

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.



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

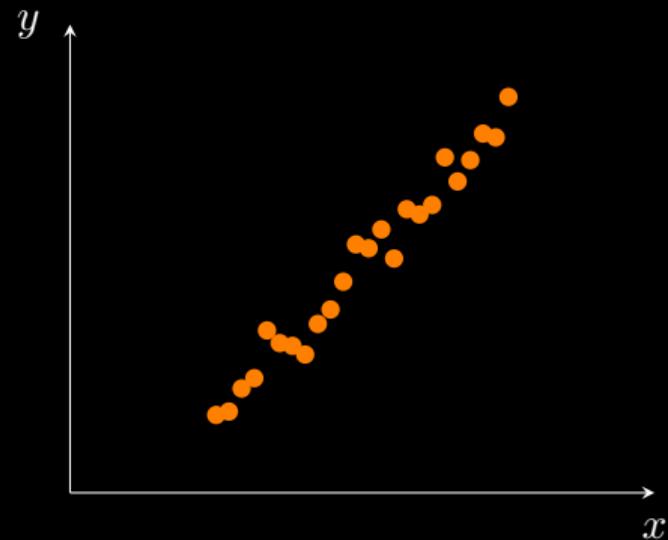
Relation to spectral decomposition

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

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.

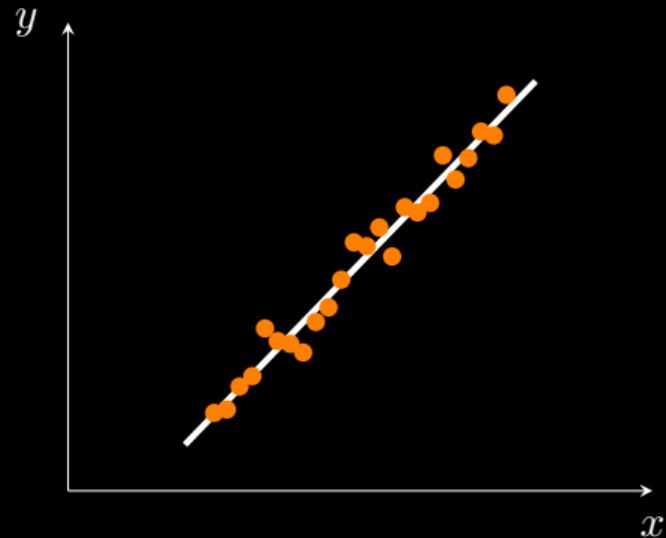
Ordinary least-squares

$$y = ax + b$$



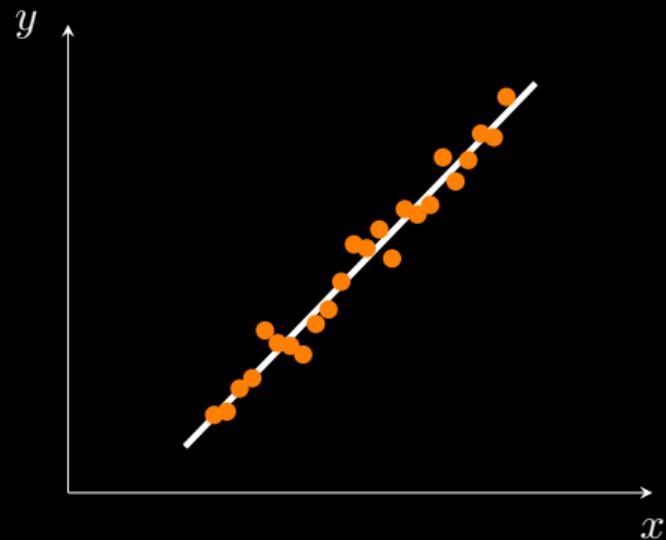
Ordinary least-squares

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



Ordinary least-squares

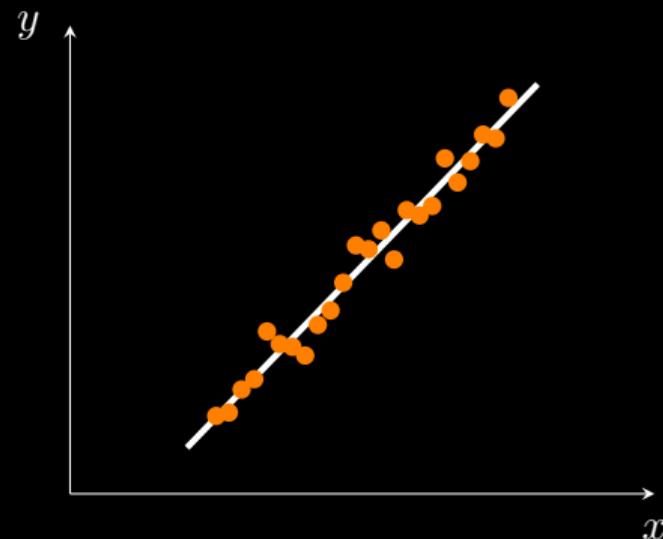
$$\underset{x}{\text{minimize}} \quad \|Ax - b\|_2^2$$



