Optimal Recovery from Inaccurate Data in Hilbert Spaces: Regularize, but what of the Parameter?

Simon Foucart* and Chunyang Liao — Texas A&M University

Abstract

In Optimal Recovery, the task of learning a function from observational data is tackled deterministically by adopting a worst-case perspective tied to an explicit model assumption made on the functions to be learned. Working in the framework of Hilbert spaces, this article considers a model assumption based on approximability. It also incorporates observational inaccuracies modeled via additive errors bounded in ℓ_2. Earlier works have demonstrated that regularization provide algorithms that are optimal in this situation, but did not fully identify the desired hyperparameter. This article fills the gap in both a local scenario and a global scenario. In the local scenario, which amounts to the determination of Chebyshev centers, the semidefinite recipe of Beck and Eldar (legitimately valid in the complex setting only) is complemented by a more direct approach, with the proviso that the observational functionals have orthonormal representers. In the said approach, the desired parameter is the solution to an equation that can be resolved via standard methods. In the global scenario, where linear algorithms rule, the parameter elusive in the works of Micchelli et al. is found as the byproduct of a semidefinite program. Additionally and quite surprisingly, in case of observational functionals with orthonormal representers, it is established that any regularization parameter is optimal.

Key words and phrases: Regularization, Chebyshev center, semidefinite programming, S-procedure, hyperparameter selection.

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

1 Introduction

1.1 Background on Optimal Recovery

This article is concerned with a central problem in Data Science, namely: a function f is acquired through point evaluations

$$y_i = f(x^{(i)}), \qquad i = 1, \dots, m, \tag{1}$$

*S. F. is supported by grants from the NSF (CCF-1934904, DMS-2053172) and from the ONR (N00014-20-1-2787).

and these data should be used to learn f—or to recover it, with the terminology preferred in this article. Importantly, the evaluation points $x^{(1)}, \dots, x^{(m)}$ are considered fixed entities in our scenario: they cannot be chosen in a favorable way, as in Information-Based Complexity [Novak and Woźniakowski, 2008], nor do they occur as independent realizations of a random variable, as in Statistical Learning Theory [Hastie, Tibshirani, and Friedman, 2009]. In particular, without an underlying probability distribution, the performance of the recovery process cannot be assessed via generalization error. Instead, it is assessed via a notion of worst-case error, central to the theory of Optimal Recovery [Micchelli and Rivlin, 1977].

To outline this theory, we make the framework slightly more abstract. Precisely, given a normed space F, the unknown function is replaced by an element $f \in F$. This element is accessible only through *a priori* information expressing an educated belief about f and *a posteriori* information akin to (1). In other words, our partial knowledge about f is summed up via

- the fact that $f \in \mathcal{K}$ for a subset $\mathcal{K}$ of F called a model set;
- the observational data $y_i = \lambda_i(f)$, $i = 1, \dots, m$, for some linear functionals $\lambda_1, \dots, \lambda_m \in F^*$ making up the observation map $\Lambda : g \in F \mapsto [\lambda_1(g); \dots; \lambda_m(g)] \in \mathbb{R}^m$.

We wish to approximate f by some $\widehat{f} \in F$ produced using this partial knowledge of f. Since the error $\|f - \widehat{f}\|$ involves the unknown f, which is only accessible via $f \in \mathcal{K}$ and $\Lambda(f) = y$, we take a worst-case perspective leading to the local worst-case error

$$\operatorname{lwce}(y, \widehat{f}) := \sup_{\substack{f \in \mathcal{K} \\ \Lambda(f) = y}} \|f - \widehat{f}\|. \tag{2}$$

Our objective consists in finding an element $\widehat{f}$ that minimizes $\operatorname{lwce}(y, \widehat{f})$. Such an $\widehat{f}$ can be described, almost tautologically, as a center of a smallest ball containing $\mathcal{K} \cap \Lambda^{-1}(\{y\})$. It is called a Chebyshev center of this set of model- and data-consistent elements. This remark, however, does not come with any practical construction of a Chebyshev center.

The term local was used above to make a distinction with the global worst-case error of a recovery map $\Delta (= \Delta_{\mathcal{K}}) : \mathbb{R}^m \to F$, defined as

$$\operatorname{gwce}(\Delta) := \sup_{y \in \Lambda(\mathcal{K})} \operatorname{lwce}(y, \Delta(y)) = \sup_{f \in \mathcal{K}} \|f - \Delta(\Lambda(f))\|. \tag{3}$$

The minimal value of $\operatorname{gwce}(\Delta)$ is called the intrinsic error (of the observation map Λ over the model set $\mathcal{K}$) and the maps Δ that achieve this minimal value are called globally optimal recovery maps. Our objective consists in constructing such maps—of course, the map that assigns to y a Chebyshev center of $\mathcal{K} \cap \Lambda^{-1}(\{y\})$ is one of them, but it may be impractical. By contrast, for model sets that are convex and symmetric, the existence of linear maps among the set of globally optimal recovery maps is guaranteed by fundamental results from Optimal Recovery in at least two settings: when

F is a Hilbert space and when F is an arbitrary normed space but the full recovery of f gives way to the recovery of a quantity of interest $Q(f)$, Q being a linear functional. We refer the readers to [Foucart, To appear, Chapter 9] for details.

1.2 The specific problem

The problem solved in this article is a quintessential Optimal Recovery problem—its specificity lies in the particular model set and in the incorporation of errors in the observation process. The underlying normed space F is a Hilbert space and is therefore denoted by H from now on. Reproducing kernel Hilbert spaces, whose usage is widespread in Data Science [Schölkopf and Smola, 2002], are of particular interest as point evaluations of type (1) make perfect sense there.

Concerning the model set, we concentrate on an approximation-based choice that is increasingly scrutinized, see e.g. [Maday, Patera, Penn, and Yano, 2015], [DeVore, Petrova, and Wojtaszczyk, 2017] and [Cohen, Dahmen, Mula, and Nichols, 2020]. Depending on a linear subspace $\mathcal{V}$ of H and on a parameter $\varepsilon > 0$, it takes the form

$$\mathcal{K} = \{f \in H : \operatorname{dist}(f, \mathcal{V}) \le \varepsilon\}.$$

Binev, Cohen, Dahmen, DeVore, Petrova, and Wojtaszczyk [2017] completely solved the Optimal Recovery problem with exact data in this situation (locally and globally). Precisely, they showed that the solution $\widehat{f}$ to

$$\underset{f \in H}{\text{minimize}} \ \operatorname{dist}(f, \mathcal{V}) \qquad \text{s.to} \quad \Lambda(f) = y, \tag{4}$$

which clearly belongs to the model- and data-consistent set $\mathcal{K} \cap \Lambda^{-1}(\{y\})$, turns out to be its Chebyshev center. Moreover, with $P_{\mathcal{V}}$ and $P_{\mathcal{V}^\perp}$ denoting the orthogonal projectors onto $\mathcal{V}$ and onto the orthogonal complement $\mathcal{V}^\perp$ of $\mathcal{V}$, the fact that $\operatorname{dist}(f, \mathcal{V}) = \|f - P_{\mathcal{V}} f\| = \|P_{\mathcal{V}^\perp} f\|$ makes the optimization program (4) tractable. It can actually be seen that $\Delta : y \mapsto \widehat{f}$ is a linear map. This is a significant advantage because Δ can then be precomputed in an offline stage knowing only $\mathcal{V}$ and Λ and the program (4) need not be solved afresh for each new data $y \in \mathbb{R}^m$ arriving in an online stage.

Concerning the observation process, instead of exact data $y = \Lambda(f) \in \mathbb{R}^m$, it is now assumed that

$$y = \Lambda(f) + e \in \mathbb{R}^m$$

for some unknown error vector $e \in \mathbb{R}^m$. This error vector is not modeled as random noise but through the deterministic ℓ_2-bound $\|e\|_2 \le \eta$. Although other ℓ_p-norms can the considered for the optimal recovery of $Q(f_0)$ when Q is a linear functional on an arbitrary normed space F (see [Ettehad and Foucart, 2021]), here the arguments rely critically on $\mathbb{R}^m$ being endowed with the

ℓ_2-norm. It will be written simply as $\|\cdot\|$ below, hoping that it does not create confusion with the Hilbert norm on H.

For our specific problem, the worst-case recovery errors (2) and (3) need to be adjusted. The local worst-case recovery error at y for $\widehat{f}$ becomes

$$\mathrm{lwce}(y,\widehat{f}) = \sup_{\substack{\|P_{\mathcal{V}^\perp} f\| \le \varepsilon \\ \|\Lambda(f)-y\| \le \eta}} \|f - \widehat{f}\|.$$

As for the global worst-case error of $\Delta : \mathbb{R}^m \to H$, it reads

$$\mathrm{gwce}(\Delta) = \sup_{\substack{\|P_{\mathcal{V}^\perp} f\| \le \varepsilon \\ \|e\| \le \eta}} \|f - \Delta(\Lambda(f)+e)\|.$$

Note that both worst-case errors are infinite if one can find a nonzero h in $\mathcal{V} \cap \ker(\Lambda)$. Indeed, the element $f_t := f + th$, $t \in \mathbb{R}$, obeys $\|P_{\mathcal{V}^\perp} f_t\| = \|P_{\mathcal{V}^\perp} f\| \le \varepsilon$ and $\|y - \Lambda(f_t)\| = \|y - \Lambda(f)\| \le \eta$, so for instance $\mathrm{lwce}(y,\widehat{f}) \ge \sup_{t\in\mathbb{R}} \|f_t - \widehat{f}\| = +\infty$. Thus, we always make the assumption that

$$\mathcal{V} \cap \ker(\Lambda) = \{0\}. \tag{5}$$

We keep in mind that the latter forces $n := \dim(\mathcal{V}) \le m$, as can be seen by dimension arguments. With Λ^* denoting the Hermitian adjoint of Λ, another assumption that we sometimes make reads

$$\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}. \tag{6}$$

This is not extremely stringent: assuming the surjectivity of Λ is quite natural, otherwise certain observations need not be collected; then the map Λ can be preprocessed into another map $\widetilde{\Lambda}$ satisfying $\widetilde{\Lambda}\widetilde{\Lambda}^* = \mathrm{Id}_{\mathbb{R}^m}$ by setting $\widetilde{\Lambda} = (\Lambda\Lambda^*)^{-1/2}\Lambda$. Incidentally, if $u_1, \dots, u_m \in H$ represent the Riesz representers of the observation functionals $\lambda_1, \dots, \lambda_m \in H^*$, characterized by $\langle u_i, f\rangle = \lambda_i(f)$ for all $f \in H$, then the assumption (6) is equivalent to the orthonormality of the system $(u_1, \dots, u_m)$. In a reproducing kernel Hilbert space with kernel K, if the λ_i's are point evaluations at some $x^{(i)}$'s, so that $u_i = K(\cdot, x^{(i)})$, then (6) is equivalent to $K(x^{(i)}, x^{(j)}) = \delta_{i,j}$ for all $i, j = 1, \dots, m$. This occurs e.g. for the Paley–Wiener space of functions with Fourier transform supported on $[-\pi, \pi]$ when the evaluations points come from an integer grid, since the kernel is given by $K(x, x') = \mathrm{sinc}(\pi(x - x'))$, $x, x' \in \mathbb{R}$.

