

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

Jean-Christophe Loiseau

May, 13<sup>th</sup> 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.



# **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

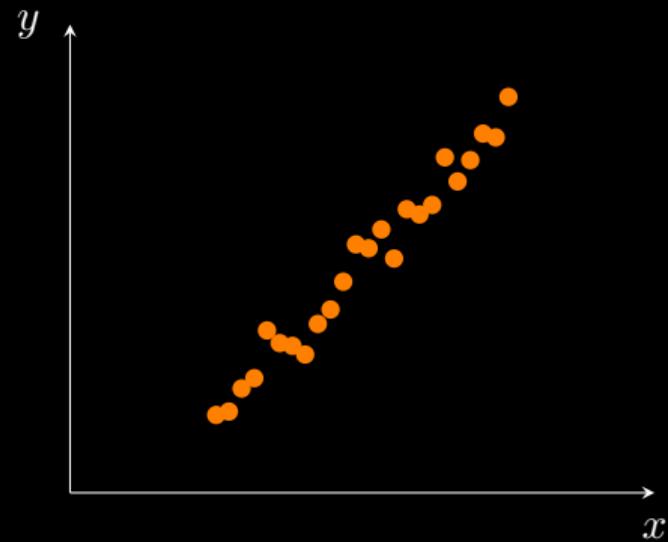
$$\begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix} = \sigma_i \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}$$

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.



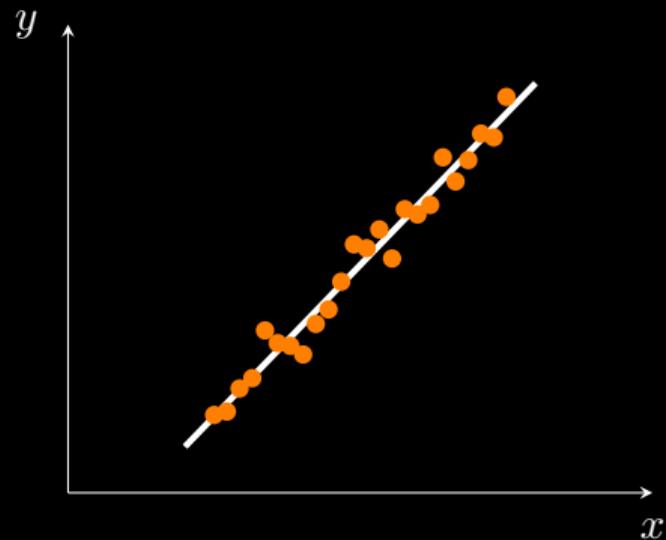
# Ordinary least-squares

$$y = ax + b$$



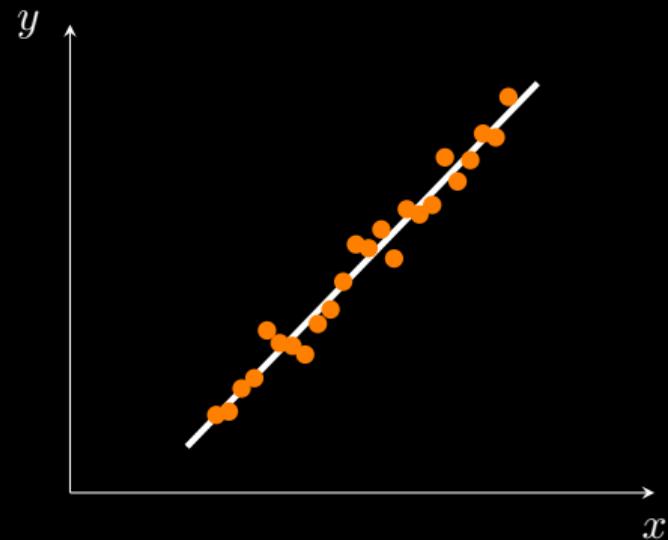
# Ordinary least-squares

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



# Ordinary least-squares

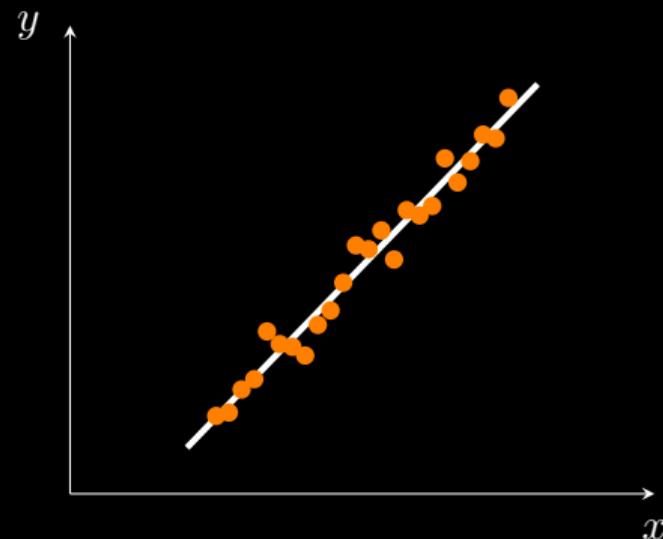
$$\underset{x}{\text{minimize}} \quad \|Ax - b\|_2^2$$



