

Optimal Algorithms for Nonlinear Estimation with Convex Models

Simon Foucart*— Texas A&M University

Abstract

A linear functional of an object from a convex symmetric set can be optimally estimated, in a worst-case sense, by a linear functional of observations made on the object. This well-known fact is extended here to a nonlinear setting: other simple functionals of the object can be optimally estimated by functionals of the observations that share a similar simple structure. This is established for the maximum of several linear functionals and even for the ℓ th largest among them. Proving the latter requires an unusual variation of the analytical Hahn–Banach theorem. The existence results are accompanied by practical recipes relying on convex optimization to construct the desired functionals, thereby justifying the term of estimation algorithms.

Key words and phrases: Optimal recovery, Minimax problems, Dominated extension theorem.

AMS classification: 41A28, 46N10, 65D15, 90C47.

1 Introduction

Suppose that an unknown object f , living in a vector space F , is acquired via a vector $\Lambda(f) \in \mathbb{R}^m$ of linear observations $\lambda_1(f), \dots, \lambda_m(f) \in \mathbb{R}$, in addition to the prior knowledge that it belongs to a prescribed model set $\mathcal{K}$. Estimating a quantity $\gamma(f)$, living in some normed space X , boils down to devising a recovery map $\delta : \mathbb{R}^m \rightarrow X$. From a worst-case perspective, its performance is measured through

$$e(\delta) := \sup_{f \in \mathcal{K}} \|\gamma(f) - \Delta(\Lambda(f))\|.$$

For a fixed observation map Λ , constructing a map δ that minimizes the above is the essence of Optimal Recovery [8], a lead-in to the field of Information-Based Complexity [13], where one is also interested in minimizing over Λ , with possibly some restrictions on the nature of the observations (see the comprehensive survey [7] for results in this direction). From here on, the Optimal Recovery framework prevails, so the dependence on the linear observation map $\Lambda : F \rightarrow \mathbb{R}^m$ —considered a

*S. F. is partially supported by a grant from the NSF (DMS-2505204).

fixed entity—will not be made explicit. For instance, it is implicit in the lower bound e_b below, asserting that for any $\delta : \mathbb{R}^m \rightarrow X$,

$$e(\delta) \geq e_b, \quad \text{where} \quad e_b := \sup_{\substack{f', f'' \in \mathcal{K} \\ \Lambda(f') = \Lambda(f'')}} \frac{\|\gamma(f') - \gamma(f'')\|}{2}.$$

This is well known and easy to derive, and so is the complementary inequality $e(\gamma \circ \Delta^{\text{cons}}) \leq 2e_b$, where $\Delta^{\text{cons}} : \mathbb{R}^m \rightarrow F$ is any data- and model-consistent map¹, i.e., one satisfying $\Lambda(\Delta^{\text{cons}}(y)) = y$ and $\Delta^{\text{cons}}(y) \in \mathcal{K}$ for all $y \in \Lambda(\mathcal{K})$. In short, the map $\gamma \circ \Delta^{\text{cons}}$ is nearly optimal (with a factor 2). To surpass this, one seeks maps $\delta^{\text{opti}} : \mathbb{R}^m \rightarrow X$ yielding $e(\delta^{\text{opti}}) \leq e_b$, which are therefore genuinely optimal (with a factor 1). This is achievable when γ is a linear map taking values in $X = \mathbb{R}$, i.e., it is a linear functional. Such a classical result is due to Smolyak (see e.g. [10, Theorem 4.7]) when the set $\mathcal{K}$ is convex and symmetric about the origin and was refined by Sukharev [12] when $\mathcal{K}$ is merely convex. These two results also guarantee the existence of a genuinely optimal estimation functional $\delta^{\text{opti}} : \mathbb{R}^m \rightarrow \mathbb{R}$ which is linear in the former case and affine in the latter case.

The goal of this article is, firstly, to exhibit examples of nonlinear yet ‘simple’ functionals $\gamma : F \rightarrow \mathbb{R}$ for which there exist ‘simple’ estimation functionals $\delta^{\text{opti}} : \mathbb{R}^m \rightarrow \mathbb{R}$ that are genuinely optimal, in the sense that

$$e(\delta^{\text{opti}}) \leq e(\delta) \quad \text{for any } \delta : \mathbb{R}^m \rightarrow \mathbb{R},$$

and secondly, to complement existence results with practical constructions whenever possible. As a prototypical example of nonlinear functionals, the maximum of a function f was considered early on, often with near optimality in mind (see e.g. [14]), but not always (see e.g. [9, p.28–31]). With genuine optimality in mind, it has also been treated recently relatively to a model set $\mathcal{K}$ consisting of Hölder functions within the space F of continuous functions, see [4]. The arguments were specific to the case at hand, so the situation where the function space F is a reproducing kernel Hilbert space, say, was not covered. The results obtained in the present article do apply to such a situation: for instance, Theorem 1 will guarantee the existence of a genuinely optimal estimation functional $\delta^{\text{opti}} : \mathbb{R}^m \rightarrow \mathbb{R}$ which is convex and Proposition 6 will later provide a practical recipe to construct it, under the proviso that the maximum is over a finite set in order for a convex optimization program to be solvable. As an extension, Theorem 3 will establish e.g. that the functional outputting the ℓ th largest value among a number of linear functionals comes with a genuinely optimal estimation functional $\delta^{\text{opti}} : \mathbb{R}^m \rightarrow \mathbb{R}$ which is the supremum of infima of affine functionals. This result is based on a variation of the Hahn–Banach dominated extension theorem, see Lemma 2.

Here is a brief outline of the organization of the article: theoretical results such as Theorem 1 and Theorem 3 are proved right below (Section 2), before computational considerations are addressed (Section 3), first at an abstract level and then specialized to two particular examples.

¹A data- and model-consistent map is usually called an interpolatory algorithm in Information-Based Complexity, see e.g. [13, p.49–53].

2 Optimal Estimation of Some Nonlinear Functionals

This section starts by uncovering the ‘simple’ form of an optimal functional for the estimation of a nonlinear functional γ which appears as the supremum of linear functionals. It continues by enlarging the result to include γ ’s that are a mixed supremum-infimum of linear functionals. In the process, an unusual variation of Hahn–Banach dominated extension theorem is exploited. The section ends with a discussion on the handling of possible errors in the observation procedure.

2.1 The supremum of linear functionals

In the result presented below, valid in an arbitrary normed space F , the nonlinear functional γ to be optimally estimated is the supremum of linear functionals indexed by a set I . Since the latter need not be finite, it transpires that any norm $\|f\| = \max\{\eta(f) : \eta \in F^* \text{ with } \|\eta\|_* = 1\}$ can be optimally estimated by way of a functional $\delta^{\text{opti}} : \mathbb{R}^m \rightarrow \mathbb{R}$ which, as the supremum of affine functionals, is a convex functional. Of course, an analog result could be obtained if γ was the infimum of linear functionals by simply applying Theorem 1 to $-\gamma$.

Theorem 1. Suppose that the model set $\mathcal{K}$ is convex and contains the origin and that the functional $\gamma : F \rightarrow \mathbb{R}$ is sup-linear, i.e., of the form

$$\gamma(f) = \sup_{i \in I} \gamma_i(f), \quad f \in F,$$

where the $\gamma_i \in F^*$ are linear functionals. Then there exists an optimal estimation functional $\delta^{\text{opti}} : \mathbb{R}^m \rightarrow \mathbb{R}$ which is also sup-affine, i.e., of the form

$$\delta^{\text{opti}}(y) = \sup_{i \in I} \left(c_0^{(i)} + \sum_{k=1}^m c_k^{(i)} y_k \right), \quad y \in \mathbb{R}^m.$$

Proof. As a first step, the previously mentioned lower bound is rewritten as

$$e_b = \sup_{\substack{f', f'' \in \mathcal{K} \\ \Lambda(f') = \Lambda(f'')}} \frac{\gamma(f') - \gamma(f'')}{2} = \sup_{\bar{f} \in \bar{\mathcal{K}} \cap \ker(\bar{\Lambda})} \frac{\bar{\gamma}'(\bar{f}) - \bar{\gamma}''(\bar{f})}{2},$$

where one has introduced the set $\bar{\mathcal{K}} := \mathcal{K} \times \mathcal{K} \subseteq \bar{F} := F \times F$, as well as the linear map $\bar{\Lambda} : \bar{F} \rightarrow \mathbb{R}^m$ and the (nonlinear) functionals $\bar{\gamma}', \bar{\gamma}'' : \bar{F} \rightarrow \mathbb{R}$ defined via

$$\bar{\Lambda}([f'; f'']) = \Lambda(f') - \Lambda(f''), \quad \bar{\gamma}'([f'; f'']) = \gamma(f'), \quad \bar{\gamma}''([f'; f'']) = \gamma(f'').$$

