldsymbol{C}_{xx} \boldsymbol{Q} = \boldsymbol{I}_r \end{aligned}$$

Canonical Correlation Analysis

$$\begin{bmatrix} C_{yy} & 0 \\ 0 & C_{xx} \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix} \Sigma = \begin{bmatrix} 0 & C_{yx} \\ C_{xy} & 0 \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix}$$

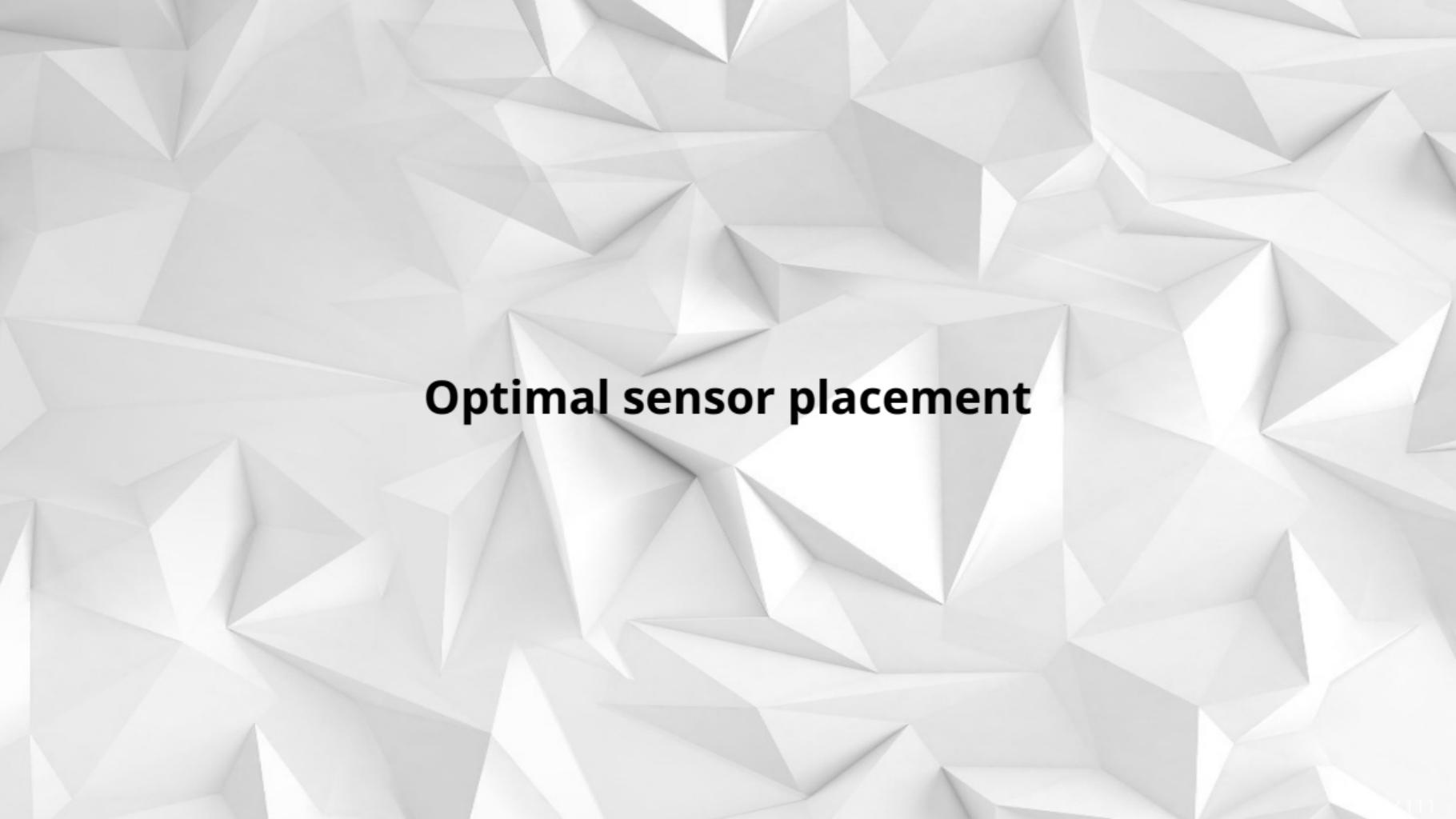
CCA relies on a *generalized eigenproblem*. P and Q describe the encoders such that the latent representations $z = Q^T x$ and $z' = P^T Y$ are as similar as possible. It is closely related to the concept of *mutual information*.

