Dictionary-based model reduction for state estimation

Alexandre PASCO

Centrale Nantes, Nantes Université, France
In collaboration with Anthony NOUY.

Context

Context

- Parametric equation in a Hilbert space U,

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

$u \in U$ the **state to recover** and $\xi \in \mathcal{P} \subset \mathbb{R}^d$ an **unknown** parameter.

Context

- Parametric equation in a Hilbert space U,

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

$u \in U$ the **state to recover** and $\xi \in \mathcal{P} \subset \mathbb{R}^d$ an **unknown** parameter.

Observations

of m continuous linear measurements
$\ell_1(u), \cdots, \ell_m(u)$, i.e. $w = P_W u$
with $W := \text{span}\{R_U \ell_i\}_{i=1}^m$

Context

- Parametric equation in a Hilbert space U,

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

$u \in U$ the **state to recover** and $\xi \in \mathcal{P} \subset \mathbb{R}^d$ an **unknown** parameter.

Observations + **Prior knowledge**

of m continuous linear measurements $\ell_1(u), \cdots, \ell_m(u)$, i.e. $w = P_W u$ with $W := \text{span}\{R_U \ell_i\}_{i=1}^m$

on the solution manifold $\mathcal{M} := \{u(\xi) : \xi \in \mathcal{P}\}$ based on model reduction

Context

- Parametric equation in a Hilbert space U,

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

$u \in U$ the **state to recover** and $\xi \in \mathcal{P} \subset \mathbb{R}^d$ an **unknown** parameter.

Observations + **Prior knowledge**

of m continuous linear measurements $\ell_1(u), \cdots, \ell_m(u)$, i.e. $w = P_W u$ with $W := \text{span}\{R_U \ell_i\}_{i=1}^m$

on the solution manifold $\mathcal{M} := \{u(\xi) : \xi \in \mathcal{P}\}$ based on model reduction

$\downarrow$

Compute a recovery $A(w) \simeq u$

One space approach

Linear MOR

One space: formulation

- Approximate $\mathcal{M}$ by a linear subspace V,

$$\mathsf{dist}(V, \mathcal{M}) \leq \varepsilon, \qquad \mathsf{dim}(V) = n$$

One space: formulation

- Approximate $\mathcal{M}$ by a linear subspace V,

$$\text{dist}(V, \mathcal{M}) \leq \varepsilon, \qquad \dim(V) = n$$

- Parameterized-Background Data-Weak (PBDW) [Maday et al., 2015]:

$$A_V(w) := v^* + \eta^*, \quad v^* := \arg\min_{v \in V} \|P_W(u - v)\|, \quad \eta^* := w - P_W v^*$$

One space: formulation

- Approximate $\mathcal{M}$ by a linear subspace V,

$$\text{dist}(V, \mathcal{M}) \leq \varepsilon, \qquad \dim(V) = n$$

- Parameterized-Background Data-Weak (PBDW) [Maday et al., 2015]:

$$A_V(w) := v^* + \eta^*, \quad v^* := \arg\min_{v \in V} \|P_W(u - v)\|, \quad \eta^* := w - P_W v^*$$

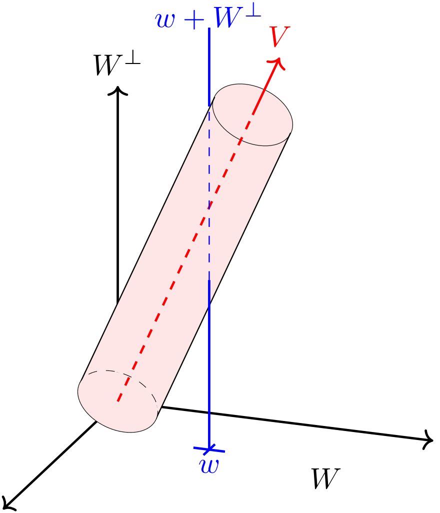
- Sharp error bound [Binev et al., 2017]: $\|u - A_V(w)\| \leq \varepsilon \mu(V, W)$

$$\mu(V, W) := \left(\inf_{v \in V} \sup_{w \in W} \frac{\langle v, w \rangle}{\|v\| \|w\|} \right)^{-1}$$

One space: Geometry

Linear MOR

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



One space: Pros and cons

Pros

- Online efficiency $\mathcal{O}(n^3)$.
- No need to know ε.
- Optimal (worst case sense) when $\mathcal{M}$ is a cylinder centered in V [Binev et al., 2017].

