

Radius of Information for Two Intersected Centered Hyperellipsoids and Implications in Optimal Recovery from Inaccurate Data

Simon Foucart* and Chunyang Liao†

Abstract

For objects belonging to a known model set and observed through a prescribed linear process, we aim at determining methods to recover linear quantities of these objects that are optimal from a worst-case perspective. Working in a Hilbert setting, we show that, if the model set is the intersection of two hyperellipsoids centered at the origin, then there is an optimal recovery method which is linear. It is specifically given by a constrained regularization procedure whose parameters, short of being explicit, can be precomputed by solving a semidefinite program. This general framework can be swiftly applied to several scenarios: the two-space problem, the problem of recovery from ℓ_2 -inaccurate data, and the problem of recovery from a mixture of accurate and ℓ_2 -inaccurate data. With more effort, it can also be applied to the problem of recovery from ℓ_1 -inaccurate data. For the latter, we reach the conclusion of existence of an optimal recovery method which is linear, again given by constrained regularization, under a computationally verifiable sufficient condition. Experimentally, this condition seems to hold whenever the level of ℓ_1 -inaccuracy is small enough. We also point out that, independently of the inaccuracy level, the minimal worst-case error of a linear recovery method can be found by semidefinite programming.

Key words and phrases: Optimal Recovery, Regularization, Semidefinite Programming, S-procedure.

AMS classification: 41A65, 46N40, 90C22, 90C47.

1 Introduction

The question “Do linear problems have linear optimal algorithms?” was surveyed by Packel [1988]. He gave the commonly accepted answer “usually but not always”. This question, central to the subject of Optimal Recovery, is also one of the main concerns of the present article. We shall start by recalling the meaning of this cryptic question and by introducing our notation, already employed in [Foucart, 2022], which is inspired by the field of Learning Theory. The concepts differ by names only from familiar concepts traditionally encountered in the field of Information-Based Complexity (IBC), see e.g. [Novak and Woźniakowski, 2008]. We try to draw parallels between the terminologies of these fields below.

*Texas A&M University, foucart@tamu.edu

†University of California, Los Angeles, liao-chunyang@math.ucla.edu

S. F. is supported by grants from the NSF (DMS-2053172) and from the ONR (N00014-20-1-2787).

Common to the two notational settings is the use of the letter f for the objects of interest, since in both cases they are primarily thought of as functions, although they could be seen as arbitrary elements from a prescribed normed space. Whereas we use the notation F for this normed space (and H when it is a Hilbert space), F typically stands for a strict subset of the said normed space in IBC. We too assume that our objects of interest live in a strict subset of F , but it is denoted by \mathcal{K} and called model set. The premise that $f \in \mathcal{K}$ is referred to as *a priori* information, since it reflects some prior scientific knowledge about realistic objects of interest. In addition, we have at our disposal some *a posteriori* information in the form $y_i = \lambda_i(f)$, $i = 1, \dots, m$, for some linear functionals $\lambda_1, \dots, \lambda_m \in F^*$. Oftentimes, these linear functionals are point evaluations, giving rise, in IBC parlance, to the standard information $y_1 = f(x^{(1)}), \dots, y_m = f(x^{(m)})$. We call $y \in \mathbb{R}^m$ the observation vector and notice that it can be written as $y = \Lambda f$ for some linear map $\Lambda : F \rightarrow \mathbb{R}^m$, referred to as observation map. From the available information, both *a priori* and *a posteriori*, the task is to recover (approximate, learn, ...) the object f in full or maybe just to estimate a quantity of interest Qf , where $Q : F \rightarrow Z$ is a linear map from F into another normed space Z . Such a map Q is called the solution operator in IBC. Our task is realized by way of a recovery map $\Delta : \mathbb{R}^m \rightarrow Z$ —we refrain from using the IBC term algorithm, since computational feasibility is not a requirement at this point. The performance of this recovery map is assessed by the (global) worst-case error defined as

$$\text{Err}_{Q,\mathcal{K}}(\Lambda, \Delta) := \sup_{f \in \mathcal{K}} \|Q(f) - \Delta(\Lambda f)\|_Z.$$

We are interested in how small the latter can be, in other words in the intrinsic error—often labeled radius of information in IBC—defined as

$$\text{Err}_{Q,\mathcal{K}}^*(\Lambda) := \inf_{\Delta : \mathbb{R}^m \rightarrow Z} \text{Err}_{Q,\mathcal{K}}(\Lambda, \Delta).$$

Moreover, our quest is concerned with optimal recovery maps, i.e., recovery maps $\Delta^{\text{opt}} : \mathbb{R}^m \rightarrow Z$ that achieve the above infimum. With the terminology settled, the initial question may now be phrased as: “among all the possible optimal recovery maps, is there one which is linear?”. It is well known that the answer is affirmative in two prototypical situations: (i) when the quantity of interest Q is a linear functional and the model set \mathcal{K} is symmetric and convex (this goes back to [Smolyak, 1965]) and (ii) when F is a Hilbert space and the model set is a centered hyperellipsoid. Another situation allowing for linear optimal recovery maps involves $F = C(\mathcal{X})$, although the existence arguments rarely turn into practical constructions, except in a handful of cases such as [Foucart, 2023].

One contribution of the present article is to uncover yet another situation where optimality of linear recovery maps occur, precisely when the model set is the intersection of two centered hyperellipsoids. We do actually construct the linear optimal recovery map: it is given by constrained regularization with parameters that are clearly determined. In fact, we determine the corresponding radius of information simultaneously: it is the optimal value of a semidefinite program. The main theoretical tool is Polyak’s S-procedure, which elucidates exactly when a quadratic inequality (with no linear

terms) is a consequence of two quadratic inequalities (with no linear terms). This S-procedure is in general not valid with more quadratic inequalities, explaining why our result pertains to the intersection of two centered hyperellipsoids only. This foremost result is established in Section 2 and some implications are derived in subsequent sections. Specifically, Section 3 lists a few simple consequences. One of them is the solution, in the so-called global Optimal Recovery setting, of the two-space problem, where the model set is based on approximability capabilities of two linear subspaces. The other consequences concern two situations for which the observations are inaccurate: first, when inaccuracies are bounded in ℓ_2 , we retrieve—and extend to infinite dimensions—some earlier results of ours; second, when some observations are exact while an ℓ_2 -bound is available for the inaccurate ones, we uncover an optimal recovery map built by constrained regularization, hence linear. Section 4 presents a more intricate consequence of Section 2, namely the scenario of inaccurate observations bounded in ℓ_1 . There, the result is not as pleasing as hoped for, but it nonetheless reveals, somewhat surprisingly, that linear recovery maps can be optimal in this scenario, too. The caveat is that the result holds conditionally on a certain sufficient condition. This condition is close to tautological, but it has the advantage of being computationally verifiable. Our numerical experiments (outlined in the reproducible files accompanying this article) indicate that the condition is likely to hold in case of small ℓ_1 -inaccuracies.

2 Solution for the two-hyperellipsoid-intersection model set

From now on, the space F where the objects of interest live will be a Hilbert space (of possibly infinite dimension), hence it shall be designated by H . There are other Hilbert spaces involved as ranges of linear maps, such as the quantity of interest Q . We will use the notation $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ indistinctly for all the associated Hilbert norms and inner products. Thus, the model set considered in this section—an intersection of two centered hyperellipsoids—takes the form

$$(1) \quad \mathcal{K} = \{f \in H : \|Rf\| \leq 1 \text{ and } \|Sf\| \leq 1\}$$

for some Hilbert-valued bounded linear maps R and S defined on H . We assume throughout that

$$(2) \quad \ker(R) \cap \ker(S) \cap \ker(\Lambda) = \{0\},$$

for otherwise the worst-case error of any recovery map Δ , i.e.,

$$(3) \quad \text{Err}_{Q,\mathcal{K}}(\Lambda, \Delta) := \sup_{\substack{\|Rf\| \leq 1 \\ \|Sf\| \leq 1}} \|Qf - \Delta(\Lambda f)\|$$

would be infinite for $Q = \text{Id}$, say. We also assume that $\Lambda : H \rightarrow \mathbb{R}^m$ is surjective, for otherwise some observations would be redundant. This allows us to define the pseudo-inverse of Λ as

$$\Lambda^\dagger = \Lambda^*(\Lambda\Lambda^*)^{-1} : \mathbb{R}^m \rightarrow H.$$

2.1 Main result

The result stated below not only provides the value of the radius of information, i.e., of the minimal worst-case error over all recovery maps, but it also identifies an optimal recovery map. The latter involves the constrained regularization maps parametrized by $a, b \geq 0$ and defined as

$$(4) \quad \Delta_{a,b} : y \in \mathbb{R}^m \mapsto \left[\operatorname{argmin}_{f \in H} a \|Rf\|^2 + b \|Sf\|^2 \quad \text{s.to } \Lambda f = y \right] \in H, \quad a, b > 0,$$

$$(5) \quad \Delta_{a,0} : y \in \mathbb{R}^m \mapsto \left[\operatorname{argmin}_{f \in H} \|Sf\|^2 \quad \text{s.to } \Lambda f = y \text{ and } Rf = 0 \right] \in \ker(R), \quad a > 0,$$

$$(6) \quad \Delta_{0,b} : y \in \mathbb{R}^m \mapsto \left[\operatorname{argmin}_{f \in H} \|Rf\|^2 \quad \text{s.to } \Lambda f = y \text{ and } Sf = 0 \right] \in \ker(S), \quad b > 0.$$

Although not obvious at first sight, the maps $\Delta_{a,b}$ are linear. For instance, when $a, b > 0$, they indeed take the form, with $\mathcal{N} := \ker \Lambda$ denoting the null space of Λ and $R_{\mathcal{N}}, S_{\mathcal{N}}$ standing for the restrictions of R, S to \mathcal{N} ,

$$(7) \quad \Delta_{a,b} = \Lambda^\dagger - [aR_{\mathcal{N}}^*R_{\mathcal{N}} + bS_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1}(aR_{\mathcal{N}}^*R + bS_{\mathcal{N}}^*S)\Lambda^\dagger,$$

where the invertibility of $aR_{\mathcal{N}}^*R_{\mathcal{N}} + bS_{\mathcal{N}}^*S_{\mathcal{N}}$ follows from (2)¹. It is also worth pointing out that

$$(8) \quad \operatorname{Id} - \Delta_{a,b}\Lambda = [aR_{\mathcal{N}}^*R_{\mathcal{N}} + bS_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1}(aR_{\mathcal{N}}^*R + bS_{\mathcal{N}}^*S).$$

The justification of both (7) and (8) can be found in the appendix. There, we also establish the convergence of $\Delta_{a,b}(y)$ to $\Delta_{a,0}(y)$ as $b \rightarrow 0$ and of $\Delta_{a,b}(y)$ to $\Delta_{0,b}(y)$ as $a \rightarrow 0$, the convergence being understood in the weak sense when $\dim(H) = \infty$.