Ordinary least-squares

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

Moore-Penrose
pseudoinverse



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

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

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

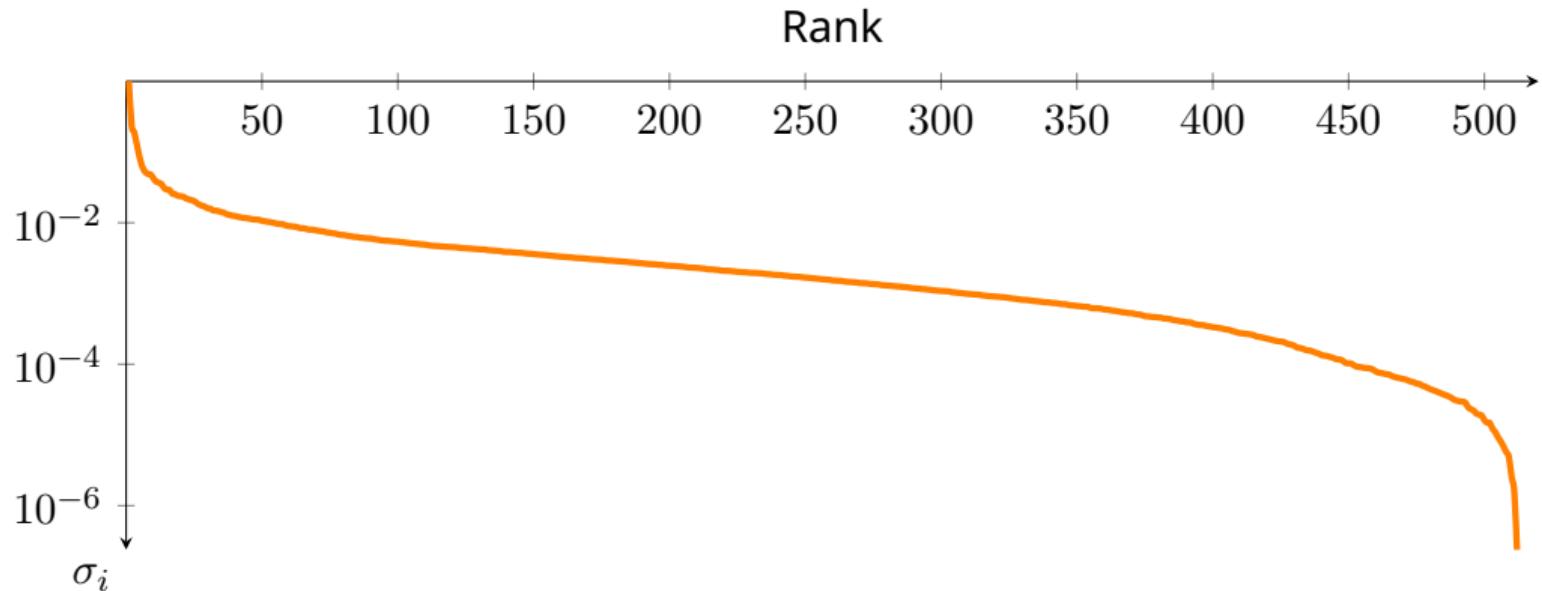
Low-rank approximation

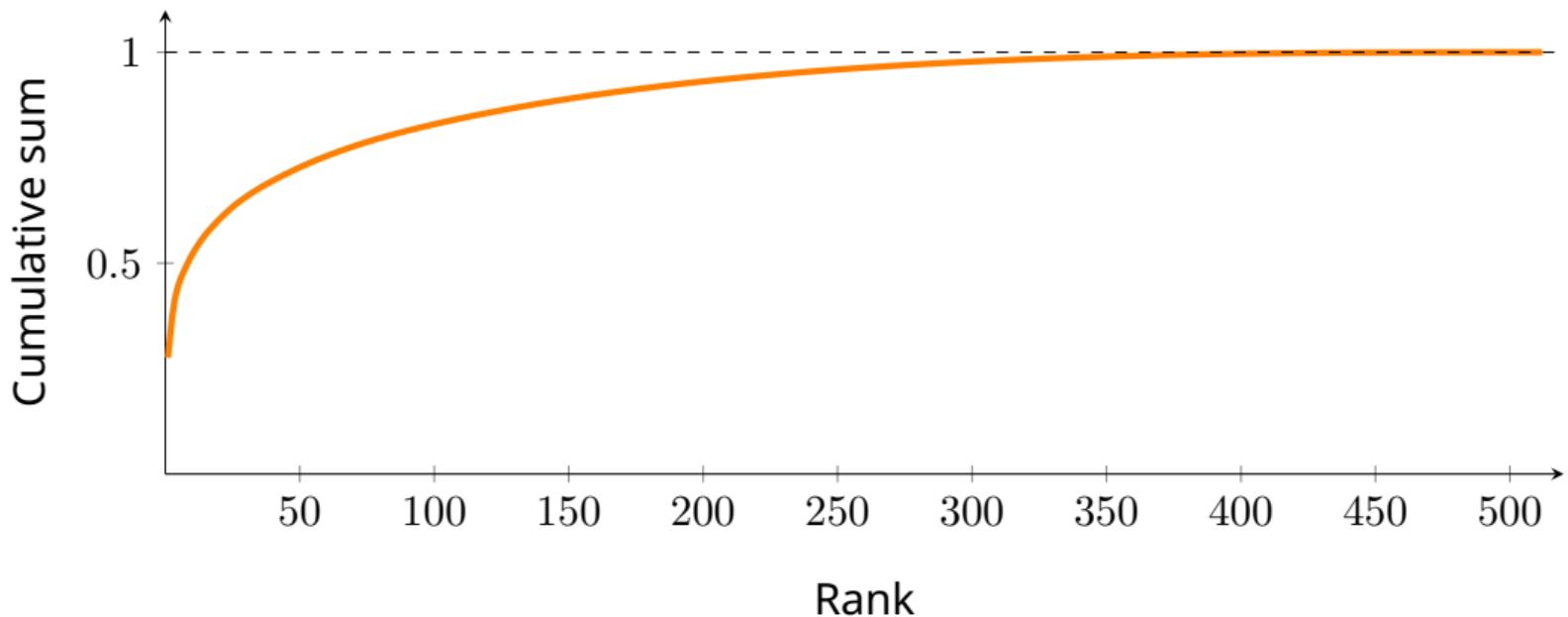


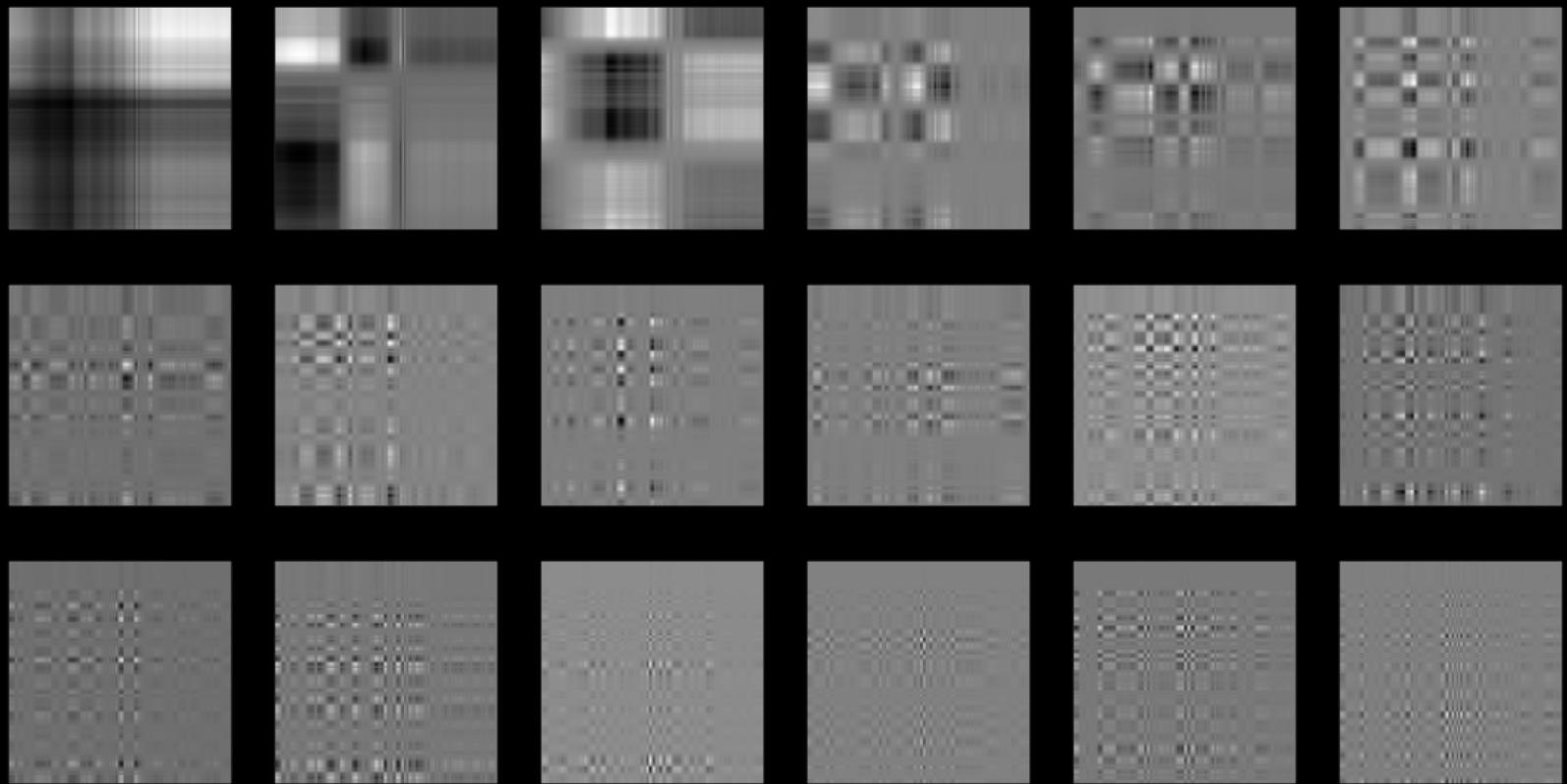
How to compress this image ?

