Greedy approximation with data assimilation

James Ashton Nichols

April 24, 2018

Laboratoire Jacques-Louis Lions, Sorbonne Université

With Albert Cohen, Olga Mula, Peter Binev

Our setting

We are motivated by elliptic parametric PDEs on a domain $D \subset \mathbb{R}^d$ of the general form

$$\mathcal{P}(u, y) = 0,$$

where $u(x, y)$ is the solution, $y \in \mathcal{Y}$ is a (high-dimensional / stochastic) parameter. We typically write $u(y)$.

In a wide variety of problems it is known that the solution map

$$y \mapsto u(y)$$

is well-defined and smooth w.r.t. $\|\cdot\|_{\mathcal{V}}$, hence the solution manifold

$$\mathcal{M} := \{u(y) \ : \ y \in \mathcal{Y}\} \subset \mathcal{V}.$$

$\mathcal{M}$ is typically smooth and compact. $\mathcal{V}$ is our ambient Hilbert-space e.g. $H_0^1(D)$ with $D \subset \mathbb{R}^d$.

Our goal:
Given that y is unknown (and likely unknowable), how do we approximate u with minimal physical measurements?

In our setting we can take m independent measurements

$$\ell_i(u), \quad i = 1, \dots, m, \quad u = u(a)$$

where the $\ell_i \in \mathcal{V}'$ are linear functionals on $\mathcal{V}$.

We know what ℓ_i are, we know the representers $\ell_i(u) = \langle u, \omega_i \rangle$. Define the measurement space

$$W_m := \mathrm{span}\{\omega_1, \dots, \omega_m\}.$$

Hence from the ℓ_i we can find

$$w = P_{W_m} u \in W_m$$

So we retain a low dimensional information on the complex manifold $\mathcal{M}$.

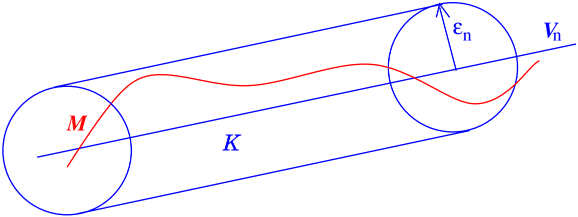
The ℓ_i represent real-world sensor / microphone response

In an *"offline"* computation we find nested subspaces

$$V_0 \subset V_1 \subset \cdots \subset V_n \subset \cdots, \quad \dim(V_n) = n,$$

that cover $\mathcal{M}$ to within $\varepsilon_0 > \varepsilon_1 > \cdots > 0$

$$\mathcal{M} \subset \mathcal{K}^n := \{v \in \mathcal{V} \,:\, \mathrm{dist}(v, V_n) \leq \varepsilon_n\}$$



E.g.

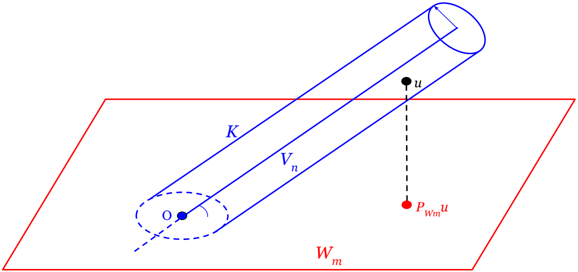
- Sparse polynomials: $u(y) \approx \sum_{\nu \in \Lambda_n} u_\nu y^\nu \in V_n := \mathrm{span}\{u_\nu \,:\, \nu \in \Lambda_n\}$.
- Reduced bases: $V_n := \mathrm{span}\{u_i \,:\, i = 1, \ldots, n\}$ with $u_i = u(y_i)$ snapshots.
- Fourier / wavelet bases etc...

Acoustic holography

A is our *lifting* from observation $w = P_{W_m} u$ to best fit point in $\mathcal{K}^n$. Evidently we require that $P_{W_m} A(w) = w$ hence $A(w) \in w + W_m^\perp$

$$\mathcal{K}_w^n := \{ v \in \mathcal{K} \ : \ P_{W_m} v = w \} = \mathcal{K}^n \cap (w + W_m^\perp)$$

which is an ellipsoid: intersection of the cylinder $\mathcal{K}^n$ with affine space $w + W_m^\perp$.



