

## Greedy Algorithms for Optimal Measurements Selection in State Estimation Using Reduced Models\*

Peter Binev<sup>†</sup>, Albert Cohen<sup>‡</sup>, Olga Mula<sup>§</sup>, and James Nichols<sup>‡</sup>

**Abstract.** We consider the problem of optimal recovery of an unknown function  $u$  in a Hilbert space  $V$  from measurements of the form  $\ell_j(u)$ ,  $j = 1, \dots, m$ , where the  $\ell_j$  are known linear functionals on  $V$ . We are motivated by the setting where  $u$  is a solution to a PDE with some unknown parameters, therefore lying on a certain manifold contained in  $V$ . Following the approach adopted in [Maday, Patera, Penn and Yano, *Int. J. Numer. Methods Engrg.*, 102 (2015), pp. 933–965, Binev, Cohen, Dahmen, DeVore, Petrova, and Wojtaszczyk, *SIAM J. Uncertainty Quantification*, 5 (2017), pp. 1–29], the prior on the unknown function can be described in terms of its approximability by finite-dimensional reduced model spaces  $(V_n)_{n \geq 1}$  where  $\dim(V_n) = n$ . Examples of such spaces include classical approximation spaces, e.g., finite elements or trigonometric polynomials, as well as reduced basis spaces which are designed to match the solution manifold more closely. The error bounds for optimal recovery under such priors are of the form  $\mu(V_n, W_m) \varepsilon_n$ , where  $\varepsilon_n$  is the accuracy of the reduced model  $V_n$  and  $\mu(V_n, W_m)$  is the inverse of an inf-sup constant that describe the angle between  $V_n$  and the space  $W_m$  spanned by the Riesz representers of  $(\ell_1, \dots, \ell_m)$ . This paper addresses the problem of properly selecting the measurement functionals, in order to control at best the stability constant  $\mu(V_n, W_m)$ , for a given reduced model space  $V_n$ . Assuming that the  $\ell_j$  can be picked from a given dictionary  $\mathcal{D}$  we introduce and analyze greedy algorithms that perform a suboptimal selection in reasonable computational time. We study the particular case of dictionaries that consist either of point value evaluations or local averages, as idealized models for sensors in physical systems. Our theoretical analysis and greedy algorithms may therefore be used in order to optimize the position of such sensors.

**Key words.** greedy algorithms, data assimilation, reduced models, sensor placement, inf-sup stability

**AMS subject classification.** 65M12

**DOI.** 10.1137/17M1157635

### 1. Introduction.

**1.1. State estimation from data in parametric PDEs.** One typical recovery problem in a Hilbert space  $V$  is the following: we observe  $m$  measurements of an unknown element  $u \in V$  and would like to recover  $u$  up to some accuracy. Specifically, we observe

$$(1.1) \quad z_i := \ell_i(u), \quad i = 1, \dots, m,$$

\*Received by the editors November 20, 2017; accepted for publication (in revised form) June 21, 2018; published electronically August 1, 2018.

<http://www.siam.org/journals/juq/6-3/M115763.html>

**Funding:** This research was supported by the Institut Universitaire de France, the ERC Adv grant BREAD, NSF grant DMS 1720297, the Simons foundation, and MFO.

<sup>†</sup>University of South Carolina, Department of Mathematics, Columbia, SC 29208 ([binev@math.sc.edu](mailto:binev@math.sc.edu)).

<sup>‡</sup>Laboratoire Jacques Louis-Lions, Sorbonne Universités, UMR 7598, 4 Place Jussieu, 75005, Paris, France ([cohen@ann.jussieu.fr](mailto:cohen@ann.jussieu.fr), [james.ashton.nichols@gmail.com](mailto:james.ashton.nichols@gmail.com)).

<sup>§</sup>Université Paris-Dauphine, PSL University, CNRS, UMR 7534, CEREMADE, 75016 Paris, France ([mula@ceremade.dauphine.fr](mailto:mula@ceremade.dauphine.fr)).

where the  $\ell_i$  are independent continuous linear functionals over  $V$ . In this paper we consider a Hilbert space  $V$  and denote by  $\|\cdot\|$  its norm and by  $\langle \cdot, \cdot \rangle$  its inner product. The knowledge of  $z = (z_i)_{i=1,\dots,m}$  is equivalent to that of the orthogonal projection  $w = P_{W_m} u$ , where

$$(1.2) \quad W_m := \text{span}\{\omega_1, \dots, \omega_m\}$$

and  $\omega_i \in V$  are the Riesz representers of the linear functionals  $\ell_i$ , that is,

$$(1.3) \quad \ell_i(v) = \langle \omega_i, v \rangle, \quad v \in V.$$

Obviously, there are infinitely many  $v \in V$  such that  $P_{W_m} v = w$ , and the only way to recover  $u$  up to a guaranteed accuracy is to combine the measurements with some a priori information on  $u$ .

One particularly relevant scenario is when  $u$  is a solution to some parameter-dependent PDE of the general form

$$(1.4) \quad \mathcal{P}(u, a) = 0,$$

where  $\mathcal{P}$  is a differential operator and  $a$  is a parameter that describes some physical property and lives in a given set  $\mathcal{A}$ . We assume for each  $a \in \mathcal{A}$  the problem is well posed, that is, there exists a solution  $u(a) \in V$ . Therefore, in such a scenario, our prior on  $u$  is that it belongs to the set

$$(1.5) \quad \mathcal{M} := \{u(a) : a \in \mathcal{A}\},$$

which is sometimes called the solution manifold. We assume that this manifold is a pre-compact set in  $V$ , which allows us to approximate it by finite-dimensional spaces.

For example, one could consider the diffusion equation

$$(1.6) \quad -\text{div}(a \nabla u) = f,$$

on a given domain  $D$  with fixed right side  $f$  and homogeneous Dirichlet boundary condition. Then with  $\mathcal{A}$  a set of symmetric matrix valued functions such that

$$(1.7) \quad rI \leq a(x) \leq RI, \quad x \in D, a \in \mathcal{A}$$

for some  $0 < r \leq R < \infty$ , the solution  $u(a)$  is well defined in  $V = H_0^1(D)$ . Assuming in addition that  $\mathcal{A}$  is compact in  $L^\infty(D)$ , the compactness of  $\mathcal{M}$  in  $V$  follows by standard continuity properties of the map  $a \rightarrow u(a)$ .

**1.2. Reduced model based estimation.** The above prior is generally not easily exploitable due to the fact that  $\mathcal{M}$  does not have a simple geometry. For example it is not a convex set, which prevents classical convex optimization techniques when trying to recover  $u$  in such a set. On the other hand, the particular PDE structure often allows one to derive interesting approximation properties of the solution manifold  $\mathcal{M}$  by linear spaces  $V_n$  of moderate dimension  $n$ . Such spaces can for example be obtained through a scalar parametrization of  $a(y)$  of  $a$  where  $y = (y_1, \dots, y_d)$ , using polynomial approximations of the form

$$(1.8) \quad u_n(y) = \sum_{\nu \in \Lambda_n} v_\nu y^\nu, \quad y^\nu := \prod_{j \geq 1} y_j^{\nu_j},$$

where  $\Lambda_n$  is a conveniently chosen set of multi-indices such that  $\#(\Lambda_n) = n$ . See in particular [7, 8] where convergence estimates of the form

$$(1.9) \quad \sup_{y \in U} \|u(y) - u_n(y)\| \leq Cn^{-s}$$

are established for some  $s > 0$  even when  $d = \infty$ . Thus, all solutions  $\mathcal{M}$  are approximated in the space  $V_n := \text{span}\{v_\nu : \nu \in \Lambda_n\}$ . Another typical instance is provided by reduced bases. In such approximations, one directly builds a family of  $n$ -dimensional spaces  $V_n := \text{span}\{v_1, \dots, v_n\}$  with  $v_i \in \mathcal{M}$ , in such a way that all solutions in  $\mathcal{M}$  are uniformly well approximated in  $V_n$ . In particular the approximation rate compares favorably with that achieved by the best  $n$ -dimensional spaces, that is, the Kolmogorov  $n$ -width of  $\mathcal{M}$ ; see [4, 9].

The common feature of these reduced models is that they yield a hierarchy of spaces  $(V_n)_{n \geq 1}$  with  $\dim(V_n) = n$  such that the solution manifold is uniformly well approximated in such spaces, in the sense that

$$(1.10) \quad \sup_{u \in \mathcal{M}} \text{dist}(u, V_n) \leq \varepsilon_n,$$

where  $\varepsilon_n$  is some known tolerance. Therefore, one natural option is to replace  $\mathcal{M}$  by the simpler prior class described by the cylinder

$$(1.11) \quad \mathcal{K} = \{v \in V : \text{dist}(v, V_n) \leq \varepsilon_n\}$$

for some given  $n$ . This point of view is adopted in [12] and further analyzed in [5] where the optimal recovery solution  $u^*(w)$  from the data  $w$  is characterized. This solution is defined as the center of the ellipsoid

$$(1.12) \quad \mathcal{K}_w = \{v \in \mathcal{K} : P_W v = w\},$$

and equivalently given by

$$(1.13) \quad u^*(w) := \text{argmin}\{\|v - P_{V_n} v\| : P_{W_m} v = w\}.$$

It can be computed from the data  $w$  by solving a finite set of linear equations. The worst case performance for this reconstruction is given by

$$(1.14) \quad \max_{u \in \mathcal{K}} \|u - u^*(P_W u)\| = \mu(V_n, W_m) \varepsilon_n,$$

where, for any pair of closed subspaces  $(X, Y)$  of  $V$ , we have set  $\mu(X, Y) := \beta(X, Y)^{-1}$ , with  $\beta(X, Y)$  the inf-sup constant

$$(1.15) \quad \beta(X, Y) := \inf_{x \in X} \sup_{y \in Y} \frac{\langle x, y \rangle}{\|x\| \|y\|} = \inf_{x \in X} \frac{\|P_Y x\|}{\|x\|} \in [0, 1].$$

Note that finiteness in  $\mu(V_n, W_m)$ , equivalent to  $\beta(V_n, W_m) > 0$ , requires that  $m \geq n$ . It is also shown in [5] that  $\beta(V_n, W_m)$  can be computed as the smallest singular value of the  $n \times m$  cross-Gramian matrix with entries  $\langle \phi_i, \psi_j \rangle$  between any pair of orthonormal bases  $(\phi_i)_{i=1, \dots, n}$  and  $(\psi_j)_{j=1, \dots, m}$  of  $V_n$  and  $W_m$ , respectively.

**Remark 1.1.** As already mentioned, the map  $u \mapsto u^*(w)$  in (1.13) is linear. Conversely it can easily be checked that any linear recovery algorithm may be rewritten in the form of (1.13) for a certain space  $V_n$ . On the other hand, let us note that these linear recovery methods apply to general solution manifolds that may result from nonlinear PDEs.

**1.3. Optimal measurement selection.** For a given reduced model space  $V_n$  with accuracy  $\varepsilon_n$ , one natural objective is therefore that  $\mu(V_n, W_m)$  is maintained of moderate size, with a number of measurements  $m \geq n$  as small as possible. Note that taking  $W_m = V_n$  would automatically give the minimal value  $\mu(V_n, W_m) = 1$  with  $m = n$ . However, in a typical data acquisition scenario, the measurements that comprise the basis of  $W_m$  are chosen from within a limited class. This is the case for example when placing  $m$  pointwise sensors at various locations within the physical domain  $D$ .

