Dictionary-Based Model Reduction for State Estimation

A. Pasco and A. Nouy

Centrale Nantes, Nantes Université, LMJL, UMR CNRS 6629

alexandre.pasco@ec-nantes.fr, anthony.nouy@ec-nantes.fr

Context

We aim to approximate an **unknown state** $u(\xi) \in U$, e.g. $U = \mathbb{R}^{\mathcal{N}}$, solution of the **parametric problem**

$$\mathcal{F}\big(u(\xi), \xi\big) = 0,$$

where $\xi \in \mathcal{P} \subset \mathbb{R}^d$ is an **unknown parameter**. We only have access to m **linear measurements** on $u(\xi)$, in a **noise-less** framework, $\ell_1(u), \cdots, \ell_m(u) \in \mathbb{R}$. Equivalently, we measure $w = P_W u(\xi)$ with $\dim(W) = m$. We then use **Model Order Reduction** on $\mathcal{M} := \{u(\xi') : \xi' \in \mathcal{P}\} \subset U$ as prior knowledge to estimate the state.

One-space approach (PBDW) [3]

Formulation

- Given a n-dimensional space V such that

$$\mathrm{dist}(V, \mathcal{M}) \leq \varepsilon.$$

- Assume finiteness of the inf-sup constant

$$\mu(V, W) := \sup_{x \in W^\perp} \frac{\|x\|}{\|x - P_V x\|}$$

One space (PBDW) recovery

$$A_V(w) := v^* + \eta^*,$$

$$\underbrace{v^* := \operatorname{argmin}_{v \in V} \|w - P_W v\|}_{\text{Best fitted to the observations in } V}, \quad \underbrace{\eta^* := w - P_W v^*}_{\text{To fit the observations}}.$$

Properties

- ▶ A priori error bound

$$\|u - A_V(w)\| \leq \mu(V, W)\varepsilon.$$

- ▶ A_V is **optimal** when $\mathcal{M}$ is a cylinder centered at V.

Issues

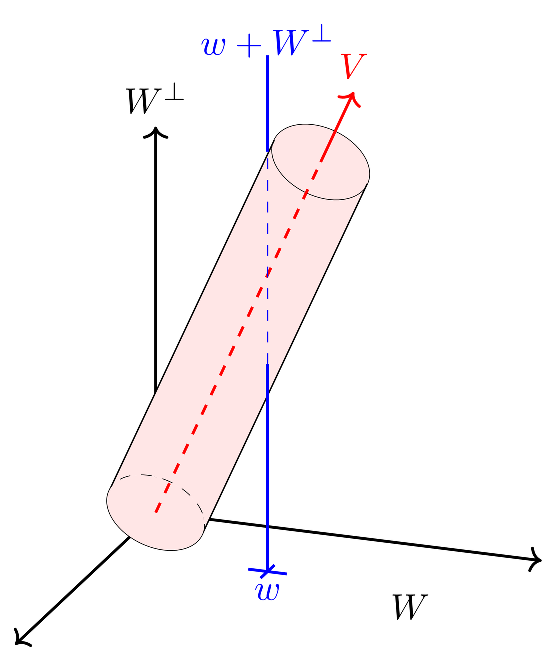
The approach may perform poorly when:

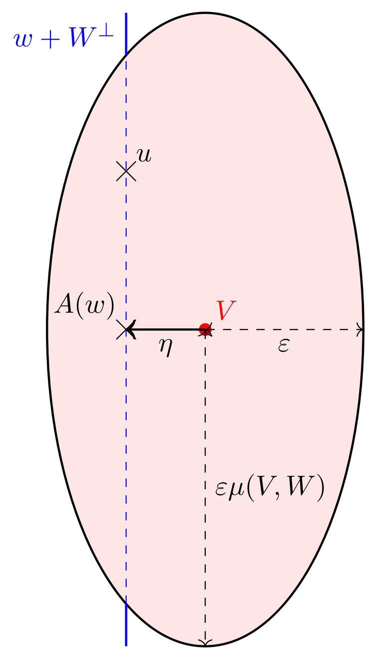
- ▷ The Kolmogorov n-width $d_n(\mathcal{M})$ has a **slow decay**,

$$d_n(\mathcal{M}) := \inf_{\dim V = n} \mathrm{dist}(V, \mathcal{M}).$$

- ▷ The number and type of measurements are restricted, which deteriorates $\mu(V, W)$.