[Maday et al., 2015] : take

$$A(w) = \operatorname{argmin} \{ \operatorname{dist}(u, V_n) : u \in \mathcal{K}_w^n \} .$$

A is linear. $A(w)$ coincides with the center of the ellipsoid $\mathcal{K}_w^n$, hence is an optimal recovery algorithm [Binev et al., 2017]

Error estimates

The error of the optimal recovery algorithm is

$$\sup_{u \in \mathcal{K}^n} \|u - A(P_{W_m} u)\| = \frac{\varepsilon_n}{\beta(V_n, W_m)}$$

Based on the inf-sup constant (the "angle" between V_n and W_m)

$$\beta(V_n, W_m) := \inf_{v \in V_n} \frac{\|P_{W_m} v\|}{\|v\|} \in [0, 1].$$

see also [Adcock and Hansen, 2012].

- $\beta(V_n, W_m) = 1$ if and only if $V_n \subset W_m$.
- We require that $n < m$, otherwise $\beta(V_n, W_m) = 0$ (as $V_n \cap W_m^\perp \neq \{0\}$).
- In general $\beta(V_n, W_m)$ and ε_n decrease as n grows... (hence stability issues)

Optimal measurement selection

- The sensors ℓ_i (or ω_i) are usually selected from a set $\mathcal{D}$ (the *dictionary*).
- If we are given a fixed budget m, what is the best choice?
- **Task:** Given given a fixed V_n, for some goal $\beta^* > 0$ find $\omega_1, \dots, \omega_m$ from $\mathcal{D}$ such that

$$\beta(V_n, W_m) \geq \beta^* > 0,$$

  with a number of measurements $m \geq n$ as small possible.

Benchmark: $m^*(\beta)$ the smallest value of $m \geq n$ such that such a selection exists (sometimes theoretically available).

Nonlinear approximation/Data-driven reduced models.
Ongoing work with A. Cohen and O. Mula.

- So far, V_n is tailored for the whole manifold $\mathcal{M}$.
- For given measurements, $\ell_1(u), \dots, \ell_m(u)$, can we find a reduced model V_n^{data} that performs better than V_n for the reconstruction of u?

Optimal measurement selection

Goal: Given $\beta^* > 0$ and given V_n, find $\omega_1, \dots, \omega_m$ from $\mathcal{D}$ such that

$$\beta(V_n, W_m) \geq \beta^* > 0,$$

with a number of measurements $m \geq n$ as small possible.

Dictionary: We pick the ω_i which span W_m from a *dictionary* $\mathcal{D}$ of $\mathcal{V}$, that is,

$$\overline{\mathcal{D}} = V \quad \text{and} \quad \|\omega\| = 1, \ \forall \omega \in \mathcal{D}.$$

The ω may represent response characteristics of real-world sensors or microphones

Examples of dictionaries: If $\mathcal{V} = H_0^1(D)$ with $D \subset \mathbb{R}^d$

- Pointwise evaluations: $\mathcal{D} = \{\ell_x \,:\, x \in D\}$, $\ell_x(f) = f(x)$ (when $d = 1$)
- Local averages: for a fixed $\epsilon > 0$, $\mathcal{D}_\epsilon = \{\ell_{x,\epsilon} \,:\, x \in D\}$ where

$$\ell_{x,\epsilon}(u) := \int_D u(y)\varphi_\epsilon(y-x)dy, \quad \varphi_\epsilon(y) := \epsilon^{-d}\varphi\left(\frac{y}{\epsilon}\right), \quad \varphi \text{ is a unit mollifier}$$

A greedy approach

Optimising W_m simultaneously over $\mathcal{D}$ is not an option. So we go greedy.

[Binev et al., 2018]: greedy *orthogonal matching pursuit* (OMP) type algorithms for selecting ω_i from $\mathcal{D}$ for the collective approximation of the elements of V_n.