1.3 Main results

There are previous works on Optimal Recovery in Hilbert spaces in the presence of observation error bounded in ℓ_2. Notably, [Beck and Eldar, 2007] dealt with the local setting, while [Melkman and Micchelli, 1979] and [Micchelli, 1993] dealt with the global setting. These works underline the importance of regularization, which is prominent in many other settings [Chen and Haykin,

2002]. They establish that the optimal recovery maps are obtained by solving the unconstrained program

$$\underset{f\in H}{\text{minimize}}\ (1-\tau)\|P_{\mathcal{V}^\perp} f\|^2 + \tau\|\Lambda f - y\|^2 \tag{7}$$

for some $\tau \in [0,1]$. It is the precise choice of this regularization parameter τ which is the purpose of this article. Assuming from now on that H is finite dimensional[1], we provide a complete (almost) picture of the local and global Optimal Recovery solutions, as summarized in the four points below, three of them being new:

L1. With H restricted here to be a complex Hilbert space, the Chebyshev center of the set $\{f \in H : \|P_{\mathcal{V}^\perp} f\| \le \varepsilon, \|\Lambda f - y\| \le \eta\}$ is the minimizer of (7) for the choice $\tau = d_\sharp/(c_\sharp + d_\sharp)$, where $c_\sharp, d_\sharp$ are solutions to the semidefinite program[2]

$$\underset{c,d,t\ge 0}{\text{minimize}}\ \varepsilon^2 c + (\eta^2 - \|y\|^2)d + t \qquad \text{s.to} \qquad cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda \succeq \mathrm{Id},$$

$$\text{and} \qquad \left[\begin{array}{c|c} cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda & -d\Lambda^* y \\ \hline -d(\Lambda^* y)^* & t \end{array}\right] \succeq 0.$$

L2. Under the orthonormal observations assumption (6) but without the above restriction on H, the Chebyshev center of the set $\{f \in H : \|P_{\mathcal{V}^\perp} f\| \le \varepsilon, \|\Lambda f - y\| \le \eta\}$ is the minimizer of (7) for the choice τ that satisfies

$$\lambda_{\min}((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda) = \frac{(1-\tau)^2\varepsilon^2 - \tau^2\eta^2}{(1-\tau)\varepsilon^2 - \tau\eta^2 + (1-\tau)\tau(1-2\tau)\delta^2}, \tag{8}$$

where δ is precomputed as $\delta = \min\{\|P_{\mathcal{V}^\perp} f\| : \Lambda f = y\} = \min\{\|\Lambda f - y\| : f \in \mathcal{V}\}$. For the distinct case $\mathcal{V} = \{0\}$, the best choice of parameter is more simply $\tau = \max\{1 - \eta/\|y\|, 0\}$.

G1. A globally optimal recovery map is provided by the linear map sending $y \in \mathbb{R}^m$ to the minimizer of (7) with parameter $\tau = d_\flat/(c_\flat + d_\flat)$, where $c_\flat, d_\flat$ are solutions to the semidefinite program

$$\underset{c,d}{\text{minimize}}\ \varepsilon^2 c + \eta^2 d \qquad \text{s.to} \quad cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda \succeq \mathrm{Id}. \tag{9}$$

G2. Under the orthonormal observations assumption (6), the linear map sending $y \in \mathbb{R}^m$ to the minimizer of (7) is a globally optimal recovery map for any choice of parameter $\tau \in [0,1]$.

[1] It is likely that the results are still valid in the infinite-dimensional case. But then it is unclear how to solve (8) and (9) numerically, so the infinite-dimensional case is not given proper scrutiny in the rest of the article.

[2] In the statement of this semidefinite program and elsewhere, the notation $T \succeq 0$ means that an operator T is positive semidefinite on H, i.e., that $\langle Tf, f\rangle \ge 0$ for all $f \in H$.

Before entering the technicalities, a few comments are in order to put these results in context. Item **L1** is the result of [Beck and Eldar, 2007] (see Corollary 3.2 there) adapted to our situation. It relies on an extension of the S-lemma involving two quadratic constraints. This extension is valid in the complex finite-dimensional setting, but not necessarily in the real setting, hence the restriction on H (this does not preclude the validity of the result in the real setting, though). It is worth pointing out the nonlinearity of the map that sends $y \in \mathbb{R}^m$ to the above Chebyshev center. Incidentally, we can safely talk about the Chebyshev center, because it is known [Garkavi, 1962] that a bounded set in a uniformly convex Banach space has exactly one Chebyshev center. A sketch of the argument adapted to our situation is presented in the appendix.

For item **L2**, working with an observation map Λ satisfying $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$ allows us to construct the Chebyshev center even in the setting of a real Hilbert space. This is possible because our argument does not rely on the extension of the S-lemma—it just uses the obvious implication. As for equation (8), it is easily solved using the bisection method or the Newton/secant method. Moreover, it gives some insight on the value of the optimal parameter τ. For instance, the proof reveals that τ is always between $1/2$ and $\varepsilon/(\varepsilon+\eta)$. When $\varepsilon \geq \eta$, say, the optimal parameter should then satisfy $\tau \geq 1/2$, which is somewhat intuitive: $\varepsilon \geq \eta$ means that there is more model mismatch than data mismatch, so the regularization should penalize model fidelity less than data fidelity by taking $1-\tau \leq \tau$, i.e., $\tau \geq 1/2$. As an aside, we point out that, here too, the map that sends $y \in \mathbb{R}^m$ to the Chebyshev center is not a linear map—if it was, then the optimal parameter should be independent of y.

In contrast, the globally optimal recovery map of item **G1** is linear. It is one of several globally optimal recovery maps, since the locally optimal one (which is nonlinear) is also globally optimal. However, as revealed in the reproducible[3] accompanying this article, it is in general the only regularization map that turns out to be globally optimal. The fact that regularization produces globally optimal recovery maps was recognized by Micchelli, who wrote in the abstract of [Micchelli, 1993] that "*the regularization parameter must be chosen with care*". However, a recipe for selecting the parameter was not given there, except on a specific example. The closest to a nonexhaustive search is found in [Plaskota, 1996, Lemma 2.6.2] for the case $\mathcal{V} = \{0\}$, but even this result does not translate into a numerically tractable recipe. The selection stemming from (9) does, at least when H is finite-dimensional, which is assumed here. Semidefinite programs can indeed be solved in MATLAB using CVX [Grant and Boyd, 2014] and in Python using CVXPY [Diamond and Boyd, 2016].

Finally, a surprise arises in item **G2**. Working with an observation map Λ satisfying $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$, the latter indeed reveals that *the regularization parameter does not need to be chosen with care* after all, since regularization maps are globally optimal no matter how the parameter $\tau \in [0,1]$ is chosen. The precise interpretation of the choices $\tau = 0$ and $\tau = 1$ will be elucidated later.

[3] MATLAB and Python files illustrating the findings of this article are located at `https://github.com/foucart/COR`.

The rest of this article is organized as follows. Section 2 gathers some auxiliary results that are used in the proofs of the main results. Section 3 elucidates item **L1** and establishes item **L2**—in other words, it is concerned with local optimality. Section 4, which is concerned with global optimality, is the place where items **G1** and **G2** are proved. Lastly, a short appendix containing some side information is included after the bibliography.

2 Technical Preparation

This section establishes (or recalls) a few results that we isolate here in order not to disrupt the flow of subsequent arguments.

2.1 S-lemma and S-procedure

Loosely speaking, the S-procedure is a relaxation technique expressing the fact that a quadratic inequality is a consequence of some quadratic constraints. In case of a single quadratic constraint, the relaxation turns out to be exact. This result, known as the S-lemma, can be stated as follows: given quadratic functions q_0 and q_1 defined on $\mathbb{K}^N$, with $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$,

$$[q_0(x) \le 0 \text{ whenever } q_1(x) \le 0] \iff [\text{there exists } a \ge 0 : q_0 \le a q_1],$$

provided $q_1(\widetilde{x}) < 0$ for some $\widetilde{x} \in \mathbb{K}^N$. With more than one quadratic constraint, $q_1, \ldots, q_k$, say, $q_0(x) \le 0$ whenever $q_1(x) \le 0, \ldots, q_k(x) \le 0$ is still a consequence of $q_0 \le a_1 q_1 + \cdots + a_k q_k$ for some $a_1, \ldots, a_k \ge 0$, but the reverse implication does not hold anymore. There is a subtlety when $k = 2$, as the reverse implication holds for $\mathbb{K} = \mathbb{C}$ but not for $\mathbb{K} = \mathbb{R}$, see [Pólik and Terlaky, 2007, Section 3]. However, if the quadratic constraints do not feature linear terms, then the reverse implication holds for $k = 2$ also when $\mathbb{K} = \mathbb{R}$. Since this result of [Polyak, 1998, Theorem 4.1] is to be invoked later, we state it formally below.

Theorem 1. Suppose that $N \ge 3$ 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

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

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

2.2 Regularization

In this subsection, we take a closer look at the regularization program (7). The result below shows that its solution depends linearly on $y \in \mathbb{R}^m$. In fact, the result covers a slightly more general

program and the linearity claim follows by taking $R = P_{\mathcal{V}^\perp}$, $r = 0$, $S = \Lambda$, and $s = y$.

Proposition 2. Let R, S be linear maps from H into other Hilbert spaces containing points r, s, respectively. For $\tau \in (0,1)$, the optimization program

$$\underset{f \in H}{\text{minimize}} \ (1-\tau)\|Rf - r\|^2 + \tau\|Sf - s\|^2 \tag{10}$$

has solutions $f_\tau \in H$ characterized by

$$\big((1-\tau)R^*R + \tau S^*S\big) f_\tau = (1-\tau)R^*r + \tau S^*s. \tag{11}$$

Moreover, if $\ker(R) \cap \ker(S) = \{0\}$, then f_τ is uniquely given by

$$f_\tau = \big((1-\tau)R^*R + \tau S^*S\big)^{-1}\big((1-\tau)R^*r + \tau S^*s\big). \tag{12}$$

Proof. The program (10) can be interpreted as a standard least squares problem, namely as

$$\underset{f \in H}{\text{minimize}} \left\| \begin{bmatrix} \sqrt{1-\tau}R \\ \sqrt{\tau}S \end{bmatrix} f - \begin{bmatrix} \sqrt{1-\tau}r \\ \sqrt{\tau}s \end{bmatrix} \right\|^2.$$

According to the normal equations, its solutions f_τ are characterized by

$$\begin{bmatrix} \sqrt{1-\tau}R \\ \sqrt{\tau}S \end{bmatrix}^* \begin{bmatrix} \sqrt{1-\tau}R \\ \sqrt{\tau}S \end{bmatrix} f_\tau = \begin{bmatrix} \sqrt{1-\tau}R \\ \sqrt{\tau}S \end{bmatrix}^* \begin{bmatrix} \sqrt{1-\tau}r \\ \sqrt{\tau}s \end{bmatrix},$$

which is a rewriting of (11). Next, if $\ker(R) \cap \ker(S) = \{0\}$, then

$$\langle ((1-\tau)R^*R + \tau S^*S)f, f\rangle = (1-\tau)\|Rf\|^2 + \tau\|Sf\|^2 \geq 0,$$

with equality only possible when $f \in \ker(R) \cap \ker(S)$, i.e., $f = 0$. This shows that $(1-\tau)R^*R + \tau S^*S$ is positive definite, and hence invertible, which allows us to write (12) as a consequence of (11). □

The expression (12) is not always the most convenient one. Under extra conditions on R and S, we shall see that f_τ, $\tau \in [0,1]$, can in fact be expressed as the convex combination $f_\tau = (1-\tau)f_0 + \tau f_1$. The elements f_0 and f_1 should be interpreted[4] as

$$f_0 = \underset{f \in H}{\operatorname{argmin}} \|Sf - s\| \qquad \text{s.to} \ \ Rf = r,$$
$$f_1 = \underset{f \in H}{\operatorname{argmin}} \|Rf - r\| \qquad \text{s.to} \ \ Sf = s.$$

[4]Intuitively, the solution to the program (10) written as the minimization of $\|Rf - r\|^2 + (\tau/(1-\tau))\|Sf - s\|^2$ becomes, as $\tau \to 1$, the mininizer of $\|Rf - r\|^2$ subject to $\|Sf - s\|^2 = 0$. This explains the interpretation of f_1. A similar argument explains the interpretation of f_0.

The requirements that $r \in \mathrm{ran}(R)$ and $s \in \mathrm{ran}(S)$ need to be imposed for f_0 and f_1 to even exist and the condition $\ker(R) \cap \ker(S) = \{0\}$ easily guarantees that f_0 and f_1 are unique. They obey

$$Rf_0 = r, \quad S^*(Sf_0 - s) \in \ker(R)^\perp, \quad Sf_1 = s, \quad R^*(Rf_1 - r) \in \ker(S)^\perp. \tag{13}$$

For instance, the identity $Rf_0 = r$ reflects the constraint in the optimization program defining f_0, while $S^*(Sf_0 - s) \in \ker(R)^\perp$ is obtained by expanding $\|S(f_0 + tu) - s\|^2 \geq \|S(f_0) - s\|^2$ around $t = 0$ for any $u \in \ker(R)$. At this point, we are ready to establish our claim under extra conditions on R and S, namely that they are orthogonal projections. These conditions will be in place when the observation map satisfies $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$. Indeed, in view of $\|w\|^2 = \langle w, \Lambda\Lambda^* w\rangle = \|\Lambda^* w\|^2$ for any $w \in \mathbb{R}^m$, the regularization program (7) also reads

$$\underset{f \in H}{\text{minimize}} \ (1-\tau)\|P_{\mathcal{V}^\perp} f\|^2 + \tau\|\Lambda^*\Lambda f - \Lambda^* y\|^2,$$

where both $P_{\mathcal{V}^\perp}$ and $\Lambda^*\Lambda$ are orthogonal projections. The result below will then be applied with $R = P_{\mathcal{V}^\perp}$, $r = 0$, $S = \Lambda^*\Lambda$, and $s = \Lambda^* y$.

