# Projection-Based & Data-Driven Reduced-Order Models with Linear Dimensionality Reduction

Mengwu Guo (University of Twente).

* Linear dimensionality reduction = linear PCA. = POD.

## §1. Basics on projection-based reduction.

Benchmark example : 1D viscous Burgers' equation

$\nu = \frac{1}{Re}$.

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu \frac{\partial^2 u}{\partial x^2} \qquad x \in \Omega = (0,L),\ t>0$$

$$u(0,t) = u(L,t) = 0. \qquad t>0.$$

$$u(x,0) = u_0(x). \qquad x \in \Omega$$

### §1.1. Weak form & Galerkin.

- weak form of this problem: Find $u(\cdot,t) \in H_0^1(\Omega)$ ?, $t>0$, such that

In fact, it's also required that $\frac{\partial}{\partial t}u(\cdot,t) \in L^2(\Omega)$.

$$\left(\frac{\partial u}{\partial t}, v\right)_\Omega + \left(u\frac{\partial u}{\partial x}, v\right)_\Omega + \nu\left(\frac{\partial u}{\partial x}, \frac{\partial v}{\partial x}\right)_\Omega = 0 \qquad \forall v \in H_0^1(\Omega).$$

(integration by part) Here $(f,g)_\Omega := \int_0^L f(x)\,g(x)\,dx$.

- Galerkin scheme for approximation.

  Let $V_h = \mathrm{span}\{\psi_1(x), \psi_2(x), \cdots, \psi_N(x)\} \subset H_0^1(\Omega)$.

Find $u_h(\cdot;t) \in V_h$, $t>0$, such that

$$\left(\frac{\partial u_h}{\partial t}, \psi_i\right)_\Omega + \left(u_h \frac{\partial u_h}{\partial x}, \psi_i\right)_\Omega + \nu\left(\frac{\partial u_h}{\partial x}, \frac{\partial \psi_i}{\partial x}\right)_\Omega = 0$$

$i = 1, 2, \cdots, N.$

- Algebraic formulation.

$$V_h \ni u_h(x,t) = \sum_{j=1}^{N} a_j(t)\psi_j(x).$$

$$\left(\sum_{j=1}^{N} \frac{da_j}{dt}\psi_j, \psi_i\right)_\Omega + \left(\sum_{k=1}^{N} a_k\psi_k \cdot \sum_{j=1}^{N} a_j \frac{\partial \psi_j}{\partial x}, \psi_i\right)_\Omega + \nu\left(\sum_{j=1}^{N} a_j\frac{\partial \psi_j}{\partial x}, \frac{\partial \psi_i}{\partial x}\right)_\Omega = 0$$

$$\sum_{j=1}^{N} (\psi_i, \psi_j)_\Omega \frac{da_j}{dt} + \sum_{j=1}^{N}\sum_{k=1}^{N}\left(\psi_i, \frac{\partial \psi_j}{\partial x}\psi_k\right)_\Omega a_j a_k + \sum_{j=1}^{N} \nu\left(\frac{\partial \psi_i}{\partial x}, \frac{\partial \psi_j}{\partial x}\right)_\Omega a_j = 0.$$

$$(*) \quad \underline{\underline{M}}\,\dot{\underline{a}} + \underline{\underline{\underline{B}}} : (\underline{a} \otimes \underline{a}) + \underline{\underline{A}}\,\underline{a} = \underline{0}. \qquad + \text{initial condition}$$

in which $M_{ij} = (\psi_i, \psi_j)_\Omega$. $B_{ijk} = \left(\psi_i, \frac{\partial \psi_j}{\partial x}\psi_k\right)_\Omega$, $A_{ij} = \nu\left(\frac{\partial \psi_i}{\partial x}, \frac{\partial \psi_j}{\partial x}\right)_\Omega$

## §1.2 Reduced-order formulation.

$$[\underline{a}^{(1)}\ \underline{a}^{(2)} \cdots \underline{a}^{(N_s)}] = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{Z}}^T \qquad \underline{\text{SVD}}$$

r-dimensional POD basis $\underline{\underline{V}} = \underline{\underline{U}}_{:,1:r} \in \mathbb{R}^{N\times r}$ $\quad \underline{r \ll N}.$