We define **two** algorithms. Assume $V_n = \mathrm{span}\{\phi_1, \dots, \phi_n\}$ is orthonormal. Having selected $\{\omega_1, \dots, \omega_m\}$ and with $W_m = \mathrm{span}\{\omega_1, \dots, \omega_m\}$, we define:

Collective OMP:

$$\omega_{m+1} := \mathrm{argmax}\left\{ \sum_{j=1}^{n} |\langle \phi_j - P_{W_m}\phi_j, \omega\rangle|^2 \ : \ \omega \in \mathcal{D} \right\}$$

Worst-case OMP:

$$v_{m+1} := \mathrm{argmax}\left\{ \|v - P_{W_m} v\| : v \in V_n, \ \|v\| = 1 \right\}$$

then

$$\omega_{m+1} := \mathrm{argmax}\left\{ |\langle v_{m+1} - P_{W_m} v_{m+1}, \omega\rangle| : \ \omega \in \mathcal{D} \right\}$$

Convergence rates

Converge (loose statement) [Binev et al., 2018]

For the sequence of measurement spaces W_m built with either greedy algorithm, we have

$$\beta(V_n, W_m) \geq \left(1 - \frac{C}{m+1}\right)^{1/2}$$

The constant C is more favorable in the collective OMP but in our numerical experiments the worst case OMP performed better.

Proof ideas: We take any orthonormal basis $\Phi = (\phi_1, \dots, \phi_n)$ of V_n and introduce the residual quantity

$$r_m^2 := \sum_{i=1}^{n} \|\phi_i - P_{W_m}\phi_i\|^2.$$

which is such that $\beta(V_n, W_m)^2 \geq 1 - r_m^2$. We can derive convergence rates for $(r_m^2)_{m\geq 1}$.

We will satisfy $\beta(V_n, W_m) \geq \beta^*$ as soon as $r_m^2 \leq 1 - (\beta^*)^2$.

Convergence rates

We introduce for any $\Phi = (\phi_1, \dots, \phi_n) \in \mathcal{V}^n$ the quantity

$$\|\Phi\|_{\ell^1(\mathcal{D})} := \inf_{c_{\omega,i}} \left\{ \sum_{\omega\in\mathcal{D}} \left(\sum_{i=1}^n |c_{\omega,i}|^2 \right)^{1/2} : \ \phi_i = \sum_{\omega\in\mathcal{D}} c_{\omega,i}\omega, \quad i = 1, \dots, n \right\}.$$

(similar to the typical approximation spaces $\mathcal{A}^1(\mathcal{D})$, but for a basis)

Convergence of (r_k) in the collective OMP [Binev et al., 2018]

Let $\Phi = (\phi_1, \dots, \phi_n)$ be an orthonormal basis of V_n with finite $\|\Phi\|_{\ell^1(\mathcal{D})}$. Then, we have for the collective OMP

$$r_m^2 \leq \frac{\|\Phi\|^2_{\ell^1(\mathcal{D})}}{\kappa^2(m+1)}.$$

and for the worst case OMP,

$$r_m^2 \leq \frac{n^2\|\Phi\|^2_{\ell^1(\mathcal{D})}}{\kappa^2(m+1)}.$$

A few notes:

- This can be extended to *any* basis $\Psi = (\psi_1, \dots, \psi_n)$ of V_n, can thus show that it applies to all of V_n
- κ is included because in practice we have a large finite but incomplete dictionary $\mathcal{D}_N \subset \mathcal{D}$ and some residual from the Galerkin projection etc... so our theory is not for $\omega_{m+1} = \mathrm{argmax}(\dots)$ but in fact ω_{m+1} satisfies Collective OMP:

$$\sum_{j=1}^{n} |\langle \phi_j - P_{W_m}\phi_j, \omega_{m+1}\rangle|^2 \geq \kappa \max\left\{ \sum_{j=1}^{n} |\langle \phi_j - P_{W_m}\phi_j, \omega\rangle|^2 \; : \; \omega \in \mathcal{D} \right\}$$