# Ordinary least-squares

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

Moore-Penrose  
pseudoinverse



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

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

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

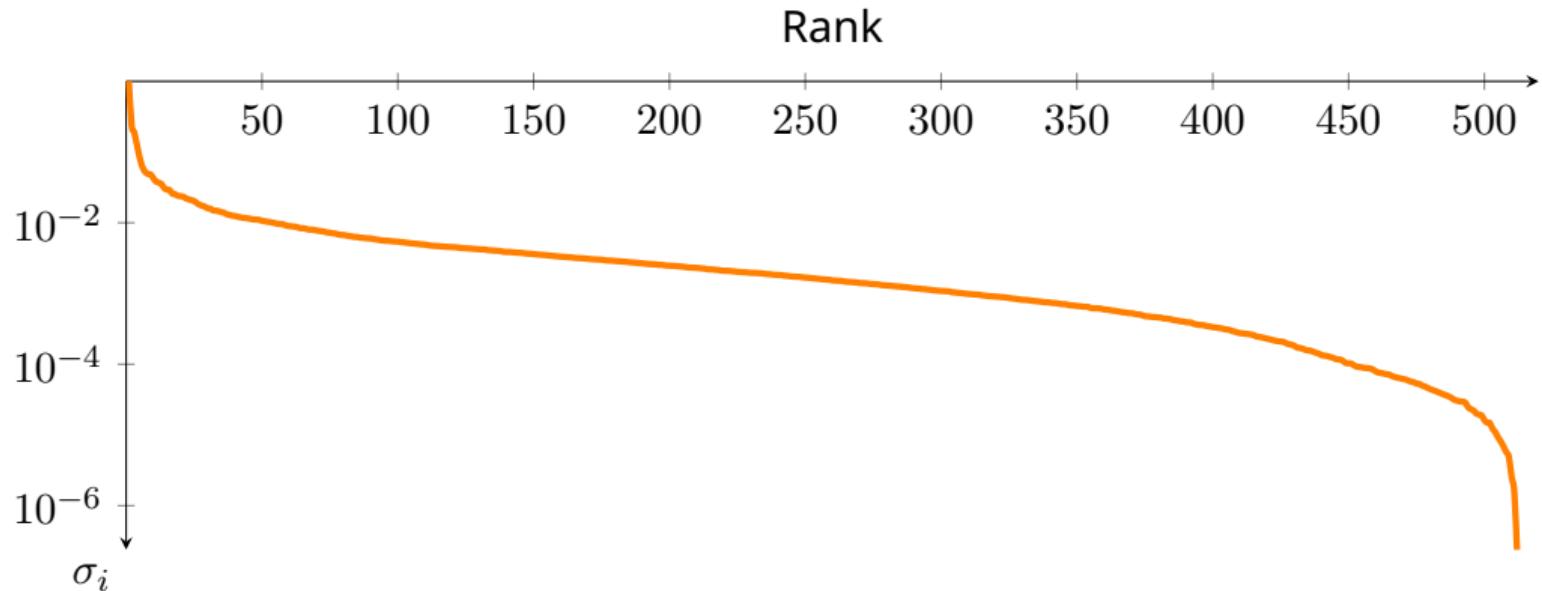
# Low-rank approximation

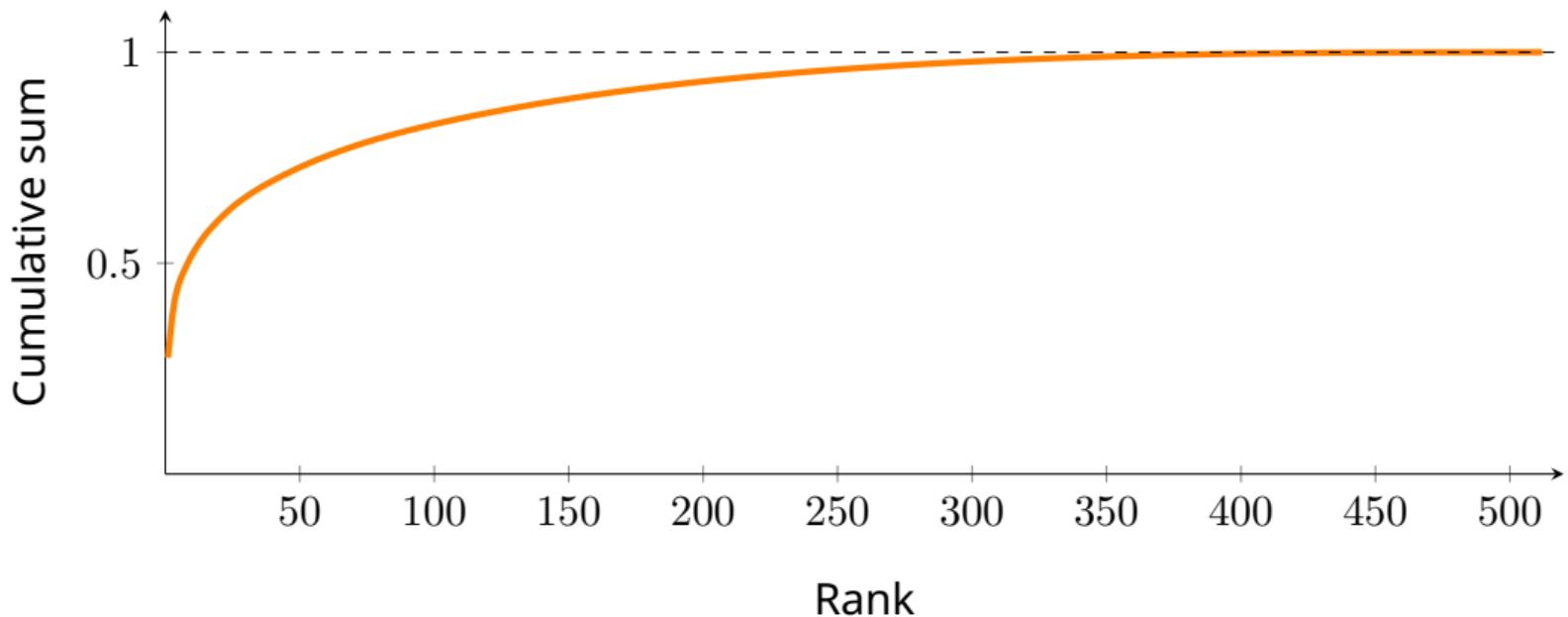


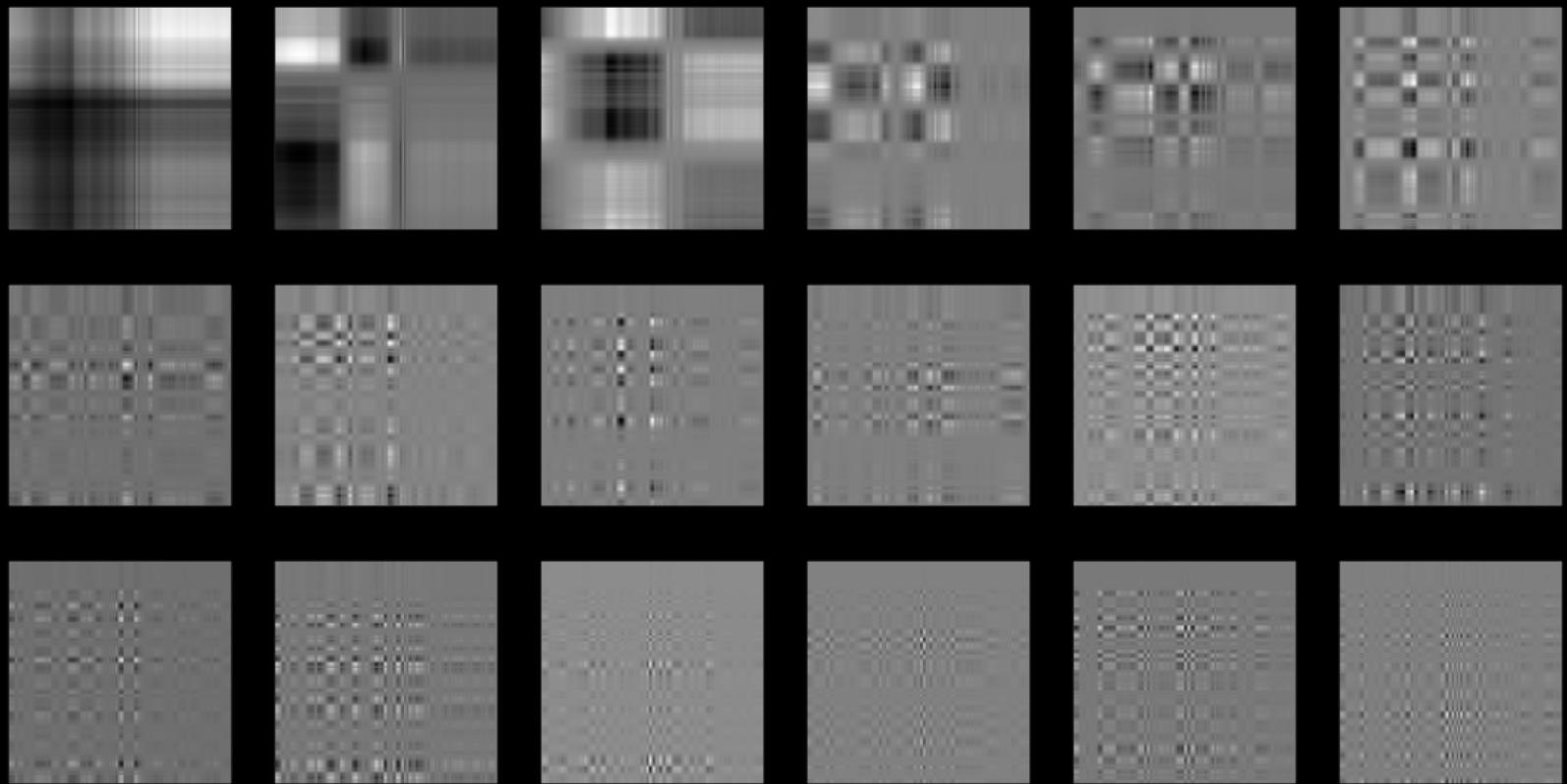
How to compress this image ?

