# Optimal reduced model algorithms for data-based state estimation

Albert Cohen, Wolfgang Dahmen, Ron DeVore, Jalal Fadili, Olga Mula and James Nichols [*]

March 20, 2019

**Abstract**

Reduced model spaces, such as reduced basis and polynomial chaos, are linear spaces $V_n$ of finite dimension $n$ which are designed for the efficient approximation of families parametrized PDEs in a Hilbert space $V$. The manifold $\mathcal{M}$ that gathers the solutions of the PDE for all admissible parameter values is globally approximated by the space $V_n$ with some controlled accuracy $\varepsilon_n$, which is typically much smaller than when using standard approximation spaces of the same dimension such as finite elements. Reduced model spaces have also been proposed in [13] as a vehicle to design a simple linear recovery algorithm of the state $u \in \mathcal{M}$ corresponding to a particular solution instance when the values of parameters are unknown but a set of data is given by $m$ linear measurements of the state. The measurements are of the form $\ell_j(u)$, $j = 1, \dots, m$, where the $\ell_j$ are linear functionals on $V$. The analysis of this approach in [2] shows that the recovery error is bounded by $\mu_n \varepsilon_n$, where $\mu_n = \mu(V_n, W)$ is the inverse of an inf-sup constant that describe the angle between $V_n$ and the space $W$ spanned by the Riesz representers of $(\ell_1, \dots, \ell_m)$. A reduced model space which is efficient for approximation might thus be uneffective for recovery if $\mu_n$ is large or infinite. In this paper, we discuss the existence and effective construction of an optimal reduced model space for this recovery method. We extend our search to affine spaces which are better adapted than linear spaces for various purposes. Our basic observation is that this problem is equivalent to the search of an optimal affine algorithm for the recovery of $\mathcal{M}$ in the worst case error sense. This allows us to peform our search by a convex optimization procedure. Numerical tests illustrate that the reduced model spaces constructed from our approach perform better than the classical reduced basis spaces.

# 1 Introduction

## 1.1 The state estimation problem

This paper is concerned with the *sensing* or *recovery* problem in a Hilbert space $V$ equiped with some norm $\|\cdot\|$ and inner product $\langle \cdot, \cdot \rangle$: we want to recover an approximation to an unknown function $u \in V$ from data given by $m$ linear measurements

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

where the $\ell_i$ are $m$ linearly independent linear functionals over $V$. This problem appears in many different setting. The particular setting that motivates our work is the case where $u = u(y)$ represents the *state* of physical system described as a solution to a parametric PDE

$$\mathcal{P}(u, y) = 0 \tag{1.2}$$

[*] This research was supported by the Institut Universitaire de France; the ERC Adv grant BREAD;

for some unknown finite or infinite dimensional parameter vector $y = (y_j)_{j\geq 1}$ picked from some admissible set $Y$. The $\ell_i$ are a mathematical model for sensors that capture some partial information on the unknown solution $u(y) \in V$.

Denoting by $\omega_i \in V$ the Riesz representers of the $\ell_i$, such that $\ell_i(v) = \langle v, \omega_i\rangle$ for all $v \in V$, and defining

$$W := \text{span}\{\omega_1, \dots, \omega_m\}, \tag{1.3}$$

the measurements are equivalently represented by

$$w = P_W u. \tag{1.4}$$

where $P_W$ is the orthongal projection from $V$ onto $W$. A *recovery algorithm* is a computable map

$$A : W \to V \tag{1.5}$$

and the approximation to $u$ obtained by this algorithm is

$$u^* = A(w) = A(P_W u). \tag{1.6}$$

**Remark 1.1** *Given any recovery algorithm $A$, we can always decompose $A(w)$ into its orthogonal components in $W$ and $W^\perp$. Since $w$ is known to us, for $A$ to be optimal it should have the form*

$$A(w) = w + B(w), \tag{1.7}$$

*where $B : W \to W^\perp$ with $W^\perp$ the orthogonal complement of $W$ in $V$. Therefore, in going further in this paper, we always require that $A$ has the form* (1.7) *and concentrate on the construction of good maps $B$.*

The construction of $A$ or $B$ should be based on the available prior information that describes the properties of the unknown $u$, and the evaluation of its performance needs to be defined in some precise sense. Two distinct approaches are usually followed:

- In the *deterministic setting*, the sole prior information is that $u$ belongs to the set

  $$\mathcal{M} := \{u(y) \ : \ y \in Y\}, \tag{1.8}$$

  of all possible solutions. The set $\mathcal{M}$ is sometimes called the *solution manifold.* The performance of an algorithm $A$ over the class $\mathcal{M}$ is usually measured by the "worst case" reconstruction error

  $$E_{\rm wc}(A, \mathcal{M}) = \sup\{\|u - A(P_W u)\| \ : \ u \in \mathcal{M}\}. \tag{1.9}$$

  The problem of finding an algorithm that minimizes $E_{\rm wc}(A)$ is called *optimal recovery.* It has been extensively studied for convex sets $\mathcal{M}$ that are balls of smoothness classes [4, 14, 15], which is not the case of (1.8).

- In the *stochastic setting*, the prior information on $u$ is described by a probability distribution $p$ on $V$, which is supported on $\mathcal{M}$, typically induced by a probability distribution on $Y$ that is assumed to be known. It is then natural to measure the performance of an algorithm in an averaged sense, for example through the mean-square error

  $$E_{\rm ms}(A, p) = \mathbb{E}(\|u - A(P_W u)\|^2) = \int_V \|u - A(P_W u)\|^2 dp(u). \tag{1.10}$$

  This stochastic setting is the starting point to *Bayesian estimation* methods [11]. Let us observe that for any algorithm $A$ one has $E_{\rm ms}(A, p) \leq E_{\rm wc}(A, \mathcal{M})^2$.

### 1.2 Optimal algorithms

The present paper concentrates on the deterministic setting according to the above distinction, although some remarks will be given on the analogies with the stochastic setting. In this setting, the benchmark for the performance of recovery algorithms is given by

$$E^*_{\rm wc}(\mathcal{M}) = \min_A E_{\rm wc}(A, \mathcal{M}),$$

where the minimum is taken over all possible maps $A$.

There is a simple mathematical description of an optimal map that meets this benchmark. For any bounded set $S \subset V$ we define its *Chebychev ball* as the smallest closed ball that contains $S$. The *Chebychev radius and center* denoted by $\mathrm{rad}(S)$ and $\mathrm{cen}(S)$ are the radius and center of this ball. Therefore, the information that we have on $u$ is that it belongs to the set

$$\mathcal{M}_w := \mathcal{M} \cap V_w, \quad V_w := \{v \in V \;:\; P_W v = w\} = w + W^\perp, \tag{1.11}$$

where $W^\perp$ is the orthogonal complement of $W$ in $V$. It follows that an optimal reconstruction map $A^*_{\rm wc}$ for the worst case error is given by

$$A^*_{\rm wc}(w) = \mathrm{cen}(\mathcal{M}_w), \tag{1.12}$$

since the Chebychev center of $\mathcal{M}_w$ minimizes the quantity $\sup\{\|u - v\|_V \;:\; u \in \mathcal{M}_w\}$ among all $v \in V$. The worst case error is therefore given by

$$E^*_{\rm wc}(\mathcal{M}) = E_{\rm wc}(A^*_{\rm wc}, \mathcal{M}) = \sup\{\mathrm{rad}(\mathcal{M}_w) \;:\; w \in P_W(\mathcal{M})\}. \tag{1.13}$$

Note that the map $A^*_{\rm wc}$ is also optimal among all algorithms for each $\mathcal{M}_w$, $w \in P_W(\mathcal{M})$, since

$$E_{\rm wc}(A^*_{\rm wc}, \mathcal{M}_w) = \min_A E_{\rm wc}(A, \mathcal{M}_w) = \mathrm{rad}(\mathcal{M}_w), \quad w \in P_W(\mathcal{M}). \tag{1.14}$$

However, there may exists other maps $A$ such that $E_{\rm wc}(A, \mathcal{M}) = E^*_{\rm wc}(\mathcal{M})$, since we also supremize over $w \in P_W(\mathcal{M})$.

### 1.3 Linear and affine algorithms based on reduced models

In practice the above map $A^*_{\rm wc}$ cannot be easily constructed due to the fact that the solution manifold $\mathcal{M}$ is a high-dimensional and geometrically complex object. One is therefore interested in designing "sub-optimal yet good" recovery algorithms and analyze their performance.

One vehicle constructing linear recovery mappings $A$ is to use *reduced modeling.* Generally speaking, reduced models consist of linear spaces $(V_n)_{n\geq 0}$ with increasing dimension $\dim(V_n) = n$ which uniformly approximate the solution manifold in the sense that

$$\mathrm{dist}(\mathcal{M}, V_n) := \max_{u\in\mathcal{M}} \min_{v\in V_n} \|u - v\|_V \leq \varepsilon_n, \tag{1.15}$$

where

$$\varepsilon_0 \geq \varepsilon_1 \geq \cdots \geq \varepsilon_n \geq \cdots \geq 0, \tag{1.16}$$

are known tolerances. Instances of reduced models for parametrized families of PDEs with provable accuracy are provided by polynomial approximations in the $y$ variable [8, 9] or reduced bases [5, 17, 16]. The construction of a reduced model is typically done offline, using a large training set of instances of $u \in \mathcal{M}$ called *snapshots*. The offline stage potentially has a high computational cost. Once this is done, the online cost of recovering $u^* = A(w)$ from any data $w$ using this reduced model should in contrast be moderate.