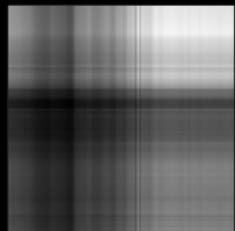
Low-rank approximation

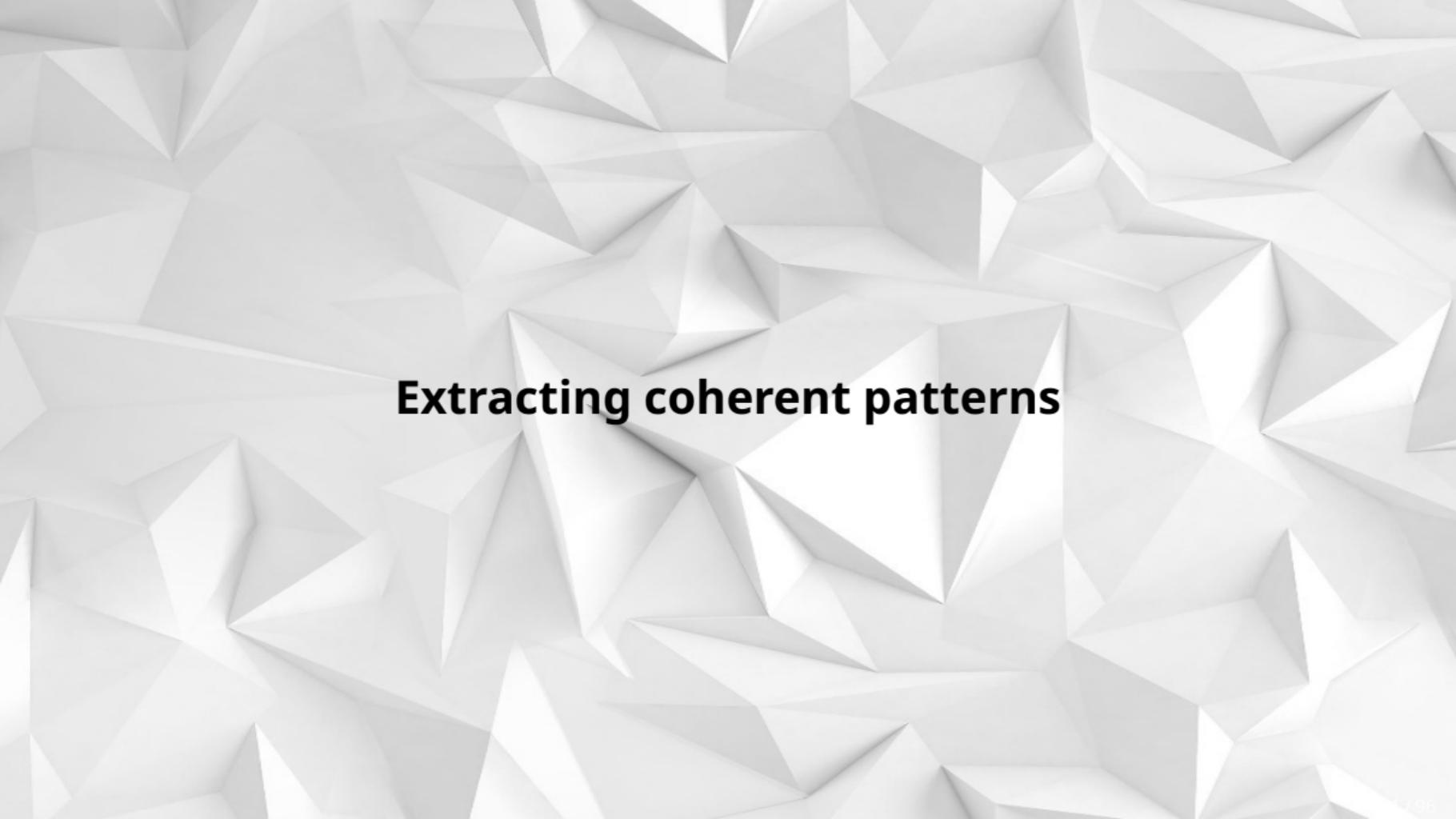
$$\begin{aligned} & \underset{\mathbf{X}}{\text{minimize}} && \|\mathbf{A} - \mathbf{X}\|_F^2 \\ & \text{subject to} && \text{rank } \mathbf{X} = r \end{aligned}$$









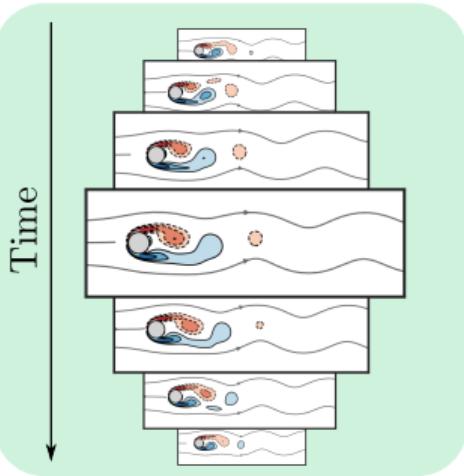
The background of the slide features a complex, abstract pattern of white polygons, likely triangles, arranged in a way that creates a sense of depth and texture. The polygons vary in size and orientation, with some appearing as sharp peaks and others as recessed valleys. The lighting is soft, highlighting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