Theorem 1. For the two-hyperellipsoid-intersection model set (1), the square of the radius of information of the observation map $\Lambda : H \rightarrow \mathbb{R}^m$ for the estimation of Q is given by the optimal value of the program

$$(9) \quad \underset{a,b \geq 0}{\text{minimize}} \quad a + b \quad \text{s.to} \quad a\|Rh\|^2 + b\|Sh\|^2 \geq \|Qh\|^2 \quad \text{for all } h \in \ker \Lambda.$$

Further, if $a_\sharp, b_\sharp \geq 0$ are minimizers of this program, then $Q \circ \Delta_{a_\sharp, b_\sharp}$ is an optimal recovery map. In short,

$$(10) \quad \operatorname{Err}_{Q,\mathcal{K}}(\Lambda, Q \circ \Delta_{a_\sharp, b_\sharp})^2 = \inf_{\Delta: \mathbb{R}^m \rightarrow Z} \operatorname{Err}_{Q,\mathcal{K}}(\Lambda, \Delta)^2 = a_\sharp + b_\sharp.$$

The proof of this result is postponed for a short while. Before that, we address the question of whether the optimization program (9) can be solved in practice. The answer is yes, at least when H is finite-dimensional. Indeed, if (h_1, \dots, h_n) denotes a basis for $\mathcal{N} = \ker \Lambda$, representing $h \in \mathcal{N}$

¹In the infinite-dimensional setting, this assumption should in fact be strengthened to the existence of $\delta > 0$ such that $\max\{\|Rh\|, \|Sh\|\} \geq \delta\|h\|$ for all $h \in \ker \Lambda$, see e.g. [Rudin, 1991, Theorem 12.12].

as $h = \sum_{i=1}^n x_i h_i$ for $x \in \mathbb{R}^n$ allows us to reformulate the constraint $a\|Rh\|^2 + b\|Sh\|^2 \geq \|Qh\|^2$ for all $h \in \mathcal{N}$ as $a\langle R'x, x \rangle + b\langle S'x, x \rangle \geq \langle Q'x, x \rangle$ for all $x \in \mathbb{R}^n$, where $R', S', Q' \in \mathbb{R}^{n \times n}$ are symmetric matrices with entries

$$(11) \quad R'_{i,j} = \langle R(h_i), R(h_j) \rangle, \quad S'_{i,j} = \langle S(h_i), S(h_j) \rangle, \quad Q'_{i,j} = \langle Q(h_i), Q(h_j) \rangle.$$

Thus, the program (9) is equivalent to the semidefinite program²

$$\underset{a, b \geq 0}{\text{minimize}} \quad a + b \quad \text{s.to} \quad aR' + bS' \succeq Q'.$$

Such a semidefinite program can be solved efficiently via a variety of solvers, e.g. the ones embedded in the MATLAB-based modeling system CVX [Grant and Boyd, 2014], although they (currently) all struggle when n is in the thousands.

2.2 Justification

The proof of Theorem 1 is broken down into three small results which we find useful to isolate as separate lemmas. The first lemma estimates the radius of information from below and the second lemma is a key step for the third lemma, which estimates the radius of information from above using constrained regularization maps. Here is the first lemma.

Lemma 2. The squared worst-case error of any recovery map Δ satisfies, with $\mathcal{N} := \ker \Lambda$,

$$(12) \quad \text{Err}_{Q, \mathcal{K}}(\Lambda, \Delta)^2 \geq \text{LB} := \sup_{h \in \mathcal{K} \cap \mathcal{N}} \|Qh\|^2$$

and this lower bound LB can be reformulated as

$$\text{LB} = \inf_{a, b \geq 0} a + b \quad \text{s.to} \quad a\|Rh\|^2 + b\|Sh\|^2 \geq \|Qh\|^2 \quad \text{for all } h \in \mathcal{N}.$$

Proof. We include the argument for the first part, even though it is very classical. It starts by considering any $h \in \mathcal{K} \cap \mathcal{N}$ and by noticing that both $+h$ and $-h$ belong to $\mathcal{K} \cap \mathcal{N}$ before observing that

$$\begin{aligned} \text{Err}_{Q, \mathcal{K}}(\Lambda, \Delta)^2 &= \sup_{f \in \mathcal{K}} \|Qf - \Delta(\Lambda f)\|^2 \geq \max_{\pm} \|Q(\pm h) - \Delta(0)\|^2 \\ &\geq \frac{1}{2} \|Qh - \Delta(0)\|^2 + \frac{1}{2} \|-Qh - \Delta(0)\|^2 = \|Qh\|^2 + \|\Delta(0)\|^2 \\ &\geq \|Qh\|^2. \end{aligned}$$

Finally, it finishes by taking the supremum over $h \in \mathcal{K} \cap \mathcal{N}$ to derive that $\text{Err}_{Q, \mathcal{K}}(\Lambda, \Delta)^2 \geq \text{LB}$.

²Here and in other places, we use the standard notation $M \succeq 0$ to signify that a matrix M is positive semidefinite. The same notation is used for operators.

The argument for the second part begins by reformulating the lower bound as

$$\text{LB} = \inf_{\gamma} \gamma \quad \text{s.to } \|Qh\|^2 \leq \gamma \text{ whenever } h \in \mathcal{N} \text{ satisfies } \|Rh\|^2 \leq 1 \text{ and } \|Sh\|^2 \leq 1.$$

By the version of Polyak's S-procedure recalled in the appendix and its extension to the infinite-dimensional case, the latter constraint is equivalent to³

$$\text{there exist } a, b \geq 0 \text{ such that } \|Qh\|^2 - \gamma \leq a(\|Rh\|^2 - 1) + b(\|Sh\|^2 - 1) \text{ for all } h \in \mathcal{N}.$$

The latter decouples as

$$\text{there exist } a, b \geq 0 \text{ such that } \gamma \geq a + b \text{ and } a\|Rh\|^2 + b\|Sh\|^2 \geq \|Qh\|^2 \text{ for all } h \in \mathcal{N}.$$

Therefore, the lower bound takes the form

$$\text{LB} = \inf_{\substack{\gamma \\ a, b \geq 0}} \gamma \quad \text{s.to } \gamma \geq a + b \text{ and } a\|Rh\|^2 + b\|Sh\|^2 \geq \|Qh\|^2 \text{ for all } h \in \mathcal{N}.$$

Since the minimal value that γ can achieve under these constraints is $a + b$, this infimum indeed reduces to the form of the lower bound announced in the statement of lemma. \square

The second lemma is reminiscent of a result already obtained (with $n = 2$) in [Foucart and Liao, 2022, Lemma 13] for $Q = \text{Id}$ and in [Foucart, Liao, and Veldt, 2023, Lemma 3] for an arbitrary linear quantity of interest Q , but the new proof presented here is more transparent, as it avoids arguments involving semidefinite matrices. As such, it is valid in infinite-dimensional Hilbert spaces, too.

Lemma 3. Let \mathcal{N} be a linear subspace of H and let R_1, \dots, R_n be Hilbert-valued linear maps defined on H . Suppose that $c_1, \dots, c_n > 0$ satisfy

$$(13) \quad \|Qh\|^2 \leq \sum_{i=1}^n c_i \|R_i h\|^2 \quad \text{for all } h \in \mathcal{N}.$$

Then, setting $T = \sum_{i=1}^n c_i R_{i,\mathcal{N}}^* R_i : H \rightarrow \mathcal{N}$ and assuming that $T_{\mathcal{N}} = \sum_{i=1}^n c_i R_{i,\mathcal{N}}^* R_{i,\mathcal{N}} : \mathcal{N} \rightarrow \mathcal{N}$ is invertible, one has

$$\left\| Q T_{\mathcal{N}}^{-1} \left(\sum_{i=1}^n c_i R_{i,\mathcal{N}}^* R_i f_i \right) \right\|^2 \leq \sum_{i=1}^n c_i \|R_i f_i\|^2 \quad \text{for all } f_1, \dots, f_n \in H.$$

Proof. To ease notation, let $h := \sum_{i=1}^n c_i R_{i,\mathcal{N}}^* R_i f_i$. Note that h belongs to \mathcal{N} , and so does $T_{\mathcal{N}}^{-1}h$. In view of (13), it is enough to prove that

$$(14) \quad \sum_{i=1}^n c_i \|R_i T_{\mathcal{N}}^{-1} h\|^2 \leq \sum_{i=1}^n c_i \|R_i f_i\|^2.$$

³To verify the applicability of the S-procedure, note that $h = 0$ satisfies the strict feasibility condition ($h \in \mathcal{N}$, $\|Rh\|^2 < 1$, $\|Sh\|^2 < 1$) and that any $a, b > 0$ satisfy the positive definiteness condition ($aR_{\mathcal{N}}^* R_{\mathcal{N}} + bS_{\mathcal{N}}^* S_{\mathcal{N}} \succ 0$).

The left-hand side of (14), which we denote by LHS for short, is manipulated as follows:

$$\begin{aligned}
\text{LHS} &= \sum_{i=1}^n c_i \|R_{i,\mathcal{N}} T_{\mathcal{N}}^{-1} h\|^2 = \sum_{i=1}^n c_i \left\langle R_{i,\mathcal{N}}^* R_{i,\mathcal{N}} T_{\mathcal{N}}^{-1} h, T_{\mathcal{N}}^{-1} h \right\rangle = \left\langle \sum_{i=1}^n c_i R_{i,\mathcal{N}}^* R_{i,\mathcal{N}} T_{\mathcal{N}}^{-1} h, T_{\mathcal{N}}^{-1} h \right\rangle \\
&= \left\langle T_{\mathcal{N}} T_{\mathcal{N}}^{-1} h, T_{\mathcal{N}}^{-1} h \right\rangle = \left\langle h, T_{\mathcal{N}}^{-1} h \right\rangle = \left\langle \sum_{i=1}^n c_i R_{i,\mathcal{N}}^* R_{i,\mathcal{N}} f_i, T_{\mathcal{N}}^{-1} h \right\rangle = \sum_{i=1}^n c_i \left\langle R_{i,\mathcal{N}}^* R_{i,\mathcal{N}} f_i, T_{\mathcal{N}}^{-1} h \right\rangle \\
&= \sum_{i=1}^n c_i \left\langle R_i f_i, R_i T_{\mathcal{N}}^{-1} h \right\rangle.
\end{aligned}$$

From the general inequality $\langle u, v \rangle \leq (\|u\|^2 + \|v\|^2)/2$, we derive that

$$\text{LHS} \leq \sum_{i=1}^n \frac{c_i}{2} \left(\|R_i f_i\|^2 + \|R_i T_{\mathcal{N}}^{-1} h\|^2 \right) = \frac{1}{2} \sum_{i=1}^n c_i \|R_i f_i\|^2 + \frac{1}{2} \text{LHS},$$

which is just a rearrangement of the desired inequality (14). \square

The third and final lemma gives an upper bound for the squared worst-case error of the constrained regularization map $\Delta_{a,b}$.