We model this restriction by asking that the  $\ell_i$  be picked from a *dictionary*  $\mathcal{D}$  of  $V'$ , that is, a set of linear functionals normalized according to

$$\|\ell\|_{V'} = 1, \quad \ell \in \mathcal{D},$$

which is *complete* in the sense that  $\ell(v) = 0$  for all  $\ell \in \mathcal{D}$  implies that  $v = 0$ . With an abuse of notation, we identify  $\mathcal{D}$  with the subset of  $V$  that consists of all Riesz representers  $\omega$  of the above linear functionals  $\ell$ . With such an identification,  $\mathcal{D}$  is the set of functions normalized according to

$$\|\omega\| = 1, \quad \omega \in \mathcal{D},$$

such that the finite linear combinations of elements of  $\mathcal{D}$  are dense in  $V$ . Our task is therefore to pick  $\{\omega_1, \dots, \omega_m\} \in \mathcal{D}$  in such a way that

$$(1.16) \quad \beta(V_n, W_m) \geq \beta^* > 0$$

for some prescribed  $0 < \beta^* < 1$ , with  $m$  larger than  $n$  but as small as possible. In particular, we may introduce

$$(1.17) \quad m^* = m^*(\beta^*, \mathcal{D}, V_n),$$

the minimal value of  $m$  such that there exists  $\{\omega_1, \dots, \omega_m\} \in \mathcal{D}$  satisfying (1.16).

We start section 2 with a discussion of the typical range of  $m^*$  compared to  $n$ . We first show the following two “extreme” results:

- For any  $V_n$  and  $\mathcal{D}$ , there exists  $\beta^* > 0$  such that  $m^* = n$ ; that is, the inf-sup condition (1.16) holds with the minimal possible number of measurements. However this  $\beta^*$  could be arbitrarily close to 0.
- For any prescribed  $\beta^* > 0$  and any model space  $V_n$ , there are instances of dictionaries  $\mathcal{D}$  such that  $m^*$  is arbitrarily large.

We then discuss particular cases of relevant dictionaries for the particular space  $V = H_0^1(D)$ , with inner product and norms

$$(1.18) \quad \langle u, v \rangle := \int_D \nabla u(x) \cdot \nabla v(x) \, dx \quad \text{and} \quad \|u\| := \|\nabla u\|_{L^2},$$

motivated by the aforementioned example of parametric elliptic PDEs. These dictionaries model local sensors, either as point evaluations (which is only when  $D$  is univariate) or as local averages. In such a case, we provide upper estimates of  $m^*$  in the case of spaces  $V_n$  that satisfy some inverse estimates, such as finite element or trigonometric polynomial spaces.

The optimal value  $m^*$  is proved to be of the same order as  $n$  when the sensors are uniformly spaced.

This a priori analysis is not possible for more general spaces  $V_n$  such as reduced basis spaces, which are preferred to finite element spaces due to their advantageous approximation properties. For such general spaces, we need a strategy to select the measurements. In practice,  $V$  is of finite but very large dimension and  $\mathcal{D}$  is of finite but very large cardinality

$$(1.19) \quad M := \#(\mathcal{D}) \gg 1.$$

For this reason, the exhaustive search of the set  $\{\omega_1, \dots, \omega_m\} \subset \mathcal{D}$  maximizing  $\beta(V_n, W_m)$  for a given  $m > 1$  is out of reach. One natural alternative is to rely on greedy algorithms where the  $\omega_j$  are picked incrementally.

Our starting point to the design of such algorithms is the observation that (1.16) is equivalent to having

$$(1.20) \quad \sigma_m = \sigma(V_n, W_m) := \sup_{v \in V_n, \|v\|=1} \|v - P_{W_m} v\| \leq \sigma^*, \quad \sigma^* := \sqrt{1 - (\beta^*)^2} < 1.$$

Therefore, our objective is to construct a space  $W_m$  spanned by  $m$  elements from  $\mathcal{D}$  that captures all unit norm vectors of  $V_n$  with the prescribed accuracy  $\sigma^* < 1$ . This leads us to study and analyze algorithms which may be thought as generalization to the well-studied orthogonal matching pursuit algorithm (OMP), equivalent to the algorithms we study here when applied to the case  $n = 1$  with a unit norm vector  $\phi_1$  that generates  $V_1$ . Two algorithms are proposed and analyzed in section 3 and section 4, respectively. In particular, we show that they always converge, ensuring that (1.16) holds for  $m$  sufficiently large, and we also give conditions on  $\mathcal{D}$  that allow us to a priori estimate the minimal value of  $m$  where this happens.

We close our paper in section 5 with numerical experiments that illustrate the ability of our greedy algorithms to pick good points. In particular, we return to the case of dictionaries of point evaluations or local averages, and show that the selection performed by the greedy algorithms is near optimal in the sense that it achieves (1.16) after a number of iteration of the same order as that established for  $m^*$  in the results of section 2 when  $V_n$  is a trigonometric polynomial space. We also illustrate the interest of the greedy selection of the measurement for the reduced basis spaces. We close by some remarks on the relevance of the method in the case of hyperbolic PDEs, for which reduced basis approximation is known to be less effective than for elliptic or parabolic problems due the presence of shocks.

*Remark 1.2.* The problem of optimal placement of sensors, which corresponds to the particular setting where the linear functionals are point evaluations or local averages as in section 2, has been extensively studied since the 1970s in control and systems theory. In this context, the state function to be estimated is the realization of a Gaussian stochastic process, typically obtained as the solution of a linear PDE with a white noise forcing term. The error is then measured in the mean square sense, rather than in the worst case performance sense (1.14) which is the point of view adopted in our work. The function to be minimized by the sensors locations is then the trace of the error covariance, while we target at minimizing the inverse inf-sup constant  $\mu(V_n, W)$ . See in particular [3] where the existence and characterization of the optimal sensor location is established in this stochastic setting. Continuous

optimization algorithms have been proposed for computing the optimal sensor location; see, e.g., [1, 6, 16]. One common feature with our approach is that the criterion to be minimized by the optimal location is nonconvex, which leads to potential difficulties when the number of sensors is large. This is our main motivation for introducing a greedy selection algorithm, which in addition allows us to consider more general dictionaries.

**2. A priori analysis of measurement selection strategies.** The goal of this section is to present several results that provide with estimates on the inf-sup constant  $\beta(V_n, W_m)$ . We first discuss general dictionaries  $\mathcal{D}$  and approximation spaces  $V_n$ , in which case it is not possible to control how large  $m$  has to be compared to  $n$  in order to ensure a prescribed inf-sup constant. We then show that this goal can be met with  $m$  proportional to  $n$  for more specific choices of dictionaries and approximation spaces.

**2.1. Two extreme results.** The following result shows that one can always achieve a positive inf-sup constant  $\beta(V_n, W_n)$  using a minimal number of measurement  $m = n$ ; however, there are no guarantees on the lower limit of the inf-sup constant, other than that it is greater than zero.

**Theorem 2.1.** *Given a space  $V_n$  of dimension  $n$  and any complete dictionary  $\mathcal{D}$ , there exists a selection  $\{\omega_1, \dots, \omega_n\}$  from  $\mathcal{D}$  such that  $\beta(V_n, W_n) > 0$ .*

*Proof.* We define the  $\omega_i$  inductively, together with certain functions  $\phi_i$  that constitute a basis of  $V_n$ . We first pick any element  $\phi_1 \in V_n$  of unit norm. Since the dictionary is complete, there exists  $\omega_1 \in \mathcal{D}$  such that

$$(2.1) \quad \langle \phi_1, \omega_1 \rangle \neq 0.$$

Assuming that  $\{\phi_1, \dots, \phi_{k-1}\}$  and  $\{\omega_1, \dots, \omega_{k-1}\}$  have been constructed for  $k \leq n$ , we pick  $\phi_k \in V_n$  of unit norm and orthogonal to  $W_{k-1} := \text{span}\{\omega_1, \dots, \omega_{k-1}\}$ . Then, we select  $\omega_k \in \mathcal{D}$  such that

$$(2.2) \quad \langle \phi_k, \omega_k \rangle \neq 0.$$

With such a selection procedure, we find that the cross-Gramian matrix  $(\langle \omega_i, \phi_j \rangle)_{i,j=1,\dots,n}$  is lower triangular with nonzero diagonal entries. These both show that  $\{\omega_1, \dots, \omega_n\}$  and  $\{\phi_1, \dots, \phi_n\}$  are bases, and that there is no nontrivial element of  $V_n$  that is orthogonal to  $W_n$ , which means that  $\beta(V_n, W_n) > 0$ . ■

On the negative side, the following result shows that, for a general space  $V_n$  and dictionary  $\mathcal{D}$ , there is no hope to control the value of  $\beta(V_n, W_m)$  by below, even for arbitrarily large  $m$  (here we assume that  $V$  is infinite dimensional).

**Theorem 2.2.** *Given any space  $V_n$  of dimension  $n > 0$ , any  $\varepsilon > 0$ , and any  $m > 0$ , there exists a dictionary  $\mathcal{D}$  such that, for any selection  $\{\omega_1, \dots, \omega_m\}$  from  $\mathcal{D}$ , one has  $\beta(V_n, W_m) \leq \varepsilon$ .*

*Proof.* It suffices to prove the result for  $n = 1$  since enriching the space  $V_n$  has the effect of lowering the inf-sup constant. We thus take  $V_1 = \mathbb{R}\phi$  for some  $\phi \in V$  of unit norm, and we take for  $\mathcal{D}$  an orthonormal basis of  $V$ . By appropriate rotation, we can always choose this basis so that  $|\langle \phi, \omega \rangle| \leq \varepsilon$  for all  $\omega \in \mathcal{D}$ . ■

**2.2. Pointwise evaluations.** In this section, we consider the particular dictionary that consists of all point evaluation functionals  $\delta_x$  for  $x \in D$ , where  $D$  is some domain of  $\mathbb{R}^d$ . This means that the Hilbert space  $V$  should be a reproducing kernel Hilbert space (RKHS) of functions defined on  $D$ , that is, a Hilbert space that continuously embeds in  $C(D)$ . Examples of such spaces are the Sobolev spaces  $H^s(D)$  for  $s > d/2$ , possibly with additional boundary conditions.

As a simple example, motivated by parametric second-order elliptic PDEs, for a bounded univariate interval  $I$  we consider  $V = H_0^1(I)$  which is continuously embedded in  $C(I)$ . Without loss of generality we take  $I = ]0, 1[$ . For every  $x \in ]0, 1[$ , the Riesz representer of  $\delta_x$  is given by the solution of  $\omega'' = \delta_x$  with zero boundary condition. Normalizing this solution  $\omega$  with respect to the  $V$ -norm (1.18), we obtain

$$(2.3) \quad \omega_x(t) = \begin{cases} \frac{t(1-x)}{\sqrt{x(1-x)}} & \text{for } t \leq x, \\ \frac{(1-t)x}{\sqrt{x(1-x)}} & \text{for } t > x. \end{cases}$$

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 with nodes at  $x_1, \dots, x_m$  that vanish at the boundary. Denoting  $x_0 := 0$  and  $x_{m+1} := 1$ , we have

$$(2.4) \quad W_m = \{\omega \in C^0([0, 1]), \omega|_{[x_k, x_{k+1}]} \in \mathbb{P}_1, 0 \leq k \leq m, \text{ and } \omega(0) = \omega(1) = 0\}.$$

As an example for the space  $V_n$ , let us consider the span of the Fourier basis (here orthonormalized in  $V$ ),

$$(2.5) \quad \phi_k := \frac{\sqrt{2}}{\pi k} \sin(k\pi x), \quad 1 \leq k \leq n.$$

With such choices, we can establish a lower bound (1.16) on  $\beta(V_n, W_m)$  with a number of measurements  $m^*$  that scales linearly with  $n$ , when using equally spaced measurements, as shown by the following result.