Proposition 3. Let R, S be two orthogonal projectors on H such that $\ker(R) \cap \ker(S) = \{0\}$ and let $r \in \mathrm{ran}(R)$, $s \in \mathrm{ran}(S)$. For $\tau \in [0,1]$, the solution f_τ to the optimization program

$$\underset{f \in H}{\text{minimize}} \ (1-\tau)\|Rf - r\|^2 + \tau\|Sf - s\|^2 \tag{14}$$

satisfies

$$f_\tau = (1-\tau) f_0 + \tau f_1. \tag{15}$$

Moreover, one has

$$\|Rf_\tau - r\| = \tau\|f_1 - f_0\| \qquad \text{and} \qquad \|Sf_\tau - s\| = (1-\tau)\|f_1 - f_0\|. \tag{16}$$

Proof. Taking the extra conditions on R and S into account, the identities (13) read

$$Rf_0 = r, \quad Sf_0 - s \in \mathrm{ran}(R), \quad Sf_1 = s, \quad Rf_1 - r \in \mathrm{ran}(S).$$

In this form, they imply that

$$\begin{aligned} \langle S(f_0 - f_1), (R-S)(f_0 - f_1)\rangle &= \langle Sf_0 - s, R(f_0 - f_1)\rangle - \langle Sf_0 - s, S(f_0 - f_1)\rangle \\ &= \langle Sf_0 - s, f_0 - f_1\rangle - \langle Sf_0 - s, f_0 - f_1\rangle \\ &= 0. \end{aligned} \tag{17}$$

In a similar fashion, by exchanging the roles of R and S, and consequently also of f_0 and f_1, we have $\langle R(f_1 - f_0), (S-R)(f_1 - f_0)\rangle = 0$, i.e., $\langle R(f_0 - f_1), (R-S)(f_0 - f_1)\rangle = 0$. Subtracting (17) from the latter yields $\|(R-S)(f_0 - f_1)\|^2 = 0$, in other words $R(f_0 - f_1) = S(f_0 - f_1)$. Then, the

element $h := f_0 - f_1 - R(f_0 - f_1) = f_0 - f_1 - S(f_0 - f_1)$ belongs to $\ker(R) \cap \ker(S)$, so that $h = 0$. In summary, we have established that

$$R(f_0 - f_1) = S(f_0 - f_1) = f_0 - f_1. \tag{18}$$

From here, we can deduce the two parts of the proposition. For the first part, we notice that

$$\begin{aligned}\big((1-\tau)R + \tau S\big)\big((1-\tau)f_0 + \tau f_1\big) &= (1-\tau)^2 Rf_0 + (1-\tau)\tau(Rf_1 + Sf_0) + \tau^2 Sf_1 \\ &= (1-\tau)^2 Rf_0 + (1-\tau)\tau(Rf_0 + Sf_1) + \tau^2 Sf_1 \\ &= (1-\tau)Rf_0 + \tau Sf_1 \\ &= (1-\tau)r + \tau s,\end{aligned}$$

which shows that $(1-\tau)f_0 + \tau f_1$ satisfies the relation (11) characterizing the minimizer f_τ of (14), so that $f_\tau = (1-\tau)f_0 + \tau f_1$, as announced in (15). For the second part, we notice that

$$Rf_\tau - r = (1-\tau)Rf_0 + \tau Rf_1 - Rf_0 = \tau R(f_1 - f_0) = \tau(f_1 - f_0),$$

so the first equality of (16) follows by taking the norm. The second equality of (16) is derived in a similar fashion. □

We complement Proposition 3 with a few additional pieces of information.

Remark. Under the assumptions of Proposition 3, the solution f_τ to (14) is also solution to

$$\underset{f \in H}{\text{minimize}} \;\; \max\big\{(1-\tau)\|Rf - r\|, \tau\|Sf - s\|\big\}.$$

Indeed, at $f = f_\tau$, the squared objective function equals $(1-\tau)^2\tau^2\|f_1 - f_0\|^2$, while at an arbitrary $f \in H$, it satisfies

$$\begin{aligned}\max\big\{(1-\tau)^2\|Rf - r\|^2, \tau^2\|Sf - s\|^2\big\} &\geq \frac{1}{\dfrac{1}{1-\tau} + \dfrac{1}{\tau}}\big((1-\tau)\|Rf - r\|^2 + \tau\|Sf - s\|^2\big) \\ &\geq (1-\tau)\tau\big((1-\tau)\|Rf_\tau - r\|^2 + \tau\|Sf_\tau - s\|^2\big) \\ &= (1-\tau)\tau\big((1-\tau)\tau^2\|f_1 - f_0\|^2 + \tau(1-\tau)^2\|f_1 - f_0\|^2\big) \\ &= (1-\tau)^2\tau^2\|f_1 - f_0\|^2.\end{aligned}$$

In the case $R = P_{\mathcal{V}^\perp}$, $r = 0$, $S = \Lambda^*\Lambda$, and $s = \Lambda^* y$, the choice $\tau = \varepsilon/(\varepsilon + \eta)$ is quite relevant, since the above optimization program becomes equivalent to

$$\underset{f \in H}{\text{minimize}} \;\; \max\left\{\frac{1}{\varepsilon}\|P_{\mathcal{V}^\perp} f\|, \frac{1}{\eta}\|\Lambda f - y\|\right\}.$$

Its solution is clearly in the model- and data-consistent set $\{f \in H : \|P_{\mathcal{V}^\perp} f\| \leq \varepsilon, \|\Lambda f - y\| \leq \eta\}$. In fact, this could have been a natural guess for its Chebyshev center, but item **L2** reveals the invalidity of such a guess. Nonetheless, the special parameter $\tau = \varepsilon/(\varepsilon+\eta)$ will make a reappearance in the argument leading to item **L2**.

Remark. The proof of Proposition 3 showcased the important identities $Rf_0 = r$, $Sf_1 = s$, and $R(f_0 - f_1) = S(f_0 - f_1) = f_0 - f_1$. In the case $R = P_{\mathcal{V}^\perp}$, $r = 0$, $S = \Lambda^*\Lambda$, and $s = \Lambda^* y$, if Δ_τ denotes the recovery map assigning to $y \in \mathbb{R}^m$ the solution f_τ to the regularization program (7), these identities read, when $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$,

$$P_{\mathcal{V}^\perp}\Delta_0 = 0, \qquad \Lambda^*\Lambda\Delta_1 = \Lambda^*, \qquad P_{\mathcal{V}^\perp}(\Delta_0 - \Delta_1) = \Lambda^*\Lambda(\Delta_0 - \Delta_1) = \Delta_0 - \Delta_1. \tag{19}$$

Remark. Considering again the case $R = P_{\mathcal{V}^\perp}$, $r = 0$, $S = \Lambda^*\Lambda$, and $s = \Lambda^* y$, Proposition 3 implies that $f_\tau \in \mathcal{V} + \mathrm{ran}(\Lambda^*)$ for any $\tau \in [0,1]$, given that the latter holds for $\tau = 0$ and for $\tau = 1$. For $\tau = 0$, this is because the constraint $P_{\mathcal{V}^\perp} f = 0$ of the optimization program defining f_0 imposes $f_0 \in \mathcal{V}$. For $\tau = 1$, this is a result established e.g. in [Foucart, Liao, Shahrampour, and Wang, 2020, Theorem 2]. The said result also provides an efficient way to compute the solution f_τ of (7) even when H is infinite dimensional, as stated in the appendix.

3 Local Optimality

Our goal in this section is to determine locally optimal recovery maps. In other words, the section is concerned with Chebyshev centers. We start by considering the situation of an arbitrary observation map Λ, but with a restriction on the space H. Next, lifting this restriction on H, we refine the result in the particular case of an observation map satisfying $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$.

3.1 Arbitrary observations

In this subsection, we reproduce a result from [Beck and Eldar, 2007], albeit with different notation, and explain how it implies the statement of item **L1**. The result in question, namely Corollary 3.2, relies on the S-procedure with two constraints, and as such cannot be claimed in the real setting.

Theorem 4. Let H be a complex Hilbert space. Let R, S be two linear maps from H into other Hilbert spaces containing points r, s, respectively. Suppose the existence of $\widetilde{f} \in H$ such that $\|R\widetilde{f} - r\| < \varepsilon$ and $\|S\widetilde{f} - s\| < \eta$ and the existence of $\tau \in [0,1]$ such that $(1-\tau)R^*R + \tau S^*S$ is positive definite. Then the Chebyshev center of $\{f \in H : \|Rf - r\| \le \varepsilon, \|Sf - s\| \le \eta\}$ equals $f_\sharp = \big(c_\sharp R^*R + d_\sharp S^*S\big)^{-1}(c_\sharp R^* r + d_\sharp S^* s)$, where $c_\sharp, d_\sharp$ are solutions to

$$\underset{c,d,t \ge 0}{\text{minimize}} \ (\varepsilon^2 - \|r\|^2)c + (\eta^2 - \|s\|^2)d + t \qquad \text{s.to} \quad cR^*R + dS^*S \succeq \mathrm{Id},$$

$$\text{and} \quad \left[\begin{array}{c|c} cR^*R + dS^*S & -cR^*r - dS^*s \\ \hline -c(R^*r)^* - d(S^*s)^* & t \end{array}\right] \succeq 0.$$

The statement made in item **L1** is of course derived by taking $R = P_{\mathcal{V}^\perp}$, $r = 0$, $S = \Lambda$, and $s = y$. Theorem 4 is indeed applicable, as $\widetilde{f} = (f_0 + f_1)/2$ satisfies the strict feasibility conditions, while the positive definiteness condition is not only fulfilled for some $\tau \in [0,1]$, but for all $\tau \in (0,1)$, since $\langle((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda)f, f\rangle = (1-\tau)\|P_{\mathcal{V}^\perp} f\|^2 + \tau\|\Lambda f\|^2 \geq 0$, with equality only possible if $f \in \mathcal{V} \cap \ker(\Lambda)$, i.e., if $f = 0$ thanks to the assumption (5). We also note that, by virtue of (12), the element $f_\sharp$ defined above is nothing else than the regularized solution with parameter $\tau = d_\sharp/(c_\sharp + d_\sharp)$.

3.2 Orthonormal observations

In this subsection, we place ourselves in the situation of an observation map satisfying $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$ and we provide a proof of the statements made item **L2**. In fact, we prove some slightly more general results and **L2** follows by taking $R = P_{\mathcal{V}^\perp}$, $r = 0$, $S = \Lambda^*\Lambda$, and $s = \Lambda^* y$. Note that we must separate the cases where $R = \mathrm{Id}$ (corresponding to $\mathcal{V} = \{0\}$) and where R is a proper orthogonal projection (corresponding to $\mathcal{V} \neq \{0\}$). We emphasize that, in each of these two cases, the optimal parameter $\tau_\sharp$ is not independent of y. Therefore, in view of (15) and of the linear dependence of f_0 and f_1 on y, the regularized solution $f_{\tau_\sharp}$ does not depend linearly on y. In other words, the locally optimal recovery map is not a linear map. The following two simple lemmas will be used to deal with both cases.