Lemma 4. Suppose that $a, b > 0$ satisfy

$$\|Qh\|^2 \leq a\|Rh\|^2 + b\|Sh\|^2 \quad \text{for all } h \in \ker \Lambda.$$

Then one has

$$\text{Err}_{Q,\mathcal{K}}(\Lambda, Q \circ \Delta_{a,b})^2 \leq a + b.$$

Proof. The squared worst-case error of the recovery map $Q \circ \Delta_{a,b}$ is

$$\begin{aligned}
\text{Err}_{Q,\mathcal{K}}(\Lambda, Q \circ \Delta_{a,b})^2 &= \sup_{f \in \mathcal{K}} \|Qf - Q \circ \Delta_{a,b}(\Lambda f)\|^2 = \sup_{\substack{\|Rf\| \leq 1 \\ \|Sf\| \leq 1}} \|Q(\text{Id} - \Delta_{a,b}\Lambda)f\|^2 \\
&= \sup_{\substack{\|Rf\| \leq 1 \\ \|Sf\| \leq 1}} \left\| Q[aR_{\mathcal{N}}^* R_{\mathcal{N}} + bS_{\mathcal{N}}^* S_{\mathcal{N}}]^{-1} (aR_{\mathcal{N}}^* Rf + bS_{\mathcal{N}}^* Sf) \right\|^2,
\end{aligned}$$

where we have made use of (8) with $\mathcal{N} = \ker \Lambda$. Then, invoking Lemma 3 for $n = 2$, $R_1 = R$, $R_2 = S$, and $f_1 = f_2 = f$, we obtain

$$\text{Err}_{Q,\mathcal{K}}(\Lambda, Q \circ \Delta_{a,b})^2 \leq \sup_{\substack{\|Rf\| \leq 1 \\ \|Sf\| \leq 1}} (a\|Rf\|^2 + b\|Sf\|^2) \leq a + b,$$

which is the announced result. \square

With this series of lemmas at hand, we are now ready to justify the main result of this section.

Proof of Theorem 1. Let $a_{\sharp}, b_{\sharp} \geq 0$ be minimizers of the optimization program (9). On the one hand, Lemma 2 guarantees that

$$(15) \quad \inf_{\Delta: \mathbb{R}^m \rightarrow Z} \text{Err}_{Q, \mathcal{K}}(\Lambda, \Delta)^2 \geq a_{\sharp} + b_{\sharp}.$$

On the other hand, by the feasibility of a_{\sharp} and b_{\sharp} , we have $\|Qh\|^2 \leq a_{\sharp}\|Rh\|^2 + b_{\sharp}\|Sh\|^2$ for all $h \in \ker \Lambda$. To deal with the possibility of a_{\sharp} or b_{\sharp} being zero, we consider, for any $\varepsilon > 0$, $a_{\sharp}^{(\varepsilon)} := a_{\sharp} + \varepsilon > 0$ and $b_{\sharp}^{(\varepsilon)} := b_{\sharp} + \varepsilon > 0$ and notice that $\|Qh\|^2 \leq a_{\sharp}^{(\varepsilon)}\|Rh\|^2 + b_{\sharp}^{(\varepsilon)}\|Sh\|^2$ for all $h \in \ker \Lambda$. Lemma 4 then guarantees that $\text{Err}_{Q, \mathcal{K}}(\Lambda, Q \circ \Delta_{a_{\sharp}^{(\varepsilon)}, b_{\sharp}^{(\varepsilon)}})^2 \leq a_{\sharp}^{(\varepsilon)} + b_{\sharp}^{(\varepsilon)}$. It is now easy to see that taking (possibly weak) limits as $\varepsilon \rightarrow 0$ yields

$$(16) \quad \text{Err}_{Q, \mathcal{K}}(\Lambda, Q \circ \Delta_{a_{\sharp}, b_{\sharp}})^2 \leq a_{\sharp} + b_{\sharp}.$$

The inequalities (15) and (16) together fully justify (10) and thus complete the proof. \square

2.3 Side results

In this section, we put forward an interpretation of the radius of information that differs from the minimal value of the program (9) and we shed light on the extremizer appearing in the expression of the lower bound from (12)—which is now known to coincide with the squared radius of information. Although these results are not used later, we include them here because they appear interesting for their own sake. Both results call upon the largest eigenvalue, denoted by λ_{\max} , of self-adjoint operators.

Proposition 5. For the two-hyperellipsoid-intersection model set (1), the radius of information of the observation map $\Lambda : H \rightarrow \mathbb{R}^m$ for the estimation of Q is also given as the optimal value λ_{\sharp} of the program

$$\underset{\tau \in [0, 1]}{\text{minimize}} \quad \lambda_{\max}(Q_{\mathcal{N}}[(1 - \tau)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau S_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1}Q_{\mathcal{N}}^*),$$

where $\mathcal{N} := \ker \Lambda$. Moreover, if $\tau_{\sharp} \in (0, 1)$ represents a minimizer of the above program, then $a_{\sharp} := (1 - \tau_{\sharp})\lambda_{\sharp}$ and $b_{\sharp} := \tau_{\sharp}\lambda_{\sharp}$ are minimizers of (9).

Proof. The foremost observation consists in reformulating the constraint in (9) as

$$(17) \quad \lambda_{\max}(Q_{\mathcal{N}}[aR_{\mathcal{N}}^*R_{\mathcal{N}} + bS_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1}Q_{\mathcal{N}}^*) \leq 1.$$

Indeed, the said constraint can be equivalently expressed in the form

$$\begin{aligned} aR_{\mathcal{N}}^*R_{\mathcal{N}} + bS_{\mathcal{N}}^*S_{\mathcal{N}} \succeq Q_{\mathcal{N}}^*Q_{\mathcal{N}} &\iff \text{Id} \succeq [aR_{\mathcal{N}}^*R_{\mathcal{N}} + bS_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1/2}Q_{\mathcal{N}}^*Q_{\mathcal{N}}[aR_{\mathcal{N}}^*R_{\mathcal{N}} + bS_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1/2} \\ &\iff 1 \geq \lambda_{\max}([aR_{\mathcal{N}}^*R_{\mathcal{N}} + bS_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1/2}Q_{\mathcal{N}}^*Q_{\mathcal{N}}[aR_{\mathcal{N}}^*R_{\mathcal{N}} + bS_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1/2}) \\ &\iff 1 \geq \lambda_{\max}(Q_{\mathcal{N}}[aR_{\mathcal{N}}^*R_{\mathcal{N}} + bS_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1}Q_{\mathcal{N}}^*). \end{aligned}$$

Thus, the above-defined $a_{\sharp} = (1 - \tau_{\sharp})\lambda_{\sharp}$ and $b_{\sharp} = \tau_{\sharp}\lambda_{\sharp}$ are feasible for (9), since then

$$(18) \quad \lambda_{\max}(Q_{\mathcal{N}}[a_{\sharp}R_{\mathcal{N}}^*R_{\mathcal{N}} + b_{\sharp}S_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1}Q_{\mathcal{N}}^*) = \frac{1}{\lambda_{\sharp}}\lambda_{\max}(Q_{\mathcal{N}}[(1 - \tau_{\sharp})R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau_{\sharp}S_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1}Q_{\mathcal{N}}^*) = 1.$$

It now remains to show that $a + b \geq a_{\sharp} + b_{\sharp}$ whenever $a, b > 0$ are feasible for (9). To see this, with $\tau = b/(a + b)$ and $1 - \tau = a/(a + b)$, notice that

$$\begin{aligned} \lambda_{\sharp} &\leq \lambda_{\max}(Q_{\mathcal{N}}[(1 - \tau)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau S_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1}Q_{\mathcal{N}}^*) = (a + b)\lambda_{\max}(Q_{\mathcal{N}}[aR_{\mathcal{N}}^*R_{\mathcal{N}} + bS_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1}Q_{\mathcal{N}}^*) \\ &\leq (a + b), \end{aligned}$$

where (17) was used for the last inequality. The desired conclusion follows from $\lambda_{\sharp} = a_{\sharp} + b_{\sharp}$. \square

Proposition 6. Under the setting of Proposition 5, recall that

$$(19) \quad \sup_{h \in \mathcal{N}} \{\|Qh\|^2 : \|Rh\|^2 \leq 1, \|Sh\|^2 \leq 1\} = \inf_{a, b \geq 0} \{a + b : aR_{\mathcal{N}}^*R_{\mathcal{N}} + bS_{\mathcal{N}}^*S_{\mathcal{N}} \succeq Q_{\mathcal{N}}^*Q_{\mathcal{N}}\}.$$

If $h_{\sharp} \in \mathcal{N}$ and $a_{\sharp}, b_{\sharp} > 0$ are extremizers of the above two programs, then

- (i) $\|Rh_{\sharp}\| = 1$ and $\|Sh_{\sharp}\| = 1$;
- (ii) $(a_{\sharp}R_{\mathcal{N}}^*R_{\mathcal{N}} + b_{\sharp}S_{\mathcal{N}}^*S_{\mathcal{N}})h_{\sharp} = Q_{\mathcal{N}}^*Q_{\mathcal{N}}h_{\sharp}$.

Proof. Setting $T_{\mathcal{N}} := a_{\sharp}R_{\mathcal{N}}^*R_{\mathcal{N}} + b_{\sharp}S_{\mathcal{N}}^*S_{\mathcal{N}}$, we already know from (18) that $\lambda_{\max}(Q_{\mathcal{N}}T_{\mathcal{N}}^{-1}Q_{\mathcal{N}}^*) = 1$. In view of $\|Rh_{\sharp}\| \leq 1$, $\|Sh_{\sharp}\| \leq 1$, $\|Qh_{\sharp}\|^2 = a_{\sharp} + b_{\sharp}$, and writing $g_{\sharp} := T_{\mathcal{N}}^{1/2}h_{\sharp}$, we observe that

$$\begin{aligned} \|g_{\sharp}\|^2 &= \langle T_{\mathcal{N}}h_{\sharp}, h_{\sharp} \rangle = \langle (a_{\sharp}R_{\mathcal{N}}^*R_{\mathcal{N}} + b_{\sharp}S_{\mathcal{N}}^*S_{\mathcal{N}})h_{\sharp}, h_{\sharp} \rangle = a_{\sharp}\|Rh_{\sharp}\|^2 + b_{\sharp}\|Sh_{\sharp}\|^2 \\ &\stackrel{(1)}{\leq} a_{\sharp} + b_{\sharp} = \|Qh_{\sharp}\|^2 = \|Q_{\mathcal{N}}T_{\mathcal{N}}^{-1/2}g_{\sharp}\|^2 = \langle (T_{\mathcal{N}}^{-1/2}Q_{\mathcal{N}}^*Q_{\mathcal{N}}T_{\mathcal{N}}^{-1/2})g_{\sharp}, g_{\sharp} \rangle \\ &\stackrel{(2)}{\leq} \lambda_{\max}(T_{\mathcal{N}}^{-1/2}Q_{\mathcal{N}}^*Q_{\mathcal{N}}T_{\mathcal{N}}^{-1/2})\|g_{\sharp}\|^2 = \lambda_{\max}(Q_{\mathcal{N}}T_{\mathcal{N}}^{-1}Q_{\mathcal{N}}^*)\|g_{\sharp}\|^2 = \|g_{\sharp}\|^2. \end{aligned}$$