**Theorem 2.3.** *For any  $0 < \beta^* < 1$ , and any  $n > 0$ , taking point evaluations at  $x_i = \frac{i}{m+1}$  for  $i = 1, \dots, m$ , we have  $\beta(V_n, W_m) \geq \beta^*$  as soon as  $m \geq \frac{n}{\sigma^*} - 1$ , where  $\sigma^* = \sqrt{1 - (\beta^*)^2}$ . Therefore,*

$$(2.6) \quad m^*(\beta^*, \mathcal{D}, V_n) \leq \left\lceil \frac{n}{\sigma^*} - 1 \right\rceil \leq \frac{n}{\sigma^*}.$$

**Proof.** We introduce the interpolation operator  $\mathcal{I}_{W_m} : V \rightarrow W_m$ , so that the projection error in the  $V$  norm is bounded by

$$(2.7) \quad \|v - P_{W_m} v\| \leq \|v - \mathcal{I}_{W_m} v\| \leq \frac{1}{\pi} h \|v''\|_{L^2},$$

with  $h := \max_{0 \leq k \leq m} |x_{k+1} - x_k|$ . The constant  $\frac{1}{\pi}$  in the second inequality is optimal for the interpolation error and can be derived from the Poincaré inequality  $\|g'\|_{L^2([x_k, x_{k+1}])} \leq \frac{x_{k+1} - x_k}{\pi} \|g''\|_{L^2([x_k, x_{k+1}])}$  for  $g'$ . On the other hand, one has the inverse estimate

$$(2.8) \quad \|v''\|_{L^2([0, 1])} \leq \pi n \|v'\|_{L^2([0, 1])}, \quad v \in V_n,$$



and therefore

$$(2.9) \quad \|v - P_{W_m} v\| \leq hn\|v\|, \quad v \in V_n.$$

When using equally spaced measurements, we have  $h = (m+1)^{-1}$  and therefore

$$(2.10) \quad \|v - P_{W_m} v\| \leq \frac{n}{m+1} \|v\|.$$

To satisfy  $\sup_{v \in V_n, \|v\|=1} \|v - P_{W_m} v\| \leq \sigma^*$ , we need

$$(2.11) \quad m \geq \frac{n}{\sigma^*} - 1$$

equispaced points. ■

Bearing in mind that  $\beta(V_n, W_m) = 0$  for  $m < n$ , it is fair to say that the estimate (2.11) gives a relatively sharp estimation of the minimal  $m$  which is required.

**Remark 2.4.** The interplay between pointwise evaluation and Fourier reconstruction has been the object of much attention in the area of sparse recovery and compressed sensing. It is known in particular that, with high probability, trigonometric polynomials of degree  $N$  which are  $n$ -sparse can be exactly reconstructed from their sampling at  $m$  randomly chosen points according to the uniform measure, if  $m$  is larger than  $n$  by some logarithmic factors. We refer to [13] for an instance of such results. In our case the setting is different since we are searching for a recovery in the fixed trigonometric space  $V_n$ , which allows us to have  $m$  exactly proportional to  $n$ .

**2.3. Local averages.** We return to the situation where  $D$  is a general domain in  $\mathbb{R}^d$ . Since real world pointwise sensors have a nonzero point spread, it is natural to model them by linear functionals that are local averages rather than point evaluations, that is,

$$(2.12) \quad \ell_{x,\tau}(u) = \int_D u(y) \varphi_\tau(y-x) \, dy,$$

where

$$(2.13) \quad \varphi_\tau(y) := \tau^{-d} \varphi\left(\frac{y}{\tau}\right)$$

for some fixed radial function  $\varphi$  compactly supported in the unit ball  $B = \{|x| \leq 1\}$  of  $\mathbb{R}^d$  and such that  $\int \varphi = 1$ , and  $\tau > 0$  representing the point spread. Here, we assume in addition that  $\varphi$  belongs to  $H_0^1(B)$ . Taking the measurement point  $x$  at a distance at least  $\tau$  from the boundary of  $D$ , we are ensured that the support of  $\varphi_\tau(\cdot - x)$  is fully contained in  $D$ .

Note that in several space dimension  $d > 1$  the point evaluation functionals are not continuous on  $H_0^1(D)$ ; however, the above local averages are. We may therefore use these functionals in the case where  $V = H_0^1(D)$ , in arbitrary multivariate dimension. The corresponding Riesz representers  $\omega_{x,\tau} \in V$  are the solutions to

$$(2.14) \quad -\Delta \omega_{x,\tau} = g_{x,\tau}, \quad g_{x,\tau} := \varphi_\tau(\cdot - x),$$

with homogeneous Dirichlet boundary conditions.



For simplicity we work in dimension  $d \leq 3$ . We consider measurement points  $x_1, \dots, x_m$  that are uniformly spaced, in the sense that they are at the vertices of a quasi-uniform mesh  $\mathcal{T}_h$  for the domain  $D$  with mesh size  $h > 0$ . Therefore

$$(2.15) \quad m \sim h^{-d}.$$

We are interested in the approximation properties of the corresponding space  $W_m$ . We cannot hope for an estimate similar to (2.7) since these approximation properties should also depend on  $\tau > 0$ . This is reflected by the following direct estimate.

**Lemma 2.5.** *One has the estimate for the projection error in the  $V$ -norm*

$$(2.16) \quad \|v - P_{W_m} v\| \leq C \left( \tau + \frac{h^2}{\tau} \right) \|v\|_{H^2}, \quad v \in H^2(D) \cap V,$$

where  $C$  is independent of  $(h, \tau)$ .

*Proof.* We shall establish the following estimate in  $H^{-1}(D)$ :

$$(2.17) \quad \min_{c_1, \dots, c_m \in \mathbb{R}^m} \left\| w - \sum_{i=1}^m c_i g_{x_i, \tau} \right\|_{H^{-1}} \leq C \left( \tau + \frac{h^2}{\tau} \right) \|w\|_{L^2}, \quad w \in L^2(D).$$

Then, (2.16) immediately follows by application of (2.17) to  $w = \Delta v$ .

In order to prove (2.17), we first introduce the nodal  $\mathbb{P}_1$  finite element basis  $\psi_i$  associated to the mesh points  $x_i$ . So, under some reasonable geometric assumptions on  $D$  (piecewise smoothness of its boundary), we have the interpolation error estimate

$$(2.18) \quad \left\| v - \sum_{i=1}^m v(x_i) \psi_i \right\|_{L^2} \leq Ch^2 \|v\|_{H^2}, \quad v \in H^2(D).$$

Here we have used the fact that  $d \leq 3$  in order to be able to apply point evaluation of the elements of  $H^2(D)$ . We introduce the dual space

$$(2.19) \quad G^{-2}(D) := H^2(D)',$$

which differs from  $H^{-2}(D)$ , and obtain by a duality argument that

$$(2.20) \quad \left\| w - \sum_{i=1}^m c_i \delta_{x_i} \right\|_{G^{-2}} \leq Ch^2 \|w\|_{L^2}, \quad c_i := \int_D w \psi_i, \quad w \in L^2.$$

In order to derive (2.17), we note that  $g_{x_i, \tau} = \delta_{x_i} * \varphi_\tau$ , from which it follows that, for any  $w \in L^2(D)$ ,

$$(2.21) \quad \left\| w - \sum_{i=1}^m c_i g_{x_i, \tau} \right\|_{H^{-1}} \leq \|w - w * \varphi_\tau\|_{H^{-1}} + \left\| \varphi_\tau * \left( w - \sum_{i=1}^m c_i \delta_{x_i} \right) \right\|_{H^{-1}}.$$

Here, the convolution of  $w$  is meant in the sense

$$(2.22) \quad (w * \varphi_\tau)(x) := \int_D w(y) \varphi_\tau(x - y) dy = \int_{\mathbb{R}^d} \tilde{w}(y) \varphi_\tau(x - y) dy,$$

where  $\tilde{w}$  is the extension of  $w$  by 0 outside of  $D$ .

For the first term on the right side of (2.21), we write

$$\begin{aligned} \|w - w * \varphi_\tau\|_{H^{-1}} &= \max_{\|\psi\|=1} \int_D (w - w * \varphi_\tau) \psi \\ &= \max_{\|\psi\|=1} \int_D w(\psi - \psi * \varphi_\tau) \\ &\leq \|w\|_{L^2} \max_{\|\psi\|=1} \|\psi - \psi * \varphi_\tau\|_{L^2}. \end{aligned}$$

We now remark that the extension  $\psi \rightarrow \tilde{\psi}$  by 0 outside of  $D$  is continuous from  $V$  to  $H^1(\mathbb{R}^d)$ . Thus, for  $\psi \in V$ , we may write

$$(2.23) \quad \|\psi - \psi * \varphi_\tau\|_{L^2(D)} \leq \|\tilde{\psi} - \tilde{\psi} * \varphi_\tau\|_{L^2(\mathbb{R}^d)} \leq C\tau \|\tilde{\psi}\|_{H^1(\mathbb{R}^d)} \leq C\tau \|\psi\|,$$

up to a change in the constant  $C$  in the last inequality, and where the second inequality is a standard estimate for regularization by convolution. It follows that

$$(2.24) \quad \|w - w * \varphi_\tau\|_{H^{-1}} \leq C\tau \|w\|_{L^2}.$$

For the second term on the right side of (2.21), we write

$$\begin{aligned} \left\| \varphi_\tau * \left( w - \sum_{i=1}^m c_i \delta_{x_i} \right) \right\|_{H^{-1}} &= \max_{\|\psi\|=1} \int_D \left( \varphi_\tau * w - \sum_{i=1}^m c_i \varphi_\tau(\cdot - x_i) \right) \psi \\ &= \max_{\|\psi\|=1} \left( \int_D w(\varphi_\tau * \psi) - \sum_{i=1}^m c_i (\varphi_\tau * \psi)(x_i) \right) \\ &\leq \max_{\|\psi\|=1} \left\| w - \sum_{i=1}^m c_i \delta_{x_i} \right\|_{G^{-2}} \|\varphi_\tau * \psi\|_{H^2} \\ &\leq Ch^2 \|w\|_{L^2} \max_{\|\psi\|=1} \|\varphi_\tau * \psi\|_{H^2}, \end{aligned}$$

where we have used (2.20). Using again the extension  $\tilde{\psi}$  of  $\psi$  by 0 outside of  $D$ , we write

$$(2.25) \quad \|\varphi_\tau * \psi\|_{H^2(D)} = \|\varphi_\tau * \tilde{\psi}\|_{H^2(\mathbb{R}^d)} \leq \|\varphi_\tau\|_{H^1(\mathbb{R}^d)} \|\tilde{\psi}\|_{H^1(\mathbb{R}^d)} \leq C\tau^{-1} \|\tilde{\psi}\|_{H^1(\mathbb{R}^d)} \leq C\tau^{-1} \|\psi\|,$$

where the first inequality is straightforward by Fourier transform, and the second inequality follows from the definition of  $\varphi_\tau$  by scaling. It follows that

$$(2.26) \quad \left\| \varphi_\tau * \left( w - \sum_{i=1}^m c_i \delta_{x_i} \right) \right\|_{H^{-1}} \leq C \frac{h^2}{\tau} \|w\|_{L^2}.$$

By summation of (2.24) and (2.26), we reach (2.17). ■

*Remark 2.6.* The estimate (2.16) deteriorates for too small or too large  $\tau$ . The choice  $\tau \sim h$  gives the optimal approximation order  $\mathcal{O}(h)$ .

