Instances of Computational Optimal Recovery: Refined Approximability Models

Simon Foucart*— Texas A&M University

Abstract

Models based on approximation capabilities have recently been studied in the context of Optimal Recovery. These models, however, are not compatible with overparametrization, since model- and data-consistent functions could then be unbounded. This drawback motivates the introduction of refined approximability models featuring an added boundedness condition. Thus, two new models are proposed in this article: one where the boundedness applies to the target functions (first type) and one where the boundedness applies to the approximants (second type). For both types of model, optimal maps for the recovery of linear functionals are first described on an abstract level before their efficient constructions are addressed. By exploiting techniques from semidefinite programming, these constructions are explicitly carried out on a common example involving polynomial subspaces of $\mathcal{C}[-1,1]$.

Key words and phrases: Optimal recovery, approximability models, semidefinite programming, overparametrization.

AMS classification: 41A65, 46N10, 49M29, 65K05, 90C22, 90C47.

1 Introduction

The objective of this article is to uncover practical methods for the optimal recovery of functions available through observational data when the underlying models based on approximability allow for overparametrization. To clarify this objective and its various challenges, we start with some background on traditional Optimal Recovery. Typically, an unknown function f defined on a domain D is observed through point evaluations $y_1 = f(x_1), \ldots, y_m = f(x_m)$ at distinct points $x_1, \ldots, x_m \in D$. More generally, an unknown object f, simply considered as an element of a normed space X, is observed through

$$y_i = \ell_i(f), \qquad i \in [1:m], \tag{1}$$

*S. F. is partially supported by NSF grants DMS-1622134 and DMS-1664803, and also acknowledges the NSF grant CCF-1934904.

where $\ell_1, \dots, \ell_m$ are linear functionals defined on X. We assume here that these data are perfectly accurate — we refer to the companion article [5] for the incorporation of observation error. The data is summarized as $y = L(f)$, where the linear map $L : X \to \mathbb{R}^m$ is called observation operator. Based on the knowledge of $y \in \mathbb{R}^m$, the task is then to recover a quantity of interest $Q(f)$, where throughout this article $Q : X \to \mathbb{R}$ is assumed to be a linear functional. The recovery procedure can be viewed as a map R from $\mathbb{R}^m$ to $\mathbb{R}$, with no concern for its practicability at this point.

Besides the observational data (which is also called *a posteriori* information), there is some *a priori* information coming from an educated belief about the properties of realistic f's. It translates into the assumption that f belongs to a model set $\mathcal{K} \subseteq X$. The choice of this model set is of course critical. When the f's indeed represent functions, it is traditionally taken as the unit ball with respect to some norm that characterizes smoothness. More recently, motivated by parametric partial differential equations, a model based on approximation capabilities has been proposed in [2]. Namely, given a linear subspace V of X and a threshold $\varepsilon > 0$, it is defined as

$$\mathcal{K} = \mathcal{K}_{V,\varepsilon} := \{f \in X : \operatorname{dist}_X(f, V) \le \varepsilon\}. \tag{2}$$

This model set is also implicit in many numerical procedures and in machine learning.

Whatever the selected model set, the performance of the recovery procedure $R : \mathbb{R}^m \to \mathbb{R}$ is measured in a worst-case setting via the (global) error of R over $\mathcal{K}$, i.e.,

$$e_{\mathcal{K},Q}(L, R) := \sup_{f \in \mathcal{K}} |Q(f) - R(L(f))|. \tag{3}$$

Obviously, one is interested in optimal recovery maps $R^{\mathrm{opt}} : \mathbb{R}^m \to \mathbb{R}$ minimizing this worst-case error, i.e., such that

$$e_{\mathcal{K},Q}(L, R^{\mathrm{opt}}) = \inf_{R:\mathbb{R}^m \to \mathbb{R}} e_{\mathcal{K},Q}(L, R). \tag{4}$$

This infimum is called the intrinsic error of the observation map L (for Q over $\mathcal{K}$). It is known, at least since Smolyak's doctoral dissertation [12], that there is a linear functional among the optimal recovery maps as soon as the set $\mathcal{K}$ is symmetric and convex, see e.g. [10, Theorem 4.7] for a proof. The practicality of such a linear optimal recovery map is not automatic, though. For the approximability set (2), Theorem 3.1 of [4] revealed that such a linear optimal recovery map takes the form $R^{\mathrm{opt}} : y \in \mathbb{R}^m \mapsto \sum_{i=1}^m a_i^{\mathrm{opt}} y_i$, where $a^{\mathrm{opt}} \in \mathbb{R}^m$ is a solution to

$$\underset{a \in \mathbb{R}^m}{\text{minimize}} \left\| Q - \sum_{i=1}^m a_i \ell_i \right\|_{X^*} \qquad \text{subject to } \sum_{i=1}^m a_i \ell_i(v) = Q(v) \quad \text{for all } v \in V, \tag{5}$$

an optimization problem that can be solved for $X = \mathcal{C}(D)$ in exact form when the observation functionals are point evaluations (see [4]) and in approximate form when they are arbitrary linear functionals (see [5] or Subsection 3.2 below).

The approximability set (2), however, presents some important restrictions. Suppose indeed that there is some nonzero $v \in \ker(L) \cap V$. Then, for a given $f_0 \in \mathcal{K}$ observed through $y = L(f_0) \in \mathbb{R}^m$, any $f_t := f_0 + tv$, $t \in \mathbb{R}$, is both model-consistent (i.e., $f_t \in \mathcal{K}$) and data-consistent (i.e., $L(f_t) = y$), so that the local error at y of any recovery map $R : \mathbb{R}^m \to \mathbb{R}$ satisfies

$$e^{\mathrm{loc}}_{\mathcal{K},Q}(L, R(y)) := \sup_{\substack{f \in \mathcal{K} \\ L(f)=y}} |Q(f) - R(y)| \geq \sup_{t \in \mathbb{R}} |Q(f_t) - R(y)| = \sup_{t \in \mathbb{R}} |(Q(f_0) - R(y)) + tQ(v)|, \tag{6}$$

which is generically infinite. Thus, for the optimal recovery problem to make sense under the approximability model (2), one must assume that $\ker(L) \cap V = \{0\}$. By a dimension argument, this imposes

$$n := \dim(V) \leq m. \tag{7}$$

In other words, we must place ourselves in an underparametrized regime for which the number n of parameters describing the model does not exceed the number m of data. This contrasts with many current studies, especially in the field of Deep Learning, which emphasize the advantages of overparametrization. In order to incorporate overparametrization in the optimal recovery problem under consideration, we must then restrict the magnitude of model- and data-consistent elements. A glaring strategy consists in altering the approximability set (2). We do so in two different ways, namely by considering a bounded approximability set *of the first type*, i.e.,

$$\mathcal{K} = \mathcal{K}^{\mathrm{I}}_{V,\varepsilon,\kappa} := \big\{ f \in X : \ \mathrm{dist}_X(f, V) \leq \varepsilon \text{ and } \|f\|_X \leq \kappa \big\}, \tag{8}$$

and a bounded approximability set *of the second type*, i.e.,

$$\mathcal{K} = \mathcal{K}^{\mathrm{II}}_{V,\varepsilon,\kappa} := \big\{ f \in X : \ \exists v \in V \text{ with } \|f - v\|_X \leq \varepsilon \text{ and } \|v\|_X \leq \kappa \big\}. \tag{9}$$

We will start by analyzing the second type of bounded approximability sets in Section 2 by formally describing the optimal recovery maps before revealing on a familiar example how the associated minimization problem is tackled in practice. The main ingredient in essence belongs to the sum-of-squares techniques from semidefinite programming. Next, we will analyze the first type of bounded approximability sets in Section 3. We will even formally describe optimal recovery maps over more general model sets consisting of intersections of approximability sets. On the prior example, we will again reveal how the associated minimization problem is tackled in practice. This time, the main ingredient in essence belongs to the moment techniques from semidefinite programming. In view of this article's emphasis on computability issues, all of the theoretical constructions are illustrated in a reproducible MATLAB file downloadable from the author's webpage.

