S-Procedure Relaxation: a Case of Exactness Involving Chebyshev Centers

Simon Foucart and Chunyang Liao

Abstract Optimal recovery is a mathematical framework for learning functions from observational data by adopting a worst-case perspective tied to model assumptions on the functions to be learned. Working in a finite-dimensional Hilbert space, we consider model assumptions based on approximability and observation inaccuracies modeled as additive errors bounded in ℓ_2. We focus on the local recovery problem, which amounts to the determination of Chebyshev centers. Earlier work by Beck and Eldar presented a semidefinite recipe for the determination of Chebyshev centers. The result was valid in the complex setting only, but not necessarily in the real setting, since it relied on the S-procedure with two quadratic constraints, which offers a tight relaxation only in the complex setting. Our contribution consists in proving that this semidefinite recipe is exact in the real setting, too, at least in the particular instance where the quadratic constraints involve orthogonal projectors. Our argument exploits a previous work of ours, where exact Chebyshev centers were obtained in a different way. We conclude by statting some open questions and by commenting on other recent results in optimal recovery.

1 Rundown on Optimal Recovery

The field of optimal recovery, arguably shaped by the influence of Kolmogorov [1] and deeply rooted in Approximation Theory [2], is experiencing a revival thanks to newly available optimization tools. Chosen as one of the core topics first investigated at CAMDA—the Center for Approximation and Mathematical Data Analytics—

Simon Foucart
Texas A&M University, College Station
e-mail: foucart@tamu.edu

Chunyang Liao
University of California, Los Angeles
e-mail: liaochunyang@math.ucla.edu

1

optimal recovery can be understood as a nonstatistical learning theory. Indeed, a function f acquired through point evaluations $y_i = f(x^{(i)})$, $i = 1, \ldots, m$, needs to be learned—or recovered, in the parlance preferred here. But one does not abide by the postulate behind statistical learning theory, stipulating that the $x^{(i)}$'s are independent realizations of a random variable. Thus, the performance of a learning/recovery procedure cannot be assessed in an average case and one opts for an assessment focusing on the worst case, relative to a model expressing some prior scientific knowledge about f.

To be more precise, the task at hand consists in recovering an element f from a Banach space F. It is typically thought of as a space of functions, although it does not have to be. Rather than recovering f in full, it can be more relevant to recover a quantity of interest $Q(f)$, where $Q : F \to Z$ represents a linear map in this article. The element f is only available through partial information, specifically:

- some *a priori* information conveyed by a modeling assumption taking the form

$$f \in \mathcal{K},$$

  where the so-called model set $\mathcal{K} \subseteq F$ reflects an educated guess about realistic objects to be recovered;
- some *a posteriori* information obtained through observational data of the form

$$y_i = \lambda_i(f), \qquad i = 1, \ldots, m,$$

  for some linear functionals $\lambda_1, \ldots, \lambda_m : F \to \mathbb{R}$. More concisely, one writes $y = \Lambda f \in \mathbb{R}^m$, where the so-called observation map $\Lambda : F \to \mathbb{R}^m$ is a linear map.

A recovery procedure is a process, perhaps partially cognizant of the model set $\mathcal{K}$, that takes in the observational data in $\mathbb{R}^m$ and returns an estimation to $Q(f)$ in Z. In other words, it is nothing but a map $\Delta : \mathbb{R}^m \to Z$. Its recovery performance can evidently be quantified by $\|Q(f) - \Delta(y)\|_Z = \|Q(f) - \Delta(\Lambda f)\|_Z$ for a fixed $f \in \mathcal{K}$ satisfying $\Lambda f = y$. However, f being unknown, one takes a worst-case perspective over all consistent f's, leading to a recovery performance quantified by

- the local worst-case error, at a fixed $y \in \mathbb{R}^m$, defined as

$$\text{lwce}(\Delta, y) = \sup_{\substack{f \in \mathcal{K} \\ \Lambda f = y}} \|Q(f) - \Delta(y)\|_Z; \tag{1}$$

- the global worst-case error defined as

$$\text{gwce}(\Delta) = \sup_{f \in \mathcal{K}} \|Q(f) - \Delta(\Lambda f)\|_Z. \tag{2}$$

A locally, respectively globally, optimal recovery map $\Delta^{\text{opt}} : \mathbb{R}^m \to Z$ is a map $\Delta : \mathbb{R}^m \to Z$ that minimizes $\text{lwce}(\Delta, y)$ at every $y \in \mathbb{R}^m$, respectively $\text{gwce}(\Delta)$. Noticing that $\text{gwce}(\Delta) = \sup\{\text{lwce}(\Delta, y), y \in \Lambda(\mathcal{K})\}$, one realizes that a locally optimal recovery map is automatically globally optimal. This somewhat makes the

global setting 'easier' than the local setting, conceivably explaining the prevalence of the latter in the standard theory of optimal recovery. There, a coveted result often consists of the assertion that there exists an optimal recovery map which is linear—of course, one strives to construct it! In the global setting, this existence result classically holds when the model set $\mathcal{K}$ is symmetric and convex and Q is a linear functional (see [3, Theorem 4.7] or [4, Theorem 9.3]) and when the model set is a centered hyperellipsoid in a Hilbert space (see [3, Theorem 4.11] or [4, Theorem 9.4]). In the latter situation, the optimal recovery map, dubbed spline algorithm, is also locally optimal. There are other situations where global optimality via linear maps stands, for instance, in the space $C(\mathcal{X})$ of continuous functions on a compact space $\mathcal{X}$ relative to the model set

$$\mathcal{K}_{\mathcal{V}} = \{f \in C(\mathcal{X}) : \mathrm{dist}_{C(\mathcal{X})}(f, \mathcal{V}) \leq \varepsilon\} \tag{3}$$

subordinate to a linear subspace $\mathcal{V}$ of $C(\mathcal{X})$ and an approximability parameter $\varepsilon \geq 0$. More details will be given in Section 6.

2 Our Contribution: Local Optimality from Inaccurate Data

From now on, we leave the global setting behind and tackle the harder local setting, starting by highlighting its geometric interpretation. Namely, considering a locally optimal recovery map $\Delta^{\mathrm{opt}} : \mathbb{R}^m \to Z$ and a fixed $y \in \mathbb{R}^m$, since $\Delta^{\mathrm{opt}}(y) \in Z$ minimizes $\sup\{\|Q(f) - z\|_Z : f \in \mathcal{K}, \Lambda f = y\}$, one can write, almost tautologically, that

$$\Delta^{\mathrm{opt}}(y) \in \underset{z \in Z, r \geq 0}{\mathrm{argmin}} \quad r \qquad \text{s.to} \quad \|Q(f) - z\|_Z \leq r \text{ whenever } f \in \mathcal{K} \text{ and } \Lambda f = y.$$

This shows that $\Delta^{\mathrm{opt}}(y)$ is a center of a smallest-radius ball containing $Q(\mathcal{K}_y)$, where $\mathcal{K}_y := \{f \in F : f \in \mathcal{K} \text{ and } \Lambda f = y\}$. It is said that $\Delta^{\mathrm{opt}}(y)$ is a Chebyshev center for the set $Q(\mathcal{K}_y)$, often eluding to mention the norm on Z.