We may use the above approximation result to estimate the number of local average measurements needed to control the inf-sup constant  $\beta(V_n, W_m)$ , provided that  $V_n$  satisfies some inverse estimate. As an example, consider the case where  $D = [0, 1]^d$  and the space  $V_n$  is the span of the Fourier basis

$$(2.27) \quad \phi_k := \sin(\pi k \cdot x), \quad k = (k_1, \dots, k_d), \quad 1 \leq k_i \leq K,$$

so that  $n = \dim(V_n) = K^d$ . Combining the direct estimate (2.16) with an inverse estimate, we obtain

$$(2.28) \quad \|v - P_{W_m} v\| \leq C \left( \tau + \frac{h^2}{\tau} \right) n^{\frac{1}{d}} \|v\|, \quad v \in V_n.$$

Using  $m = J^d$  equally spaced measurements at points on the tensorized grid

$$(2.29) \quad \frac{1}{J+1}(j_1, \dots, j_d), \quad 1 \leq j_i \leq J,$$

we find that  $h \sim m^{-1/d}$  and therefore

$$(2.30) \quad \|v - P_{W_m} v\| \leq C \left( \tau + \frac{m^{-2/d}}{\tau} \right) n^{\frac{1}{d}} \|v\|, \quad v \in V_n.$$

This shows that  $\beta(V_n, W_m) \geq \beta^*$  can be achieved provided that

$$(2.31) \quad C \left( \tau + \frac{m^{-2/d}}{\tau} \right) n^{\frac{1}{d}} \leq \sigma^*,$$

where  $\sigma^* = \sqrt{1 - (\beta^*)^2}$ .

In this analysis, the number  $m$  of required measurements deteriorates for large or small values of  $\tau$ , due to the deterioration of the approximation estimate already noted in Remark 2.6. This fact will be confirmed in the numerical experiments given in section 5.2. In the case where  $\tau$  is of the optimal order of  $\tau \sim h \sim m^{-1/d}$ , the above condition becomes  $C(n/m)^{\frac{1}{d}} \leq \sigma^*$ , which shows that  $\beta(V_n, W_m) \geq \beta^*$  can be achieved with a number of measurements  $m^*$  that scales linearly with  $n$ , similar to univariate pointwise evaluation described by Theorem 2.3. Note that we are not able to recover the results on univariate pointwise evaluation from the results on local averages, since the estimate (2.30) deteriorates as  $\tau \rightarrow 0$  even when  $d = 1$ . This hints that there is some room for improvement in this estimate.

**3. A collective OMP algorithm.** In this section we discuss a first numerical algorithm for the incremental selection of the spaces  $W_m$ , inspired by the orthonormal matching pursuit (OMP) algorithm which is recalled below. More precisely, our algorithm may be viewed as applying the OMP algorithm for the collective approximation of the elements of an orthonormal basis of  $V_n$  by linear combinations of  $m$  members of the dictionary.

Our objective is to reach a bound (1.20) for the quantity  $\sigma_m$ . Note that this quantity can also be written as

$$\sigma_m = \|(I - P_{W_m})|_{V_n}\|_{\mathcal{L}(V_n, V)},$$

that is,  $\sigma_m$  is the spectral norm of  $I - P_{W_m}$  restricted to  $V_n$ .

**3.1. Description of the algorithm.** When  $n = 1$ , there is only one unit vector  $\phi_1 \in V_1$  up to a sign change. A commonly used strategy for approximating  $\phi_1$  by a small combination of elements from  $\mathcal{D}$  is to apply a greedy algorithm, the most prominent one being the orthogonal matching pursuit (OMP): we iteratively select

$$(3.1) \quad \omega_k = \operatorname{argmax}_{\omega \in \mathcal{D}} |\langle \omega, \phi_1 - P_{W_{k-1}} \phi_1 \rangle|,$$

where  $W_{k-1} := \operatorname{span}\{\omega_1, \dots, \omega_{k-1}\}$  and  $W_0 := \{0\}$ . In practice, one often relaxes the above maximization, by taking  $\omega_k$  such that

$$(3.2) \quad |\langle \omega_k, \phi_1 - P_{W_{k-1}} \phi_1 \rangle| \geq \kappa \max_{\omega \in \mathcal{D}} |\langle \omega, \phi_1 - P_{W_{k-1}} \phi_1 \rangle|$$

for some fixed  $0 < \kappa < 1$ , for example  $\kappa = \frac{1}{2}$ . This is known as the weak OMP algorithm, but we refer to it as OMP, as well. It has been studied in [2, 10]; see also [14] for a complete survey on greedy approximation.

For a general value of  $n$ , one natural strategy is to define our greedy algorithm as follows: we iteratively select

$$(3.3) \quad \omega_k = \operatorname{argmax}_{\omega \in \mathcal{D}} \max_{v \in V_n, \|v\|=1} |\langle \omega, v - P_{W_{k-1}} v \rangle| = \operatorname{argmax}_{\omega \in \mathcal{D}} \|P_{V_n}(\omega - P_{W_{k-1}} \omega)\|.$$

Note that in the case  $n = 1$  we obtain the original OMP algorithm applied to  $\phi_1$ .

As to the implementation of this algorithm, we take  $(\phi_1, \dots, \phi_n)$  to be any orthonormal basis of  $V_n$ . Then

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

Therefore, at every step  $k$ , we have

$$\omega_k = \operatorname{argmax}_{\omega \in \mathcal{D}} \sum_{i=1}^n |\langle \phi_i - P_{W_{k-1}} \phi_i, \omega \rangle|^2,$$

which amounts to a stepwise optimization of a similar nature as in the standard OMP. Note that, while the basis  $(\phi_1, \dots, \phi_n)$  is used for the implementation, the actual definition of the greedy selection algorithm is independent of the choice of this basis in view of (3.3). It only involves  $V_n$  and the dictionary  $\mathcal{D}$ . Similar to OMP, we may weaken the algorithm by taking  $\omega_k$  such that

$$\sum_{i=1}^n |\langle \phi_i - P_{W_{k-1}} \phi_i, \omega_k \rangle|^2 \geq \kappa^2 \max_{\omega \in \mathcal{D}} \sum_{i=1}^n |\langle \phi_i - P_{W_{k-1}} \phi_i, \omega \rangle|^2$$

for some fixed  $0 < \kappa < 1$ .

For such a basis, we introduce the residual quantity

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

This quantity allows us to control the validity of (1.16) since we have

$$\sigma_m = \sup_{v \in V_n, \|v\|=1} \|v - P_{W_m} v\| = \sup_{\sum_{i=1}^n c_i^2 = 1} \left\| \sum_{i=1}^n c_i (\phi_i - P_{W_m} \phi_i) \right\| \leq r_m^{1/2},$$

and therefore (1.16) holds provided that  $r_m \leq \sigma^2 = 1 - \gamma^2$ .

**Remark 3.1.** The quantity  $r_m^{1/2}$  is the Hilbert–Schmidt norm of the operator  $I - P_{W_m}$  restricted to  $V_n$ . The inequality  $\sigma_m \leq r_m^{1/2}$  simply expresses the fact that the Hilbert–Schmidt norm controls the spectral norm. On the other hand, in dimension  $n$ , the Hilbert–Schmidt norm can be up to  $n^{1/2}$  times the spectral norm. This lack of sharpness is one principle limitation in our convergence analysis which uses the fact that we can estimate the decay of  $r_m$ , but not directly that of  $\sigma_m$ .

**3.2. Convergence analysis.** By analogy to the analysis of OMP provided in [10], we introduce for any  $\Psi = (\psi_1, \dots, \psi_n) \in V^n$  the quantity

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

or equivalently, denoting  $c_\omega := \{c_{\omega,i}\}_{i=1}^n$ ,

$$\|\Psi\|_{\ell^1(\mathcal{D})} := \inf_{c_\omega} \left\{ \sum_{\omega \in \mathcal{D}} \|c_\omega\|_2 : \Psi = \sum_{\omega \in \mathcal{D}} c_\omega \omega \right\}.$$

This quantity is a norm on the subspace of  $V^n$  on which it is finite.

Given that  $\Phi = (\phi_1, \dots, \phi_n)$  is any orthonormal basis of  $V_n$ , we write

$$J(V_n) := \|\Phi\|_{\ell^1(\mathcal{D})}.$$

This quantity is indeed independent on the orthonormal basis  $\Phi$ : if  $\tilde{\Phi} = (\tilde{\phi}_1, \dots, \tilde{\phi}_n)$  is another orthonormal basis, we have  $\tilde{\Phi} = U\Phi$  where  $U$  is unitary. Therefore any representation  $\Phi = \sum_{\omega \in \mathcal{D}} c_\omega \omega$  induces the representation

$$\tilde{\Phi} = \sum_{\omega \in \mathcal{D}} \tilde{c}_\omega \omega, \quad \tilde{c}_\omega = U c_\omega,$$

with the equality

$$\sum_{\omega \in \mathcal{D}} \|\tilde{c}_\omega\|_2 = \sum_{\omega \in \mathcal{D}} \|c_\omega\|_2,$$

so that  $\|\Phi\|_{\ell^1(\mathcal{D})} = \|\tilde{\Phi}\|_{\ell^1(\mathcal{D})}$ .

One important observation is that if  $\Phi = (\phi_1, \dots, \phi_n)$  is an orthonormal basis of  $V_n$  and if  $\Phi = \sum_{\omega \in \mathcal{D}} c_\omega \omega$ , one has

$$n = \sum_{i=1}^n \|\phi_i\| \leq \sum_{i=1}^n \sum_{\omega \in \mathcal{D}} |c_{\omega,i}| = \sum_{\omega \in \mathcal{D}} \|c_\omega\|_1 \leq \sum_{\omega \in \mathcal{D}} n^{1/2} \|c_\omega\|_2.$$

Therefore, we always have

$$J(V_n) \geq n^{1/2}.$$

Using the quantity  $J(V_n)$ , we can generalize the result of [10] on the OMP algorithm in the following way.

**Theorem 3.2.** *Assuming that  $J(V_n) < \infty$ , the collective OMP algorithm satisfies*

$$(3.4) \quad r_m \leq \frac{J(V_n)^2}{\kappa^2} (m+1)^{-1}, \quad m \geq 0.$$

*Proof.* For  $m = 0$ , we have

$$r_0 = \sum_{i=1}^n \|\phi_i\|^2 = n \leq J(V_n)^2.$$

We then write

$$r_m = r_{m-1} - \sum_{i=1}^n \|(P_{W_m} - P_{W_{m-1}})\phi_i\|^2 \leq r_{m-1} - \sum_{i=1}^n |\langle \phi_i - P_{W_{m-1}}\phi_i, \omega_m \rangle|^2.$$

On the other hand, if  $\Phi = \sum_{\omega \in \mathcal{D}} c_\omega \omega$ , we have

$$r_{m-1} = \sum_{i=1}^n |\langle \phi_i - P_{W_{m-1}}\phi_i, \phi_i \rangle| \leq \sum_{\omega \in \mathcal{D}} \sum_{i=1}^n |c_{\omega,i}| |\langle \phi_i - P_{W_{m-1}}\phi_i, \omega \rangle|,$$

which by the Cauchy–Schwarz inequality implies

$$\begin{aligned} r_{m-1} &\leq \sum_{\omega \in \mathcal{D}} \|c_\omega\|_2 \left( \sum_{i=1}^n |\langle \phi_i - P_{W_{m-1}}\phi_i, \omega \rangle|^2 \right)^{1/2} \\ &\leq \kappa^{-1} \left( \sum_{i=1}^n |\langle \phi_i - P_{W_{m-1}}\phi_i, \omega_m \rangle|^2 \right)^{1/2} \sum_{\omega \in \mathcal{D}} \|c_\omega\|_2, \end{aligned}$$

and therefore

$$\sum_{i=1}^n |\langle \phi_i - P_{W_{m-1}}\phi_i, \omega_m \rangle|^2 \geq \frac{\kappa^2 r_{m-1}^2}{J(V_n)^2}.$$

This implies that

$$r_m \leq r_{m-1} \left( 1 - \frac{\kappa^2}{J(V_n)^2} r_{m-1} \right),$$

from which (3.4) follows by the same elementary induction argument as used in [10]. ■

**Remark 3.3.** Note that the right side of (3.4) is always larger than  $n(m+1)^{-1}$ , which is consistent with the fact that  $\beta(V_n, W_m) = 0$  if  $m < n$ .