Add cylinder flow and pressure coefficient

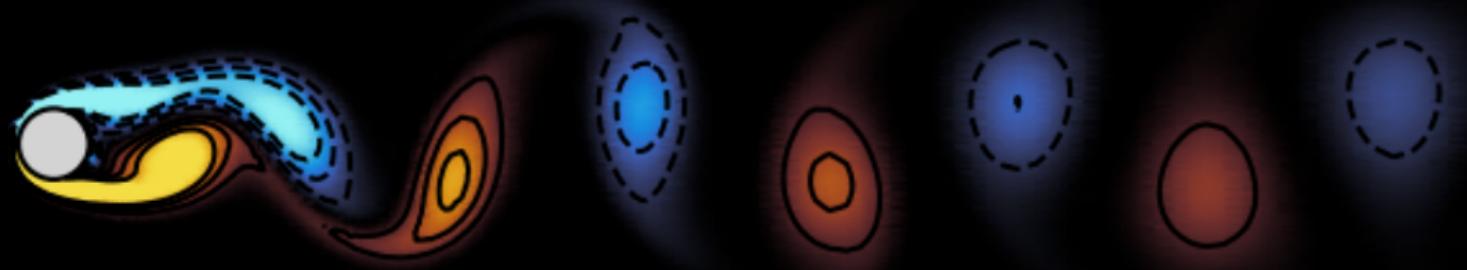
- Relies on extremely efficient numerical linear algebra techniques.
- The singular value distribution provides a simple diagnostic to estimate the dimensional of the embedding Euclidean space.
- Component analysis is well understood from a statistical point of view and benefits from good properties.



- The dimension inferred from the singular value distribution tends to over-estimate the intrinsic dimension of the underlying low-dimensional manifold.
 - This is directly related to the fact the linear dimensionality reduction techniques do not account for *nonlinear correlations*, only linear ones..
-

The background of the image is a white surface composed of numerous small, sharp-edged triangles, creating a geometric, crystalline or mountain-like texture.

Optimal sensor placement



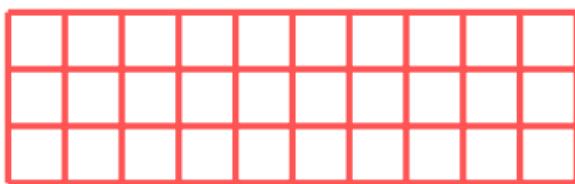
$$\mathbf{y} = \mathbf{C} \mathbf{x}$$

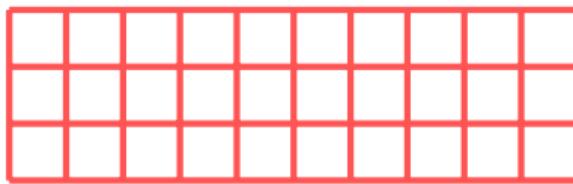
Measurement operator

Full state

Observations

The diagram illustrates a linear relationship between observations \mathbf{y} and the full state \mathbf{x} . The equation $\mathbf{y} = \mathbf{C} \mathbf{x}$ is shown with three colored brackets: a red bracket above the matrix \mathbf{C} labeled 'Measurement operator', a green bracket below the matrix \mathbf{x} labeled 'Full state', and a blue bracket to the left of the vector \mathbf{y} labeled 'Observations'.

y C x  \sim 

y C U z  \approx 

A large curly brace spanning the width of the red and green grids, labeled with the symbol Θ .

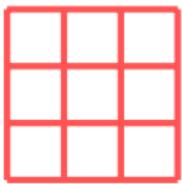
y

Θ

z



\sim



$$\underset{\boldsymbol{z}}{\text{minimize}} \quad \|\boldsymbol{y} - \boldsymbol{\Theta}\boldsymbol{z}\|_2$$

$$z = \Theta^{-1}y$$

$$\underset{\boldsymbol{C}}{\text{maximize}} \quad |\det(\boldsymbol{C}\boldsymbol{U})|$$

$$\begin{aligned} & \underset{\boldsymbol{C}}{\text{maximize}} && |\det(\boldsymbol{C}\boldsymbol{U})| \\ & \text{subject to} && \boldsymbol{C}_i \in \{\boldsymbol{e}_j\}_{j=1,n} \end{aligned}$$