Lemma 5. Let R, S be two linear maps from H into other Hilbert spaces containing points r, s, respectively. Given $f_\sharp \in H$, let

$$h_\sharp \in \operatorname*{argmax}_{h \in H} \|h\| \qquad \text{s.to} \begin{cases} \|Rf_\sharp - r + Rh\| \leq \varepsilon, \\ \|Sf_\sharp - s + Sh\| \leq \eta. \end{cases}$$

If the orthogonality conditions

$$\langle R^*(Rf_\sharp - r), h_\sharp\rangle = 0 \qquad \text{and} \qquad \langle S^*(Sf_\sharp - s), h_\sharp\rangle = 0 \tag{20}$$

are fulfilled, then $f_\sharp$ is the Chebyshev center of the set $\{f \in H : \|Rf - r\| \leq \varepsilon, \|Sf - s\| \leq \eta\}$, i.e., for any $g \in H$,

$$\sup_{\substack{\|Rf-r\|\leq\varepsilon \\ \|Sf-s\|\leq\eta}} \|f - g\| \geq \sup_{\substack{\|Rf-r\|\leq\varepsilon \\ \|Sf-s\|\leq\eta}} \|f - f_\sharp\|. \tag{21}$$

Proof. First, writing $f = f_\sharp + h$, we easily see that the right-hand side of (21) reduces to $\|h_\sharp\|$. Second, let us remark that the orthogonality conditions guarantee that $f_\pm := f_\sharp \pm h_\sharp$ both satisfy $\|Rf_\pm - r\| \leq \varepsilon$ and $\|Sf_\pm - s\| \leq \eta$. For instance, we have

$$\|Rf_\pm - r\|^2 = \|Rf_\sharp - r \pm Rh_\sharp\|^2 = \|Rf_\sharp - r\|^2 + \|Rh_\sharp\|^2 = \|Rf_\sharp - r + Rh_\sharp\|^2 \leq \varepsilon^2, \tag{22}$$

where the latter inequality reflects the feasibility of $h_\sharp$. Therefore, the left-hand side of (21) is bounded below by

$$\max_{\pm} \|f_\pm - g\| \geq \frac{1}{2}\big(\|f_+ - g\| + \|f_- - g\|\big) \geq \frac{1}{2}\|(f_+ - g) - (f_- - g)\| = \frac{1}{2}\|2h_\sharp\| = \|h_\sharp\|, \tag{23}$$

i.e., by the right-hand side of (21). □

The next lemma somehow relates to the S-procedure. However, it does not involve the coveted (and usually invalid) equivalence, but only the straightforward implication.

Lemma 6. Let R, S be two linear maps from H into other Hilbert spaces containing points r, s, respectively. Given $f_\sharp \in H$ and $h_\sharp \in H$, suppose that

$$\|Rf_\sharp - r + Rh_\sharp\|^2 = \varepsilon^2 \qquad \text{and} \qquad \|Sf_\sharp - s + Sh_\sharp\|^2 = \eta^2, \tag{24}$$

and that there exist $a, b \geq 0$ such that

$$aR^*R + bS^*S \succeq \mathrm{Id} \tag{25}$$

as well as

$$aR^*(Rf_\sharp - r) + bS^*(Sf_\sharp - s) + (aR^*R + bS^*S)h_\sharp = h_\sharp. \tag{26}$$

Then, one has

$$h_\sharp \in \operatorname*{argmax}_{h \in H} \|h\| \qquad \text{s.to} \begin{cases} \|Rf_\sharp - r + Rh\| \leq \varepsilon, \\ \|Sf_\sharp - s + Sh\| \leq \eta. \end{cases} \tag{27}$$

Proof. By writing the variable in the optimization program (27) as $h = h_\sharp + g$, the constraints on h transform into constraints on g. Thanks to (24), the latter constraints read

$$\langle R^*Rg, g\rangle + 2\langle R^*(Rf_\sharp - r + Rh_\sharp), g\rangle \leq 0 \quad \text{and} \quad \langle S^*Sg, g\rangle + 2\langle S^*(Sf_\sharp - s + Sh_\sharp), g\rangle \leq 0.$$

Combining these constraints—specifically, multiplying the first by a, the second by b, and summing—implies that

$$\begin{aligned} 0 &\geq \langle (aR^*R + bS^*S)g, g\rangle + 2\langle aR^*(Rf_\sharp - r) + bS^*(Sf_\sharp - s) + (aR^*R + bS^*S)h_\sharp, g\rangle \\ &\geq \langle g, g\rangle + 2\langle h_\sharp, g\rangle, \end{aligned}$$

where (25) and (26) were exploited in the last step. In other words, one has $0 \geq \|h_\sharp + g\|^2 - \|h_\sharp\|^2$, i.e., $\|h\|^2 \leq \|h_\sharp\|^2$, under the constraints on h, proving that $h_\sharp$ is indeed a maximizer in (27). □

3.2.1 The case $R = \mathrm{Id}$

As mentioned earlier, the case $R = \mathrm{Id}$ corresponds to the choice $\mathcal{V} = \{0\}$, i.e., to a model set $\mathcal{K}$ being an origin-centered ball in H, and to regularizations being classical Tikhonov regularizations. The arguments are slightly less involved here than for the case $R \neq \mathrm{Id}$. Here is the main result.

Theorem 7. Let S be an orthogonal projector on H with $\ker(S) \neq \{0\}$ and let $r \in H$, $s \in \mathrm{ran}(S)$. The solution $f_{\tau_\sharp}$ to the regularization program (14) with parameter

$$\tau_\sharp = \max\left\{1 - \frac{\eta}{\|Sr - s\|}, 0\right\}$$

is the Chebyshev center of the set $\{f \in H : \|f - r\| \leq \varepsilon, \|Sf - s\| \leq \eta\}$.

Proof. Before separating two cases, we remark that $\|Sr - s\| \leq \varepsilon + \eta$ is implicitly assumed for the above set to be nonempty. Now, we first consider the case $\|Sr - s\| > \eta$. Defining $f_\sharp := f_{\tau_\sharp}$ with $\tau_\sharp = 1 - \eta/\|Sr - s\| \in (0, 1)$, our objective is to find $h_\sharp \in H$ and $a, b \geq 0$ for which conditions (24), (25), and (26) of Lemma 6 are fulfilled, so that $h_\sharp$ is a maximizer appearing in Lemma 5, and then to verify that the orthogonality conditions (20) hold, so that $f_\sharp$ is indeed the required Chebyshev center. We take any $h_\sharp \in \ker(S)$, with a normalization will be decided later, and $a = 1$, $b = \tau_\sharp/(1 - \tau_\sharp)$. In this way, since $R = \mathrm{Id}$, condition (25) is automatic, and condition (26) follows from the characterization (11) written here as $(1 - \tau_\sharp)(f_\sharp - r) = -\tau_\sharp(Sf_\sharp - s)$. This characterization also allows us to deduce (20) only from $\langle Sf_\sharp - s, h_\sharp \rangle = 0$, which holds because the spaces $\mathrm{ran}(S)$ and $\ker(S)$ are orthogonal. The remaining condition (24) now reads $\|f_\sharp - r\|^2 + \|h_\sharp\|^2 = \varepsilon^2$ and $\|Sf_\sharp - s\|^2 = \eta^2$. Recalling from Proposition 3 that $f_\sharp = (1 - \tau_\sharp)f_0 + \tau_\sharp f_1$, while taking into account that $f_0 = r$ here and that $f_1 = f_0 + S(f_1 - f_0) = r + s - Sr$ thanks to (18), we have $f_\sharp - r = \tau_\sharp(s - Sr)$ and $Sf_\sharp - s = -(1 - \tau_\sharp)(s - Sr)$. Thus, condition (24) reads

$$\tau_\sharp^2 \|s - Sr\|^2 + \|h_\sharp\|^2 = \varepsilon^2 \qquad \text{and} \qquad (1 - \tau_\sharp)^2 \|s - Sr\|^2 = \eta^2.$$

The latter is justified by our choice of $\tau_\sharp$, while the former can simply be achieved by normalizing $h_\sharp$, so long as $\varepsilon \geq \tau_\sharp \|s - Sr\|$, i.e., $\varepsilon \geq \|s - Sr\| - \eta$, which is our implicit assumption for nonemptiness of the set under consideration.

Next, we consider the case $\|Sr - s\| \leq \eta$. We note that this implies that r belongs to the set $\{f \in H : \|f - r\| \leq \varepsilon, \|Sf - s\| \leq \eta\}$—we are going to show that r is actually the Chebyshev center of this set. In other words, since $r = f_0$, this means that $f_{\tau_\sharp}$ with $\tau_\sharp = 0$ is the Chebyshev center. To this end, we shall establish that, for any $g \in H$,

$$\sup_{\substack{\|f - r\| \leq \varepsilon \\ \|Sf - s\| \leq \eta}} \|f - g\| \geq \sup_{\substack{\|f - r\| \leq \varepsilon \\ \|Sf - s\| \leq \eta}} \|f - r\|.$$

On the one hand, the right-hand side is obviously bounded above by ε. On the other hand, selecting $h \in \ker(S)$ with $\|h\| = \varepsilon$, we define $f_\pm := r \pm h$ to obtain $\|f_\pm - r\| = \|h\| = \varepsilon$ and $\|Sf_\pm - s\| = \|Sr - s\| \leq \eta$. Thus, the left-hand side is bounded below by

$$\max_\pm \|f_\pm - g\| \geq \frac{1}{2}\|f_+ - g\| + \frac{1}{2}\|f_- - g\| \geq \frac{1}{2}\|(f_+ - g) - (f_- - g)\| = \frac{1}{2}\|2h\| = \varepsilon.$$

This proves that the left-hand side is larger than or equal to the right-hand side, as required. □

3.2.2 The case $R \neq \mathrm{Id}$

We now assume that R is a proper orthogonal projection, i.e., that $R \neq \mathrm{Id}$, which corresponds to the case $\mathcal{V} \neq \{0\}$. The main result is stated below. To apply it in practice, the optimal parameter τ needs to be computed by solving an equation involving the smallest eigenvalue of self-adjoint operator depending on τ. This can be done using an all purpose routine. We could also devise our own bisection method, Newton method (since the derivative $d\lambda_{\min}/d\tau$ is accessible, see the appendix), or secant method.

Theorem 8. Let $R \neq \mathrm{Id}, S \neq \mathrm{Id}$ be two orthogonal projectors on H such that $\ker(R) \cap \ker(S) = \{0\}$ and let $r \in \mathrm{ran}(R)$, $s \in \mathrm{ran}(S)$. Consider $\tau_\sharp$ to be a (often unique) τ between $1/2$ and $\varepsilon/(\varepsilon + \eta)$ such that

$$\lambda_{\min}((1-\tau)R + \tau S) - \frac{(1-\tau)^2\varepsilon^2 - \tau^2\eta^2}{(1-\tau)\varepsilon^2 - \tau\eta^2 + (1-\tau)\tau(1-2\tau)\delta^2} = 0, \tag{28}$$

where δ is precomputed as $\delta = \min\{\|Rf - r\| : Sf = s\} = \min\{\|Sf - s\| : Rf = r\}$. Then the solution $f_{\tau_\sharp}$ of the regularization program (14) with parameter $\tau_\sharp$ is the Chebyshev center of the set $\{f \in H : \|Rf - r\| \leq \varepsilon, \|Sf - s\| \leq \eta\}$.

Remark. It there is no observation error, i.e., if $\eta = 0$, then the parameter solving equation (28) is $\tau_\sharp = 1$. In case $R = P_{\mathcal{V}^\perp}$, $r = 0$, $S = \Lambda^*\Lambda$, and $s = \Lambda^* y$, this means that the Chebyshev center is $f_1 = \mathrm{argmin}\, \|P_{\mathcal{V}^\perp} f\|$ s.to $\Lambda f = y$ and we thus retrieve the result of [Binev, Cohen, Dahmen, DeVore, Petrova, and V 2017].

The proof of Theorem 8 requires an additional result that gives information about the norms of the projections Rh and Sh when h is an eigenvector of the positive semidefinite operator $(1-\tau)R+\tau S$. This result will be applied for the eigenvector associated with the smallest eigenvalue.