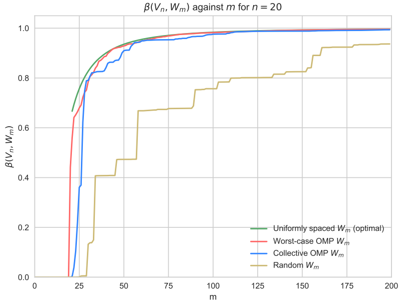
Wost-case OMP

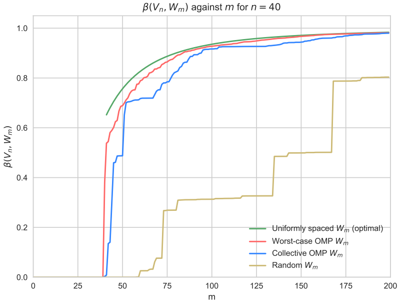
$$|\langle v_{m+1} - P_{W_m} v_{m+1}, \omega_{m+1}\rangle| \geq \kappa \max\left\{ |\langle v_{m+1} - P_{W_m} v_{m+1}, \omega\rangle| : \; \omega \in \mathcal{D} \right\}$$

- **Ambient space:** $\mathcal{V} = H_0^1(]0,1[)$
- **Reduced model:** $V_n = \mathrm{span}\left\{ \frac{\sqrt{2}}{\pi k} \sin(k\pi x) \right\}_{k=1}^{n}$
- **Dictionary of pointwise evaluation:** $\mathcal{D} = \{\ell_x : x \in]0,1[\}$, $\ell_x(f) = f(x)$.
- In this case we know equispaced evaluation points are optimal:

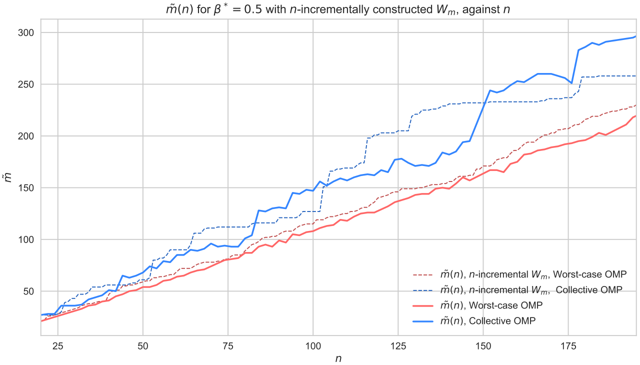
$$W_m^{\mathrm{opt}} = \left\{ \omega_x \ : \ x \in \left\{ \frac{1}{m+1}, \dots, \frac{m}{m+1} \right\} \right\}$$

But the greedy algo cannot choose equispaced points at every step.





What is the minimum m to get $\beta(V_n, W_m) > \beta^* > 0$? We see it is almost linear with n.



Open problem: can a greedy algorithm achieve some fixed lower bound β with a number of measurements $m(\beta)$ of comparable size as $m^*(\beta)$?

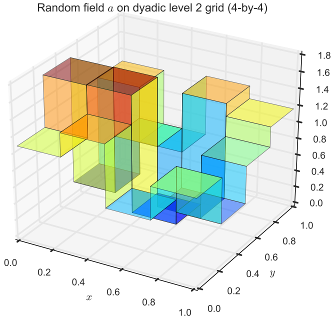
E.g.

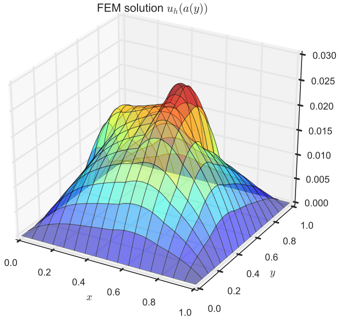
$$-\mathrm{div}(a\nabla u) = 1 \quad \text{on} \quad [0,1]^2$$

