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

Jean-Christophe Loiseau

May, 13^{th} 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

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

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

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

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

Diagonal matrix

Relation to spectral decomposition

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

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.

Schematic of SVD

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

$$\underset{\boldsymbol{x}}{\text{minimize}} \ \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2$$

Ordinary least-squares

$$x = \boxed{\left(A^T A\right)^{-1} A^T} b$$

Moore-Penrose
pseudoinverse

$$\boldsymbol{A}^{\dagger} = \left(\boldsymbol{A}^{T}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{T}$$

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

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

$$\sigma_i^{-1} = \begin{cases} \dfrac{1}{\sigma_i} & \text{if} \quad \sigma_i > \varepsilon \\ 0 & \text{otherwise.} \end{cases}$$

- `np.linalg.lstsq(A, b)`
- `A\b`

Low-rank approximation

How to compress this image ?

Low-rank approximation

$$\begin{array}{ll} \underset{\boldsymbol{X}}{\text{minimize}} & \|\boldsymbol{A} - \boldsymbol{X}\|_F^2 \\ \text{subject to} & \operatorname{rank} \boldsymbol{X} = r \end{array}$$

Low-rank approximation

Singular value distribution

Low-rank approximation

Reconstructed image for different rank

Low-rank approximation

Canonical images

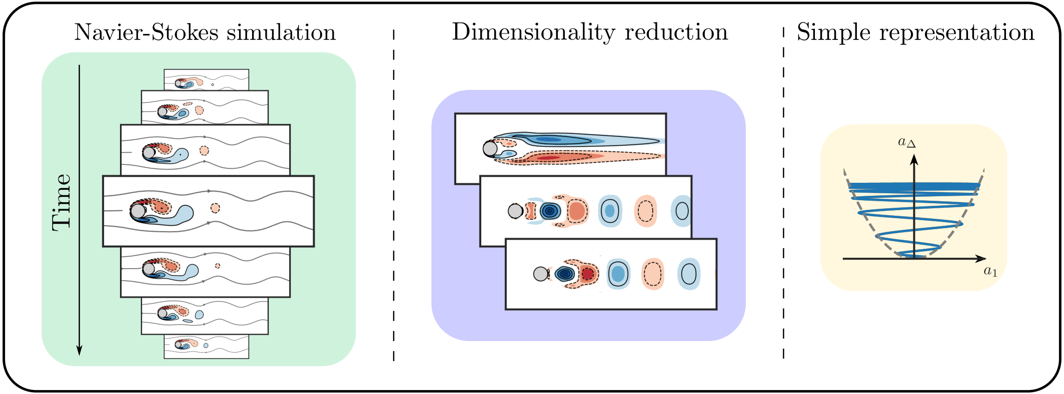
Low-rank approximation

Image with more fine scale details

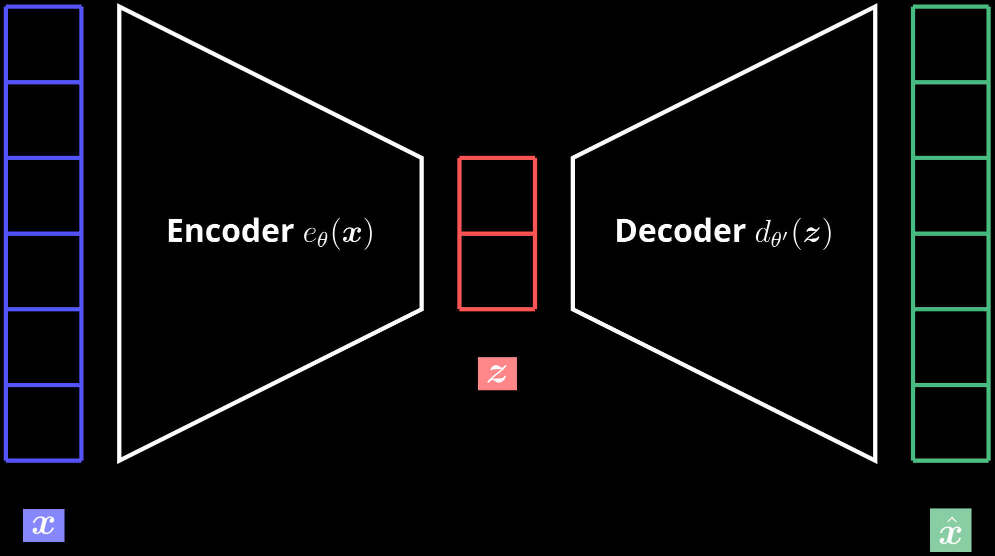
Extracting coherent patterns

Add Yale B faces

Add video of the cavity



Navier-Stokes simulation
Dimensionality reduction
Simple representation
Time
a_Δ
a_1



Encoder $e_\theta(x)$
Decoder $d_{\theta'}(z)$
z
x
$\hat{x}$

Estimate

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

Ground truth

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

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

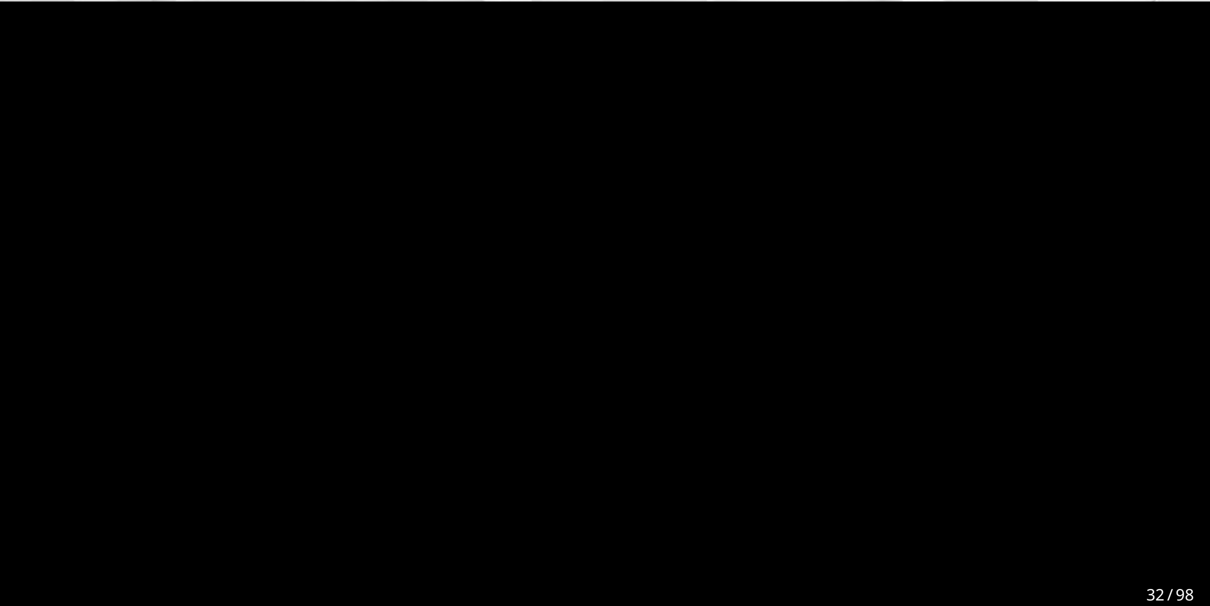
$$\begin{aligned} \underset{\boldsymbol{P}}{\text{minimize}} \quad & \|\boldsymbol{X} - \boldsymbol{P}\boldsymbol{P}^T\boldsymbol{X}\|_F^2 \\ \text{subject to} \quad & \boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{I}_r \end{aligned}$$

Proper Orthogonal Decomposition

$$\boldsymbol{P}\boldsymbol{\Lambda} = \boldsymbol{C}_{\boldsymbol{xx}}\boldsymbol{P}$$