2 Bounded approximability set of the second type

We concentrate in this section on the bounded approximability set of the second type, i.e., on

$$\mathcal{K} = \big\{ f \in X : \ \exists v \in V \text{ with } \|f - v\|_X \leq \varepsilon \text{ and } \|v\|_X \leq \kappa \big\}. \tag{10}$$

We shall first describe optimal recovery maps before showing how they can be computed in practice.

2.1 Description of an optimal recovery map

The result below reveals how [4, Theorem 3.1] extends from the model set (2) to the model set (10).

Theorem 1. If $Q : X \to \mathbb{R}$ is a linear functional, then an optimal recovery map over the bounded approximability set (10) is the linear functional

$$R^{\mathrm{opt}} : y \in \mathbb{R}^m \mapsto \sum_{i=1}^m a_i^{\mathrm{opt}} y_i \in \mathbb{R}, \tag{11}$$

where the optimal weight $a^{\mathrm{opt}} \in \mathbb{R}^m$ are precomputed as a solution to

$$\underset{a \in \mathbb{R}^m}{\mathrm{minimize}} \left[\varepsilon \times \Big\| Q - \sum_{i=1}^m a_i \ell_i \Big\|_{X^*} + \kappa \times \max_{v \in B_V} \Big| Q(v) - \sum_{i=1}^m a_i \ell_i(v) \Big| \right]. \tag{12}$$

Proof. Since the model set (10) is symmetric and convex, there exists an optimal recovery map $R^{\mathrm{opt}} : \mathbb{R}^m \to \mathbb{R}$ which is linear, i.e., of the form $R^{\mathrm{opt}}(y) = \sum_{i=1}^m a_i^{\mathrm{opt}} y_i$. The vector $a^{\mathrm{opt}} \in \mathbb{R}^m$ minimizes in particular the worst-case error $e := \max\{|Q(f) - \sum_{i=1}^m a_i \ell_i(f)| : f \in \mathcal{K}\}$ among all $a \in \mathbb{R}^m$. Thus, it is sufficient to transform this worst-case error into the expression featured between square brackets in (12). This is done by writing

$$\begin{aligned} e &= \max_{f \in X} \Big\{ \Big| Q(f) - \sum_{i=1}^m a_i \ell_i(f) \Big| : \|f - v\|_X \le \varepsilon \text{ for some } v \in V \text{ with } \|v\|_X \le \kappa \Big\} \\ &= \max_{\substack{f \in X \\ v \in V}} \Big\{ \Big| Q(f) - \sum_{i=1}^m a_i \ell_i(f) \Big| : \|f - v\|_X \le \varepsilon,\ \|v\|_X \le \kappa \Big\} \\ &= \max_{\substack{h \in X \\ v \in V}} \Big\{ \Big| Q(h+v) - \sum_{i=1}^m a_i \ell_i(h+v) \Big| : \|h\|_X \le \varepsilon,\ \|v\|_X \le \kappa \Big\} \\ &= \max_{h \in X} \Big\{ \Big| Q(h) - \sum_{i=1}^m a_i \ell_i(h) \Big| : \|h\|_X \le \varepsilon \Big\} + \max_{v \in V} \Big\{ \Big| Q(v) - \sum_{i=1}^m a_i \ell_i(v) \Big| : \|v\|_X \le \kappa \Big\}. \end{aligned} \tag{13}$$

By homogeneity, the latter is readily seen to coincide with the required expression. □

Remark. The approximability set (2) where the condition $\|v\|_X \le \kappa$ is not imposed can be viewed as an instantiation of (10) with $\kappa = \infty$. In this instantiation, if $\max_{v \in B_V} \Big| Q(v) - \sum_{i=1}^m a_i \ell_i(v) \Big|$ was nonzero, then the objective function would be infinite. Therefore, the infimum will be attained with the constraint $\max_{v \in B_V} \Big| Q(v) - \sum_{i=1}^m a_i \ell_i(v) \Big| = 0$ in effect. This argument constitutes another way of deriving the form of the optimal recovery map over the original approximability set (2). Let us note in passing that, while the optimization program (5) was independent of $\varepsilon > 0$, adding the condition $\|v\|_X \le \kappa$ does create a dependence on $\varepsilon > 0$ in the optimization program (12), unless κ is proportional to ε.

Remark. In the presence of observation error $e \in \mathbb{R}^m$ in $y = L(f) + e$, modeled as in [5] by the bounded uncertainty set

$$\mathcal{E} = \{e \in \mathbb{R}^m : \|e\|_p \le \eta\}, \tag{14}$$

an optimal recovery map for a linear functional $Q : X \to \mathbb{R}$ over $\mathcal{K}$ and $\mathcal{E}$ simultaneously still consists of a linear functional $R^{\mathrm{opt}} : y \in \mathbb{R}^m \mapsto \sum_{i=1}^m a_i^{\mathrm{opt}} y_i$, but now the optimal weights $a^{\mathrm{opt}} \in \mathbb{R}^m$ are solution to the optimization program

$$\underset{a \in \mathbb{R}^m}{\mathrm{minimize}} \left[\varepsilon \times \left\| Q - \sum_{i=1}^m a_i \ell_i \right\|_{X^*} + \kappa \times \max_{v \in B_V} \left| Q(v) - \sum_{i=1}^m a_i \ell_i(v) \right| + \eta \times \|a\|_{p'} \right], \tag{15}$$

where $p' = p/(p-1)$ is the conjugate exponent to $p \in [1, \infty]$. The argument, which follows the ideas presented in [5], is left to the reader. We do point out that the program (15) is solvable in practice as soon as soon as the program (12) itself is solvable in practice, for instance as in the forthcoming example.