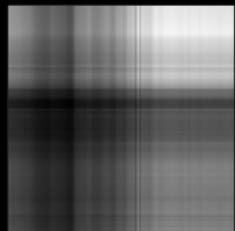
# Low-rank approximation

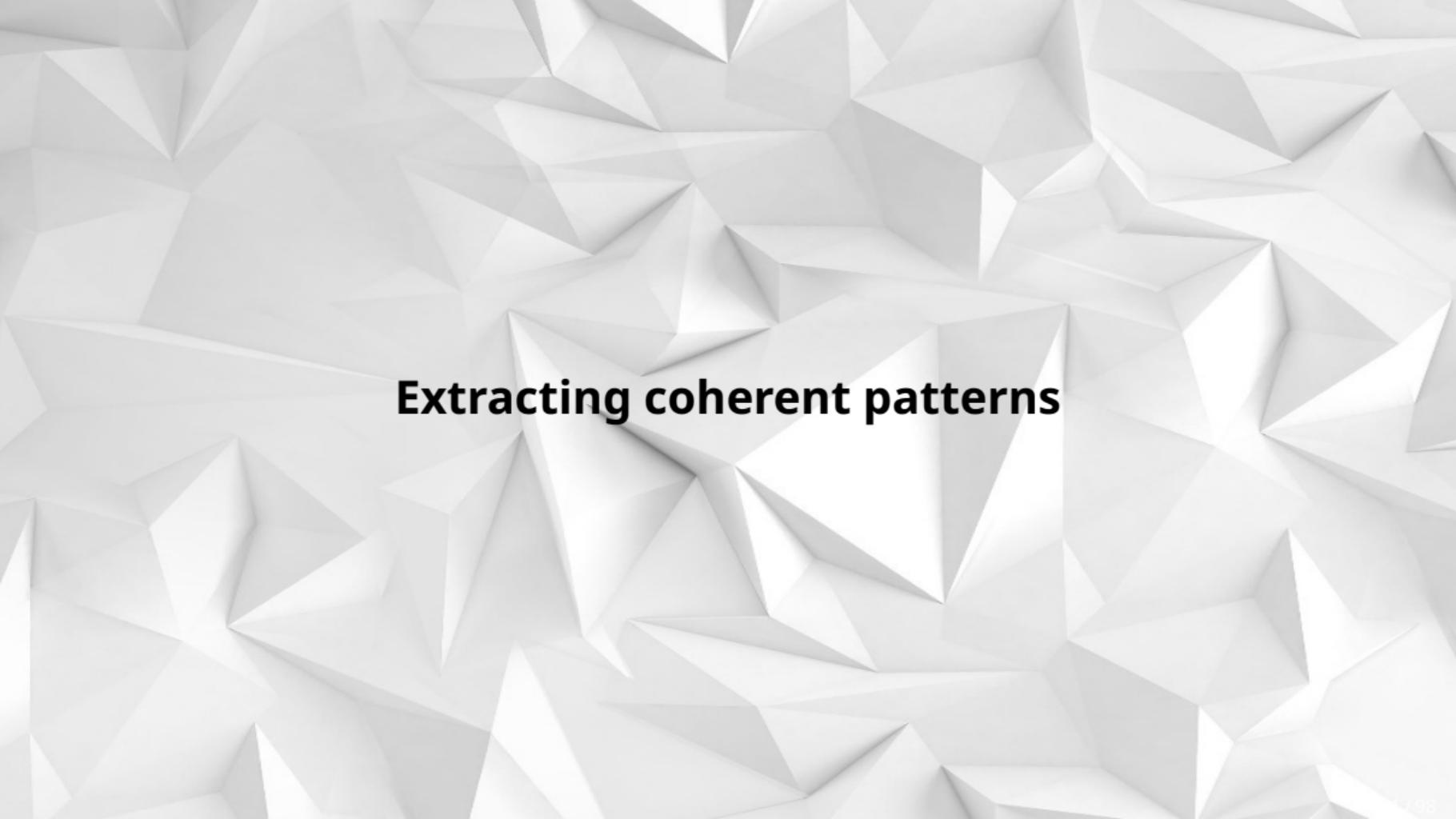
$$\begin{aligned} & \underset{\mathbf{X}}{\text{minimize}} && \|\mathbf{A} - \mathbf{X}\|_F^2 \\ & \text{subject to} && \text{rank } \mathbf{X} = r \end{aligned}$$









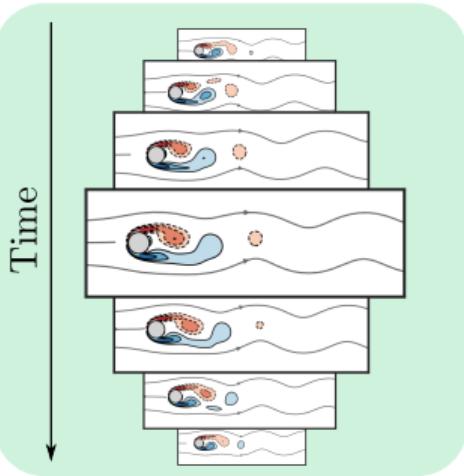
The background of the slide features a complex, abstract pattern of white polygons, likely triangles, arranged in a way that creates a sense of depth and texture. The polygons vary in size and orientation, with some appearing as sharp peaks and others as recessed valleys. The lighting is soft, highlighting the edges of the polygons and creating a subtle gradient across the surface.

