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

Peter Binev, Albert Cohen, Olga Mula, James Nichols

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

Peter Binev, Albert Cohen, Olga Mula, James Nichols *

November 17, 2017

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 ℓ_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 [9, 3], 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 ε_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 ℓ_j can be picked from a given dictionary $\mathcal{D}$ we introduce and analyze greedy algorithms that perform a sub-optimal 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.

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

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

where the ℓ_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

$$W_m := \operatorname{span}\{\omega_1, \dots, \omega_m\} \tag{1.2}$$

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

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

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

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

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

where $\mathcal{P}$ is a differential operator, a is a parameter that describes some physical property and lives in a given set $\mathcal{A}$. We assume that there exists a unique solution $u = u(a)$ for each $a \in \mathcal{A}$, but do not know the particular solution of interest. For example, one could consider the diffusion equation

$$-\mathrm{div}(a \nabla u) = f, \tag{1.5}$$

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

$$rI \leq a(x) \leq RI, \quad x \in D, a \in \mathcal{A}, \tag{1.6}$$

for some $0 < r \leq R < \infty$, the solution $u(a)$ is well defined in $V = H_0^1(D)$.

Therefore, in such a scenario, our prior on u is that it belongs to the set

$$\mathcal{M} := \{u(a) \ : \ a \in \mathcal{A}\}. \tag{1.7}$$

which is sometimes called the solution manifold.

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 $y = (y_1, \ldots, y_d)$, using polynomial approximations of the form

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

where Λ_n is a conveniently chosen set of multi-indices such that $\#(\Lambda_n) = n$. See in particular [4, 5] where convergence estimates of the form

$$\sup_{y \in U} \|u(y) - u_n(y)\| \leq C n^{-s}, \tag{1.9}$$

are established for some $s > 0$ even when $d = \infty$. Thus, all solutions $\mathcal{M}$ are approximated in the space $V_n := \mathrm{span}\{v_\nu \ : \ \nu \in \Lambda_n\}$. Another typical instance is provided by reduced bases. In such approximations, one directly builds a family of n-dimensional spaces $V_n := \mathrm{span}\{v_1, \ldots, v_n\}$ 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 [2, 6].

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

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

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

$$\mathcal{K} = \{v \in V \ : \ \operatorname{dist}(v, V_n) \leq \varepsilon_n\}. \tag{1.11}$$

for some given n. This point of view is adopted in [9] and further analyzed in [3] where the optimal recovery solution $u^*(w)$ from the data w is characterized. This solution is defined as the center of the ellipsoid

$$\mathcal{K}_w = \{v \in \mathcal{K} \ : \ P_W v = w\}, \tag{1.12}$$

and equivalently given by

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

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

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

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

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

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 [3] that $\beta(V_n, W_m)$ can be computed as the smallest singular value of the $n \times m$ cross-Grammian 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.

1.3 Optimal measurement selection

For a given reduced model space V_n with accuracy ε_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 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 ℓ_i are picked within 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 ω of the above linear functionals ℓ. With such an identification, $\mathcal{D}$ is 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 $\{\omega_1, \ldots, \omega_m\} \in \mathcal{D}$ in such a way that

$$\beta(V_n, W_m) \geq \beta^* > 0, \tag{1.16}$$

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

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

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

We start §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 β^* 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

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

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

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

For this reason, the exhaustive search of the set $\{\omega_1, \ldots, \omega_m\} \subset \mathcal{D}$ 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 ω_j are picked incrementally.

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

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

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 ϕ_1 that generates V_1. Two algorithms are proposed and analyzed in §3 and §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 §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 §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.