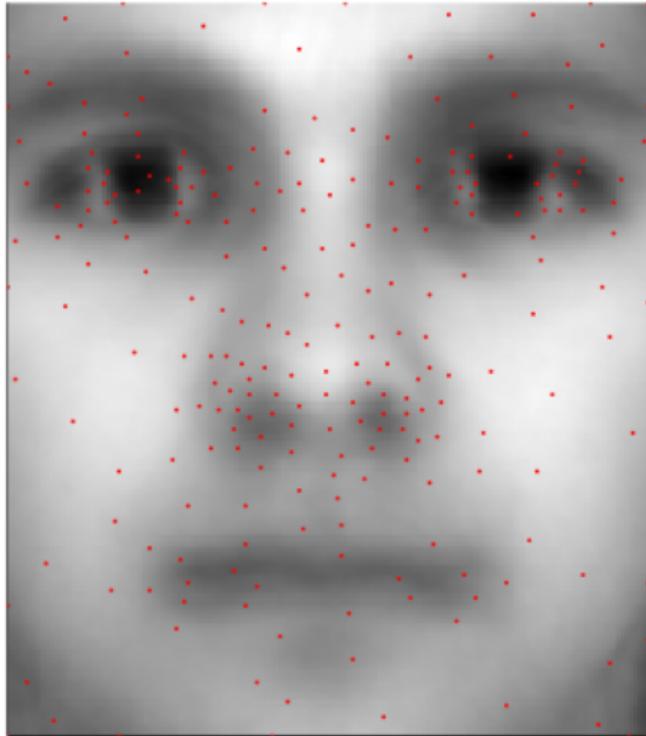
$$\mathbf{U}^T \mathbf{P} = \mathbf{Q} \mathbf{R}$$

Permutation matrix

Low-rank basis

Upper triangular matrix
with $|r_{i-1}| \geq |r_i|$

The diagram illustrates the QR decomposition of a matrix. It shows the equation $\mathbf{U}^T \mathbf{P} = \mathbf{Q} \mathbf{R}$. Above the equation, a red bracket labeled "Permutation matrix" points to the red box containing \mathbf{P} . To the left of the equation, a blue bracket labeled "Low-rank basis" points to the blue box containing \mathbf{U}^T . To the right of the equation, a green bracket labeled "Upper triangular matrix with $|r_{i-1}| \geq |r_i|$ " points to the green box containing \mathbf{R} .