with $u_{|\partial D} = 0$,

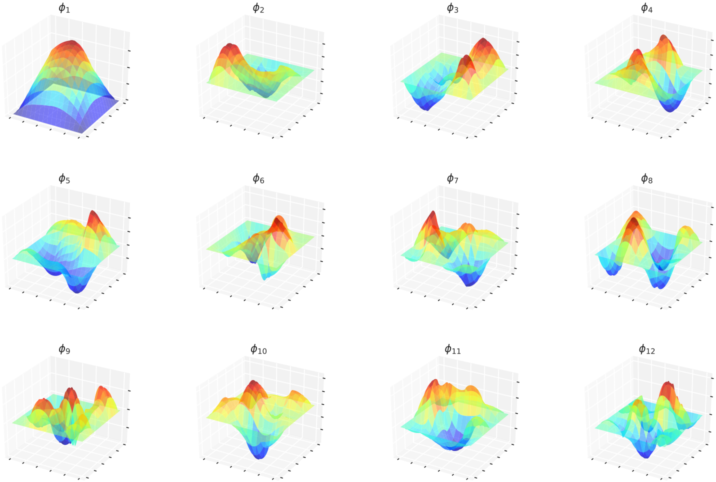
$$a = a(y) = 1 + 0.9\sum_{j=1}^{16} y_j \chi_{D_j}, \quad y = (y_j) \in [-1,1]^{16}.$$

We solve using FEM





We can produce a random sequence of $y^{(i)} \in [-1,1]^{16}$ for $y = 1, \dots, n$ then calculate snapshots $u(y^{(i)})$ and orthonormalise them to produce V_n:

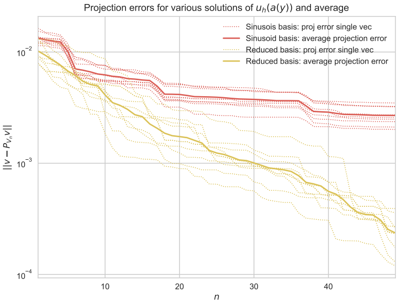


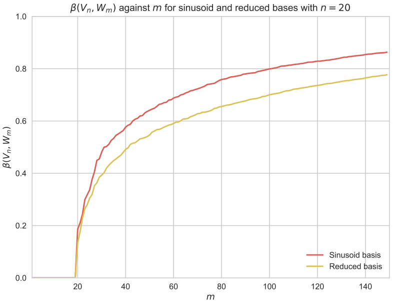
Using a dictionary $\mathcal{D}$ of local averages $\ell_{x_0,\varepsilon}$ for any $x_0 \in \mathcal{D}$ we perform WC-OMP greedy algo:

Recall

$$\sup_{u \in \mathcal{K}^n} \|u - A(P_{W_m} u)\| = \beta^{-1}(V_n, W_m)\, \varepsilon_n$$

Reduced basis V_n has smaller avg. projection error than a sinusoid basis...





...but slightly worse $\beta(W_m, V_n)$

Goal:

- Now W_m **and** $w = P_{W_m} u$ **are given**
- We look for a good $V_n = \mathrm{span}\{\phi_i\}_{i=1}^n$ to reconstruct u.
- The ϕ_i are sought in a set $\mathcal{D} \subset V$ e.g. set of snapshots from $\mathcal{M}$, ideally $\mathcal{M} \subseteq \overline{\mathcal{D}}$

Greedy algorithms to build V_n:

1. **Pure greedy (not data driven):**
   For $k = 1$, we choose
   $$\phi_1 = \mathrm{argmax}_{v \in \mathcal{D}} ||v||$$
   and set $V_1 := \mathrm{span}\{\phi_1\}$.
   For $n > 1$, given $V_n = \mathrm{span}\{\phi_1, \dots, \phi_n\}$, we look for
   $$\phi_{n+1} = \mathrm{argmax}_{v \in \mathcal{D}} ||v - P_{V_n} v||$$
   and set $V_{N+1} := \mathrm{span}\{V_n, \phi_{n+1}\}$.

Data-driven greedy algorithms for V_n

1. **Measurement based OMP:** For $n = 1$, we choose

$$\phi_1 = \mathrm{argmax}_{v\in\mathcal{D}} \langle w, v\rangle$$

and set $V_1 := \mathrm{span}\{\phi_1\}$.
For $n > 1$, given $V_n = \mathrm{span}\{\phi_1, \dots, \phi_n\}$, we look for

$$\phi_{n+1} = \mathrm{argmax}_{v\in\mathcal{D}} \left\langle w - P_{P_{W_m}V_n} w, \frac{P_{W_m} v}{\|P_{W_m} v\|} \right\rangle$$

and set $V_{n+1} := \mathrm{span}\{V_n, \phi_{n+1}\}$.

