# Greedy measurement selection for state estimation

**P. Binev[1], A. Cohen[2], O. Mula[3] & J. Nichols[2]**

University of South Carolina[1], Université Pierre et Marie Curie[2], Université Paris Dauphine[3]

Contact: `mula@ceremade.dauphine.fr, nichols@ljll.math.upmc.fr`

## 1 State estimation from data in parametric PDEs

**Recovery problem:** find an element $u$ of a Hilbert space $V$ from $m$ observations

$$z_i := \ell_i(u), \quad i = 1, \dots, m.$$

The $\ell_i$ are independent continuous linear functionals on $V$, whose Riesz representers are $\omega_i \in V$. From the $z := (z_i)_{i=1,\dots,m}$ we can find $w = P_{W_m}u$, which is the projection on the space

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

Since there are infinitely many $v \in V$ such that $P_{W_m}v = w$, we have to impose additional conditions on u to determine the one we recover.

**Prior assumption:** We assume that $u$ is a solution to a parametric PDE of the form

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

where $\mathcal{P}$ is a differential operator, and $a$ is a parameter which lives in some set $\mathcal{A}$. For any $a \in \mathcal{A}$ there exists a unique solution $u(a)$. However, in the recovery problem we do not know the parameter $a$ that is associated with the particular solution $u$ that we measure. Therefore, our prior on $u$ is that it belongs to the solution manifold

$$\mathcal{M} := \{u(a) \ : \ a \in \mathcal{A}\}.$$

In general, $\mathcal{M}$ does not have a simple geometry but the PDE structure often allows us to derive good approximations of $\mathcal{M}$ by linear spaces $V_n$ of moderate dimension $n$. Therefore, one possible option is to replace $\mathcal{M}$ by the simpler prior class described by the cylinder

$$\mathcal{K} = \{v \in V \ : \ \text{dist}(v, V_n) \le \varepsilon_n\},$$

for some given $n$.

**Approximation:** In this setting, we can combine the observed measurements $z$ and $V_n$ to produce an approximation $u^*$ of $u$ up to accuracy (see [2, 1])

$$\|u - u^*\| \le \beta(V_n, W_m)^{-1} \,\text{dist}(u, V_n),$$

where

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

plays the role of a stability constant.

**Goal**

For a given $V_n$, we would like to find $W_m$ such that

$$\beta(V_n, W_m) \ge \beta^* > 0, \tag{1.1}$$

with a number of measurements $m \ge n$ as small possible. Since the $\ell_i$ (or the $\omega_i$) which form $W_m$ represent sensors in reality, we will pick them from a *dictionary* $\mathcal{D}$ of $V'$ (resp. of $V$).

## 2 Greedy algorithms

Given an approximation space $V_n$, we need a general strategy to select the measurements from $\mathcal{D}$. Here we analyze two greedy algorithms.

### 2.1 A Collective Orthogonal Matching Pursuit algorithm

For a given $V_n$ ($n \ge 1$), we iteratively select $\omega_k$ such that

$$\|P_{V_n}(\omega_k - P_{W_{k-1}}\omega_k)\| \ge \kappa \max_{\omega \in \mathcal{D}} \|P_{V_n}(\omega - P_{W_{k-1}}\omega)\|, \quad m \ge 1, \tag{C-OMP}$$

for some fixed $0 < \kappa \le 1$. Note that in the case $n = 1$, we obtain the original OMP algorithm applied to the basis function of $V_1$. For the practical implementation, we take an orthonormal basis $(\phi_1, \dots, \phi_n)$ of $V_n$. Then

$$\|P_{V_n}(\omega - P_{W_{k-1}}\omega)\|^2 = \sum_{i=1}^n |\langle \omega - P_{W_{k-1}}\omega, \phi_i\rangle|^2 = \sum_{i=1}^n |\langle \phi_i - P_{W_{k-1}}\phi_i, \omega\rangle|^2.$$

Note that the basis $(\phi_1, \dots, \phi_n)$ is used only for the implementation, the definition of the greedy selection is independent of the choice of this basis in view of (C-OMP). For such a basis, we introduce the residual quantity