- Directly start with the algebraic formulation.

step 1. approximate : $\underline{a}(t) \approx \underline{\underline{V}}\,\underline{q}(t)$. $\quad \underline{q} \in \mathbb{R}^r$.

step 2. substitute : $\underline{\underline{M}}\,\underline{\underline{V}}\,\dot{\underline{q}} + \underline{\underline{\underline{B}}} : (\underline{\underline{V}}\,\underline{q} \otimes \underline{\underline{V}}\,\underline{q}) + \underline{\underline{A}}\,\underline{\underline{V}}\,\underline{q} = \underline{0}$

(r unknowns, N equations).

Step.3 (Galerkin) project : $\underline{\underline{V}}^T \underline{\underline{M}}\,\underline{\underline{V}}\,\dot{\underline{q}} + \underline{\underline{V}}^T \underline{\underline{\underline{B}}} : (\underline{\underline{V}}\underline{q} \otimes \underline{\underline{V}}\underline{q}) + \underline{\underline{V}}^T \underline{\underline{A}}\,\underline{\underline{V}}\,\underline{q} = \underline{0}.$

( r unknown, r equations).

$$(*) \quad \hat{\underline{\underline{M}}}\,\dot{\underline{q}} + \hat{\underline{\underline{\underline{B}}}}\,(\underline{q} \otimes \underline{q}) + \hat{\underline{\underline{A}}}\,\underline{q} = \underline{0}.$$

System size $= r$. (REDUCED FROM N !!!)

$$\hat{\underline{\underline{M}}} = \underline{\underline{V}}^T \underline{\underline{M}}\,\underline{\underline{V}}. \quad \hat{\underline{\underline{\underline{B}}}} = \underline{\underline{V}}^T \underline{\underline{\underline{B}}} : (\underline{\underline{V}} \otimes \underline{\underline{V}}) . \quad \hat{\underline{\underline{A}}} = \underline{\underline{V}}^T \underline{\underline{A}}\,\underline{\underline{V}}.$$

what does this mean?

- Intepretation from the Galerkin formulation. (weak form).

$$V_r = \text{span}\{\phi_1, \phi_2, \cdots, \phi_r\}. \subset V_h$$

This is a function space!

$$\phi_\ell(x) = \sum_{i=1}^{N} V_{i\ell}\,\psi_i(x)$$

This is the POD basis matrix.

$$V_r \ni u_r(x,t) = \sum_{\ell=1}^{r} q_\ell(t)\,\phi_\ell(x)$$

$$= \sum_{i=1}^{N} \underbrace{\sum_{\ell=1}^{r} V_{i\ell}\, q_\ell(t)}_{a_i(t)}\, \psi_i(x)$$

i.e., $\underline{a}(t) = \underline{\underline{V}}\,\underline{q}(t)$.

Find $u_r(\cdot,t) \in V_r$, $t>0$, such that

$$\left(\frac{\partial u_r}{\partial t}, \phi_i\right)_\Omega + \left(u_r \frac{\partial u_r}{\partial x}, \phi_i\right)_\Omega + \nu\left(\frac{\partial u_r}{\partial x}, \frac{\partial \phi_i}{\partial x}\right)_\Omega = 0$$

$i = 1, 2, \cdots, r.$

similar with before;

$$\sum_{j=1}^{r} (\phi_i, \phi_j)_\Omega \frac{d}{dt} q_j + \sum_{j,k=1}^{r} \left(\phi_i, \frac{\partial \phi_j}{\partial x}\phi_k\right)_\Omega q_j q_k + \sum_{j=1}^{r} \nu \left(\frac{\partial \phi_i}{\partial x}, \frac{\partial \phi_j}{\partial x}\right)_\Omega q_j = 0.$$

$$\hat{\underline{\underline{M}}}\,\dot{\underline{q}}(t) + \hat{\underline{\underline{\underline{B}}}} : \underline{q} \otimes \underline{q} + \hat{\underline{\underline{A}}}\,\underline{q} = 0.$$