One natural strategy for selecting the measurement space  $W_m$  is therefore to apply the above described greedy algorithm until the first value  $\tilde{m} = \tilde{m}(n)$  is met such that  $\beta(V_n, W_m) \geq \gamma$ . According to (3.4), this value satisfies

$$(3.5) \quad m(n) \leq \frac{J(V_n)^2}{\kappa^2 \sigma^2}.$$

For a general dictionary  $\mathcal{D}$  and space  $V_n$  we have no control on the quantity  $J(V_n)$ , which could even be infinite, and therefore the above result does not guarantee that the above selection strategy eventually meets the target bound  $\beta(V_n, W_m) \geq \gamma$ . In order to treat this case, we establish a perturbation result similar to that obtained in [2] for the standard OMP algorithm.

**Theorem 3.4.** *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*

$$(3.6) \quad r_m \leq 4 \frac{\|\Psi\|_{\ell^1(\mathcal{D})}^2}{\kappa^2} (m+1)^{-1} + \|\Phi - \Psi\|^2, \quad m \geq 1,$$

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

*Proof.* We introduce

$$t_m := r_m - \|\Phi - \Psi\|^2,$$

so that, by the same argument as in the proof of Theorem 3.2, we have

$$(3.7) \quad t_m \leq t_{m-1} - \sum_{i=1}^n |\langle \phi_i - P_{W_{m-1}} \phi_i, \omega_m \rangle|^2.$$

Next, we write

$$r_{m-1} = \sum_{i=1}^n \langle \phi_i - P_{W_{m-1}} \phi_i, \psi_i \rangle + \sum_{i=1}^n \langle \phi_i - P_{W_{m-1}} \phi_i, \phi_i - \psi_i \rangle$$

By the same argument as in the proof of Theorem 3.2, the first term is bounded by

$$\sum_{i=1}^n \langle \phi_i - P_{W_{m-1}} \phi_i, \psi_i \rangle \leq \kappa^{-1} \left( \sum_{i=1}^n |\langle \phi_i - P_{W_{m-1}} \phi_i, \omega_m \rangle|^2 \right)^{1/2} \|\Psi\|_{\ell^1(\mathcal{D})}.$$

By Cauchy–Schwarz and Young inequalities, the second term is bounded by

$$\sum_{i=1}^n \langle \phi_i - P_{W_{m-1}} \phi_i, \phi_i - \psi_i \rangle \leq r_{m-1}^{1/2} \|\Phi - \Psi\| \leq \frac{1}{2} (r_{m-1} + \|\Phi - \Psi\|^2).$$

It follows that

$$t_{m-1} \leq 2\kappa^{-1} \left( \sum_{i=1}^n |\langle \phi_i - P_{W_{m-1}} \phi_i, \omega_m \rangle|^2 \right)^{1/2} \|\Psi\|_{\ell^1(\mathcal{D})},$$

which combined with (3.7) gives

$$t_m \leq t_{m-1} \left( 1 - \frac{\kappa^2}{4\|\Psi\|_{\ell^1(\mathcal{D})}^2} t_{m-1} \right).$$

We may apply the same induction argument as for  $r_m$  in the proof of Theorem 3.2 to conclude that

$$(3.8) \quad t_m \leq \frac{4\|\Psi\|_{\ell^1(\mathcal{D})}^2}{\kappa^2} (m+1)^{-1},$$



which is the announced result. This argument requires that for the first step we have  $t_0 \leq \frac{4\|\Psi\|_{\ell^1(\mathcal{D})}^2}{\kappa^2}$ . However, if  $t_0 \geq \frac{4\|\Psi\|_{\ell^1(\mathcal{D})}^2}{\kappa^2}$ , then the above recursion shows that  $t_1 \leq 0$ , which implies the result for all values of  $m \geq 1$  by monotonicity of  $m \mapsto t_m$ . ■

As an immediate consequence of the above result, we obtain that the collective OMP converges for any space  $V_n$ , even when  $J(V_n)$  is not finite.

**Corollary 3.5.** *For any  $n$ -dimensional space  $V_n$ , the application of the collective OMP algorithm on the space  $V_n$  gives that  $\lim_{m \rightarrow +\infty} r_m = 0$ .*

*Proof.* By completeness of  $\mathcal{D}$ , for any  $\delta$ , there exists a finite subset  $\mathcal{F} \subset \mathcal{D}$  and vector coefficients  $c_\omega$  such that

$$\|\Phi - \Psi\| \leq (\delta/2)^{1/2}, \quad \Psi := \sum_{\omega \in \mathcal{F}} c_\omega \omega.$$

One obviously has  $\|\Psi\|_{\ell^1(\mathcal{D})} < \infty$ , and therefore one has

$$4 \frac{\|\Psi\|_{\ell^1(\mathcal{D})}^2}{\kappa^2} (m+1)^{-1} \leq \frac{\delta}{2}$$

for  $m \geq m(\delta)$  sufficiently large. It follows that  $r_m \leq \delta$  for  $m \geq m(\delta)$ . ■

The above corollary shows that if  $\gamma > 0$ , one has  $\beta(V_n, W_m) \geq \gamma$  for  $m$  large enough.

**Remark 3.6.** One alternative strategy to select the measurements could be to apply the OMP algorithm separately on each basis element  $\phi_j$ . This leads for each  $j \in \{1, \dots, n\}$  to the selection of  $\omega_{1,j}, \dots, \omega_{m_j,j} \in \mathcal{D}$  and to a space

$$W_m := \text{span}\{\omega_{k,j} : k = 1, \dots, m_j, j = 1, \dots, m\}, \quad m_1 + \dots + m_j = m.$$

The following argument shows that, even when optimizing the choice of the number of iterations  $m_j$  used for each basis element, this strategy leads to convergence bounds that are not as good as those achieved by the collective OMP algorithm. Since the residual satisfies

$$r_m = \sum_{j=1}^n \|\phi_j - P_{W_m} \phi_j\|^2 \leq \sum_{j=1}^n \kappa^{-2} (m_j + 1)^{-1} \|\phi_j\|_{\ell^1(\mathcal{D})}^2,$$

we optimize by choosing  $m_j := \lfloor m \|\phi_j\|_{\ell^1(\mathcal{D})} (\sum_{j=1}^n \|\phi_j\|_{\ell^1(\mathcal{D})})^{-1} \rfloor$ . This leads to the convergence bound

$$r_m \leq \kappa^{-2} m^{-1} \sum_{j=1}^n \|\phi_j\|_{\ell^1(\mathcal{D})}.$$

This bound is not as good as (3.4), since we have

$$(3.9) \quad \sum_{j=1}^n \|\phi_j\|_{\ell^1(\mathcal{D})} = \inf \left\{ \sum_{\omega \in \mathcal{D}} \|c_\omega\|_1 : \sum_{\omega \in \mathcal{D}} c_\omega \omega = \phi \right\} \leq \|\Phi\|_{\ell^1(\mathcal{D})} = J(V_n).$$

**Remark 3.7.** In the case  $n = 1$ , a well-known variant to the OMP algorithm, also discussed in [2], is the so-called *relaxed greedy algorithm*. This variant avoids the computation of the projection onto  $W_k$ : the approximation of  $\phi_1$  is updated by

$$(3.10) \quad A_k \phi_1 = \alpha_k A_{k-1} \phi_1 + \beta_k \langle \phi_1 - A_{k-1} \phi_1, \omega_k \rangle \omega_k,$$

where  $\omega_k$  is selected by maximizing  $|\langle \phi_1 - A_{k-1} \phi_1, \omega \rangle|$  over  $\omega \in \mathcal{D}$  and  $(\alpha_k, \beta_k)$  are appropriate weights. This strategy is generalized in [11] to the collective setting (with  $\alpha_k = 1 - (k+1)^{-1}$  and  $\beta_k$  minimizing the norm of the residual), and is proved to achieve similar convergence properties as the collective OMP algorithm.

**4. A worst case OMP algorithm.** We present in this section a variant of the previous collective OMP algorithm. In our numerical experiments presented in section 5 this variant performs better than the collective OMP algorithm; however, its analysis is more delicate. In particular we do not obtain convergence bounds that are as good.

**4.1. Description of the algorithm.** We first take

$$(4.1) \quad v_k := \operatorname{argmax} \left\{ \|v - P_{W_{k-1}} v\| : v \in V_n, \|v\| = 1 \right\},$$

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

$$(4.2) \quad |\langle v_k - P_{W_{k-1}} v_k, \omega_k \rangle| \geq \kappa \max \left\{ |\langle v_k - P_{W_{k-1}} v_k, \omega \rangle| : \omega \in \mathcal{D} \right\}$$

for some fixed  $0 < \kappa < 1$ .

**Remark 4.1.** A variant to this algorithm was priorily suggested in [12] in the particular case where the dictionary consists of the  $\omega_{x,\tau}$  in (2.14) associated to the local average functionals in (2.12) with  $\varphi_\tau$  a Gaussian of fixed width (see algorithm 2 in [12], called SGreedy, therein). In this variant, the selection is done by searching for the point  $x_k$  where  $|v_k - P_{W_{k-1}} v_k|$  takes its maximum, and taking  $\omega_k = \omega_{x_k,\tau}$ . There is no evidence provided that this selection process has convergence properties similar to those that we prove next for the selection by (4.2).

**4.2. Convergence analysis.** The first result gives a convergence rate of  $r_m$  under the assumption that  $J(V_n) < \infty$ , similar to Theorem 3.2, but with a multiplicative constant that is inflated by  $n^2$ .