With $\rho_{\bar{\mathcal{K}}}$ denoting the Minkowski functional of $\bar{\mathcal{K}}$ —which is a sublinear (i.e., positively homogeneous and convex) functional by virtue of the fact that $\bar{\mathcal{K}}$ is convex and contains the origin—one has

$\bar{\gamma}'(\bar{f}) - \bar{\gamma}''(\bar{f}) \leq 2e_b \rho_{\bar{\mathcal{K}}}(\bar{f})$ whenever $\bar{f} \in \ker(\bar{\Lambda})$. Also introducing linear functionals $\bar{\gamma}'_i, \bar{\gamma}''_i : \bar{F} \rightarrow \mathbb{R}$ defined for all $i \in I$ via $\bar{\gamma}'_i([f'; f'']) = \gamma_i(f')$ and $\bar{\gamma}''_i([f'; f'']) = \gamma_i(f'')$, so that $\bar{\gamma}'(\bar{f}) = \sup_{i \in I} \bar{\gamma}'_i(\bar{f})$, the previous inequality ensures that, for all $i_\star \in I$, one has $\bar{\gamma}'_{i_\star}(\bar{f}) \leq 2e_b \rho_{\bar{\mathcal{K}}}(\bar{f}) + \bar{\gamma}''(\bar{f})$ whenever $\bar{f} \in \ker(\bar{\Lambda})$. Fixing $i_\star$ and noticing that $2e_b \rho_{\bar{\mathcal{K}}} + \bar{\gamma}''$ is a sublinear functional, the Hahn–Banach dominated extension theorem (see e.g. [11, Theorem 3.2]) applies to guarantee the existence of a linear functional $\mu_{i_\star} \in \bar{F}^*$ such that $\mu_{i_\star}|_{\ker(\bar{\Lambda})} = \bar{\gamma}'_{i_\star}|_{\ker(\bar{\Lambda})}$ and $\mu_{i_\star}(\bar{f}) \leq 2e_b \rho_{\bar{\mathcal{K}}}(\bar{f}) + \bar{\gamma}''(\bar{f})$ for all $\bar{f} \in \bar{F}$. In view of the equivalence between the fact that $\bar{\gamma}'_{i_\star} - \mu_{i_\star}$ vanishes on $\ker(\bar{\Lambda})$ and the existence of $c^{(i_\star)} \in \mathbb{R}^m$ such that $\bar{\gamma}'_{i_\star} - \mu_{i_\star} = \sum_{k=1}^m c_k^{(i_\star)} \bar{\lambda}_k$ (see e.g. [11, Lemma 3.9]), one derives that

$$\bar{\gamma}'_{i_\star}(\bar{f}) - \sum_{k=1}^m c_k^{(i_\star)} \bar{\lambda}_k(\bar{f}) \leq 2e_b \rho_{\bar{\mathcal{K}}}(\bar{f}) + \bar{\gamma}''(\bar{f}) \quad \text{for all } \bar{f} \in \bar{F}.$$

In other words, for any $i_\star \in I$, it holds that

$$\gamma_{i_\star}(f') - \sum_{k=1}^m c_k^{(i_\star)} (\lambda_k(f') - \lambda_k(f'')) \leq 2e_b \max\{\rho_{\mathcal{K}}(f'), \rho_{\mathcal{K}}(f'')\} + \gamma(f'') \quad \text{for all } f', f'' \in F,$$

which immediately yields the key observation that

$$(1) \quad \left(\gamma_{i_\star}(f') - \sum_{k=1}^m c_k^{(i_\star)} \lambda_k(f') \right) - \left(\gamma(f'') - \sum_{k=1}^m c_k^{(i_\star)} \lambda_k(f'') \right) \leq 2e_b \quad \text{for all } f', f'' \in \mathcal{K}.$$

At this point, for each $i_\star \in I$, one defines

$$(2) \quad c_0^{(i_\star)} = \inf_{f \in \mathcal{K}} \left(\gamma(f) - \sum_{k=1}^m c_k^{(i_\star)} \lambda_k(f) + e_b \right).$$

This definition readily implies that, for each $i_\star \in I$, one has $\gamma(f) - \left(c_0^{(i_\star)} + \sum_{k=1}^m c_k^{(i_\star)} \lambda_k(f) \right) \geq -e_b$ for all $f \in \mathcal{K}$. One can now take the supremum over $i_\star \in I$ to arrive at

$$(3) \quad \gamma(f) - \sup_{i \in I} \left(c_0^{(i)} + \sum_{k=1}^m c_k^{(i)} \lambda_k(f) \right) \geq -e_b \quad \text{for all } f \in \mathcal{K}.$$

Next, for each $i_\star \in I$, one selects $f'' \in \mathcal{K}$ such that $\gamma(f'') - \sum_{k=1}^m c_k^{(i_\star)} \lambda_k(f'') + e_b$ is equal (or is arbitrarily close) to $c_0^{(i_\star)}$. Writing (1) for this f'' and for f' being an arbitrary $f \in \mathcal{K}$ leads to $\gamma_{i_\star}(f) \leq e_b + \left(c_0^{(i_\star)} + \sum_{k=1}^m c_k^{(i_\star)} \lambda_k(f) \right)$, and hence $\gamma_{i_\star}(f) \leq e_b + \sup_{i \in I} \left(c_0^{(i)} + \sum_{k=1}^m c_k^{(i)} \lambda_k(f) \right)$. Taking the supremum over $i_\star \in I$ yields

$$(4) \quad \gamma(f) - \sup_{i \in I} \left(c_0^{(i)} + \sum_{k=1}^m c_k^{(i)} \lambda_k(f) \right) \leq e_b \quad \text{for all } f \in \mathcal{K}.$$

Now defining the estimation functional $\delta^{\text{opti}} : \mathbb{R}^m \rightarrow \mathbb{R}$ by $\delta^{\text{opti}}(y) = \sup_{i \in I} \left(c_0^{(i)} + \sum_{k=1}^m c_k^{(i)} y_k \right)$ and combining inequalities (3) and (4) show that

$$e(\delta^{\text{opti}}) = \sup_{f \in \mathcal{K}} |\gamma(f) - \delta^{\text{opti}}(\Lambda(f))| \leq e_b.$$

This proves the genuine optimality of the sup-affine functional δ^{opti} . $\square$

Remark. Although natural from an Optimal Recovery point of view, invoking Hahn–Banach dominated extension theorem may appear quite abstract, so building intuition via an alternative route could be welcome. Here is a sketch of such a convex-duality route, pointed out by a referee. Start by writing $2e_b = \sup\{\gamma(f') - \gamma(f'') : f', f'' \in \mathcal{K}, \Lambda(f') = \Lambda(f'')\}$ as

$$2e_b = \sup_{i \in I} \mu_i, \quad \text{where} \quad \mu_i := \sup_{\substack{f', f'' \in \mathcal{K} \\ \Lambda(f') = \Lambda(f'')}} (\gamma_i(f') - \gamma(f'')).$$

Since the program defining $-\mu_i$ is convex, consider its dual program to obtain (after some work, in particular to justify strong duality)

$$\mu_i = \inf_{c^{(i)} \in \mathbb{R}^m} \sup_{f', f'' \in \mathcal{K}} \left(\gamma_i(f') - \gamma(f'') - \sum_{k=1}^m c_k^{(i)} (\lambda_k(f') - \lambda_k(f'')) \right).$$

With $c^{(i)}$ representing for simplicity a minimizer of the above, observe that

$$\mu_i - e_b = \sup_{f' \in \mathcal{K}} \left(\gamma_i(f') - \sum_{k=1}^m c_k^{(i)} \lambda_k(f') \right) + \sup_{f'' \in \mathcal{K}} \left(-\gamma(f'') + \sum_{k=1}^m c_k^{(i)} \lambda_k(f'') - e_b \right).$$

Denote the latter supremum as $-c_0^{(i)}$, in accordance with (2), and deduce that

$$e_b = \sup_{i \in I} (\mu_i - e_b) = \sup_{f' \in \mathcal{K}} \sup_{i \in I} \left(\gamma_i(f') - \sum_{k=1}^m c_k^{(i)} \lambda_k(f') - c_0^{(i)} \right) \geq \sup_{f' \in \mathcal{K}} \left(\gamma(f') - \delta^{\text{opti}}(\Lambda f') \right)$$

where $\delta^{\text{opti}}(y) = \sup_{i \in I} \left(\sum_{k=1}^m c_k^{(i)} y_k + c_0^{(i)} \right)$. This inequality is one of two that explain the optimality of the estimation functional δ^{opti} . The other one, namely $\sup_{f'' \in \mathcal{K}} \left(-\gamma(f'') + \delta^{\text{opti}}(\Lambda f'') \right) \leq e_b$, is a direct consequence of the definition of $c_0^{(i)}$.