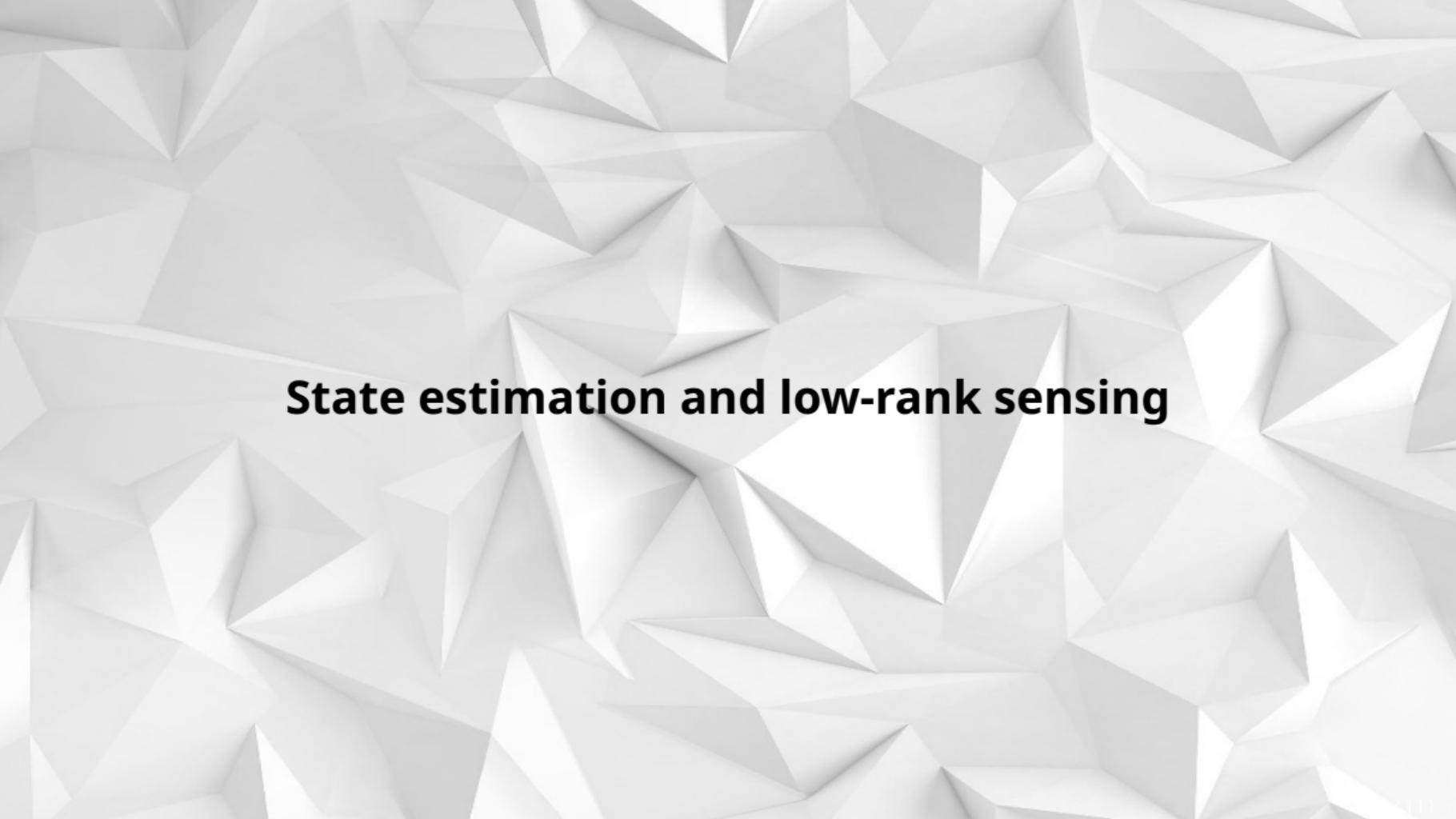
Shear-driven cavity flow

- Relies on extremely efficient numerical linear algebra techniques.
- Greedy solution to an otherwise intractable combinatorial problem.
- Relevant in many situations and benefits from good empirical performances.



- No easy way to estimate how far from the optimum the greedy solution is.

—

The background of the slide features a complex, abstract pattern of white polygons, resembling a low-poly 3D model or a crystal lattice. The polygons are of various sizes and orientations, creating a sense of depth and texture. The lighting is soft, with subtle shadows and highlights that emphasize the three-dimensional nature of the surface.

State estimation and low-rank sensing

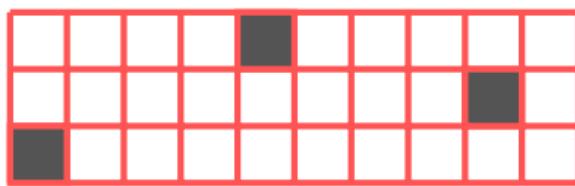
$$\mathbf{y} = \mathbf{C} \mathbf{x}$$

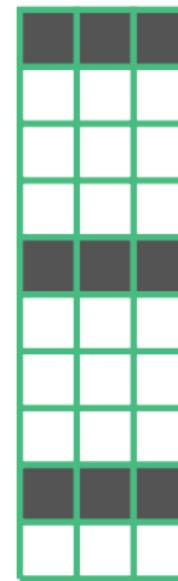
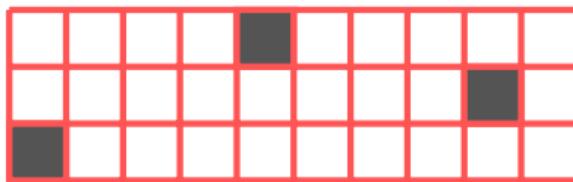
Measurement operator

Full state

Observations

The diagram illustrates a linear system $\mathbf{y} = \mathbf{C} \mathbf{x}$. A red bracket labeled "Measurement operator" points down to the matrix \mathbf{C} . A green bracket labeled "Full state" points up to the matrix \mathbf{x} . A blue bracket labeled "Observations" points left to the vector \mathbf{y} .

y C x  \sim 

y C U z  \sim  Θ

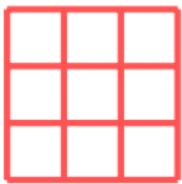
y

Θ

z



\approx



Underdetermined problem

$$\begin{aligned} & \underset{\mathbf{z}}{\text{minimize}} && \|\mathbf{z}\|_2 \\ & \text{subject to} && \mathbf{y} = \Theta\mathbf{z} \end{aligned}$$

Overdetermined problem

$$\underset{\mathbf{z}}{\text{minimize}} \quad \|\mathbf{y} - \Theta\mathbf{z}\|_2^2$$

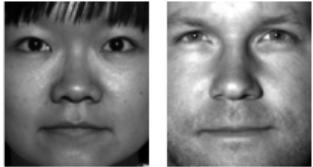
Regularized problem

$$\underset{\mathbf{z}}{\text{minimize}} \quad \|\mathbf{y} - \Theta\mathbf{z}\|_2^2 + \lambda\|\mathbf{z}\|_2^2$$