But the above context is not quite where our the current investigations take place. Indeed, the discussion so far assumed that the observational data were accurate. In realistic situations, they are contaminated by additive noise, so that $y_i = \lambda_i(f) + e_i$, $i = 1, \ldots, m$. In short, we write $y = \Lambda f + e$ for some error vector $e \in \mathbb{R}^m$. We shall model this vector deterministically through $e \in \mathcal{E}$ for a so-called uncertainty set $\mathcal{E} \subseteq \mathbb{R}^m$. Thus, the *a posteriori* information now takes the form

$$y - \Lambda f \in \mathcal{E}.$$

This leads to an updated notion of local worst-case error, as defined by

$$\mathrm{lwce}(\Delta, y) = \sup_{\substack{f \in \mathcal{K} \\ y - \Lambda f \in \mathcal{E}}} \|Q(f) - \Delta(y)\|_Z.$$

Our objective of finding an optimal recovery map $\Delta^{\mathrm{opt}} : \mathbb{R}^m \to Z$ now becomes the determination, for each $y \in \mathbb{R}^m$, of a solution $\Delta^{\mathrm{opt}}(y)$ to the optimization program

$$\operatorname*{minimize}_{z \in Z} \quad \sup_{\substack{f \in \mathcal{K} \\ y - \Lambda f \in \mathcal{E}}} \|Q(f) - z\|_Z.$$

As before, one interprets geometrically $\Delta^{\mathrm{opt}}(y)$ as a Chebyshev center for the set $Q(\mathcal{K}_{y,\mathcal{E}})$, where $\mathcal{K}_{y,\mathcal{E}} = \{f \in F : f \in \mathcal{K} \text{ and } y - \Lambda f \in \mathcal{E}\}$, i.e., as a center for a smallest-radius ball containing $Q(\mathcal{K}_{y,\mathcal{E}})$.

To achieve our objective—at least partially—we place ourselves in a Hilbert framework, i.e., we assume from now on that F is a finite-dimensional Hilbert space, hence it is denoted by H instead of a generic F. Let us state our first contribution, before placing it in the context of the current knowledge. Note that we can safely talk about *the* Chebyshev center in this statement, as it was known as early as [5] that a bounded set in a uniformly convex Banach space possesses a unique Chebyshev center.

Theorem 1 *In a finite-dimensional Hilbert space H, consider a model set $\mathcal{K}$ and an uncertainty set $\mathcal{E}$ given by*

$$\mathcal{K} = \{f \in H : \|Pf\|_H \le \varepsilon\} \qquad and \qquad \mathcal{E} = \{e \in \mathbb{R}^m : \|e\|_2 \le \varepsilon\}.$$

If P is an orthogonal projection, if $\Lambda\Lambda^* = \mathrm{Id}_m$, and if $\ker(P) \cap \ker(\Lambda) = \{0\}$, then, for any $y \in \mathbb{R}^m$, the Chebyshev radius of $\mathcal{K}_{y,\mathcal{E}} = \{f \in H : f \in \mathcal{K} \text{ and } y - \Lambda f \in \mathcal{E}\}$ is equal to the optimal value of the semidefinite program

$$\operatorname*{minimize}_{\substack{c,d \ge 0 \\ t \in \mathbb{R}}} \quad c\varepsilon^2 + d\eta^2 - d\|y\|_2^2 + t \qquad s.to \quad cP + d\Lambda^*\Lambda \ge \mathrm{Id}_H, \tag{4}$$

$$and \quad \left[\begin{array}{c|c} cP + d\Lambda^*\Lambda & d\Lambda^*y \\ \hline dy^*\Lambda & t \end{array} \right] \ge 0.$$

Moreover, the Chebyshev center is the solution to a regularization program with specified parameters, namely it is given by

$$f_{\widetilde{c},\widetilde{d}} := \operatorname*{argmin}_{f \in H} \ \widetilde{c}\|Pf\|_H^2 + \widetilde{d}\|y - \Lambda f\|_2^2, \tag{5}$$

where $\widetilde{c}, \widetilde{d} \ge 0$ solve the semidefinite program (4).

For the sake of the following discussion, we select the model and uncertainty sets as arbitrary hyperellipsoids. Specifically, with R, S representing linear maps from H into other Hilbert spaces (all norms now being written as $\| \cdot \|$ for ease of notation) and with ε, η representing positive parameters, we consider

$$\mathcal{K} = \{f \in H : \|Rf\| \le \varepsilon\} \subseteq H, \tag{6}$$

$$\mathcal{E} = \{e \in \mathbb{R}^m : \|Se\| \le \eta\} \subseteq \mathbb{R}^m. \tag{7}$$

In this case, Micchelli and Melkman [6, 7] already observed that some regularization (with unspecified parameters) provides linear recovery maps that are optimal, albeit globally. Locally, a similar conclusion was derived by Beck and Eldar [8] for the full estimation problem, i.e., when $Q = \mathrm{Id}_H$. In other words, they—almost—established Theorem 1 in the more general situation of the model and uncertainty sets (6)-(7). There is a subtlety, and an important one: technically, the Chebyshev center was determined in the complex setting only, but not necessarily in the real setting. This incongruity occurs because the main tool used in the argument, i.e., the S-procedure with two constraints, is only exact in the complex setting, see next section for details. In the real setting, the S-procedure is only a relaxation, merely leading to an overestimation of the Chebyshev radius rather than to the genuine Chebyshev radius. Our main contribution in this article therefore consists in showing that the supposed overestimation actually agrees with the Chebyshev radius (i.e., the minimal local worst-case error), at least in a particular instance.

The basis of the argument is the fact that the exact Chebyshev radius and center have already been obtained, in a different form, in case P is an orthogonal projection and $\Lambda\Lambda^* = \mathrm{Id}_m$. Indeed, as an extension to the local optimal recovery problem in Hilbert spaces with approximability model (3) and accurate data (i.e., with $P = P_{\mathcal{V}^\perp}$ and $\eta = 0$), which was settled in [9], we established in [10, Theorem 8] that the Chebyshev center arises from a regularization program with explicitly described parameter $\tau_\sharp \in (0, 1)$. Precisely, for $\tau \in (0, 1)$, let us define

$$f_\tau := \underset{f \in H}{\mathrm{argmin}} \ (1 - \tau)\|Pf\|^2 + \tau\|y - \Lambda f\|^2 \qquad (8)$$

as the solution to a regularization program akin to (5). The Chebyshev center is $f_{\tau_\sharp}$, where the desired parameter $\tau_\sharp$ is the unique τ between $1/2$ and $\varepsilon/(\varepsilon + \eta)$ satisfying the implicit equation

$$\lambda_{\min}((1 - \tau)P + \tau\Lambda^*\Lambda) = \frac{(1 - \tau)^2\varepsilon^2 - \tau^2\eta^2}{(1 - \tau)\varepsilon^2 - \tau\eta^2 + (1 - \tau)\tau(1 - 2\tau)\delta^2}, \qquad (9)$$

in which δ is precomputed as $\delta = \min\{\|Pf\| : \Lambda f = y\} = \min\{\|\Lambda f - y\| : Pf = 0\}$. As explained in [10, Appendix], the above equation can be solved efficiently via the Newton method. According to the yet-to-be-established Theorem 1, the Chebyshev center can alternatively be determined by solving the semidefinite program (4). We have not seriously compared these two options, but we would instinctively favor solving (9) to bypass the black-box nature of semidefinite solvers.