2 A-priori analysis of measurement selection strategies

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 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 $\{\omega_1, \ldots, \omega_n\}$ from $\mathcal{D}$ such that $\beta(V_n, W_n) > 0$.*

Proof: We define the ω_i inductively, together with certain functions ϕ_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

$$\langle \phi_1, \omega_1 \rangle \neq 0. \tag{2.1}$$

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

$$\langle \phi_k, \omega_k \rangle \neq 0. \tag{2.2}$$

With such a selection procedure, we find that the cross-Grammian matrix $(\langle \omega_i, \phi_j \rangle)_{i,j=1,\ldots,n}$ is lower-triangular with non-zero diagonal entries. These both show that $\{\omega_1, \ldots, \omega_n\}$ and $\{\phi_1, \ldots, \phi_n\}$ are bases, and that there is no non-trivial 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 $\{\omega_1, \ldots, \omega_n\}$ from $\mathcal{D}$, one has $\beta(V_n, W_n) \leq \varepsilon$.*

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 δ_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 δ_x is given by the solution of $\omega'' = \delta_x$ with zero boundary condition. Normalising this solution ω it with respect to the V-norm (1.18), we obtain

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

For any set of m distinct points $0 < x_1 < \cdots < x_m < 1$, the associated measurement space $W_m = \operatorname{span}\{\omega_{x_1}, \ldots, \omega_{x_m}\}$ coincides with the space of piecewise affine polynomials with nodes at $x_1, \ldots, x_m$ that vanish at the boundary. Denoting $x_0 := 0$ and $x_{m+1} := 1$, we have

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

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

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

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 $i = 1, \ldots, m$, 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,*

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

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

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

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

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

and therefore

$$\|v - P_{W_m} v\| \le hn\|v\|, \quad v \in V_n. \tag{2.9}$$

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

$$\|v - P_{W_m} v\| \le \frac{n}{m+1}\|v\|. \tag{2.10}$$

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

$$m \ge \frac{n}{\sigma^*} - 1 \tag{2.11}$$

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.

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 non-zero point spread, it is natural to model them by linear functionals that are local averages rather than point evaluations, that is

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

where

$$\varphi_\tau(y) := \tau^{-d}\varphi\Big(\frac{y}{\tau}\Big), \tag{2.13}$$

for some fixed radial function φ compactly supported in the unit ball $B = \{|x| \le 1\}$ of $\mathbb{R}^d$ and such that $\int \varphi = 1$, and $\tau > 0$ representing the point spread. Here, we assume in addition that φ belongs to $H_0^1(B)$. Taking the measurement point x at a distance at least τ 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

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

with homogeneous Dirichlet boundary conditions.

For simplicity we work in dimension $d \leq 3$. We consider measurement points $x_1, \ldots, x_m$ 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 meshsize $h > 0$. Therefore

$$m \sim h^{-d}. \tag{2.15}$$

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 2.4 *One has the estimate for the projection error in the V-norm*

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

where C is independent of (h, τ).

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

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

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 ψ_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

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

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

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

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

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

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 $w \in L^2(D)$,

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

Here, the convolution of w is meant in the sense

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

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

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

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

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

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

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

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

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

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

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

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

Remark 2.5 *The estimate* (2.16) *deteriorates for too small or too large* τ*. 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

$$\phi_k \coloneqq \sin(\pi k \cdot x), \quad k = (k_1, \ldots, k_d), \quad 1 \leq k_i \leq K, \tag{2.27}$$

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

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

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

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

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

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

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

$$C\Big(\tau + \frac{m^{-2/d}}{\tau}\Big) n^{\frac{1}{d}} \le \sigma^*, \tag{2.31}$$

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

In this analysis, the number m of required measurement deteriorates for large or small values of τ, due to the deterioration of the approximation estimate already noted in Remark 2.5. This fact will be confirmed in the numerical experiments given in §5.2. In the case where τ is of the optimal order of $\tau \sim h \sim m^{-1/d}$, the above condition becomes $C(n/m)^{\frac{1}{d}} \le \sigma^*$, which shows that $\beta(V_n, W_m) \ge \beta^*$ can be achieved with a number of measurement m^* that scales linearly with n, similar to univariate pointwise evaluation described by Theorem 2.3.

3 A collective OMP algorithm

3.1 Description of the algorithm

Note that

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

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

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 ϕ_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

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

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 ω_k such that

$$|\langle \omega_k, \phi_1 - P_{W_{k-1}} \phi_1 \rangle| \ge \kappa \max_{\omega \in \mathcal{D}} |\langle \omega, \phi_1 - P_{W_{k-1}} \phi_1 \rangle|, \tag{3.2}$$

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 [1, 7], see also [10] 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 iterately select

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

Note that in the case $n = 1$, we obtain the original OMP algorithm applied to ϕ_1.