Specifically

$$\hat{M}_{ij} = (\phi_i, \phi_j)_\Omega = \Big( \sum_{m=1}^{N} V_{mi} \psi_m , \sum_{p=1}^{N} V_{pj} \psi_p \Big)_\Omega$$

$$= \sum_{m=1}^{N} \sum_{p=1}^{N} V_{mi} (\psi_m, \psi_p)_\Omega V_{pj}.$$

i.e., $\hat{\underline{\underline{M}}} = \underline{\underline{V}}^T \underline{\underline{M}}\,\underline{\underline{V}}.$

$$\hat{A}_{ij} = \nu \Big( \frac{\partial \phi_i}{\partial x}, \frac{\partial \phi_j}{\partial x} \Big)_\Omega = \sum_{m=1}^{N} \sum_{p=1}^{N} V_{mi} \Big( \frac{\partial \psi_m}{\partial x}, \frac{\partial \psi_p}{\partial x} \Big)_\Omega V_{pj}$$

i.e., $\hat{\underline{\underline{A}}} = \underline{\underline{V}}^T \underline{\underline{A}}\,\underline{\underline{V}}.$

$$\hat{B}_{ijk} = \Big( \phi_i , \frac{\partial \phi_j}{\partial x} \phi_k \Big)_\Omega = \Big( \sum_{m=1}^{N} V_{mi} \psi_m , \Big( \sum_{p=1}^{N} V_{pj} \frac{\partial \psi_p}{\partial x} \Big) \cdot \Big( \sum_{s=1}^{N} V_{sk} \psi_s \Big) \Big)_\Omega$$

$$= \sum_{m,p,s=1}^{N} V_{mi} \Big( \psi_m , \frac{\partial \psi_p}{\partial x} \psi_s \Big)_\Omega V_{pj} V_{sk} \quad \circledast.$$

i.e., $\hat{\underline{\underline{\underline{B}}}} = \underline{\underline{V}}^T \underline{\underline{\underline{B}}} : (\underline{\underline{V}} \otimes \underline{\underline{V}}).$

## §2. Parametric reduced-order surrogate modeling.

Full-order:
$$\dot{\underline{a}}(t) = -\underline{\underline{M}}^{-1} \underline{\underline{A}}\,\underline{a}(t) - \underline{\underline{M}}^{-1} \underline{\underline{\underline{B}}} : (\underline{a}(t) \otimes \underline{a}(t))$$

$$:= \underline{\underline{K}}\,\underline{a}(t) + \underline{\underline{\underline{H}}} : (\underline{a}(t) \otimes \underline{a}(t)).$$

In fact, $\underline{\underline{K}}$ and $\underline{\underline{\underline{H}}}$ can be parametric. $\quad \underline{\mu} \in D \subset \mathbb{R}^d.$

$$\dot{\underline{a}}(t;\underline{\mu}) = \underline{\underline{K}}(\underline{\mu})\,\underline{a}(t;\underline{\mu}) + \underline{\underline{\underline{H}}}(\underline{\mu}) : (\underline{a}(t;\underline{\mu}) \otimes \underline{a}(t;\underline{\mu})).$$

### §2.1. Low-rank approximation with the POD basis.

$$\underline{a}(t;\underline{\mu}) \approx \sum_{l=1}^{r} \hat{q}_l(t;\underline{\mu}) \overbrace{\underline{V}_l}^{l\text{-th column of } \underline{\underline{V}}} = \underline{\underline{V}}\,\hat{\underline{q}}(t;\underline{\mu}).$$

$$\hat{\underline{q}}(t;\underline{\mu}) = \arg\min_{\underline{b}\in\mathbb{R}^r} \left\|\underline{a}(t;\underline{\mu}) - \underline{\underline{V}}\,\underline{b}\right\|^2. \text{ (least-squares).}$$

---

least squares $\min_{x\in\mathbb{R}^r} \|\underline{\underline{D}}\,\underline{x} - \underline{y}\|^2$. $\underline{y}\in\mathbb{R}^N$. $\underline{\underline{D}}\in\mathbb{R}^{N\times r}$. $r\leq N$.