Since the left-hand and right-hand sides are identical, equality must hold throughout. In particular, equality in (1) implies (i). As for equality in (2), it imposes that g_{\sharp} is an eigenvector associated with the eigenvalue $\lambda_{\max}(T_{\mathcal{N}}^{-1/2}Q_{\mathcal{N}}^*Q_{\mathcal{N}}T_{\mathcal{N}}^{-1/2}) = 1$, meaning that $(T_{\mathcal{N}}^{-1/2}Q_{\mathcal{N}}^*Q_{\mathcal{N}}T_{\mathcal{N}}^{-1/2})g_{\sharp} = g_{\sharp}$, i.e., $Q_{\mathcal{N}}^*Q_{\mathcal{N}}h_{\sharp} = T_{\mathcal{N}}h_{\sharp}$, which is (ii). \square

Remark. The result of Proposition 6 is, in a sense, a characterization of the equality between the supremum and the infimum in (19). Indeed, the argument can easily be turned around: given minimizers $a_{\sharp}, b_{\sharp} > 0$, if we can find $h_{\sharp} \in \mathcal{N}$ satisfying (i) and (ii), then the supremum equals the infimum (it is always “at most” by the trivial part of the S-procedure and it is “at least” thanks to the existence of h_{\sharp}). Such an approach was used, in essence, to determine explicit solutions of specific differential-equation-inspired Optimal Recovery problems featuring two quadratics constraints but without invoking Polyak’s S-procedure, see [Magaril-Il’yaev, Osipenko, and Tikhomirov, 2004,

Vvedenskaya, 2009]. The same circle of ideas extends to the intersection of $n > 2$ hyperellipsoids. Indeed, leaving the details to the reader, we state the loose equivalence between the equality

$$\sup_{h \in \mathcal{N}} \{ \|Qh\|^2 : \|R_i h\|^2 \leq 1, i = 1, \dots, n \} = \inf_{c_1, \dots, c_n \geq 0} \left\{ \sum_{i=1}^n c_i : c_i R_{i,\mathcal{N}}^* R_{i,\mathcal{N}} \succeq Q_{\mathcal{N}}^* Q_{\mathcal{N}} \right\}$$

and, with $c_1^\sharp, \dots, c_n^\sharp > 0$ denoting minimizers of the latter, the existence of $h_\sharp \in \mathcal{N}$ such that

$$\|R_i h_\sharp\| = 1, \quad i = 1, \dots, n, \quad \text{and} \quad \left(\sum_{i=1}^n c_i^\sharp R_{i,\mathcal{N}}^* R_{i,\mathcal{N}} \right) h_\sharp = Q_{\mathcal{N}}^* Q_{\mathcal{N}} h_\sharp.$$

This gives us a practical way of deciding whether Theorem 1 extends to $n > 2$ hyperellipsoids: after solving a semidefinite program, construct a candidate h_\sharp by solving an eigenvalue problem and test if the $\|R_i h_\sharp\|$ are all equal. As observed numerically, this occurs in some situations, but certainly not in all, in particular not when the R_i are orthogonal projectors (as in the multispace problem described below). The moral is that this article deals with the intersection of $n = 2$ hyperellipsoids not only because the strategy based on Polyak's S-procedure does not apply to $n > 2$, but also because the natural extension is not valid for $n > 2$.

3 Three easy consequences

The two-hyperellipsoid-intersection framework has direct implications for optimal recovery from (partially) inaccurate data, to be discussed later, and even more directly for optimal recovery from accurate data under the two-space approximability model, to be elucidated right now.

3.1 The two-space problem

A model set based on approximability by a linear subspace V of F with parameter $\varepsilon > 0$, namely

$$\mathcal{K} = \{f \in F : \text{dist}(f, V) \leq \varepsilon\},$$

gained traction after the work of Binev, Cohen, Dahmen, DeVore, Petrova, and Wojtaszczyk [2017]. When F is a Hilbert space, these authors completely solved the full recovery ($Q = \text{Id}$) problem even in the local setting—the present article deals solely with the global setting. They also raised the question of the multispace problem, a particular case of which being the two-space problem where

$$(20) \quad \mathcal{K} = \{f \in H : \text{dist}(f, V) \leq \varepsilon \text{ and } \text{dist}(f, W) \leq \eta\}.$$

For the multispace problem, they proposed two iterative algorithms which, in the limit, produce model- and data-consistent objects. As such, these algorithms yield worst-case errors that are near-optimal by a factor at most two.

The two-space problem—in fact, even the multispace problem—in an arbitrary normed space F was solved in [Foucart, 2021] but only when the quantity of interest Q is a linear functional. For more general linear maps Q , but when F is a Hilbert space, the two-space problem is a special case of our two-hyperellipsoid-intersection problem. Indeed, the model set (20) is an instantiation of the model set (1) with R and S being scaled versions of the orthogonal projectors onto the orthogonal complements of V and W , precisely $R = (1/\varepsilon)P_{V^\perp}$ and $S = (1/\eta)P_{W^\perp}$. Thus, Theorem 1 applies directly and we arrive, through the change of optimization variables $a = c\varepsilon^2$ and $b = d\eta^2$, at the result stated below for completeness.

Theorem 7. For the two-space model set (20), the square of the radius of information of the observation map $\Lambda : H \rightarrow \mathbb{R}^m$ for the estimation of Q is given by the optimal value of the program

$$(21) \quad \underset{c, d \geq 0}{\text{minimize}} \quad c\varepsilon^2 + d\eta^2 \quad \text{s.to} \quad c\|P_{V^\perp}h\|^2 + d\|P_{W^\perp}h\|^2 \geq \|Qh\|^2 \quad \text{for all } h \in \ker \Lambda.$$

Further, if $c_\sharp, d_\sharp \geq 0$ are minimizers of this program and if $\Delta_{c_\sharp, d_\sharp}$ is the map defined for $y \in \mathbb{R}^m$ by

$$\Delta_{c_\sharp, d_\sharp}(y) = \left[\underset{f \in H}{\operatorname{argmin}} \quad c_\sharp\|P_{V^\perp}f\|^2 + d_\sharp\|P_{W^\perp}f\|^2 \quad \text{s.to} \quad \Lambda f = y \right]$$

(and interpreted via continuity in case $c_\sharp = 0$ or $d_\sharp = 0$), then the linear map $Q \circ \Delta_{c_\sharp, d_\sharp}$ provides an optimal recovery map.

3.2 Recovery from ℓ_2 -inaccurate data

Suppose now that the observations made on the objects of interest $f \in \mathcal{K}$ are not accurate anymore, but rather of the form $y = \Lambda f + e$ with an error vector $e \in \mathbb{R}^m$ belonging to some uncertainty set \mathcal{E} . We then need to adjust the notion of worst-case error of a recovery map $\Delta : \mathbb{R}^m \rightarrow Z$ and thus define the quantity

$$(22) \quad \operatorname{Err}_{Q, \mathcal{K}, \mathcal{E}}(\Lambda, \Delta) := \sup_{\substack{f \in \mathcal{K} \\ e \in \mathcal{E}}} \|Qf - \Delta(\Lambda f + e)\|.$$

As already remarked in [Micchelli and Rivlin, 1977], this setting can formally be reduced to the setting of accurate data. When Q is a linear functional, the existence of a linear optimal recovery map follows for data perturbation not only bounded in ℓ_2 , see e.g. [Magaril-Il'yayev and Osipenko, 1991, Plaskota, 1996, Ettehad and Foucart, 2021].

In this subsection, both the model set and uncertainty set are hyperellipsoids, i.e.,

$$(23) \quad \mathcal{K} = \{f \in H : \|Rf\| \leq \varepsilon\},$$

$$(24) \quad \mathcal{E} = \{e \in \mathbb{R}^m : \|Se\| \leq \eta\}.$$

In this situation, the problem at hand reduces to the two-hyperellipsoid-intersection problem with accurate data. Indeed, considering the compound variable $\tilde{f} = (f, e)$ belonging to the extended Hilbert space $\tilde{H} := H \times \mathbb{R}^m$, let us introduce linear maps $\tilde{\Lambda}$, \tilde{Q} , \tilde{R} , and \tilde{S} defined on \tilde{H} by

$$(25) \quad \tilde{\Lambda}((f, e)) = \Lambda f + e, \quad \tilde{Q}((f, e)) = Qf, \quad \tilde{R}((f, e)) = (1/\varepsilon)Rf, \quad \tilde{S}((f, e)) = (1/\eta)Se.$$

The worst-case error (22) of the recovery map Δ is then expressed as

$$\text{Err}_{Q, \mathcal{K}, \varepsilon}(\Lambda, \Delta) = \sup_{\substack{\|\tilde{R}\tilde{f}\| \leq 1 \\ \|\tilde{S}\tilde{f}\| \leq 1}} \|\tilde{Q}\tilde{f} - \Delta(\tilde{\Lambda}\tilde{f})\|,$$

i.e., exactly as in (3). Exploiting this analogy, Theorem 1 yields the result stated below. Note that it is not entirely new. Indeed, the fact that regularization produces a (linear) optimal recovery map was already recognized in [Mekman and Micchelli, 1979], see also [Micchelli, 1993]. However, a recipe for selecting the regularization parameters was not given there, except on a specific example. Such a recipe was uncovered in [Foucart and Liao, 2022] with $Q = \text{Id}$, R being an orthogonal projector, and $S = \text{Id}$, and later in [Foucart, Liao, and Veldt, 2023] with arbitrary Q and R but still with $S = \text{Id}$. The extension to $S \neq \text{Id}$ is minor—more pertinent is the fact that the result is now valid in infinite dimensions (although solving (26) in practice would then be a challenge).

Theorem 8. For the hyperellipsoidal model and uncertainty sets (23) and (24), the square of the radius of information of the observation map $\Lambda : H \rightarrow \mathbb{R}^m$ for the estimation of Q is given by the optimal value of the program

$$(26) \quad \underset{c, d \geq 0}{\text{minimize}} \quad c\varepsilon^2 + d\eta^2 \quad \text{s.to} \quad c\|Rf\|^2 + d\|S\Lambda f\|^2 \geq \|Qf\|^2 \quad \text{for all } f \in H.$$

Further, if $c_{\sharp}, d_{\sharp} \geq 0$ are minimizers of this program and if $\Delta_{c_{\sharp}, d_{\sharp}}$ is the map defined for $y \in \mathbb{R}^m$ by

$$(27) \quad \Delta_{c_{\sharp}, d_{\sharp}}(y) = \left[\underset{f \in H}{\text{argmin}} \quad c_{\sharp}\|Rf\|^2 + d_{\sharp}\|S(y - \Lambda f)\|^2 \right]$$

(and interpreted via continuity in case $c_{\sharp} = 0$ or $d_{\sharp} = 0$), then the linear map $Q \circ \Delta_{c_{\sharp}, d_{\sharp}}$ provides an optimal recovery map.