2.2 The mixed supremum-infimum of linear functionals

It is now acquired that there is a ‘simple’ optimal algorithm for the estimation of the (pointwise) largest value of linear functionals $\gamma_1, \dots, \gamma_d$, say. Is there also a ‘simple’ optimal algorithm for the estimation of the (pointwise) ℓ th largest value of these linear functionals? The answer is yes, as established in Theorem 3 below. ‘Simple’ in this case means the mixed sup-inf of affine functionals. This mirrors the sup-inf/inf-sup nature of the functional $\gamma_\ell^\downarrow$ associating to $f \in F$ the ℓ th largest value of $\gamma_1(f), \dots, \gamma_d(f)$, namely

$$\gamma_\ell^\downarrow(f) = \sup_{|I|=\ell} \inf_{i \in I} \gamma_i(f) = \inf_{|J|=d+1-\ell} \sup_{j \in J} \gamma_j(f).$$

The essential feature is to have both a sup-inf and an inf-sup representation of the nonlinear functional at hand, as e.g. in Courant–Fischer characterization of the ℓ th largest eigenvalue. Thus, the result will also apply to the difference of suprema of linear functionals, since

$$\sup_{i \in I} \mu_i(f) - \sup_{j \in J} \nu_j(f) = \sup_{i \in I} \inf_{j \in J} (\mu_i - \nu_j)(f) = \inf_{j \in J} \sup_{i \in I} (\mu_i - \nu_j)(f).$$

A crucial part of the argument will rely on the following unusual, yet simple, variation on Hahn–Banach dominated extension theorem.

Lemma 2. Let V be a vector space and let U be a finite codimensional subspace of V given as $U = \ker(\eta_1) \cap \dots \cap \ker(\eta_m)$ for some linear functionals $\eta_1, \dots, \eta_m \in V^*$. If $\mu_i \in V^*$, $i \in I$, are linear functionals on V and if ρ is a sublinear functional on V such that

$$\inf_{i \in I} \mu_i(u) \leq \rho(u) \quad \text{for all } u \in U,$$

then there exist scalars $c_1, \dots, c_m \in \mathbb{R}$ such that

$$\inf_{i \in I} \mu_i(v) + \sum_{k=1}^m c_k \eta_k(v) \leq \rho(v) \quad \text{for all } v \in V.$$

Proof. The assumption says that the zero linear functional is dominated on the subspace U by the sublinear functional $\rho - \inf_{i \in I} \mu_i$. Applying the standard Hahn–Banach dominated extension theorem then yields a linear functional ω such that $\omega \leq \rho - \inf_{i \in I} \mu_i$ on the whole of V and $\omega|_U = 0$. Since the latter is equivalent to the existence of scalars $c_1, \dots, c_m \in \mathbb{R}$ such that $\omega = \sum_{k=1}^m c_k \eta_k$ (see e.g. [11, Lemma 3.9]), the announced result follows. $\square$

With this technical lemma at hand, the main result of the subsection can now be stated and proved.

Theorem 3. Suppose that the model set $\mathcal{K}$ is convex and contains the origin and that the functional $\gamma : F \rightarrow \mathbb{R}$ can be expressed as

$$\gamma(f) = \sup_{a \in A} \inf_{i \in I_a} \gamma_i(f) = \inf_{b \in B} \sup_{j \in J_b} \gamma_j(f), \quad f \in F,$$

where the $\gamma_i \in F^*$ are linear functionals. Then there exists an optimal estimation functional $\delta^{\text{opti}} : \mathbb{R}^m \rightarrow \mathbb{R}$ which has the form

$$\delta^{\text{opti}}(y) = \sup_{a \in A} \inf_{b \in B} \left(c_0^{(a,b)} + \sum_{k=1}^m c_k^{(a,b)} y_k \right), \quad y \in \mathbb{R}^m.$$

Proof. One starts from the lower bound e_b expressed once again as

$$e_b = \sup_{\substack{f', f'' \in \mathcal{K} \\ \Lambda(f') = \Lambda(f'')}} \frac{\gamma(f') - \gamma(f'')}{2} = \sup_{\bar{f} \in \bar{\mathcal{K}} \cap \ker(\bar{\Lambda})} \frac{\bar{\gamma}'(\bar{f}) - \bar{\gamma}''(\bar{f})}{2},$$

where it is recalled that the ‘bar’-notation involves $\bar{\mathcal{K}} := \mathcal{K} \times \mathcal{K} \subseteq \bar{F} := F \times F$, which has Minkowski functional $\rho_{\bar{\mathcal{K}}}$, and $\bar{\Lambda} : \bar{F} \rightarrow \mathbb{R}^m$, $\bar{\gamma}', \bar{\gamma}'', \bar{\gamma}'_i, \bar{\gamma}''_i : \bar{F} \rightarrow \mathbb{R}$, which are defined via

$$\begin{aligned} \bar{\Lambda}([f'; f'']) &= \Lambda(f') - \Lambda(f''), & \bar{\gamma}'([f'; f'']) &= \gamma(f'), & \bar{\gamma}''([f'; f'']) &= \gamma(f''), \\ \bar{\gamma}'_i([f'; f'']) &= \gamma_i(f'), & \bar{\gamma}''_i([f'; f'']) &= \gamma_i(f''). \end{aligned}$$

In this way, one has $\bar{\gamma}'(\bar{f}) \leq 2e_b \rho_{\bar{\mathcal{K}}}(\bar{f}) + \bar{\gamma}''(\bar{f})$ whenever $\bar{f} \in \ker(\bar{\Lambda})$. Therefore, for all $a_\star \in A$ and all $b_\star \in B$,

$$\inf_{i \in I_{a_\star}} \bar{\gamma}'_i(\bar{f}) \leq 2e_b \rho_{\bar{\mathcal{K}}}(\bar{f}) + \sup_{j \in J_{b_\star}} \bar{\gamma}''_j(\bar{f}) \quad \text{whenever } \bar{f} \in \ker(\bar{\Lambda}).$$

This is the time to invoke Lemma 2, justifying the existence of $c^{(a_\star, b_\star)} \in \mathbb{R}^m$ such that

$$\inf_{i \in I_{a_\star}} \bar{\gamma}'_i(\bar{f}) - \sum_{k=1}^m c_k^{(a_\star, b_\star)} \bar{\lambda}_k(\bar{f}) \leq 2e_b \rho_{\bar{\mathcal{K}}}(\bar{f}) + \sup_{j \in J_{b_\star}} \bar{\gamma}''_j(\bar{f}) \quad \text{for all } \bar{f} \in \bar{F}.$$

It follows that, for any $a_\star \in A$ and $b_\star \in B$, one has

$$(5) \quad \left(\inf_{i \in I_{a_\star}} \gamma_i(f') - \sum_{k=1}^m c_k^{(a_\star, b_\star)} \lambda_k(f') \right) - \left(\sup_{j \in J_{b_\star}} \gamma_j(f'') - \sum_{k=1}^m c_k^{(a_\star, b_\star)} \lambda_k(f'') \right) \leq 2e_b \quad \text{for all } f', f'' \in \mathcal{K}.$$

At this point, for each $a_\star \in A$ and $b_\star \in B$, one defines

$$c_0^{(a_\star, b_\star)} := \inf_{f \in \mathcal{K}} \left(\sup_{j \in J_{b_\star}} \gamma_j(f) - \sum_{k=1}^m c_k^{(a_\star, b_\star)} \lambda_k(f) + e_b \right).$$

For each $a_\star \in A$ and $b_\star \in B$, one immediately obtains that, for all $f \in \mathcal{K}$,

$$\sup_{j \in J_{b_\star}} \gamma_j(f) \geq -e_b + \left(c_0^{(a_\star, b_\star)} + \sum_{k=1}^m c_k^{(a_\star, b_\star)} \lambda_k(f) \right) \geq -e_b + \inf_{b \in B} \left(c_0^{(a_\star, b)} + \sum_{k=1}^m c_k^{(a_\star, b)} \lambda_k(f) \right).$$

Now, taking the infimum over $b_\star \in B$ on the left-hand side yields

$$\gamma(f) \geq -e_b + \inf_{b \in B} \left(c_0^{(a_\star, b)} + \sum_{k=1}^m c_k^{(a_\star, b)} \lambda_k(f) \right) \quad \text{for all } f \in \mathcal{K},$$

and finally taking the supremum over $a_\star \in A$ on the right-hand side and rearranging gives

$$(6) \quad \gamma(f) - \sup_{a \in A} \inf_{b \in B} \left(c_0^{(a, b)} + \sum_{k=1}^m c_k^{(a, b)} \lambda_k(f) \right) \geq -e_b \quad \text{for all } f \in \mathcal{K}.$$