In [13], a simple reduced-model based recovery algorithm was proposed, in terms of the map

$$A_n(w) := \operatorname{argmin}\{\operatorname{dist}(v, V_n) \; : \; v \in V_w\}, \tag{1.17}$$

which is well defined provided that $V_n \cap W^\perp = \{0\}$. It turns out that $A_n$ is a linear mapping and so these algorithms are linear. In agreement with the terminology of these previous papers, we call the algorithms of the form $A_n$ a *one space algorithm*. It was shown in [2] that $A_n$ has a simple interpretation in terms of the cylinder

$$\mathcal{K}_n := \{v \in V \; : \; \operatorname{dist}(v, V_n) \leq \varepsilon_n\}, \tag{1.18}$$

that contains the solution manifold $\mathcal{M}$. Namely, the algorithm $A_n$ is also given by

$$A_n(w) = \operatorname{cen}(\mathcal{K}_{n,w}), \quad \mathcal{K}_{n,w} := \mathcal{K}_n \cap V_w. \tag{1.19}$$

It is therefore the optimal map when $\mathcal{M}$ is replaced by the simpler containement set $\mathcal{K}_n$. The substantial advantage of this approach is that, in contrast to $A^*_{\mathrm{wc}}$, the map $A_n$ can be easily computed by solving simple least-squares minimization problems which amount to finite linear systems. In turn $A_n$ is a linear map from $W$ to $V$. This map depends on $V_n$ and $W$, but not on $\varepsilon_n$ in view of (1.17). We refer to $A_n$ as the *one space algorithm* based on the space $V_n$.

This algorithm satisfies the performance bound

$$\|u - A_n(P_W u)\|_V \leq \mu_n \operatorname{dist}(u, V_n \oplus (V_n^\perp \cap W)) \leq \mu_n \operatorname{dist}(u, V_n) \leq \mu_n \varepsilon_n, \tag{1.20}$$

where the last inequality holds when $u \in \mathcal{M}$. Here

$$\mu_n = \mu(V_n, W) := \max_{v \in V_n} \frac{\|v\|}{\|P_W v\|}, \tag{1.21}$$

is the inverse of the inf-sup constant $\beta_n := \min_{v \in V_n} \max_{w \in W} \frac{\langle v,w \rangle}{\|v\| \, \|w\|}$ which describes the angle between $V_n$ and $W$. In particular $\mu_n = \infty$ in the event where $V_n \cap W^\perp$ is non-trivial.

An important observation is that the one space algorithm (1.17) has a simple extension to the setting where $V_n$ is an affine space rather than a linear space, namely, when

$$V_n = \overline{u} + \tilde{V}_n, \tag{1.22}$$

with $\widetilde{V}_n$ a linear space of dimension $n$ and $\overline{u}$ a given offset that is known to us.

**Remark 1.2** *The motivation for using affine reduced models is that they are more accurate when the solution manifold $\mathcal{M}$ is not localized near the origin. This typically happens when the parametric solution $u(y)$ is a perturbation of a nominal solution $\overline{u} = u(\overline{y})$ for some $\overline{y} \in Y$. Another perspective, currently under investigation, is to agglomerate local affine models in order to generate nonlinear reduced model. This can be executed, for example, by decomposing the parameter domain $Y$ into $K$ subdomains $Y_k$ and using different affine reduced models for approximating the resulting subsets $\mathcal{M}_k = u(Y_k)$.*

### 1.4 Objective and outline

The standard constructions of reduced models are targeted at making the spaces $V_n$ as efficient as possible for approximating $\mathcal{M}$, that is, making $\varepsilon_n$ as small as possible for each given $n$. For example, for the reduced basis spaces, it is known [1, 10] that a certain greedy selection of snapshots generates spaces $V_n$ such that $\operatorname{dist}(\mathcal{M}, V_n)$ decays at the same rate (polynomial or exponential) as the Kolmogorov $n$-width

$$\delta_n(\mathcal{M}) := \inf\{\operatorname{dist}(\mathcal{M}, E) \ : \ \dim(E) = n\}. \tag{1.23}$$

However these constructions do not ensure the control of $\mu_n$ and therefore these reduced spaces may be much less efficient when using the one space algorithm for the recovery problem.

In view of the above observations, the objective of this paper is to discuss the construction of reduced models that are better targeted towards the recovery task. In other words, we want to build the spaces $V_n$ to make the recovery algorithm $A_n$ as efficient as possible, given the measurement space $W$. Note that a different problem is, given $\mathcal{M}$, to optimize the choice of the measurement functionals $\ell_i$ picked from some admissible dictionary, which amounts to optimizing the space $W$, as discussed for example in [3]. Here, we consider our measurement system to be imposed to us, and therefore $W$ to be fixed once and for all. We extend our discussion to the version of the one space algorithm where $V_n$ is an affine space as in (1.22).

The rest of our paper is organized as follows. In §2, we detail the affine map $A_n$ associated to $V_n$, that can be computed in a similar way as in the linear case. Conversely, we show that any affine recovery map may be interpreted as a one space algorithm for a certain affine reduced model $V_n$. We show for a general set $\mathcal{M}$ the existence of an optimal affine recovery map $A^*_{\rm wca}$ for the worst case error, thus equivalent to that of an optimal reduced space for the recovery problem. We draw a short comparison with the stochastic setting in which the optimal affine map $A^*_{\rm msa}$ for the mean-square error (1.10) is derived explicitely from the second order statistics of $u$.

In §3, we compute an approximation of $A^*_{\rm wca}$ by convex optimization, based on a training set of snapshots. Two algorithms are considered: subgradient descent and primal-dual proximal splitting. Our numerical results illustrate the superiority of the latter for this problem. The optimal affine map $A^*_{\rm wca}$ significantly outperforms the one space algorithm $A_{n^*}$ when standard reduced basis spaces $V_n$ are used and an optimal value $n^*$ is selected using the training set. It also outperforms the affine map $A^*_{\rm msa}$ computed from second order statistics of the training set. All three maps significanly outperform the minimal $V$-norm recovery given by $A(w) = w = P_W u$.

## 2 Affine one space recovery

In this section, we show that any linear algorithm is given by a one space algorithm and therefore a similar result holds for any affine algorithm. We then go on to describe the optimal one space algorithms by exploiting this fact.

### 2.1 The one space algorithm

We begin by discussing in more detail the one space algorithm for a linear space $V_n$ of dimension $n \leq m$. As shown in [2], the map $A_n$ associated to $V_n$ has a simple expression after a proper choice of favorable bases has been made for $W$ and $V_n$ through an SVD applied to the cross-grammian of an initial pair of orthonormal bases. The resulting *favorable bases* $\{\psi_1, \dots, \psi_m\}$ for $W$ and $\{\varphi_1, \dots, \varphi_n\}$ for $V_n$ satisfy the equations

$$\langle \psi_i, \varphi_j \rangle = s_i \delta_{i,j}, \tag{2.1}$$

where

$$1 \geq s_1 \geq s_2 \geq \cdots \geq s_n > 0, \tag{2.2}$$

are the singular values of the cross-grammian. Then, if $w$ is in $W$, we can write $w = \sum_{j=1}^m w_j \psi_j$ in the favorable basis, and find that

$$A_n(w) = \sum_{j=1}^n s_j^{-1} w_j \varphi_j + \sum_{j=n+1}^m w_j \psi_j. \tag{2.3}$$

Let us observe that the functions $\psi_j$ in the second sum span the space $V_n^\perp \cap W$.

Now consider any linear recovery algorithm $A$. Then, $A$ has the form $A(w) = w + B(w)$ where $B$ is a linear map from $W^\perp$ into $V$. Our next observation is that $A$ can always be interpreted as a one space algorithm $A_n$ for a certain space $V_n$ with $n \leq m$.

**Proposition 2.1** *Let $A$ be any linear map of the form* (1.7). *Then, there exists a space $V_n$ of dimension $n \leq m$ such that $A$ coincides with the one space algorithm* (2.3) *for $V_n$.*

**Proof:** By considering the SVD of the linear transform $B$, there exists an orthonormal basis $\{\psi_1, \dots, \psi_m\}$ of $W$ and an orthonormal system $\{\omega_1, \dots, \omega_m\}$ in $W^\perp$ such that, with $w = \sum_{j=1}^m w_j \psi_j$,

$$Bw = \sum_{j=1}^m \alpha_j w_j \omega_j, \quad w \in W, \tag{2.4}$$

for some numbers $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_m \geq 0$. Defining the functions

$$\varphi_j = s_j(\psi_j + \alpha_j \omega_j), \quad s_j = (1 + \alpha_j^2)^{-1/2}, \tag{2.5}$$

and defining $V_n$ as the span of those $\varphi_j$ for which $\alpha_j \neq 0$, we recover the exact form (2.3) of the one space algorithm expressed in favorable bases. □

These results can be readily extended to the case where $V_n$ is an affine space given by (1.22) for some given $n$-dimensional linear space $\tilde{V}_n$ and offset $\overline{u}$. In what follows, we systematically use the notation

$$\tilde{u} = u - \overline{u}, \tag{2.6}$$

for the recentered state, and likewise $\tilde{w} = w - \overline{w}$ with $\overline{w} = P_W \overline{u}$ for the recentered observation. The one space algorithm associated to $V_n$ has the form

$$A_n(w) = \overline{u} + \tilde{A}_n(\tilde{w}), \tag{2.7}$$

where $\tilde{A}_n$ is the one space linear algorithm associated to $\tilde{V}_n$.

Performances bounds similar to those of the linear are derived in the same way as in [2]: the reconstruction satisfies

$$\|u - A_n(P_W u)\| \leq \mu_n \text{dist}(u, \overline{u} + \tilde{V}_n \oplus (\tilde{V}_n^\perp \cap W)) \leq \mu_n \text{dist}(u, V_n), \tag{2.8}$$

where

$$\mu_n = \mu(\tilde{V}_n, W) = \max_{v\in\tilde{V}_n} \frac{\|v\|}{\|P_W v\|} = s_n^{-1} < \infty. \tag{2.9}$$

The map $A_n$ is optimal for the cylinders of the form

$$\mathcal{K}_n = \{u \in V \ : \ \mathrm{dist}(u, V_n) \le \varepsilon_n\}, \tag{2.10}$$

since it coincides with the Chebychev center of $\mathcal{K}_{n,w} = \mathcal{K}_n \cap V_w$. In particular, one has

$$E^*_{\rm wc}(\mathcal{K}_n) = E_{\rm wc}(A_n, \mathcal{K}_n) = \mu_n \varepsilon_n. \tag{2.11}$$

For a solution manifold $\mathcal{M}$ contained in $\mathcal{K}_n$, one has

$$E^*_{\rm wc}(\mathcal{M}) \le E_{\rm wc}(A_n, \mathcal{M}) \le \mu_n \mathrm{dist}(\mathcal{M}, \tilde{V}_n \oplus (\tilde{V}_n^\perp \cap W)) \le \mu_n \mathrm{dist}(\mathcal{M}, V_n) \le \mu_n \varepsilon_n, \tag{2.12}$$

and these inequalities are generally strict.