2.2 Computational realization for $X = \mathcal{C}[-1, 1]$

For practical purposes, the result of Theorem 1 is close to useless if the minimization cannot be performed efficiently. We show below that in the important case $X = \mathcal{C}[-1, 1]$, choosing V as the space $\mathcal{P}_n$ of algebraic polynomials of degree $< n$ leads to an optimization problem which can be solved exactly via semidefinite programming. For that, we also assume that the observation functionals are distinct point evaluations and that the quantity of interest Q is another point evaluation or the normalized integral. These restrictions can be lifted if we trade exact solutions for quantifiably approximate solutions, see Subsection 3.2. In the statement below, the notation $\mathrm{Toep}(x)$ represents the symmetric Toeplitz matrix built from a vector $x \in \mathbb{R}^d$, i.e.,

$$\mathrm{Toep}(x) := \begin{bmatrix} x_1 & x_2 & \cdots & \cdots & x_d \\ x_2 & x_1 & x_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & x_2 & x_1 & x_2 \\ x_d & \cdots & \cdots & x_2 & x_1 \end{bmatrix}, \tag{16}$$

and the polynomials T_j, $j \in [0 : n-1]$, denote the jth Chebyshev polynomials of the first kind.

Theorem 2. Assuming that $V = \mathcal{P}_n \subseteq \mathcal{C}[-1, 1]$ and that $\ell_1, \ldots, \ell_m$ are point evaluations at distinct points $x_1, \ldots, x_m \in [-1, 1]$, an optimal recovery map over the bounded approximability set (10) for the quantity of interest $Q(f) = f(x_0)$, $x_0 \notin \{x_1, \ldots, x_m\}$, or $Q(f) = (1/2) \int_{-1}^1 f(x) dx$ is the linear functional

$$R^{\mathrm{opt}} : y \in \mathbb{R}^m \mapsto \sum_{i=1}^m a_i^{\mathrm{opt}} y_i \in \mathbb{R}, \tag{17}$$

where the optimal weights $a^{\mathrm{opt}} \in \mathbb{R}^m$ are precomputed as a solution to the semidefinite program

$$\underset{\substack{a,s\in\mathbb{R}^m\\ u\in\mathbb{R}^n}}{\text{minimize}} \left[\varepsilon \times \sum_{i=1}^m s_i + \kappa \times u_1\right] \quad \text{subject to } \operatorname{Toep}(u + Ca - b) \succeq 0,\ \operatorname{Toep}(u - Ca + b) \succeq 0, \tag{18}$$
$$\text{and } s + a \geq 0,\ s - a \geq 0.$$

Here, $b \in \mathbb{R}^n$ and $C \in \mathbb{R}^{n\times m}$ have entries $b_j = Q(T_j)$ and $C_{j,i} = \ell_i(T_j)$, $i \in [1:m]$, $j \in [0:n-1]$.

Proof. The work consists in recasting the objective function of (12) into manageable form. Under the assumptions on $\ell_1, \ldots, \ell_m$ and on Q, the first term is not a problem, by virtue of

$$\Big\|Q - \sum_{i=1}^m a_i \ell_i\Big\|_{\mathcal{C}[-1,1]^*} = 1 + \sum_{i=1}^m |a_i|. \tag{19}$$

We now turn to the second term, i.e., the one involving the maximum over the unit ball B_V of V. The idea, common in Robust Optimization [1], relies on duality to change the maximum into a minimum, which is then integrated into a larger minimization problem. This is possible essentially when B_V admits a linear or semidefinite description, which is the case for $V = \mathcal{P}_n$. Indeed, as already observed in [6, Subsection 5.3], following ideas formulated in [8], the unit ball of $\mathcal{P}_n$ admits the semidefinite description

$$B_{\mathcal{P}_n} = \Big\{ \sum_{j=0}^{n-1} \operatorname{tr}[D_j(P-M)]T_j \text{ for some positive semidefinite matrices } M, P \in \mathbb{R}^{n\times n} \tag{20}$$
$$\text{that satisfy } \operatorname{tr}[D_j(P+M)] = \delta_{0,j} \Big\},$$

where, for each $j \in [0:n-1]$, the symmetric matrix

$$D_j := \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 & 1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 1 & 0 & \ddots & \ddots & \ddots & 0 & 1 \\ 0 & 1 & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n\times n} \tag{21}$$

has 1's on the jth subdiagonal and superdiagonal and 0's elsewhere — in particular D_0 is the $n \times n$ identity matrix. Thus, for a fixed $a \in \mathbb{R}^m$, with $Q_a := Q - \sum_{i=1}^m a_i \ell_i$, the maximum over B_V reads

$$\max_{M,P\in\mathbb{R}^{n\times n}} \Big\{ \operatorname{tr}\Big[\Big(\sum_{j=0}^{n-1} Q_a(T_j) D_j\Big)(P-M)\Big] : M, P \succeq 0,\ \operatorname{tr}[D_j(P+M)] = \delta_{0,j} \Big\}. \tag{22}$$

Invoking duality in semidefinite programming (see e.g. [3, p.265-266]), the latter can be transformed into

$$\min_{u\in\mathbb{R}^n} u_1 \qquad \text{subject to } \sum_{j=0}^{n-1} u_j D_j \pm \sum_{j=0}^{n-1} Q_a(T_j) D_j \succeq 0. \tag{23}$$

Since $Q_a(T_j) = b_j - (Ca)_j$ for any $j \in [0:n-1]$, the constraint in (23) can be condensed to $\mathrm{Toep}(u \pm (Ca-b)) \succeq 0$. Then, combining the minimization over $u \in \mathbb{R}^n$ with the minimization over $a \in \mathbb{R}^m$, the optimization program (12) becomes equivalent to

$$\underset{\substack{a\in\mathbb{R}^m\\ u\in\mathbb{R}^n}}{\text{minimize}} \left[\varepsilon \times \sum_{i=1}^m |a_i| + \kappa \times u_1\right] \qquad \text{subject to } \mathrm{Toep}(u \pm (Ca-b)) \succeq 0. \tag{24}$$

The final step is the introduction of slack variables $s \in \mathbb{R}^m$ such that $|a_i| \le s_i$, i.e., $-s_i \le a_i \le s_i$, for all $i \in [1:m]$. □

3 Bounded approximability set of the first type

We concentrate in this section on the bounded approximability set of the first type, i.e., on

$$\mathcal{K} = \big\{f \in X : \ \mathrm{dist}_X(f,V) \le \varepsilon \text{ and } \|f\|_X \le \kappa\big\}. \tag{25}$$

Once again, we shall first describe optimal recovery maps before showing how they can be computed in practice.

3.1 Description of an optimal recovery map

The result below reveals how [4, Theorem 3.1] extends from the model set (2) to the model set (25).