Regularized and constrained problem

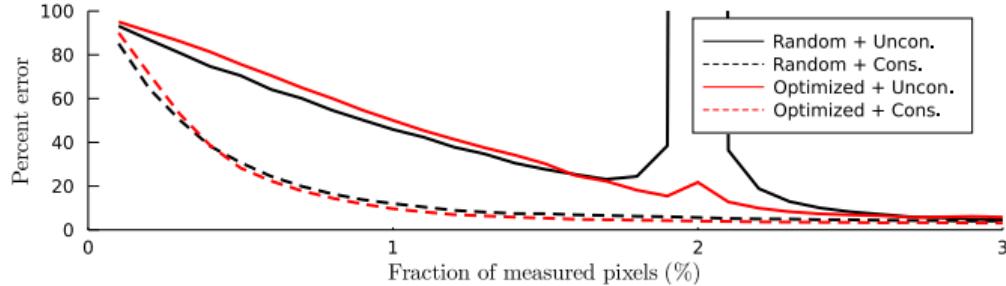
$$\begin{aligned} & \underset{\mathbf{z}}{\text{minimize}} && \|\mathbf{y} - \Theta\mathbf{z}\|_2^2 + \lambda\|\mathbf{z}\|_2^2 \\ & \text{subject to} && |z_i| \leq 2\sigma_i \quad \forall i \end{aligned}$$

Ground truth



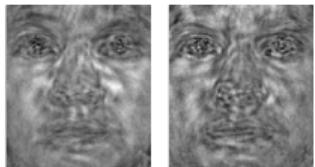
Underdetermined

Overdetermined

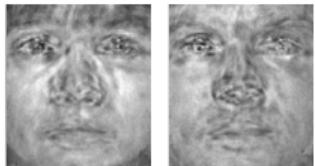


Unconstrained

0.5%



1.4%

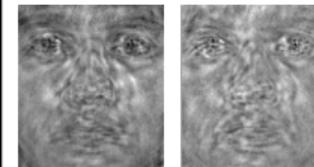


Random sensors

Box-constrained

Unconstrained

Box-constrained



QR-optimized sensors



- Low-rank sensing benefits from strong theoretical guarantees.
- Relies on simple yet efficient numerical procedures, both during the training and deployment stages.
- Requires much less data than naïve deep learning alternatives.



- Data needs to be standardized and characterized by an underlying low-rank structure.
- The state estimator is a **static map**. It does not account for the dynamics (if they exist) of the generating process.
- The measurement equation needs to be (approximately) linear.

—

Low-rank structures and sparse sensing for classification

Is it a cat or a dog ?

Linear classifier

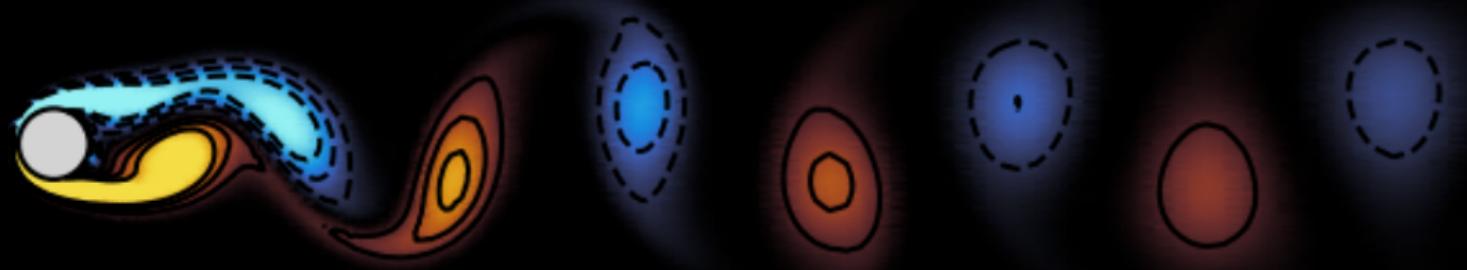
$$\text{sign}(\mathbf{w}^T \mathbf{x} + b) = \begin{cases} 1 & \text{then it is a cat,} \\ -1 & \text{otherwise a dog.} \end{cases}$$