Next, for each $a_\star \in A$ and $b_\star \in B$, one selects $f'' \in \mathcal{K}$ achieving (or coming arbitrarily close to) the infimum defining $c_0^{(a_\star, b_\star)}$, so writing (5) for this f'' and for f' being an arbitrary $f \in \mathcal{K}$ leads to

$$\inf_{i \in I_{a_\star}} \gamma_i(f) - \left(c_0^{(a_\star, b_\star)} + \sum_{k=1}^m c_k^{(a_\star, b_\star)} \lambda_k(f) \right) \leq e_b,$$

from where it follows, by taking the infimum over $b_\star \in B$, that

$$\inf_{i \in I_{a_\star}} \gamma_i(f) \leq e_b + \inf_{b \in B} \left(c_0^{(a_\star, b)} + \sum_{k=1}^m c_k^{(a_\star, b)} \lambda_k(f) \right) \leq e_b + \sup_{a \in A} \inf_{b \in B} \left(c_0^{(a, b)} + \sum_{k=1}^m c_k^{(a, b)} \lambda_k(f) \right).$$

Finally, taking the supremum over $a_\star \in A$ leads to

$$(7) \quad \gamma(f) - \sup_{a \in A} \inf_{b \in B} \left(c_0^{(a,b)} + \sum_{k=1}^m c_k^{(a,b)} \lambda_k(f) \right) \leq e_b \quad \text{for all } f \in \mathcal{K}.$$

Defining the estimation functional $\delta^{\text{opti}} : \mathbb{R}^m \rightarrow \mathbb{R}$ by $\delta^{\text{opti}}(y) = \sup_{a \in A} \inf_{b \in B} (c_0^{(a,b)} + \sum_{k=1}^m c_k^{(a,b)} y_k)$ and combining inequalities (6) and (7) show that

$$e(\delta^{\text{opti}}) = \sup_{f \in \mathcal{K}} |\gamma(f) - \delta^{\text{opti}}(\Lambda(f))| \leq e_b.$$

This proves the optimality of the sup-inf-affine functional δ^{opti} . $\square$

2.3 Handling observation error

In realistic situations, the observations made on the unknown object f are not exact. They typically take the form $\lambda_k(f) + e_k$ for some unknown $e \in \mathbb{R}^m$. This vector can be modeled stochastically, as done in [2] but not considered here, or deterministically via an assumption $e \in \mathcal{E}$, as considered here, where $\mathcal{E}$ is convex set containing the origin. It is folklore to remark that this ‘inaccurate scenario’ formally reduces to the ‘accurate scenario’ by focusing on the compound object (f, e) belonging to the model set $\mathcal{K} \times \mathcal{E}$ and acquired via the linear observations $\tilde{\lambda}_k((f, e)) = \lambda_k(f) + e_k$. For instance, since a sup-linear functional γ acting on f is also a sup-linear functional acting on (f, e) , Theorem 1 will still guarantee that there exists a sup-affine functional which is optimal for the estimation of γ . Here is a more explicit statement for the counterpart of Theorem 3.

Proposition 4. Suppose that the sets $\mathcal{K} \subseteq F$ and $\mathcal{E} \subseteq \mathbb{R}^m$ are convex and contains the origins and that the functional $\gamma : F \rightarrow \mathbb{R}$ can be expressed as

$$\gamma(f) = \sup_{a \in A} \inf_{i \in I_a} \gamma_i(f) = \inf_{b \in B} \sup_{j \in J_b} \gamma_j(f), \quad f \in F,$$

where the $\gamma_i \in F^*$ are linear functionals. Then there exists a functional $\delta^{\text{opti}} : \mathbb{R}^m \rightarrow \mathbb{R}$ of the form

$$\delta^{\text{opti}}(y) = \sup_{a \in A} \inf_{b \in B} \left(c_0^{(a,b)} + \sum_{k=1}^m c_k^{(a,b)} y_k \right), \quad y \in \mathbb{R}^m,$$

which is optimal for the worst-case estimation of γ , in the sense that, for all $\delta : \mathbb{R}^m \rightarrow \mathbb{R}$,

$$\sup_{\substack{f \in \mathcal{K} \\ e \in \mathcal{E}}} |\gamma(f) - \delta^{\text{opti}}(\Lambda(f) + e)| \leq \sup_{\substack{f \in \mathcal{K} \\ e \in \mathcal{E}}} |\gamma(f) - \delta(\Lambda(f) + e)|.$$

The practical construction of the optimal estimation functional is addressed in the next section—thanks to the formal reduction, it will be enough to consider the ‘accurate scenario’ there. A central role shall be played by the support function of the set $\mathcal{K}$, as defined by

$$\|\eta\|_{\mathcal{K}} := \sup_{f \in \mathcal{K}} \eta(f), \quad \eta \in F^*.$$

In the ‘inaccurate scenario’, one should use the support function of the set $\mathcal{K} \times \mathcal{E}$ instead. To fix the ideas, suppose as an example that $\mathcal{E} = \varepsilon B_p^m$ is an origin-centered ball of radius ε in ℓ_p -space for some $p \in [1, \infty]$ with conjugate exponent $p' \in [1, \infty]$. Then, the observation that

$$\left\| \sum_{k=1}^m c_k \tilde{\lambda}_k \right\|_{\mathcal{K} \times \varepsilon B_p^m} = \sup_{(f,e) \in \mathcal{K} \times \varepsilon B_p^m} \sum_{k=1}^m c_k (\lambda_k(f) + e_k) = \left\| \sum_{k=1}^m c_k \lambda_k \right\|_{\mathcal{K}} + \varepsilon \|c\|_{p'}$$

enables the application of the forthcoming Proposition 7, say, to determine the coefficients $c_k^{(a,b)}$ of the optimal estimation functional δ^{opti} .

While an analysis of the dependence of these coefficients on the level ε of observation error is beyond the scope of this article, a rough estimate of the dependence on ε of the minimal worst-case error is not hard to come by. Precisely, with $e_\varepsilon(\delta) := \sup\{|\gamma(f) - \delta(\Lambda f + e)| : f \in \mathcal{K}, e \in \varepsilon B_p^m\}$ denoting the worst-case error of an arbitrary $\delta : \mathbb{R}^m \rightarrow \mathbb{R}$ at level ε , one has

$$\inf_{\delta} e_\varepsilon(\delta) \leq \inf_{\delta} e_\zeta(\delta) \leq \inf_{\delta} e_\varepsilon(\delta) + C_\varepsilon (\zeta - \varepsilon) \quad \text{for any } \zeta \geq \varepsilon \geq 0,$$

where the constant C_ε can be taken as the largest $\ell_{p'}$ -norm of the coefficient vectors $c^{(a,b)} \in \mathbb{R}^m$ appearing in the expression of the optimal estimation functional δ^{opti} at level ε . Indeed, the leftmost inequality follows from the simple fact that $e_\varepsilon(\delta) \leq e_\zeta(\delta)$ for any $\delta : \mathbb{R}^m \rightarrow \mathbb{R}$. As for the rightmost inequality, it relies on a fact easily verifiable from the Lipschitz properties of $\inf$ and $\sup$, namely from $|\delta^{\text{opti}}(y) - \delta^{\text{opti}}(y')| \leq \sup_{a \in A} \sup_{b \in B} \left| \sum_{k=1}^m c_k^{(a,b)} (y_k - y'_k) \right|$, as used in

$$\begin{aligned} \inf_{\delta} e_\zeta(\delta) &\leq e_\zeta(\delta^{\text{opti}}) = \sup_{\substack{f \in \mathcal{K} \\ \|e\|_p \leq \zeta}} |\gamma(f) - \delta^{\text{opti}}(\Lambda(f) + e)| \\ &\leq \sup_{\substack{f \in \mathcal{K} \\ \|e\|_p \leq \zeta}} \left| \gamma(f) - \delta^{\text{opti}}\left(\Lambda(f) + \frac{\varepsilon}{\zeta} e\right) \right| + \sup_{\substack{f \in \mathcal{K} \\ \|e\|_p \leq \zeta}} \left| \delta^{\text{opti}}(\Lambda(f) + e) - \delta^{\text{opti}}\left(\Lambda(f) + \frac{\varepsilon}{\zeta} e\right) \right| \\ &\leq e_\varepsilon(\delta^{\text{opti}}) + \sup_{\|e\|_p \leq \zeta} \sup_{\substack{a \in A \\ b \in B}} \left| \sum_{k=1}^m c_k^{(a,b)} \left(1 - \frac{\varepsilon}{\zeta}\right) e_k \right| = \inf_{\delta} e_\varepsilon(\delta) + \sup_{\substack{a \in A \\ b \in B}} \|c^{(a,b)}\|_{p'} \left(1 - \frac{\varepsilon}{\zeta}\right) \zeta. \end{aligned}$$

This indeed reduces to the desired inequality.