Lemma 9. Let R, S be two orthogonal projectors on H. For $\tau \in (0,1)$, let $h \in H$ be an eigenvector of $(1-\tau)R + \tau S$ corresponding to an eigenvalue $\lambda \neq 1/2$. Then

$$\|Rh\|^2 = \frac{(\tau - \lambda)\lambda}{(1-\tau)(1-2\lambda)}\|h\|^2 \qquad \text{and} \qquad \|Sh\|^2 = \frac{(1-\tau-\lambda)\lambda}{\tau(1-2\lambda)}\|h\|^2. \tag{29}$$

Proof. We notice, on the one hand, that

$$(30) \qquad (1-\tau)\|Rh\|^2 + \tau\|Sh\|^2 = (1-\tau)\langle Rh, h\rangle + \tau\langle Sh, h\rangle = \langle((1-\tau)R + \tau S)h, h\rangle \\ = \lambda\|h\|^2,$$

and, on the other hand, that

$$(1-\tau)^2\|Rh\|^2 - \tau^2\|Sh\|^2 = \langle(1-\tau)Rh + \tau Sh, (1-\tau)Rh - \tau Sh\rangle = \langle\lambda h, (1-\tau)Rh - \tau Sh\rangle \\ = \lambda(1-\tau)\|Rh\|^2 - \lambda\tau\|Sh\|^2.$$

Rearranging the latter yields

$$(31) \qquad (1-\tau)(1-\tau-\lambda)\|Rh\|^2 - \tau(\tau-\lambda)\|Sh\|^2 = 0.$$

Together, the equaltions (30) and (31) form a two-by-two linear system in the unknowns $\|Rh\|^2$ and $\|Sh\|^2$ with determinant $-(1-\tau)\tau(1-2\lambda) \neq 0$. Its solutions are easily verified to be the ones given in (29). □

Remark. Because $\|Rh\|^2$, $\|Sh\|^2$, and $\|h\|^2$ are all nonnegative, Lemma 9 implicitly guarantees that $\tau - \lambda$ and $1 - \tau - \lambda$ have the same sign as $1 - 2\lambda \neq 0$. These quantities are nonnegative when $R \neq \mathrm{Id}$, $S \neq \mathrm{Id}$, and λ is the smallest eigenvalue—the case of application of the lemma. Indeed, taking $f \in \ker(R)$ with $\|f\| = 1$ (which is possible because $R \neq \mathrm{Id}$), one has

$$\lambda_{\min} := \lambda_{\min}((1-\tau)R + \tau S) \leq \|(1-\tau)Rf + \tau Sf\| = \tau\|Sf\| \leq \tau,$$

i.e., $\tau - \lambda_{\min} \geq 0$. The inequality $\lambda_{\min} \leq 1 - \tau$, i.e., $1 - \tau - \lambda_{\min} \geq 0$, is obtained in a similar fashion. These inequalities sum up to give $1 - 2\lambda_{\min} \geq 0$. The latter is in fact (strictly) positive when $\tau \neq 1/2$, since either τ or $1 - \tau$ is smaller than $1/2$, so that $\lambda_{\min} < 1/2$.

With the above result at hand, we are ready to fully justify the main result of this subsection.

Proof of Theorem 8. Let us temporarily take for granted the existence of a solution $\tau_\sharp$ to (28). Defining $f_\sharp := f_{\tau_\sharp}$, our objective is again to find $h_\sharp \in H$ and $a, b \geq 0$ for which conditions (24), (25), and (26) of Lemma 6 are fulfilled, so that $h_\sharp$ is a maximizer appearing in Lemma 5, and then to verify that the orthogonality conditions (20) hold, so that $f_\sharp$ is indeed the required Chebyshev center. Writing $\lambda_\sharp := \lambda_{\min}((1-\tau_\sharp)R + \tau_\sharp S)$, we choose $h_\sharp$ to be a (so far unnormalized) eigenvector of $(1-\tau_\sharp)R + \tau_\sharp S$ corresponding to the eigenvalue $\lambda_\sharp$. Setting $a := (1-\tau_\sharp)/\lambda_\sharp$ and $b := \tau_\sharp/\lambda_\sharp$, conditions (25) is swiftly verified, since $RR^* = R$, $SS^* = S$, and

$$aR + bS = \frac{(1-\tau_\sharp)R + \tau_\sharp S}{\lambda_{\min}((1-\tau_\sharp)R + \tau_\sharp S)} \succeq \mathrm{Id}.$$

Then, the characterization $(1-\tau_\sharp)R(f_\sharp - r) = -\tau_\sharp S(f_\sharp - s)$ of the regularization solution $f_\sharp$, see (11), allows us to validate condition (26) via

$$\begin{aligned} aR(f_\sharp - r) + bS(f_\sharp - s) + (aR+bS)h_\sharp &= \frac{1}{\lambda_\sharp}\Big((1-\tau_\sharp)R(f_\sharp - r) + \tau_\sharp S(f_\sharp - s) + ((1-\tau_\sharp)R + \tau_\sharp S)h_\sharp\Big) \\ &= \frac{1}{\lambda_\sharp}\,(0 + \lambda_\sharp h_\sharp) = h_\sharp. \end{aligned}$$

The orthogonality conditions (20) are also swiftly verified: the second one follows from the first one using (11); the first one holds because, while $h_\sharp$ is an eigenvector of $(1-\tau_\sharp)R + \tau_\sharp S$ corresponding to its smallest eigenvalue, $R(f_\sharp - r) = -\tau_\sharp/(1-\tau_\sharp)S(f_\sharp - s)$ is an eigenvector corresponding to the largest eigenvalue (i.e., to one), since it is invariant when applying both R and S. Thus, it remains to verify that the two conditions of (24) are fulfilled. In view of the orthogonality conditions (20), they read

$$\|Rf_\sharp - r\|^2 + \|Rh_\sharp\|^2 = \varepsilon^2 \qquad \text{and} \qquad \|Sf_\sharp - s\|^2 + \|Sh_\sharp\|^2 = \eta^2. \tag{32}$$

Now, invoking Proposition 3, as well as Lemma 9, the two conditions of (24) become

$$\tau_\sharp^2\delta^2 + \frac{(\tau_\sharp - \lambda_\sharp)\lambda_\sharp}{(1-\tau_\sharp)(1-2\lambda_\sharp)}\|h_\sharp\|^2 = \varepsilon^2 \tag{33}$$

$$(1-\tau_\sharp)^2\delta^2 + \frac{(1-\tau_\sharp-\lambda_\sharp)\lambda_\sharp}{\tau_\sharp(1-2\lambda_\sharp)}\|h_\sharp\|^2 = \eta^2. \tag{34}$$

After some simplification work, starting by forming the combinations $(1-\tau_\sharp)\times(33) - \tau_\sharp^2\times(34)$ and $(1-\tau_\sharp-\lambda_\sharp)(1-\tau_\sharp)\times(33) - (\tau_\sharp-\lambda_\sharp)(\tau_\sharp)\times(34)$, these two conditions are seen to be equivalent to

$$\|h_\sharp\|^2 = \frac{1-2\lambda_\sharp}{(2\tau_\sharp - 1)\lambda_\sharp^2}\big((1-\tau_\sharp)^2\varepsilon^2 - \tau_\sharp^2\eta^2\big), \tag{35}$$

$$\lambda_\sharp = \frac{(1-\tau_\sharp)^2\varepsilon^2 - \tau_\sharp^2\eta^2}{(1-\tau_\sharp)\varepsilon^2 - \tau_\sharp\eta^2 + (1-\tau_\sharp)\tau_\sharp(1-2\tau_\sharp)\delta^2}. \tag{36}$$

These two conditions can be fulfilled: the latter is the condition that defined $\tau_\sharp$, i.e., (28), while the former is simply guaranteed by properly normalizing the eigenvector $h_\sharp$.

Before establishing the existence $\tau_\sharp$, we point out that its uniqueness holds when $f_0 \neq f_1$, i.e., when there is no $f \in H$ such that $Rf = r$ and $Sf = s$—such an f would solve the regularization program for any $\tau \in [0,1]$. Indeed, if $\tau \neq \tau'$ were two solutions to (28), then the previous argument would imply that f_τ and $f_{\tau'}$ are both Chebyshev centers, which could only happen if they were equal, i.e., if $f_0 = f_1$ by (15). Now, for the existence of $\tau_\sharp$, it will be justified by the fact that the function

$$\theta : \tau \mapsto \lambda_{\min}((1-\tau)R + \tau S) - \frac{(1-\tau)^2\varepsilon^2 - \tau^2\eta^2}{(1-\tau)\varepsilon^2 - \tau\eta^2 + (1-\tau)\tau(1-2\tau)\delta^2}$$

is continuous between $1/2$ and $\varepsilon/(\varepsilon+\eta)$ and takes values of different signs there. To see the difference in sign, notice that $\lambda_{\min}((1-\tau)R + \tau S) \in [0, 1/2]$ by the remark after Lemma 9—this is

where the assumption $R \neq \mathrm{Id}$ is critical—so that

$$\theta\left(\frac{1}{2}\right) \leq \frac{1}{2} - \frac{1}{2} \leq 0 \qquad \text{and} \qquad \theta\left(\frac{\varepsilon}{\varepsilon+\eta}\right) \geq 0 - 0 \geq 0.$$

To see the continuity, we need the continuity of the smallest eigenvalue as a function of τ and the nonvanishing of the denominator $(1-\tau)\varepsilon^2 - \tau\eta^2 + (1-\tau)\tau(1-2\tau)\delta^2$ between $1/2$ and $\varepsilon/(\varepsilon+\eta)$. The former is a consequence of Weyl's inequality, yielding

$$|\lambda_{\min}((1-\tau)R+\tau S) - \lambda_{\min}((1-\tau')R+\tau' S)| \leq \|((1-\tau)R+\tau S) - ((1-\tau')R+\tau' S)\| = |\tau-\tau'|\,\|R-S\|.$$

The latter is less immediate. We start by using (16) and recalling the very definition of f_τ to write

$$(1-\tau)\tau\delta^2 = (1-\tau)\|Rf_\tau - r\|^2 + \tau\|Sf_\tau - r\|^2 \leq (1-\tau)\varepsilon^2 + \tau\eta^2.$$

Therefore, if the denominator vanished for some $\tau \in (0,1) \setminus \{1/2\}$, we would have

$$\begin{aligned} 0 &= \frac{(1-\tau)\varepsilon^2 - \tau\eta^2}{1-2\tau} + (1-\tau)\tau\delta^2 \leq \frac{(1-\tau)\varepsilon^2 - \tau\eta^2}{1-2\tau} + (1-\tau)\varepsilon^2 + \tau\eta^2 = \frac{2(1-\tau)^2\varepsilon^2 - 2\tau^2\eta^2}{1-2\tau} \\ &= \frac{\big((1-\tau)\varepsilon + \tau\eta\big)\big((1-\tau)\varepsilon - \tau\eta\big)}{1/2-\tau} = \big((1-\tau)\varepsilon + \tau\eta\big)\big(\varepsilon+\eta\big)\frac{\varepsilon/(\varepsilon+\eta) - \tau}{1/2-\tau}. \end{aligned}$$

This would force $\varepsilon/(\varepsilon+\eta) - \tau$ and $1/2 - \tau$ to have the same sign, contrary to the assumption that τ runs between $1/2$ and $\varepsilon/(\varepsilon+\eta)$. Thus, the nonvanishing of the denominator is explained, concluding the proof. □

Remark. The above arguments contain the value of the minimal local worst-case error, i.e., of the Chebyshev radius of the set $\mathcal{C} = \{f \in H : \|Rf - r\| \leq \varepsilon, \|Sf - s\| \leq \eta\}$. Indeed, we recall from the proof of Lemma 5 that this radius equals $\|h_\sharp\|$, whose value was derived in (35). This expression can be simplified with the help of (36) by noticing that

$$\frac{1-2\lambda_\sharp}{\lambda_\sharp} = (2\tau_\sharp - 1)\frac{(1-\tau_\sharp)\varepsilon^2 + \tau_\sharp\eta^2 - (1-\tau_\sharp)\tau_\sharp\delta^2}{(1-\tau_\sharp)^2\varepsilon^2 - \tau_\sharp^2\eta^2}.$$

As a consequence, we deduce that the Chebyshev radius satisfies

$$\mathrm{radius}(\mathcal{C})^2 = \frac{1-\tau_\sharp}{\lambda_\sharp}\varepsilon^2 + \frac{\tau_\sharp}{\lambda_\sharp}\eta^2 - \frac{(1-\tau_\sharp)\tau_\sharp}{\lambda_\sharp}\delta^2, \qquad \lambda_\sharp := \lambda_{\min}((1-\tau_\sharp)R + \tau_\sharp S).$$

4 Global Optimality

Our goal in this section is to uncover some favorable globally optimal recovery maps—favorable in the sense that they are linear maps. We start by considering the situation of an arbitrary observation map Λ before moving to the particular case where it satisfies $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$.