One space: Pros and cons

Pros

- Online efficiency $\mathcal{O}(n^3)$.
- No need to know ε.
- Optimal (worst case sense) when $\mathcal{M}$ is a cylinder centered in V [Binev et al., 2017].

Cons

- Requires $n \leq m$.
- Trade-off between ε and $\mu(V, W)$.
- Limited by the Kolmogorov m-width, $\varepsilon \geq d_m(\mathcal{M})$ where

$$d_n(\mathcal{M}) := \inf_{\substack{X \subset U \\ \dim X = n}} \max_{u \in \mathcal{M}} \operatorname{dist}(X, u).$$

Multi-space approach

Library-based MOR

Multi-space: library-based MOR

- Consider a library of spaces $\mathcal{L}_n^N := \{V_1, \cdots, V_N\}$

$$\text{dist}\Big(\bigcup_{k=1}^{N} V_k, \mathcal{M}\Big) \leq \varepsilon, \qquad \text{dim}(V_k) \leq n \leq m.$$

Multi-space: library-based MOR

- Consider a library of spaces $\mathcal{L}_n^N := \{V_1, \cdots, V_N\}$

$$\mathrm{dist}\Big(\bigcup_{k=1}^{N} V_k, \mathcal{M}\Big) \leq \varepsilon, \qquad \mathrm{dim}(V_k) \leq n \leq m.$$

- New benchmark: non-linear Kolmogorov (n, N)-width [Temlyakov, 1998]

$$d_n(\mathcal{M}, N) := \inf_{\#\mathcal{L}_n^N = N} \max_{u \in \mathcal{M}} \min_{V \in \mathcal{L}_n^N} \mathrm{dist}(V, u),$$

which is expected to decay much faster than $d_n(\mathcal{M})$.

Multi-space: library-based MOR

- Consider a library of spaces $\mathcal{L}_n^N := \{V_1, \cdots, V_N\}$

$$\text{dist}\Big(\bigcup_{k=1}^{N} V_k, \mathcal{M}\Big) \leq \varepsilon, \qquad \text{dim}(V_k) \leq n \leq m.$$

- New benchmark: non-linear Kolmogorov (n, N)-width [Temlyakov, 1998]

$$d_n(\mathcal{M}, N) := \inf_{\#\mathcal{L}_n^N = N} \max_{u \in \mathcal{M}} \min_{V \in \mathcal{L}_n^N} \text{dist}(V, u),$$

  which is expected to decay much faster than $d_n(\mathcal{M})$.
- Aim for low ε with a low n, thus better stability.

Multi-space: How ?

Each space $V \in \mathcal{L}_n^N$ gives a one-space estimate $A_V(w)$...

Multi-space: How ?

Each space $V \in \mathcal{L}_n^N$ gives a one-space estimate $A_V(w)$...

How to select $V^*(w)$ among this library ?

Multi-space: Selection

Idea [Cohen et al., 2022]: select the "closest" to $\mathcal{M}$

Multi-space: Selection

Idea [Cohen et al., 2022]: select the "closest" to $\mathcal{M}$

- Assume we have $\mathcal{S}$ such that for any $v \in U$,

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

Multi-space: Selection

Idea [Cohen et al., 2022]: select the "closest" to $\mathcal{M}$

- Assume we have $\mathcal{S}$ such that for any $v \in U$,

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

- Select $V^*(w)$ as

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

Multi-space: Selection

Proposition (Near optimal selection [Cohen et al., 2022])

Assuming that P_W is injective on $\mathcal{M}$ and that $\mu(\mathcal{M}, W) < \infty$,

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

$\mu(\mathcal{M}, W)$ reflects how well $\mathcal{M}$ and W are aligned.

Dictionary approach

Dictionary-based MOR

Dictionary approach: Library

- Dictionary of K vectors (or snapshots),

$$\mathcal{D}_K = \{v^{(1)}, \cdots, v^{(K)}\}$$

Dictionary approach: Library

- Dictionary of K vectors (or snapshots),

$$\mathcal{D}_K = \{v^{(1)}, \cdots, v^{(K)}\}$$

- Take as library $\mathcal{L}_m^N = \mathcal{L}_m(\mathcal{D}_K)$, containing all the subspaces spanned by at most m vectors from $\mathcal{D}_K$,

$$\mathcal{L}_m(\mathcal{D}_K) := \Big\{ \sum_{k=1}^{K} x_k v^{(k)} : \; x \in \mathbb{R}^K, \; \|x\|_0 \leq m \Big\},$$

Dictionary approach: Library

$\mathcal{L}_m(\mathcal{D}_K)$ is **large** $\longrightarrow$ low ε but not fully explorable.