2. **Measurement based Projection Pursuit:** For $n = 1$, we choose

$$\phi_1 = \mathrm{argmin}_{v\in\mathcal{D}} ||w - P_{P_{W_m}(v)}(w)||$$

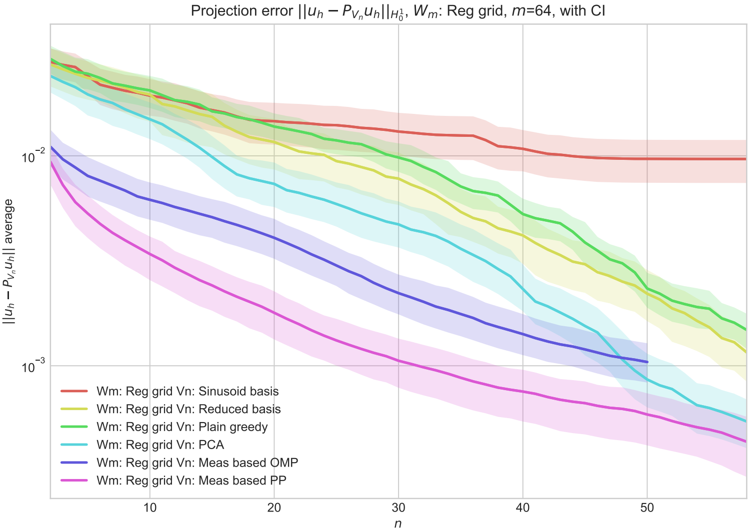
and set $V_1 := \mathrm{span}\{\phi_1\}$.
For $n > 1$, given $V_n = \mathrm{span}\{\phi_1, \dots, \phi_n\}$, we look for

$$\phi_{n+1} = \mathrm{argmin}_{v\in\mathcal{D}} ||w - P_{P_{W_m}(V_n\oplus\mathbb{R}v)} w||$$

and set $V_{n+1} := \mathrm{span}\{V_n, \phi_{n+1}\}$.

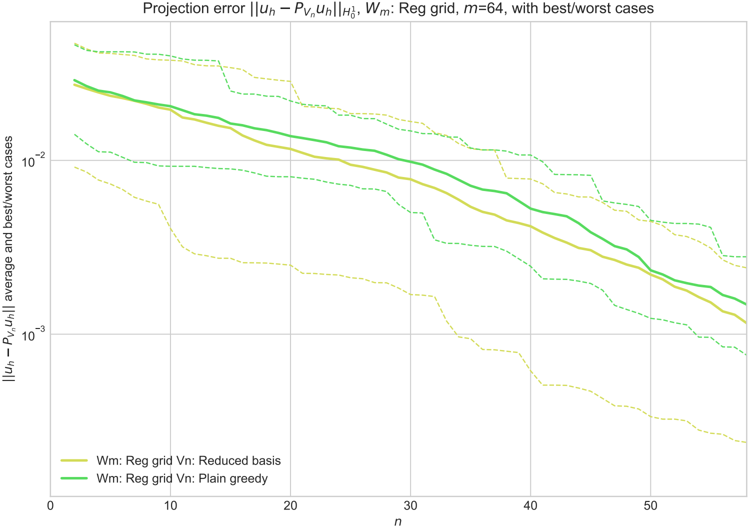
*** Both these algorithms operate in $\mathbb{R}^m$ hence relatively cheap ***