3 Computational Realizations

The purpose of this section is to provide computational recipes for the practical construction of the optimal estimation functional appearing in Theorem 1 and, to a lesser extent, Theorem 3. At first, a rather abstract optimization program outputting the desired coefficient vectors will be presented. Then, for the sake of implementation in a reproducible MATLAB file², two situations of specific interest will be spelled out in separate subsections.

² Accessible from the author’s webpage or at <https://github.com/foucart/COR>.

3.1 Generic convex optimization programs

In most of this subsection and throughout the next two, the nonlinear functional $\gamma : F \rightarrow \mathbb{R}$ denotes a sup-linear functional, i.e., it has the form $\gamma(f) = \sup_{i \in I} \gamma_i(f)$ with the γ_i 's being linear functionals. It was established in Theorem 1 that the functionals $\delta : \mathbb{R}^m \rightarrow \mathbb{R}$ minimizing the estimation error $e(\delta) = \sup\{|\gamma(f) - \delta(\Lambda(f))| : f \in \mathcal{K}\}$ can be sought among sup-affine functionals. As a preliminary, for a fixed such δ , one highlights how $e(\delta)$ can be computed by solving a convex optimization program. The crucial point is that $\nu := \delta \circ \Lambda : F \rightarrow \mathbb{R}$ is itself a sup-affine functional in this case. The statement below features the standard simplex associated with the index set I , defined when I is finite by

$$\mathcal{S}^I := \left\{ \sigma \in \mathbb{R}^I : \sigma_i \geq 0 \text{ for all } i \in I \text{ and } \sum_{i \in I} \sigma_i = 1 \right\}.$$

Proposition 5. Let $\mathcal{K}$ be a convex set. For a sup-linear functional $\gamma : F \rightarrow \mathbb{R}$ and a sup-affine functional $\nu : F \rightarrow \mathbb{R}$ given, for $f \in F$, by

$$\gamma(f) = \sup_{i \in I} \gamma_i(f) \quad \text{and} \quad \nu(f) = \sup_{i \in I} (\nu_i(f) + b_i),$$

with the $\gamma_i, \nu_i \in F^*$ being linear functionals and the $b_i \in \mathbb{R}$ being scalars, one has

$$(8) \quad \sup_{f \in \mathcal{K}} |\gamma(f) - \nu(f)| = \inf_{\substack{e \in \mathbb{R} \\ \sigma^{(i_\star)}, \tau^{(i_\star)} \in \mathcal{S}^I}} e \quad \text{s.to} \quad \begin{cases} \|\gamma_{i_\star} - \sum_{i \in I} \sigma_i^{(i_\star)} \nu_i\|_{\mathcal{K}} \leq e + \sum_{i \in I} \sigma_i^{(i_\star)} b_i, \\ \|\nu_{i_\star} - \sum_{i \in I} \tau_i^{(i_\star)} \gamma_i\|_{\mathcal{K}} \leq e - b_{i_\star}, \end{cases}$$

where the constraints over the scalar e and the simplex-vectors $\sigma^{(i_\star)}, \tau^{(i_\star)}$ are understood to be imposed for all $i_\star \in I$.

Proof. The supremum appearing on the left-hand side of (8) is nothing else than the infimum value of e subject to $|\gamma(f) - \nu(f)| \leq e$ for all $f \in \mathcal{K}$. The latter decouples as the two constraints

$$(9) \quad \gamma(f) - \nu(f) \leq e \quad \text{for all } f \in \mathcal{K} \quad \text{and} \quad \nu(f) - \gamma(f) \leq e \quad \text{for all } f \in \mathcal{K}.$$

In view of $\gamma(f) = \sup_{i \in I} \gamma_i(f)$ and of

$$-\nu(f) = -\sup_{i \in I} (\nu_i(f) + b_i) = \inf_{i \in I} (-\nu_i(f) - b_i) = \inf_{\sigma \in \mathcal{S}^I} \sum_{i \in I} \sigma_i (-\nu_i(f) - b_i),$$

the first constraint in (9) is equivalent to a set of constraints indexed by $i_\star \in I$, namely

$$\gamma_{i_\star}(f) + \inf_{\sigma \in \mathcal{S}^I} \sum_{i \in I} \sigma_i (-\nu_i(f) - b_i) \leq e \quad \text{for all } f \in \mathcal{K}.$$

Each of these individual constraints reads

$$\sup_{f \in \mathcal{K}} \inf_{\sigma \in \mathcal{S}^I} \left(\gamma_{i_\star}(f) + \sum_{i \in I} \sigma_i (-\nu_i(f) - b_i) \right) \leq e.$$

Since the inner expression depends affinely of $f \in \mathcal{K}$, which is a convex set, and also affinely on $\sigma \in \mathcal{S}^I$, which is a convex and compact set, von Neumann minimax theorem legitimizes the exchange of sup and inf, so that the above constraint reads

$$\inf_{\sigma \in \mathcal{S}^I} \sup_{f \in \mathcal{K}} \left(\gamma_{i_\star}(f) + \sum_{i \in I} \sigma_i (-\nu_i(f) - b_i) \right) \leq e.$$

All in all, the first constraint in (9) is equivalent to the existence of $\sigma^{(i_\star)} \in \mathcal{S}^I$, $i_\star \in I$, such that

$$\sup_{f \in \mathcal{K}} \left(\gamma_{i_\star}(f) - \sum_{i \in I} \sigma_i^{(i_\star)} \nu_i(f) \right) \leq e + \sum_{i \in I} \sigma_i^{(i_\star)} b_i,$$

which is tautologically equivalent to the first set of convex constraints indexed by $i_\star$ appearing in (8). Likewise, with very similar details left to the reader, the second constraint in (9) is equivalent to the existence of $\tau^{(i_\star)} \in \mathcal{S}^I$, $i_\star \in I$, obeying the second set of convex constraints indexed by $i_\star$ appearing in (8). Incorporating all the $\sigma^{(i_\star)}$ and $\tau^{(i_\star)}$ as optimization variables, one arrives at the announced convex optimization program. $\square$

Turning to the minimization of the estimation error $e(\delta)$ over all possible functionals $\delta : \mathbb{R}^m \rightarrow \mathbb{R}$, or in fact only over all functionals of the form $\delta_c(y) = \sup_{i \in I} (c_0^{(i)} + \sum_{k=1}^m c_k^{(i)} y_k)$, $y \in \mathbb{R}^m$, one notices that this error becomes

$$e(\delta_c) = \sup_{f \in \mathcal{K}} |\gamma(f) - \nu_c(f)|, \quad \text{with } \nu_c := \delta_c \circ \Lambda : F \rightarrow \mathbb{R} \text{ being a sup-affine functional.}$$

At a fixed $c \in (\mathbb{R} \times \mathbb{R}^m)^I$, Proposition 5 just indicated how to compute $e(\delta_c)$ by solving a convex optimization program. Regrettably, a straightforward minimization over c as well seems out of reach, due to the presence of products $\sigma_i^{(i_\star)} c_k^{(i)}$ of optimization variables generated through the first set of constraints. An alternative route, pursued below, consists in translating the constructive argument of Theorem 1 into a manageable optimization program.

Proposition 6. Given a convex set $\mathcal{K}$ containing the origin, an optimal functional $\delta^{\text{opti}} : \mathbb{R}^m \rightarrow \mathbb{R}$ for the estimation of a sup-linear functional $\gamma(f) = \sup_{i \in I} \gamma_i(f)$, $f \in F$, is obtained as the sup-affine functional $\delta^{\text{opti}}(y) = \sup_{i \in I} (\hat{c}_0^{(i)} + \sum_{k=1}^m \hat{c}_k^{(i)} y_k)$, $y \in \mathbb{R}^m$. The ‘hat’-notation denotes solutions to the convex optimization program

$$(10) \quad \begin{array}{ll} \underset{\substack{e \in \mathbb{R}, e', e'' \in \mathbb{R}^I \\ c^{(i_\star)} \in \mathbb{R}^m \\ \sigma^{(i_\star)} \in \mathcal{S}^I}}{\text{minimize}} & e \quad \text{s.to} \quad \begin{cases} e'_{i_\star} + e''_{i_\star} \leq 2e, \\ \left\| \gamma_{i_\star} - \sum_{k=1}^m c_k^{(i_\star)} \lambda_k \right\|_{\mathcal{K}} \leq e'_{i_\star}, \\ \left\| -\sum_{i \in I} \sigma_i^{(i_\star)} \gamma_i + \sum_{k=1}^m c_k^{(i_\star)} \lambda_k \right\|_{\mathcal{K}} \leq e''_{i_\star}, \end{cases} \end{array}$$

where the constraints on the scalars $e, e'_{i_\star}, e''_{i_\star}$, the vectors $c^{(i_\star)}$, and the simplex-vectors $\sigma^{(i_\star)}$ are understood to be imposed for all $i_\star \in I$. Subsequently, the remaining coefficients $\hat{c}_0^{(i)} \in \mathbb{R}$ are obtained as $\hat{c}_0^{(i_\star)} = \hat{e} - \hat{e}''_{i_\star}$ for all $i_\star \in I$.