Proof. With the change of optimization variables $a = c\varepsilon^2$ and $b = d\eta^2$, the program (9) for $\tilde{\Lambda}$, \tilde{Q} , \tilde{R} , and \tilde{S} becomes

$$\underset{c, d \geq 0}{\text{minimize}} \quad c\varepsilon^2 + d\eta^2 \quad \text{s.to} \quad c\|Rf\|^2 + d\|Se\|^2 \geq \|Qf\|^2 \quad \text{when } f \in H, e \in \mathbb{R}^m \text{ satisfy } \Lambda f + e = 0.$$

Eliminating $e \in \mathbb{R}^m$ from the above yields the program (26). As for the constrained regularization map $\Delta_{a_{\sharp}, b_{\sharp}}$ from (4), it is to be replaced by

$$y \in \mathbb{R}^m \mapsto \left[\underset{f \in H, e \in \mathbb{R}^m}{\text{argmin}} \quad c_{\sharp}\|Rf\|^2 + d_{\sharp}\|Se\|^2 \quad \text{s.to } \Lambda f + e = y \right].$$

Again, eliminating $e \in \mathbb{R}^m$ from the above leads to (27). \square

3.3 Recovery for mixed accurate and ℓ_2 -inaccurate data

In some situations, parts of the observations on the objects of interest $f \in \mathcal{K}$ can be made accurately, while other parts are subject to errors. Such a situation occurs e.g. when learning the parameters of partial differential equations (using kernel methods, say) from example solutions that can be perfectly evaluated at points on the boundary of the domain but imprecisely at points inside the domain, see e.g [Anandkumar, Azizzadenesheli, Bhattacharya, Kovachki, Li, Liu, and Stuart, 2020, Long, Mrvaljevic, Zhe, and Hosseini, 2023]. To cover this possibility, we can decompose the error vector $e \in \mathbb{R}^m$ as $e = (e', e'') \in \mathbb{R}^{m'} \times \mathbb{R}^{m''}$, with $e' = 0$ and $\|e''\| \leq \eta$. More generally, we shall assume that $S'e = 0$ and $\|S''e\| \leq \eta$ for some Hilbert-valued linear maps S', S'' defined on \mathbb{R}^m . We shall therefore consider model and uncertainty sets of the form

$$(28) \quad \mathcal{K} = \{h \in H : \|Rf\| \leq \varepsilon\},$$

$$(29) \quad \mathcal{E} = \{e \in \ker(S') : \|S''e\| \leq \eta\}.$$

This time working with the different extended space $\tilde{H} = H \times \ker(S')$, we still introduce linear maps $\tilde{\Lambda}$, \tilde{Q} , \tilde{R} , and \tilde{S} defined on the compound variable $\tilde{f} = (f, e) \in \tilde{H}$ almost as in (25), but with one slight modification for \tilde{S} , namely

$$(30) \quad \tilde{\Lambda}((f, e)) = \Lambda f + e, \quad \tilde{Q}((f, e)) = Qf, \quad \tilde{R}((f, e)) = (1/\varepsilon)Rf, \quad \tilde{S}((f, e)) = (1/\eta)S''e.$$

The worst-case error (22) of a recovery map for this mixed error scenario is still identifiable with the worst-case error (3) for the two-hyperellipsoid-intersection scenario, so we can once more leverage Theorem 1 to derive the following result.

Theorem 9. For the model set (28) and the mixed-uncertainty set (29), the square of the radius of information of the observation map $\Lambda : H \rightarrow \mathbb{R}^m$ for the estimation of Q is given by the optimal value of the program

$$(31) \quad \underset{c, d \geq 0}{\text{minimize}} \quad c\varepsilon^2 + d\eta^2 \quad \text{s.to} \quad c\|Rf\|^2 + d\|S''\Lambda f\|^2 \geq \|Qf\|^2 \quad \text{for all } f \in \ker(S'\Lambda).$$

Further, if $c_\sharp, d_\sharp \geq 0$ are minimizers of this program and if $\Delta_{c_\sharp, d_\sharp}$ is the map defined for $y \in \mathbb{R}^m$ by

$$(32) \quad \Delta_{c_\sharp, d_\sharp}(y) = \left[\underset{f \in H}{\operatorname{argmin}} \quad c_\sharp\|Rf\|^2 + d_\sharp\|S''(y - \Lambda f)\|^2 \quad \text{s.to } S'\Lambda f = S'y \right]$$

(interpreted via continuity in case $c_\sharp = 0$ or $d_\sharp = 0$), then the linear map $Q \circ \Delta_{c_\sharp, d_\sharp}$ provides an optimal recovery map.

Proof. With the change of optimization variables $a = c\varepsilon^2$ and $b = d\eta^2$, the program (9) for $\tilde{\Lambda}$, \tilde{Q} , \tilde{R} , and \tilde{S} becomes

$$\underset{c, d \geq 0}{\text{minimize}} \quad c\varepsilon^2 + d\eta^2 \quad \text{s.to} \quad c\|Rf\|^2 + d\|S''e\|^2 \geq \|Qf\|^2 \quad \text{when } f \in H, e \in \ker(S') \text{ satisfy } \Lambda f + e = 0.$$

The form of the program (31) is obtained by eliminating $e \in \mathbb{R}^m$ from the above via $e = -\Lambda f$ and noticing that $e \in \ker(S')$ means that $S'\Lambda f = 0$, i.e., $f \in \ker(S'\Lambda)$. The constrained regularization map is now to be replaced by

$$y \in \mathbb{R}^m \mapsto \left[\operatorname{argmin}_{f \in H, e \in \ker(S')} c_{\sharp} \|Rf\|^2 + d_{\sharp} \|S''e\|^2 \quad \text{s.to } \Lambda f + e = y \right].$$

The form of the program (32) is obtained by eliminating $e \in \mathbb{R}^m$ from the above via $e = y - \Lambda f$ and noticing that $e \in \ker(S')$ means that $S'\Lambda f = S'y$. \square

It is worth making this result more explicit for our motivating example where $e \in \mathbb{R}^m$ is decomposed as $e = (e', e'') \in \mathbb{R}^{m'} \times \mathbb{R}^{m''}$ and we have $S'e = e'$, $S''e = e''$. The observation process on the object of interest $f \in H$ satisfying $\|Rf\| \leq \varepsilon$ is itself decomposed as $y' = \Lambda'f$ and $y'' = \Lambda''f + e''$, where $\Lambda' : H \rightarrow \mathbb{R}^{m'}$, $\Lambda'' : H \rightarrow \mathbb{R}^{m''}$ are linear maps and where $\|e''\| \leq \eta$. In this case, a linear optimal recovery map is obtained, maybe somewhat intuitively, via the constrained regularization

$$\operatorname{minimize}_{f \in H} c_{\sharp} \|Rf\|^2 + d_{\sharp} \|y'' - \Lambda''f\|^2 \quad \text{s.to } \Lambda'f = y'.$$

Our more significant contribution consists in uncovering a principled way of selecting the parameters $c_{\sharp}, d_{\sharp} \geq 0$, namely as solutions to the program

$$\operatorname{minimize}_{c, d \geq 0} c\varepsilon^2 + d\eta^2 \quad \text{s.to } c\|Rf\|^2 + d\|\Lambda''f\|^2 \geq \|Qf\|^2 \quad \text{for all } f \in \ker(\Lambda').$$

4 One intricate consequence: recovery from ℓ_1 -inaccurate data

In this final section, we contemplate yet another scenario of optimal recovery from inaccurate data which borrows from the results of Section 2. The situation is more delicate than in Section 3, though, because the observation error is not modeled through an ℓ_2 -bound but an ℓ_1 -bound. Thus, the objects of interest f from a Hilbert space H are acquired via inaccurate linear observations of the form $y = \Lambda f + e \in \mathbb{R}^m$, where the model set for f and the uncertainty set for e are given relative to some parameter $\varepsilon > 0$ and $\eta > 0$ by

$$(33) \quad \mathcal{K} = \{f \in H : \|Rf\| \leq \varepsilon\},$$

$$(34) \quad \mathcal{E} = \{e \in \mathbb{R}^m : \|e\|_1 \leq \eta\}.$$

Towards the goal of optimally estimating a Hilbert-valued linear quantity of interest $Q : H \rightarrow Z$, the worst-case error of a recovery map $\Delta : \mathbb{R}^m \rightarrow Z$ is defined as

$$(35) \quad \operatorname{Err}_{Q, \mathcal{K}, \mathcal{E}}(\Lambda, \Delta) = \sup_{\substack{\|Rf\| \leq \varepsilon \\ \|e\|_1 \leq \eta}} \|Qf - \Delta(\Lambda f + e)\|.$$

We will reveal that, conditionally on a checkable sufficient condition, the radius of information can still be computed and a constrained-regularization-based optimal recovery map—turning out, perhaps surprisingly, to be linear—can still be constructed efficiently. The sufficient condition is not vacuous: numerically, it even appears to hold whenever η is small enough. Unfortunately, we were not able to establish this fact theoretically.

4.1 Main result

The result presented below involves constrained regularization maps $\Delta_{c,d}^{(j)}$ defined, for $j = 1, \dots, m$ and for $c, d > 0$, by

$$(36) \quad \Delta_{c,d}^{(j)} : y \in \mathbb{R}^m \mapsto \left[\operatorname{argmin}_{f \in H} c\|Rf\|^2 + d\|y - \Lambda f\|^2 \quad \text{s.to } \lambda_i(f) = y_i \text{ for } i \neq j \right] \in H,$$

with the usual interpretation when $c = 0$ or $d = 0$. These constrained regularization maps are linear. Indeed, as a consequence of Lemma 16 in the appendix, they are given by

$$\Delta_{c,d}^{(j)} = \Lambda^\dagger - [cR_{\mathcal{N}_j}^* R_{\mathcal{N}_j} + d\Lambda_{\mathcal{N}_j}^* \Lambda_{\mathcal{N}_j}]^{-1} (cR_{\mathcal{N}_j}^* R \Lambda^\dagger), \quad \mathcal{N}_j = \bigcap_{i \neq j} \ker(\lambda_i).$$

For each $j = 1, \dots, m$, fixing from now on an element $u_j \in H$ such that⁴ $\Lambda u_j = e_j$ (e.g. $u_j = \Lambda^\dagger e_j$), we compute

$$(37) \quad \text{lb}'_j := \min_{c,d \geq 0} c\varepsilon^2 + d\eta^2 \quad \text{s.to } c\|R(h - \theta u_j)\|^2 + d\theta^2 \geq \|Q(h - \theta u_j)\|^2 \quad \text{for all } h \in \ker \Lambda \text{ and } \theta \in \mathbb{R}$$

and we let $c_j, d_j \geq 0$ denote extremizers of this optimization program. In addition, we consider (and we shall compute some of) the quantities $M_{i,j}$ defined for $i, j = 1, \dots, m$ by

$$M_{i,j} := \min_{c,d \geq 0} c\varepsilon^2 + d\eta^2 \\ \text{s.to } c\|R(h - \theta u_i)\|^2 + d\theta^2 \geq \|Q(h - \theta u_i) - Q\Delta_{c_j,d_j}^{(j)} \Lambda h\|^2 \quad \text{for all } h \in H \text{ and } \theta \in \mathbb{R}.$$

The main result of this section can now be stated as follows.