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

Eigenfaces



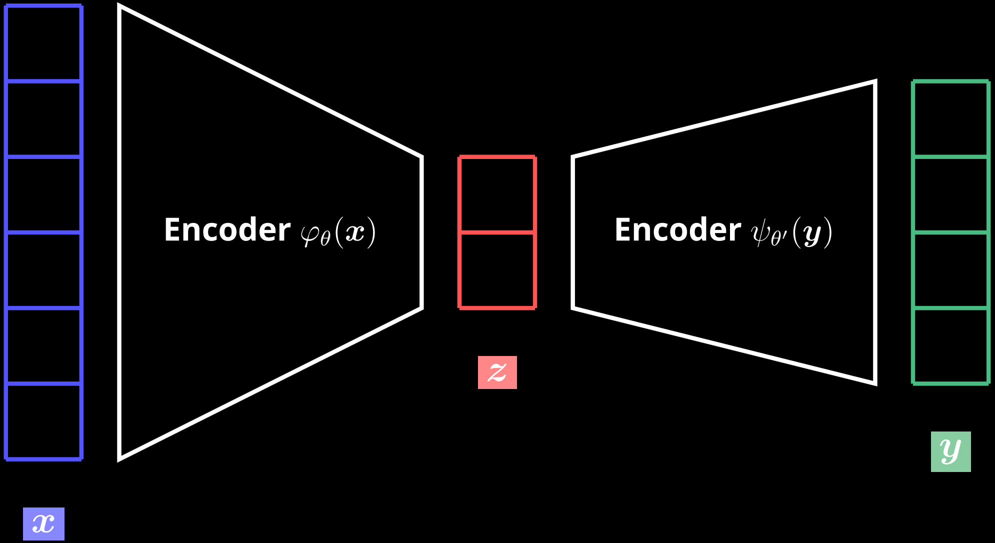
Shear-driven cavity POD modes

Add POD modes

Shear-driven cavity POD modes

Add phase portraits

Add cylinder flow and pressure coefficient


Encoder $\varphi_\theta(x)$
Encoder $\psi_{\theta'}(y)$
z
x
y

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

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

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

Canonical Correlation Analysis

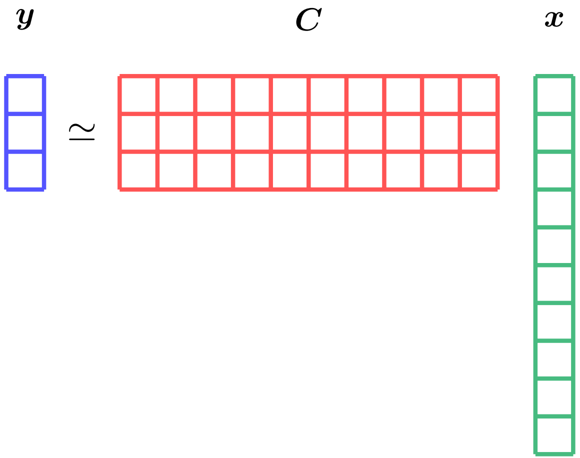
$$\begin{bmatrix} \boldsymbol{C}_{yy} & 0 \\ 0 & \boldsymbol{C}_{xx} \end{bmatrix} \begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{Q} \end{bmatrix} \boldsymbol{\Sigma} = \begin{bmatrix} 0 & \boldsymbol{C}_{yx} \\ \boldsymbol{C}_{xy} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{Q} \end{bmatrix}$$