As to the implementation of this algorithm, we take $(\phi_1, \ldots, \phi_n)$ 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 $(\phi_1, \ldots, \phi_n)$ 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 ω_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

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

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 σ_m.*

3.2 Convergence analysis

By analogy to the analysis of OMP provided in [7], we introduce for any $\Psi = (\psi_1, \ldots, \psi_n) \in V^n$ the quantity

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

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

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

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

Given that $\Phi = (\phi_1, \ldots, \phi_n)$ 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 Φ: if $\tilde{\Phi} = (\tilde{\phi}_1, \ldots, \tilde{\phi}_n)$ 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 $\Phi = (\phi_1, \ldots, \phi_n)$ 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 [7] on the OMP algorithm in the following way.

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

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

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 Cauchy-Schwarz inequality implies

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

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

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

from which (3.4) follows by the same elementary induction argument as used in [7]. □

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

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

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 [1] for the standard OMP algorithm.

Theorem 3.4 *Let* $\Phi = (\phi_1, \ldots, \phi_n)$ *be an orthonormal basis of* V_n *and* $\Psi = (\psi_1, \ldots, \psi_n) \in V^n$ *be arbitrary. Then the application of the collective OMP algorithm on the space* V_n *gives*

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

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

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

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

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

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

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

which combined with (3.7) gives

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

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

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

which is the announced result. This argument requires that for the first step we have $t_0 \leq \frac{4\|\Psi\|^2_{\ell^1(\mathcal{D})}}{\kappa^2}$. However, if $t_0 \geq \frac{4\|\Psi\|^2_{\ell^1(\mathcal{D})}}{\kappa^2}$, 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* $\lim_{m\to+\infty} r_m = 0$.

Proof: By completeness of $\mathcal{D}$, for any δ, there exists a finite subset $\mathcal{F} \subset \mathcal{D}$ and vector coefficients c_ω 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

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

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 ϕ_j. This leads for each $j \in \{1, \ldots, n\}$ to the selection of $\omega_{1,j}, \ldots, \omega_{m_j,j} \in \mathcal{D}$ and to a space*

$$W_m := \operatorname{span}\{\omega_{k,j} \,:\, k = 1, \ldots, m_j,\ j = 1, \ldots m\}, \quad m_1 + \cdots + m_j = m.$$

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

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

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*

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

Remark 3.7 *In the case $n = 1$, a well-known variant to the OMP algorithm, also discussed in [1], is the so-called* relaxed greedy algorithm*. This variant avoids the computation of the projection onto W_k: the approximation of ϕ_1 is updated by*

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

where ω_k is selected by maximizing $|\langle \phi_1 - A_{k-1}\phi_1, \omega \rangle|$ over $\omega \in \mathcal{D}$ and (α_k, β_k) are appropriate weights. This strategy is generalized in [8] to the collective setting (with $\alpha_k = 1 - (k+1)^{-1}$ and β_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

4.1 Description of the algorithm

In this algorithm, we first take

$$v_k := \operatorname{argmax}\Big\{\|v - P_{W_{k-1}} v\| \,:\, v \in V_n,\, \|v\| = 1\Big\}, \tag{4.1}$$

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

$$|\langle v_k - P_{W_{k-1}} v_k, \omega_k\rangle| \geq \kappa \max\Big\{|\langle v_k - P_{W_{k-1}} v_k, \omega\rangle| \,:\, \omega \in \mathcal{D}\Big\}, \tag{4.2}$$

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

In our numerical experiments, this algorithm 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.

Remark 4.1 *A variant to this algorithm was priorily suggested in [9] 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 φ_τ a Gaussian of fixed width (see algorithm 2, 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, however 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*

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

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

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

We have $v_m = \sum_{i=1}^n a_j \phi_j$ for a vector $a = (a_1, \ldots, a_n)$ 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

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

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

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

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

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

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

$$v_m = \sum_{i=1}^n a_j \phi_j = \sum_{i=1}^n a_j \psi_j + \sum_{i=1}^n a_j (\phi_j - \psi_i),$$

for a vector $a = (a_1, \ldots, a_n)$ 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 inequality, 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 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* $\lim_{m\to+\infty} r_m = 0$.