In view of (2.7) the map $A_n$ is affine. A general affine recovery map takes form

$$A(w) = w + Bw + c, \tag{2.13}$$

where $B : W \to W^\perp$ is linear and $c = A(0) \in W^\perp$. The following result is a direct consequence of Proposition 2.1.

**Corollary 2.2** *Let $A$ be an affine map of the form* (2.13). *Then, there exists an affine space $V_n = \overline{u} + \tilde{V}_n$ such that $A$ coincides with the one space algorithm* (2.7).

### 2.2 The best affine map

In view of this result, the search for an affine reduced model $V_n$ that is best taylored for the recovery problem is equivalent to the search of an optimal affine map. Our next result is that such a map always exist when $\mathcal{M}$ is a bounded set.

**Theorem 2.3** *Let $\mathcal{M}$ be a bounded set. Then there exists a map $A^*_{\rm wca}$ that minimizes $E_{\rm wc}(A, \mathcal{M})$ among all affine maps $A$.*

**Proof:** We consider affine map of the form (2.13), so that the error is given by

$$E_{\rm wc}(A, \mathcal{M}) = \sup\{u \in \mathcal{M} \ : \ \|P_{W^\perp} u - c - BP_W u\|\} = F(c, B). \tag{2.14}$$

We begin by remarking that for each $(c, B) \in W^\perp \times \mathcal{L}(W, W^\perp)$, the map $u \mapsto \|P_{W^\perp} u - c - BP_W u\|$ is uniformly bounded on the bounded set $\mathcal{M}$. Its supremum $F(c, B)$ is thus a finite positive number, which we may write as

$$F(c, B) = \sup_{u\in\mathcal{M}} F_u(c, B), \tag{2.15}$$

where $F_u(c, B) = \|P_{W^\perp} u - c - BP_W u\|$. Each $F_u$ is convex and satisfies the Lipschitz bound

$$|F_u(c, B) - F_u(c', B')| \le \|c - c'\| + M\|B - B'\|_S, \tag{2.16}$$

with

$$\|B\|_S = \max\{\|Bv\| \ : \ v \in W, \ \|v\| = 1\}, \tag{2.17}$$

the spectral norm and $M := \sup\{\|P_W u\| \;:\; u \in \mathcal{M}\} < \infty$. This implies that the function $F$ is convex and satisfies the same Lipschitz bound.

We note that the linear maps of $\mathcal{L}(W, W^\perp)$ are of rank at most $m$ and therefore, given any orthonormal basis $(e_1, \dots, e_m)$ of $W$, we can equip $\mathcal{L}(W, W^\perp)$ with the Hilbert-Schmidt norm

$$\|B\|_{HS} := \Big(\sum_{i=1}^m \|Be_i\|^2\Big)^{1/2}, \tag{2.18}$$

which is equivalent to the spectral norm since

$$\|B\|_S \le \|B\|_{HS} \le \sqrt{m}\|B\|_S, \quad B \in \mathcal{L}(W, W^\perp). \tag{2.19}$$

In particular $F$ is continuous with respect to the Hilbertian norm

$$\|(c, B)\|_H := \Big(\|c\|^2 + \sum_{i=1}^m \|Be_i\|^2\Big)^{1/2}. \tag{2.20}$$

The function $F$ may not be infinite at infinity: this happens if there exists a non-trivial pair $(c, B)$ such that

$$c + BP_W u = 0, \quad u \in \mathcal{M}.$$

In order to fix this problem, we define the subspace

$$S_0 := \Big\{(c, B) \in W^\perp \times \mathcal{L}(W, W^\perp) \;:\; c + BP_W u = 0,\ u \in \mathcal{M}\Big\}. \tag{2.21}$$

and we denote by $S_1$ its orthogonal complement in $W^\perp \times \mathcal{L}(W, W^\perp)$ for the inner product associated to the above Hilbertian norm $\|\cdot\|_H$. The function $F$ is constant in the direction of $S_0$ and therefore we are left to prove the existence of the minimum of $F$ on $S_1$. For any $(c, B) \in S_1$, there exists $u \in \mathcal{M}$ such that $c + BP_W u \neq 0$. This implies that

$$\lim_{|t| \to +\infty} \|P_{W^\perp} u - tc - tBP_W u\| = +\infty, \tag{2.22}$$

and therefore that $\lim_{|t| \to +\infty} F_u(t(c, B)) = +\infty$. This shows that $F$ is infinite at infinity when restricted to $S_1$. Any convex and continuous function in a Hilbert space is weakly lower semi-continuous, and admits a minimum when it is infinite at infinity. We thus concludes in the existence of a minimizer $(c^*, B^*)$ of $F$ and therefore

$$A^*_{\mathrm{wca}}(w) = w + c^* + B^* w, \tag{2.23}$$

is an optimal affine recovery map. □

## 2.3 Comparison with the stochastic setting

In the stochastic setting, assuming that $u$ has finite second order moments, the optimal map that minimizes the mean square error (1.10) is given by the conditional expectation

$$A^*_{\mathrm{ms}}(w) = \mathbb{E}(u \mid P_W u = w), \tag{2.24}$$

that is, the expectation of posterior distribution $p_w$ of $u$ conditioned to the observation of $w$. Various sampling strategies have been developed in order to approximate the posterior and its expectation, see [11] for a survey. These approaches come at a significant computational cost they require a specific sampling for each instance $w$ of observed data. In the parametric PDE setting, each sample requires one solve of the forward problem.

On the other hand, it is well known that an optimal affine map $A^*_{\rm msa}$ for the mean square error can be explicitely derived from the first and second order statistics of $u$. We briefly recall this derivation by using an arbitrary orthonormal basis $(e_1, \dots, e_m)$ of $W$ that we complement into an orthonormal basis $(e_j)_{j\geq 1}$ of $V$. We write

$$u = \sum_{j\geq 1} w_j e_j \quad \text{and} \quad \overline{u} = \mathbb{E}(u) = \sum_{j\geq 1} \overline{w}_j e_j, \quad \overline{w}_j := \mathbb{E}(w_j), \tag{2.25}$$

as well as

$$\tilde{u} = u - \overline{u} = \sum_{j\geq 1} \tilde{w}_j e_j, \quad \tilde{w}_j := w_j - \overline{w}_j. \tag{2.26}$$

An affine recovery map of the form (2.13) leaves the coordinates $w_1, \dots, w_m$ unchanged and recovers for each $i \geq 1$

$$w^*_{m+i} = c_i + \sum_{j=1}^m b_{i,j} w_j, \tag{2.27}$$

which can be rewritten as

$$w^*_{m+i} = \overline{w}_{m+i} + d_i + \sum_{j=1}^m b_{i,j} \tilde{w}_j. \tag{2.28}$$

Since $E_{\rm ms}(A) = \sum_{i\geq 1} \mathbb{E}(|w^*_{m+i} - w_{m+i}|^2)$, the numbers $d_i$ and $b_{i,j}$ are found by separately minimizing each term. By Pythagoras theorem one has

$$\mathbb{E}(|w^*_{m+i} - w_{m+i}|^2) = |d_i|^2 + \mathbb{E}\Big(|\sum_{j=1}^m b_{i,j}\tilde{w}_j - \tilde{w}_{m+i}|^2\Big), \tag{2.29}$$

which shows that we should take $d_i = 0$. Minimizing the second term leads to the orthogonal projection equations

$$\sum_{j=1}^m b_{i,j} t_{j,l} = t_{m+j,l}, \quad l = 1, \dots, m. \tag{2.30}$$

which involve the entries of the covariance matrix

$$\mathbf{S} := (t_{i,j}), \quad t_{i,j} := \mathbb{E}(\tilde{w}_i \tilde{w}_j). \tag{2.31}$$

Therefore, with the block decomposition

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{1,1} & \mathbf{S}_{1,2} \\ \mathbf{S}_{2,1} & \mathbf{S}_{2,2} \end{pmatrix}, \tag{2.32}$$

corresponding to the splitting of rows and columns from $\{1, \dots, m\}$ and $\{m+1, m+2, \dots\}$, one obtains that the matrix $\mathbf{B} = (b_{i,j})$ that defines the optimal affine map satisfies $\mathbf{S}_{1,1}\mathbf{B}^{\rm T} = \mathbf{S}_{1,2}$ and therefore,

$$\mathbf{B} = \mathbf{S}_{2,1}\mathbf{S}_{1,1}^{-1} \tag{2.33}$$

where we have used the symmetry of $\mathbf{S}$. In other words,

$$A^*_{\rm msa}(w) = w + P_{W^\perp}\overline{u} + B\tilde{w}, \tag{2.34}$$

where the linear transform $B \in \mathcal{L}(W, W^\perp)$ is represented by the matrix $\mathbf{B}$ in the basis $(e_j)_{j\geq 1}$.