4.1 Arbitrary observations

In this subsection, we first recall a standard lower bound for the global worst-case error. This lower bound, already exploited e.g. in [Micchelli, 1993], shall be expressed as the minimal value of a certain semidefinite program. This expression will allow us to demonstrate that the lower bound is achieved by the regularization map

$$\Delta_\tau : y \in \mathbb{R}^m \mapsto \underset{f \in H}{\operatorname{argmin}} \ (1-\tau)\|P_{\mathcal{V}^\perp} f\|^2 + \tau \|\Lambda f - y\|^2$$

for some parameter $\tau \in (0,1)$ to be explicitly determined. Here is a precise formulation of the result.

Theorem 10. Given the approximability set $\mathcal{K} = \{f \in H : \operatorname{dist}(f, \mathcal{V}) \le \varepsilon\}$ and the uncertainty set $\mathcal{E} = \{e \in \mathbb{R}^m : \|e\| \le \eta\}$, define $\tau_\flat := d_\flat/(c_\flat + d_\flat)$ where $c_\flat, d_\flat \ge 0$ are solutions to

$$\underset{c,d \ge 0}{\text{minimize}} \ c\varepsilon^2 + d\eta^2 \qquad \text{s.to} \quad cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda \succeq \mathrm{Id}.$$

Then the regularization map $\Delta_{\tau_\flat}$ is a globally optimal recovery map over $\mathcal{K}$ and $\mathcal{E}$, i.e.,

$$\operatorname{gwce}(\Delta_{\tau_\flat}) = \inf_{\Delta : \mathbb{R}^m \to H} \operatorname{gwce}(\Delta). \tag{37}$$

The proof relies on three lemmas given below, the first of which introducing the said lower bound.

Lemma 11. For any recovery map $\Delta : \mathbb{R}^m \to H$, one has $\operatorname{gwce}(\Delta) \ge \mathrm{lb}$, where

$$\mathrm{lb} := \sup_{\substack{\|P_{\mathcal{V}^\perp} h\| \le \varepsilon \\ \|\Lambda h\| \le \eta}} \|h\|.$$

Proof. As a reminder, the global worst-case error of Δ is defined by

$$\operatorname{gwce}(\Delta) = \sup_{\substack{\|P_{\mathcal{V}^\perp} f\| \le \varepsilon \\ \|e\| \le \eta}} \|f - \Delta(\Lambda f + e)\|.$$

For any $h \in H$ such that $\|P_{\mathcal{V}^\perp} h\| \le \varepsilon$ and $\|\Lambda h\| \le \eta$, since $f_\pm = \pm h$ satisfies $\|P_{\mathcal{V}^\perp} f_\pm\| \le \varepsilon$ and $e_\pm = \mp \Lambda h$ satisfies $\|e_\pm\| \le \eta$, we have

$$\begin{aligned} \operatorname{gwce}(\Delta) &\ge \max_\pm \|f_\pm - \Delta(\Lambda f_\pm + e_\pm)\| = \max_\pm \|f_\pm - \Delta(0)\| \ge \frac{1}{2}\|f_+ - \Delta(0)\| + \frac{1}{2}\|f_- - \Delta(0)\| \\ &\ge \frac{1}{2}\|(f_+ - \Delta(0)) - (f_- - \Delta(0))\| = \frac{1}{2}\|2h\| = \|h\|. \end{aligned}$$

Taking the supremum over h leads to the required inequality $\operatorname{gwce}(\Delta) \ge \mathrm{lb}$. ∎

The second lemma expresses the square of the lower bound as the minimal value of a semidefinite program. In passing, the square of the global worst-case error of a linear recovery map is also related to the minimal value of a semidefinite program.

Lemma 12. One has

$$\mathrm{lb}^2 = \min_{c,d\geq 0} \ c\varepsilon^2 + d\eta^2 \qquad \text{s.to} \quad cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda \succeq \mathrm{Id}. \tag{38}$$

Moreover, if a recovery map $\Delta : \mathbb{R}^m \to H$ is linear, one also has

$$\mathrm{gwce}(\Delta)^2 \leq \min_{c,d\geq 0} \ c\varepsilon^2 + d\eta^2 \qquad \text{s.to} \quad \left[\begin{array}{c|c} cP_{\mathcal{V}^\perp} & 0 \\ \hline 0 & d\,\mathrm{Id}_{\mathbb{R}^m} \end{array}\right] \succeq \left[\begin{array}{c} \mathrm{Id} - \Lambda^*\Delta^* \\ \hline \Delta^* \end{array}\right] \Big[\mathrm{Id} - \Delta\Lambda \mid \Delta\Big]. \tag{39}$$

Proof. The first semidefinite characterization is based on the version of the S-procedure stated in Theorem 1. Precisely, we write the square of the lower bound as

$$\begin{aligned}
\mathrm{lb}^2 = & \sup_h \|h\|^2 \quad \text{s.to } \|P_{\mathcal{V}^\perp} h\|^2 \leq \varepsilon^2 \text{ and } \|\Lambda h\|^2 \leq \eta^2 \\
= & \inf_\gamma \gamma \qquad \text{s.to } \|h\|^2 \leq \gamma \text{ whenever } \|P_{\mathcal{V}^\perp} h\|^2 \leq \varepsilon^2 \text{ and } \|\Lambda h\|^2 \leq \eta^2 \\
= & \inf_\gamma \gamma \qquad \text{s.to } \exists\, c,d \geq 0 : \ \|h\|^2 - \gamma \leq c(\|P_{\mathcal{V}^\perp} h\|^2 - \varepsilon^2) + d(\|\Lambda h\|^2 - \eta^2) \text{ for all } h \in H \\
= & \inf_{\substack{\gamma \\ c,d\geq 0}} \gamma \qquad \text{s.to } c\langle P_{\mathcal{V}^\perp} h, h\rangle + d\langle \Lambda^*\Lambda h, h\rangle - \langle h, h\rangle + \gamma - c\varepsilon^2 - d\eta^2 \geq 0 \text{ for all } h \in H.
\end{aligned}$$

The validity of Theorem 1 in ensured by the facts that $\|P_{\mathcal{V}^\perp}\widetilde{h}\|^2 - \varepsilon^2 < 0$ and $\|\Lambda\widetilde{h}\|^2 - \eta^2 < 0$ for $\widetilde{h} = 0$ and that $P_{\mathcal{V}^\perp} + \Lambda^*\Lambda \succ 0$. Note that the resulting constraint decouples as $\langle cP_{\mathcal{V}^\perp} h + d\Lambda^*\Lambda h - h, h\rangle \geq 0$ for all $h \in H$, i.e., $cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda - \mathrm{Id} \succeq 0$, and $\gamma - c\varepsilon^2 - d\eta^2 \geq 0$. Taking the minimal value of γ under the latter constraint, namely $c\varepsilon^2 + d\eta^2$, leads to the expression of lb^2 given in (38).

As for (39), we start by remarking that the linearity of the recovery map Δ allows us to write

$$\begin{aligned}
\mathrm{gwce}(\Delta)^2 &= \sup_{f,e} \|f - \Delta\Lambda f - \Delta e\|^2 \quad \text{s.to } \|P_{\mathcal{V}^\perp} f\|^2 \leq \varepsilon^2 \text{ and } \|e\|^2 \leq \eta^2 \\
&= \inf_\gamma \gamma \quad \text{s.to } \|f - \Delta\Lambda f - \Delta e\|^2 \leq \gamma \text{ whenever } \|P_{\mathcal{V}^\perp} f\|^2 \leq \varepsilon^2 \text{ and } \|e\|^2 \leq \eta^2.
\end{aligned}$$

The latter constraint can be expressed in terms of the combined variable $v = (f, -e) \in H \times \mathbb{R}^m$ as

$$\left\| \Big[\mathrm{Id} - \Delta\Lambda \mid \Delta\Big] v \right\|^2 \leq \gamma \text{ whenever } \left\| \Big[P_{\mathcal{V}^\perp} \mid 0\Big] v \right\|^2 \leq \varepsilon^2 \text{ and } \left\| \Big[0 \mid \mathrm{Id}_{\mathbb{R}^m}\Big] v \right\|^2 \leq \eta^2. \tag{40}$$

Although the proviso of Theorem 1 is not fulfiled here, the constraint (40) is still a consequence of (but is not equivalent to) the existence of $c, d \geq 0$ such that

$$\left\| \Big[\mathrm{Id} - \Delta\Lambda \mid \Delta\Big] v \right\|^2 - \gamma \leq c\Big(\left\| \Big[P_{\mathcal{V}^\perp} \mid 0\Big] v \right\|^2 - \varepsilon^2\Big) + d\Big(\left\| \Big[0 \mid \mathrm{Id}_{\mathbb{R}^m}\Big] v \right\|^2 - \eta^2\Big) \quad \text{for all } v \in H \times \mathbb{R}^m.$$

The latter can also be written as the existence of $c, d \geq 0$ such that, for all $v \in H \times \mathbb{R}^m$,

$$\left\langle \left(c \left[P_{\mathcal{V}^\perp} \mid 0 \right]^* \left[P_{\mathcal{V}^\perp} \mid 0 \right] + d \left[0 \mid \mathrm{Id}_{\mathbb{R}^m} \right]^* \left[0 \mid \mathrm{Id}_{\mathbb{R}^m} \right] - \left[\mathrm{Id} - \Delta\Lambda \mid \Delta \right]^* \left[\mathrm{Id} - \Delta\Lambda \mid \Delta \right] \right) v, v \right\rangle + \gamma - c\varepsilon^2 - d\eta^2 \geq 0.$$

Therefore, we obtain the inequality (instead of the equality)

$$\mathrm{gwce}(\Delta)^2 \leq \inf_{\substack{\gamma \\ c,d \geq 0}} \gamma \quad \text{s.to} \quad c \left[\begin{array}{c|c} P_{\mathcal{V}^\perp} & 0 \\ \hline 0 & 0 \end{array}\right] + d \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & \mathrm{Id}_{\mathbb{R}^m} \end{array}\right] - \left[\begin{array}{c} \mathrm{Id} - \Lambda^*\Delta^* \\ \hline \Delta^* \end{array}\right] \left[\mathrm{Id} - \Delta\Lambda \mid \Delta \right] \succeq 0$$
$$\text{and} \quad \gamma - c\varepsilon^2 - d\eta^2 \geq 0.$$

The variable γ can be eliminated from this optimization program by assigning it the value $c\varepsilon^2 + d\eta^2$, thus arriving at the semidefinite program announced in (39). □

The third and final lemma relates the constraints of (38) and (39): while the constraint of (39) with any regularization map Δ_τ implies the constraint of (38), see the appendix, we need the partial converse that the constraint of (38) implies the constraint of (39) for a specific regularization map Δ_τ.

Lemma 13. If $cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda \succeq \mathrm{Id}$, then setting $\tau = d/(c+d)$ yields

$$\left[\begin{array}{c|c} cP_{\mathcal{V}^\perp} & 0 \\ \hline 0 & d\,\mathrm{Id}_{\mathbb{R}^m} \end{array}\right] \succeq \left[\begin{array}{c} \mathrm{Id} - \Lambda^*\Delta_\tau^* \\ \hline \Delta_\tau^* \end{array}\right] \left[\mathrm{Id} - \Delta_\tau\Lambda \mid \Delta_\tau \right].$$

Proof. We recall from Proposition 2 adapted to the current situation that, for any $\tau \in (0,1)$,

$$\Delta_\tau = \big((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda\big)^{-1}(\tau\Lambda^*), \quad \text{hence} \quad \mathrm{Id} - \Delta_\tau\Lambda = \big((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda\big)^{-1}((1-\tau)P_{\mathcal{V}^\perp}).$$

We now notice that the hypothesis $cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda \succeq \mathrm{Id}$ is equivalent to $\lambda_{\min}(cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda) \geq 1$. With our particular choice of τ, this reads $\lambda_{\min}((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda) \geq 1/(c+d)$. It follows that

$$\lambda_{\max}\big(((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda)^{-1}\big) = \frac{1}{\lambda_{\min}((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda)} \leq c + d.$$

The inverse appearing above can be written as

$$\big((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda\big)^{-1} \left[\sqrt{1-\tau}P_{\mathcal{V}^\perp} \mid \sqrt{\tau}\Lambda^* \right] \left[\begin{array}{c} \sqrt{1-\tau}P_{\mathcal{V}^\perp} \\ \hline \sqrt{\tau}\Lambda \end{array}\right] \big((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda\big)^{-1}$$

and since AB and BA always have the same nonzero eigenvalues, we derive that

$$\lambda_{\max}\left(\left[\begin{array}{c} \sqrt{1-\tau}P_{\mathcal{V}^\perp} \\ \hline \sqrt{\tau}\Lambda \end{array}\right] \big((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda\big)^{-2} \left[\sqrt{1-\tau}P_{\mathcal{V}^\perp} \mid \sqrt{\tau}\Lambda^* \right] \right) \leq c + d.$$