Proof. Keeping the proof of Theorem 1 in mind, recall that the key resided in (1) being satisfied for all $i_\star \in I$, i.e.,

$$\sup_{f' \in \mathcal{K}} \left(\gamma_{i_\star}(f') - \sum_{k=1}^m c_k^{(i_\star)} \lambda_k(f') \right) + \sup_{f'' \in \mathcal{K}} \left(-\gamma(f'') + \sum_{k=1}^m c_k^{(i_\star)} \lambda_k(f'') \right) \leq 2e$$

for some $c^{(i_\star)} \in \mathbb{R}^m$ and for $e \in \mathbb{R}$ as small as possible, i.e., $e = e_b$. Thus, introducing slack variables $e'_{i_\star}, e''_{i_\star} \in \mathbb{R}$ bounding the above suprema, the task consists in minimizing e subject to $e'_{i_\star} + e''_{i_\star} \leq 2e$, to $\sup_{f' \in \mathcal{K}} (\gamma_{i_\star}(f') - \sum_{k=1}^m c_k^{(i_\star)} \lambda_k(f')) \leq e'_{i_\star}$, and to $\sup_{f'' \in \mathcal{K}} (-\gamma(f'') + \sum_{k=1}^m c_k^{(i_\star)} \lambda_k(f'')) \leq e''_{i_\star}$ for all $i_\star \in I$. The first of these constraints are exactly the first constraints in (10), while the second ones reduce to the second constraints in (10) simply by the definition of the support function. Considering now the third and last constraints, in view of

$$-\gamma(f'') = -\sup_{i \in I} \gamma_i(f'') = \inf_{i \in I} (-\gamma_i(f'')) = \inf_{\sigma \in \mathcal{S}^I} \sum_{i \in I} \sigma_i (-\gamma_i(f'')),$$

one obtains

$$\begin{aligned} \sup_{f'' \in \mathcal{K}} \left(-\gamma(f'') + \sum_{k=1}^m c_k^{(i_\star)} \lambda_k(f'') \right) &= \sup_{f'' \in \mathcal{K}} \inf_{\sigma \in \mathcal{S}^I} \left(-\sum_{i \in I} \sigma_i \gamma_i(f'') + \sum_{k=1}^m c_k^{(i_\star)} \lambda_k(f'') \right) \\ &= \inf_{\sigma \in \mathcal{S}^I} \sup_{f'' \in \mathcal{K}} \left(-\sum_{i \in I} \sigma_i \gamma_i(f'') + \sum_{k=1}^m c_k^{(i_\star)} \lambda_k(f'') \right), \end{aligned}$$

where the exchange of sup and inf relied on von Neumann minimax theorem, in the same way as for the proof of Proposition 5. As a result, the constraints can be rephrased as the existence of $\sigma^{(i_\star)} \in \mathcal{S}^I$ such that

$$\left\| -\sum_{i \in I} \sigma_i^{(i_\star)} \gamma_i + \sum_{k=1}^m c_k^{(i_\star)} \lambda_k \right\|_{\mathcal{K}} \leq e''_{i_\star}.$$

Incorporating all the $\sigma^{(i_\star)} \in \mathcal{S}^I$ as optimization variables, one arrives at the third constraints in (10). All in all, it has been justified that the convex optimization program (10) outputs e_b (i.e., the minimal error $e(\delta)$ over all possible estimation functionals δ) as $\widehat{e}$ and the coefficient vectors making the linear part of δ^{opti} as the $\widehat{c}^{(i_\star)} \in \mathbb{R}^m$. Finally, the announced expression for the remaining coefficients $\widehat{c}_0^{(i_\star)} \in \mathbb{R}$ follows from their very definition (2), written as

$$\widehat{c}_0^{(i_\star)} = -\sup_{f'' \in \mathcal{K}} \left(-\gamma(f'') + \sum_{k=1}^m \widehat{c}_k^{(i_\star)} \lambda_k(f'') \right) + e_b,$$

and from the identification of the above supremum as $\widehat{e}''_{i_\star}$ and of e_b as $\widehat{e}$. $\square$

When estimating a mixed sup-inf-linear functional, it is as also possible to translate the proof of Theorem 3 into a practical recipe for the construction of the optimal sup-inf-affine functional. The key resides again in fulfilling the inequalities (5), which are treated as above. More precisely,

one proceeds in exactly the same way (save for the notation) to recast the bound on the second summand in the left-hand side of (5) by a slack variable e''_{a_*, b_*} , and the argument will also apply when recasting a bound on the first summand by a slack variable e'_{a_*, b_*} . With details left out, one arrives at the convex optimization program displayed below.

Proposition 7. Under the setting of Theorem 3, the optimal estimation functional $\delta^{\text{opti}} : \mathbb{R}^m \rightarrow \mathbb{R}$ has the sup-inf-affine form $\delta^{\text{opti}}(y) = \sup_{a \in A} \inf_{b \in B} (\hat{c}_0^{(a,b)} + \sum_{k=1}^m \hat{c}_k^{(a,b)} y_k)$, $y \in \mathbb{R}^m$. The ‘hat’-notation denotes solutions to the convex optimization program

$$\begin{aligned} \underset{\substack{e \in \mathbb{R}, e', e'' \in \mathbb{R}^{A \times B} \\ c^{(a_*, b_*)} \in \mathbb{R}^m \\ \sigma^{(a_*, b_*)} \in \mathcal{S}^{I_{a_*}} \\ \tau^{(a_*, b_*)} \in \mathcal{S}^{I_{b_*}}} }{\text{minimize}} \quad & e \quad \text{s.to} \quad \begin{cases} e'_{a_*, b_*} + e''_{a_*, b_*} \leq 2e, \\ \left\| \sum_{i \in I_{a_*}} \sigma_i^{(a_*, b_*)} \gamma_i - \sum_{k=1}^m c_k^{(a_*, b_*)} \lambda_k \right\|_{\mathcal{K}} \leq e'_{a_*, b_*}, \\ \left\| -\sum_{j \in J_{b_*}} \tau_j^{(a_*, b_*)} \gamma_j + \sum_{k=1}^m c_k^{(a_*, b_*)} \lambda_k \right\|_{\mathcal{K}} \leq e''_{a_*, b_*}, \end{cases} \end{aligned}$$

where the constraints on the scalars $e, e'_{a_*, b_*}, e''_{a_*, b_*}$, the vectors $c^{(a_*, b_*)}$, and the simplex-vectors $\sigma^{(a_*, b_*)}, \tau^{(a_*, b_*)}$ are understood to be imposed for all $a_* \in A, b_* \in B$. Subsequently, the coefficients $\hat{c}_0^{(a,b)} \in \mathbb{R}$ are obtained as $\hat{c}_0^{(a,b)} = \hat{e} - \hat{e}''_{a_*, b_*}$ for all $a_* \in A, b_* \in B$.

Arguably, Propositions 6 and 7 are still abstract statements. Whether they really translate into practical constructions depends on the specific model set $\mathcal{K}$ and its amenability to numerical computations. Concerning Proposition 7, there is the added issue of the number of constraints, which is $|A| \times |B|$. In the example of the ℓ th largest value among d linear functionals, one has

$$|A| = |\{I \in [1:d] : |I| = \ell\}| = \binom{d}{\ell} \quad \text{and} \quad |B| = |\{J \in [1:d] : |J| = d+1-\ell\}| = \binom{d}{\ell-1},$$

so $|A| \times |B|$ behaves exponentially in d unless ℓ or $d - \ell$ is a small constant. For this reason, one concentrates only on the estimation of the maximum of linear functionals when expounding on the two relevant examples presented in the next subsections.