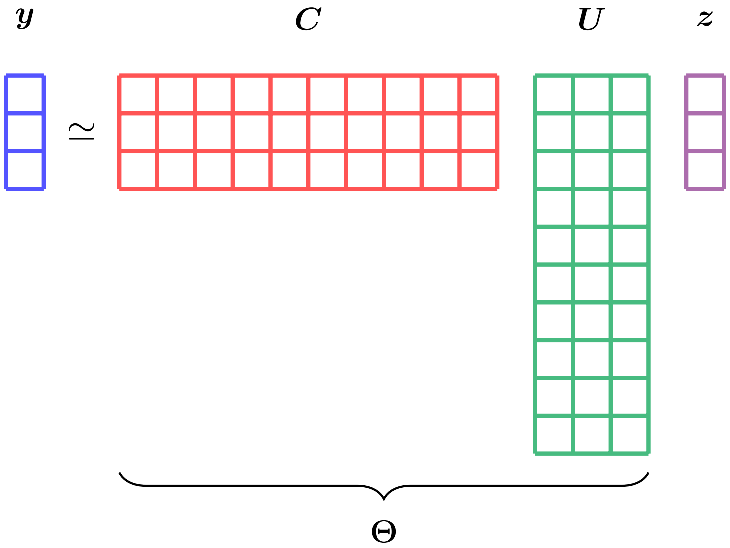
CCA relies on a *generalized eigenproblem*. $\boldsymbol{P}$ and $\boldsymbol{Q}$ describe the encoders such that the latent representations $\boldsymbol{z} = \boldsymbol{Q}^T \boldsymbol{x}$ and $\boldsymbol{z}' = \boldsymbol{P}^T \boldsymbol{Y}$ are as similar as possible. It is closely related to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

Add cylinder flow and pressure coefficient

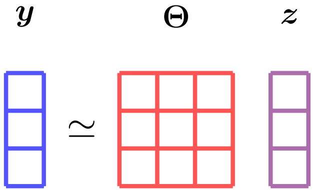
Optimal sensor placement

$$\underbrace{y}_{\text{Observations}} = \overbrace{C}^{\text{Measurement operator}}\underbrace{x}_{\text{Full state}}$$


$\boldsymbol{y}$
$\simeq$
$\boldsymbol{C}$
$\boldsymbol{x}$



$\boldsymbol{y}$
$\simeq$
$\boldsymbol{C}$
$\boldsymbol{U}$
$\boldsymbol{z}$
$\boldsymbol{\Theta}$


$\boldsymbol{y}$
$\boldsymbol{\Theta}$
$\boldsymbol{z}$
$\simeq$

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

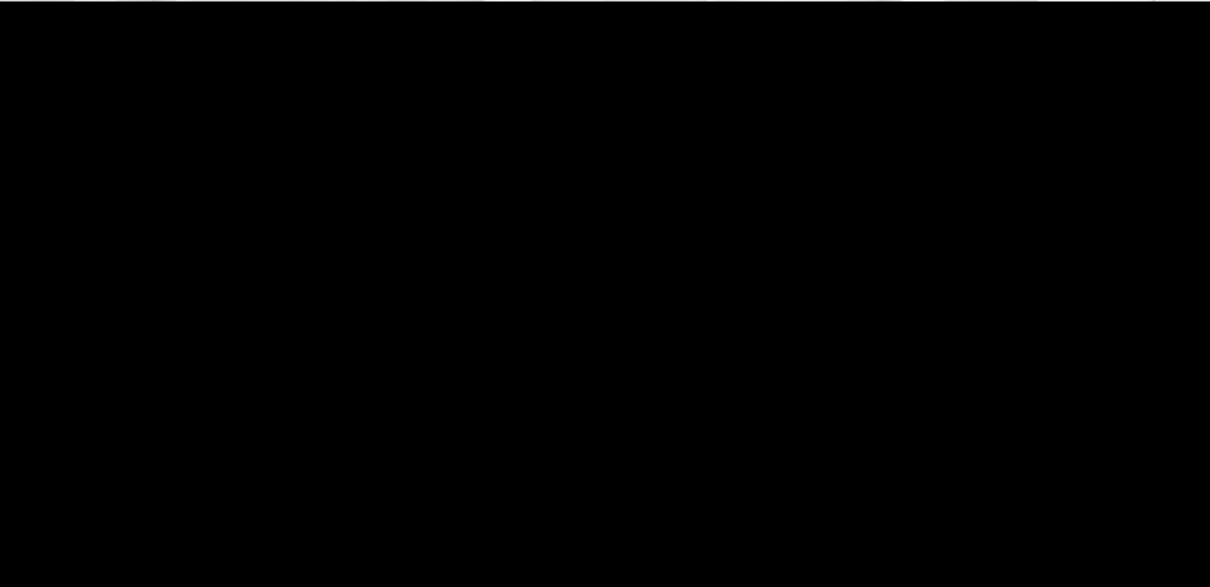
$$\boldsymbol{z} = \boldsymbol{\Theta}^{-1}\boldsymbol{y}$$