**Extracting coherent patterns**

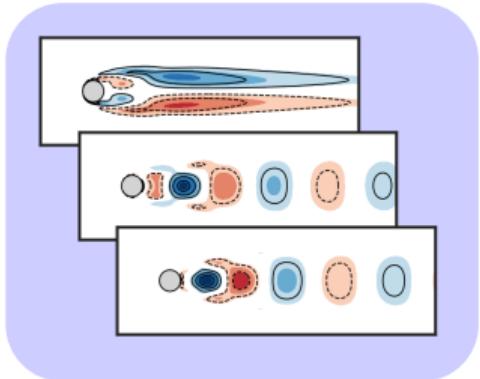
Add Yale B faces

Add video of the cavity

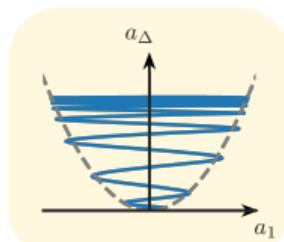
Navier-Stokes simulation

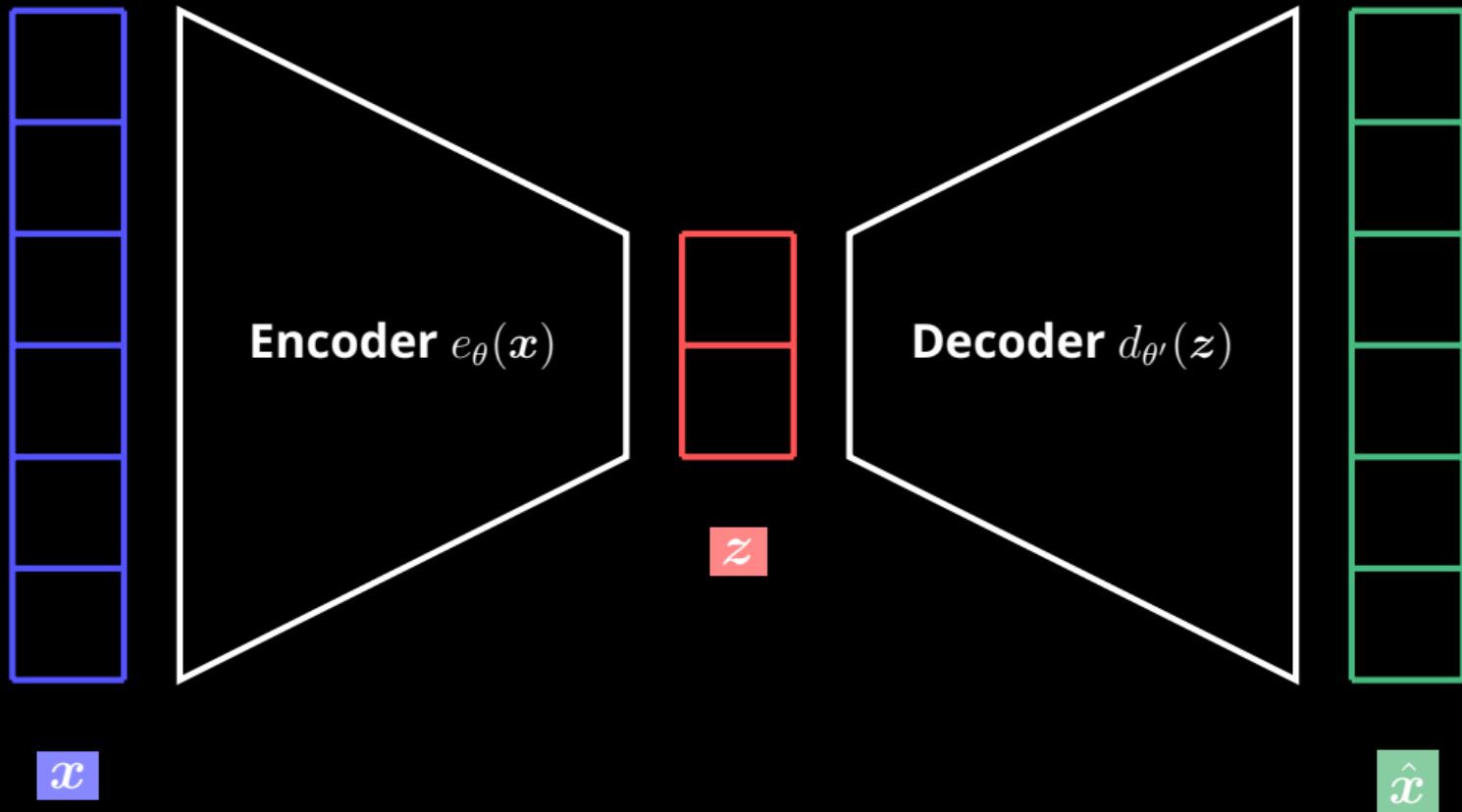


Dimensionality reduction



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  $\Lambda = \Sigma^2$ .