3.2 Polytopal models in ℓ_∞ -spaces

Working in $F = \ell_\infty^N$, the convex model set containing the origin (but not necessarily symmetric) considered here is the polytope

$$\mathcal{K}^{\text{poly}} := \{f \in \ell_\infty^N : \langle a_\ell, f \rangle \leq 1 \text{ for all } \ell \in L\}$$

described through prescribed vectors $a_\ell \in \ell_1^N$ indexed by a finite set L . Using duality in linear programming, it is routine to verify that the support function of $\mathcal{K}^{\text{poly}}$, evaluated at a linear functional η defined for $f \in \ell_\infty^N$ by $\eta(f) = \langle g, f \rangle$, can be expressed as

$$\begin{aligned} \|\eta\|_{\mathcal{K}^{\text{poly}}} &= \sup_{f \in \ell_\infty^N} \langle g, f \rangle \quad \text{s.to} \quad \langle a_\ell, f \rangle \leq 1 \text{ for all } \ell \in L \\ &= \inf_{s \in \mathbb{R}^L} \sum_{\ell \in L} s_\ell \quad \text{s.to} \quad s \geq 0 \text{ and } \sum_{\ell \in L} s_\ell a_\ell = g. \end{aligned}$$

Therefore, an inequality constraint $\|\eta\|_{\mathcal{K}^{\text{poly}}} \leq \kappa$ found as a constraint in an optimization program can be rephrased as the existence of some $s \in \mathbb{R}_+^L$ such that $\sum_{\ell \in L} s_\ell \leq \kappa$ and $\sum_{\ell \in L} s_\ell a_\ell = g$, and subsequently s can be incorporated in the optimization variables. For instance, with $u_k \in \ell_1^N$ representing the linear functional λ_k and $w_i \in \ell_1^N$ representing the linear functional γ_i , so that

$$\lambda_k(f) = \langle u_k, f \rangle \quad \text{and} \quad \gamma_i(f) = \langle w_i, f \rangle \quad \text{for all } f \in \ell_\infty^N,$$

the convex optimization program from Proposition 6 becomes:

Linear program for the optimal estimation of $\gamma = \sup_{i \in I} \gamma_i$ with model $\mathcal{K}^{\text{poly}}$.

$$\begin{aligned} & \underset{\substack{e \in \mathbb{R}, e', e'' \in \mathbb{R}^I \\ c^{(i_\star)} \in \mathbb{R}^m \\ \sigma^{(i_\star)} \in \mathcal{S}^I \\ s^{(i_\star)}, t^{(i_\star)} \in \mathbb{R}_+^L}}{\text{minimize}} & e \quad \text{s.to} & \left\{ \begin{array}{l} e'_{i_\star} + e''_{i_\star} \leq 2e, \sum_{\ell \in L} s_\ell^{(i_\star)} \leq e'_{i_\star}, \sum_{\ell \in L} t_\ell^{(i_\star)} \leq e''_{i_\star} \\ \sum_{\ell \in L} s_\ell^{(i_\star)} a_\ell = w_{i_\star} - \sum_{k=1}^m c_k^{(i_\star)} u_k \\ \sum_{\ell \in L} t_\ell^{(i_\star)} a_\ell = -\sum_i \sigma_i^{(i_\star)} w_i + \sum_{k=1}^m c_k^{(i_\star)} u_k \end{array} \right\} \quad \forall i_\star \in I. \end{aligned}$$

After solving this linear program, one can construct an optimal estimation functional δ^{opti} which has the generic sup-affine form

$$(11) \quad \delta_{b,z}(y) = \sup_{i \in I} (b_i + \langle z_i, y \rangle), \quad y \in \mathbb{R}^m.$$

For good measure, one can verify (as done in the reproducible) that the minimal value of e found above agrees with $e(\delta^{\text{opti}})$ when the latter is computed by solving the convex program from Proposition 5. In the present case, it becomes:

Linear program to compute $e(\delta_{b,z})$ for a fixed sup-affine functional with model $\mathcal{K}^{\text{poly}}$.

$$(12) \quad \underset{\substack{e \in \mathbb{R} \\ \sigma^{(i_\star)}, \tau^{(i_\star)} \in \mathcal{S}^I \\ s^{(i_\star)}, t^{(i_\star)} \in \mathbb{R}_+^L}}{\text{minimize}} \quad e \quad \text{s.to} \quad \left\{ \begin{array}{l} \sum_{\ell \in L} s_\ell^{(i_\star)} \leq e + \sum_{i \in I} \sigma_i^{(i_\star)} b_i, \sum_{\ell \in L} t_\ell^{(i_\star)} \leq e - b_{i_\star} \\ \sum_{\ell \in L} s_\ell^{(i_\star)} a_\ell = w_{i_\star} - \sum_{i \in I} \sigma_i^{(i_\star)} \Lambda^* z_i \\ \sum_{\ell \in L} t_\ell^{(i_\star)} a_\ell = \Lambda^* z_{i_\star} - \sum_{i \in I} \tau_i^{(i_\star)} w_i \end{array} \right\} \quad \forall i_\star \in I.$$

Remark. In $F = \ell_\infty^N$ and with a convex model, it is somewhat known that there exists an affine map $\Delta^{\text{opti}} : \mathbb{R}^m \rightarrow \ell_\infty^N$ which is genuinely optimal for the estimation of Id_F , i.e., for the full approximation problem. This is essentially Sukharev's result applied componentwise. Leaving out the details, this optimal map has the form

$$\Delta^{\text{opti}} : y \in \mathbb{R}^m \mapsto \left[\hat{c}^{(0)} + \sum_{k=1}^m y_k \hat{c}^{(k)} \right] \in \ell_\infty^N,$$

where the vectors $\hat{c}^{(0)}, \hat{c}^{(1)}, \dots, \hat{c}^{(m)} \in \ell_\infty^N$ are solutions to the following linear program (involving

the standard basis $(v_1, \dots, v_N)$ for $\mathbb{R}^N$:

$$\begin{aligned} & \underset{\substack{e \in \mathbb{R} \\ c^{(0)}, c^{(k)} \in \mathbb{R}^N \\ s^{(j)}, t^{(j)} \in \mathbb{R}_+^L}}{\text{minimize}} \quad e \quad \text{s.to} \quad \left\{ \begin{array}{l} \sum_{\ell \in L} s_\ell^{(j)} \leq e + c_j^{(0)}, \quad \sum_{\ell \in L} t_\ell^{(j)} \leq e - c_j^{(0)} \\ \sum_{\ell \in L} s_\ell^{(j)} a_\ell = v_j - \sum_{k=1}^m c_j^{(k)} u_k \\ \sum_{\ell \in L} t_\ell^{(j)} a_\ell = -v_j + \sum_{k=1}^m c_j^{(k)} u_k \end{array} \right\} \quad \forall j \in [1:N] \end{aligned}$$

As an initially guess, a natural candidate for the estimation of $\gamma = \sup_{i \in I} \gamma_i$ could have been the ‘plug-in’ functional $\delta^{\text{plug}} = \gamma \circ \Delta^{\text{opti}} : \mathbb{R}^m \rightarrow \mathbb{R}$. Since it has the generic sup-affine form (11) with $b_i = \gamma_i(\hat{c}^{(0)})$ and $(z_i)_k = \gamma_i(\hat{c}^{(k)})$, the estimation error $e(\delta^{\text{plug}})$ of this ‘plug-in’ functional can be computed by solving the linear program (12). Numerical experiments available in the reproducible file confirm that this guess is not optimal, i.e., that $e(\delta^{\text{plug}})$ is larger than $e(\delta^{\text{opti}})$.

3.3 Approximability models in Hilbert spaces

Working now in a Hilbert space $F = H$, the convex model set containing the origin (but not necessarily symmetric) considered here is the set of elements that are approximated with prescribed accuracy $\varepsilon > 0$ by an affine subspace $g + V$, where V is an n -dimensional linear subspace of H and $g \in H \setminus V$ is a fixed element satisfying $\text{dist}(g, V) < \varepsilon$. More precisely, one considers

$$\mathcal{K}^{\text{appr}} := \{f \in H : \text{dist}(f, g + V) \leq \varepsilon\} = \{f \in H : \|P_{V^\perp}(f - g)\| \leq \varepsilon\},$$

where $P_{V^\perp}$ denotes the orthogonal projector onto the orthogonal complement of V . Motivated by Optimal Recovery questions, such approximability sets (with $g = 0$) were introduced in [1]. Based on an observation made (with $g = 0$) in [5, Subsection 3.3], their support function evaluated at a linear functional η can be expressed as

$$\|\eta\|_{\mathcal{K}^{\text{appr}}} = \eta(g) + \begin{cases} \varepsilon \|\eta\|_* & \text{if } \eta|_V = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Consequently, still using the notation $u_k \in H$ for the Riesz representer of the linear functional λ_k and $w_i \in H$ for the Riesz representers of the linear functional γ_i —in a reproducing kernel Hilbert space with kernel $\mathbf{k}$, note that $w_i = \mathbf{k}(\cdot, x^{(i)})$ if γ_i is the evaluation at a point $x^{(i)}$ —the convex optimization program from Proposition 6 becomes:

Second-order cone program for the optimal estimation of $\gamma = \sup_{i \in I} \gamma_i$ with model $\mathcal{K}^{\text{appr}}$.

$$\begin{aligned} & \underset{\substack{e \in \mathbb{R}, e', e'' \in \mathbb{R}^I \\ c^{(i_\star)} \in \mathbb{R}^m \\ \sigma^{(i_\star)} \in \mathcal{S}^I}}{\text{minimize}} \quad e \\ & \text{s.to} \quad \left\{ \begin{array}{l} e'_{i_\star} + e''_{i_\star} \leq 2e \\ w_{i_\star} - \sum_{k=1}^m c_k^{(i_\star)} u_k \in V^\perp, \quad -\sum_{i \in I} \tau_i^{(i_\star)} w_i + \sum_{k=1}^m c_k^{(i_\star)} u_k \in V^\perp \\ \varepsilon \|w_{i_\star} - \sum_{k=1}^m c_k^{(i_\star)} u_k\| \leq e'_{i_\star} - \langle w_{i_\star} - \sum_{k=1}^m c_k^{(i_\star)} u_k, g \rangle \\ \varepsilon \|\sigma^{(i_\star)} - \sum_{i \in I} \sigma_i^{(i_\star)} w_i + \sum_{k=1}^m c_k^{(i_\star)} u_k\| \leq e''_{i_\star} - \langle -\sum_{i \in I} \sigma_i^{(i_\star)} w_i + \sum_{k=1}^m c_k^{(i_\star)} u_k, g \rangle \end{array} \right\} \quad \forall i_\star \in I. \end{aligned}$$

As in the previous subsection, one can construct an optimal estimation functional δ^{opti} of the generic sup-affine form (11) after solving the above second-order cone program. And again, for good measure, one can verify (as done in the reproducible) that the minimal value of e just found agrees with $e(\delta^{\text{opti}})$ when the latter is computed by solving the convex program from Proposition 5. In the present case, it becomes:

Second-order cone program to compute $e(\delta_{b,z})$ for a fixed sup-affine functional with model $\mathcal{K}^{\text{appr}}$.

$$(13) \quad \begin{aligned} & \underset{\substack{e \in \mathbb{R} \\ \sigma^{(i_*)}, \tau^{(i_*)} \in \mathcal{S}^I}}{\text{minimize}} && e \\ & \text{s.to} && \left\{ \begin{aligned} & w_{i_*} - \sum_{i \in I} \sigma_i^{(i_*)} \Lambda^* z_i \in V^\perp, \quad \Lambda^* z_{i_*} - \sum_{i \in I} \tau_i^{(i_*)} w_i \in V^\perp \\ & \varepsilon \|w_{i_*} - \sum_{i \in I} \sigma_i^{(i_*)} \Lambda^* z_i\| \leq e + \sum_{i \in I} \sigma_i^{(i_*)} b_i - \langle w_{i_*} - \sum_{i \in I} \sigma_i^{(i_*)} \Lambda^* z_i, g \rangle \\ & \varepsilon \|\Lambda^* z_{i_*} - \sum_{i \in I} \tau_i^{(i_*)} w_i\| \leq e - b_{i_*} - \langle \Lambda^* z_{i_*} - \sum_{i \in I} \tau_i^{(i_*)} w_i, g \rangle \end{aligned} \right\} \quad \forall i_* \in I. \end{aligned}$$

Remark. In $F = H$ and with the model set $\mathcal{K}^{\text{appr}}$, there also exists an affine map $\Delta^{\text{opti}} : \mathbb{R}^m \rightarrow H$ which is genuinely optimal for the estimation of Id_F . Specifically, it was established in [1] that,

$$\Delta^{\text{cheb}} : y \in \mathbb{R}^m \mapsto \left[\underset{f \in H}{\text{argmin}} \|P_{V^\perp} f\| \quad \text{s.to } \Lambda(f) = y \right] \in H$$

is an optimal³ recovery map in the case $g = 0$. It then easily follows that

$$\Delta^{\text{opti}} : y \in \mathbb{R}^m \mapsto g + \Delta^{\text{cheb}}(y - \Lambda(g)) \in H$$

is an optimal recovery map with $g \neq 0$. It is also affine, since Δ^{cheb} is a linear map which admits the expression (not immediate from above, see [3, Section 10.2] for details)

$$\Delta^{\text{cheb}}(y) = \sum_{k=1}^m a_k u_k + \sum_{\ell=1}^n b_\ell v_\ell, \quad a = [I_m - C(C^\top C)^{-1}C^\top]y, \quad b = [(C^\top C)^{-1}C^\top]y,$$

where the u_k are assumed to be orthonormal, the v_ℓ form a basis for V , and $C \in \mathbb{R}^{m \times n}$ is the cross-gramian matrix with entries $\langle u_k, v_\ell \rangle$. Given the above, an initial candidate for the estimation of $\gamma = \sup_{i \in I} \gamma_i$ is the ‘plug-in’ functional $\delta^{\text{plug}} = \gamma \circ \Delta^{\text{opt}} : \mathbb{R}^m \rightarrow \mathbb{R}$. The latter has the generic sup-affine form (11) with $b_i = \langle w_i, g - \Delta^{\text{cheb}} \Lambda g \rangle$ and $z_i = (\Delta^{\text{cheb}})^* w_i$, so the estimation error $e(\delta^{\text{plug}})$ of this ‘plug-in’ functional can be computed by solving the second-order cone program (13). Numerical experiments available in the reproducible file hold a surprise: on all the examples tested, the ‘plug-in’ functional was optimal, i.e., $e(\delta^{\text{plug}})$ was equal to $e(\delta^{\text{opti}})$. In Hilbert spaces, for model sets of the form $\{f \in H : \|Tf\| \leq 1\}$ with T not necessarily equal to $(1/\varepsilon)P_{V^\perp}$, it was known (see [6, Lemma 6]) that ‘plug-in’ maps are indeed optimal for the estimation of linear maps, but the indication that this phenomenon could remain true for the estimation of some nonlinear maps is quite striking.

³In fact, it was even shown that $\Delta^{\text{cheb}}(y)$ is the Chebyshev center for the set $\mathcal{K}^{\text{appr}} \cap \Lambda^{-1}(\{y\})$ —i.e., the center of the smallest ball containing the set—for any $y \in \Lambda(\mathcal{K}^{\text{appr}})$, which is stronger than the notion of optimality examined so far.

References

- [1] Binev, P., Cohen, A., Dahmen, W., DeVore, R., Petrova, G., and Wojtaszczyk, P. (2017). *Data assimilation in reduced modeling*. SIAM/ASA Journal on Uncertainty Quantification, 5(1), 1–29.
- [2] Donoho, D. L. (1994). *Statistical estimation and optimal recovery*. The Annals of Statistics, 22(1), 238–270.
- [3] Foucart, S. (2022). *Mathematical Pictures at a Data Science Exhibition*. Cambridge University Press.
- [4] Foucart, S. (2026). *Learning the maximum of a Hölder function from inexact data*. Proceedings of the American Mathematical Society, 54(3), 1077–1091.
- [5] Foucart, S. (2026). *Optimal prediction of vector-valued functions from point samples*. Journal of Complexity, 92, 101981.
- [6] Foucart, S. and Hengartner, N. (2025). *Worst-case learning under a multifidelity model*. SIAM/ASA Journal on Uncertainty Quantification, 13(1), 171–194.
- [7] Krieg, D. and Ullrich, M. (2026). *Approximation of functions: optimal sampling and complexity*. Acta Numerica, 35, 273–457.
- [8] Micchelli, C. A. and Rivlin, T. J. (1985). *Lectures on optimal recovery*. In: Numerical Analysis Lancaster 1984, 21–93. Springer.
- [9] Novak, E. (2006). *Deterministic and Stochastic Error Bounds in Numerical Analysis*. Springer.
- [10] Novak, E. and Woźniakowski, H. (2010). *Tractability of Multivariate Problems (Vol. 1)*. European Mathematical Society.
- [11] Rudin, W. (1991). *Functional Analysis*. Second edition. McGraw-Hill.
- [12] Sukharev, A. G. (1986). *On the existence of optimal affine methods for approximating linear functionals*. Journal of Complexity, 2(4), 317–322.
- [13] Traub, J. F., Wasilkowski, G. W., and Woźniakowski, H. (1988). *Information-Based Complexity*. Academic Press.
- [14] Wasilkowski, G. W. (1984). *Some nonlinear problems are as easy as the approximation problem*. Computers & Mathematics with Applications, 10(4-5), 351–363.