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

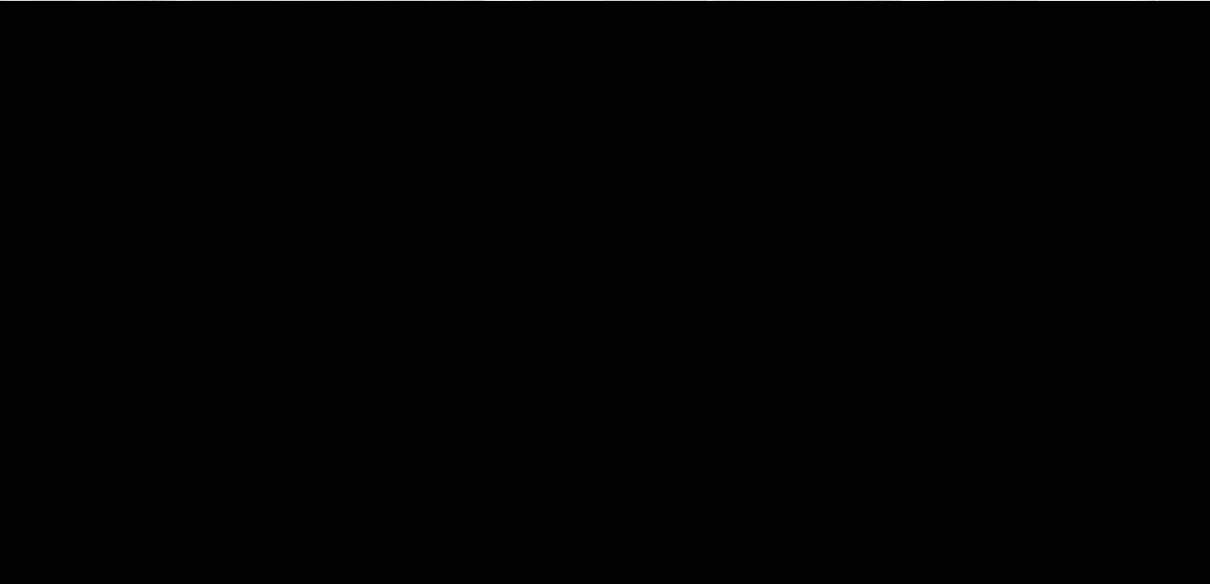
$$\begin{aligned} \underset{\boldsymbol{C}}{\text{maximize}} \quad & |\det(\boldsymbol{C}\boldsymbol{U})| \\ \text{subject to} \quad & \boldsymbol{C}_i \in \{\boldsymbol{e}_j\}_{j=1,n} \end{aligned}$$