# Eigenfaces

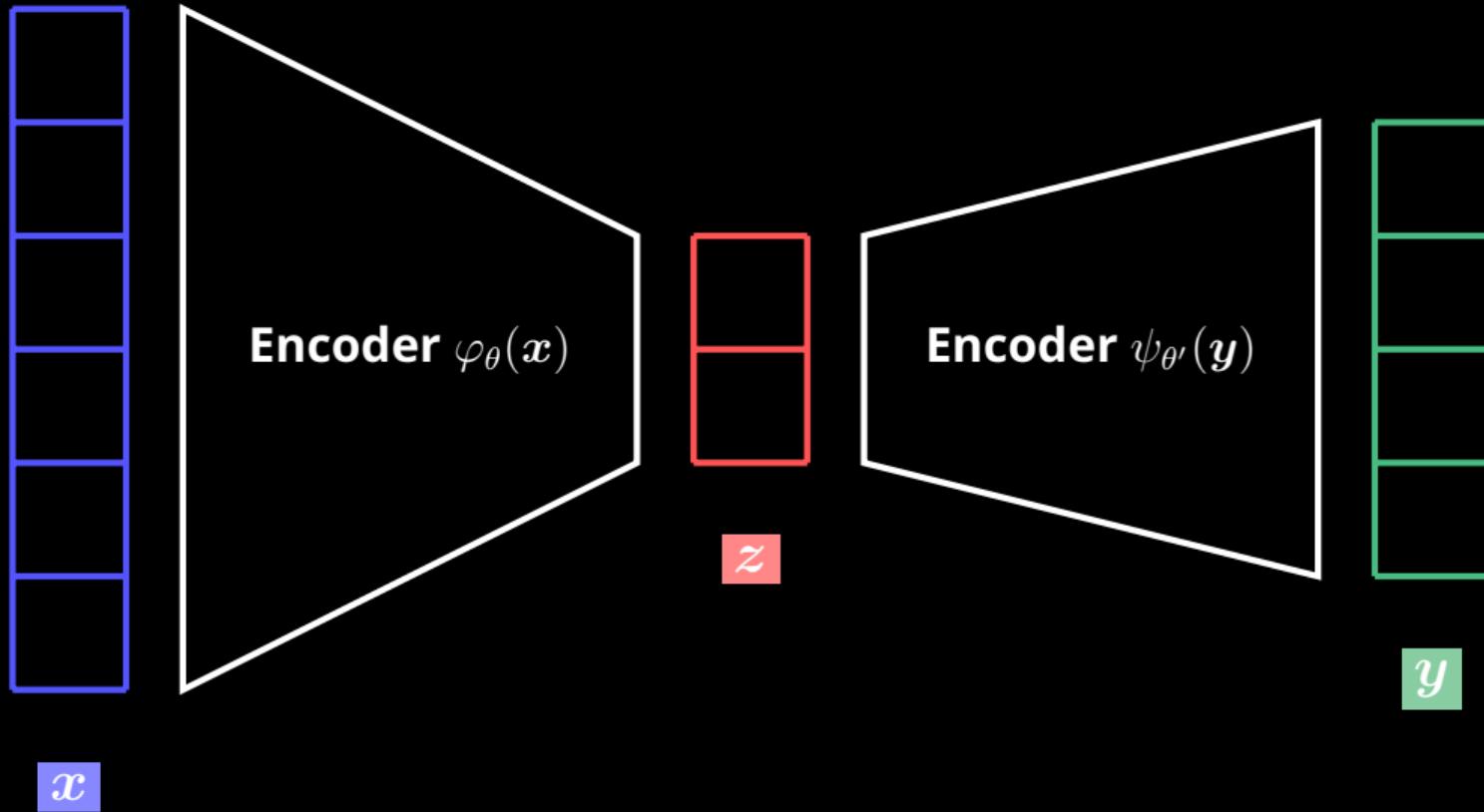
# Shear-driven cavity POD modes

Add POD modes

# Shear-driven cavity POD modes

Add phase portraits

Add cylinder flow and pressure coefficient



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

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

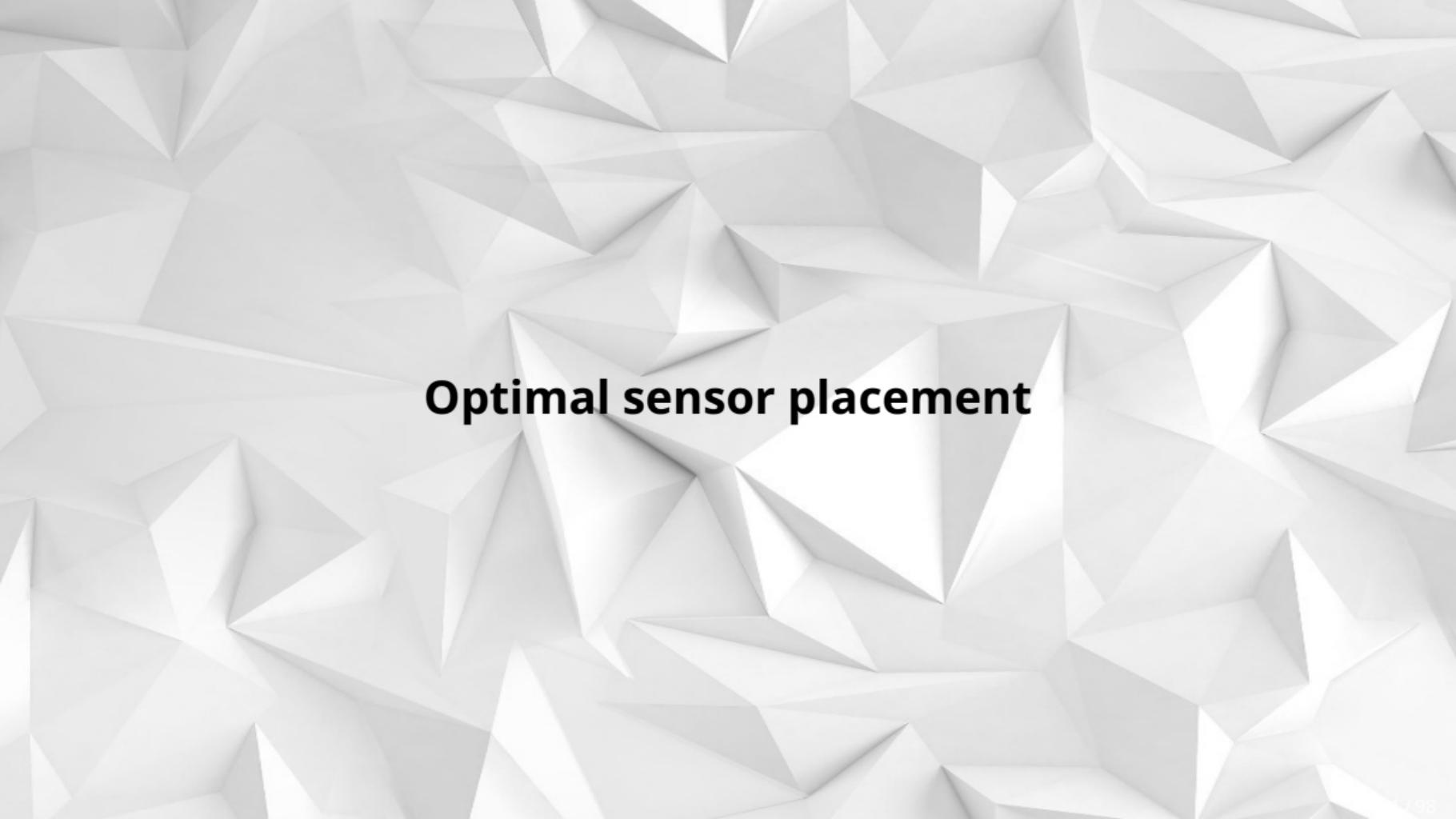
$$\begin{aligned} & \underset{\boldsymbol{P}, \boldsymbol{Q}}{\text{minimize}} && \|\boldsymbol{P}^T \boldsymbol{Y} - \boldsymbol{Q}^T \boldsymbol{X}\|_F^2 \\ & \text{subject to} && \boldsymbol{P}^T \boldsymbol{C}_{yy} \boldsymbol{P} = \boldsymbol{Q}^T \boldsymbol{C}_{xx} \boldsymbol{Q} = \boldsymbol{I}_r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

## Canonical Correlation Analysis

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

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

Add cylinder flow and pressure coefficient

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 vary in size and orientation, 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.