The optimal affine recovery map $A^*_{\rm msa}$ agrees with the optimal map $A^*_{\rm ms}$ in the particular case where $u$ has gaussian distribution, therefore entirely characterized by its average $\overline{u}$ and covariance matrix $\mathbf{S}$. To see this, assume for simplicity that $V$ is finite dimensional. The distribution of $\mathbf{u} = (w_j)_{j\geq 1}$ has density proportional to $\exp(-\frac{1}{2}\langle \mathbf{T}\tilde{\mathbf{u}}, \tilde{\mathbf{u}}\rangle)$ where $\mathbf{T} = \mathbf{S}^{-1}$. We expand the quadratic form into

$$\frac{1}{2}\langle \mathbf{T}\tilde{\mathbf{u}}, \tilde{\mathbf{u}}\rangle = \frac{1}{2}\langle \mathbf{T}_{1,1}\tilde{\mathbf{w}}, \tilde{\mathbf{w}}\rangle + \langle \mathbf{T}_{2,1}\tilde{\mathbf{w}}, \tilde{\mathbf{w}}_\perp\rangle + \frac{1}{2}\langle \mathbf{T}_{2,2}\tilde{\mathbf{w}}_\perp, \tilde{\mathbf{w}}_\perp\rangle, \tag{2.35}$$

where $\tilde{\mathbf{w}}_\perp = (\tilde{w}_{m+j})_{j\geq 1}$ and $\tilde{\mathbf{w}} = (\tilde{w}_j)_{j=1,\dots,m}$, and where

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{1,1} & \mathbf{T}_{1,2} \\ \mathbf{T}_{2,1} & \mathbf{T}_{2,2} \end{pmatrix}, \tag{2.36}$$

is a block decomposition similar to that of $\mathbf{S}$. The distribution of the vector $\tilde{\mathbf{w}}_\perp$ conditional to the observation of $\tilde{\mathbf{w}}$ is also gaussian and its expectation coincides with the minimum of the quadratic form

$$Q_{\mathbf{w}}(\tilde{\mathbf{w}}_\perp) = \frac{1}{2}\langle \mathbf{T}_{2,2}\tilde{\mathbf{w}}_\perp, \tilde{\mathbf{w}}_\perp\rangle + \langle \mathbf{T}_{2,1}\tilde{\mathbf{w}}, \tilde{\mathbf{w}}_\perp\rangle. \tag{2.37}$$

Therefore

$$\mathbb{E}(\tilde{\mathbf{w}}_\perp \,|\, \tilde{\mathbf{w}}) = -\mathbf{T}_{2,2}^{-1}\mathbf{T}_{2,1}\tilde{\mathbf{w}} = \mathbf{S}_{2,1}\mathbf{S}_{1,1}^{-1}\tilde{\mathbf{w}} = \mathbf{B}\tilde{\mathbf{w}}, \tag{2.38}$$

which shows that

$$A^*_{\rm ms}(w) = \mathbb{E}(u \,|\, P_W u = w) = A^*_{\rm msa}(w). \tag{2.39}$$

An analogy can be drawn with affine recovery in the deterministic setting: consider the particular case where $\mathcal{M}$ is an ellipsoid described by an equation of the form

$$\langle \mathbf{T}\tilde{\mathbf{u}}, \tilde{\mathbf{u}}\rangle \leq 1, \tag{2.40}$$

for a symmetric positive matrix $\mathbf{T}$. Then, the set $\mathcal{M}_w = \mathcal{M} \cap V_w$ is also an ellipsoid associated with the above quadratic form $Q_{\mathbf{w}}$. The coordinates of its center are therefore given by the same equation $\tilde{\mathbf{w}}_\perp = -\mathbf{T}_{2,2}^{-1}\mathbf{T}_{2,1}\tilde{\mathbf{w}}$ as the above conditional expectation. This shows that, in the particular case of an ellipsoid, the optimal map $A^*_{\rm wc}$ agrees with the optimal affine recovery map $A^*_{\rm wca}$ for the worst case error, and it also agrees with the above described optimal map $A^*_{\rm msa}$ for the mean square error.

# 3 Algorithms for optimal affine recovery

## 3.1 Discretization and truncation

We have seen that the optimal affine recovery map is obtained by minimizing the convex function

$$F(c, B) = \sup_{u\in\mathcal{M}} \|P_{W^\perp}u - c - BP_W u\|, \tag{3.1}$$

over $W^\perp \times \mathcal{L}(W, W^\perp)$. This optimization problem cannot be solved exactly for two reasons:

(i) The sets $W^\perp$ as well as $\mathcal{L}(W, W^\perp)$ are infinite dimensional when $V$ is infinite dimensional.

(ii) One single evaluation of $F(c, B)$ requires in principle to explore the entire manifold $\mathcal{M}$.

The first difficulty is solved by replacing $V$ by a finite dimensional space that approximates the solution manifold $\mathcal{M}$ with an accuracy of smaller order than that expected for the recovery error. One possibility is to work in a fine grid finite element space $V_h$, however its dimension $N_h$ needed to reach the accuracy could still be quite large. An alternative is to use reduced model spaces $V_N$ of dimension $N$ which are more efficient for the approximation of $\mathcal{M}$. We therefore minimize $F(c, B)$ over $\tilde{W}^\perp \times \mathcal{L}(W, \tilde{W}^\perp)$, where $\tilde{W}^\perp$ is the orthogonal complement of $W$ in the space $W + V_N$, and obtain an affine map $\tilde{A}_{\rm wca}$.

In order to compare its performance with that of $A^*_{\rm wca}$, we first observe that

$$\|P_{W^\perp}u - P_{\tilde{W}^\perp}u\| \leq \varepsilon_N := \sup_{u\in\mathcal{M}} \operatorname{dist}(u, V_N). \tag{3.2}$$

For any $(c, B) \in W^\perp \times \mathcal{L}(W, W^\perp)$, we define $(\tilde{c}, \tilde{B}) \in \tilde{W}^\perp \times \mathcal{L}(W, \tilde{W}^\perp)$ by $\tilde{c} = P_{\tilde{W}^\perp} c$ and $\tilde{B} = P_{\tilde{W}^\perp} B$. Then, for any $u \in \mathcal{M}$,

$$\begin{aligned} \|P_{W^\perp}u - \tilde{c} - \tilde{B}u\| &\leq \|P_{\tilde{W}^\perp}(P_{W^\perp}u - c - BP_W u)\| + \|P_{W^\perp}u - P_{\tilde{W}^\perp}u\| \\ &\leq \|P_{W^\perp}u - c - BP_W u\| + \varepsilon_N. \end{aligned}$$

It follows that we have the framing

$$E(A^*_{\rm wca}, \mathcal{M}) \leq E(\tilde{A}_{\rm wca}, \mathcal{M}) \leq E(A^*_{\rm wca}, \mathcal{M}) + \varepsilon_N, \tag{3.3}$$

which shows that the loss in the recovery error is at most of the order $\varepsilon_N$.

To understand how large $N$ should be, let us observe that a recovery map $A$ of the form (2.13) takes it value in the linear space

$$F_{m+1} = \mathbb{R}c + \operatorname{ran}(B), \tag{3.4}$$

which has dimension $m+1$. It follows that the recovery error is always larger than the approximation error by such a space. Therefore

$$E_{wc}(A^*_{\rm wca}, \mathcal{M}) \geq \delta_{m+1}(\mathcal{M}), \tag{3.5}$$

where $\delta_{m+1}(\mathcal{M})$ is the Kolmogorov $n$-width defined by (1.23) for $n = m + 1$. Therefore, if we could use the reduced model spaces $V_n := E_n$ that exactly achieve the infimum in (1.23), we would be ensured that, with $N = m + 1$, the additional error $\varepsilon_N = \delta_{m+1}(\mathcal{M})$ in (3.3) is of smaller order than $E_{wc}(A^*_{\rm wca}, \mathcal{M})$. In practice, since we do not have access to the $n$-width spaces, we choose instead the reduced basis spaces which are expected to have comparable approximation performances in view of the results from [1, 10]. We take $N$ larger than $m$ but of comparable order.

The second difficulty is solved by replacing the set $\mathcal{M}$ in the supremum that defines $F(c, B)$ by a discrete training set $\tilde{\mathcal{M}}$, which corresponds to a discretization $\tilde{Y}$ of the parameter domain $Y$, that is

$$\tilde{\mathcal{M}} := \{u(y) \,:\, y \in \tilde{Y}\}, \tag{3.6}$$

with finite cardinality.

We therefore minimize over $\tilde{W}^\perp \times \mathcal{L}(W, \tilde{W}^\perp)$ the function

$$\tilde{F}(c, B) = \sup_{u\in\tilde{\mathcal{M}}} \|P_{W^\perp}u - c - BP_W u\|, \tag{3.7}$$

which is computable. The additional error resulting from this discretization can be controlled from the resolution of the discretization. Let $\varepsilon > 0$ be the smallest value such that $\tilde{\mathcal{M}}$ is an $\varepsilon$-approximation net of $\mathcal{M}$, that is, $\mathcal{M}$ is covered by the $V$-balls $B(u, \varepsilon)$ for $u \in \tilde{\mathcal{M}}$. Then, we find that

$$\tilde{F}(c, B) \leq F(c, B) \leq \tilde{F}(c, B) + \varepsilon \|B\|_{\mathcal{L}(W, \tilde{W}^\perp)}, \tag{3.8}$$

which shows that the additional recovery error will be of the order of $\varepsilon$ amplified by the norm of the linear part of the optimal recovery map.

One difficulty is that the cardinality of $\varepsilon$-approximation nets become potentially untractable for small $\varepsilon$ as the parameter dimension becomes large, due to the curse of dimensionality. This difficulty also occurs when constructing reduced basis by a greedy selection process which also needs to be performed in a sufficiently dense discretized sets. Recent results obtained in [7] show that in certain relevant instances $\varepsilon$ approximation nets can be replaced by random training sets of smaller cardinality. One direction under investigation is to apply similar ideas in the context of the present paper.

### 3.2 Optimization algorithms

As already brought up in the previous section, the practical computation of $\tilde{A}_{\mathrm{wc}}$ consists in solving

$$\min_{(c,B) \in \tilde{W}^\perp \times \mathcal{L}(W, \tilde{W}^\perp)} \sup_{u \in \tilde{\mathcal{M}}} \|P_{W^\perp} u - c - B P_W u\|^2, \tag{3.9}$$

The numerical solution of this problem is challenging due to its lack of smoothness (the objective function is convex but non differentiable) and its high dimensionality (for a given target accuracy $\varepsilon_N$, the cardinality of $\tilde{\mathcal{M}}$ might be large). One could use classical subgradient methods, which are simple to implement. However these schemes only guarantee a very slow $O(k^{-1/2})$ convergence rate of the objective function, where $k$ is the number of iterations. This approach did not give satisfactory results in our case: due to the slow convergence, the solution update of one iteration falls below machine precision before approaching the minimum close enough, see Figure 3.1. This has motivated the use of a primal-dual splitting method which is known to ensure a $O(1/k)$ convergence rate on the partial duality gap. We next describe this method but only briefly as a detailed analysis would make us deviate too far from the main topic of this paper. A complete analysis with further examples of application will be presented in a forthcoming work [12].