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 estimate for $J(V_n)$. In this simple case,

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

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

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

Thus, using the basis functions ϕ_k defined by (2.5), we have

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

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 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 §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 ω_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 ω_x and ϕ_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 β 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.

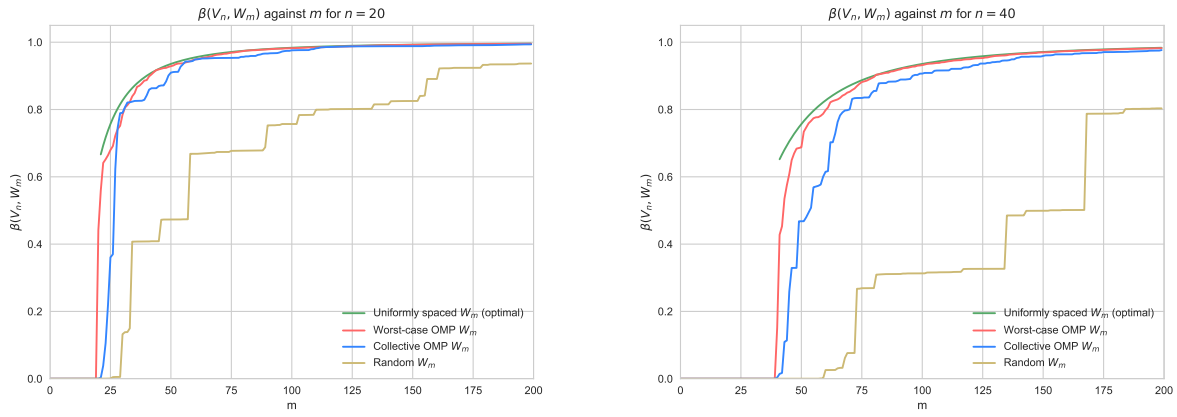

Figure 5.1: Comparisons of collective OMP, worst-case OMP, uniformly spaced and equally spaced point selection, using a Fourier basis V_n, with $n = 20$ or $n = 40$, and increasing values of $m \geq n$.

Results in Fig 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

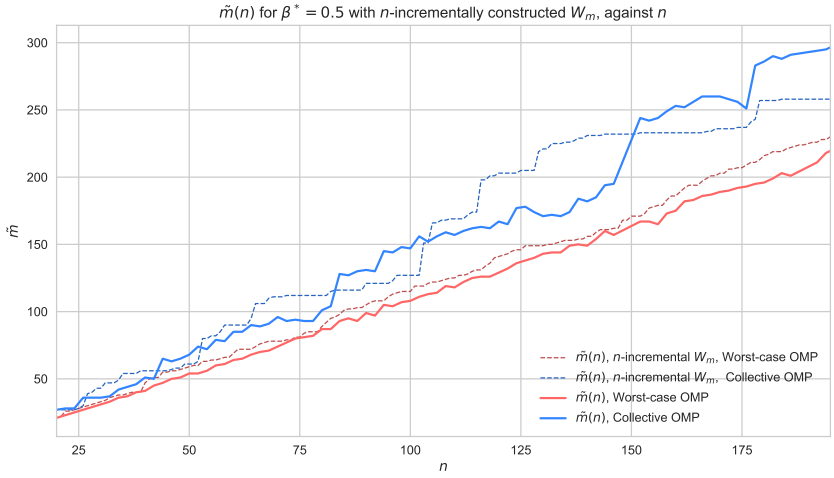

Figure 5.2: Minimum $\tilde{m} = \tilde{m}(n)$ required for $\beta(V_n, W_{\tilde{m}}) > \beta^* = 0.5$ against a variety of n, using collective OMP and worst-case OMP, and intertwining the growth of V_n and W_m (dashed).

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.

Fig 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 §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 algorithm. In this fashion we construct a sequence of $W_{\tilde{m}(1)} \subset \ldots \subset W_{\tilde{m}(n)} \subset \ldots$ that pair with each V_n and always ensure an inf-sup constant larger than β^*. This strategy bears similarity to the generalized empirical interpolation method [11] 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 Fig 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 that 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.