Theorem 10. Aiming at estimating a Hilbert-valued linear map $Q : H \rightarrow Z$ from the observation map $\Lambda : \mathbb{H} \rightarrow \mathbb{R}^m$ under the hyperellipsoid model set (33) and the ℓ_1 -uncertainty set (34), assume that

$$(38) \quad M_{i,k} \leq M_{k,k} \quad \text{for all } i = 1, \dots, m, \quad \text{where } k := \operatorname{argmax}_{j=1,\dots,m} \text{lb}'_j.$$

Then the square of the radius of information is equal to lb'_k and the linear map $Q \circ \Delta_{c_k,d_k}^{(k)} : \mathbb{R}^m \rightarrow Z$ provides an optimal recovery map.

⁴In this section, the notation e_j does not represent the j th entry of the error vector, but the j th element of the canonical basis for \mathbb{R}^m .

The proof of this result is given in the next subsection. Before getting there, we reiterate that the sufficient condition (38) seems to occur whenever η is small enough, as supported by the numerical experiments presented in the reproducible files accompanying this article.

4.2 Justification

Much like the proof of Theorem 1, the proof of Theorem 10 is divided into three separate lemmas: one which establishes lower bounds for the radius of information, one which indicates that each lower bound is achieved by an associated constrained regularization map, and one that establishes a key property of such constrained regularization maps. Finally, these three ingredients will be put together while incorporating the sufficient condition (38). Throughout the argument, it will be convenient to work with the linear maps Γ , $Q^{(1)}, \dots, Q^{(m)}$, $R^{(1)}, \dots, R^{(m)}$, and S defined for $g = (h, \theta)$ in the extended Hilbert space $H \times \mathbb{R}$ via

$$(39) \quad \Gamma(g) = \Lambda h, \quad Q^{(j)}(g) = Q(h - \theta u_j), \quad R^{(j)}(g) = (1/\varepsilon)R(h - \theta u_j), \quad S(g) = (1/\eta)\theta.$$

Let us now state the first of our series of three lemmas.

Lemma 11. The squared global worst-case error of any recovery map Δ satisfies

$$\text{Err}_{Q, \mathcal{K}, \varepsilon}(\Lambda, \Delta)^2 \geq \max_{j=1, \dots, m} \text{lb}_j(\Delta) \geq \max_{j=1, \dots, m} \text{lb}'_j,$$

where the lower bounds are expressed as

$$(40) \quad \text{lb}_j(\Delta) = \sup_{\substack{\|Rf\| \leq \varepsilon \\ |\theta| \leq \eta}} \|Qf - \Delta(\Lambda f + \theta e_j)\|^2$$

$$(41) \quad = \sup_{\substack{\|R^{(j)}g\| \leq 1 \\ \|Sg\| \leq 1}} \|Q^{(j)}g - \Delta(\Gamma g)\|^2,$$

$$(42) \quad \text{lb}'_j = \sup_{\substack{\|R^{(j)}g\| \leq 1 \\ \|Sg\| \leq 1 \\ g \in \ker \Gamma}} \|Q^{(j)}g\|^2$$

$$(43) \quad = \inf_{a, b \geq 0} a + b \quad \text{s.to} \quad a\|R^{(j)}g\|^2 + b\|Sg\|^2 \geq \|Q^{(j)}g\|^2 \quad \text{for all } g \in \ker \Gamma.$$

Proof. For $j = 1, \dots, m$, noticing in (35) that the supremum over all $e \in \mathbb{R}^m$ satisfying $\|e\|_1 \leq \eta$ is larger than or equal to the supremum over all $e = \theta e_j$ with $|\theta| \leq \eta$ leads to the lower bound on $\text{Err}_{Q, \mathcal{K}, \varepsilon}(\Lambda, \Delta)^2$ expressed in (40). To arrive at (41), we make the change of variable $h = f + \theta u_j$, so that

$$\text{lb}_j(\Delta) = \sup_{\substack{\|R(h - \theta u_j)\| \leq \varepsilon \\ |\theta| \leq \eta}} \|Q(h - \theta u_j) - \Delta(\Lambda h)\|^2,$$

and (41) follows by setting $g = (h, \theta) \in H \times \mathbb{R}$ and taking the expressions (39) for Γ , $Q^{(j)}$, $R^{(j)}$, and S into account. At this point, it is apparent that $\text{lb}_j(\Delta)$ coincides with the worst-case error of Δ for the estimation of $Q^{(j)}g$ from Γg under the two-hyperellipsoid-intersection model assumption $\|R^{(j)}g\| \leq 1$ and $\|Sg\| \leq 1$. Thus, the further lower bound (42) and its reformulation (43) follow from an application of Lemma 2. \square

The lemma below, although not used explicitly later, gives us an idea of the coveted optimal recovery map by looking at the case of equality in $\text{lb}_j(\Delta) \geq \text{lb}'_j$. For the latter, note that the expression (43) is seen to coincide with the expression (37) by making the change of optimization variables $a = c\varepsilon^2$, $b = d\eta^2$.

Lemma 12. For each $j = 1, \dots, m$, if $\Delta_{c_j, d_j}^{(j)}$ is the constrained regularization map defined in (36) and its parameters $c_j, d_j \geq 0$ are extremizers of the expression (37) for lb'_j , then

$$\text{lb}_j(Q \circ \Delta_{c_j, d_j}^{(j)}) = \text{lb}'_j.$$

Proof. According to the results of Section 2, we know that equality in $\text{lb}_j(\Delta) \geq \text{lb}'_j$ occurs for $\Delta_b = Q^{(j)} \circ \Delta_{j, a_j, b_j}$, where $a_j, b_j > 0$ are extremizers of the program

$$\underset{a, b \geq 0}{\text{minimize}} \quad a + b \quad \text{s.to} \quad a\|R^{(j)}g\|^2 + b\|Sg\|^2 \geq \|Q^{(j)}g\|^2 \quad \text{for all } g \in \ker \Gamma$$

and where the recovery map $\Delta_{j, a_j, b_j} : \mathbb{R}^m \rightarrow H \times \mathbb{R}$ is defined, for $y \in \mathbb{R}^m$, by

$$\Delta_{j, a_j, b_j}(y) = \left[\underset{g \in H \times \mathbb{R}}{\text{argmin}} \quad a_j\|R^{(j)}g\|^2 + b_j\|Sg\|^2 \quad \text{s.to} \quad \Gamma g = y \right].$$

It now remains to verify that Δ_b agrees with $Q \circ \Delta_{c_j, d_j}^{(j)}$. First, according to the expressions (39) for Γ , $Q^{(j)}$, $R^{(j)}$, and S , and in view of the relations $a = c\varepsilon^2$ and $b = d\eta^2$, it is easily seen that $a_j = c_j\varepsilon^2$ and $b_j = d_j\eta^2$. Then, writing $\Delta_{j, a_j, b_j}(y) = (h_b, \theta_b)$ with $h_b \in H$ and $\theta_b \in \mathbb{R}$, we see that

$$(h_b, \theta_b) = \left[\underset{(h, \theta) \in H \times \mathbb{R}}{\text{argmin}} \quad c_j\|R(h - \theta u_j)\|^2 + d_j\theta^2 \quad \text{s.to} \quad \Lambda h = y \right],$$

and consequently, making the change $f = h - \theta u_j$, we deduce that

$$(h_b - \theta_b u_j, \theta_b) = \left[\underset{(f, \theta) \in H \times \mathbb{R}}{\text{argmin}} \quad c_j\|Rf\|^2 + d_j\theta^2 \quad \text{s.to} \quad \Lambda f + \theta e_j = y \right].$$

Since the constraint $\Lambda f + \theta e_j = y$ decomposes as $\lambda_j(f) + \theta = y_j$ and $\lambda_i(f) = y_i$ for $i \neq j$, we obtain, after eliminating θ ,

$$\begin{aligned} h_b - \theta_b u_j &= \left[\underset{f \in H}{\text{argmin}} \quad c_j\|Rf\|^2 + d_j(y_j - \lambda_j(f))^2 \quad \text{s.to} \quad \lambda_i(f) = y_i \text{ for } i \neq j \right] \\ &= \Delta_{c_j, d_j}^{(j)}(y), \end{aligned}$$

where that last equality is simply the definition of $\Delta_{c_j, d_j}^{(j)}$. The remaining justification is settled by remarking that $\Delta_b(y) = Q^{(j)}(\Delta_{j, a_j, b_j}(y)) = Q^{(j)}((h_b, \theta_b)) = Q(h_b - \theta_b u_j) = Q(\Delta_{c_j, d_j}^{(j)}(y))$. \square

The third lemma is a step towards the determination of the worst-case error of the constrained regularization map $\Delta_{c_j, d_j}^{(j)}$.

Lemma 13. For each $j = 1, \dots, m$, one has

$$\sup_{\substack{\|Rf\| \leq \varepsilon \\ |\theta| \leq \eta}} \|Qf - Q\Delta_{c_j, d_j}^{(j)}(\Lambda f + \theta e_j)\|^2 = \text{lb}'_j.$$

Proof. Setting $g = (f + \theta u_j, \theta) \in H \times \mathbb{R}$, the quantity under consideration becomes

$$\sup_{\substack{\|R^{(j)}g\| \leq 1 \\ \|Sg\| \leq 1}} \|Q^{(j)}g - Q\Delta_{c_j, d_j}^{(j)}(\Gamma g)\|^2 = \sup_{\substack{\|R^{(j)}g\| \leq 1 \\ \|Sg\| \leq 1}} \|Q^{(j)}g - Q^{(j)}\Delta_{j, a_j, b_j}(\Gamma g)\|^2,$$

where we have borrowed from the previous proof the observation that $Q \circ \Delta_{c_j, d_j}^{(j)} = Q^{(j)} \circ \Delta_{j, a_j, b_j}$ with $a_j = c_j \varepsilon^2$ and $b_j = d_j \eta^2$. Thus, our quantity appears to be the worst-case error of $Q^{(j)} \circ \Delta_{j, a_j, b_j}$ for the estimation of $Q^{(j)}g$ from Γg given that $\|R^{(j)}g\| \leq 1$ and $\|Sg\| \leq 1$. We know from the results of Section 2 that the latter is equal to $a_j + b_j = c_j \varepsilon^2 + d_j \eta^2$, i.e., to lb'_j , as announced. \square

Having these three lemmas at our disposal, we now turn to the justification of the main result of this section.