**Theorem 4.2.** *Assuming that  $J(V_n) < \infty$ , the worst case OMP algorithm satisfies*

$$(4.3) \quad r_m \leq \frac{n^2 J(V_n)^2}{\kappa^2} (m+1)^{-1}, \quad m \geq 0.$$

**Proof.** As in the proof of Theorem 3.2, we use the inequality

$$r_m \leq r_{m-1} - \sum_{i=1}^n |\langle \phi_i - P_{W_{m-1}} \phi_i, \omega_m \rangle|^2,$$

and try to bound  $\sum_{i=1}^n |\langle \phi_i - P_{W_{m-1}} \phi_i, \omega_m \rangle|^2$  from below. For this purpose, we write

$$(4.4) \quad r_{m-1} = \sum_{j=1}^n \|\phi_j - P_{W_{m-1}} \phi_j\|^2 \leq n \|v_m - P_{W_{m-1}} v_m\|^2 = n \langle v_m - P_{W_{m-1}} v_m, v_m \rangle$$

We have  $v_m = \sum_{i=1}^n a_i \phi_i$  for a vector  $a = (a_1, \dots, a_n)$  such that  $\|a\|_2 = 1$ . Thus, if  $\Phi = \sum_{\omega \in \mathcal{D}} c_\omega \omega$  we may write

$$v_m = \sum_{\omega \in \mathcal{D}} d_\omega \omega, \quad d_\omega := \langle c_\omega, a \rangle_2$$

and therefore

$$\begin{aligned} r_{m-1} &\leq n \sum_{\omega \in \mathcal{D}} |d_\omega| |\langle v_m - P_{W_{m-1}} v_m, \omega \rangle| \\ &\leq n |\langle v_m - P_{W_{m-1}} v_m, \omega_m \rangle| \sum_{\omega \in \mathcal{D}} |d_\omega| \\ &\leq n |\langle v_m - P_{W_{m-1}} v_m, \omega_m \rangle| \sum_{\omega \in \mathcal{D}} \|c_\omega\|_2. \end{aligned}$$

Next we find by Cauchy–Schwartz that

$$|\langle v_m - P_{W_{m-1}} v_m, \omega_m \rangle| = \left| \sum_{j=1}^n a_j \langle \phi_j - P_{W_{m-1}} \phi_j, \omega_m \rangle \right| \leq \left( \sum_{j=1}^n |\langle \phi_j - P_{W_{m-1}} \phi_j, \omega_m \rangle|^2 \right)^{1/2}.$$

We have thus obtained the lower bound

$$\sum_{i=1}^n |\langle \phi_i - P_{W_{m-1}} \phi_i, \omega_m \rangle|^2 \geq \frac{\kappa^2 r_{m-1}^2}{n^2 J(V_n)^2},$$

from which we conclude in a similar way as in Theorem 3.2. ■

For the general case, we establish a perturbation result similar to Theorem 3.4, with again a multiplicative constant that depends on the dimension of  $V_n$ .

**Theorem 4.3.** *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 worst case OMP algorithm on the space  $V_n$  gives*

$$(4.5) \quad r_m \leq 4 \frac{n^2 \|\Psi\|_{\ell^1(\mathcal{D})}^2}{\kappa^2} (m+1)^{-1} + n^2 \|\Phi - \Psi\|^2, \quad m \geq 1,$$

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

*Proof.* We introduce

$$t_m := r_m - n^2 \|\Phi - \Psi\|^2,$$

for which we have

$$(4.6) \quad t_m \leq t_{m-1} - \sum_{i=1}^n |\langle \phi_i - P_{W_{m-1}} \phi_i, \omega_m \rangle|^2.$$

We next write, as in the proof of Theorem 4.2,

$$v_m = \sum_{i=1}^n a_i \phi_i = \sum_{i=1}^n a_i \psi_i + \sum_{i=1}^n a_i (\phi_i - \psi_i)$$

for a vector  $a = (a_1, \dots, a_n)$  such that  $\|a\|_2 = 1$ . If  $\Psi = \sum_{\omega \in \mathcal{D}} c_\omega \omega$ , using (4.4), we now reach

$$r_{m-1} \leq n |\langle v_m - P_{W_{m-1}} v_m, \omega_m \rangle| \sum_{\omega \in \mathcal{D}} \|c_\omega\|_2 + n \sum_{j=1}^n a_j |\langle v_m - P_{W_{m-1}} v_m, \phi_j - \psi_j \rangle|.$$

By Cauchy–Schwartz and Young inequalities, the second term is bounded by  $\frac{1}{2} r_{m-1} + \frac{1}{2} n^2 \|\Phi - \Psi\|^2$ . By subtracting, we thus obtain

$$t_{m-1} \leq 2n |\langle v_m - P_{W_{m-1}} v_m, \omega_m \rangle| \sum_{\omega \in \mathcal{D}} \|c_\omega\|_2.$$

Proceeding in a similar way as in the proof of Theorem 4.2, we obtain the lower bound

$$\sum_{i=1}^n |\langle \phi_i - P_{W_{m-1}} \phi_i, \omega_m \rangle|^2 \geq \frac{\kappa^2 t_{m-1}^2}{4n^2 J(V_n)^2},$$

and we conclude in the same way as in the proof of Theorem 3.4. ■

By the exact same argument as in the proof of Corollary 3.5 we find that the worst case OMP converges for any space  $V_n$ , even when  $J(V_n)$  is not finite.

**Corollary 4.4.** *For any  $n$ -dimensional space  $V_n$ , the application of the worst case OMP algorithm on the space  $V_n$  gives that  $\lim_{m \rightarrow +\infty} r_m = 0$ .*

**4.3. Application to point evaluation.** Let us now estimate  $m(n)$  if we choose the points with the greedy algorithms that we have introduced. This boils down to estimating for  $J(V_n)$ . In this simple case,

$$J(V_n) := \|\Phi\|_{\ell^1(\mathcal{D})} = \inf \left\{ \int_{x \in [0,1]} \|c_x\|_2 \, dx : \Phi = \int_{x \in [0,1]} c_x \omega_x \, dx \right\}$$

and we can derive  $c_x$  for every  $x \in [0, 1]$  by differentiating twice the components of  $\Phi$  since

$$\Phi''(x) = \int_{y \in [0,1]} c_y \omega_y''(x) \, dy = - \int_{y \in [0,1]} c_y \delta_y(x) \, dx = -c_x.$$

Thus, using the basis functions  $\phi_k$  defined by (2.5), we have

$$J(V_n) = \int_{x \in [0,1]} \left( \sum_{k=1}^n |\phi_k''(x)|^2 \right)^{1/2} \, dx = \int_{x \in [0,1]} \left( \sum_{k=1}^n 2k\pi |\sin(k\pi x)|^2 \right)^{1/2} \, dx \sim n^{3/2}.$$

Estimate (3.5) for the convergence of the collective OMP approach yields

$$m(n) \gtrsim \frac{n^3}{\kappa^2 \sigma^2},$$

while, for the worst case OMP, estimate (4.3) gives

$$m(n) \gtrsim \frac{n^5}{\kappa^2 \sigma^2}.$$

As already anticipated, these bounds deviate from the optimal estimation (2.11) due to the use of the Hilbert–Schmidt norm in the analysis. The numerical results in the next section reveal that greedy algorithms actually behave much better in this case.

**5. Numerical tests.** All numerical tests presented in this section were performed in Python using the standard NumPy and SciPy packages.

**5.1. Pointwise evaluation.** In this test we consider the setting as outlined in section 2.2. That is, we have  $V = H_0^1(I)$ , where the interval  $I = ]0, 1[$ , the approximation space is the Fourier basis,  $V_n = \text{span}\{\phi_k : k \in 1, \dots, n\}$  where  $\phi_k(t) = \frac{\sqrt{2}}{\pi k} \sin(k\pi t)$ , and we pick the measurement functionals that define the space  $W_m$  from a dictionary of pointwise evaluations in  $]0, 1[$ . The dictionary  $\mathcal{D}$  is thus composed of elements  $\omega_x$ , the normalized Riesz representers of function evaluation at the point  $x$ , given explicitly in (2.3).

The selection of the evaluation points performed both using the collective OMP and worst case OMP algorithms. The inner product between  $\omega_x$  and  $\phi_k$  can be computed using an explicit formula, hence there is no discretization or approximation in this aspect of our implementation. In both algorithms we require a search of the dictionary  $\mathcal{D}$  to find the element that maximizes either (3.2) or (4.2). Evidently we require  $\mathcal{D}$  to be of finite size to perform this search, so we use

$$\mathcal{D} = \{\omega_x : x \in 1/M, \dots, (M-1)/M\},$$

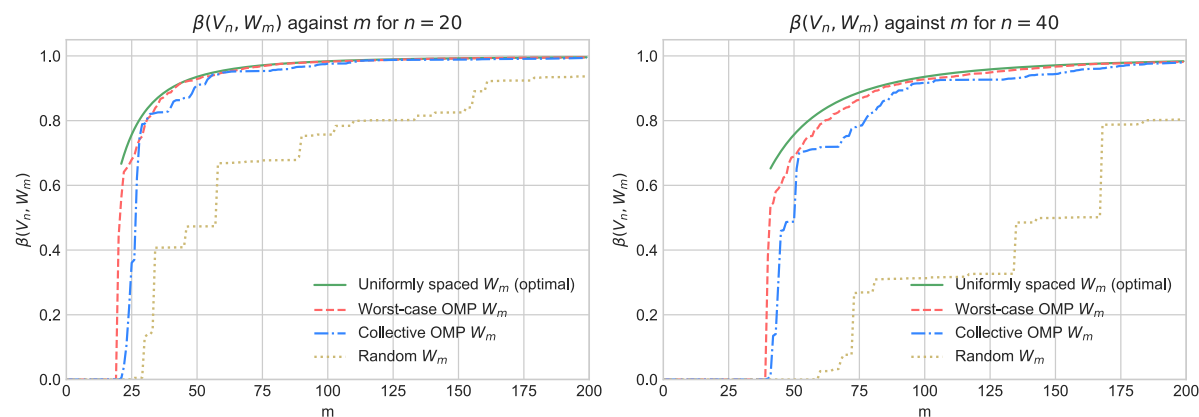
where we take  $M$  to be some large number. In practice we found that  $M = 10^4$  offered similar results to any larger number, so we kept this value.

We compare the inf-sup constant  $\beta(V_n, W_m)$  for the evaluation points picked by these two OMP algorithms with those obtained when these points are picked at random with uniform law on  $]0, 1[$ . In all cases, the selected points are nested as we increase  $m \geq n$ , that is,  $W_m \subset W_{m+1}$ , so that  $\beta$  is monotone increasing. We also compare with the value of  $\beta(V_n, W_m)$  obtained for equally spaced points, which in view of Theorem 2.3 are expected to be a near optimal choice, but are not nested.

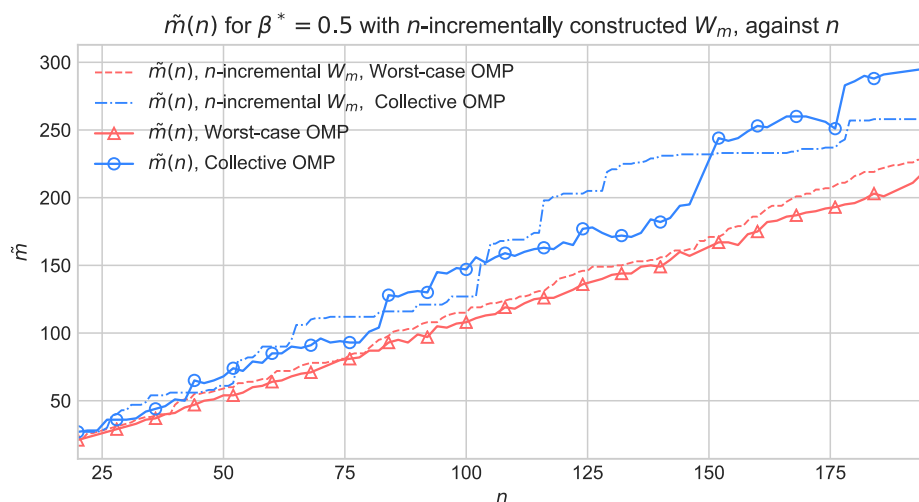
Results in Figure 5.1 show the behavior of  $\beta(V_n, W_m)$  as  $m$  increases for two representative values of  $n$ . They reveal that the worst case OMP algorithm produces a slightly better behaved inf-sup constant than the collective OMP algorithm, not far from the near optimal value obtained with evenly spaced evaluations. In contrast, the random selection is clearly suboptimal.