Theorem 3. If $Q : X \to \mathbb{R}$ is a linear functional, then an optimal recovery map over the bounded approximability set (25) is the linear functional

$$R^{\mathrm{opt}} : y \in \mathbb{R}^m \mapsto \sum_{i=1}^m a_i^{\mathrm{opt}} y_i \in \mathbb{R}, \tag{26}$$

where the optimal weights $a^{\mathrm{opt}} \in \mathbb{R}^m$ are precomputed as a solution to

$$\underset{\substack{a\in\mathbb{R}^m\\ \mu,\nu\in X^*}}{\text{minimize}} \left[\varepsilon \times \|\mu\|_{X^*} + \kappa \times \|\nu\|_{X^*}\right] \qquad \text{subject to } \mu + \nu = Q - \sum_{i=1}^m a_i \ell_i \ \text{ and } \ \mu_{|V} = 0. \tag{27}$$

As a matter of fact, Theorem 3 is a corollary of Theorem 4 below. The setting of the more general result involves subspaces $V_1, \dots, V_K$ of a linear space X equipped with possibly distinct norms $\|\cdot\|_{(1)}, \dots, \|\cdot\|_{(K)}$. The model set is then defined, for some parameters $\varepsilon_1, \dots, \varepsilon_K > 0$, by

$$\mathcal{K} = \{f \in X : \operatorname{dist}_{\|\cdot\|_{(1)}}(f, V_1) \le \varepsilon_1, \dots, \operatorname{dist}_{\|\cdot\|_{(K)}}(f, V_K) \le \varepsilon_K\}. \tag{28}$$

It corresponds to what was called the multispace problem in [2, Section 3]. One works under the assumption that

$$\ker(L) \cap V_1 \cap \ldots \cap V_K = \{0\}. \tag{29}$$

This assumption holds for the bounded approximability set of the first type, obtained by taking $V_1 = V$, $V_2 = \{0\}$, and $\|\cdot\|_{(1)} = \|\cdot\|_{(2)} = \|\cdot\|_X$.

Theorem 4. If $Q : X \to \mathbb{R}$ is a linear functional, then an optimal recovery map over the model set (28) is the linear functional

$$R^{\mathrm{opt}} : y \in \mathbb{R}^m \mapsto \sum_{i=1}^m a_i^{\mathrm{opt}} y_i \in \mathbb{R}, \tag{30}$$

where the optimal weights $a^{\mathrm{opt}} \in \mathbb{R}^m$ are precomputed as a solution to

$$\underset{\substack{a \in \mathbb{R}^m \\ \lambda_1, \dots, \lambda_K \in X^*}}{\text{minimize}} \ \left[\varepsilon_1 \|\lambda_1\|_{(1)}^* + \cdots + \varepsilon_K \|\lambda_K\|_{(K)}^*\right] \quad \text{subject to } \lambda_1 + \cdots + \lambda_K = Q - \sum_{i=1}^m a_i \ell_i \text{ and } \lambda_{1|V_1} = 0, \dots, \lambda_{K|V_K} = 0. \tag{31}$$

Proof. We first notice that, replacing the norms $\|\cdot\|_{(k)}$ by $\|\cdot\|_{(k)}/\varepsilon_k$, we can assume that $\varepsilon_k = 1$. Next, since the model set $\mathcal{K}$ is symmetric and convex, there exists an optimal recovery map which is linear, i.e., of the form $y \in \mathbb{R}^m \mapsto \sum_{i=1}^m a_i y_i \in \mathbb{R}$. An optimal weight vector $a^{\mathrm{opt}} \in \mathbb{R}^m$ is then obtained as a solution to the optimization problem

$$\underset{a \in \mathbb{R}^m}{\text{minimize}} \max_{f \in X} \left\{ \left| Q(f) - \sum_{i=1}^m a_i \ell_i(f) \right| : \operatorname{dist}_{\|\cdot\|_{(k)}}(f, V_k) \le 1 \text{ for all } k \in [1:K] \right\}. \tag{32}$$

We claim that an optimal weight vector $a^{\mathrm{opt}} \in \mathbb{R}^m$ can also be obtained as a solution to the optimization problem

$$\underset{a \in \mathbb{R}^m}{\text{minimize}} \min_{\lambda_1, \dots, \lambda_K \in X^*} \left\{ \|\lambda_1\|_{(1)}^* + \cdots + \|\lambda_K\|_{(K)}^* : \lambda_1 + \cdots + \lambda_K = Q - \sum_{i=1}^m a_i \ell_i, \ \lambda_{k|V_k} = 0 \text{ for all } k \in [1:K] \right\}. \tag{33}$$

In other words, we shall prove in two steps that the minimal values of (32) and (33) coincide.

Firstly, we shall justify that the objective function in (32) is bounded by the objective function in (33) — a property which holds independently of $a \in \mathbb{R}^m$. To do so, let us consider $f \in X$ such that $\|f - v_1\|_{(1)} \le 1, \ldots, \|f - v_K\|_{(K)} \le 1$ for some $v_1 \in V_1$, …, $v_K \in V_K$. Let us also consider $\lambda_1, \ldots, \lambda_K \in X^*$ such that $\lambda_1 + \cdots + \lambda_K = Q - \sum_{i=1}^m a_i \ell_i$ and $\lambda_{1|V_1} = 0$, …, $\lambda_{K|V_K} = 0$. We have

$$\begin{aligned} \left| Q(f) - \sum_{i=1}^m a_i \ell_i(f) \right| &= |\lambda_1(f) + \cdots + \lambda_K(f)| = |\lambda_1(f - v_1) + \cdots + \lambda_K(f - v_K)| \\ &\le \|\lambda_1\|_{(1)}^* \|f - v_1\|_{(1)} + \cdots + \|\lambda_K\|_{(1)}^* \|f - v_K\|_{(K)} \\ &\le \|\lambda_1\|_{(1)}^* + \cdots + \|\lambda_K\|_{(K)}^*. \end{aligned} \tag{34}$$

Taking the infimum over $\lambda_1, \ldots, \lambda_K$ and the supremum over f yields the desired result.

Secondly, we shall justify that the minimal value of (33) is bounded by the minimal value of (32). To do so, let us consider the linear space $Z := X \times \cdots \times X$ equipped with the norm

$$\|(f_1, \ldots, f_K)\|_Z := \max_{k \in [1:K]} \|f_k\|_{(k)}. \tag{35}$$

Introducing the subspace U of Z given by

$$U := \{(h, \ldots, h), h \in \ker(L)\}, \tag{36}$$

the assumption (29) is equivalent to $U \cap (V_1 \times \cdots \times V_K) = \{0\}$. Thus, we can define a linear functional λ on $U \oplus (V_1 \times \cdots \times V_K)$ by

$$\lambda((h, \ldots, h)) = Q(h) \qquad \text{for } h \in \ker(L), \tag{37}$$

$$\lambda((v_1, \ldots, v_K)) = 0 \qquad \text{for } (v_1, \ldots, v_K) \in V_1 \times \cdots \times V_K. \tag{38}$$