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

Since

$$\beta(V_n, W_k)^2 \ge 1 - r_k,$$

the inequality (1.1) holds provided that $r_k \le 1 - (\beta^*)^2$. To derive a convergence rate for the sequence $(r_k)_{k \ge 1}$, we introduce for any $\Psi = (\psi_1, \dots, \psi_n) \in V^n$ the quantity

$$\|\Psi\|_{\ell^1(\mathcal{D})} := \inf\Big\{ \sum_{\omega \in \mathcal{D}} \Big(\sum_{i=1}^n |c_{\omega,i}|^2\Big)^{1/2} \ : \ \psi_i = \sum_{\omega \in \mathcal{D}} c_{\omega,i}\omega, \quad i = 1, \dots, n\Big\}.$$

**Theorem C-OMP**

Let $\Phi = (\phi_1, \dots, \phi_n)$ be an orthonormal basis of $V_n$ and $\Psi = (\psi_1, \dots, \psi_n) \in V^n$ be arbitrary. Then the application of the Collective OMP algorithm on the space $V_n$ gives

$$r_k \le 4 \frac{\|\Psi\|^2_{\ell^1(\mathcal{D})}}{\kappa^2}(k+1)^{-1} + \|\Phi - \Psi\|^2, \quad k \ge 1. \tag{2.1}$$

where $\|\Phi - \Psi\|^2 := \|\Phi - \Psi\|^2_{V^n} = \sum_{i=1}^n \|\phi_i - \psi_i\|^2$.

### 2.2 A Worst Case OMP algorithm

In this algorithm, at each step we take

$$v_k := \text{argmax}\Big\{\|v - P_{W_{k-1}}v\| \ : \ v \in V_n, \ \|v\| = 1\Big\},$$

which is the vector in the unit ball of $V_n$ that is least well captured by $W_{k-1}$ and then define $\omega_k$ by applying one step of OMP to this vector, that is

$$|\langle v_k - P_{W_{k-1}}v_k, \omega_k\rangle| \ge \kappa \max_{\omega \in \mathcal{D}} |\langle v_k - P_{W_{k-1}}v_k, \omega\rangle|. \tag{WC-OMP}$$

In our numerical experiments, this algorithm performs better than the Collective OMP algorithm but the analysis is more delicate and the convergence bounds are not as good. Similarly as for theorem 2.1, here we can prove

**Theorem WC-OMP**

$$r_k \le 4 \frac{n^2\|\Psi\|^2_{\ell^1(\mathcal{D})}}{\kappa^2}(k+1)^{-1} + n^2\|\Phi - \Psi\|^2, \quad k \ge 1. \tag{2.2}$$

## 3 Numerical results

In the following, $V = H^1_0(]0,1[)$ and $V_n$ is the Fourier basis spanned by the orthonormal functions