Figure 5.2 displays the minimum value  $\tilde{m} = \tilde{m}(n)$  required to make  $\beta(V_n, W_{\tilde{m}}) > \beta^* := 0.5$  for a variety of  $n$ . It shows a clear linear trend, with the rate of increase almost equal to 1 for the worst case OMP algorithm. This shows that the estimates on  $\tilde{m}(n)$  obtained in section 4.3 from our theoretical results are in this case too pessimistic.

We note that the obtained curves (solid lines) are not monotone increasing, due to the fact that the selected points for different values of  $n$  have no nestedness properties. An alternative strategy that leads to a nested sequence consist in intertwining the greedy algorithm with the growth of  $V_n$ : assuming that we have selected  $\tilde{m}(n)$  points such that  $\beta(V_n, W_{\tilde{m}(n)}) > \beta^*$ , we apply the greedy algorithm for  $V_{n+1}$  to enrich  $W_{\tilde{m}(n)}$  until we reach the first value  $\tilde{m}(n+1)$  such that  $\beta(V_{n+1}, W_{\tilde{m}(n+1)}) > \beta^*$ . We may again use either the collective or worst case OMP



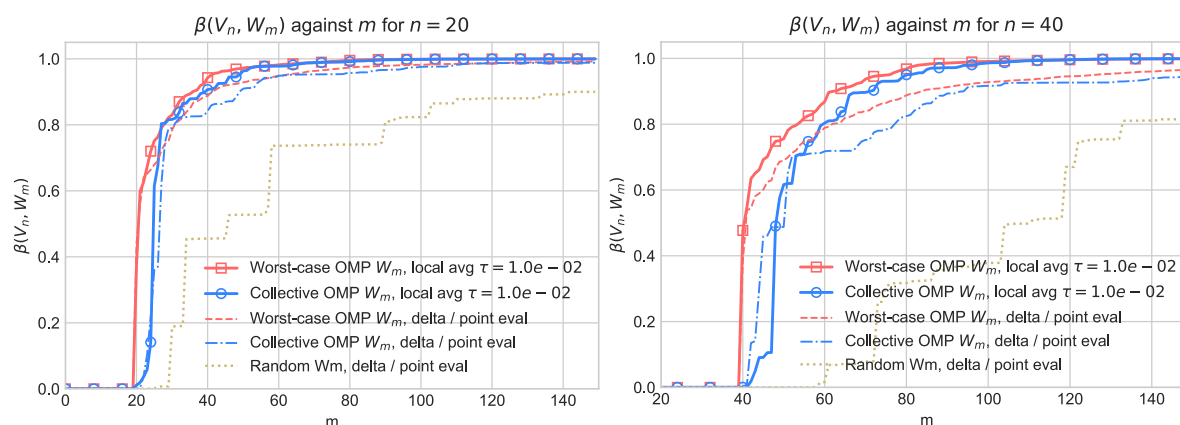
**Figure 5.1.** Comparisons of collective OMP, worst case OMP, uniformly spaced and equally spaced point selection, using a Fourier basis  $V_n$ , with  $n = 20$  or  $n = 40$ , and increasing values of  $m \geq n$ .



**Figure 5.2.** Minimum  $\tilde{m} = \tilde{m}(n)$  required for  $\beta(V_n, W_{\tilde{m}}) > \beta^* = 0.5$  against a variety of  $n$ , using collective OMP and worst case OMP, and intertwining the growth of  $V_n$  and  $W_m$  (dashed).

algorithm. In this fashion we construct a sequence of  $W_{\tilde{m}(1)} \subset \cdots \subset W_{\tilde{m}(n)} \subset \cdots$  that pair with each  $V_n$  and always ensure an inf-sup constant larger than  $\beta^*$ . This strategy bears similarity to the generalized empirical interpolation method [15] where, at each step, one adds a new function to  $V_n$  and a new linear functional to  $W_m$ . In that case, we always have  $m = n$ , but no theoretical guarantee that  $\beta(V_n, W_n)$  remains bounded away from zero.

The resulting curves are also plotted on Figure 5.2 (dashed lines). We see that we do not pay a significant penalty, as  $\tilde{m}(n)$  is not significantly worse for this incremental method than when  $W_m$  is built from nothing for each  $V_n$ . We again find that the worst case algorithm is slightly superior to the collective algorithm.



**Figure 5.3.** Comparisons of collective OMP, worst-case OMP for a dictionary of local average functions, as well as randomly selected local average locations, using a Fourier basis  $V_n$ , with  $n = 20$  or  $n = 40$ , and increasing values of  $m \geq n$ .

**5.2. Local averages.** In this test we perform the collective and worst case OMP algorithms on dictionaries of local averages. Here we build our measurement spaces  $W_m$  against a Fourier space  $V_n$ , working on the unit interval  $I = ]0, 1[$  as in section 5.1.

The dictionary  $\mathcal{D}$  is the collection of Riesz representers  $\omega_{x,\tau}$  of  $M$  local averages of width  $\tau$ , with  $x$  equispaced between  $\tau/2$  and  $1 - \tau/2$ , that is,

$$(5.1) \quad \mathcal{D} = \left\{ \omega_{x,\tau} : x = \frac{\tau}{2}, \frac{1-\tau}{M-1} + \frac{\tau}{2}, \dots, \frac{(M-2)(1-\tau)}{M-1} + \frac{\tau}{2}, 1 - \frac{\tau}{2} \right\}.$$

Again, the value  $M = 10^4$  appeared to produce satisfactory results, similar to those at any larger  $M$ .

Figure 5.3 illustrates the behavior of both greedy algorithms for the particular value  $\tau = 10^{-2}$  and shows that they are slightly better than what had been previously obtained with point values (which correspond to  $\tau = 0$ ).

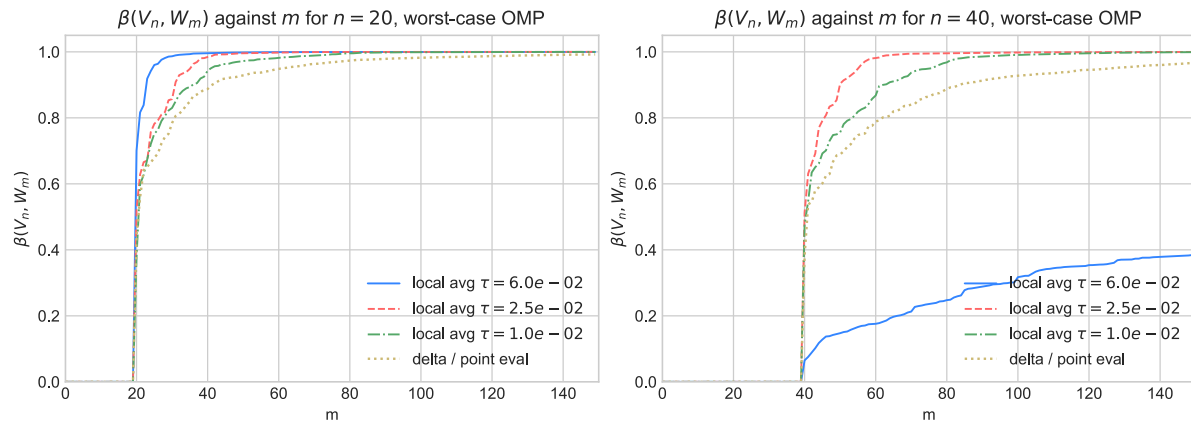
The dependence on  $\tau$ , which appears in the theoretical analysis in section 2.3, is reflected in Figure 5.4: increasing  $\tau$  first allows both OMP algorithms to obtain better  $\beta$  values, until a certain value where  $\beta$  deteriorates as  $\tau$  gets larger. In this particular case, we actually notice that if  $\tau$  is a multiple of  $2/n$ , then we have  $\langle \omega_{x,\tau}, \phi_n \rangle = 0$  for any  $x$  appearing in (5.1), and hence  $\beta(V_n, W_m) = 0$ .

**5.3. Reduced bases.** In our last test we consider the elliptic problem proposed in section 1.1, on the unit square  $D = ]0, 1]^2$  with Dirichlet boundary conditions, and a parameter dependence in the field  $a$ , that is,

$$(5.2) \quad -\operatorname{div}(a(y)\nabla u) = f \quad \text{for all } x \in D \text{ with } u(x) = 0 \text{ on } \partial D.$$

In this example we consider “checkerboard” random fields where  $a(y)$  is piecewise constant on a dyadic subdivision of the unit square. That is, for a given level  $j \geq 0$ , we consider the dyadic partition





**Figure 5.4.** The resulting  $\beta(V_n, W_m)$  against  $m$  for  $n = 20$  and  $n = 40$  and a selection of measurement widths  $\tau$ .

$$D = \bigcup_{k,l=0}^{2^j-1} S_{k,l}^{(j)}$$

with

$$S_{k,l}^{(j)} := [k 2^{-j}, (k+1) 2^{-j}] \times [\ell 2^{-j}, (\ell+1) 2^{-j}], \quad k, \ell \in 0, \dots, 2^j - 1.$$

The random field is defined as

$$(5.3) \quad a(y) = 1 + \frac{1}{2} \sum_{k,\ell=0}^{2^j-1} \chi_{S_{k,l}^{(j)}} y_{k,\ell},$$

where  $\chi_{S_{k,l}^{(j)}}$  is the indicator function on  $S_{k,l}^{(j)}$ , and the  $y_{k,\ell}$  are random coefficients that are independent, each with identical uniform distribution on  $[-1, 1]$ .

For the reduced basis space, we generate  $n$  random parameters  $y^{(1)}, \dots, y^{(n)}$ , with each  $y^{(k)} \in [-1, 1]^{2^{2j}}$ , and solve the variational form of (5.2) using  $\mathbb{P}_1$  finite elements to produce the corresponding solutions  $u_h(y^{(k)})$ . In our numerical test, we take  $j = 2$ , that is, 16 parameters, and we use for the finite element space a triangulation  $\mathcal{T}_h$  on a regular grid of mesh size  $h = 2^{-7}$ . Our approximation space is then defined as

$$(5.4) \quad V_n = \text{span} \left\{ u_h(y^{(1)}), \dots, u_h(y^{(n)}) \right\}.$$

Regarding the dictionary  $\mathcal{D}$ , we consider local averages by the nodal basis functions of the finite element space, so the linear form  $\ell_x$  in this dictionary are indexed by the mesh points of  $\mathcal{T}_h$ .

Here, we compare the performance of the above reduced basis space which we label here  $V_n^{\text{red}}$ , with the trigonometric polynomial spaces

$$V_n^{\text{sin}} = \text{span} \{ \phi_{k,\ell} : 1 \leq k \times \ell \leq n \},$$

where the  $\phi_{k,\ell}$  are given by the linear interpolation on  $\mathcal{T}_h$  of  $\frac{\sqrt{2}}{\pi\sqrt{k^2+\ell^2}} \sin(k\pi x_1) \sin(\ell\pi x_2)$  for  $k, \ell = 1, \dots, r$  and  $n = r^2$ .