3 Overestimate of the Chebyshev Radius via the S-Procedure

Our goal in this section is to extend a result of [8] to a quantity of interest Q which is an arbitrary linear map between two Hilbert spaces, instead of just $Q = \mathrm{Id}_H$. This extension is not really difficult, but our arguments differ slightly from the ones

of [8]. The result itself, which provides an upper bound for the Chebyshev radius of $Q(\mathcal{K}_{y,\mathcal{E}})$, as well a candidate for its Chebyshev center, is based on the S-prodecure relaxation. When this relaxation is exact, the upper bound agrees with the Chebyshev radius and the candidate is the genuine Chebyshev center. It is in the next section that we establish the exactness of the S-procedure relaxation in our particular instance. Here, we simply explain where the upper bound is coming from.

To this end, we start by recalling the gist of the S-procedure and point to the survey [11] for more details. Given $K + 1$ quadratic functions defined on H, say

$$q_k(h) = \langle A_k h, h \rangle + 2\langle a_k, h \rangle + \alpha_k, \qquad k = 0, 1, \ldots, K,$$

where the A_k's are self-adjoint operators, the a_k's are vectors, and the α_k's are scalars, we consider the two assertions

$$q_0(h) \leq 0 \text{ whenever } q_1(h) \leq 0, \ldots, q_K(h) \leq 0, \tag{10}$$

$$\text{there exist } c_1, \ldots, c_K \geq 0 : q_0(h) \leq c_1 q_1(h) + \cdots + c_K q_K(h) \text{ for all } h \in H. \tag{11}$$

Obviously, if assertion (11) holds, then assertion (10) holds as well. This, in essence, is the S-procedure. The question of its exactness is whether (10) and (11) are in reality equivalent. It is the case for $K = 1$: this is Yakubovich S-lemma [12]. We are actually interested in $K = 2$ here. In this situation, it was shown in [13] that the S-procedure is exact when the scalar field is $\mathbb{C}$, but not necessarily when the scalar field is $\mathbb{R}$, which is our primary concern. Still, under mild assumptions, exactness holds for $\mathbb{R}$ and $K = 2$ in the absence of linear terms, i.e., when $a_0 = a_1 = a_2 = 0$. The latter result, established by Polyak in [14], was the key for us to settle the global optimality problem in [10]. But in general, we make do with the mere relaxation: this leads to the overestimation derived below (which turns into an exact evaluation if the scalar field is $\mathbb{C}$).

Theorem 2 *Let $Q : H \to Z$ be a linear map between finite-dimensional Hilbert spaces. Given Hilbert-space-valued linear maps R and S defined on H and $\mathbb{R}^m$, respectively, and satisfying $\ker(R) \cap \ker(S\Lambda) = \{0\}$, consider the model set (6) and uncertainty set (7), i.e.,*

$$\mathcal{K} = \{f \in H : \|Rf\| \leq \varepsilon\} \subseteq H \qquad and \qquad \mathcal{E} = \{e \in \mathbb{R}^m : \|Se\| \leq \eta\} \subseteq \mathbb{R}^m.$$

For $y \in \mathbb{R}^m$, consider $\widetilde{\gamma}$ and $\widetilde{c}, \widetilde{d}, \widetilde{t}$ to be the minimal value and minimizers of the semidefinite program

$$\underset{\substack{c,d \geq 0 \\ t \in \mathbb{R}}}{\text{minimize}} \ c\varepsilon^2 + d\eta^2 - d\|Sy\|_2^2 + t \quad s.to \quad cR^*R + d\Lambda^*S^*S\Lambda \geq Q^*Q, \tag{12}$$

$$and \quad \begin{bmatrix} cR^*R + d\Lambda^*S^*S\Lambda & d\Lambda^*S^*Sy \\ \hline dy^*S^*S\Lambda & t \end{bmatrix} \geq 0.$$

Then the Chebyshev radius of $Q(\mathcal{K}_{y,\mathcal{E}})$, $\mathcal{K}_{y,\mathcal{E}} := \{f \in H : f \in \mathcal{K} \text{ and } y - \Lambda f \in \mathcal{E}\}$, is upper bounded as

$$\inf_{z \in Z} \sup_{\substack{f \in \mathcal{K} \\ y - \Lambda f \in \mathcal{E}}} \|Q(f) - z\| \;\leq\; \sup_{\substack{f \in \mathcal{K} \\ y - \Lambda f \in \mathcal{E}}} \|Q(f) - Q(f_{\widetilde{c},\widetilde{d}})\| \;=\; \widetilde{\gamma}^{1/2}, \qquad (13)$$

where $f_{\widetilde{c},\widetilde{d}}$ is the solution to a regularization program with parameters $\widetilde{c}, \widetilde{d}$, namely

$$f_{\widetilde{c},\widetilde{d}} := \operatorname*{argmin}_{f \in H} \; \widetilde{c}\,\|Rf\|^2 + \widetilde{d}\,\|S(y - \Lambda f)\|^2. \qquad (14)$$

Proof We drop the dependence on y throughout the argument below. For $c, d \geq 0$, we also use the notation

$$f_{c,d} := \operatorname*{argmin}_{f \in H} \; c\,\|Rf\|^2 + d\,\|S(y - \Lambda f)\|^2.$$

It is essential to keep in mind that $f_{c,d}$ is characterized by

$$cR^*Rf_{c,d} + d\Lambda^*S^*S(\Lambda f_{c,d} - y) = 0. \qquad (15)$$

As $cR^*R + d\Lambda^*S^*S\Lambda$ is invertible thanks to the assumption $\ker(R) \cap \ker(S\Lambda) = \{0\}$, this characterization can be rewritten as

$$f_{c,d} = [cR^*R + d\Lambda^*S^*S\Lambda]^{-1}(d\Lambda^*S^*Sy). \qquad (16)$$

Moreover, working preferentially with squared norms, we introduce the quantities (sv is for squared value, ub is for upper bound, and lub is for least upper bound):

$$\mathsf{sv}(z) = \sup_{f \in H} \|Q(f) - z\|^2 \qquad \text{s.to} \quad f \in \mathcal{K} \text{ and } y - \Lambda f \in \mathcal{E},$$

$$\mathsf{ub}(z) = \inf_{\gamma, c, d \geq 0} \gamma \qquad\qquad \text{s.to} \quad \text{the constraint elucidated in (17) below,}$$

$$\mathsf{lub} = \inf_{c, d \geq 0} \mathsf{Obj} \qquad\qquad \text{s.to} \quad \text{the constraint elucidated in (18) below,}$$

where the objective function is $\mathsf{Obj} = c(\varepsilon^2 - \|Rf_{c,d}\|^2) + d(\eta^2 - \|S(y - \Lambda f_{c,d})\|^2)$. As for the constraints, they read

$$\gamma - c(\varepsilon^2 - \|Rf_{c,d}\|^2) - d(\eta^2 - \|S(y - \Lambda f_{c,d})\|^2) + c\|Rh\|^2 + d\|S\Lambda h\|^2 - \|Qh\|^2$$
$$\geq \|Q(f_{c,d}) - z\|^2 + 2\langle Q(f_{c,d}) - z, Qh \rangle \qquad \text{for all } h \in H, \qquad (17)$$

and
$$c\|Rh\|^2 + d\|S\Lambda h\|^2 - \|Qh\|^2 \geq 0 \qquad \text{for all } h \in H. \qquad (18)$$

We now divide the argument into the proofs of several facts, namely:

 (i) for all $z \in Z$, $\mathsf{sv}(z) \leq \mathsf{ub}(z)$;
 (ii) $\inf_{z \in Z} \mathsf{ub}(z) = \mathsf{lub} = \mathsf{ub}(Q(f_{\widetilde{c},\widetilde{d}}))$;
(iii) the program defining lub is equivalent to the program (12).