Geometric interpretation





Multi-space approach [2]

Formulation

- Approximate $\mathcal{M}$ by a library of spaces of dim $\leq n$,

$$\mathcal{L}_n^N := \{V_1, \cdots, V_N\}, \quad \mathrm{dist}(\mathcal{M}, \cup_{k=1}^N V_k) \leq \varepsilon.$$

- Assume we have $\mathcal{S}(\cdot, \mathcal{M})$ such that for all $v \in U$,

$$c \ \mathrm{dist}(v, \mathcal{M}) \leq \mathcal{S}(v, \mathcal{M}) \leq C \ \mathrm{dist}(v, \mathcal{M}).$$

Multi-space recovery

Select the recovery $A_{\mathcal{S}}^{\mathrm{mult}}(w) := A_{V_{\mathcal{S}}}(w)$ with

$$V_{\mathcal{S}} = V_{\mathcal{S}}(w) := \operatorname{argmin}_{V \in \mathcal{L}_n^N} \mathcal{S}(A_V(w), \mathcal{M})$$

Properties

- ▶ **Better prior** approximation of $\mathcal{M}$ with low-dimensional subspaces, thus better stability expected.
- ▶ **Near-optimal selection** if P_W is injective on $\mathcal{M}$,

$$\|u - A_{\mathcal{S}}^{\mathrm{mult}}(w)\| \leq 2\frac{C}{c}\mu(\mathcal{M}, W) \min_{V \in \mathcal{L}_n^N} \|u - A_V(w)\|,$$

where $\mu(\mathcal{M}, W)$ reflects the orientation between $\mathcal{M}$ and W.

Issues

- ▷ The stability constant $\mu(\mathcal{M}, W)$ is generally not computable and may be infinite.

Dictionary-based multi-space [4]

Formulation

- Given a dictionary $\mathcal{D}_K := \{v^{(1)}, \cdots, v^{(K)}\}$, consider the library $\mathcal{L}_m(\mathcal{D}_K)$ containing the subspaces spanned by at most m vectors from $\mathcal{D}_K$.
- For $\alpha > 0$ consider the Lasso problem

$$\mathbf{x}_\alpha(w) \in \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^K} \frac{1}{2}\|w - \sum_{i=1}^K \mathbf{x}_i P_W v^{(i)}\|^2 + \alpha\|\mathbf{x}\|_1,$$

use the non-zero components to build the subspace

$$V_\alpha(w) := \mathrm{span}\{v^{(i)} : \mathbf{x}_\alpha(w)_i \neq 0\} \in \mathcal{L}_m(\mathcal{D}_K).$$

Dictionary-based multi-space recovery

LARS algorithm solves the Lasso for all α and gives a finite sub-library $\mathcal{L}^{\mathrm{dic}}(w) := \{V_\alpha(w) : \alpha > 0\}$ among which we select the recovery $A_{\mathcal{S}}^{\mathrm{dic}}(w) := A_{V_{\mathcal{S}}}(w)$ with

$$V_{\mathcal{S}} = V_{\mathcal{S}}(w) := \operatorname{argmin}_{V \in \mathcal{L}^{\mathrm{dic}}(w)} \mathcal{S}(A_V(w), \mathcal{M})$$

Properties

- ▶ **Wide range** of possible background spaces V.
- ▶ **Near-optimal selection** among the sub-library.
- ▶ **Fast online** sub-library generation with LARS.

Issues

- ▷ Same issue as multi-space.

Randomized approach for linear PDEs

Linear PDE framework

- Consider $U = \mathbb{R}^{\mathcal{N}}$, $B(\xi) \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}}$, $f(\xi) \in \mathbb{R}^{\mathcal{N}}$ such that

$$B(\xi)u(\xi) = f(\xi)$$

with singular values of $B(\cdot)$ bounded by $0 < c \leq C$.

- Consider the residual-based quantity [2]

$$\mathcal{S}(v, \mathcal{M}) := \min_{\xi \in \mathcal{P}} \|B(\xi)v - f(\xi)\|, \quad v \in U$$

Properties

- ▶ When B and f are **parameter separable**, $\mathcal{S}(\cdot, \mathcal{M})$ can be computed by solving a non-linear least-squares problem, built via an **offline-online** decomposition.

Issues

- ▷ The previous **offline cost** may be prohibitive (esp. for dic-based approach) and sensitive to **round-off errors**.

Randomized approach

- Consider a random (e.g. gaussian) matrix $\Theta \in \mathbb{R}^{k \times \mathcal{N}}$

Randomized residual-based distance

Use the randomized quantity

$$\mathcal{S}^\Theta(v, \mathcal{P}) := \min_{\xi \in \mathcal{P}} \|\Theta\big(B(\xi)v - f(\xi)\big)\|.$$

Properties

- ▶ for any $v \in U$, with high probability,

$$\sqrt{1 - \epsilon}\ \mathcal{S}(v, \mathcal{P}) \leq \mathcal{S}^\Theta(v, \mathcal{P}) \leq \sqrt{1 + \epsilon}\ \mathcal{S}(v, \mathcal{P})$$