Dictionary approach: Compressive sensing

Sparse approximation given a few linear measurements

Dictionary approach: Compressive sensing

Sparse approximation given a few linear measurements $\rightarrow$ Compressive Sensing

Dictionary approach: Compressive sensing

Sparse approximation given a few linear measurements $\rightarrow$ Compressive Sensing

- Consider the Basis Pursuit Denoising problem for $\alpha > 0$

$$\mathbf{x}_\alpha(w) := \underset{\mathbf{x}\in\mathbb{R}^K}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{C}\mathbf{x} - \mathbf{w}\|_2^2 + \alpha\|\mathbf{x}\|_1,$$

whose solution is unique and m-sparse under some assumptions.

Dictionary approach: Compressive sensing

Sparse approximation given a few linear measurements $\rightarrow$ Compressive Sensing

- Consider the Basis Pursuit Denoising problem for $\alpha > 0$

$$\mathbf{x}_\alpha(w) := \underset{\mathbf{x}\in\mathbb{R}^K}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{C}\mathbf{x} - \mathbf{w}\|_2^2 + \alpha\|\mathbf{x}\|_1,$$

whose solution is unique and m-sparse under some assumptions.

- Focus on the space spanned by the support,

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

Dictionary approach: Selection

- Use $\mathcal{S}$ to select $V_{\mathcal{S}}(w)$ among the **smaller** library $\mathcal{L}(w) := \{V_\alpha(w) : \alpha > 0\}$ and define

$$A_{\mathcal{S}}^{\mathsf{dic}}(w) := A_{V_{\mathcal{S}(w)}}(w)$$

Dictionary approach: Selection

- Use $\mathcal{S}$ to select $V_{\mathcal{S}}(w)$ among the **smaller** library $\mathcal{L}(w) := \{V_\alpha(w) : \alpha > 0\}$ and define

$$A_{\mathcal{S}}^{\mathsf{dic}}(w) := A_{V_{\mathcal{S}}(w)}(w)$$

- In practice for efficiency and numerical stability reasons, we only compute $\hat{\mathcal{L}}(w) \subset \mathcal{L}(w)$ with $\#\hat{\mathcal{L}}(w) \leq \tau K$.

Dictionary approach: Selection

- Use $\mathcal{S}$ to select $V_{\mathcal{S}}(w)$ among the **smaller** library $\mathcal{L}(w) := \{V_\alpha(w) : \alpha > 0\}$ and define

$$A_{\mathcal{S}}^{\text{dic}}(w) := A_{V_{\mathcal{S}}(w)}(w)$$

- In practice for efficiency and numerical stability reasons, we only compute $\hat{\mathcal{L}}(w) \subset \mathcal{L}(w)$ with $\#\hat{\mathcal{L}}(w) \leq \tau K$.

Proposition (Near optimal selection)

Assuming that P_W is injective on $\mathcal{M}$ and that $\mu(\mathcal{M}, W) < \infty$,

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

Parameterized PDEs

Offline-online decomposition for dictionary-based multi-space

Parameterized PDEs: Offline-online decomposition

- Discrete framework $U = \mathbb{R}^{\mathcal{N}}$ with large $\mathcal{N}$.
- **Offline**: Heavy pre-computations independently on w.
- **Online**: Fast computation of $A_S^{\text{dic}}(w)$.