Once all these facts are justified, we are able to conclude via

$$\inf_{z \in Z} \mathsf{sv}(z) \leq \mathsf{sv}(Q(f_{\widetilde{c},\widetilde{d}})) \underset{\text{(i)}}{\leq} \mathsf{ub}(Q(f_{\widetilde{c},\widetilde{d}})) \underset{\text{(ii)}}{=} \mathsf{lub} \underset{\text{(iii)}}{=} \widetilde{\gamma},$$

which, up to taking square roots, is the result announced in (13).

<u>Justification of (i).</u> For $z \in Z$, the definition of $\mathsf{sv}(z)$ specified to our situation yields

$$\mathsf{sv}(z) = \sup_{f \in H} \{\|Q(f) - z\|^2 : \|Rf\|^2 \leq \varepsilon^2, \|S(y - \Lambda f)\|^2 \leq \eta^2\}$$

$$= \inf_{\gamma \geq 0} \{\gamma : \|Q(f) - z\|^2 \leq \gamma \text{ whenever } \|Rf\|^2 \leq \varepsilon^2 \text{ and } \|S(y - \Lambda f)\|^2 \leq \eta^2\}.$$

In the spirit of the S-procedure, the latter constraint is satisfied if, for some $c, d \geq 0$,

$$\|Q(f) - z\|^2 - \gamma \leq c(\|Rf\|^2 - \varepsilon^2) + d(\|S(y - \Lambda f)\|^2 - \eta^2) \qquad \text{for all } f \in H. \quad (19)$$

Thus, fixing such $c, d \geq 0$, we obtain

$$\mathsf{sv}(z) \leq \inf_{\gamma} \gamma \quad \text{s.to the constraint (19).}$$

Reparametrizing $f \in H$ as $f = f_{c,d} + h$ with variable $h \in H$, (19) becomes

$$\|Q(f_{c,d}) - z\|^2 - \gamma + 2\langle Q(f_{c,d}) - z, Qh \rangle + \|Qh\|^2$$
$$\leq c(\|Rf_{c,d}\|^2 - \varepsilon^2) + d(\|S(y - \Lambda f_{c,d})\|^2 - \eta^2)$$
$$+ 2(c\langle Rf_{c,d}, Rh \rangle + d\langle S(\Lambda f_{c,d} - y), S\Lambda h \rangle)$$
$$+ c\|Rh\|^2 + d\|S\Lambda h\|^2 \qquad \text{for all } h \in H.$$

The linear term in the right-hand side is, according to (15),

$$2\left\langle cR^*Rf_{c,d} + d\Lambda^*S^*S(\Lambda f_{c,d} - y),\, h \right\rangle = 0.$$

Therefore, for any $c, d \geq 0$, we arrive at

$$\mathsf{sv}(z) \leq \inf_{\gamma \geq 0} \quad \gamma$$
$$\text{s.to} \quad \|Q(f_{c,d}) - z\|^2 + 2\langle Q(f_{c,d}) - z, Qh \rangle$$
$$\leq \gamma - c(\varepsilon^2 - \|Rf_{c,d}\|^2) - d(\eta^2 - \|S(y - \Lambda f_{c,d})\|^2)$$
$$+ c\|Rh\|^2 + d\|S\Lambda h\|^2 - \|Qh\|^2 \qquad \text{for all } h \in H.$$

Taking the infimum over $c, d \geq 0$, we recognize the desired inequality $\mathsf{sv}(z) \leq \mathsf{ub}(z)$. Note that this inequality turns into an equality in case where the S-procedure is exact—in particular, if the scalar field is $\mathbb{C}$. The subsequent steps (ii)-(iii) would then establish that $Q(f_{\widetilde{c},\widetilde{d}})$ is the Chebyshev center.

<u>Justification of (ii): Part 1.</u> Here, we prove that $\inf_{z \in Z} \mathsf{ub}(z) \geq \mathsf{lub}$. Towards this end, for $z \in Z$, we consider the constraint (17) in the defining expression of $\mathsf{ub}(z)$,

which we write succinctly as

$$\text{LHS}(h) \geq \|Q(f_{c,d}) - z\|^2 + 2\langle Q(f_{c,d}) - z, Qh \rangle \qquad \text{for all } h \in H.$$

Since $\text{LHS}(-h) = \text{LHS}(h)$, averaging the above inequality for h and $-h$ leads to $\text{LHS}(h) \geq \|Q(f_{c,d}) - z\|^2$ for all $h \in H$, and hence to $\text{LHS}(h) \geq 0$ for all $h \in H$. Having loosened the constraint, we deduce that

$$\text{ub}(z) \geq \inf_{\gamma, c, d \geq 0} \gamma \qquad \text{s.to} \quad \text{LHS}(h) \geq 0 \quad \text{for all } h \in H.$$

Taking the explicit form of $\text{LHS}(h)$ into account, we see that the above constraint decouples as

$$\gamma - c(\varepsilon^2 - \|Rf_{c,d}\|^2) - d(\eta^2 - \|S(y - \Lambda f_{c,d})\|^2) \geq 0$$

and

$$c\|Rh\|^2 + d\|S\Lambda h\|^2 - \|Qh\|^2 \geq 0 \qquad \text{for all } h \in H.$$

The former reads $\gamma \geq \text{Obj}$ and the latter is the constraint (18). Thus we arrive at

$$\text{ub}(z) \geq \inf_{\gamma, c, d \geq 0} \{\gamma \quad \text{s.to} \quad \gamma \geq \text{Obj} \ \text{and} \ (18)\} = \inf_{c, d \geq 0} \{\text{Obj} \quad \text{s.to} \quad (18)\}.$$

This is the desired inequality $\text{ub}(z) \geq \text{lub}$, valid for any $z \in Z$.