$$\phi_k(x) = \frac{\sqrt{2}}{\pi k}\sin(k\pi x), \quad 1 \le k \le n.$$

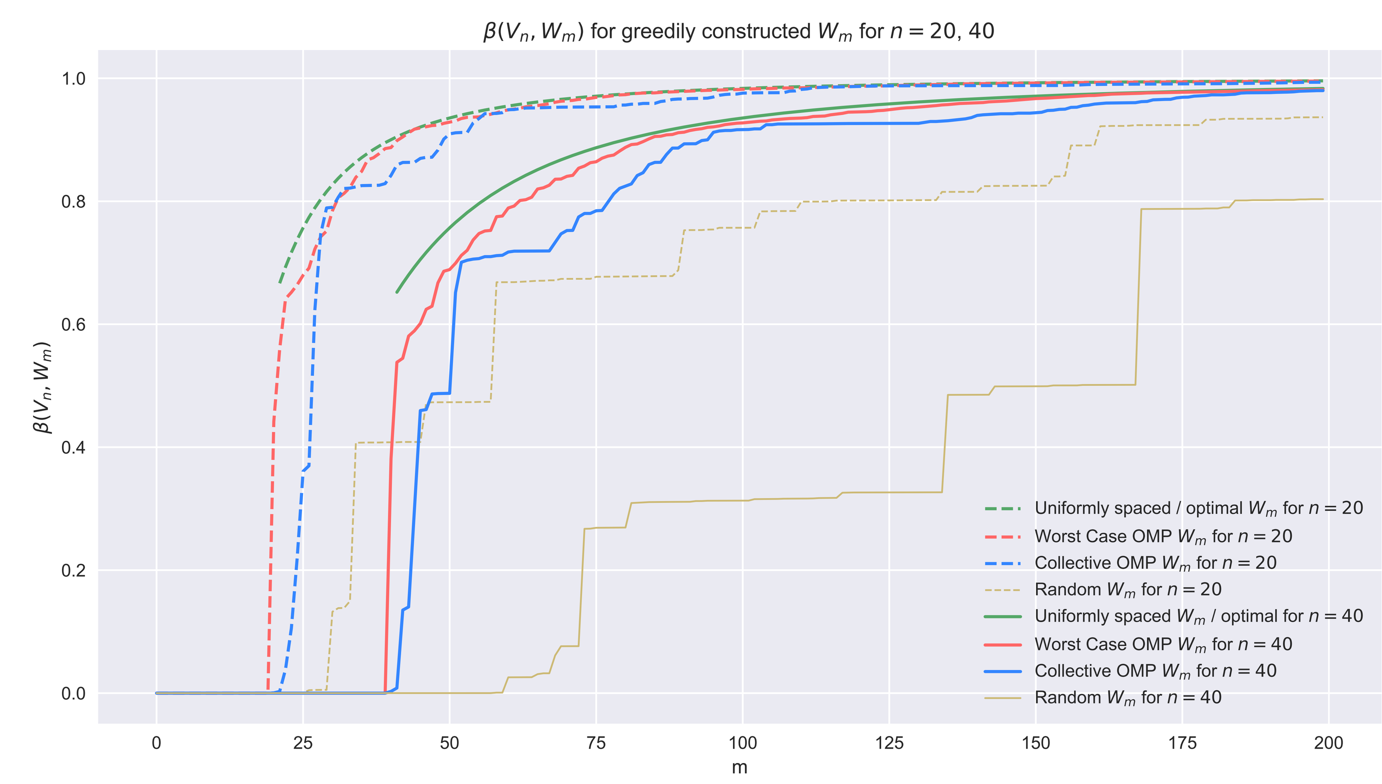
We consider the dictionary of point evaluation functionals $\mathcal{D} = \{\delta_x \ : \ x \in ]0,1[\}$. For any set of $m$ distinct points $0 < x_1 < \dots < x_m < 1$, the associated measurement space $W_m = \text{span}\{\omega_{x_1}, \dots, \omega_{x_m}\}$ coincides with the space of piecewise affine polynomials that vanish at the boundary: with $x_0 := 0$ and $x_{m+1} := 1$, we have

$$W_m = \{\omega \in \mathcal{C}^0([0,1]), \ \omega|_{[x_k, x_{k+1}]} \in \mathbb{P}_1, \ 0 \le k \le m, \text{ and } \omega(0) = \omega(1) = 0\}.$$

We can prove that in this case, a near-optimal choice is to take equispaced points. We have $\beta(V_n, W_m) \ge \beta^*$ as soon as

$$m \ge \frac{\pi}{\sqrt{2}} \frac{n}{(1 - (\beta^*)^2)^{1/2}}.$$

We next compare the performance of our two greedy algorithms with the case of equispaced points.