Parameterized PDEs: Framework

- Consider that u solves the operator equation

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

where the singular values of $B(\xi)$ are **uniformly bounded**,

$$0 < c \leq \min_{v \in U} \frac{\|B(\xi)v\|}{\|v\|} \leq \max_{v \in U} \frac{\|B(\xi)v\|}{\|v\|} \leq C < \infty,$$

Parameterized PDEs: Framework

- Consider that u solves the operator equation

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

where the singular values of $B(\xi)$ are **uniformly bounded**,

$$0 < c \leq \min_{v \in U} \frac{\|B(\xi)v\|}{\|v\|} \leq \max_{v \in U} \frac{\|B(\xi)v\|}{\|v\|} \leq C < \infty,$$

- Take $\mathcal{S}$ as a **residual norm** as in [Cohen et al., 2022],

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

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

Parameterized PDEs: Affine decomposition

- Assume the **affine decompositions**

$$B(\xi) = B_0 + \sum_{q=1}^{d} \theta_q^{(B)}(\xi) B_q \quad \text{and} \quad f(\xi) = f_0 + \sum_{q=1}^{m_f} \theta_q^{(f)}(\xi) f_q,$$

with $B_q : U \longrightarrow U'$ linear operators and $f_q \in U'$.

Parameterized PDEs: Affine decomposition

- Assume the **affine decompositions**

$$B(\xi) = B_0 + \sum_{q=1}^{d} \theta_q^{(B)}(\xi) B_q \quad \text{and} \quad f(\xi) = f_0 + \sum_{q=1}^{m_f} \theta_q^{(f)}(\xi) f_q,$$

with $B_q : U \longrightarrow U'$ linear operators and $f_q \in U'$.

- Then computing $\mathcal{S}$ requires solving the l.s. system

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

where $\theta(\xi) \in \mathbb{R}^{m_B+m_f}$, $G(v) : \mathbb{R}^{m_B+m_f} \longrightarrow U'$ and $g(v) \in U'$.

Parameterized PDEs: Offline-online decomposition

- **Problem**: precomputing the normal equation for $\mathcal{S}$ costs

$$\mathcal{O}\big((m_B K + m_f)^2 \mathcal{N}\big)$$

Parameterized PDEs: Offline-online decomposition

- **Problem**: precomputing the normal equation for $\mathcal{S}$ costs

$$\mathcal{O}\big((m_B K + m_f)^2 \mathcal{N}\big)$$

- **Instead**, we use a random embedding $\Theta \in \mathbb{R}^{k \times \mathcal{N}}$ and consider

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

where $G^{\Theta}(v) \in \mathbb{R}^{k \times (m_B + m_f)}$ and $g^{\Theta}(v) \in \mathbb{R}^{k}$.

Parameterized PDEs: Offline-online decomposition

- **Problem**: precomputing the normal equation for $\mathcal{S}$ costs

$$\mathcal{O}\Big((m_B K + m_f)^2 \mathcal{N}\Big)$$

- **Instead**, we use a random embedding $\Theta \in \mathbb{R}^{k \times \mathcal{N}}$ and consider

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

where $G^{\Theta}(v) \in \mathbb{R}^{k \times (m_B + m_f)}$ and $g^{\Theta}(v) \in \mathbb{R}^k$.

Proposition

With $k = \mathcal{O}\Big(\epsilon^{-2}\ \big(m_B + m_f + \log(\delta^{-1})\big)\Big)$, *for any* $v \in U$, *with probability at least* $1 - \delta$ *we have*

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

Parameterized PDEs: Computational aspects

- **Offline cost**: Using a structured embedding Θ (e.g. SRHT),

$$\mathcal{O}\Big(\underbrace{mK\mathcal{N}}_{\text{compute } \mathbf{C}} + \underbrace{(m_B K + m_f)\mathcal{N}\log(k)}_{\text{pre-compute } G^{\Theta}(v) \text{ and } g^{\Theta}(v)} \Big)$$

- **Online cost**: Considering $k = \mathcal{O}(m_B + m_f)$,

$$\mathcal{O}\Big(\underbrace{m^2K}_{\text{LARS}} + \underbrace{k(mm_B + m_f)K}_{\text{prepare l.s.}} + \underbrace{C_{ls}K}_{\text{solve l.s.}} \Big)$$

- **Numerical stability** : few affine terms, thus robust to round-off errors.

Parameterized PDEs: Computational aspects

Example when $\theta(\xi) = \xi$ and $f(\xi) = f_0$.