Let then $\widetilde{\lambda} \in Z^*$ denote a Hahn–Banach extension of λ to the whole Z. With linear functionals $\lambda_1, \ldots, \lambda_K \in X^*$ defined for each $k \in [1:K]$ and $f \in X$ by $\lambda_k(f) = \widetilde{\lambda}((0, \ldots, 0, f, 0, \ldots, 0))$, where f appears at the kth position, we have $\lambda_1(f) + \cdots + \lambda_K(f) = \widetilde{\lambda}((f, \ldots, f))$ for all $f \in X$, hence in particular $Q - (\lambda_1 + \cdots + \lambda_K)$ vanishes on $\ker(L)$. This implies (see e.g. [11, Lemma 3.9]) that $Q - (\lambda_1 + \cdots + \lambda_K) = \sum_{i=1}^m a_i^\sharp \ell_i$ for some $a^\sharp \in \mathbb{R}^m$. In other words, the first constraint in (33) is satisfied by $a^\sharp$ and $\lambda_1, \ldots, \lambda_K$. The second constraint is also satisfied: indeed, for $v_k \in V_k$, $\lambda_k(v_k) = \widetilde{\lambda}((0, \ldots, 0, v_k, 0, \ldots, 0)) = 0$ since $(0, \ldots, 0, v_k, 0, \ldots, 0) \in V_1 \times \cdots \times V_K$. Therefore, the minimal value of (33) is bounded by

$$\begin{aligned} \|\lambda_1\|_{(1)}^* + \cdots + \|\lambda_K\|_{(K)}^* &= \max_{\|f_1\|_{(1)} \le 1} \lambda_1(f_1) + \cdots + \max_{\|f_K\|_{(K)} \le 1} \lambda_K(f_K) \\ &= \max_{\|f_1\|_{(1)} \le 1, \ldots, \|f_K\|_{(K)} \le 1} \lambda_1(f_1) + \cdots + \lambda_K(f_K) \\ &= \max_{\|(f_1, \ldots, f_K)\|_Z \le 1} \widetilde{\lambda}((f_1, \ldots, f_K)) \\ &= \|\widetilde{\lambda}\|_Z^*. \end{aligned} \tag{39}$$

The latter equals the norm of λ on $U \oplus (V_1 \times \cdots \times V_K)$, by virtue of $\widetilde{\lambda}$ being a Hahn–Banach extension of λ, so that

$$\begin{aligned} (40) \qquad \|\lambda_1\|_{(1)}^* + \cdots + \|\lambda_K\|_{(K)}^* &= \max_{\substack{u=(h,\dots,h)\in U \\ v=(v_1,\dots,v_K)\in V_1\times\cdots\times V_K}} \big\{\lambda(u-v) : \|u-v\|_Z \le 1\big\} \\ &= \max_{\substack{h\in\ker(L) \\ v_k\in V_k}} \big\{Q(h) : \|h-v_k\|_{(k)} \le 1 \text{ for all } k \in [1:K]\big\}. \end{aligned}$$

It follows that, for any $a \in \mathbb{R}^m$,

$$\begin{aligned} (41) \quad \|\lambda_1\|_{(1)}^* + \cdots + \|\lambda_K\|_{(K)}^* &= \max_{\substack{h\in\ker(L) \\ v_k\in V_k}} \left\{ Q(h) - \sum_{i=1}^m a_i\ell_i(h) : \|h-v_k\|_{(k)} \le 1 \text{ for all } k \in [1:K] \right\} \\ &\le \max_{f\in X} \left\{ \left| Q(f) - \sum_{i=1}^m a_i\ell_i(f) \right| : \operatorname{dist}_{\|\cdot\|_{(k)}}(f, V_k) \le 1 \text{ for all } k \in [1{:}K] \right\}. \end{aligned}$$

Taking the minimum over all $a \in \mathbb{R}^m$ shows that$\|\lambda_1\|_{(1)}^* + \cdots + \|\lambda_K\|_{(K)}^*$ is less than or equal to the minimal value of (32), and in turn that the same is true for the minimal value of (33). □

Remark. The approximability set (2) where the condition $\|f\|_X \le \kappa$ is not imposed can be viewed as an instantiation of (25) with $\kappa = \infty$. In this instantiation, if $\|\nu\|_{X^*}$ was nonzero, then the objective function in (27) would be infinite. Therefore, the minimum will be attained with the constraint $\|\nu\|_{X^*} = 0$ in effect, leading to $\mu = Q - \sum_{i=1}^m a_i\ell_i$ and in turn to the constraint $(Q - \sum_{i=1}^m a_i\ell_i)_{|V} = 0$. We do retrieve the minimization of (5), as expected. We note in passing that, while the optimization program (5) was independent of $\varepsilon > 0$, adding the condition $\|f\|_X \le \kappa$ does create a dependence on $\varepsilon > 0$ in the optimization problem (27), unless κ is proportional to ε.

Remark. In the presence of observation error $e \in \mathbb{R}^m$ in $y = L(f) + e$, again modeled as in [5] by the bounded uncertainty set

$$(42) \qquad \mathcal{E} = \{e \in \mathbb{R}^m : \|e\|_p \le \eta\},$$

an optimal recovery map for a linear functional $Q : X \to \mathbb{R}$ over $\mathcal{K}$ and $\mathcal{E}$ simultaneously still consists of a linear functional $R^{\mathrm{opt}} : y \in \mathbb{R}^m \mapsto \sum_{i=1}^m a_i^{\mathrm{opt}} y_i$, but now the optimal weights $a^{\mathrm{opt}} \in \mathbb{R}^m$ are solution to the optimization program

$$(43) \quad \underset{\substack{a\in\mathbb{R}^m \\ \mu,\nu\in X^*}}{\text{minimize}} \ \big[\varepsilon\times\|\mu\|_{X^*} + \kappa\times\|\nu\|_{X^*} + \eta\times\|a\|_{p'}\big] \qquad \text{subject to } \mu+\nu = Q - \sum_{i=1}^m a_i\ell_i \ \text{ and } \ \mu_{|V} = 0.$$

The argument follows the ideas presented in [5] and, although more subtle, is once again left to the reader. We do point out that the program (43) is solvable in practice as soon as soon as the program (27) itself is solvable in practice, for instance as in the forthcoming example.

3.2 Computational realization for $X = \mathcal{C}[-1,1]$

As before, the high-level results of Theorems 3 and 4 are of little practical use if the minimizations (27) and (31) cannot be performed efficiently. In the important situation $X = \mathcal{C}[-1,1]$, the dual functionals appearing as optimization variables are identified with measures. Despite involving infinite dimensional objects, minimizations over measures can be tackled via semidefinite programming, see e.g. [9]. Although such minimizations are in general not solved exactly, their accuracy can be quantifiably estimated in our specific case. For ease of presentation, we illustrate the approach by concentrating on the optimization program (27) rather than (31). We also assume that $V = \mathcal{P}_n$ and we write the observation functionals $\ell_1, \ldots, \ell_m$, as well as the quantity of interest Q, as

$$\ell_i(f) = \int_{-1}^1 f(x) d\lambda_i(x), \qquad Q(f) = \int_{-1}^1 f(x) d\rho(x), \qquad f \in \mathcal{C}[-1,1], \tag{44}$$

for some signed Borel measures $\lambda_1, \ldots, \lambda_m, \rho$ defined on $[-1,1]$. In this way, passing from linear functionals to signed Borel measures as optimization variables, the program (27) reads

$$\underset{\substack{a \in \mathbb{R}^m \\ \mu, \nu}}{\text{minimize}} \int_{-1}^1 \varepsilon \, d|\mu| + \kappa \, d|\nu| \qquad \text{subject to } \mu + \nu = \rho - \sum_{i=1}^m a_i \lambda_i \text{ and } \int_{-1}^1 v(x) d\mu(x) = 0 \text{ for all } v \in \mathcal{P}_n. \tag{45}$$