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 §5.1.

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

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

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

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

The dependence on τ, which appears in the theoretical analysis in §2.3 is reflected in Figure 5.4: increasing τ first allows both OMP algorithms to obtain better β values, until a certain value where β deteriorates as τ gets larger. In this particular case, we actually notice that if τ 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 $\beta(V_n, W_m) = 0$.

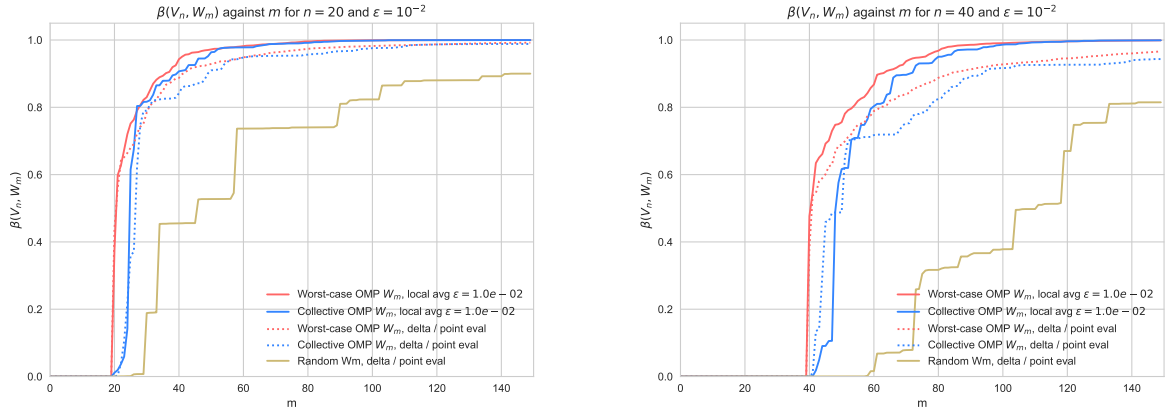

Figure 5.3: Comparisons of collective OMP, worst-case OMP for a dictionary of local average functions, as well as randomly selected local average locations, using a Fourier basis V_n, with $n = 20$ or $n = 40$, and increasing values of $m \geq n$.

5.3 Reduced bases

In our last test we consider the elliptic problem proposed in §1.1, on the unit square $D =]0,1[^2$ with Dirichelet boundary conditions, and a parameter dependence in the field a, that is

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

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

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

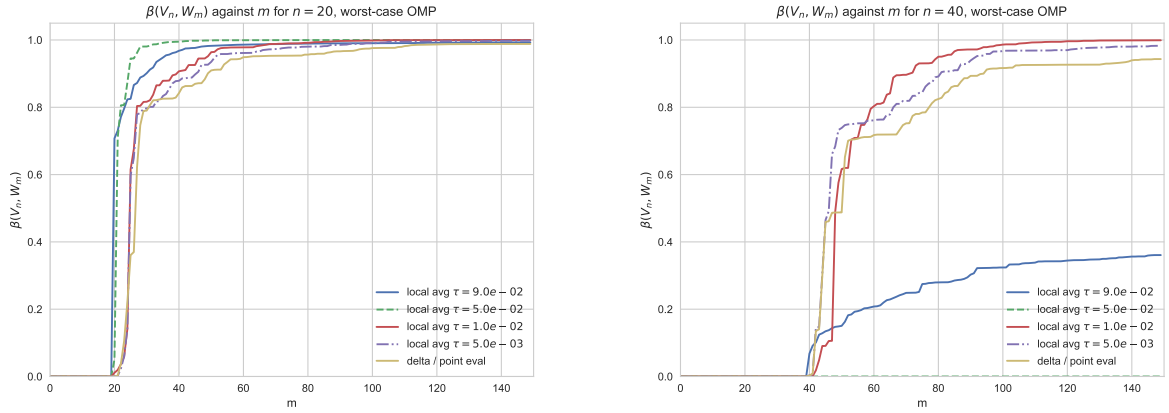

Figure 5.4: The resulting $\beta(V_n, W_m)$ against m for $n = 20$ and $n = 40$ and a selection of measurement widths τ.

with

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

The random field is defined as

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

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

$$V_n = \operatorname{span}\{u_h(y^{(1)}), \dots, u_h(y^{(n)})\} \tag{5.4}$$

As to the dictionary $\mathcal{D}$, we consider local averages by the nodal basis functions of the finite element space, so the linear form ℓ_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^{red}, with the trigonometric polynomial spaces

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

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 $k, l = 1, \dots, r$ and $n = r^2$.