We assume without loss of generality that $\dim(W + V_N) = m + N$ and that $\dim \tilde{W}^\perp = N$. Let $\{\psi_i\}_{i=1}^{m+N}$ be an orthonormal basis of $W + V_N$ such that $W = \mathrm{span}\{\psi_1, \dots, \psi_m\}$. Since for any $u \in V$,

$$P_{W+V_N} u = \sum_{i=1}^{m+N} u_i \psi_i,$$

the components of $u$ in $W$ can be given in terms of the vector $\mathbf{w} = (u_i)_{i=1}^m$ and the ones in $\tilde{W}^\perp$ with $\mathbf{u} = (u_{i+m})_{i=1}^N$.

We now consider the finite training set

$$\tilde{\mathcal{M}} := \{u^1, \dots, u^J\}, \quad J := \#(\tilde{\mathcal{M}}) < \infty, \tag{3.10}$$

and denote by $\mathbf{w}^j$ and $\mathbf{u}^j$ the vectors associated to the snapshot functions $u^j$ for $j = 1, \dots, J$. One may express the problem (3.9) as the search for

$$\min_{\substack{(\mathbf{R},\mathbf{b})\in \\ \mathbb{R}^{N\times m}\times\mathbb{R}^N}} \max_{j=1,\dots,J} \|\mathbf{u}^j - \mathbf{R}\mathbf{w}^j - \mathbf{b}\|_2^2. \tag{3.11}$$

Concatenating the matrix and vector variables $(\mathbf{R}, \mathbf{b})$ into a single $\mathbf{x} \in \mathbb{R}^{m(N+1)}$, we rewrite the above problem as

$$\min_{\mathbf{x}\in\mathbb{R}^{m(N+1)}} \max_{j=1,\dots,J} f_j(\mathbf{Q}_i\mathbf{x}), \tag{3.12}$$

where $\mathbf{Q}_i \in \mathbb{R}^{N\times m(N+1)}$ is a sparse matrix built using the coefficients of $\mathbf{w}^j$ and $f_j(\mathbf{y}) := \|\mathbf{u}^j - \mathbf{y}\|_2^2$.

The key observation to build our algorithm is that problem (3.12) can be equivalently written as a minimization problem on the epigraphs, i.e.,

$$\begin{aligned} &\min_{(\mathbf{x},t)\in\mathbb{R}^{m(N+1)}\times\mathbb{R}^+} t \quad \text{subject to} \quad f_j(\mathbf{Q}_j\mathbf{x}) \le t, \quad j = 1, \dots, J \\ \iff &\min_{(\mathbf{x},t)\in\mathbb{R}^{m(N+1)}\times\mathbb{R}^+} t \quad \text{subject to} \quad (\mathbf{Q}_j\mathbf{x}, t) \in \mathrm{epi}_{f_j}, \quad j = 1, \dots, J, \end{aligned} \tag{3.13}$$

or, in a more compact (and implicit) form,

$$\min_{(\mathbf{x},t)\in\mathbb{R}^{m(N+1)}\times\mathbb{R}^+} t + \sum_{j=1}^{J} \iota_{\mathrm{epi}_{f_j}}(\mathbf{Q}_j\mathbf{x}, t)\,. \tag{P$_{\text{epi}}$}$$

where, for any non-empty set $S$ the indicator function $\iota_S$ has value 0 on $S$ and $+\infty$ on $S^c$.

This problem takes the following canonical expression, which is amenable to a primal-dual proximal splitting algorithm

$$\min_{(\mathbf{x},t)\in\mathbb{R}^{m(N+1)}\times\mathbb{R}} G(\mathbf{x}, t) + F \circ L(\mathbf{x}, t). \tag{3.14}$$

Here, $G$ is the projection map for the second variable

$$G(\mathbf{x}, t) = t, \tag{3.15}$$

the linear operator $L$ is defined by

$$L(\mathbf{x}, t) := ((\mathbf{Q}_1\mathbf{x}, t), (\mathbf{Q}_2\mathbf{x}, t), \cdots, (\mathbf{Q}_J\mathbf{x}, t)) \tag{3.16}$$

and acts from $\mathbb{R}^{m(N+1)} \times \mathbb{R}$ to $\times_{j=1}^{J}(\mathbb{R}^N \times \mathbb{R})$ and the function $F$ acting from $\times_{j=1}^{J}(\mathbb{R}^N \times \mathbb{R})$ to $\mathbb{R}$ is defined by

$$F\Big((\mathbf{v}_1, t_1), \cdots, (\mathbf{v}_J, t_J)\Big) := \sum_{j=1}^{J} \iota_{\mathrm{epi}_{f_j}}(\mathbf{v}_j, t_j)\,. \tag{3.17}$$

Note that $F$ is the indicator function of the cartesian product of epigraphs.

Before introducing the primal-dual algorithm, some remarks are in order:

(i) We recall that if $\phi$ is a proper closed convex function on $\mathbb{R}^d$, its proximal mapping $\mathrm{prox}_\phi$ is defined by

$$\mathrm{prox}_\phi(y) = \mathrm{argmin}_{\mathbb{R}^d}\Big(\phi(x) + \frac{1}{2}\|x - y\|_2^2\Big). \tag{3.18}$$

(ii) The adjoint operator $L^*$ is given by

$$L^*\Big((\mathbf{v}_1, t_1), \cdots, (\mathbf{v}_J, t_J)\Big) := \left(\sum_{i=1}^{J} \mathbf{Q}_j^T \mathbf{v}_j, \sum_{j=1}^{J} t_j\right). \tag{3.19}$$

It can be easily shown that the operator norm of $L$ verifies $\|L\|^2 \leq J + \sum_{j=1}^{J} \|\mathbf{Q}_j\|^2$.

(iii) Both $G$ and $F$ are simple functions in the sense that their proximal mappings, $\mathrm{prox}_G$ and $\mathrm{prox}_F$, can be computed in closed form. See [12] for details.

The iterations of our primal-dual splitting method read for $k \geq 0$,

$$\begin{aligned} (\mathbf{x}, t)^{k+1} &= \mathrm{prox}_{\gamma_G G}\Big((\mathbf{x}, t)^k - \gamma_G L^*\Big(\Big((\mathbf{v}_1, \xi_1), \ldots, (\mathbf{v}_J, \xi_J)\Big)^k\Big)\Big), \\ (\bar{\mathbf{x}}, \bar{t})^{k+1} &= (\mathbf{x}, t)^{k+1} + \theta\Big((\mathbf{x}, t)^{k+1} - (\mathbf{x}, t)^k\Big), \\ \Big((\mathbf{v}_1, \xi_1), \ldots, (\mathbf{v}_J, \xi_J)\Big)^{k+1} &= \mathrm{prox}_{\gamma_F \hat{F}}\Big(\Big((\mathbf{v}_1, \xi_1), \ldots, (\mathbf{v}_J, \xi_J)\Big)^k + \gamma_F L(\bar{\mathbf{x}}, \bar{t})^{k+1}\Big), \end{aligned} \tag{3.20}$$

where $\hat{F}$ is the Fenchel-Legendre transform of $F$, $\gamma_G > 0$ and $\gamma_F > 0$ are such that $\gamma_G \gamma_F < 1/\|L\|^2$, and $\theta \in [-1, +\infty[$ (it is generally set to $\theta = 1$ as in [6]).

To illustrate the relevance of this algorithm for our purposes, we compare its performance with a standard subgradient method. Figure 3.1 plots the convergence history of the objective function across the iterations of both optimization methods in the example described in the next section ($m = 40$, $N = 110$ and $J = 10^3$). Two different reconstruction maps have been considered as starting guesses: the minimal $V$-norm recovery map given by $A(w) = w = P_W u$, and the one space algorithm $A_{n^*}$ based on reduced basis spaces $V_n$ with an optimal choice $n^*$ for $n$. The convergence plot shows the superiority of the primal-dual method which has converged to the same minimal value of the objective function after $10^5$ regardless of the intialization, while the subgradient method fails to reach it since its increments fall below machine precision.

For the same numerical example described next, we vary $m$ and consider $m = 10, 20, 30, 40, 50$. Figure 3.2 gives the convergence of the reconstruction error over the training set $\widetilde{\mathcal{M}}$ across the primal-dual iterations (for simplicity, we took $P_{W_m}$ as the starting guess for $A_{\mathrm{wca}}^{(m)}$). To make sure that we reach convergence, we perform $10^6$ iterations for each case. As expected, we observe in this figure that the final value of the objective function decreases as we increase the value of $m$ (the reconstruction error decreases as we increase the number of measurements).

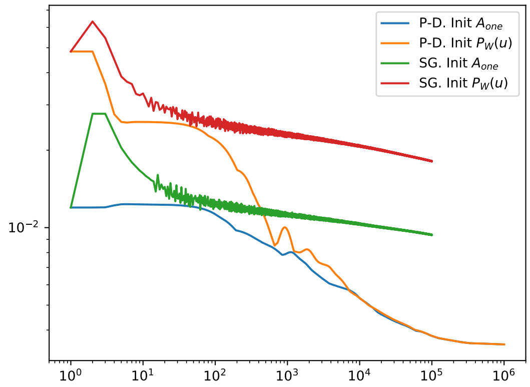


Figure 3.1: Convergence of the objective function for two different optimization algorithms and starting guesses. P.D. = Primal-Dual splitting. S.G.=Subgradient. Here, $m = 40$.

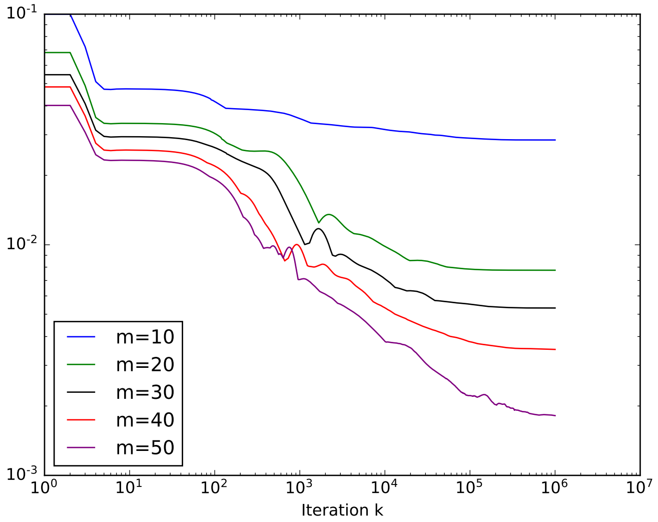


Figure 3.2: Convergence of the objective function in the primal-dual iterations for $m = 10, 20, 30, 40, 50$.