let $\nabla_{\underline{x}}\left(\|\underline{\underline{D}}\,\underline{x} - \underline{y}\|^2\right) = 2\left(\underline{\underline{D}}^T\underline{\underline{D}}\,\underline{x} - \underline{\underline{D}}^T\underline{y}\right) = 0.$

$$\underline{x}^* = \left(\underline{\underline{D}}^T\underline{\underline{D}}\right)^{-1}\underline{\underline{D}}^T\underline{y} \quad (\text{if } \underline{\underline{D}} \text{ is full column-rank})$$

---

$$= \left(\underline{\underline{V}}^T\underline{\underline{V}}\right)\underline{\underline{V}}^T\underline{a}(t;\underline{\mu}) = \underline{\underline{V}}^T\underline{a}(t;\underline{\mu}).$$

Direct projection onto the POD basis $\underline{\underline{V}}$.

## §2.2 Surrogate approximation of the expansion coefficients

$$S: [0,T]\times D \to \mathbb{R}^r,\ (t,\underline{\mu}) \mapsto \underline{\underline{V}}^T\underline{a}(t;\underline{\mu}).$$

Black-box.

- approximated by vector-valued neural networks.
- Component-wise approximated by scalar-valued Gaussian processes.

## §2.3 Gaussian process surrogate modeling.

---

Conditional Gaussian.

$$\begin{pmatrix}\underline{a}\\ \underline{b}\end{pmatrix} \sim N\left(\begin{pmatrix}\underline{\theta}_a\\ \underline{\theta}_b\end{pmatrix}, \begin{pmatrix}\underline{\underline{C}}_{aa} & \underline{\underline{C}}_{ab}\\ \underline{\underline{C}}_{ba} & \underline{\underline{C}}_{bb}\end{pmatrix}\right).$$

$$\Rightarrow \underline{b}\,|\,\underline{a} \sim N\left(\underline{\theta}_b + \underline{\underline{C}}_{ba}\underline{\underline{C}}_{aa}^{-1}(\underline{a}-\underline{\theta}_a),\ \underline{\underline{C}}_{bb} - \underline{\underline{C}}_{ba}\underline{\underline{C}}_{aa}^{-1}\underline{\underline{C}}_{ab}\right).$$

---

$$f(\underline{x}) \sim GP(m(\underline{x}), k(\underline{x}, \underline{x}')) \qquad \text{random field.}$$

$$y(\underline{x}) = f(\underline{x}) + \varepsilon. \qquad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2) \text{ white noise.}$$

Input-output data pairs

$$(\underline{\underline{X}}, \underline{y}) = \{(\underline{x}^{(i)}, y^{(i)})\}_{i=1}^{M}.$$

$$\begin{pmatrix} \underline{y} \\ f(\underline{x}^*) \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \underline{m} \\ m(\underline{x}^*) \end{pmatrix}, \begin{pmatrix} \underline{\underline{K}} + \sigma_\varepsilon^2 \underline{\underline{I}}_M & \underline{c}(\underline{x}^*) \\ \underline{c}(\underline{x}^*)^T & k(\underline{x}^*, \underline{x}^*) \end{pmatrix} \right)$$

in which $\underline{m}_i = m(\underline{x}^{(i)})$ $\quad \underline{c}_i(\underline{x}^*) = k(\underline{x}^{(i)}, \underline{x}^*)$

$$\underline{\underline{K}}_{ij} = k(\underline{x}^{(i)}, \underline{x}^{(j)}). \qquad i, j = 1, 2, \cdots, M.$$

Then conditional Gaussian gives.

$$f(\underline{x}^*) | \underline{y} \sim \mathcal{N}(m^*(\underline{x}^*), v^*(\underline{x}^*))$$

with $m^*(\underline{x}^*) = m(\underline{x}^*) + \underline{c}(\underline{x}^*)^T (\underline{\underline{K}} + \sigma_\varepsilon^2 \underline{\underline{I}}_M)^{-1} (\underline{y} - \underline{m})$.

and $v^*(\underline{x}^*) = k(\underline{x}^*, \underline{x}^*) - \underline{c}(\underline{x}^*)^T (\underline{\underline{K}} + \sigma_\varepsilon^2 \underline{\underline{I}}_M)^{-1} \underline{c}(\underline{x}^*)$