Let us introduce as slack variables the nonnegative Borel measures μ^+, μ^-, ν^+, and ν^- involved in the Jordan decompositions $\mu = \mu^+ - \mu^-$ and $\nu = \nu^+ - \nu^-$, so that the problem (45) is recast as

$$\underset{\substack{a \in \mathbb{R}^m \\ \mu^\pm, \nu^\pm}}{\text{minimize}} \int_{-1}^1 \varepsilon \, d(\mu^+ + \mu^-) + \kappa \, d(\nu^+ + \nu^-) \quad \text{s.to } \mu^+ - \mu^- + \nu^+ - \nu^- = \rho - \sum_{i=1}^m a_i \lambda_i \text{ and } \int_{-1}^1 v(x) d(\mu^+ - \mu^-)(x) = 0 \text{ for all } v \in \mathcal{P}_n. \tag{46}$$

Next, replacing the measures $\mu^\pm$ and $\nu^\pm$ by the infinite sequences of moments $w^\pm = \mathcal{M}_\infty(\mu^\pm) \in \mathbb{R}^\mathbb{N}$ and $z^\pm = \mathcal{M}_\infty(\nu^\pm) \in \mathbb{R}^\mathbb{N}$ of moments defined for $k \ge 1$ by

$$w_k^\pm = \int_{-1}^1 T_{k-1}(x) d\mu^\pm(x), \qquad z_k^\pm = \int_{-1}^1 T_{k-1}(x) d\nu^\pm(x), \tag{47}$$

the problem (46) is equivalent[1] to the infinite semidefinite program

$$\underset{\substack{a \in \mathbb{R}^m \\ w^\pm, z^\pm \in \mathbb{R}^\mathbb{N}}}{\text{minimize}} \ \varepsilon \, (w_1^+ + w_1^-) + \kappa \, (z_1^+ + z_1^-), \quad \text{s.to } w^+ - w^- + z^+ - z^- = \mathcal{M}_\infty\Big(\rho - \sum_{i=1}^m a_i \lambda_i\Big), \text{ and } w_j^+ - w_j^- = 0 \text{ for all } j \in [1:n], \text{ and } \mathrm{Toep}_\infty(w^\pm) \succeq 0, \ \mathrm{Toep}_\infty(z^\pm) \succeq 0. \tag{48}$$

[1] the equivalence is based on the discrete trigonometric moment problem, see [7] for details.

Instead of solving this infinite optimization program, we truncate it to a level $N \geq n$ and solve instead the resulting finite semidefinite program

$$\begin{aligned} (49) \qquad \underset{\substack{a \in \mathbb{R}^m \\ w^\pm, z^\pm \in \mathbb{R}^N}}{\text{minimize}} \ & \varepsilon\,(w_1^+ + w_1^-) + \kappa\,(z_1^+ + z_1^-), \qquad \text{s.to } w^+ - w^- + z^+ - z^- = \mathcal{M}_N\Big(\rho - \sum_{i=1}^m a_i \lambda_i\Big) \\ & \qquad\qquad \text{and } w_j^+ - w_j^- = 0 \text{ for all } j \in [1:n], \\ & \qquad\qquad \text{and } \mathrm{Toep}_N(w^\pm) \succeq 0, \ \mathrm{Toep}_N(z^\pm) \succeq 0. \end{aligned}$$

The rest of this section is devoted to justifying in a quantitative way that the minimal value of this truncated problem converges to the minimal value of the original problem. We also justify, although not quantitatively, that the vectors $a^{(N)} \in \mathbb{R}^m$ obtained by solving (49) converge as $N \to \infty$ to a vector $a^{\mathrm{opt}} \in \mathbb{R}^m$ solving (27) — we do not analyze the behavior of the optimal measures since only the vector a^{opt} is required in (26). From now on, we also work under the assumption of linear independence for the restrictions $\ell_{1|\mathcal{P}_m}, \ldots, \ell_{m|\mathcal{P}_m}$ of the observation functionals to the space $\mathcal{P}_m$ of polynomials of degree $< m$. This assumption is easily seen to be equivalent to the invertibility of (the transpose of) the moment matrix $M \in \mathbb{R}^{m \times m}$ defined by

$$(50) \qquad M_{j,i} = \ell_i(T_{j-1}), \qquad i,j \in [1:m].$$

This holds e.g. when the observation functionals are evaluations at m distinct points in $[-1,1]$.

Theorem 5. Suppose that the system $(\ell_{1|\mathcal{P}_m}, \ldots, \ell_{m|\mathcal{P}_m})$ is linearly independent. If there is a unique $a^{\mathrm{opt}} \in \mathbb{R}^m$ yielding a minimizer $(a^{\mathrm{opt}}, \mu^{\mathrm{opt}}, \nu^{\mathrm{opt}})$ of (27), then a^{opt} is the limit of any sequence $(a^{(N)})_{N \geq n}$ obtained by solving (49) for each $N \geq n$. Without uniqueness, it still holds that any subsequence of $(a^{(N)})_{N \geq n}$ admits a subsequence converging to the first component of a minimizer of (27).

Proof. The first part of the theorem follows from the second part: it is indeed well-known that the convergence of a sequence towards a given point is guaranteed as soon as any of its subsequences admits a subsequence converging to that point.

To establish the second part, let $\alpha^{(N)} \in \mathbb{R}$ and $(a^{(N)}, w^{\pm,(N)}, z^{\pm,(N)}) \in \mathbb{R}^m \times (\mathbb{R}^N)^4$ denote, for each $N \geq n$, the minimum value and some minimizer of (49), respectively. We write $w^{\pm,((N))} \in \mathbb{R}^{\mathbb{N}}$ and $z^{\pm,((N))} \in \mathbb{R}^{\mathbb{N}}$ for the infinite vectors obtained by padding the finite vectors $w^{\pm,(N)} \in \mathbb{R}^N$ and $z^{\pm,(N)} \in \mathbb{R}^N$ with zeros. Let us now consider a subsequence $\big((a^{(N_k)}, w^{\pm,((N_k))}, z^{\pm,((N_k))})\big)_{k \geq 1}$ of the $(\ell_\infty^m \times (\ell_\infty^{\mathbb{N}})^4)$-valued sequence $\big((a^{(N)}, w^{\pm,(N)}, z^{\pm,(N)})\big)_{N \geq n}$. Our objective is to show that there exist a subsequence $\big((a^{(N_{k_\ell})}, w^{\pm,((N_{k_\ell}))}, z^{\pm,((N_{k_\ell}))})\big)_{\ell \geq 1}$ and a minimizer $(\widetilde{a}, \widetilde{w}^\pm, \widetilde{z}^\pm)$ of (48) such that $a^{(N_{k_\ell})}$ converges to $\widetilde{a}$ as $\ell \to \infty$. To this end, we start by observing that the sequence $(\alpha^{(N_k)})_{k \geq 1}$ is nondecreasing and bounded by the minimal value α^{opt} of (48): firstly, the inequality $\alpha^{(N_k)} \leq \alpha^{(N_{k+1})}$ follows from the feasibility of $(a^{(N_{k+1})}, w^{\pm,(N_{k+1})}_{[1:N_k]}, z^{\pm,(N_{k+1})}_{[1:N_k]})$ for (49) specified to $N = N_k$, so that