### 3.3 Numerical tests

We present some numerical experiments, aiming primarily at comparing in terms of the maximum reconstruction error the three above discussed recovery maps: the one space affine map $A_n$, the best affine map $A^*_{\rm msa}$ for the mean-square error, and the best affine map $A^*_{\rm wca}$. for the worst case error. In addition, we also consider the mimimum $V$ norm reconstruction map $A(w) = w = P_W u$. The results highlight the superiority of the best affine algorithm, which comes at the cost of a computationally intensive training phase as previously described.

We consider the elliptic problem

$$\begin{cases} -\mathrm{div}\Big(a(y)\nabla u\Big) & = f, \quad x \in D \\ u(x) & = 0, \quad x \in \partial D \end{cases} \tag{3.21}$$

on the unit square $D = ]0,1[^2$, with a certain parameter dependence in the field $a$. More precisely, for a given $p \geq 1$, we consider "checkerboard" random fields where $a(y)$ is piecewise constant on a $p \times p$ subdivision of the unit-square.

$$D = \bigcup_{i,j=0}^{p-1} S_{i,j},$$

with

$$S_{i,j} := \Big[\frac{i}{p}, \frac{i+1}{p}\Big[ \times \Big[\frac{j}{p}, \frac{j+1}{p}\Big[, \qquad i,j \in 0, \dots, p-1.$$

The random field is defined as

$$a(y) = 1 + \frac{1}{2} \sum_{i,j=0}^{p-1} \chi_{S_{i,j}} y_{i,j}, \tag{3.22}$$

where $\chi_S$ denotes the characteristic function of a set $S$, and the $y_{i,j}$ are random coefficients that are independent, each with identical uniform distribution on $[-1,1]$. Thus, our vector of parameters is

$$\mathbf{y} = (y_{i,j})_{i,j=0}^{p-1} \in \mathbb{R}^{p \times p}.$$

In our numerical tests, we take $p = 4$, that is 16 parameters, and work in the ambient space $V = H_0^1(D)$. All the sets of snapshots used for training and validating the reconstruction algorithms have been computed by first generating a certain number $J$ of random parameters $\mathbf{y}^1, \dots, \mathbf{y}^J$, with each $\mathbf{y}^i \in [-1,1]^{p \times p}$, and then solving the variational form of (3.21) in $V = H_0^1(D)$ using $\mathbb{P}_1$ finite elements on a regular grid of mesh size $h = 2^{-7}$. This gives the corresponding solutions $u_h^i = u_h(\mathbf{y}^i)$ that are used in the computations. To ease the reading, in the following we drop the dependence on $h$ in the notation.

The sensor measurements are modelled with linear functionals that are local averages of the form

$$\ell_{\mathbf{x},\tau}(u) = \int_D u(\mathbf{r}) \varphi_\tau(\mathbf{r} - \mathbf{x}) \, \mathrm{d}\mathbf{r}, \tag{3.23}$$

where

$$\varphi_\tau(\mathbf{r}) \propto \exp(-||\mathbf{r}||/2\tau^2) \tag{3.24}$$

is a radial function such that $\int \varphi_\tau = 1$. The parameter $\tau > 0$ represents the spread around the center $\mathbf{x}$. For the observation space $W$ of our example, we randomly select $m = 50$ centers $\mathbf{x}_i \in [0.1, 0.9]^2$ and spreads $\tau_i \in [0.05, 0.1]$, and compute the Riesz representers $\omega_{\mathbf{x}_i,\tau}$ of $\ell_{\mathbf{x}_i,\tau}$ in $H_0^1(D)$. We then set

$$W := \{\omega_{\mathbf{x}_i,\tau}\}_{i=1}^M$$

which is a space of dimension $m = 50$. Figure 3.3 shows the $m$ centers $\mathbf{x}_i$. As an example, the figure also plots the function $\omega_{\mathbf{x}_i,\tau}$ for $i = 10$, which has center $\mathbf{x}_i = (0.23, 0.75)$ and spread $\tau_i = 0.06$.

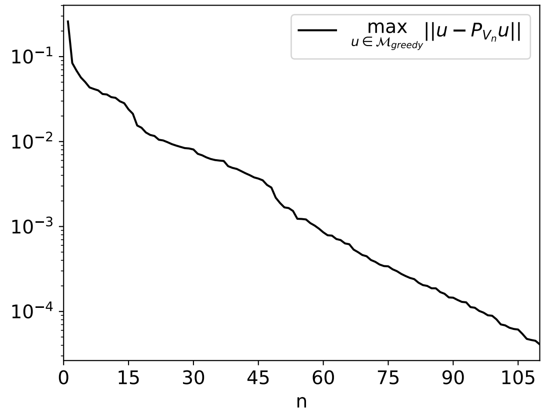


Figure 3.4: Greedy algorithm: decay of the error $e_n^{(\text{greedy})} = \max_{u \in \mathcal{M}_{\text{greedy}}} ||u - P_{V_n}u||$.

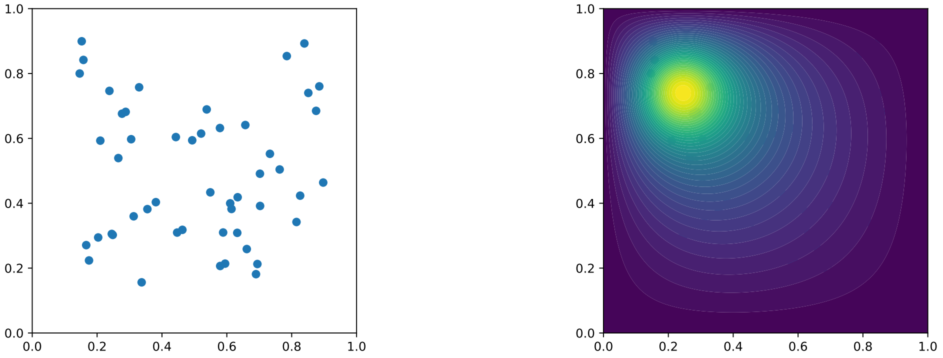


Figure 3.3: Sensor locations and the function $\omega_{\mathbf{x}_i,\tau_i}$ for $i = 10$ ($\mathbf{x}_i = (0.23, 0.75)$ and $\tau_i = 0.06$).

As explained in section 3.1, the first step to compute the best algorithm in practice consists in replacing $V = H_0^1(D)$ by a finite dimensional space that approximates the solution manifold $\mathcal{M}$ at an accuracy smaller than the one expected for the recovery error. Here, we replace $V$ by $W + V_N$ where $V_N$ is a reduced basis of dimension $N = 110$ that has been generated by running the classical greedy algorithm from [5] over a training set $\mathcal{M}_{\text{greedy}}$ of $10^3$ snapshots. We recall that it is defined for $n \geq 1$ as

$$u_n \in \underset{u \in \mathcal{M}_{\text{greedy}}}{\operatorname{argmax}} \|u - P_{V_{n-1}}u\|, \quad V_n \coloneqq V_{n-1} \oplus \mathbb{R}u_n = \operatorname{span}\{u_1, \dots, u_n\}, \tag{3.25}$$

with the convention $V_0 \coloneqq \{0\}$. Figure 3.4 gives the decay of the error

$$e_n^{(\text{greedy})} = \max_{u \in \mathcal{M}_{\text{greedy}}} \|u - P_{V_n}u\|$$

across the greedy iterations.

We next estimate the truncation accuracy $\varepsilon_N$ defined in (3.2). This has been done by computing the maximum of the error $\|u - P_{V_N}u\|$ over the training set $\mathcal{M}_{\text{greedy}}$ supplemented by a test set $\mathcal{M}_{\text{test}}$, also of $10^3$ snapshots. We obtain the estimate

$$\varepsilon_N \approx 5.10^{-5}.$$

In the comparison of the three different reconstruction algorithms, we want to illustrate the impact of the number of measurements that are used. To do this, we consider the nested subspaces

$$W_m = \text{span}\{\omega_{\mathbf{x}_i,\tau_i}\}_{i=1}^m \subset W$$

for $m = 10, 20, 30, 40, 50$ so that $W_{50} = W$.

For the computation of the best affine algorithm, we generate a new training set $\tilde{\mathcal{M}}$ of $10^3$ snapshots which we project into $W + V_N$. This projected set, which we denote by $P_{W+V_N}\tilde{\mathcal{M}}$ with a slight abuse of notation, is used to compute

$$\tilde{A}^{(m)}_{\text{wca}}(u) = \tilde{c}^{(m)} + \tilde{B}^{(m)}P_{W_m}u, \quad m = 10,\, 20, \dots,\, 50,$$

by running the primal-dual algorithm of section 3.2. We have added the indices $m$ to stress that the algorithm depends on it.