QR sensor placement algorithm

Extended Yale B Face dataset



Shear-driven cavity flow



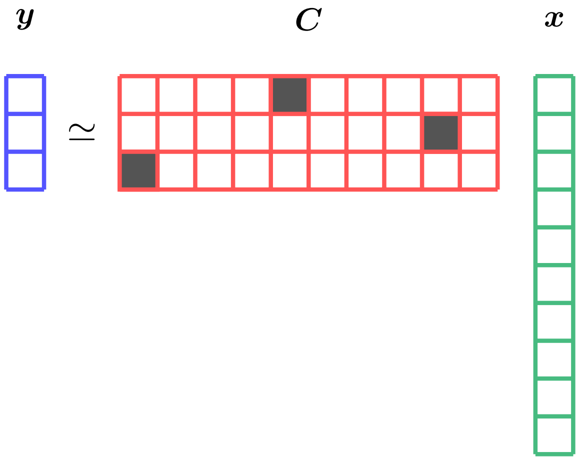
State estimation and low-rank sensing

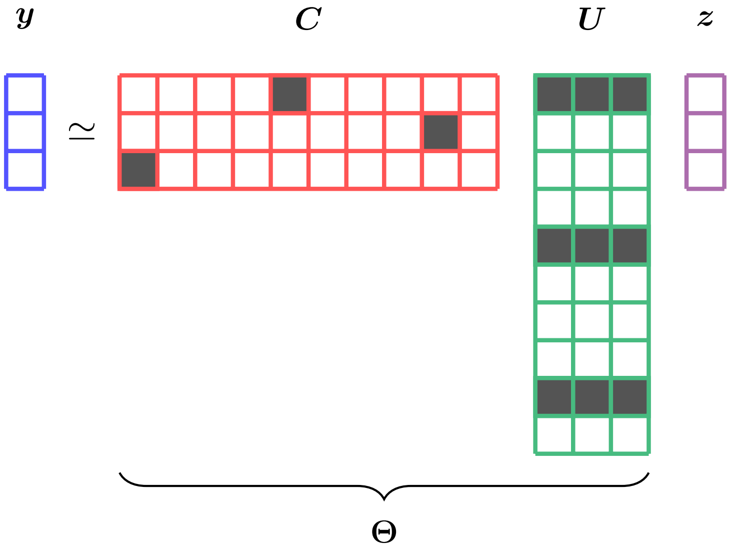
Measurement operator

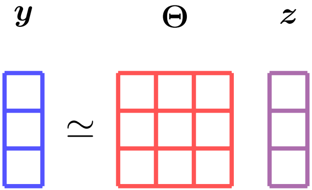
$$\boxed{y} = \boxed{C}\boxed{x}$$

Full state

Observations


$\boldsymbol{y}$
$\simeq$
$\boldsymbol{C}$
$\boldsymbol{x}$


$\boldsymbol{y}$
$\simeq$
$\boldsymbol{C}$
$\boldsymbol{U}$
$\boldsymbol{z}$
$\boldsymbol{\Theta}$


$\boldsymbol{y}$
$\simeq$
$\boldsymbol{\Theta}$
$\boldsymbol{z}$

Undertermined problem

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

Overdetermined problem

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

Regularized problem

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

Regularized and constrained problem

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

Add figures from SIAM paper

Reduced-order modeling

Natural dynamics

Actuators

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}$$
$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{u}$$

Measurements

Feedthrough

Controlability Gramian

$$\boldsymbol{W}_{\mathcal{C}} = \int_0^\infty e^{\tau \boldsymbol{A}} \boldsymbol{B} \boldsymbol{B}^* e^{\tau \boldsymbol{A}^*} \, \mathrm{d}\tau$$

Observability Gramian

$$\boldsymbol{W}_{\mathcal{O}} = \int_0^{\infty} e^{\tau \boldsymbol{A}^*} \boldsymbol{C}^* \boldsymbol{C} e^{\tau \boldsymbol{A}} \, \mathrm{d}\tau$$

Cross Gramian

$$\boldsymbol{W}_{\mathcal{X}} = \int_0^\infty e^{\tau \boldsymbol{A}} \boldsymbol{B} \boldsymbol{C} e^{\tau \boldsymbol{A}} \, \mathrm{d}\tau$$

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

Sylvester equation[1]

$$\boldsymbol{A}\boldsymbol{W}_{\mathcal{X}} + \boldsymbol{W}_{\mathcal{X}}\boldsymbol{A} = -\boldsymbol{B}\boldsymbol{C}$$

[1] Its resolution is tractable only for low-dimensional systems.

Balanced Proper Orthogonal Decomposition

1. For each actuator, compute the corresponding impulse reponse

$$\boldsymbol{X}_i = \begin{bmatrix} \boldsymbol{B}_i & e^{\Delta t \boldsymbol{A}} \boldsymbol{B}_i & e^{2\Delta t \boldsymbol{A}} \boldsymbol{B}_i & \cdots & e^{n\Delta t \boldsymbol{A}} \boldsymbol{B}_i \end{bmatrix}$$

   and assemble the data matrix $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_p \end{bmatrix}$.