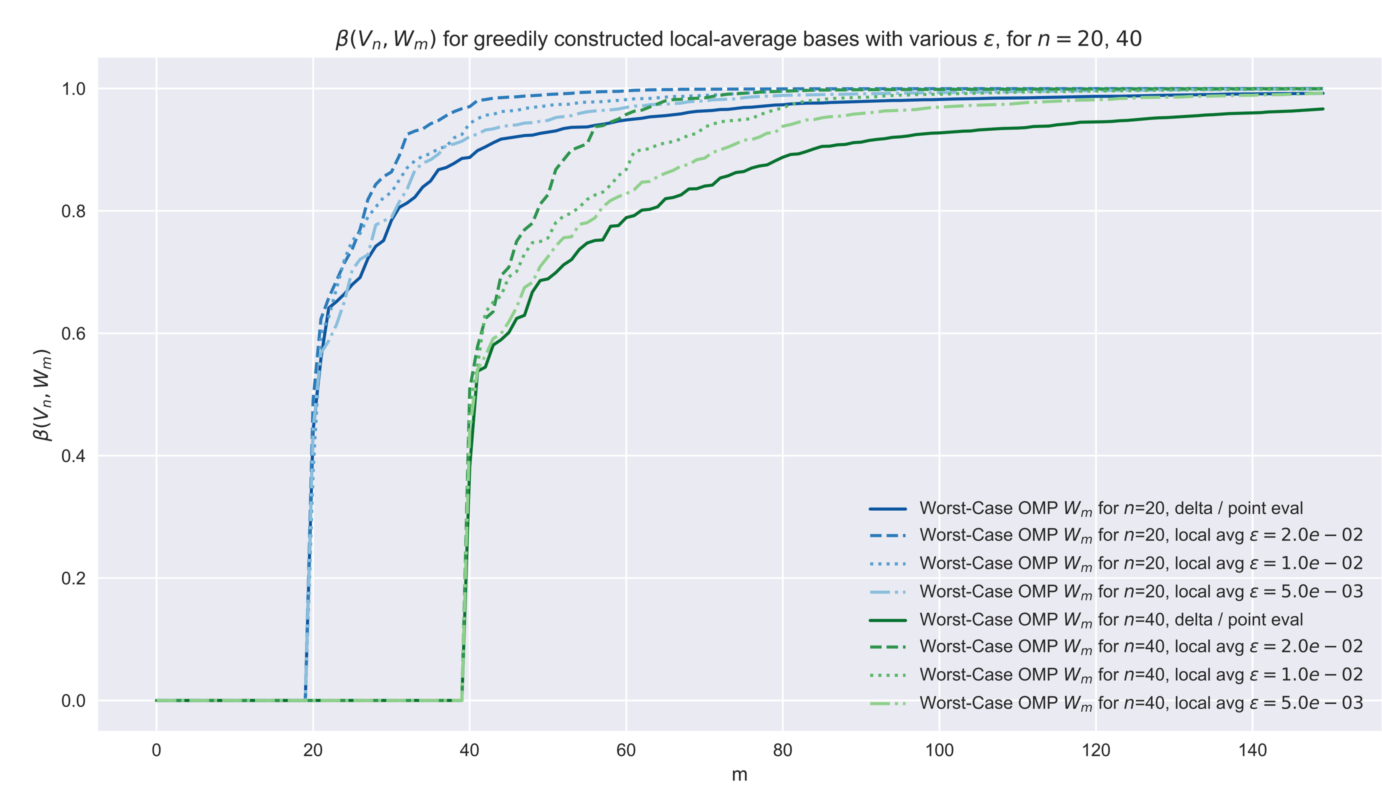



We now consider a more realistic setting where the sensors are modelled by local averages rather than point evaluations so that

$$\ell_{x,\epsilon}(u) = \int_D u(y)\varphi_\epsilon(y - x)dy, \tag{3.1}$$

where $\varphi_\epsilon(y) := \epsilon^{-d}\varphi\left(\frac{y}{\epsilon}\right)$ and $\varphi$ is some fixed radial function compactly supported in the unit ball and such that $\int \varphi = 1$. The parameter $\epsilon > 0$ denotes the width of the support. In the results below we greedily construct $W_m = \text{span}\{\omega_{x_1,\epsilon}, \dots, \omega_{x_m,\epsilon}\}$ where $\omega_{x_i,\epsilon}$ is the representer of $\ell_{x,\epsilon}$ above.

The results show an optimal value of $\epsilon$ regarding the convergence of $\beta(V_n, W_m)$ with the greedy algorithm. The existence of this optimal value can be predicted in the analysis.




## References

[1] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Data assimilation in reduced modeling. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):1–29, 2017.

[2] 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. *International Journal for Numerical Methods in Engineering*, 102(5):933–965, 2015.