With $m = 64$ evenly spaced measurements in $[0,1]^2$



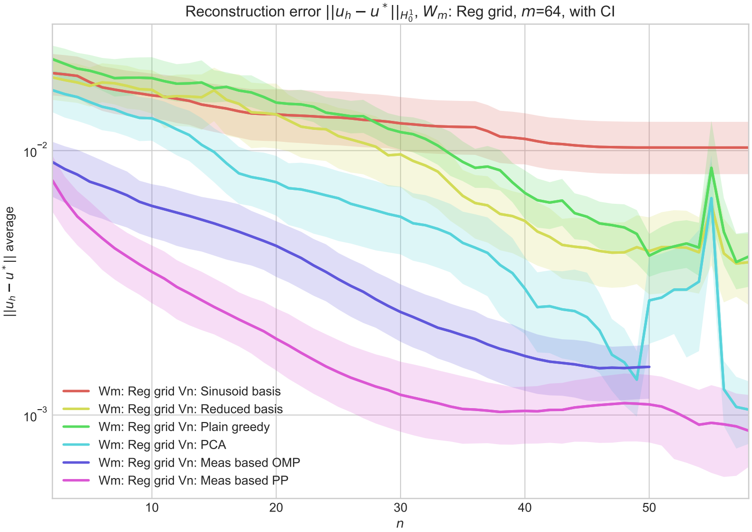
Projection error on checkerboard elliptic problem

With $m = 64$ evenly spaced measurements in $[0,1]^2$

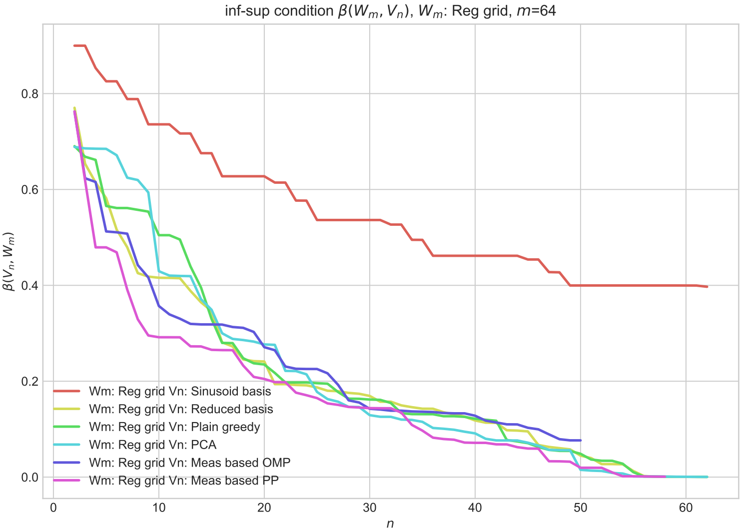


Reconstruction error on checkerboard elliptic problem

With $m = 64$ evenly spaced measurements in $[0,1]^2$



With $m = 64$ evenly spaced measurements in $[0, 1]^2$



Conclusions

Conclusions: In the well-known setting of state estimation with measurements and reduced models we

- For a given approximation space V_n perform measurement selection to build W_m with greedy algorithms.
- For W_m and $\ell_i(u)$ given, we built data based reduced models. Better accuracy than non data driven models in numerical experiments. Difficult theoretical justification.

Future directions / open questions

- "Dictionary width"? Quantify deviation of greedy algorithm with optimal choice in general cases.
- Building $W_{m(n)}, W_{m(n+1)}, \ldots$ parallel with $V_n, V_{n+1}, \ldots$ incrementally. Any guarantees?
- Non-linear measurements, sensor failure, noisy data.
- Greedy $L_\infty(Y, V)$ opt. bases vs low-rank $L_2(Y, V)$ opt. bases
- Sparsity in V_n? Links to compressed sensing? Inverse estimates?

References I

Adcock, B. and Hansen, A. C. (2012).

A Generalized Sampling Theorem for Stable Reconstructions in Arbitrary Bases.

Journal of Fourier Analysis and Applications, 18(4):685–716.

Binev, P., Cohen, A., Dahmen, W., DeVore, R., Petrova, G., and Wojtaszczyk, P. (2017).

Data assimilation in reduced modeling.

SIAM/ASA Journal on Uncertainty Quantification, 5(1):1–29.

Binev, P., Cohen, A., Mula, O., and Nichols, J. (2018).

Greedy algorithms for optimal measurements selection in state estimation using reduced models.

submitted to SIAM/ASA Journal on Uncertainty Quantification.

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.

International Journal for Numerical Methods in Engineering, 102(5):933–965.