Writing the latter as

$$\begin{bmatrix} \sqrt{1-\tau} P_{\mathcal{V}^\perp} \\ \hline \sqrt{\tau}\Lambda \end{bmatrix} \big((1-\tau)P_{\mathcal{V}^\perp} + \tau \Lambda^* \Lambda\big)^{-2} \Big[\sqrt{1-\tau} P_{\mathcal{V}^\perp} \mid \sqrt{\tau}\Lambda^*\Big] \preceq (c+d)\mathrm{Id}$$

and multiplying on both sides by $\begin{bmatrix} \sqrt{1-\tau} P_{\mathcal{V}^\perp} & | & 0 \\ \hline 0 & | & \sqrt{\tau}\,\mathrm{Id}_{\mathbb{R}^m} \end{bmatrix}$ yields

$$\begin{bmatrix} (1-\tau) P_{\mathcal{V}^\perp} \\ \hline \tau\Lambda \end{bmatrix} \big((1-\tau)P_{\mathcal{V}^\perp} + \tau \Lambda^* \Lambda\big)^{-2} \Big[(1-\tau) P_{\mathcal{V}^\perp} \mid \tau\Lambda^*\Big] \preceq (c+d) \begin{bmatrix} (1-\tau) P_{\mathcal{V}^\perp} & | & 0 \\ \hline 0 & | & \tau\,\mathrm{Id}_{\mathbb{R}^m} \end{bmatrix}.$$

Taking the expressions of Δ_τ and $\mathrm{Id} - \Delta_\tau \Lambda$ into account, we conclude that

$$\begin{bmatrix} \mathrm{Id} - \Lambda^* \Delta_\tau^* \\ \hline \Delta_\tau^* \end{bmatrix} \Big[\mathrm{Id} - \Delta_\tau \Lambda \mid \Delta_\tau\Big] \preceq \begin{bmatrix} c P_{\mathcal{V}^\perp} & | & 0 \\ \hline 0 & | & d\,\mathrm{Id}_{\mathbb{R}^m} \end{bmatrix},$$

as announced. □

With the above three lemmas at hand, the main result of this subsection follows easily.

Proof of Theorem 10. Since Lemma 11 guarantees that $\inf\{\mathrm{gwce}(\Delta), \Delta : \mathbb{R}^m \to H\} \geq \mathrm{lb}$, we only need to show that $\mathrm{gwce}(\Delta_{\tau_\flat}) \leq \mathrm{lb}$. By the first part of Lemma 12, we have $\mathrm{lb}^2 = c_\flat \varepsilon^2 + d_\flat \eta^2$ with $c_\flat$ and $d_\flat$ satisfying $c_\flat P_{\mathcal{V}^\perp} + d_\flat \Lambda\Lambda^* \succeq \mathrm{Id}$. By Lemma 13, the latter implies that

$$\begin{bmatrix} c_\flat P_{\mathcal{V}^\perp} & | & 0 \\ \hline 0 & | & d_\flat \mathrm{Id}_{\mathbb{R}^m} \end{bmatrix} \succeq \begin{bmatrix} \mathrm{Id} - \Lambda^* \Delta_{\tau_\flat}^* \\ \hline \Delta_{\tau_\flat}^* \end{bmatrix} \Big[\mathrm{Id} - \Delta_{\tau_\flat} \Lambda \mid \Delta_{\tau_\flat}\Big].$$

By the second part of Lemma 12, it follows that $\mathrm{gwce}(\Delta_{\tau_\flat})^2 \leq c_\flat \varepsilon^2 + d_\flat \eta^2 = \mathrm{lb}^2$, which is the required inequality. □

Remark. When $\mathcal{V} = \{0\}$, so that $P_{\mathcal{V}^\perp} = \mathrm{Id}$, we obtain $c_\flat = 1$ and $d_\flat = 0$, resulting in a minimal global worst-case error equal to ε and achieved for the regularization map $\Delta_0 = 0$. This result can be seen directly from $\mathrm{gwce}(\Delta) \geq \sup\{\|h\| : \|h\| \leq \varepsilon, \|\Lambda h\| \leq \eta\} = \varepsilon$ for any $\Delta : \mathbb{R}^m \to H$, while $\mathrm{gwce}(\Delta_0) = \sup\{\|f\| : \|f\| \leq \varepsilon\} = \varepsilon$.

4.2 Orthonormal observations

In this subsection, we demonstrate that the use of orthonormal observations guarantees, rather unexpectedly, that regularization provides optimal recovery maps even without a careful parameter selection. The main result reads as follows.

Theorem 14. Given the approximability set $\mathcal{K} = \{f \in H : \operatorname{dist}(f, \mathcal{V}) \leq \varepsilon\}$ and the uncertainty set $\mathcal{E} = \{e \in \mathbb{R}^m : \|e\| \leq \eta\}$, if $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$, then all the regularization maps Δ_τ are optimal recovery maps, i.e., for all $\tau \in [0,1]$,

$$\operatorname{gwce}(\Delta_\tau) = \inf_{\Delta : \mathbb{R}^m \to H} \operatorname{gwce}(\Delta). \tag{41}$$

The proof strategy consists in establishing that the constraints in (38) and in (39) with $\Delta = \Delta_\tau$ are in fact equivalent for any $\tau \in [0,1]$. This yields the inequality $\operatorname{gwce}(\Delta_\tau) \leq \mathrm{lb}$, which proves the required result, given that lb was introduced as a lower bound on $\operatorname{gwce}(\Delta)$ for every Δ. While the constraint in (39) implies the constraint in (38) for any observation map Λ (see the appendix), the reverse implication relies on the fact that $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$, e.g. via the identity $\Delta_\tau = (1-\tau)\Delta_0 + \tau\Delta_1$ derived in Proposition 3. The following realization is also a crucial point of our argument.

Lemma 15. Assume that $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$. For $c, d \geq 0$, let h be an eigenvector of $cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda$ associated with an eigenvalue λ. For any $\tau \in [0,1]$, one has

- if $\lambda \neq c + d$, then
$$(\mathrm{Id} - \Lambda^*\Delta_\tau^*)h = \frac{c}{\lambda} P_{\mathcal{V}^\perp} h \qquad \text{and} \qquad \Lambda^*\Delta_\tau^* h = \frac{d}{\lambda}\Lambda^*\Lambda h;$$
- if $\lambda = c + d$, then
$$(\mathrm{Id} - \Lambda^*\Delta_\tau^*)h = (1-\tau)h \qquad \text{and} \qquad \Lambda^*\Delta_\tau^* h = \tau h.$$

Proof. Multiplying the eigenequation defining h on the left by $\Lambda^*\Delta_\tau^*$, we obtain

$$c\Lambda^*\Delta_\tau^* P_{\mathcal{V}^\perp} h + d\Lambda^*\Delta_\tau^*\Lambda^*\Lambda h = \lambda\Lambda^*\Delta_\tau^* h. \tag{42}$$

According to (19), we have $\Delta_0^* P_{\mathcal{V}^\perp} = 0$, $\Delta_1^* P_{\mathcal{V}^\perp} = \Delta_1^* - \Delta_0^*$, $\Delta_1^*\Lambda^*\Lambda = \Lambda$, and $\Delta_0^*\Lambda^*\Lambda = \Delta_0^* - \Delta_1^* + \Lambda$. Thus, the relation (42) specified to $\tau = 0$ and to $\tau = 1$ yields

$$d\Lambda^*\Delta_0^* h - d\Lambda^*\Delta_1^* h + d\Lambda^*\Lambda h = \lambda\Lambda^*\Delta_0^* h, \tag{43}$$

$$c\Lambda^*\Delta_1^* h - c\Lambda^*\Delta_0^* h + d\Lambda^*\Lambda h = \lambda\Lambda^*\Delta_1^* h. \tag{44}$$

Subtracting (44) from (43) yields $(c+d)(\Lambda^*\Delta_0^* h - \Lambda^*\Delta_1^* h) = \lambda(\Lambda^*\Delta_0^* h - \Lambda^*\Delta_1^* h)$. Therefore, we derive that $\Lambda^*\Delta_0^* h = \Lambda^*\Delta_1^* h$ provided $\lambda \neq c + d$. In this case, the equations (43)-(44) reduce to $\Lambda^*\Delta_0^* h = \Lambda^*\Delta_1^* h = (d/\lambda)\Lambda^*\Lambda h$. In view of $\Delta_\tau = (1-\tau)\Delta_0 + \tau\Delta_1$, we arrive at $\Lambda^*\Delta_\tau^* h = (d/\lambda)\Lambda^*\Lambda h$ for any $\tau \in [0,1]$. The relation $(\mathrm{Id} - \Lambda^*\Delta_\tau^*)h = (c/\lambda)P_{\mathcal{V}^\perp} h$ follows from the eigenequation rewritten as $(c/\lambda)P_{\mathcal{V}^\perp} h + (d/\lambda)\Lambda^*\Lambda h = h$.

It remains to deal with the case $\lambda = c + d$. Notice that this case is not vacuous, as it is equivalent to $h \in \mathcal{V}^\perp \cap \operatorname{ran}(\Lambda^*\Lambda)$, which is nontrivial by a dimension argument involving assumption (5). To see this equivalence, notice that $h \in \mathcal{V}^\perp \cap \operatorname{ran}(\Lambda^*\Lambda)$ clearly implies $cP_{\mathcal{V}^\perp} h + d\Lambda^*\Lambda h = (c+d)h$, while

the latter eigenequation forces $c\|P_{\mathcal{V}^\perp}h\|^2 + d\|\Lambda^*\Lambda h\|^2 = (c+d)\|h\|^2$, hence $\|P_{\mathcal{V}^\perp}h\|^2 = \|h\|^2$ and $\|\Lambda^*\Lambda h\|^2 = \|h\|^2$, i.e., $h \in \mathcal{V}^\perp$ and $h \in \operatorname{ran}(\Lambda^*\Lambda)$. We now consider such an eigenvector h associated with the eigenvalue $c+d$: in view of $h \in \mathcal{V}^\perp \cap \operatorname{ran}(\Lambda^*\Lambda)$, we remark that $\Delta_0^* h = \Delta_0^* P_{\mathcal{V}^\perp} h = 0$ and that $\Delta_1^* h = \Delta_1^* \Lambda^* \Lambda h = \Lambda h$. We deduce that $\Lambda^* \Delta_\tau^* h = (1-\tau)\Lambda^*\Delta_0^* h + \tau \Lambda^* \Delta_1^* h = \tau \Lambda^* \Lambda h = \tau h$ and in turn that $(\mathrm{Id} - \Lambda^*\Delta_\tau^*)h = (1-\tau)h$. □

We are now ready to establish the main result of this subsection.