Proof of Theorem 10. According to Lemma 11, there holds

$$\inf_{\Delta: \mathbb{R}^m \rightarrow Z} \text{Err}_{Q, \mathcal{K}, \mathcal{E}}(\Lambda, \Delta)^2 \geq \text{lb}'_k,$$

where we recall that the index k is obtained as the maximizer of lb'_j over all $j = 1, \dots, m$. In order to prove our result, we have to show that this infimum is actually achieved for the linear recovery map $Q \circ \Delta_{c_k, d_k}^{(k)}$. To this end, we notice that the linearity of $Q \circ \Delta_{c_k, d_k}^{(k)}$ guarantees that

$$\begin{aligned} \text{Err}_{Q, \mathcal{K}, \mathcal{E}}(\Lambda, Q \circ \Delta_{c_k, d_k}^{(k)})^2 &= \sup_{\substack{\|Rf\| \leq \varepsilon \\ \|e\|_1 \leq \eta}} \|Qf - Q\Delta_{c_k, d_k}^{(k)}(\Lambda f + e)\|^2 \\ &= \max_{i=1, \dots, m} \sup_{\substack{\|Rf\| \leq \varepsilon \\ |\theta| \leq \eta}} \|Qf - Q\Delta_{c_k, d_k}^{(k)}(\Lambda f + \theta e_i)\|^2 \\ &= \max_{i=1, \dots, m} \sup_{\substack{\|R(h - \theta u_i)\| \leq \varepsilon \\ |\theta| \leq \eta}} \|Q(h - \theta u_i) - Q\Delta_{c_k, d_k}^{(k)}\Lambda h\|^2. \end{aligned}$$

It has become familiar, relying on Polyak's S-procedure, to transform the latter supremum into

$$\inf_{c, d \geq 0} c\varepsilon^2 + d\eta^2 \quad \text{s.t. } c\|R(h - \theta u_i)\|^2 + d\theta^2 \geq \|Q(h - \theta u_i) - Q\Delta_{c_k, d_k}^{(k)}\Lambda h\|^2 \text{ for all } h \in H \text{ and } \theta \in \mathbb{R},$$

which we recognize as the quantity $M_{i,k}$. Now, calling upon the sufficient condition (38), we derive that

$$\text{Err}_{Q,\mathcal{K},\mathcal{E}}(\Lambda, Q \circ \Delta_{c_k, d_k}^{(k)})^2 = \max_{i=1, \dots, m} M_{i,k} = M_{k,k} = \text{lb}'_k,$$

where the last step was due to Lemma 13. This equality completes the proof. \square

4.3 Side result

When the sufficient condition (38) fails, it is not anymore guaranteed that the linear map $Q \circ \Delta_{c_k, d_k}^{(k)}$ provides an optimal recovery map. Regardless, we can always solve a semidefinite program to obtain a linear recovery map with minimal worst-case error, according to the result stated below with the notation introduced in (39).

Proposition 14. The squared worst-case error of a linear recovery maps $\Delta^{\text{lin}} : \mathbb{R}^m \rightarrow Z$ can be computed as the optimal value of a semidefinite program, namely as

$$(44) \quad \text{Err}_{Q,\mathcal{K},\mathcal{E}}(\Lambda, \Delta^{\text{lin}})^2 = \min_{\substack{\gamma \in \mathbb{R} \\ a_1, b_1, \dots, a_m, b_m \geq 0}} \gamma \quad \text{s.to} \quad \left[\begin{array}{c|c} \text{Id} & Q^{(i)} - \Delta^{\text{lin}} \Gamma \\ \hline (Q^{(i)} - \Delta^{\text{lin}} \Gamma)^* & a_i R^{(i)*} R^{(i)} + b_i S^* S \end{array} \right] \succeq 0$$

and $a_i + b_i \leq \gamma$ for all $i = 1, \dots, m$.

This quantity can further be minimized over all linear maps $\Delta^{\text{lin}} : \mathbb{R}^m \rightarrow Z$, yielding a linear recovery map with smallest worst-case error.

Proof. For a linear recovery map $\Delta^{\text{lin}} : \mathbb{R}^m \rightarrow Z$, we have

$$(45) \quad \begin{aligned} \text{Err}_{Q,\mathcal{K},\mathcal{E}}(\Lambda, \Delta^{\text{lin}})^2 &= \max_{i=1, \dots, m} \sup_{\substack{\|Rf\| \leq \varepsilon \\ |\theta| \leq \eta}} \|Qf - \Delta^{\text{lin}}(\Lambda f + \theta e_i)\|^2 \\ &= \inf_{\gamma \in \mathbb{R}} \gamma \quad \text{s.to} \quad \sup_{\substack{\|Rf\| \leq \varepsilon \\ |\theta| \leq \eta}} \|Qf - \Delta^{\text{lin}}(\Lambda f + \theta e_i)\|^2 \leq \gamma \text{ for all } i = 1, \dots, m. \end{aligned}$$

Note that the above i -dependent suprema can also be expressed as

$$\begin{aligned} \sup_{\substack{\|R(h - \theta u_i)\| \leq \varepsilon \\ |\theta| \leq \eta}} \|Q(h - \theta u_i) - \Delta^{\text{lin}} \Lambda h\|^2 &= \sup_{\substack{\|R^{(i)} g\| \leq 1 \\ \|Sg\| \leq 1}} \|Q^{(i)} g - \Delta^{\text{lin}} \Gamma g\|^2 \\ &= \inf_{a_i, b_i \geq 0} a_i + b_i \quad \text{s.to} \quad a_i \|R^{(i)} g\|^2 + b_i \|Sg\|^2 \geq \|Q^{(i)} g - \Delta^{\text{lin}} \Gamma g\|^2 \text{ for all } g \in H \times \mathbb{R} \\ &= \inf_{a_i, b_i \geq 0} a_i + b_i \quad \text{s.to} \quad a_i R^{(i)*} R^{(i)} + b_i S^* S \succeq (Q^{(i)} - \Delta^{\text{lin}} \Gamma)^* (Q^{(i)} - \Delta^{\text{lin}} \Gamma). \end{aligned}$$

Therefore, the i -dependent constraint in (45) is equivalent to the existence of $a_i, b_i \geq 0$ such that $a_i + b_i \leq \gamma$ and $a_i R^{(i)*} R^{(i)} + b_i S^* S \succeq (Q^{(i)} - \Delta^{\text{lin}} \Gamma)^* (Q^{(i)} - \Delta^{\text{lin}} \Gamma)$. As such, we arrive at

$$\text{Err}_{Q, \mathcal{K}, \mathcal{E}}(\Lambda, \Delta^{\text{lin}})^2 = \inf_{\substack{\gamma \in \mathbb{R} \\ a_1, b_1, \dots, a_m, b_m \geq 0}} \gamma \quad \text{s.to} \quad a_i R^{(i)*} R^{(i)} + b_i S^* S \succeq (Q^{(i)} - \Delta^{\text{lin}} \Gamma)^* (Q^{(i)} - \Delta^{\text{lin}} \Gamma) \\ \text{and} \quad a_i + b_i \leq \gamma \quad \text{for all } i = 1, \dots, m.$$

Using Schur complements, the above i -dependent semidefinite constraints can each be rephrased as

$$\left[\begin{array}{c|c} \text{Id} & Q^{(i)} - \Delta^{\text{lin}} \Gamma \\ \hline (Q^{(i)} - \Delta^{\text{lin}} \Gamma)^* & a_i R^{(i)*} R^{(i)} + b_i S^* S \end{array} \right] \succeq 0,$$

leading to $\text{Err}_{Q, \mathcal{K}, \mathcal{E}}(\Lambda, \Delta^{\text{lin}})^2$ being expressed as in (44). We finally note that the linear dependence on Δ^{lin} of the constraints in (44) allows us to further view the minimization of $\text{Err}_{Q, \mathcal{K}, \mathcal{E}}(\Lambda, \Delta^{\text{lin}})^2$ over all linear maps Δ^{lin} as a semidefinite program. \square

We should remark that, even when η is small and the sufficient condition (38) holds, the linear recovery map with smallest worst-case error obtained by semidefinite programming may differ from the optimal recovery map $Q \circ \Delta_{c_k, d_k}^{(k)}$, illustrating the nonuniqueness of optimal recovery maps. Moreover, when η is not small, our numerical experiments (available from the reproducible files) suggest that $Q \circ \Delta_{c_k, d_k}^{(k)}$ may not be optimal among linear recovery maps anymore.

Appendix

In this appendix, we provide justifications for a few facts not fully explained in the main text.

Polyak's S-procedure. Given quadratic functions q_0, q_1, \dots, q_n , the statement $q_0(x) \leq 0$ whenever $q_1(x) \leq 0, \dots, q_n(x) \leq 0$ holds if there exists $a_1, \dots, a_n \geq 0$ such that $q_0 \leq a_1 q_1 + \dots + a_n q_n$. The following result, paraphrased from [Polyak, 1998, Theorem 4.1] establishes that this sufficient condition is also necessary when $n = 2$ and the q_i 's contain no linear terms.

Theorem 15. Suppose that $N \geq 3$ and that quadratic functions q_0, q_1, q_2 on \mathbb{R}^N take the form $q_i(x) = \langle A_i x, x \rangle + \alpha_i$ for symmetric matrices $A_0, A_1, A_2 \in \mathbb{R}^{N \times N}$ and scalars $\alpha_0, \alpha_1, \alpha_2 \in \mathbb{R}$. Then

$$[q_0(x) \leq 0 \text{ whenever } q_1(x) \leq 0 \text{ and } q_2(x) \leq 0] \iff [\text{there exist } a_1, a_2 \geq 0 : q_0 \leq a_1 q_1 + a_2 q_2],$$

provided $q_1(\tilde{x}) < 0$ and $q_2(\tilde{x}) < 0$ for some $\tilde{x} \in \mathbb{R}^N$ (strict feasibility) and $b_1 A_1 + b_2 A_2 \succ 0$ for some $b_1, b_2 \in \mathbb{R}$ (positive definiteness).

As established in [Contino, Fongi, and Muro, 2022, Proposition 5.2], such a result remains valid when \mathbb{R}^N is replaced by an arbitrary Hilbert space H —even of infinite dimension—and the A_i 's are self-adjoint bounded linear operators on H . This generalized version is the one called upon in the main text.

Constrained Regularization. The goal here is to justify the identities (7) and (8), which are consequences of the general observation below.