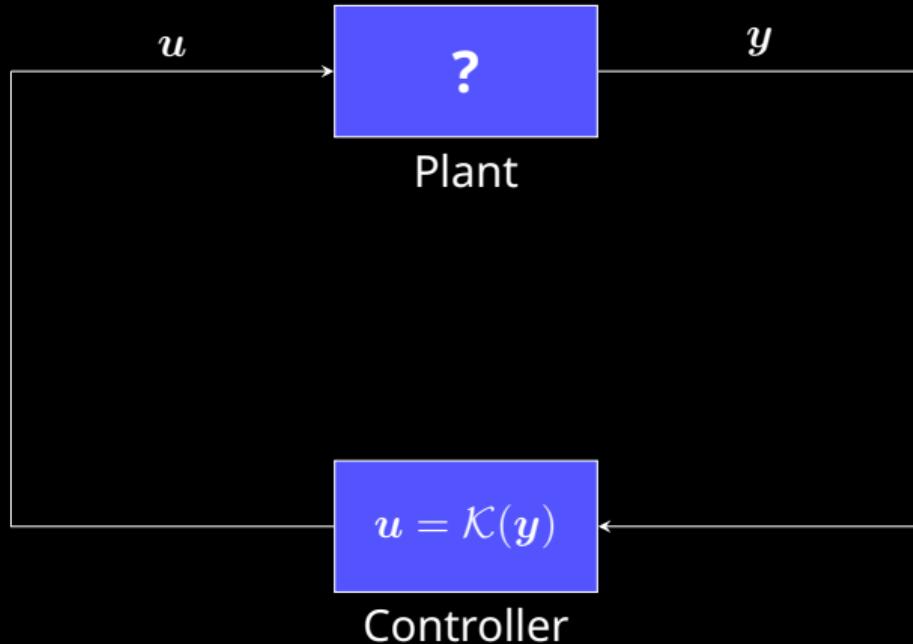
Eigenpets

$\mathbf{w} \in \mathbb{R}^{20}$, hence only 20 pixels would be necessary to classify the image. But how to choose these most informative pixels ?

$$\begin{aligned} & \underset{\mathbf{s}}{\text{minimize}} && \|\mathbf{s}\|_1 \\ & \text{subject to} && \mathbf{U}^T \mathbf{s} = \mathbf{w} \end{aligned}$$

System identification





Convolution model

$$y_i = f(u_i, u_{i-1}, u_{i-2}, \dots)$$

$$\begin{array}{l} \text{Natural dynamics} \\ \hline \\ \xrightarrow{\hspace{1cm}} \quad \quad \quad \downarrow \\ \boldsymbol{x}_{i+1} = \boxed{\boldsymbol{A}} \boldsymbol{x}_i + \boxed{\boldsymbol{B}} \boldsymbol{u}_i \\ \quad \quad \quad \uparrow \\ \text{Measurements} \end{array} \quad \quad \quad \begin{array}{l} \text{Actuators} \\ \hline \\ \downarrow \quad \quad \quad \uparrow \\ \boldsymbol{y}_i = \boxed{\boldsymbol{C}} \boldsymbol{x}_i + \boxed{\boldsymbol{D}} \boldsymbol{u}_i \\ \quad \quad \quad \uparrow \\ \text{Feedthrough} \end{array}$$

Input u_i	State x_i	Output y_i
u_0	0	Du_0
u_1	Bu_0	$CBu_0 + Du_1$
u_2	$ABu_0 + Bu_1$	$CABu_0 + CBu_1 + Du_2$
u_3	$A^2Bu_0 + ABu_1 + Bu_2$	$CA^2Bu_0 + CABu_1 + CBu_2 + Du_3$
\vdots	\vdots	\vdots
u_k	$\sum_{i=1}^k A^{k-i} Bu_{k-i}$	$\sum_{i=1}^k CA^{k-i} Bu_{k-i} + Du_k$

$$\text{Output} \quad \text{Unknown conv. kernel}$$
$$y(t) = h(t) * u(t)$$
$$\text{Input}$$

```
graph TD; Output[Output] --> y["y(t)"]; Kernel[Unknown conv. kernel] --> h["h(t)"]; Input[Input] --> u["u(t)"]
```

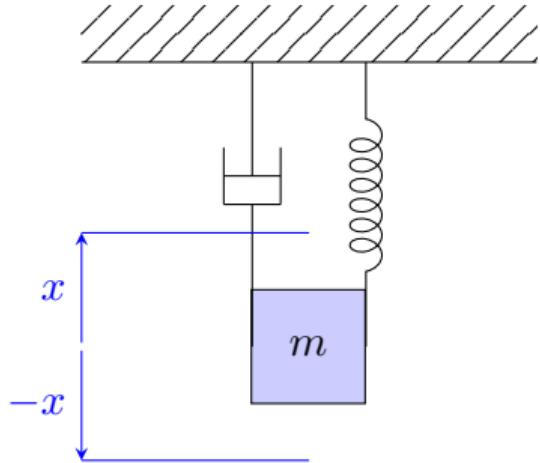
$$\begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} u_0 & \cdots & & & \\ u_1 & u_0 & \cdots & & \\ u_2 & u_1 & u_0 & \cdots & \\ u_3 & u_2 & u_1 & u_0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} D \\ CB \\ CAB \\ CA^2B \\ CA^3B \\ \vdots \end{bmatrix}$$

$$\underset{h}{\text{minimize}} \quad \| \mathbf{y} - \mathcal{T}(\mathbf{u}) \mathbf{h} \|_p + \mathcal{R}(h)$$