$$(51) \qquad \alpha^{(N_k)} \leq \varepsilon\,(w_1^{+,(N_{k+1})} + w_1^{-,(N_{k+1})}) + \kappa\,(z_1^{+,(N_{k+1})} + z_1^{-,(N_{k+1})}) = \alpha^{(N_{k+1})};$$

secondly, the inequality $\alpha^{(N_k)} \le \alpha^{\mathrm{opt}}$ similarly follows from the feasibility of $(a^{\mathrm{opt}}, w^{\pm,\mathrm{opt}}_{[1:N_k]}, z^{\pm,\mathrm{opt}}_{[1:N_k]})$ for (49) specified to $N = N_k$, where evidently $(a^{\mathrm{opt}}, w^{\pm,\mathrm{opt}}, z^{\pm,\mathrm{opt}})$ represents some minimizer of (48). We continue by remarking that the positive semidefiniteness of $\mathrm{Toep}_{N_k}(w^{\pm,(N_k)})$ and $\mathrm{Toep}_{N_k}(z^{\pm,(N_k)})$ yields, for any $j \in [1:N_k]$,

$$\text{(52)} \quad \big|w_j^{\pm,(N_k)}\big| \le w_1^{\pm,(N_k)} \le \frac{1}{\varepsilon}\big(\varepsilon\,(w_1^{+,(N_k)} + w_1^{-,(N_k)}) + \kappa\,(z_1^{+,(N_k)} + z_1^{-,(N_k)})\big) \le \frac{1}{\varepsilon}\alpha^{(N_k)} \le \frac{1}{\varepsilon}\alpha^{\mathrm{opt}},$$

$$\text{(53)} \quad \big|z_j^{\pm,(N_k)}\big| \le z_1^{\pm,(N_k)} \le \frac{1}{\kappa}\big(\varepsilon\,(w_1^{+,(N_k)} + w_1^{-,(N_k)}) + \kappa\,(z_1^{+,(N_k)} + z_1^{-,(N_k)})\big) \le \frac{1}{\kappa}\alpha^{(N_k)} \le \frac{1}{\kappa}\alpha^{\mathrm{opt}}.$$

Thus, the $\ell_\infty^{\mathbb{N}}$-valued sequences $(w^{\pm,((N_k))})_{k\ge1}$ and $(z^{\pm,((N_k))})_{k\ge1}$ are bounded. This guarantees, by the sequential compactness Banach–Alaoglu theorem, that $(w^{\pm,((N_k))})_{k\ge1}$ and $(z^{\pm,((N_k))})_{k\ge1}$ admit convergent subsequences in the weak-star topology of $\ell_\infty^{\mathbb{N}}$. We denote the resulting convergent subsequence and its limit by $\big((w^{\pm,((N_{k_\ell}))}, z^{\pm,((N_{k_\ell}))})\big)_{\ell\ge1}$ and $(\widetilde{w}^{\pm}, \widetilde{z}^{\pm})$, respectively. They come with associated $(a^{(N_{k_\ell})})_{\ell\ge1}$ and $\widetilde{a}$. Indeed, the constraint imposed on minimizers of (49) yields in particular, for $N_{k_\ell} \ge m$,

$$\text{(54)} \qquad M a^{(N_{k_\ell})} = \mathcal{M}_m(\rho) - (w^{+,(N_{k_\ell})} - w^{-,(N_{k_\ell})} + z^{+,(N_{k_\ell})} - z^{-,(N_{k_\ell})})_{[1:m]}.$$

In view of $w_j^{\pm,(N_{k_\ell})} \to \widetilde{w}_j^{\pm}$ and $z_j^{\pm,(N_{k_\ell})} \to \widetilde{z}_j^{\pm}$ for all $j \ge 1$, which is a consequence of the weak-star convergence, and of the invertibility assumption for the moment matrix, we see that

$$\text{(55)} \qquad a^{(N_{k_\ell})} \underset{\ell\to\infty}{\longrightarrow} \widetilde{a} := M^{-1}\big(\mathcal{M}_m(\rho) - (\widetilde{w}^{+} - \widetilde{w}^{-} + \widetilde{z}^{+} - \widetilde{z}^{-})_{[1:m]}\big).$$

It remains to prove that the quintuple $(\widetilde{a}, \widetilde{w}^{\pm}, \widetilde{z}^{\pm})$ is a minimizer of (48). Writing the constraint of (49) satisfied by $(a^{(N_{k_\ell})}, w^{\pm,(N_{k_\ell})}, z^{\pm,(N_{k_\ell})})$ and passing to the limit as $\ell \to \infty$ shows that the quintuple is feasible for (48). It is also a minimizer for this program, by virtue of

$$\begin{aligned}\text{(56)} \qquad \varepsilon\,(\widetilde{w}_1^{+} + \widetilde{w}_1^{-}) + \kappa\,(\widetilde{z}_1^{+} + \widetilde{z}_1^{-}) &= \lim_{\ell\to\infty} \big(\varepsilon\,(w_1^{+,(N_{k_\ell})} + w_1^{-,(N_{k_\ell})}) + \kappa\,(z_1^{+,(N_{k_\ell})} + z_1^{-,(N_{k_\ell})})\big) \\ &= \lim_{\ell\to\infty} \alpha^{(N_{k_\ell})} \le \alpha^{\mathrm{opt}}.\end{aligned}$$

Our objective is now established, so the second part of theorem is proved. □

Theorem 5 does not tell us how to choose N in order to reach a prescribed accuracy on $\|a^{\mathrm{opt}} - a^{(N)}\|$, not even on $\alpha^{\mathrm{opt}} - \alpha^{(N)}$. We intend to provide an accuracy estimate for the latter, which we do under the further restriction that the observation functionals and the quantity of interest are point evaluations at distinct $x_1, \dots, x_m \in [-1,1]$ and at $x_0 \in [-1,1] \setminus \{x_1, \dots, x_m\}$, respectively. Moreover, the estimate is an *a posteriori* one, in the sense that we need to solve problem (49) for a particular N first. As a matter of fact, we also need to solve an extra linear program subordinated

to a particular grid $\underline{t} = (t_1, \ldots, t_K)$ avoiding the x_i's. This program is

$$\underset{\substack{a\in\mathbb{R}^m \\ u,v\in\mathbb{R}^{1+m+K} \\ r,s\in\mathbb{R}^{1+m+K}}}{\text{minimize}} \ \varepsilon \sum_h r_h + \kappa \sum_h s_h \qquad \text{s.to } u+v = \begin{bmatrix} 1 \\ \hline -a \\ \hline 0 \end{bmatrix}, \ \Big[\, b \,|\, C \,|\, D \,\Big]\, u = 0, \tag{57}$$

$$\text{and } r+u \ge 0,\ r-u \ge 0, s+v \ge 0,\ s-v \ge 0.$$

Here, the vector $b \in \mathbb{R}^n$, the matrix $C \in \mathbb{R}^{n\times m}$ (both encountered before), and the matrix $D \in \mathbb{R}^{n\times K}$ have entries $b_j = T_{j-1}(x_0)$, $C_{j,i} = T_{j-1}(x_i)$, and $D_{j,k} = T_{j-1}(t_k)$, $i \in [1:m]$, $j \in [1:n]$, $k \in [1:K]$.