Justification of (iii). In view of

$$c\|Rh\|^2 + d\|S\Lambda h\|^2 - \|Qh\|^2 = \left\langle (cR^*R + d\Lambda^*S^*S\Lambda - Q^*Q)h, \, h \right\rangle,$$

we instantly see that the constraint (18) is equivalent to $cR^*R + d\Lambda^*S^*S\Lambda - Q^*Q \geq 0$. Therefore, the program defining lub is equivalent to

$$\underset{c, d \geq 0}{\text{minimize}} \ \text{Obj} \qquad \text{s.to} \quad cR^*R + d\Lambda^*S^*S\Lambda \geq Q^*Q. \tag{20}$$

We now transform Obj by observing that

$$\begin{aligned}
c\varepsilon^2 + d\eta^2 - \text{Obj} &= c\|Rf_{c,d}\|^2 + d\|S(\Lambda f_{c,d} - y)\|^2 \\
&= c\langle R^*Rf_{c,d}, f_{c,d} \rangle + d\langle S^*S(\Lambda f_{c,d} - y), \Lambda f_{c,d} - y \rangle \\
&= \langle cR^*Rf_{c,d} + d\Lambda^*S^*S(\Lambda f_{c,d} - y), f_{c,d} \rangle - d\langle S^*S\Lambda f_{c,d}, y \rangle + d\langle S^*Sy, y \rangle \\
&= -d\langle S^*S\Lambda f_{c,d}, y \rangle + d\langle S^*Sy, y \rangle,
\end{aligned}$$

where the last step made use of the characterization (15). It follows that

$$\begin{aligned}
\text{Obj} &= c\varepsilon^2 + d\eta^2 - d\|Sy\|^2 + d\langle S^*S\Lambda f_{c,d}, y \rangle \\
&= \inf_{t} \ c\varepsilon^2 + d\eta^2 - d\|Sy\|^2 + t \quad \text{s.to} \quad t \geq d\langle S^*S\Lambda f_{c,d}, y \rangle. \tag{21}
\end{aligned}$$

According to (16), the latter inequality can be written as

$$t \geq dy^* S^* S \Lambda f_{c,d} = (dy^* S^* S \Lambda)[cR^* R + d\Lambda^* S^* S \Lambda]^{-1}(d\Lambda^* S^* Sy),$$

or equivalently as the positive semidefiniteness of a Schur complement, namely as

$$\left[\begin{array}{c|c} cR^* R + d\Lambda^* S^* S \Lambda & d\Lambda^* S^* Sy \\ \hline dy^* S^* S \Lambda & t \end{array} \right] \succeq 0. \tag{22}$$

Substituting (21) into (20) while imposing the additional constraint (22) shows that the program defining lub is indeed equivalent to (12).

Justification of (ii): Part 2. It now remains to prove that $\mathsf{ub}(Q(f_{\widetilde{c},\widetilde{d}})) \leq \mathsf{lub}$, where we recall that $\widetilde{c}, \widetilde{d}, \widetilde{t}$ represent minimizers of (12). By (iii), this also means that $\widetilde{c}, \widetilde{d}$ are minimizers of the problem defining lub. Thus, the feasibility constraint (18) is met for $c = \widetilde{c}$ and $d = \widetilde{d}$, so choosing $\gamma = \widetilde{c}(\varepsilon^2 - \|Rf_{\widetilde{c},\widetilde{d}}\|^2) + \widetilde{d}(\eta^2 - \|S(y - \Lambda f_{\widetilde{c},\widetilde{d}})\|^2)$, we see that the constraint (17) associated to $\mathsf{ub}(z)$ is met with $z = Q(f_{\widetilde{c},\widetilde{d}})$. We deduce that $\mathsf{ub}(Q(f_{\widetilde{c},\widetilde{d}})) \leq \gamma = \widetilde{c}(\varepsilon^2 - \|Rf_{\widetilde{c},\widetilde{d}}\|^2) + \widetilde{d}(\eta^2 - \|S(y - \Lambda f_{\widetilde{c},\widetilde{d}})\|^2)$, which is the minimal value of Obj under the constraint (17). In other words, we have shown that $\mathsf{ub}(Q(f_{\widetilde{c},\widetilde{d}})) \leq \mathsf{lub}$, as desired.   □

4 Exactness of the S-Procedure: Proof of Theorem 1

Our goal in this section is to show that the overestimation of Theorem 2 becomes an exact evaluation if $Q = \mathrm{Id}_H$, P is an orthogonal projection, $S = \mathrm{Id}_m$, and $\Lambda\Lambda^* = \mathrm{Id}_m$, thus proving Theorem 1. We rely on duality in semidefinite programming. Retaining full generality for the moment, our primal program is the rewriting of (12) in the form

$$\mathsf{lub} = \min_{\substack{c,d \geq 0 \\ t \in \mathbb{R}}} c\varepsilon^2 + d(\eta^2 - \|Sy\|^2) + t \tag{23}$$

$$\text{s.to } M_{c,d,t} := \left[\begin{array}{c|c|c} cR^* R + d\Lambda^* S^* S \Lambda - Q^* Q & 0 & 0 \\ \hline 0 & cR^* R + d\Lambda^* S^* S \Lambda & d\Lambda^* S^* Sy \\ \hline 0 & dy^* S^* S \Lambda & t \end{array} \right] \succeq 0.$$

According to e.g. [15, Example 5.11], its dual program reads

$$\mathsf{lub}' = \max_{X \succeq 0} \ \mathrm{tr}\left(\left[\begin{array}{c|c|c} Q^*Q & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right] X\right) \quad \text{s.to} \quad \mathrm{tr}\left(\left[\begin{array}{c|c|c} R^*R & 0 & 0 \\ \hline 0 & R^*R & 0 \\ \hline 0 & 0 & 0 \end{array}\right] X\right) = \varepsilon^2, \tag{24}$$

$$\mathrm{tr}\left(\left[\begin{array}{c|c|c} \Lambda^*S^*S\Lambda & 0 & 0 \\ \hline 0 & \Lambda^*S^*S\Lambda & \Lambda^*S^*Sy \\ \hline 0 & y^*S^*S\Lambda & 0 \end{array}\right] X\right) = \eta^2 - \|Sy\|^2, \quad \mathrm{tr}\left(\left[\begin{array}{c|c|c} 0 & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline 0 & 0 & 1 \end{array}\right] X\right) = 1.$$

It is well known (and easy to verify) that the inequality $\mathsf{lub} \geq \mathsf{lub}'$ always holds. Furthermore, if $\widehat{c}, \widehat{d}, \widehat{t}$ are feasible for (23) and if $\widehat{X}$ is feasible for (24), while $\mathrm{tr}(M_{\widehat{c},\widehat{d},\widehat{t}}\widehat{X}) = 0$, then the equality $\mathsf{lub} = \mathsf{lub}'$ actually holds, and in fact

$$\mathsf{lub} = \widehat{c}\varepsilon^2 + \widehat{d}(\eta^2 - \|Sy\|^2) + \widehat{t} = \mathrm{tr}\left(\left[\begin{array}{c|c|c} Q^*Q & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right]\widehat{X}\right) = \mathsf{lub}'.$$