Lemma 16. Let $A : H \rightarrow H'$ and $B : H \rightarrow H''$ be two bounded linear maps between Hilbert spaces. Assume that there exists $\delta > 0$ such that $\|Az\| \geq \delta\|z\|$ for all $z \in \mathcal{B} := \ker(B)$, so that $A_{\mathcal{B}}^* A_{\mathcal{B}} : \mathcal{B} \rightarrow \mathcal{B}$ is invertible, where $A_{\mathcal{B}} : \mathcal{B} \rightarrow H'$ denotes the restriction of A to \mathcal{B} . Given $a \in H'$ and $b \in H''$, the solution $x^\sharp \in H$ to

$$\underset{x \in H}{\text{minimize}} \|Ax - a\|^2 \quad \text{s.to} \quad Bx = b$$

can be expressed, for any \bar{x} such that $B\bar{x} = b$, as

$$x^\sharp = \bar{x} - [A_{\mathcal{B}}^* A_{\mathcal{B}}]^{-1} A_{\mathcal{B}}^* (A\bar{x} - a).$$

Proof. Writing the optimization variable $x \in H$ as $x = \bar{x} - z$ with $z \in \mathcal{B}$ and the minimizer x^\sharp as $x^\sharp = \bar{x} - z^\sharp$ with $z^\sharp \in \mathcal{B}$, we see that z^\sharp is solution to

$$\underset{z \in \mathcal{B}}{\text{minimize}} \|A\bar{x} - a - Az\|^2.$$

This solution is characterized by the orthogonality condition $\langle A\bar{x} - a - Az^\sharp, Az \rangle = 0$ for all $z \in \mathcal{B}$, which is equivalent to $A_{\mathcal{B}}^* (A\bar{x} - a - Az^\sharp) = 0$, or to $A_{\mathcal{B}}^* A_{\mathcal{B}} z^\sharp = A_{\mathcal{B}}^* (A\bar{x} - a)$. Left-multiplying by $[A_{\mathcal{B}}^* A_{\mathcal{B}}]^{-1}$ to obtain z^\sharp and substituting into $x^\sharp = \bar{x} - z^\sharp$ yields the announced result. \square

It follows as a consequence that, if R_1, \dots, R_n are Hilbert-valued bounded linear maps defined on H such that there exists $\delta > 0$ with $\max\{\|R_1 z\|, \dots, \|R_n z\|\} \geq \delta\|z\|$ for all $z \in \mathcal{N} := \ker(\Lambda)$ ⁵ and if $c_1, \dots, c_n > 0$, then, for any $y \in \mathbb{R}^m$,

$$\begin{aligned} (46) \quad \Delta_{c_1, \dots, c_n}(y) &:= \left[\underset{x \in H}{\text{argmin}} \sum_{i=1}^n c_i \|R_i x\|^2 \quad \text{s.to} \quad \Lambda x = y \right] \\ &= \Lambda^\dagger y - \left[\sum_{i=1}^n c_i R_{i, \mathcal{N}}^* R_{i, \mathcal{N}} \right]^{-1} \left(\sum_{i=1}^n c_i R_{i, \mathcal{N}}^* R_i \right) \Lambda^\dagger y. \end{aligned}$$

To arrive at this identity, which reduces to (7) when $n = 2$, it suffices to apply Lemma 16 with

$$A = \begin{bmatrix} \sqrt{c_1} R_1 \\ \vdots \\ \sqrt{c_n} R_n \end{bmatrix}, \quad a = 0, \quad B = \Lambda, \quad b = y, \quad \bar{x} = \Lambda^\dagger y.$$

⁵This assumption simply reduces to $\ker(R_1) \cap \dots \cap \ker(R_n) \cap \ker(\Lambda) = \{0\}$ when H is finite dimensional.

Furthermore, if $y = \Lambda x$ for some $x \in H$, taking $\bar{x} = x$ instead of $\bar{x} = \Lambda^\dagger y$ leads, after rearrangement, to

$$x - \Delta_{c_1, \dots, c_n} \Lambda x = \left[\sum_{i=1}^n c_i R_{i, \mathcal{N}}^* R_{i, \mathcal{N}} \right]^{-1} \left(\sum_{i=1}^n c_i R_{i, \mathcal{N}}^* R_i \right) x.$$

The latter reduces to (8) when $n = 2$.

Finally, we want to justify the statement made in Section 2 that $\Delta_{a,b}(y)$ converges weakly to $\Delta_{a,0}(y)$ as $b \rightarrow 0$ for any fixed $y \in \mathbb{R}^m$. We shall do so under the working assumption that there exists $\delta > 0$ such that $\max\{\|Rz\|, \|Sz\|\} \geq \delta\|z\|$ for all $z \in \mathcal{N} = \ker(\Lambda)$ ⁶. Supposing without loss of generality that $a = 1$, we thus want to establish that

$$x_b := \operatorname{argmin}_{x \in H} [\|Rx\|^2 + b\|Sx\|^2 \quad \text{s.t. } \Lambda x = y] \xrightarrow{b \rightarrow 0} x_0 := \operatorname{argmin}_{x \in H} [\|Sx\|^2 \quad \text{s.t. } \Lambda x = y, Rx = 0].$$

If this was not the case, there would exist $v \in H$, $\varepsilon > 0$, and a sequence $(b_k)_{k \geq 1}$ decreasing to zero such that $|\langle x_{b_k} - x_0, v \rangle| \geq \varepsilon$ for each $k \geq 1$. Now, from the optimality property of x_{b_k} , we have $\|Rx_{b_k}\|^2 + b_k\|Sx_{b_k}\|^2 \leq \|Rx_0\|^2 + b_k\|Sx_0\|^2$, which yields, in view of $Rx_0 = 0$,

$$\|Rx_{b_k}\|^2 \leq b_k\|Sx_0\|^2 \quad \text{and} \quad \|Sx_{b_k}\|^2 \leq \|Sx_0\|^2.$$

Thanks to our working assumption, it follows that the sequence $(x_{b_k} - x_0)_{k \geq 1}$ of elements in \mathcal{N} is bounded, and then so is the sequence $(x_{b_k})_{k \geq 1}$. As such, it possesses a subsequence weakly converging to some $\tilde{x} \in H$, say. We still write $(x_{b_k})_{k \geq 1}$ for this subsequence and we note that $|\langle \tilde{x} - x_0, v \rangle| \geq \varepsilon$. Next, in view of $\Lambda x_k = y$ for all $k \geq 1$, we derive that $\Lambda \tilde{x} = y$ from $x_{b_k} \rightharpoonup \tilde{x}$. From there, we also obtain $Rx_{b_k} \rightharpoonup R\tilde{x}$ and $Sx_{b_k} \rightharpoonup S\tilde{x}$, and in turn $\|R\tilde{x}\| \leq \liminf \|Rx_{b_k}\| = 0$ and $\|S\tilde{x}\| \leq \liminf \|Sx_{b_k}\| = \|Sx_0\|$. These facts imply that \tilde{x} is also a minimizer for the program defining x_0 , so that $\tilde{x} = x_0$ by uniqueness of the minimizer. This is of course incompatible with $|\langle \tilde{x} - x_0, v \rangle| \geq \varepsilon$ and provides the required contradiction.

References

- A. Anandkumar, K. Azizzadenesheli, K. Bhattacharya, N. Kovachki, Z. Li, B. Liu, and A. Stuart. Neural operator: graph kernel network for partial differential equations. In *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*, 2020.
- P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Data assimilation in reduced modeling. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):1–29, 2017.
- M. Contino, G. Fongi, and S. Muro. Polyak’s theorem on Hilbert spaces. *Optimization*, pages 1–15, 2022.

⁶This assumption reduces to $\ker(R) \cap \ker(S) \cap \ker(\Lambda) = \{0\}$ when H is finite dimensional.

- M. Etehad and S. Foucart. Instances of computational optimal recovery: dealing with observation errors. *SIAM/ASA Journal on Uncertainty Quantification*, 9(4):1438–1456, 2021.
- S. Foucart. Instances of computational optimal recovery: refined approximability models. *Journal of Complexity*, 62:101503, 2021.
- S. Foucart. *Mathematical Pictures at a Data Science Exhibition*. Cambridge University Press, 2022.
- S. Foucart. Full recovery from point values: an optimal algorithm for Chebyshev approximability prior. *Advances in Computational Mathematics*, 49(4):57, 2023.
- S. Foucart and C. Liao. Optimal recovery from inaccurate data in Hilbert spaces: regularize, but what of the parameter? *Constructive Approximation*, pages 1–32, 2022.
- S. Foucart, C. Liao, and N. Veldt. On the optimal recovery of graph signals. In *2023 International Conference on Sampling Theory and Applications (SampTA)*, pages 1–5, 2023. doi: 10.1109/SampTA59647.2023.10301205.
- M. Grant and S. Boyd. CVX: MATLAB software for disciplined convex programming, version 2.1. <http://cvxr.com/cvx>, March 2014.
- D. Long, N. Mrvaljevic, S. Zhe, and B. Hosseini. A kernel approach for PDE discovery and operator learning. *arXiv preprint arXiv:2210.08140*, 2023.
- G. G. Magaril-Il’yaev and K. Yu. Osipenko. Optimal recovery of functionals based on inaccurate data. *Math. Notes*, 50(6):1274–1279, 1991.
- G. G. Magaril-Il’yaev, K. Yu. Osipenko, and V. M. Tikhomirov. On optimal recovery of heat equation solutions. *Approximation theory: a volume dedicated to B. Bojanov*, pages 163–175, 2004.
- A. A. Melkman and C. A. Micchelli. Optimal estimation of linear operators in Hilbert spaces from inaccurate data. *SIAM Journal on Numerical Analysis*, 16(1):87–105, 1979.
- C. A. Micchelli. Optimal estimation of linear operators from inaccurate data: a second look. *Numerical Algorithms*, 5(8):375–390, 1993.
- C. A. Micchelli and T. J. Rivlin. A survey of optimal recovery. In *Optimal Estimation in Approximation Theory*, pages 1–54. Springer, 1977.
- E. Novak and H. Woźniakowski. *Tractability of Multivariate Problems: Linear Information*. European Mathematical Society, 2008.
- E. W. Packel. Do linear problems have linear optimal algorithms? *SIAM Review*, 30(3):388–403, 1988.
- L. Plaskota. *Noisy Information and Computational Complexity*. Cambridge University Press, 1996.

- B. T. Polyak. Convexity of quadratic transformations and its use in control and optimization. *Journal of Optimization Theory and Applications*, 99(3):553–583, 1998.
- W. Rudin. *Functional Analysis*. McGraw-Hill, second edition, 1991.
- S.A. Smolyak. *Optimal Recovery of Functions and Functionals of Them*. Candidate dissertation, Moscow State University, 1965.
- E. V. Vvedenskaya. On the optimal recovery of a solution of a system of linear homogeneous differential equations. *Differential Equations*, 45(2):262–266, 2009.