Extracting coherent patterns

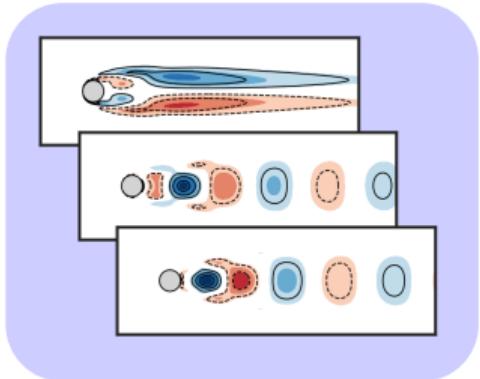
Add Yale B faces

Add video of the cavity

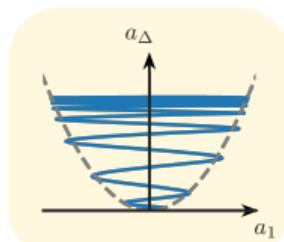
Navier-Stokes simulation

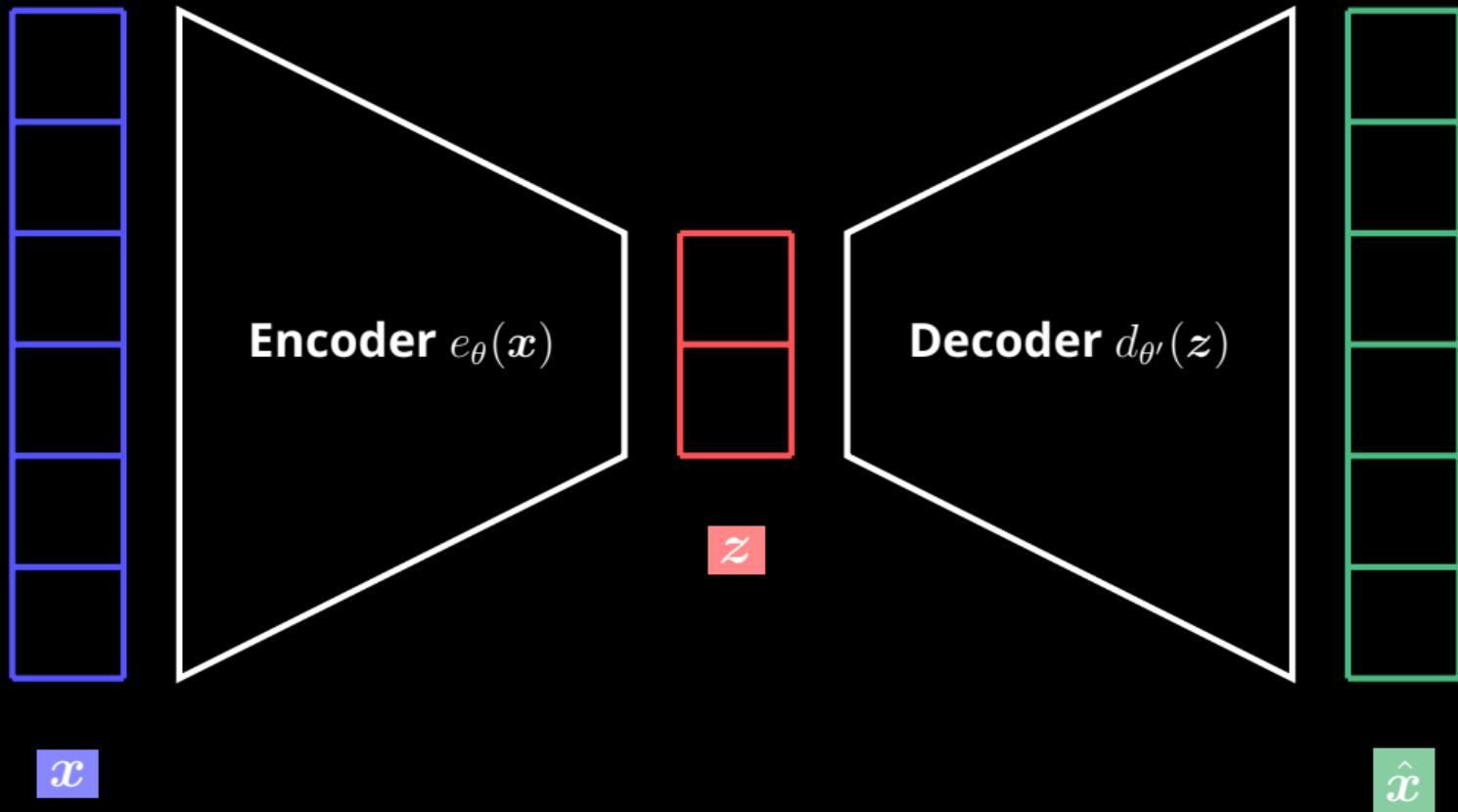


Dimensionality reduction



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

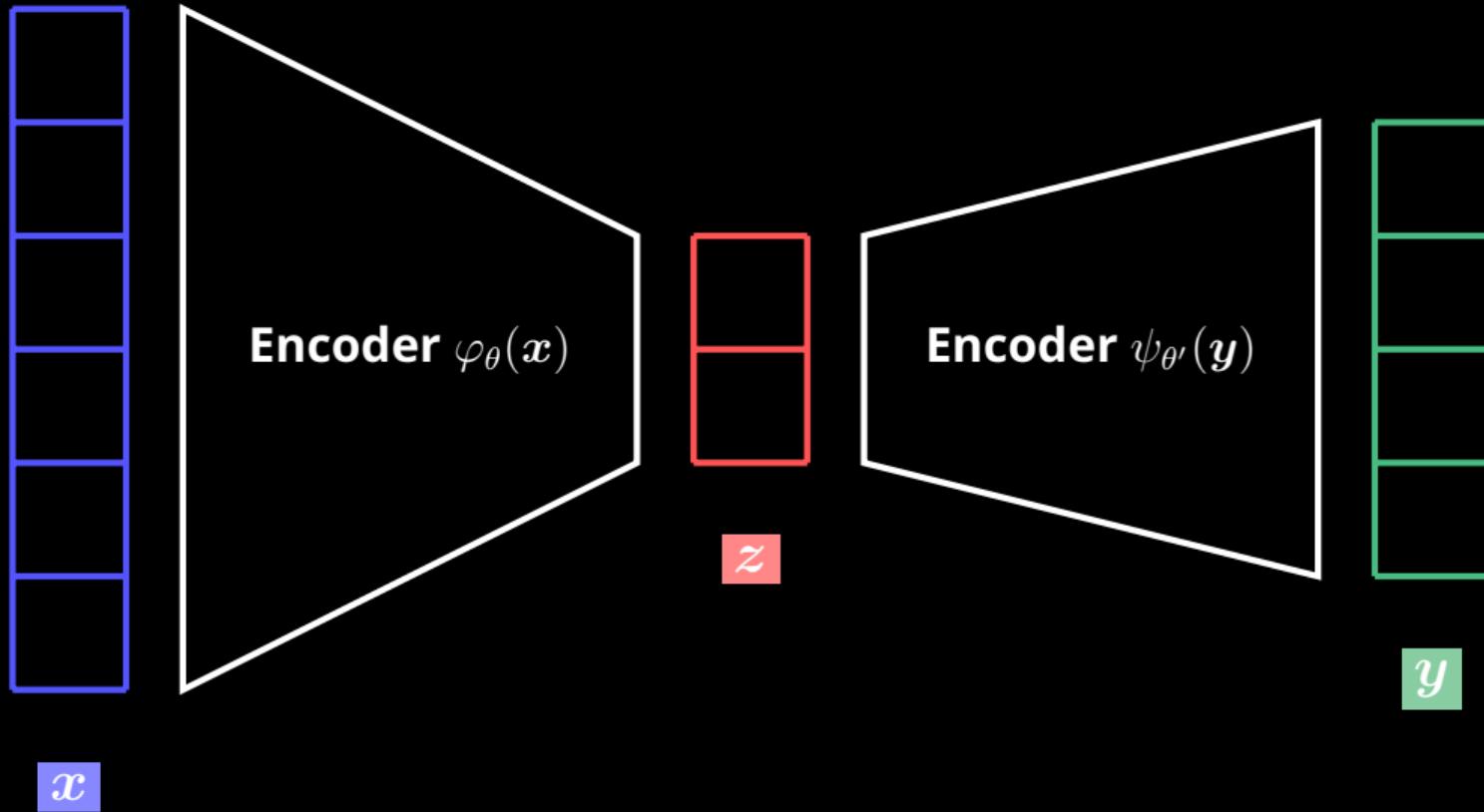
Shear-driven cavity POD modes

Add POD modes

Shear-driven cavity POD modes

Add phase portraits

Add cylinder flow and pressure coefficient