This simply is a consequence of equalities throughout the chain of inequalities

$$0 \leq \mathsf{lub} - \mathsf{lub}' \leq \widehat{c}\varepsilon^2 + \widehat{d}(\eta^2 - \|Sy\|^2) + t - \mathrm{tr}\left(\left[\begin{array}{c|c|c} Q^*Q & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right]\widehat{X}\right)$$

$$= \widehat{c}\,\mathrm{tr}\left(\left[\begin{array}{c|c|c} R^*R & 0 & 0 \\ \hline 0 & R^*R & 0 \\ \hline 0 & 0 & 0 \end{array}\right]\widehat{X}\right) + \widehat{d}\,\mathrm{tr}\left(\left[\begin{array}{c|c|c} \Lambda^*S^*S\Lambda & 0 & 0 \\ \hline 0 & \Lambda^*S^*S\Lambda & \Lambda^*S^*Sy \\ \hline 0 & y^*S^*S\Lambda & 0 \end{array}\right]\widehat{X}\right) + \widehat{t}\,\mathrm{tr}\left(\left[\begin{array}{c|c|c} 0 & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline 0 & 0 & 1 \end{array}\right]\widehat{X}\right)$$

$$= \mathrm{tr}(M_{\widehat{c},\widehat{d},\widehat{t}}\widehat{X}) = 0.$$

With the aim of choosing such suitable $\widehat{c}, \widehat{d}, \widehat{t}$ and $\widehat{X}$, it is now time to lose generality and consider our specific situation where $Q = \mathrm{Id}_H$, P is an orthogonal projection, $S = \mathrm{Id}_m$, and $\Lambda\Lambda^* = \mathrm{Id}_m$. As pointed out in Section 2, this situation was settled in [10]. There, we showed that the set $\mathcal{K}_{y,\varepsilon} = \{f \in H : \|Pf\| \leq \varepsilon$ and $\|\Lambda f - y\| \leq \eta\}$ admits $f_\sharp \in H$ as its Chebyshev center as soon as one can find $c_\sharp, d_\sharp \geq 0$ and $h_\sharp \in H$ such that

(a) $c_\sharp P + d_\sharp \Lambda^*\Lambda \succeq \mathrm{Id}$,
(b) $c_\sharp P f_\sharp + d_\sharp \Lambda^*(\Lambda f_\sharp - y) + (c_\sharp P + d_\sharp \Lambda^*\Lambda)h_\sharp = h_\sharp$,
(c) $\langle Pf_\sharp, h_\sharp \rangle = 0$, $\quad \langle \Lambda^*(\Lambda f_\sharp - y), h_\sharp \rangle = 0$,
(d) $\|Pf_\sharp + Ph_\sharp\|^2 = \varepsilon^2$, $\quad \|\Lambda f_\sharp - y + \Lambda h_\sharp\|^2 = \eta^2$.

These four sufficient conditions were verified for $f_\sharp = f_{\tau_\sharp}$, where $\tau_\sharp \in (0,1)$ was selected as the solution to the implicit equation (9) and where $f_{\tau_\sharp}$ was selected as in (8) with $\tau = \tau_\sharp$. We also made the choices $c_\sharp = (1 - \tau_\sharp)/\lambda_{\min}$ and $d_\sharp = \tau_\sharp/\lambda_{\min}$, where $\lambda_{\min}$ was the smallest eigenvalue of $(1 - \tau_\sharp)P + \tau_\sharp\Lambda^*\Lambda$. Finally, we took $h_\sharp$ as an associated eigenvector, so that $(c_\sharp P + d_\sharp\Lambda^*\Lambda)h_\sharp = h_\sharp$. It was normalized to satisfy (d)—note that it is the specific choice of $\tau_\sharp$ that made it possible to satisfy both equalities in (d). Keeping these recollections in mind, we now set

$$\widehat{c} := c_\sharp, \quad \widehat{d} := d_\sharp, \quad \widehat{t} = d_\sharp \langle \Lambda f_\sharp, y \rangle, \qquad \text{and} \quad \widehat{X} = \begin{bmatrix} h_\sharp h_\sharp^* & 0 & 0 \\ \hline 0 & f_\sharp f_\sharp^* & -f_\sharp \\ \hline 0 & -f_\sharp^* & 1 \end{bmatrix} \geq 0.$$

The feasibility of $\widehat{c}, \widehat{d}, \widehat{t}$ for (23) follows from (a), combined with the inequality $\widehat{t} \geq d_\sharp \langle \Lambda f_\sharp, y \rangle$ reformulated via the Schur complement as in the justification of (iii)—note that $f_\sharp = f_{c_\sharp, d_\sharp}$. The feasibility of $\widehat{X}$ for (24) is a consequence of (c) and (d): while the third part of the constraint is obvious, the first two parts require some work. For the first part, we observe that

$$\text{tr}\left(\begin{bmatrix} P & 0 & 0 \\ \hline 0 & P & 0 \\ \hline 0 & 0 & 0 \end{bmatrix} \widehat{X} \right) = \text{tr}(P h_\sharp h_\sharp^*) + \text{tr}(P f_\sharp f_\sharp^*) = \|P h_\sharp\|^2 + \|P f_\sharp\|^2$$

$$\underset{(c)}{=} \|P h_\sharp + P f_\sharp\|^2 \underset{(d)}{=} \varepsilon^2.$$

For the second part, we observe that

$$\text{tr}\left(\begin{bmatrix} \Lambda^*\Lambda & 0 & 0 \\ \hline 0 & \Lambda^*\Lambda & \Lambda^* y \\ \hline 0 & y^*\Lambda & 0 \end{bmatrix} \widehat{X} \right) = \text{tr}(\Lambda^*\Lambda h_\sharp h_\sharp^*) + \text{tr}(\Lambda^*\Lambda f_\sharp f_\sharp^*) - 2\langle \Lambda^* y, f_\sharp \rangle$$

$$= \|\Lambda h_\sharp\|^2 + \|\Lambda f_\sharp\|^2 - 2\langle y, \Lambda f_\sharp \rangle = \|\Lambda h_\sharp\|^2 + \|\Lambda f_\sharp - y\|^2 - \|y\|^2$$

$$\underset{(c)}{=} \|\Lambda f_\sharp - y + \Lambda h_\sharp\|^2 - \|y\|^2 \underset{(d)}{=} \eta^2 - \|y\|^2.$$

Having verified the feasibility conditions, we turn to the complementary slackness condition. Using the eigenvalue property of $h_\sharp$ and the characterization of $f_\sharp$ (which together accounted for (b)), we obtain

$$\text{tr}(M_{\widehat{c},\widehat{d},\widehat{t}} \widehat{X}) = \text{tr}((\widehat{c} P + \widehat{d} \Lambda^*\Lambda - \text{Id}) h_\sharp h_\sharp^*) + \text{tr}((\widehat{c} P + \widehat{d} \Lambda^*\Lambda) f_\sharp f_\sharp^*) - 2\widehat{d}\langle \Lambda^* y, f_\sharp \rangle + \widehat{t}$$

$$= \langle (c_\sharp P + d_\sharp \Lambda^*\Lambda - \text{Id}) h_\sharp, h_\sharp \rangle + \langle (c_\sharp P + d_\sharp \Lambda^*(\Lambda f_\sharp - y)), f_\sharp \rangle$$

$$= 0,$$

as expected. At this point, we are guaranteed that

$$\text{lub} = \text{lub}' = \text{tr}\left(\begin{bmatrix} \text{Id}_H & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline 0 & 0 & 0 \end{bmatrix} \widehat{X} \right) = \text{tr}(h_\sharp h_\sharp^*) = \|h_\sharp\|^2 = \|f_\sharp + h_\sharp - f_\sharp\|^2$$

$$\underset{(d)}{\leq} \sup_{f \in H} \|f - f_\sharp\|^2 \quad \text{s.to} \quad \|P f\|^2 \leq \varepsilon^2 \text{ and } \|\Lambda f - y\|^2 \leq \eta^2$$

$$= \text{sv}(f_\sharp) = \inf_{g \in H} \text{sv}(g),$$

where the last equality expresses the result of [10] that $f_\sharp$ is the Chebyshev. Now, with $f_{\widetilde{c},\widetilde{d}}$ denoting the candidate Chebyshev center from Theorem 2, we recall that

$\mathsf{sv}(f_{\widetilde{c},\widetilde{d}}) \leq \mathsf{ub}(f_{\widetilde{c},\widetilde{d}}) = \mathsf{lub}$. Thus, we finally derive that

$$\mathsf{sv}(f_{\widetilde{c},\widetilde{d}}) \leq \inf_{g \in H} \mathsf{sv}(g),$$

i.e., that $f_{\widetilde{c},\widetilde{d}}$ is a Chebyshev center of $\{f \in H : \|Pf\| \leq \varepsilon \text{ and } \|y - \Lambda f\| \leq \eta \}$—actually, by uniqueness, it is the Chebyshev center. This concludes the full proof of Theorem 1.