**Optimal sensor placement**



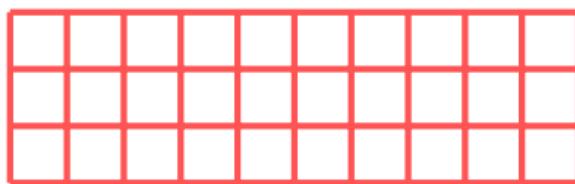
$$\mathbf{y} = \mathbf{C} \mathbf{x}$$

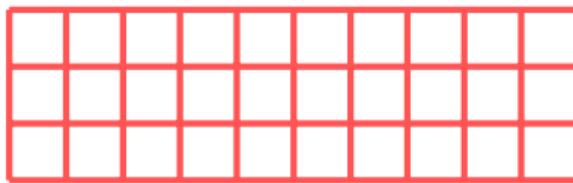
Measurement operator

Full state

Observations

The diagram illustrates a linear relationship between observations  $\mathbf{y}$ , a measurement operator  $\mathbf{C}$ , and a full state  $\mathbf{x}$ . The equation  $\mathbf{y} = \mathbf{C} \mathbf{x}$  is shown in the center. A red bracket labeled "Measurement operator" points to the matrix  $\mathbf{C}$ . A green bracket labeled "Full state" points to the matrix  $\mathbf{x}$ . A blue bracket labeled "Observations" points to the vector  $\mathbf{y}$ .

$y$  $C$  $x$  $\sim$ 

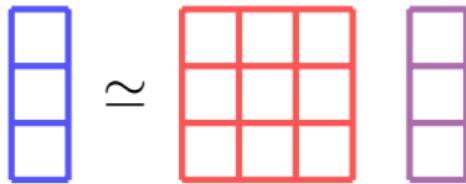
$y$  $C$  $U$  $z$  $\approx$ 

A large curly brace under the red and green rectangles, labeled  $\Theta$ .

*y*

$\Theta$

*z*



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

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

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

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

QR sensor placement algorithm

# Extended Yale B Face dataset

# Shear-driven cavity flow

# **State estimation and low-rank sensing**

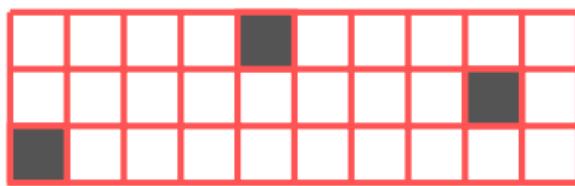
$$\mathbf{y} = \mathbf{C} \mathbf{x}$$

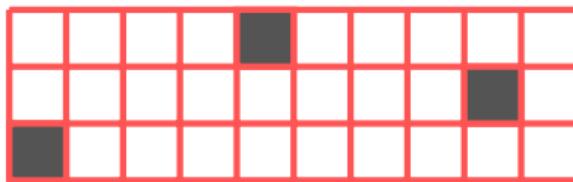
Measurement operator

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 components: the measurement operator  $\mathbf{C}$ , the full state  $\mathbf{x}$ , and the resulting observations  $\mathbf{y}$ . Red arrows indicate the flow from the measurement operator to the product term, and a green arrow indicates the flow from the full state to the same term. A blue double-headed vertical arrow connects the terms  $\mathbf{y}$  and  $\mathbf{C} \mathbf{x}$ , representing the relationship between the observed data and the measured state.

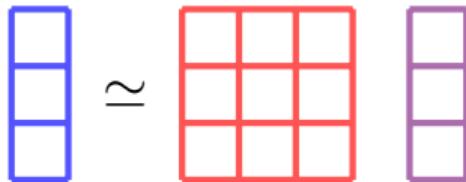
$y$  $C$  $x$  $\sim$ 

$y$  $C$  $U$  $z$  $\approx$  $\Theta$

*y*

$\Theta$

*z*



## 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

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

Add figures from SIAM paper

# **Reduced-order modeling**



$$\begin{array}{c} \text{Natural dynamics} \\ \hline \frac{dx}{dt} = \boxed{A}x + \boxed{B}u \\ y = \boxed{C}x + \boxed{D}u \\ \text{Measurements} \quad \quad \quad \text{Feedthrough} \end{array}$$

## Controlability Gramian

$$W_{\mathcal{C}} = \int_0^{\infty} e^{\tau A} B B^* e^{\tau A^*} d\tau$$

## Observability Gramian

$$W_{\mathcal{O}} = \int_0^{\infty} e^{\tau A^*} C^* C e^{\tau A} d\tau$$

## Cross Gramian

$$W_{\mathcal{X}} = \int_0^{\infty} e^{\tau A} B C e^{\tau A} d\tau$$

$$\begin{aligned} & \underset{\boldsymbol{U}, \boldsymbol{V}}{\text{maximize}} && \text{Tr} (\boldsymbol{U}^* \boldsymbol{W}_{\mathcal{X}} \boldsymbol{V}) \\ & \text{subject to} && \boldsymbol{U}^* \boldsymbol{U} = \boldsymbol{V}^* \boldsymbol{V} = \boldsymbol{I}_r \end{aligned}$$

## Sylvester equation<sup>1</sup>

$$AW_{\mathcal{X}} + W_{\mathcal{X}}A = -BC$$

---

<sup>1</sup>Its resolution is tractable only for low-dimensional systems.

## **Balanced Proper Orthogonal Decomposition**

---

1. For each actuator, compute the corresponding impulse response

$$\mathbf{X}_i = [\mathbf{B}_i \quad e^{\Delta t \mathbf{A}} \mathbf{B}_i \quad e^{2\Delta t \mathbf{A}} \mathbf{B}_i \quad \cdots \quad e^{n\Delta t \mathbf{A}} \mathbf{B}_i]$$

and assemble the data matrix  $\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdots \quad \mathbf{X}_p]$ .