Balanced Proper Orthogonal Decomposition

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

$$\boldsymbol{Y}_i = \begin{bmatrix} \boldsymbol{C}_i^* & e^{\Delta t \boldsymbol{A}^*} \boldsymbol{C}_i^* & e^{2\Delta t \boldsymbol{A}^*} \boldsymbol{C}_i^* & \cdots & e^{n\Delta t \boldsymbol{A}^*} \boldsymbol{C}_i^* \end{bmatrix}$$

   and assemble the data matrix $\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 & \boldsymbol{Y}_2 & \cdots & \boldsymbol{Y}_q \end{bmatrix}$.

Balanced Proper Orthogonal Decomposition

3. Approximate the cross Gramian $\boldsymbol{W}_{\mathcal{X}}$ as

$$\boldsymbol{W}_{\mathcal{X}} \simeq \boldsymbol{X}\boldsymbol{Y}^T.$$

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

Petrov-Galerkin projection

$$\frac{\mathrm{d}\hat{\boldsymbol{x}}}{\mathrm{d}t} = \hat{\boldsymbol{A}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{B}}\boldsymbol{u}$$
$$\hat{\boldsymbol{y}} = \hat{\boldsymbol{C}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{D}}\boldsymbol{u}$$

Example for the Ginzburg Landau equation

Example for the shear-driven cavity

System identification

Natural dynamics

Actuators

$$\boldsymbol{x}_{i+1} = \boxed{\boldsymbol{A}}\boldsymbol{x}_i + \boxed{\boldsymbol{B}}\boldsymbol{u}_i$$

$$\boldsymbol{y}_i = \boxed{\boldsymbol{C}}\boldsymbol{x}_i + \boxed{\boldsymbol{D}}\boldsymbol{u}_i$$

Measurements

Feedthrough

$$\mathcal{O}_k = \begin{bmatrix} C \\ CA \\ CA^2 \\ CA^3 \\ \vdots \\ CA^{k-1} \end{bmatrix}$$

Observability

$$\mathcal{C}_k = \begin{bmatrix} B & AB & A^2B & A^3B & \cdots & A^{k-1}B \end{bmatrix}$$

Controlability

$$\mathcal{H}_k = \begin{bmatrix} D & CB & CAB & CA^2B & CA^3B & \cdots & CA^{k-1}B \end{bmatrix}$$

Markov parameters of the system

Schematic mass spring damper

Impulse response

EigenRealization Algorithm

$$\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} \end{bmatrix}$$

EigenRealization Algorithm

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

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

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

EigenRealization Algorithm

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

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

EigenRealization Algorithm

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

EigenRealization Algorithm

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

EigenRealization Algorithm

Natural dynamics

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

Actuators

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

Measurements

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

Feedthrough

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

Mass spring damper example

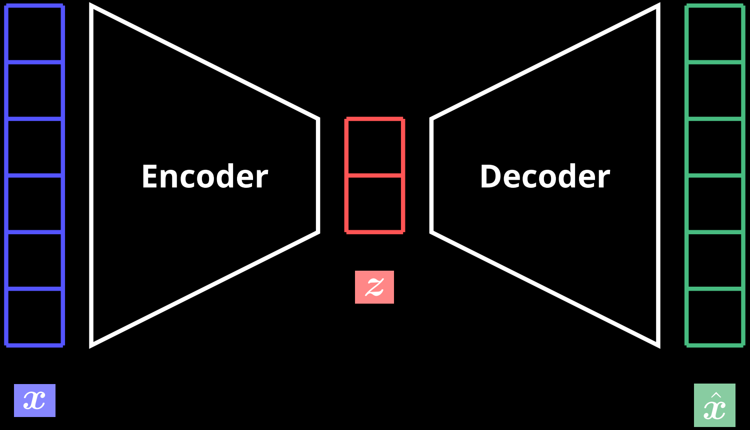
Cylinder flow example

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.