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

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

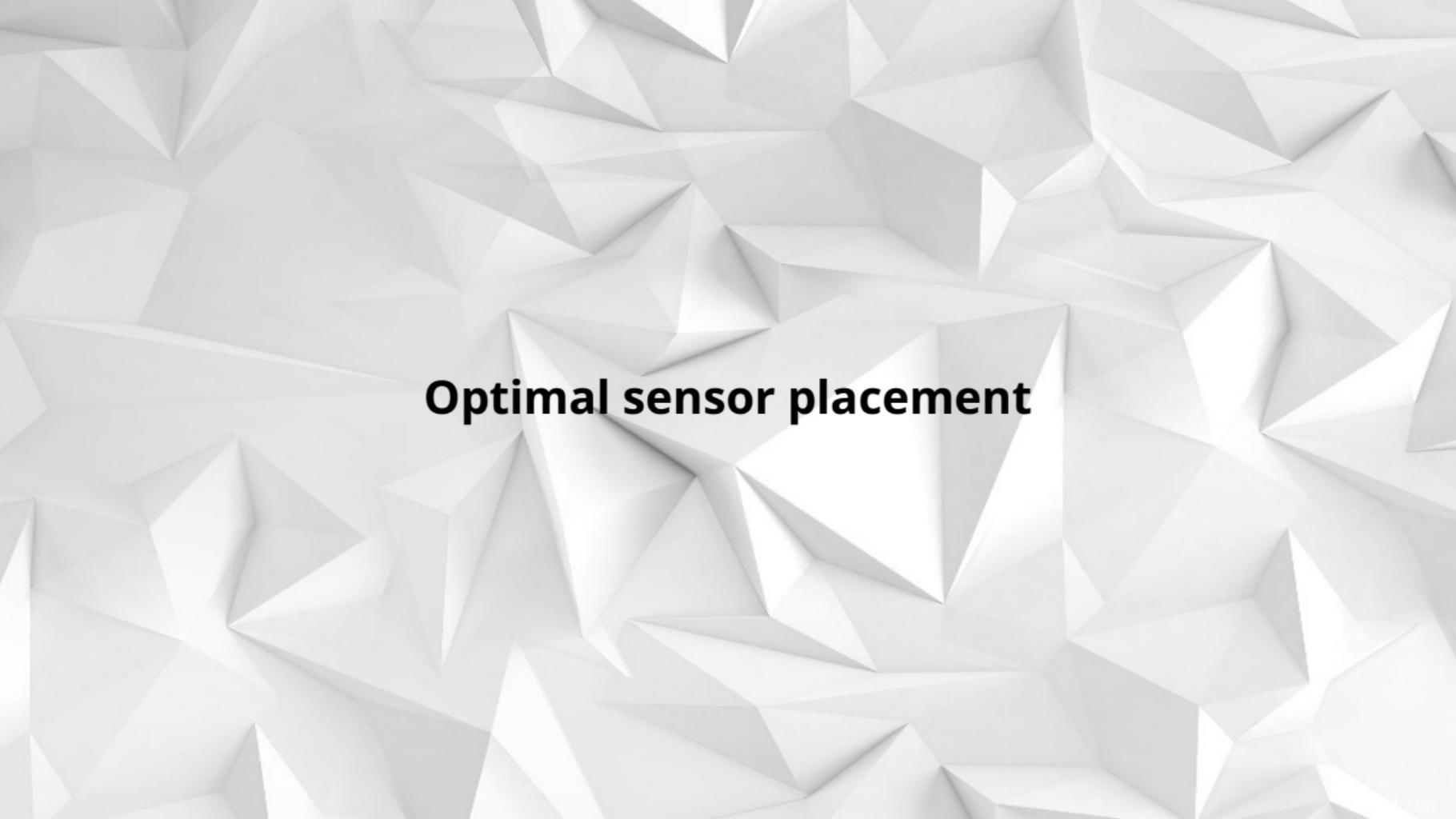
$$\begin{aligned} & \underset{\boldsymbol{P}, \boldsymbol{Q}}{\text{minimize}} && \|\boldsymbol{P}^T \mathbf{Y} - \boldsymbol{Q}^T \mathbf{X}\|_F^2 \\ & \text{subject to} && \boldsymbol{P}^T \boldsymbol{C}_{yy} \boldsymbol{P} = \boldsymbol{Q}^T \boldsymbol{C}_{xx} \boldsymbol{Q} = \boldsymbol{I}_r \end{aligned}$$