## Balanced Proper Orthogonal Decomposition

2. For each sensor, compute the corresponding **adjoint** impulse response

$$\mathbf{Y}_i = [C_i^* \quad e^{\Delta t \mathbf{A}^*} C_i^* \quad e^{2\Delta t \mathbf{A}^*} C_i^* \quad \cdots \quad e^{n\Delta t \mathbf{A}^*} C_i^*]$$

and assemble the data matrix  $\mathbf{Y} = [\mathbf{Y}_1 \quad \mathbf{Y}_2 \quad \cdots \quad \mathbf{Y}_q]$ .

## **Balanced Proper Orthogonal Decomposition**

3. Approximate the cross Gramian  $W_{\mathcal{X}}$  as

$$W_{\mathcal{X}} \simeq \mathbf{X} \mathbf{Y}^T.$$

where  $\mathbf{X}$  and  $\mathbf{Y}$  are the data matrices obtained in the previous steps.

## Petrov-Galerkin projection

$$\frac{d\hat{x}}{dt} = \hat{A}\hat{x} + \hat{B}u$$

$$\hat{y} = \hat{C}\hat{x} + \hat{D}u$$

Example for the Ginzburg Landau equation

Example for the shear-driven cavity

# **System identification**



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

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

Observability

Controlability

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

Markov parameters of the system

Schematic mass  
spring damper      Impulse response

# EigenRealization Algorithm

$$\mathbf{y} = [y_1 \ y_2 \ y_3 \ y_4 \ y_5 \ y_6 \ y_7 \ y_8 \ y_9 \ y_{10}]$$

# EigenRealization Algorithm

$$\mathbf{H}_1 = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 & y_5 \\ y_2 & y_3 & y_4 & y_5 & y_6 \\ y_3 & y_4 & y_5 & y_6 & y_7 \\ y_4 & y_5 & y_6 & y_7 & y_8 \\ y_5 & y_6 & y_7 & y_8 & y_9 \end{bmatrix}$$

# 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

$$\mathbf{H}_2 = \begin{bmatrix} y_2 & y_3 & y_4 & y_5 & y_6 \\ y_3 & y_4 & y_5 & y_6 & y_7 \\ y_4 & y_5 & y_6 & y_7 & y_8 \\ y_5 & y_6 & y_7 & y_8 & y_9 \\ y_6 & y_7 & y_8 & y_9 & y_{10} \end{bmatrix}$$

# 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

$$\mathbf{A} = \mathcal{O}^\dagger \mathbf{H}_2 \mathcal{C}^\dagger$$

Actuators

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

Measurements

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

Feedthrough

$$\mathbf{D} = \mathbf{y}_0$$

Mass spring damper example

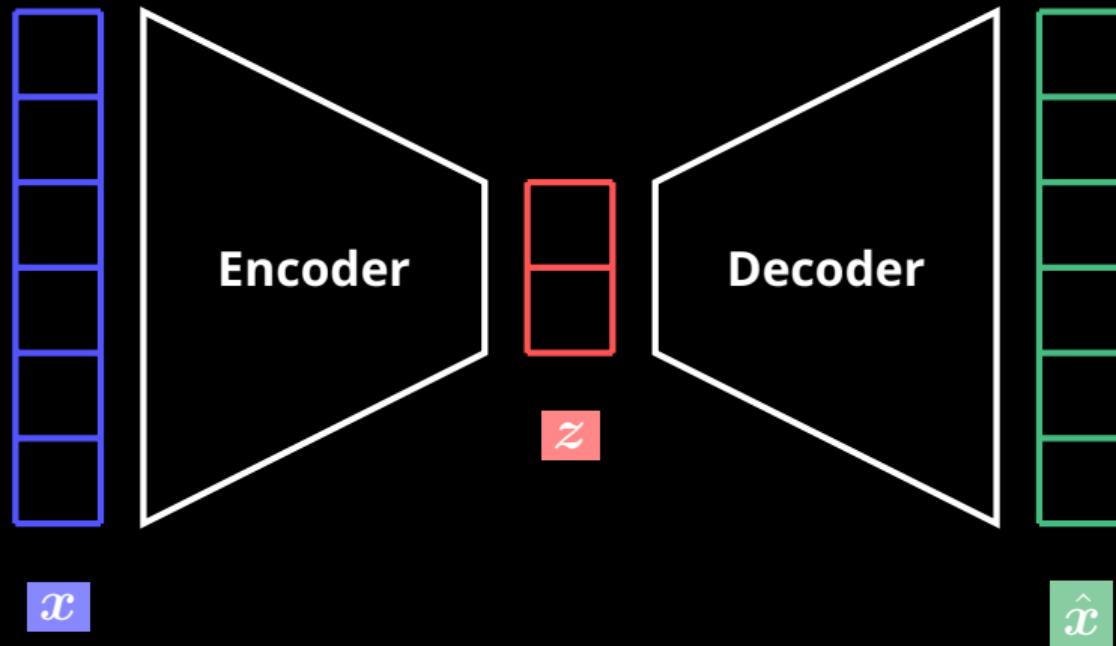
Cylinder flow example

# **Conclusion**

Since the work of G. Golub *et al.* in the late 1960s, SVD plays a pivotal role in numerical linear algebra.

It is widely used in control theory to characterize various properties of input-output linear dynamical systems or for system identification purposes.

It also lays the foundation for the mathematical description of *quantum entanglement* in particle physics.



Many (linear) dimensionality reduction techniques in machine learning can actually be re-interpreted as variations around the theme of SVD.