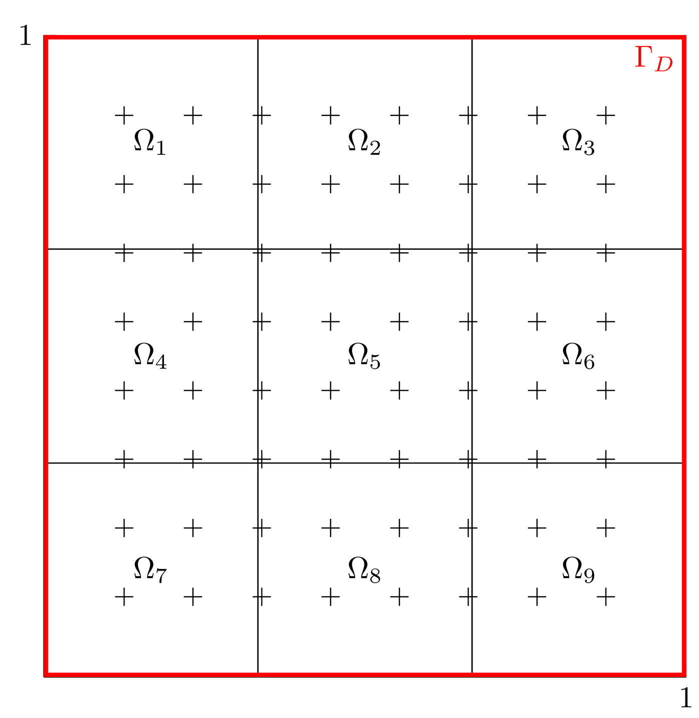
where $\epsilon \in (0, 1)$ and a small number k of rows of Θ [1].

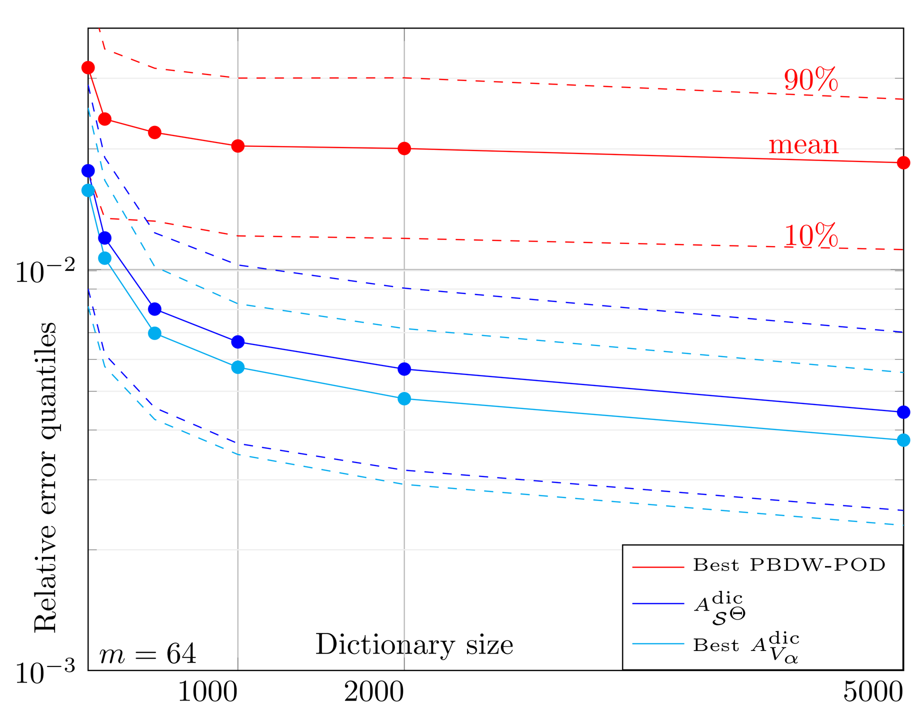
- ▶ Same as the residual-based quantity, but $\mathcal{S}^\Theta(\cdot, \mathcal{M})$ is computed with a **more efficient** and **more stable** offline-online decomposition.
- ▶ **Near-optimal selection** with high probability.

Conclusion and future work

- ▶ More details on the arXiv preprint (QR-code).
- ▷ Noisy observations case ?
- ▷ More nonlinear recovery maps (auto-encoders) ?

Numerical example





2D diffusion

Thermal block problem with $d = 9$ subdomains,

$$\begin{cases} -\nabla \cdot (\kappa \nabla u) = 1 & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma_D, \\ \kappa = \xi_i & \text{in } \Omega_i, \end{cases}$$

with $m = 64$ local convolution sensors, and a random matrix Θ with $k = 100$ rows.

References

[1] O. Balabanov and A. Nouy. Randomized linear algebra for model reduction. Part I: Galerkin methods and error estimation. *Adv Comput Math*, 45(5-6):2969–3019, Dec. 2019.

[2] A. Cohen, W. Dahmen, O. Mula, and J. Nichols. Nonlinear Reduced Models for State and Parameter Estimation. *SIAM/ASA J. Uncertainty Quantification*, 10(1):227–267, Mar. 2022.

[3] Y. Maday, A. T. Patera, J. D. Penn, and M. Yano. A parameterized-background data-weak approach to variational data assimilation: Formulation, analysis, and application to acoustics. *Int. J. Numer. Meth. Engng*, 102(5):933–965, May 2015.

[4] A. Nouy and A. Pasco. Dictionary-based model reduction for state estimation, 2023.