5 Some Open Questions

This work, as well as the previous work [10], considered the model set $\mathcal{K}$ and the uncertainty set $\mathcal{E}$ given by

$$\mathcal{K} = \{f \in H : \|Pf\|_H \leq \varepsilon\} \qquad \text{and} \qquad \mathcal{E} = \{e \in \mathbb{R}^m : \|e\|_2 \leq \varepsilon\}. \tag{25}$$

For the quantity of interest $Q = \mathrm{Id}_H$, they settled the Chebyshev center issue under the specific conditions that P is an orthogonal projector and that $\Lambda\Lambda^* = \mathrm{Id}_m$. Without these conditions, however, there are several issues that remain unsettled. Here is a list of questions, all concerning the model set and the uncertainty set from (25), to which we do not know the answers—we cannot even foretell if any of them are likely to be easily solved...

1. *Does the Chebyshev center always belong to $\mathcal{V} + \mathrm{ran}(\Lambda^*)$?*
   Under the the specific conditions that P is an orthogonal projector and that $\Lambda\Lambda^* = \mathrm{Id}_m$, we showed in [10] that the solution f_τ to (8) depends affinely on $\tau \in [0, 1]$. Since $f_0 \in \mathcal{V} + \mathrm{ran}(\Lambda^*)$ and $f_1 \in \mathcal{V}$, any f_τ belongs to $\mathcal{V} + \mathrm{ran}(\Lambda^*)$ and in particular the Chebyshev center does. Can this be proved directly without the specific conditions, even if the Hilbert space H is not finite-dimensional? An affirmative answer would allow one to deal with the infinite-dimensional setting by restricting the problem to the finite-dimensional space $\mathcal{V} + \mathrm{ran}(\Lambda^*)$.

2. *Is the Chebyshev center always obtained via regularization?*
   From [10], we know that the answer is affirmative under the specific conditions mentioned above, since the Chebyshev center equals the regularizer $f_{\tau_\sharp}$ with parameter $\tau_\sharp \in (0, 1)$ satisfying the implicit equation (9). Without the specific conditions, we also know from [8] that the answer is affirmative if one considered the complex setting rather than the real setting.

3. *What is the Chebyshev center relative to a quantity of interest $Q \neq \mathrm{Id}_H$?*
   Here, we do not have any partial answers in the real setting, even under the specific conditions above. Indeed, the result of [10] only considered the case $Q = \mathrm{Id}_H$ and initial attempts to extend it to the case $Q \neq \mathrm{Id}_H$ were not conclusive. An obvious guess (valid in the complex setting) is that the Chebyshev center equals $Q(f_{\widetilde{c},\widetilde{d}})$, where $f_{\widetilde{c},\widetilde{d}}$ is the regularizer obtained from the relaxation of Theorem 2.

4. *Does relaxation always lead to near-optimality?*

   Should one fail to find the Chebyshev center, one may be content with the state-
   ment that there exists a parameter $\tau_\natural \in (0, 1)$ such that the regularizer $f_{\tau_\natural}$ admits
   a near-minimal local worst-case error, in the sense that

   $$\sup_{\substack{f \in \mathcal{K} \\ y - \Lambda f \in \mathcal{E}}} \|Q(f) - Q(f_{\tau_\natural})\| \leq C \times \inf_{z \in Z} \sup_{\substack{f \in \mathcal{K} \\ y - \Lambda f \in \mathcal{E}}} \|Q(f) - z\|$$

   for some constant $C \geq 1$. This was proved in [16] under no specific conditions.
   We wonder if the same holds for $f_{\widetilde{c}, \widetilde{d}}$ instead of $f_{\tau_\natural}$.

6 Some Other Optimal Recovery Results Obtained at CAMDA

As alluded to in the introduction, investigations in the field of optimal recovery
constituted one of the pilot research projects during the academic year 2022-23, i.e.,
the initial year of the Center for Approximation and Mathematical Data Analytics,
better known as CAMDA. Below is a synopsis of results generated during this time.

- The article [16] considered a scenario of optimal recovery from inaccurate data
  very close to the present one, the only difference being that it targeted a global
  worst-case error akin to (2) rather than a local worst-case error akin to (1). In
  spite of a focus on graph signals (i.e., functions defined on the set of vertices of a
  graph), it settled the global optimality question in the general setting of the model
  and uncertainty sets from (25) without any specific conditions on P and Λ and for
  an arbitrary linear map $Q : H \to Z$ as quantity of interest. It established that the
  a globally optimal recovery map is given by the linear map $Q \circ \Delta_{\tau_\sharp} : \mathbb{R}^m \to Z$,
  where

  $$\Delta_{\tau_\sharp} : y \in \mathbb{R}^m \mapsto \left[\operatorname*{argmin}_{f \in H}\ (1 - \tau_\sharp)\|Pf\|^2 + \tau_\sharp\|y - \Lambda f\|^2\right] \in H$$

  and where the parameter $\tau_\sharp \in (0, 1)$ is the ratio $\tau_\sharp = d_\sharp/(c_\sharp + d_\sharp)$ featuring the
  solutions $c_\sharp, d_\sharp \geq 0$ to the semidefinite program

  $$\operatorname*{minimize}_{c, d \geq 0}\ c\varepsilon^2 + d\eta^2 \qquad \text{s.to } cP^*P + d\Lambda^*\Lambda \geq Q^*Q.$$

  Precisely, this means that

  $$\sup_{\substack{f \in \mathcal{K} \\ e \in \mathcal{E}}} \|Q(f) - (Q \circ \Delta_{\tau_\sharp})(\Lambda f + e)\| = \inf_{\Delta: \mathbb{R}^m \to Z} \sup_{\substack{f \in \mathcal{K} \\ e \in \mathcal{E}}} \|Q(f) - \Delta(\Lambda f + e)\|.$$

- The article [17] viewed the observation errors differently, since $e \in \mathbb{R}^m$ was
  considered to be a random vector there. It advocated for the global error

$$\mathrm{ge}_p^{\mathrm{or}}(\Delta) := \left(\mathbb{E}\left[\sup_{f \in \mathcal{K}} \|Q(f) - \Delta(\Lambda f + e)\|_Z^p\right]\right)^{1/p}, \qquad p \geq 1,$$

to be used in optimal recovery, as opposed to another notion appearing more often in statistical estimation, namely

$$\mathrm{ge}_p^{\mathrm{se}}(\Delta) := \left(\sup_{f \in \mathcal{K}} \mathbb{E}\left[\|Q(f) - \Delta(\Lambda f + e)\|_Z^p\right]\right)^{1/p}, \qquad p \geq 1.$$

In an arbitrary Banach space F, if $Q \in F^*$ is a linear functional, if the model set $\mathcal{K}$ is symmetric and convex, and if the random vector $e \in \mathbb{R}^m$ is merely log-concave, it was proved that linear maps are near optimal relative to $\mathrm{ge}^{\mathrm{or}}$, in the sense that

$$\inf_{\Delta:\mathbb{R}^m \to \mathbb{R} \text{ linear}} \mathrm{ge}_p^{\mathrm{or}}(\Delta^{\mathrm{lin}}) \leq C_p \times \inf_{\Delta:\mathbb{R}^m \to \mathbb{R}} \mathrm{ge}_p^{\mathrm{or}}(\Delta) \qquad (26)$$

for some constant C_p depending on p. This extends the seminal result of [18], which dealt with $\mathrm{ge}^{\mathrm{se}}$ instead of $\mathrm{ge}^{\mathrm{or}}$ and considered only gaussian vectors $e \in \mathbb{R}^m$.