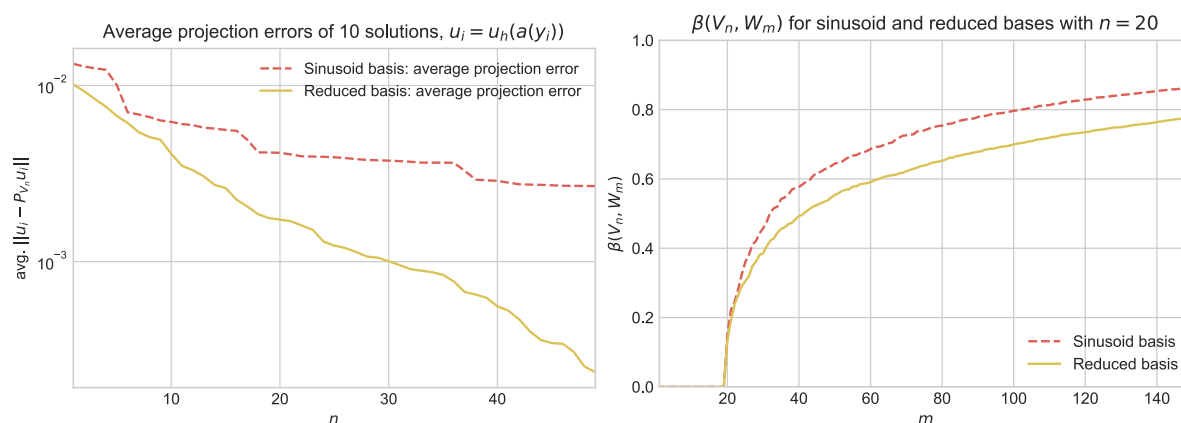


Figure 5.5. Results on the unit square.

We recall that the worst case performance of the state estimation algorithm as defined in (1.13) is given by the product of the inverse inf-sup constant  $\mu(V_n, W_m)$  by the approximation error  $\varepsilon_n = \text{dist}(\mathcal{M}, V_n)$ . Since the exact computation of  $\varepsilon_n$  is out of reach, we instead study the average projection error for a collection of solutions  $u_h(a(y))$  to  $V_n = V_n^{\text{red}}$  or  $V_n^{\text{sin}}$ . The left side of Figure 5.5 shows that the reduced bases outperform the trigonometric polynomial spaces by several order of magnitude, with respect to the decay of this approximation error. On the other hand, the right side of Figure 5.5 shows (here in the case  $n = 20$ ) that when applying the greedy algorithm, the inf-sup constant  $\beta(V_n, W_m)$  is better behaved for the trigonometric polynomial spaces, but only by a moderate factor of around 1.1. Therefore the final tradeoff is clearly in favor of reduced basis spaces.

**5.4. An adversary case: Hyperbolic PDEs.** In the case of parametrized elliptic PDEs, reduced bases are known to bring significant improvements over more conventional approximation spaces, as illustrated in the previous example by the left side of Figure 5.5. In this example, the solutions typically exhibit singularities which prevent the Fourier method from performing; however, these singularities are at fixed locations independently of the parameter value, which make the reduced basis particularly effective for capturing them.

The situation is quite different when the locations of the singularities vary together with the parameters, in which case reduced basis approximation cannot be as effective. This typically occurs for hyperbolic PDEs when parameters influence the transport velocity and therefore the positions of shocks. As a toy example, let us consider the univariate transport equation

$$(5.5) \quad \partial_t u(x, t) + y \partial_x u(x, t) = 0, \quad x \in \mathbb{R}, \quad t \geq 0,$$

with initial value  $u_0(x) = \chi_{[-\infty, 0]}$  and parameter  $y \in [0, 1]$ . We consider the parametrized family of solutions at time  $t = 1$  restricted to  $x \in [0, 1]$ , that is,

$$(5.6) \quad \mathcal{M} = \{\chi_{[0, y]} : y \in [0, 1]\}.$$

Here, we choose to work in the Hilbert space  $V = L^2([0, 1])$  which contains such discontinuous functions. Due to the presence of discontinuities, it can easily be seen that if  $V_n$  is the Fourier space generated by the functions  $\phi_k$  in (2.5), one has the approximation rate

$$(5.7) \quad \sup_{u \in \mathcal{M}} \|u - P_{V_n} u\| \sim n^{-1/2}.$$

On the other hand, a reduced basis space is of the form

$$(5.8) \quad V_n = \text{span}\{\chi_{[0, y_i]} : i = 1, \dots, n\}$$

for some points  $y_1, \dots, y_n \in [0, 1]$ , and therefore a space of piecewise constant functions on the intervals  $[y_i, y_{i+1}]$ , assuming that these points have been increasingly ordered. By taking a  $y$  to be the midpoint of the largest of such intervals, that has length larger than  $n^{-1}$ , it is easily checked that

$$(5.9) \quad \|\chi_{[0, y]} - P_{V_n} \chi_{[0, y]}\| \geq \frac{1}{2} n^{-1/2},$$

and therefore the approximation rate is not better than with the Fourier space. More generally it can be checked that the Kolmogorov  $n$ -width of  $\mathcal{M}$  in  $V$  decays like  $n^{-1/2}$ , that is, any linear approximation method cannot have a better rate.

This is illustrated by the left side of Figure 5.6, which shows the slow decay of the approximation error for the reduced basis spaces, slightly worse than when using Fourier spaces. It is of course still possible to apply the collective and worst case OMP algorithms in order to select measurements in this setting. Here, we use a dictionary of local averages

$$(5.10) \quad \mathcal{D} = \{\ell_{x, \varepsilon} : x \in [0, 1]\}, \quad \ell_{x, \varepsilon}(v) = \frac{1}{2\varepsilon} \int_{x-\varepsilon}^{x+\varepsilon} v,$$

which are continuous linear functionals on  $L^2$ . The behavior of the inf-sup constant is displayed on the right side of Figure 5.6, and shows that the greedy algorithm performs well, for both Fourier and reduced bases. However the recovery performance is affected by the fact that both of these linear spaces have poor approximation properties over the class  $\mathcal{M}$ .

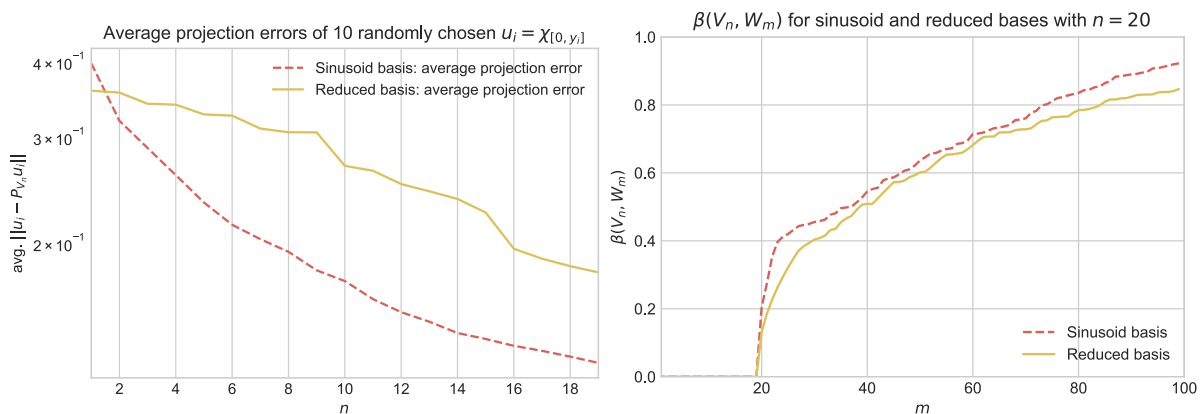


Figure 5.6. Results on the unit square.

More generally, we stress that for this class of problems, linear recovery methods can be highly suboptimal. Consider for example the case where the data are given by a single measurement in the form of the global average

$$(5.11) \quad \ell(v) = \int_0^1 v.$$

Then, we find that  $\ell(\chi_{[0,y]}) = y$ , so that an optimal reconstruction map from  $w = \ell(u)$  that gives exact recovery is simply given by

$$(5.12) \quad A^*(w) = \chi_{[0,w]}.$$

This map is nonlinear since  $\chi_{[0,w_1]} + \chi_{[0,w_2]}$  obviously differs from  $\chi_{[0,w_1+w_2]}$ .

## REFERENCES

- [1] S. E. AIDAROUS, M. R. GEVERS, AND M. J. INSTALLE, *Optimal sensors allocation strategies for a class of stochastic distributed systems*, Int. J. Control, 22 (1975), pp. 197–213.
- [2] A. BARRON, A. COHEN, W. DAHMEN, AND R. DEVORE, *Approximation and learning by greedy algorithms*, Ann. Statist., 2008.
- [3] A. BENSOUSSAN, *Optimization of sensors location in a linear filtering problem*, in Stability of Stochastic Dynamical Systems, Proceedings of the International Symposium, 1972, Lect. Notes Math. 294, Springer, Berlin, 1972, pp. 62–84.
- [4] P. BINEV, A. COHEN, W. DAHMEN, R. DEVORE, G. PETROVA, AND P. WOJTASZCZYK, *Convergence rates for greedy algorithms in reduced basis methods*, SIAM J. Math. Anal., 43 (2011), pp. 1457–1472.
- [5] P. BINEV, A. COHEN, W. DAHMEN, R. DEVORE, G. PETROVA, AND P. WOJTASZCZYK, *Data assimilation in reduced modeling*, SIAM J. Uncertainty Quantification, 5 (2017), pp. 1–29.
- [6] J. R. CANNON AND R. E. KLEIN, *Optimal selection of measurement locations in a conductor for approximate determination of temperature distributions*, J. Dyn. Syst. Meas. Control, 93 (1971), pp. 193–199.
- [7] A. COHEN AND R. DEVORE, *Approximation of high dimensional parametric pdes*, Acta Numerica, 2015.
- [8] A. COHEN, R. DEVORE, AND C. SCHWAB, *Analytic regularity and polynomial approximation of parametric stochastic elliptic PDEs*, Anal. Appl., 9 (2011), pp. 11–47.
- [9] R. DEVORE, G. PETROVA, AND P. WOJTASZCZYK, *Greedy algorithms for reduced bases in Banach spaces*, Constr. Approx., 37 (2013), pp. 455–466.
- [10] R. DEVORE AND V. TEMLYAKOV, *Some remarks on greedy algorithms*, Adv. Comput. Math., 5 (1996), pp. 173–187.
- [11] M. GRIEBEL AND P. OSWALD, *Stochastic subspace correction in Hilbert space*, preprint, Univ. Bonn, 2017.
- [12] Y. MADAY, A. T. PATERA, J. D. PENN, AND M. YANO, *A parametrized-background data-weak approach to variational data assimilation: Formulation, analysis, and application to acoustics*, Int. J. Numer. Methods Engrg., 102 (2015), pp. 933–965.
- [13] H. RAUHUT, *Random sampling of sparse trigonometric polynomials*, Appl. Comput. Harmon. Anal., 22 (2007), pp. 16–42.
- [14] V. TEMLYAKOV, *Greedy Approximation*, Cambridge University Press, Cambridge, 2012.
- [15] Y. MADAY, O. MULA, A. T. PATERA, AND M. YANO, *The generalized empirical interpolation method: Stability theory on Hilbert spaces with an application to the Stokes equation*, Comput. Methods Appl. Mech. Engrg., 287 (2015), pp. 310–334.
- [16] T. K. YU AND J. H. SEINFELD, *Observability and optimal measurement location in linear distributed parameter systems*, Int. J. Control, 18 (1973), pp. 785–799.