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 $\varepsilon_n = \operatorname{dist}(\mathcal{M}, V_n)$. Since the exact computation of ε_n is out of reach, we instead study the the

average projection error for a collection of solutions $u_h(a(y))$ to $V_n = V_n^{\text{red}}$ or V_n^{sin}. The left side of Figure 5.5 shows that the reduced bases outperform the trigonometric polynomial spaces by several order of magnitude, as 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, however only by a moderate factor of around 1.1. Therefore the final trade-off is clearly in favor of reduced basis spaces.

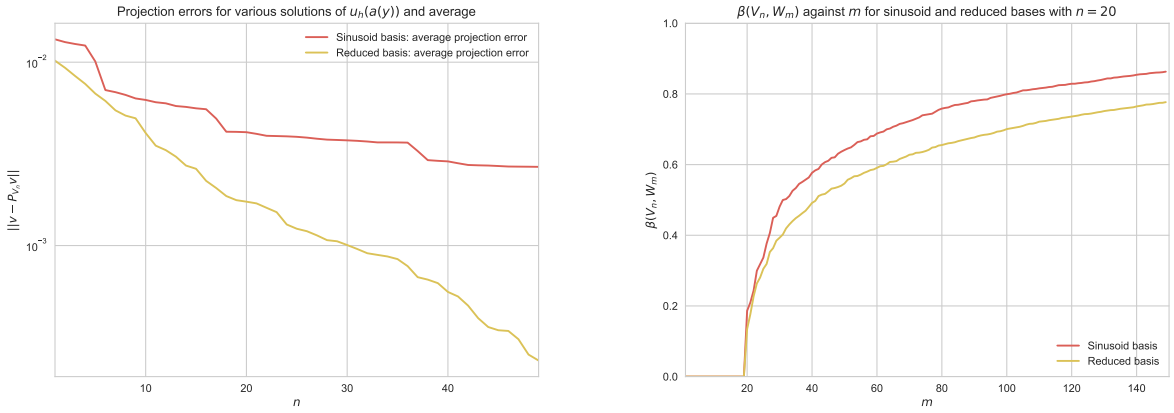

Figure 5.5: Results on unit square.

References

[1] A. Barron, A. Cohen, W. Dahmen and R. DeVore, *Approximation and learning by greedy algorithms*, Annals of Statistics, 2008.

[2] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Convergence Rates for Greedy Algorithms in Reduced Basis Methods*, SIAM Journal of Mathematical Analysis, **43**, 1457-1472, 2011.

[3] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Data assimilation in reduced modeling*, SIAM J. on Uncertainty Quantification, **5**, 1-29, 2017.

[4] A. Cohen and R. DeVore, *Approximation of high dimensional parametric pdes*, Acta Numerica, 2015.

[5] A. Cohen, R. DeVore and C. Schwab, *Analytic Regularity and Polynomial Approximation of Parametric Stochastic Elliptic PDEs*, Analysis and Applications **9**, 11-47, 2011.

[6] R. DeVore, G. Petrova, and P. Wojtaszczyk, *Greedy algorithms for reduced bases in Banach spaces*, Constructive Approximation, **37**, 455-466, 2013.

[7] R. DeVore and V. Temlyakov, *Some remarks on greedy algorithms*, Adv. Comp. Math. 5, 173-187, 1996.

[8] M. Griebel and P. Oswald, *Stochastic subspace correction in Hilbert space*, preprint, Univ. Bonn, 2017.

[9] 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. Meth. Eng., **102**, 933-965, 2015.

[10] V. Temlyakov, *Greedy approximation*, Cambridge University Press, Cambridge,

[11] 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*, Comp. Meth. Appl. Mech. Eng., **287**, 310-334, 2015.