Theorem 6. Suppose that the observation functionals $\ell_1, \ldots, \ell_m$ and the quantity of interest Q take the form $\ell_i(f) = f(x_i)$, $i \in [1:m]$, and $Q(f) = f(x_0)$ for distinct points $x_0, x_1, \ldots, x_m \in [-1,1]$. For any $N \ge n$ and any grid $\underline{t}$ of $[-1,1]$ not intersecting $\{x_0, x_1, \ldots, x_m\}$, the minimal value α^{opt} of (27) satisfies

$$\alpha^{(N)} \le \alpha^{\mathrm{opt}} \le \beta^{(\underline{t})}, \tag{58}$$

where $\alpha^{(N)}$ is the minimal value of (49) and $\beta^{(\underline{t})}$ is the minimal value of (57).

Proof. The leftmost inequality of (58) was already justified (implicitly) in the proof of Theorem 5 and actually does not rely on any assumption on the observation functionals or the quantity of interest. For the rightmost inequality of (58), we keep in mind that α^{opt} is the minimal value of the optimization program (45). Since this is a minimal value over all signed Borel measures, an upper bound is provided by the minimal value $\beta^{(\underline{t})}$ over the subset of all signed Borel measures consisting of linear combinations of Dirac measures at the distinct points $x_0, x_1, \ldots, x_m, t_1, \ldots, t_K$ — the points $x_0, x_1, \ldots, x_m$ are included in order to make the constraint of (45) feasible. Writing such measures as

$$\mu = u\,\delta_{x_0} + \sum_{i=1}^m u_i' \delta_{x_i} + \sum_{k=1}^K u_k'' \delta_{t_k}, \tag{59}$$

$$\nu = v\,\delta_{x_0} + \sum_{i=1}^m v_i' \delta_{x_i} + \sum_{k=1}^K v_k'' \delta_{t_k}, \tag{60}$$

we see that the upper bound $\beta^{(\underline{t})}$ takes the form

$$\beta^{(\underline{t})} = \min_{\substack{a\in\mathbb{R}^m, u,v\in\mathbb{R} \\ u',v'\in\mathbb{R}^m, u'',v''\in\mathbb{R}^K}} \Big\{ \varepsilon(|u|+\|u'\|_1 + \|u''\|_1) + \kappa(|v| + \|v'\|_1 + \|v''\|_1) : \tag{61}$$

$$u+v=1,\ u'+v'=-a,\ u''+v''=0,\ ub + Cu' + Du'' = 0 \Big\}.$$

By gathering $u \in \mathbb{R}$, $u' \in \mathbb{R}^m$, and $u'' \in \mathbb{R}^K$ into a single vector $\widehat{u} = [u; u'; u''] \in \mathbb{R}^{1+m+K}$, which we later rename u, and similarly for $v \in \mathbb{R}$, $v' \in \mathbb{R}^m$, and $v'' \in \mathbb{R}^K$, the objective function simply reads $\varepsilon\|\widehat{u}\|_1 + \kappa\|\widehat{v}\|_1$, while the constraints read $\widehat{u} + \widehat{v} = [1; -a; 0]$ and $[b, C, D]\widehat{u} = 0$. The final transformation applied to arrive at the linear program of (57) consists in introducing slack variables $r, s \in \mathbb{R}^{1+m+K}$ such that $|\widehat{u}| \le r$ and $|\widehat{v}| \le s$. □

Remark. When the observation functionals and the quantity of interest are point evaluations, a solution $a^{(\underline{t})}$ to the linear program (57) yields a linear functional $R^{(\underline{t})} : y \in \mathbb{R}^m \mapsto \sum_{i=1}^m a_i^{(\underline{t})} y_i \in \mathbb{R}$ which turns out to be a near-optimal recovery map. Indeed, remembering that the first step of the proof of Theorem 4 is valid for $a^{(\underline{t})}$, which comes with atomic measures $\mu^{(\underline{t})}$ and $\nu^{(\underline{t})}$ such that $\mu^{(\underline{t})} + \nu^{(\underline{t})} = \rho - \sum_{i=1}^m a_i^{(\underline{t})} \lambda_i$ and $\int_{-1}^1 v(x) d\mu^{(\underline{t})}(x) = 0$ for all $v \in \mathcal{P}_n$, we derive that the error of $R^{(\underline{t})}$ over $\mathcal{K}$ satisfies

$$\sup_{f \in \mathcal{K}} |Q(f) - R^{(\underline{t})}(L(f))| \le \int_{-1}^1 \varepsilon \, d|\mu^{(\underline{t})}| + \kappa \, d|\nu^{(\underline{t})}| = \beta^{(\underline{t})}. \tag{62}$$

This estimate matches the intrinsic error α^{opt} with error at most $\beta^{(\underline{t})} - \alpha^{(N)}$. This quantity, which is available after solving (49) and (57), is small when the truncation parameter N and the size of the grid $\underline{t}$ are large.

References

[1] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. Robust Optimization. Princeton University Press, 2009.

[2] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. *Data assimilation in reduced modeling.* SIAM/ASA Journal on Uncertainty Quantification 5.1 (2017): 1–29.

[3] S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2004.

[4] R. DeVore, S. Foucart, G. Petrova, and P. Wojtaszczyk. *Computing a quantity of interest from observational data.* Constructive Approximation 49.3 (2019): 461–508.

[5] M. Ettehad and S. Foucart. *Instances of computational optimal recovery: dealing with observation errors.* Preprint.

[6] S. Foucart. *Computation of minimal projections and extensions.* Numerical Functional Analysis and Optimization 37.2 (2016): 159–185.

[7] S. Foucart and J. B. Lasserre. *Computation of Chebyshev polynomials for union of intervals.* Computational Methods and Function Theory, 19.4 (2019): 625–641.

[8] S. Foucart and V. Powers. *BASC: constrained approximation by semidefinite programming.* IMA Journal of Numerical Analysis 37.2 (2017): 1066–1085.

[9] J. Lasserre. Moments, Positive Polynomials and Their Applications. Vol. 1. Imperial College Press, 2009.

[10] E. Novak and H. Woźniakowski. Tractability of Multivariate Problems: Linear Information. Vol. 6. European Mathematical Society, 2008.

[11] W. Rudin. Functional Analysis (2nd edition). International Series in Pure and Applied Mathematics, McGraw-Hill. 1991.

[12] S. A. Smolyak. *On optimal restoration of functions and functionals of them.* Candidate dissertation, Moscow State University, 1965.