## §3. Reduced-order operator inference.

$$\hat{\underline{\underline{M}}}\, \dot{\underline{q}}(t) + \hat{\underline{\underline{\underline{B}}}} : (\underline{q}(t) \otimes \underline{q}(t)) + \hat{\underline{\underline{A}}}\, \underline{q}(t) = \underline{0}.$$

$$\Rightarrow \quad \dot{\underline{q}}(t) = \hat{\underline{\underline{K}}}\, \underline{q}(t) + \hat{\underline{\underline{\underline{H}}}} : (\underline{q}(t) \otimes \underline{q}(t))$$

$$\hat{\underline{\underline{K}}} = -(\underline{\underline{V}}^T \underline{\underline{M}}\, \underline{\underline{V}})^{-1} (\underline{\underline{V}}^T \underline{\underline{A}}\, \underline{\underline{V}}).$$

$$\hat{\underline{\underline{\underline{H}}}} = -(\underline{\underline{V}}^T \underline{\underline{M}}\, \underline{\underline{V}})^{-1} (\underline{\underline{V}}^T \underline{\underline{\underline{B}}} : (\underline{\underline{V}} \otimes \underline{\underline{V}}))$$

What if we have no access to $\underline{\underline{M}}$, $\underline{\underline{A}}$, or $\underline{\underline{B}}$?

We directly estimate $\hat{\underline{\underline{E}}}$ and $\hat{\underline{\underline{H}}}$ from snapshot data.

§3.1. Reformulation.

$$\begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \vdots \\ \dot{q}_r \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1r} \\ A_{21} & A_{22} & \cdots & A_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rr} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_r \end{bmatrix} + \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1g} \\ H_{21} & H_{22} & \cdots & H_{2g} \\ \vdots & \vdots & \ddots & \vdots \\ H_{r1} & H_{r2} & \cdots & H_{rg} \end{bmatrix} \begin{bmatrix} q_1^2 \\ q_1 q_2 \\ q_1 q_3 \\ \vdots \\ q_2^2 \\ q_2 q_3 \\ \vdots \\ q_r^2 \end{bmatrix}$$

$$g(r) = \frac{(1+r)r}{2}.$$

For each $q_i$: $1 \le i \le r$.

$$\dot{q}_i(t) = \left(q_1, \cdots, q_r, q_1^2, q_1 q_2, \cdots, q_2^2, q_2 q_3, \cdots, q_r^2\right)\left(A_{i1}, \cdots, A_{ir}, H_{i1}, \cdots, H_{ig}\right)^T$$

$$= \underline{d}^T(t) \quad \underline{O}_i. \qquad \underline{d}, \underline{O}_i \in \mathbb{R}^{r+g(r)}.$$

§3.2 linear regression problem.

$$\dot{q}_i(t) = \underline{d}^T(t) \quad \underline{O}_i$$

From the snapshot data $\{\underline{q}(t^{(1)}), \underline{q}(t^{(2)}), \cdots, \underline{q}(t^{(N_s)})\}$.

one has $\underline{r}_i = \left[\dot{q}_i(t^{(1)}), \dot{q}_i(t^{(2)}), \cdots, \dot{q}_i(t^{(N_s)})\right]^T \in \mathbb{R}^{N_s}$

by finite difference.

and $\underline{\underline{D}} = \left[\underline{d}(t^{(1)}), \underline{d}(t^{(2)}), \cdots, \underline{d}(t^{(N_s)})\right]^T \in \mathbb{R}^{N_s \times (r+g(r))}$

$(r+g(r) < N_s)$

leading to a linear model $\underline{r}_i = \underline{\underline{D}}\, \underline{o}_i + \underline{\varepsilon}$.