For the comparison with the three other reconstruction algorithms, we evaluate

$$e^{(m)}_{\text{wca}} = \max_{u\in\mathcal{M}_{\text{test}}} ||u - \tilde{A}^{(m)}_{\text{wca}}(P_{W_m}u)||, \quad m = 10,\, 20, \dots,\, 50.$$

We stress on the fact that the three sets $\mathcal{M}_{\text{greedy}}$, $\tilde{\mathcal{M}}$ and $\mathcal{M}_{\text{test}}$ are *different*. We compare this value with the performance of a straightforward reconstruction with the minimal $V$-norm recovery map,

$$e^{(m)}_{\text{mvn}} = \max_{u\in\mathcal{M}_{\text{test}}} ||u - P_{W_m}u||, \quad m = 10,\, 20, \dots,\, 50,$$

with the mean square approach,

$$e^{(m)}_{\text{msa}} = \max_{u\in\mathcal{M}_{\text{test}}} ||u - \tilde{A}^{(m)}_{\text{msa}}(P_{W_m}u)||, \quad m = 10,\, 20, \dots,\, 50,$$

and with the best one space affine algorithm,

$$e^{(m)}_{\text{one}} = \min_{1\le n\le m} e^{(m,n)}_{\text{one}},$$

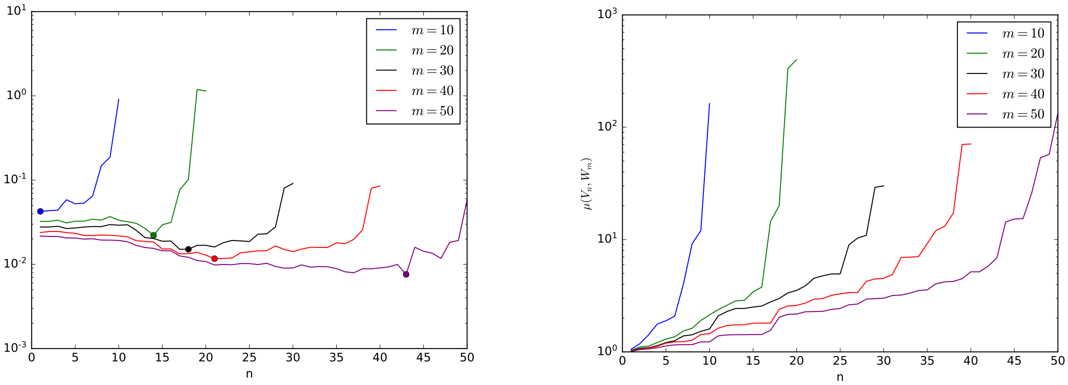
where

$$e^{(m,n)}_{\text{one}} = \max_{u\in\mathcal{M}_{\text{test}}} ||u - A^{(m)}_n(P_{W_m}u)||, \quad m = 10,\, 20, \dots,\, 50. \tag{3.26}$$

Some remarks on the computation of the one space algorithm are in order. First of all, we have used the average

$$\bar{u} := \frac{1}{\#\mathcal{M}_{\text{greedy}}} \sum_{u\in\mathcal{M}_{\text{greedy}}} u$$

as our offset. For $m \le M$ and $n \le m$ given, the one space affine algorithm $A^{(m)}_n$ is the one involving the spaces $W_m$ and $\tilde{V}_n = \bar{u} + V_n$, where $V_n = \text{span}\{u_1, \dots, u_n\}$. Its performance is given by $e^{(m,n)}_{\text{one}}$ in formula (3.26). Figure 3.5a shows $e^{(m,n)}_{\text{one}}$ as a function of $n$ and $m$. Note that, for a fixed $m$,



(a) $e_{\text{one}}^{(m,n)}$: reconstruction error with $A_n^{(m)}$.

(b) Stability factor $\mu(\tilde{V}_n, W_m)$.

Figure 3.5: One space algorithm.

the error $e_{\text{one}}^{(m,n)}$ reaches a minimal value $e_{\text{one}}^{m} = e_{\text{one}}^{(m,n^*)}$ for a certain dimension $n^* = n^*(m)$ of the reduced model, given by a thick dot in the figure. This behavior is due to the trade-off between the increase of the approximation properties of $\tilde{V}_n$ as $n$ grows and the degradation of the stability of the algorithm, given by the increase of $\mu(\tilde{V}_n, W_m)$ with $n$. For our comparison purpose, we use $A_{\text{one}}^{(m)} = A_{n^*(m)}^{(m)}$, that is, the best possible one space algorithm based on the reduced basis spaces.

Figure 3.6 shows the reconstruction errors $e_{\text{wca}}^{(m)}$, $e_{\text{mvn}}^{(m)}$, $e_{\text{msa}}^{(m)}$ and $e_{\text{one}}^{(m)}$ of the four different approaches for $m = 10, 20, \ldots, 50$. We also append a table with the values. We observe that a straightforward reconstruction with the minimal $V$-norm algorithm performs poorly in terms of approximation error and its quality improves only very mildly as we increase the number $m$ of measurements. This justifies considering our three other, more sophisticated, reconstruction algorithms. In this respect, the results confirm first of all that $\tilde{A}_{\text{wca}}^{(m)}$ is the best reconstruction algorithm. The mean square approach appears to be slightly superior to the one space algorithm but still worse than the best affine algorithm. Note that the accuracy improvement between the best affine algorithm and the one space and mean square algorithms is of about a half order of magnitude for each $m$.

Last but not least, we give some illustrations on the reconstruction algorithms applied to a particular snapshot function $u$ from the test set $\mathcal{M}_{\text{test}}$. The target function is given in Figure 3.7 and Figures 3.8 and 3.9 show the resulting reconstructions of $u$ from $P_{W_m}u$ with our four different algorithms and for $m = 20$ and 40. Visually, the reconstructed functions look very similar. However, the difference in quality can be better appreciated in the plots of the spatial errors $|u(\mathbf{x}) - A^{(m)}(u)(\mathbf{x})|$ as well as in the derivatives and their corresponding spatial errors.

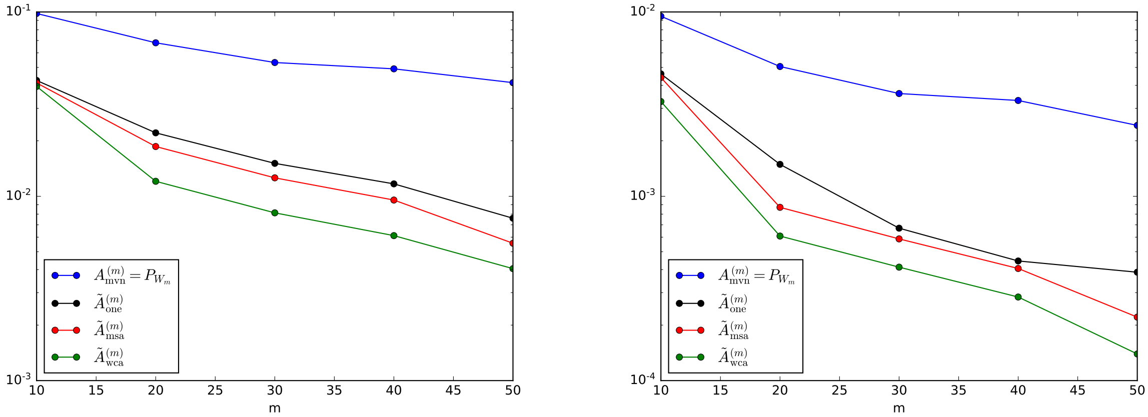


Figure 3.6: Comparison of the reconstruction errors (left: $H^1_0(D)$ norm; right: $L^2(D)$ norm).

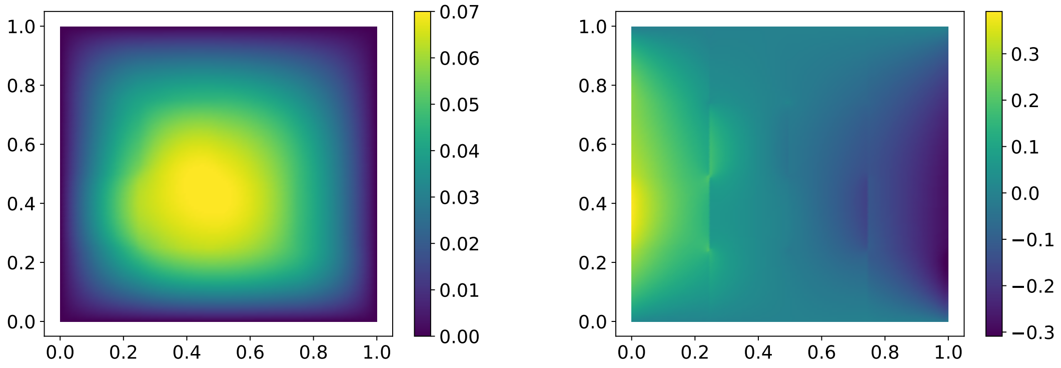
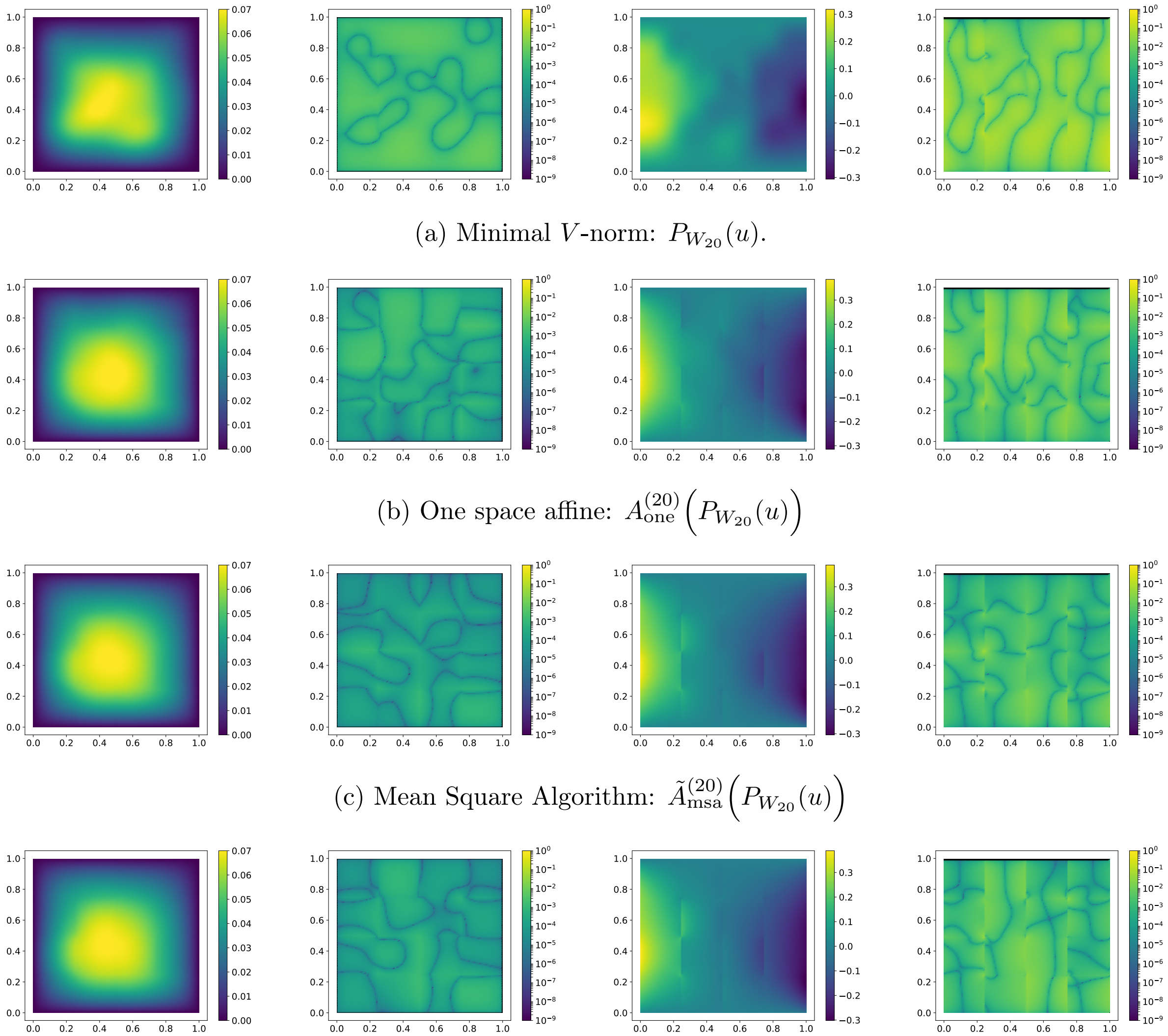