Output sequence

Unknown conv. kernel

Toeplitz matrix built
using the input sequence



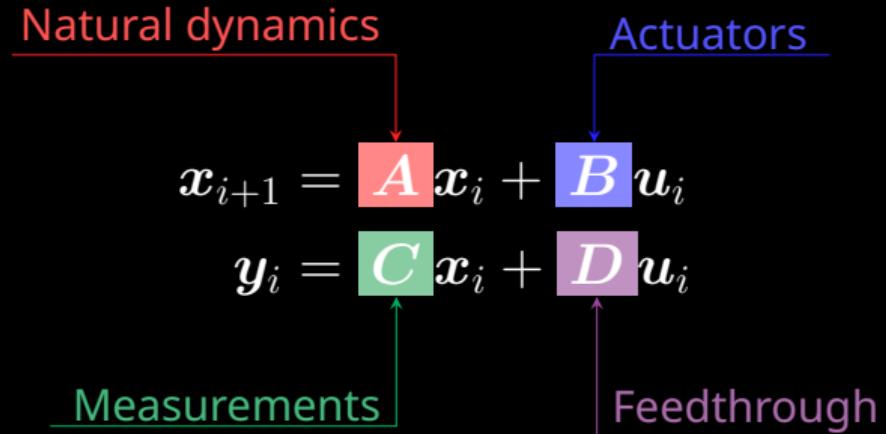
$$\frac{d}{dt} \begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & -2\zeta \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$
$$y = [1 \ 0] \begin{bmatrix} x \\ v \end{bmatrix}$$

- Relies on simple linear regression.
- Easy to extend to nonlinear systems (e.g. LSTM)
- Prior knowledge can be included through regularization or constraints.



- The length of the convolution kernel is *a priori* unknown.
- The longer the kernel, the more data is needed to obtain converged estimates.
- It is *a priori* unrelated to the number of degrees of freedom in the system.

—



$$\mathcal{O}_k = \begin{bmatrix} C \\ CA \\ CA^2 \\ CA^3 \\ \vdots \\ CA^{k-1} \end{bmatrix} \quad \mathcal{C}_k = [B \ AB \ A^2B \ A^3B \ \dots \ A^{k-1}B]$$

Observability

Controlability

$$h_k = [D \ CB \ CAB \ CA^2B \ CA^3B \ \dots \ CA^{k-1}B]$$

Markov parameters of the system

EigenRealization Algorithm

$$\mathbf{H}_1 = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 & h_5 \\ h_2 & h_3 & h_4 & h_5 & h_6 \\ h_3 & h_4 & h_5 & h_6 & h_7 \\ h_4 & h_5 & h_6 & h_7 & h_8 \\ h_5 & h_6 & h_7 & h_8 & h_9 \end{bmatrix}$$

EigenRealization Algorithm

$$H_1 = \begin{bmatrix} CB & CAB & CA^2B & CA^3B & CA^4B \\ CAB & CA^2B & CA^3B & CA^4B & CA^5B \\ CA^2B & CA^3B & CA^4B & CA^5B & CA^6B \\ CA^3B & CA^4B & CA^5B & CA^6B & CA^7B \\ CA^4B & CA^5B & CA^6B & CA^7B & CA^8B \end{bmatrix}$$

EigenRealization Algorithm

$$H_1 = \begin{bmatrix} C \\ CA \\ CA^2 \\ CA^3 \\ CA^4 \end{bmatrix} [B \ AB \ A^2B \ A^3B \ A^4B]$$

EigenRealization Algorithm

Observability: $\mathcal{O} = U\Sigma^{\frac{1}{2}}$

Controlability: $\mathcal{C} = \Sigma^{\frac{1}{2}}V^T$

EigenRealization Algorithm

$$\mathbf{H}_2 = \begin{bmatrix} h_2 & h_3 & h_4 & h_5 & h_6 \\ h_3 & h_4 & h_5 & h_6 & h_7 \\ h_4 & h_5 & h_6 & h_7 & h_8 \\ h_5 & h_6 & h_7 & h_8 & h_9 \\ h_6 & h_7 & h_8 & h_9 & h_{10} \end{bmatrix}$$

EigenRealization Algorithm

$$H_2 = \begin{bmatrix} C \\ CA \\ CA^2 \\ CA^3 \\ CA^4 \end{bmatrix} A [B \ AB \ A^2B \ A^3B \ A^4B]$$