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

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, highlighting the edges of the polygons and casting subtle shadows, which gives the background a three-dimensional appearance.

Optimal sensor placement

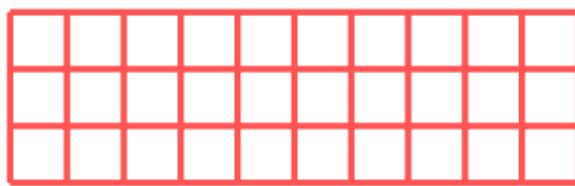
$$\mathbf{y} = \mathbf{C} \mathbf{x}$$

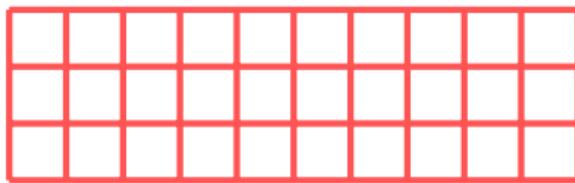
Measurement operator

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 annotations: 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 

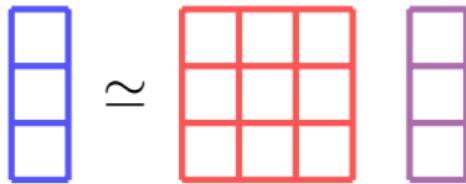
y C U z  \approx 

A large curly brace underneath the red and green rectangles, labeled with the symbol Θ .

y

Θ

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

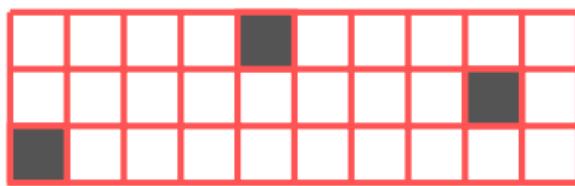
$$\mathbf{y} = \mathbf{C} \mathbf{x}$$

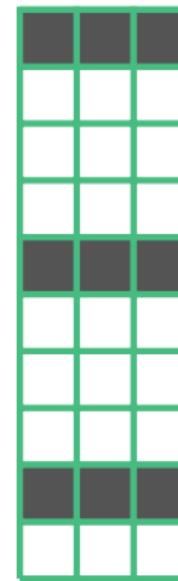
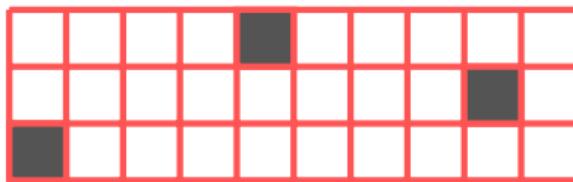
Measurement operator

Full state

Observations

The diagram illustrates a linear 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} .

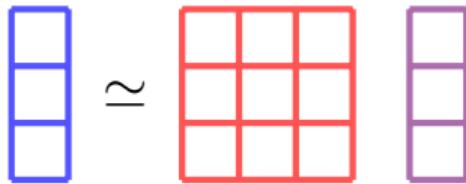
y C x  \sim 

y C U z  \approx  Θ

y

Θ

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

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

Balancing transform / Balanced truncation

$$\begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{P} \\ \mathbf{Q} \end{bmatrix} \Sigma = \begin{bmatrix} \mathbf{W}_{\mathcal{O}} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{\mathcal{C}} \end{bmatrix} \begin{bmatrix} \mathbf{P} \\ \mathbf{Q} \end{bmatrix}$$

BT discards modes which are highly controllable but poorly observable, and vice versa. The transformed system is **balanced**. Its Gramians are given by $\hat{\mathbf{W}}_{\mathcal{O}} = \hat{\mathbf{W}}_{\mathcal{C}} = \Sigma$ where Σ are the *Hankel singular values* of the system.