Figure 3.7: Function $u$ (left) and $\partial u/\partial x$ (right). The reconstruction of this function is given in Figures 3.8 and 3.9.



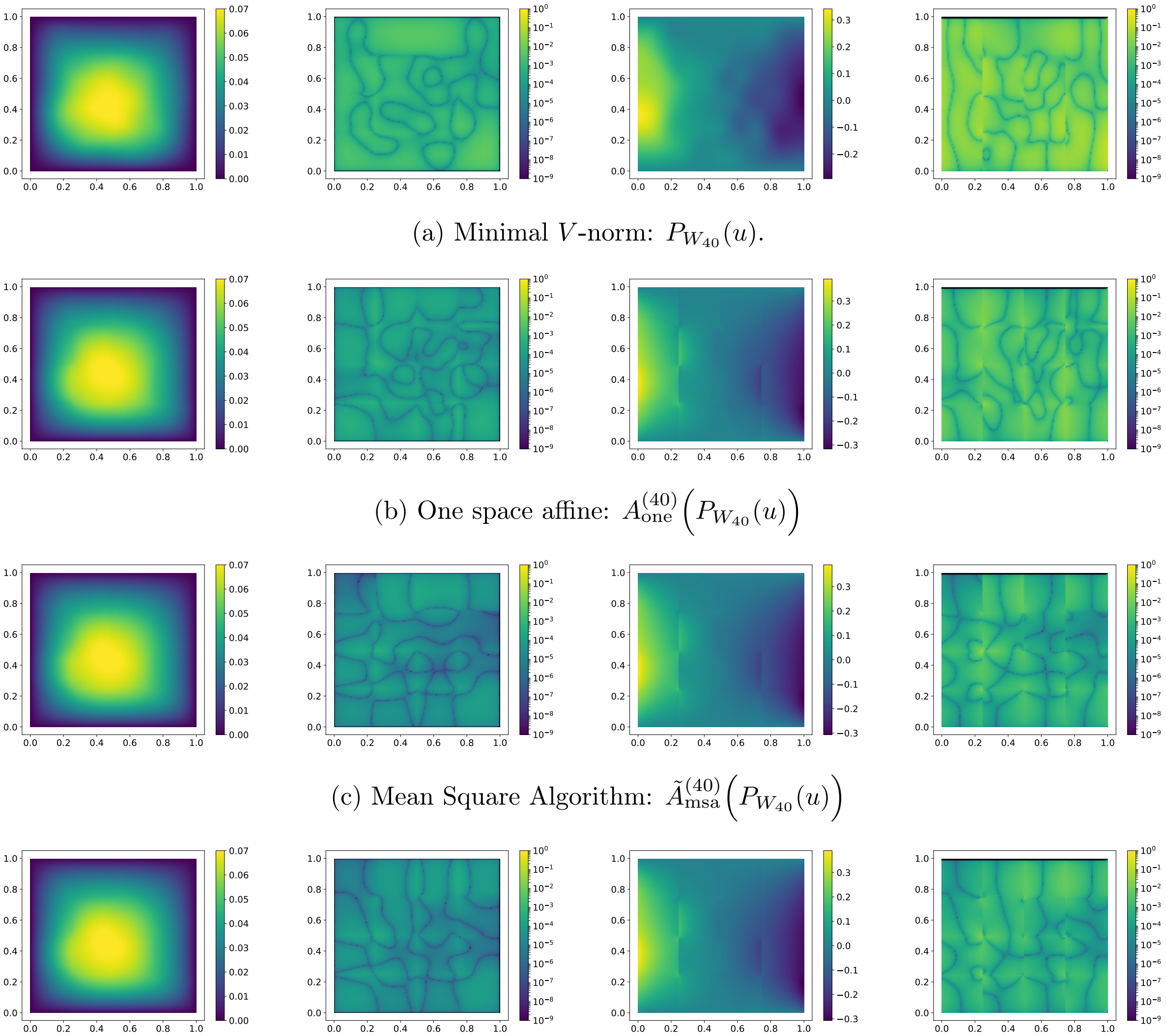
(a) Minimal $V$-norm: $P_{W_{20}}(u)$.

(b) One space affine: $A^{(20)}_{\text{one}}\left(P_{W_{20}}(u)\right)$

(c) Mean Square Algorithm: $\tilde{A}^{(20)}_{\text{msa}}\left(P_{W_{20}}(u)\right)$

(d) Best affine: $\tilde{A}^{(20)}_{\text{wca}}\left(P_{W_{20}}(u)\right)$

Figure 3.8: Reconstruction of the given function $u$ ($m = 20$). For each reconstruction strategy: (i) the two first figures are $A^{(m)}(u)(\mathbf{x})$ and the spatial errors $|u(\mathbf{x}) - A^{(m)}(u)(\mathbf{x})|$, (ii) the two last figures are $\frac{\partial A^{(m)}(u)}{\partial x}(\mathbf{x})$ and the spatial errors $|\frac{\partial u}{\partial x}(\mathbf{x}) - \frac{\partial A^{(m)}(u)}{\partial x}(\mathbf{x})|$.



(a) Minimal $V$-norm: $P_{W_{40}}(u)$.

(b) One space affine: $A_{\text{one}}^{(40)}\Big(P_{W_{40}}(u)\Big)$

(c) Mean Square Algorithm: $\tilde{A}_{\text{msa}}^{(40)}\Big(P_{W_{40}}(u)\Big)$

(d) Best affine: $\tilde{A}_{\text{wca}}^{(40)}\Big(P_{W_{40}}(u)\Big)$

Figure 3.9: Reconstruction of the given function $u$ ($m = 40$). For each reconstruction strategy: (i) the two first figures are $A^{(m)}(u)(\mathbf{x})$ and the spatial errors $|u(\mathbf{x}) - A^{(m)}(u)(\mathbf{x})|$, (ii) the two last figures are $\frac{\partial A^{(m)}(u)}{\partial x}(\mathbf{x})$ and the spatial errors $|\frac{\partial u}{\partial x}(\mathbf{x}) - \frac{\partial A^{(m)}(u)}{\partial x}(\mathbf{x})|$.

# References

[1] 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.

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

[3] P. Binev, A. Cohen, O. Mula and J. Nichols, *Greedy algorithms for optimal measurements selection in state estimation using reduced models*, SIAM Journal on Uncertainty Quantification **43**, 1101-1126, 2018.

[4] B. Bojanov, *Optimal recovery of functions and integrals.* First European Congress of Mathematics, Vol. I (Paris, 1992), 371-390, Progr. Math., 119, Birkhauser, Basel, 1994.

[5] A. Buffa, Y. Maday, A.T. Patera, C. Prud'homme, and G. Turinici, *A Priori convergence of the greedy algorithm for the parameterized reduced basis*, Mathematical Modeling and Numerical Analysis **46**, 595-603, 2012.

[6] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, **40**, 120-145, 2011.

[7] A. Cohen, W. Dahmen and R. DeVore, *Reduced basis greedy selection using random training sets*, submitted, 2018.

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

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

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

[11] M. Dashti and A.M. Stuart, *The Bayesian Approach to Inverse Problems*, Handbook of Uncertainty Quantification, Editors R. Ghanem, D. Higdon and H. Owhadi, Springer, 2015.

[12] J. Fadili and O. Mula, *Primal-dual splitting for the max of convex functions*, in progress.

[13] 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., submitted, 2014.

[14] C.A. Micchelli, T.J. Rivlin, *Lectures on optimal recovery.* Numerical analysis, Lancaster 1984 (Lancaster, 1984), 21-93, Lecture Notes in Math., **1129**, Springer, Berlin, 1985.

[15] E. Novak and H. Wozniakowski, *Tractability of Multivariate Problems, Volume I: Linear Information*, EMS Tracts in Mathematics, Vol. 6 Eur. Math. Soc. Publ. House, Zurich 2008

[16] G. Rozza, D.B.P. Huynh, and A.T. Patera, *Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations Ñ application to transport and continuum mechanics*, Archive of Computational Methods in Engineering **15**, 229–275, 2008.

[17] S. Sen, *Reduced-basis approximation and a posteriori error estimation for many-parameter heat conduction problems*, Numerical Heat Transfer B-Fund **54**, 369–389, 2008.

[18] A. M. Stuart, *Inverse problems: A Bayesian perspective* Acta Numerica 19, 451-559, 2010.