Original least-squares estimate of $\underline{o}_i$.

$$\underline{o}_i = \underset{\underline{o}_i \in \mathbb{R}^{(r+g(r))}}{\arg\min} \| \underline{r}_i - \underline{\underline{D}}\, \underline{o}_i \|^2 .$$

$$= \left(\underline{\underline{D}}^T \underline{\underline{D}}\right)^{-1} \underline{\underline{D}}^T \underline{r}_i \quad (\text{if } \underline{\underline{D}} \text{ is full column-rank}).$$

§3.3. What if $D$ is NOT full column rank

or $D^T D$ (Gram matrix) is ill-conditioned?

We can use a Tikhonov regularizer. (ridge regression).

$$\underline{o}_i = \underset{\underline{o}_i \in \mathbb{R}^{r+g(r)}}{\arg\min} \left\{ \| \underline{r}_i - \underline{\underline{D}}\, \underline{o}_i \|^2 + \lambda \|\underline{o}_i\|^2 \right\}$$

$$= \left(\underline{\underline{D}}^T \underline{\underline{D}} + \lambda \underline{\underline{I}}\right)^{-1} \underline{\underline{D}}^T \underline{r}_i .$$

§3.4 Benefits from operator inference.

① non-intrusive, but not black-box.

② preserved formulation structure.

③ can not only reconstruct training state data, but also predict for the future states.

## §3.5 Uncertainty quantification with Bayesian inference.

$$\underline{r}_i = \underline{\underline{D}}\,\underline{o}_i + \underline{\varepsilon} \qquad \underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{N_s})^T$$

all white noise $\sim \mathcal{N}(0, \sigma^2)$.

- likelihood $\underline{r}_i \,|\, \underline{\underline{D}}, \underline{o}_i \sim \mathcal{N}(\underline{\underline{D}}\,\underline{o}_i, \sigma^2 \underline{\underline{I}}_{N_s})$.

  i.e., $p(\underline{r}_i | \underline{\underline{D}}, \underline{o}_i) \propto \exp\left(-\frac{1}{2\sigma^2} \| \underline{r}_i - \underline{\underline{D}}\,\underline{o}_i \|^2\right)$.

- prior. $\underline{o}_i \sim \mathcal{N}\left(\underline{0}, \frac{\sigma^2}{\lambda} \underline{\underline{I}}_{(r+g(r))}\right)$.

  i.e., $p(\underline{o}_i) \propto \exp\left(-\frac{\lambda}{2\sigma^2} \| \underline{o}_i \|^2\right)$.

- posterior. (Bayes' rule).

$$p(\underline{o}_i | \underline{\underline{D}}, \underline{r}_i) = \frac{p(\underline{o}_i)\, p(\underline{r}_i | \underline{\underline{D}}, \underline{o}_i)}{p(\underline{r}_i | \underline{\underline{D}})}$$

$$\propto p(\underline{o}_i)\, p(\underline{r}_i | \underline{\underline{D}}, \underline{o}_i).$$

$$\propto \exp\left[-\frac{1}{\sigma^2}\left(\| \underline{r}_i - \underline{\underline{D}}\,\underline{o}_i \|^2 + \lambda \| \underline{o}_i \|^2\right)\right]$$

$$\propto \exp\left(-(\underline{o}_i - \underline{\mu}_i)^T \underline{\underline{\Sigma}}_i^{-1} (\underline{o}_i - \underline{\mu}_i)\right).$$

i.e., $\underline{o}_i | \underline{\underline{D}}, \underline{r}_i \sim \mathcal{N}(\underline{\mu}_i, \underline{\underline{\Sigma}}_i)$.

with $\underline{\mu}_i = (\underline{\underline{D}}^T \underline{\underline{D}} + \lambda \underline{\underline{I}})^{-1} \underline{\underline{D}}^T \underline{r}_i$

$\underline{\underline{\Sigma}}_i = \sigma^2 (\underline{\underline{D}}^T \underline{\underline{D}} + \lambda \underline{\underline{I}})^{-1}$

- Ridge regression
- Maximum A Posteriori (MAP).

# Outline

## Projection-based ROM

| Strong form | Weak form/Galerkin | Algebraic | |
|---|---|---|---|
| PDE | | 1 | Full-order |
| — | 2 | | Reduced-order |

## Data-driven ROM

black-box ◦ Parametric interpolation — Surrogate modeling

physics-involved ◦ Temporal extrapolation — Operator inference.