- **Offline cost**:

With random sketching $\mathcal{O}\big(mK\mathcal{N} + dK\mathcal{N}\log(k)\big)$ VS With normal equation $\mathcal{O}\big(mK\mathcal{N} + d^2K^2\mathcal{N}\big)$

- **Online cost**: Considering $k = \mathcal{O}(d)$,

With random sketching $\mathcal{O}\big(m^2K + md^2K + d^3K\big)$ VS With normal equation $\mathcal{O}\big(m^2K + m^2d^2K + d^3K\big)$

Numerical Example

Finite Element space $U \subset \mathcal{H}^1_{\Gamma_D}(\Omega)$ endowed with norm $\|\nabla \cdot\|_{\mathcal{L}^2(\Omega)}$

Numerical: Thermal block

$$\mathcal{N} \sim 8\,000 \quad \text{and} \quad \begin{cases} -\nabla \cdot (\kappa \nabla u) = 1 & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma_D, \\ \kappa = \xi_i & \text{in } \Omega_i,\ 1 \le i \le 9, \end{cases}$$

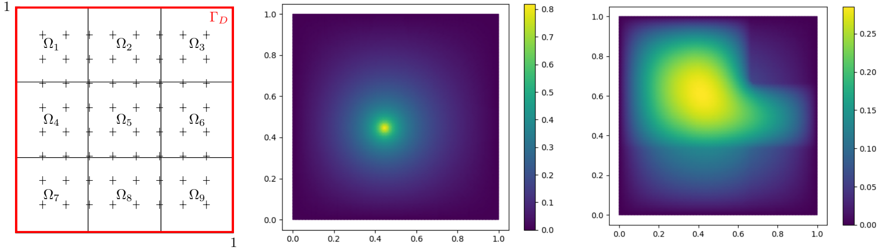


Figure: Left: geometry, with sensors locations (crosses). Middle: Riesz representative of a sensor. Right: a snapshot.

Numerical: Thermal block

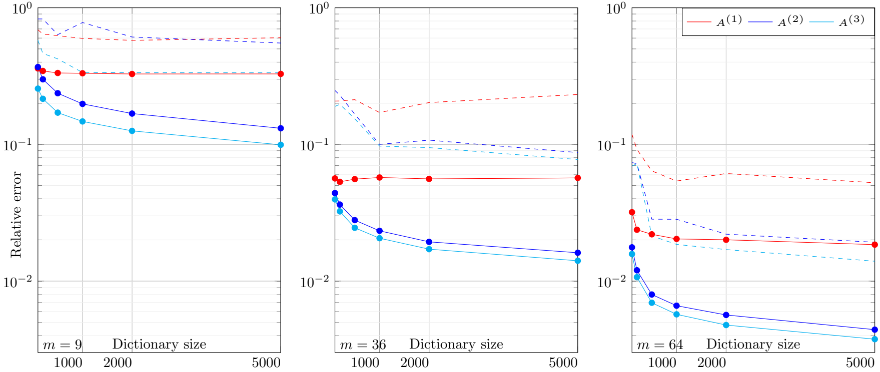
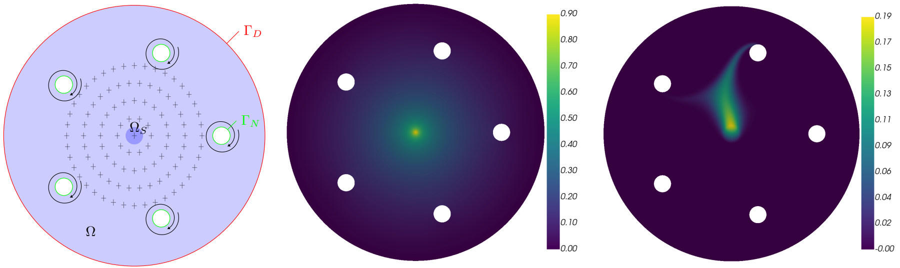


Figure: Evolution of the recovery errors in U-norm, on 500 test snapshots, with growing dictionary sizes K, for $m \in \{9, 36, 64\}$.

Numerical: Advection diffusion

$$\mathcal{N} \sim 150\ 000 \quad \text{and} \quad \begin{cases} -0.01\Delta u + \mathcal{V}(\xi)\cdot\nabla u = \dfrac{100}{\pi}\mathbb{1}_{\Omega_S} & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma_D, \\ n\cdot\nabla u = 0 & \text{on } \Gamma_N, \end{cases}$$

$$\mathcal{V}(\xi) = \sum_{i=1}^{5} \frac{1}{\|x - x^{(i)}\|}\left(\xi_i e_r(x^{(i)}) + \xi_{i+5} e_\theta(x^{(i)})\right)$$