- The article [20] did not take place in a Hilbert space H but in the Banach space of continuous functions on a compact set $\mathcal{X}$, i.e., $F = C(\mathcal{X})$. It targeted, again from a global optimality perspective, the full recovery problem ($Q = \mathrm{Id}_{C(\mathcal{X})}$) relative to a model set based on approximation capabilities relative to some linear subspace $\mathcal{V} \subseteq C(\mathcal{X})$ and to a parameter $\varepsilon > 0$, namely

$$\mathcal{K}_\mathcal{V} = \{f \in C(\mathcal{X}) : \mathrm{dist}_{C(\mathcal{X})}(f, \mathcal{V}) \leq \varepsilon\}.$$

The observational data were assumed to be acquired as accurate evaluations at points $x^{(1)}, \ldots, x^{(m)} \in \mathcal{X}$. In this situation, it was known (see [19]) how to optimally estimate point evaluations $\delta_x \in C(\mathcal{X})^*$ at each $x \in \mathcal{X}$, and how to subsequently produce an optimal recovery map for $Q = \mathrm{Id}_{C(\mathcal{X})}$—a linear one, to boot. However, the construction was not practical. A practical construction, which amounts to solving about m standard-form linear programs via the simplex method, was conceived in [20] under the proviso that $\mathcal{V}$ is a Chebyshev space. Although this is a strong proviso—by Mairhuber–Curtis theorem, it essentially excludes multivariate functions—it brings hope that genuinely optimal recovery maps in $C(\mathcal{X})$ can be computationally constructed in specific instances.

- The article [21] studied the numerical solutions to partial differential equations. With Ω being a bounded Lipschitz domain in $\mathbb{R}^d$ with $d = 2$ or $d = 3$, it considered the elliptic problem

$$-\Delta u = f \quad \text{in } \Omega, \qquad u = g \quad \text{on } \partial\Omega,$$

where $f \in H^{-1}(\Omega)$ is known but $g \in H^{1/2}(\partial\Omega)$ is not known and when linear observations $y = \Lambda u$ made on the solution $u \in H^1(\Omega)$ are available. Finding an approximant $\widehat{u}$ is viewed as an optimal recovery problem relative to the model set

$$\mathcal{K}_s = \{u \in H^1(\Omega) : -\Delta u = f \text{ and } u_{|\partial\Omega} = g \text{ for some } g \in B_{H^s(\partial\Omega)}\}, \quad s > 1/2.$$

Based on minimum-norm interpolation (aka spline algorithm), an implementable procedure was conceived to generate an output to $\widehat{u}$ which is locally near-optimal, in the sense that $\sup\{\|u - \widehat{u}\|_{H^1(\Omega)} : u \in \mathcal{K}_s, \Lambda u = y\}$ is a most a multiplicative constant away from the Chebyshev radius of the set $\{u \in \mathcal{K}_s : \Lambda u = y\}$.

References

1. Tikhomirov, V. M.: A. N. Kolmogorov and Approximation Theory. Russian Mathematical Surveys, **44**, 101 (1989).
2. Micchelli, C. A., Rivlin, T. J.: A survey of optimal recovery. In: Optimal estimation in approximation theory. pp.1–54. (1977)
3. Novak, E., Wozniakowski, H. : Tractability of Multivariate Problems, Volume I. European Mathematical Society, Zürich (2010)
4. Foucart, S.: Mathematical Pictures at a Data Science Exhibition. Cambridge University Press (2022).
5. Garkavi., A. L. : On the optimal net and best cross-section of a set in a normed space. Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya **26**, 87–106 (1962)
6. Melkman, A. A. , Micchelli, C. A. : Optimal estimation of linear operators in Hilbert spaces from inaccurate data. SIAM Journal on Numerical Analysis **16**, 87–105 (1979)
7. Micchelli, C. A. : Optimal estimation of linear operators from inaccurate data: a second look. Numerical Algorithms **5**, 375–390 (1993)
8. Beck, A., Eldar, Y.C.: Regularization in regression with bounded noise: a Chebyshev center approach. SIAM Journal on Matrix Analysis and Applications **29**, 606–625 (2007)
9. Binev, P. , Cohen, A., Dahmen, W., DeVore, R. , Petrova, G., Wojtaszczyk, P.: Data assimilation in reduced modeling. SIAM/ASA Journal on Uncertainty Quantification **5**, 1–29 (2017)
10. Foucart, S. , Chunyang, L. : Optimal recovery from inaccurate data in Hilbert spaces: regularize, but what of the parameter? Constructive Approximation **57**, 489–520 (2023)
11. Pólik. I., Terlaky, T.: A survey of the S-lemma. SIAM Review **49**, 371–418 (2007)
12. Yakubovich, V. A. : S-procedure in nonlinear control theory. Vestnik Leningrad. Univ. **1**, 62–77 (1971)
13. Beck, A. , Eldar, Y. C.: Strong duality in nonconvex quadratic optimization with two quadratic constraints. SIAM J. Optim. **17**, 844–860 (2006)
14. Polyak, B. T.: Convexity of quadratic transformations and its use in control and optimization. J. Optim. Theory Appl. **99**, 553–583 (1998)
15. Boyd, S. P. , Vandenberghe, L. : Convex Optimization. Cambridge University Press (2004)
16. Foucart, S., Liao, C., and Veldt, N.: On the optimal recovery of graph signals. In: International Conference on Sampling Theory and Applications (SampTA), 2023.
17. Foucart, S. and Paouris, G.: Near-optimal estimation of linear functionals with log-concave observation errors. Information and Inference. To appear.
18. Donoho, D. L.: Statistical estimation and optimal recovery. The Annals of Statistics **22**, 238–270 (1994).
19. DeVore, R., Foucart, S., Petrova, G., Wojtaszczyk, P.: Computing a quantity of interest from observational data. Constructive Approximation **49**, 461–508 (2019)
20. Foucart, S.: Full recovery from point values: an optimal algorithm for Chebyshev approximability prior. Advances in Computational Mathematics **49**, 57 (2023)

21. Binev, P. , Bonito, A., Cohen, A., Dahmen, W., DeVore, R. , Petrova, G.: Solving PDEs with incomplete information. arXiv:2301.05540 (2023).