EigenRealization Algorithm

Natural dynamics

$$A = \mathcal{O}^\dagger H_2 \mathcal{C}^\dagger$$

Actuators

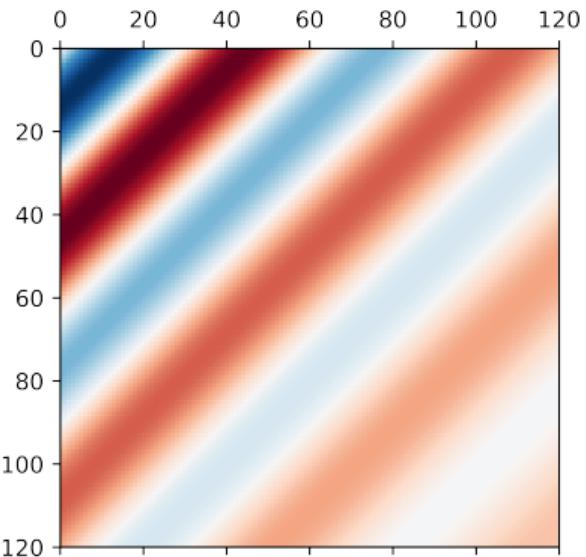
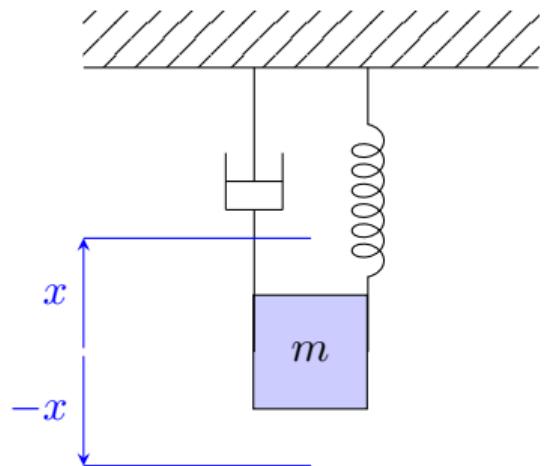
$$B = \left[\Sigma^{\frac{1}{2}} V^T \right]_{:,1:p}$$

Measurements

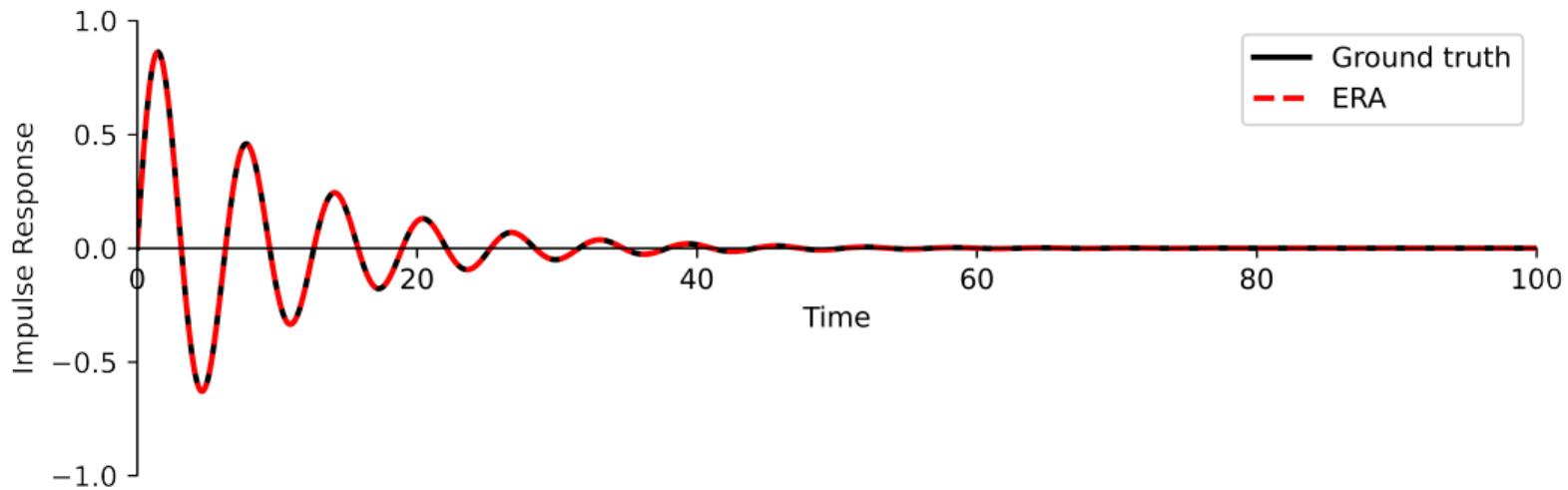
$$C = \left[U \Sigma^{\frac{1}{2}} \right]_{1:q,:}$$

Feedthrough

$$D = h_0$$



$$\Sigma = [27.68 \quad 22.62 \quad 0 \quad 0 \quad \dots]$$



Cylinder flow example

- Relies solely on the impulse response of the system and efficient numerical linear algebra.
- Identifies the minimal number of degrees of freedom needed to represent the dynamics.
- Benefits from strong theoretical guarantees and is well-grounded in control theory.



- Need a really good estimate of the impulse response.
- Extension to **interpretable** nonlinear systems is not straightforward.
- Parametric variations can be easily included into this framework.

—

Conclusion