Numerical: Advection diffusion

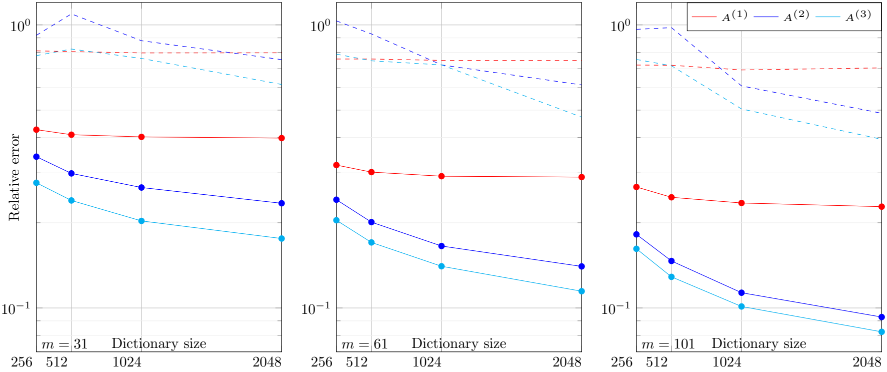


Figure: Evolution of the recovery errors in U-norm, on 500 test snapshots, with growing dictionary sizes K, for $m \in \{31, 61, 101\}$.

Conclusion

Conclusion

- Dictionary-based multi-space approach for state estimation
- Efficient offline-online decomposition using randomized linear algebra.
- More details in [Nouy and Pasco, 2023] available at https://arxiv.org/abs/2303.10771
- Python repo available at https://github.com/alexandre-pasco/rla4mor/inverse-problems

Thank you !

Appendix

Appendix: Offline stage

- Write $A_\alpha(w) = \mathbf{U}\mathbf{a}$ with $\mathbf{a} = \mathbf{a}(w) \in \mathbb{R}^{m+K}$ and $\mathbf{U} = \big(\mathbf{W} \mid \mathbf{V}\big)$.
- Write the l.s. terms to compute $\mathcal{S}^\Theta$ as

$$G^\Theta(A_\alpha(w)) := \Big(\Theta B^{(1)}\mathbf{U}\mathbf{a} \mid \cdots \mid \Theta B^{(d)}\mathbf{U}\mathbf{a}\Big)$$
$$g^\Theta(A_\alpha(w)) := \Theta f - \Theta B^{(0)}\mathbf{U}\mathbf{a}.$$

- **Offline**: compute $\underbrace{\mathbf{C}}_{\mathcal{O}(mK\mathcal{N})} \in \mathbb{R}^{m\times K}$ and $\underbrace{\Theta B^{(q)}\mathbf{U}}_{\mathcal{O}(K\mathcal{N}\log(\|))} \in \mathbb{R}^{k\times(m+K)}$
- Total offline cost (without snapshot computation) is

$$\mathcal{O}\Big((d\log(\mathcal{N}) + m)K\mathcal{N}\Big) \quad \text{VS} \quad \mathcal{O}\Big(d^2K^2\mathcal{N}\Big)$$

Appendix: Online stage

- **Step 0**: Observe $w = P_w u$.
- **Step 1**: Run LARS to generate $\mathcal{O}(K)$ subspaces, costing $\mathcal{O}(m^2 K)$.
- **Step 2**: Prepare the l.s. systems, each costing $\mathcal{O}(kmd)$.
- **Step 3**: Solve them, each costing $\mathcal{O}(kd^2)$.
- Total online cost with $k = \mathcal{O}(d)$

$$\mathcal{O}\big(\underbrace{m^2K}_{\text{LARS}} + \underbrace{md^2K}_{\text{prepare l.s.}} + \underbrace{d^3K}_{\text{solve l.s.}}\big) = \mathcal{O}\big((m^2 + md^2 + d^3)K\big).$$

References I

Binev, P., Cohen, A., Dahmen, W., DeVore, R., Petrova, G., and Wojtaszczyk, P. (2017).
Data Assimilation in Reduced Modeling.
SIAM/ASA J. Uncertainty Quantification, 5(1):1–29.

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

Maday, Y., Patera, A. T., Penn, J. D., and Yano, M. (2015).
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.

Nouy, A. and Pasco, A. (2023).
Dictionary-based model reduction for state estimation.

References II

Temlyakov, V. N. (1998).
Nonlinear Kolmogorov widths.
Math Notes, 63(6):785–795.