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

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

and assemble the data matrix $Y = [Y_1 \quad Y_2 \quad \cdots \quad Y_q]$.

Balanced Proper Orthogonal Decomposition

3. Compute the SVD of $\mathbf{Y}^T \mathbf{X}$

$$\mathbf{Y}^T \mathbf{X} = \mathbf{U} \Sigma \mathbf{V}^T$$

where Σ are the *Hankel singular values* of the system.

Balanced Proper Orthogonal Decomposition

4. Compute the first r columns and rows of the balancing transform as

$$P = X V \Sigma^{-\frac{1}{2}} \quad \text{and} \quad Q = Y U \Sigma^{-\frac{1}{2}}$$

and proceed with the construction of the reduced-order model.

Petrov-Galerkin projection

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

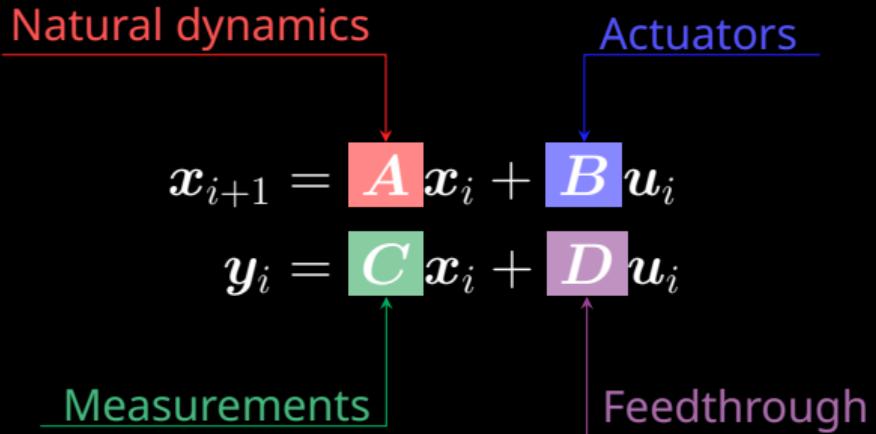
$$\hat{y} = \hat{C}\hat{x} + \hat{D}u$$

Petrov-Galerkin projection

$$\left(\begin{array}{c|c} Q^T AP & Q^T B \\ \hline CP & D \end{array} \right)$$

Example for the shear-driven cavity

System identification



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

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

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

Actuators

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

Measurements

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

Feedthrough

$$D = y_0$$

Mass spring damper example

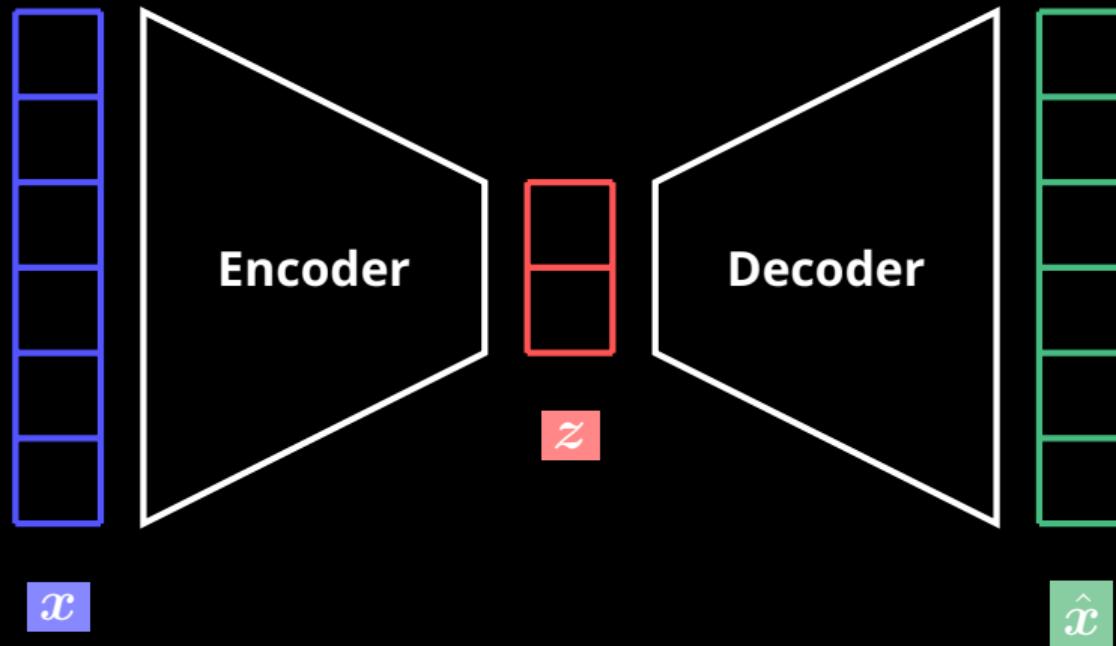
Cylinder flow example

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.