Proof of Theorem 14. Let $\tau \in [0,1]$ be fixed throughout. As announced earlier, our objective is to establish that, thanks to $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$, the condition $cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda \succeq \mathrm{Id}$ implies the condition

$$\left[\begin{array}{c|c} cP_{\mathcal{V}^\perp} & 0 \\ \hline 0 & d\,\mathrm{Id}_{\mathbb{R}^m} \end{array}\right] \succeq \left[\begin{array}{c} \mathrm{Id} - \Lambda^*\Delta_\tau^* \\ \hline \Delta_\tau^* \end{array}\right] \left[\mathrm{Id} - \Delta_\tau\Lambda \mid \Delta_\tau\right],$$

or equivalently the condition

$$\left[\begin{array}{c|c} cP_{\mathcal{V}^\perp} & 0 \\ \hline 0 & d\Lambda^*\Lambda \end{array}\right] \succeq \left[\begin{array}{c} \mathrm{Id} - \Lambda^*\Delta_\tau^* \\ \hline \Lambda^*\Delta_\tau^* \end{array}\right] \left[\mathrm{Id} - \Delta_\tau\Lambda \mid \Delta_\tau\Lambda\right].$$

The equivalence of these conditions is seen as follows: the former implies the latter by multiplying on the left by $\left[\begin{array}{c|c} \mathrm{Id} & 0 \\ \hline 0 & \Lambda^* \end{array}\right]$ and on the right by $\left[\begin{array}{c|c} \mathrm{Id} & 0 \\ \hline 0 & \Lambda \end{array}\right]$, while the latter implies the former under the assumption $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$ by multiplying on the left by $\left[\begin{array}{c|c} \mathrm{Id} & 0 \\ \hline 0 & \Lambda \end{array}\right]$ and on the right by $\left[\begin{array}{c|c} \mathrm{Id} & 0 \\ \hline 0 & \Lambda^* \end{array}\right]$. As a matter of fact, according to a classical result about Schur complements, see e.g. [Boyd and Vandenberghe, 2004, Section A.5.5], the latter is further equivalent to

$$\left[\begin{array}{c|c|c} \mathrm{Id} & \mathrm{Id} - \Delta_\tau\Lambda & \Delta_\tau\Lambda \\ \hline \mathrm{Id} - \Lambda^*\Delta_\tau^* & cP_{\mathcal{V}^\perp} & 0 \\ \hline \Lambda^*\Delta_\tau^* & 0 & d\Lambda^*\Lambda \end{array}\right] \succeq 0.$$

Thus, considering $f, g, h \in H$, our objective is to prove the nonnegativity of the inner product

$$\begin{aligned} \mathrm{ip} &:= \left\langle \left[\begin{array}{c|c|c} \mathrm{Id} & \mathrm{Id} - \Delta_\tau\Lambda & \Delta_\tau\Lambda \\ \hline \mathrm{Id} - \Lambda^*\Delta_\tau^* & cP_{\mathcal{V}^\perp} & 0 \\ \hline \Lambda^*\Delta_\tau^* & 0 & d\Lambda^*\Lambda \end{array}\right] \left[\begin{array}{c} f \\ \hline g \\ \hline h \end{array}\right], \left[\begin{array}{c} f \\ \hline g \\ \hline h \end{array}\right] \right\rangle \\ &= \langle f, f\rangle + c\langle P_{\mathcal{V}^\perp} g, g\rangle + d\langle \Lambda^*\Lambda h, h\rangle + 2\langle (\mathrm{Id} - \Lambda^*\Delta_\tau^*)f, g\rangle + 2\langle \Lambda^*\Delta_\tau^* f, h\rangle. \end{aligned}$$

Let us decompose f, g, and h as $f = f' + f''$, $g = g' + g''$, and $h = h' + h''$, where f', g', and h' belong to the space H' spanned by eigenvectors of $cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda$ corresponding to eigenvalues $\lambda \neq c+d$ and where f'', g'', and h'' belong to the eigenspace H'' of $cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda$ corresponding to the eigenvalue

$\lambda = c + d$, i.e., $H'' = \mathcal{V}^\perp \cap \mathrm{ran}(\Lambda^*\Lambda)$. We take notice of the fact that the spaces H' and H'' are orthogonal. With this decomposition, the above inner product becomes

$$\mathrm{ip} = \mathrm{ip}' + \mathrm{ip}'' + \mathrm{ip}''',$$

where we have set

$$\begin{aligned}
\mathrm{ip}' &= \langle f', f'\rangle + c\langle P_{\mathcal{V}^\perp} g', g'\rangle + d\langle \Lambda^*\Lambda h,' h'\rangle + 2\langle (\mathrm{Id} - \Lambda^*\Delta_\tau^*) f', g'\rangle + 2\langle \Lambda^*\Delta_\tau^* f', h'\rangle,\\
\mathrm{ip}'' &= \langle f'', f''\rangle + c\langle P_{\mathcal{V}^\perp} g'', g''\rangle + d\langle \Lambda^*\Lambda h'', h''\rangle + 2\langle (\mathrm{Id} - \Lambda^*\Delta_\tau^*) f'', g''\rangle + 2\langle \Lambda^*\Delta_\tau^* f'', h''\rangle,\\
\mathrm{ip}''' &= 2\langle f', f''\rangle + 2\langle P_{\mathcal{V}^\perp} g', g''\rangle + 2\langle \Lambda^*\Lambda h', h''\rangle + 2\langle (\mathrm{Id} - \Lambda^*\Delta_\tau^*) f', g''\rangle + 2\langle \Lambda^*\Delta_\tau^* f', h''\rangle\\
&\qquad + 2\langle (\mathrm{Id} - \Lambda^*\Delta_\tau^*) f'', g'\rangle + 2\langle \Lambda^*\Delta_\tau^* f'', h'\rangle.
\end{aligned}$$

We first remark that the terms in ip''' are all zero: first, it is clear that $\langle f', f''\rangle = 0$; then, one has $\langle P_{\mathcal{V}^\perp} g', g''\rangle = \langle g', P_{\mathcal{V}^\perp} g''\rangle = \langle g', g''\rangle = 0$ and $\langle \Lambda^*\Lambda h', h''\rangle = 0$ is obtained similarly; next, Lemma 15 ensures that $\langle (\mathrm{Id} - \Lambda^*\Delta_\tau^*) f'', g'\rangle = (1-\tau)\langle f'', g'\rangle = 0$ and $\langle \Lambda^*\Delta_\tau^* f'', h'\rangle = 0$ is obtained similarly; last, writing $f' = \sum_i f_i$ where the $f_i \in H'$ are orthogonal eigenvectors of $cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda$ corresponding to eigenvalues $\lambda_i < c + d$, we derive from Lemma 15 that

$$\langle (\mathrm{Id} - \Lambda^*\Delta_\tau^*) f', g''\rangle = \sum_i \frac{c}{\lambda_i} \langle P_{\mathcal{V}^\perp} f_i, g''\rangle = \sum_i \frac{c}{\lambda_i} \langle f_i, P_{\mathcal{V}^\perp} g''\rangle = \sum_i \frac{c}{\lambda_i} \langle f_i, g''\rangle = 0,$$

and $\langle \Lambda^*\Delta_\tau^* f', h''\rangle = 0$ is obtained similarly. As a result, we have $\mathrm{ip}''' = 0$.

We now turn to the quantity ip'. Exploiting Lemma 15 again, we write

$$\begin{aligned}
\mathrm{ip}' &= \langle f', f'\rangle + c\langle P_{\mathcal{V}^\perp} g', g'\rangle + d\langle \Lambda^*\Lambda h,' h'\rangle + 2\left\langle \sum_i \frac{c}{\lambda_i} P_{\mathcal{V}^\perp} f_i, g'\right\rangle + 2\left\langle \sum_i \frac{d}{\lambda_i} \Lambda^*\Lambda f_i, h'\right\rangle\\
&= \langle f', f'\rangle + c\left(\langle P_{\mathcal{V}^\perp} g', P_{\mathcal{V}^\perp} g'\rangle + 2\left\langle \sum_i \frac{1}{\lambda_i} P_{\mathcal{V}^\perp} f_i, P_{\mathcal{V}^\perp} g'\right\rangle\right)\\
&\qquad + d\left(\langle \Lambda^*\Lambda h', \Lambda^*\Lambda h'\rangle + 2\left\langle \sum_i \frac{1}{\lambda_i} \Lambda^*\Lambda f_i, \Lambda^*\Lambda h'\right\rangle\right)\\
&= \langle f', f'\rangle + c\left(\left\| P_{\mathcal{V}^\perp} g' + \sum_i \frac{1}{\lambda_i} P_{\mathcal{V}^\perp} f_i\right\|^2 - \left\| \sum_i \frac{1}{\lambda_i} P_{\mathcal{V}^\perp} f_i\right\|^2\right)\\
&\qquad + d\left(\left\| \Lambda^*\Lambda h' + \sum_i \frac{1}{\lambda_i} \Lambda^*\Lambda f_i\right\|^2 - \left\| \sum_i \frac{1}{\lambda_i} \Lambda^*\Lambda f_i\right\|^2\right).
\end{aligned}$$

At this point, we can bound ip' from below as

$$\begin{aligned}\mathrm{ip}' &\geq \langle f', f' \rangle - \left(c \left\| P_{\mathcal{V}^\perp} \Big(\sum_i \frac{1}{\lambda_i} f_i \Big) \right\|^2 + d \left\| \Lambda^* \Lambda \Big(\sum_i \frac{1}{\lambda_i} f_i \Big) \right\|^2 \right) \\ &= \langle f', f' \rangle - \left\langle \Big(c P_{\mathcal{V}^\perp} + d \Lambda^* \Lambda \Big) \Big(\sum_i \frac{1}{\lambda_i} f_i \Big), \Big(\sum_i \frac{1}{\lambda_i} f_i \Big) \right\rangle \\ &= \sum_i \|f_i\|^2 - \left\langle \sum_i f_i, \sum_i \frac{1}{\lambda_i} f_i \right\rangle = \sum_i \|f_i\|^2 \Big(1 - \frac{1}{\lambda_i} \Big).\end{aligned}$$

This shows that $\mathrm{ip}' \geq 0$ since the condition $c P_{\mathcal{V}^\perp} + d \Lambda^* \Lambda \succeq \mathrm{Id}$ ensures that $\lambda_i \geq 1$ for every i.

Finally, Lemma 15 also helps us to bound the quantity ip'' from below according to

$$\begin{aligned}\mathrm{ip}'' &= \|f''\|^2 + c\|g''\|^2 + d\|h''\|^2 + 2(1-\tau)\langle f'', g'' \rangle + 2\tau \langle f'', h'' \rangle \\ &= (1-\tau)\big(\|f''\|^2 + 2\langle f'', g'' \rangle \big) + \tau \big(\|f''\|^2 + 2\langle f'', h'' \rangle \big) + c\|g''\|^2 + d\|h''\|^2 \\ &\geq -(1-\tau)\|g''\|^2 - \tau \|h''\|^2 + c\|g''\|^2 + d\|h''\|^2.\end{aligned}$$

This allows us to obtain $\mathrm{ip}'' \geq 0$ since the condition $c P_{\mathcal{V}^\perp} + d \Lambda^* \Lambda \succeq \mathrm{Id}$ ensures that $c \geq 1$ and $d \geq 1$. Altogether, we have shown that $\mathrm{ip} = \mathrm{ip}' + \mathrm{ip}'' + \mathrm{ip}''' \geq 0$, which concludes the proof. □

Remark. The value of the minimal global worst-case error can, in general, be computed by solving the semidefinite program (38) characterizing the lower bound lb. In the case where $\Lambda \Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$, it can also be computed without resorting to semidefinite programming. Precisely, if $\tau_\sharp$ denotes the (unique) τ between $1/2$ and $\varepsilon/(\varepsilon + \eta)$ such that

$$\lambda_{\min}((1-\tau) P_{\mathcal{V}^\perp} + \tau \Lambda^* \Lambda) = \frac{(1-\tau)^2 \varepsilon^2 - \tau^2 \eta^2}{(1-\tau)\varepsilon^2 - \tau \eta^2} \tag{45}$$

and if $\lambda_\sharp$ denotes $\lambda_{\min}((1-\tau_\sharp) P_{\mathcal{V}^\perp} + \tau_\sharp \Lambda^* \Lambda)$, then we claim that, for any $\tau \in [0,1]$,

$$\mathrm{gwce}(\Delta_\tau)^2 = \frac{1-\tau_\sharp}{\lambda_\sharp} \varepsilon^2 + \frac{\tau_\sharp}{\lambda_\sharp} \eta^2.$$

Indeed, since we now know that the global worst-case error $\mathrm{gwce}(\Delta_\tau)$ equals its lower bound lb independently of $\tau \in [0,1]$ and since $c_\sharp := (1-\tau_\sharp)/\lambda_\sharp$ and $d_\sharp := \tau_\sharp / \lambda_\sharp$ are feasible for the semidefinite program (38) characterizing lb, we obtain

$$\mathrm{gwce}(\Delta_\tau)^2 \leq \frac{1-\tau_\sharp}{\lambda_\sharp} \varepsilon^2 + \frac{\tau_\sharp}{\lambda_\sharp} \eta^2. \tag{46}$$

Moreover, going back to the proof of Theorem 8, we recognize that the choice of $\tau_\sharp$ here corresponds to the instance $y = 0$ there. This instance comes with $f_\sharp$ being equal to zero and with $h_\sharp$ being equal to a properly normalized eigenvector of $(1-\tau_\sharp) P_{\mathcal{V}^\perp} + \tau_\sharp \Lambda^* \Lambda$ corresponding to the eigenvalue $\lambda_\sharp$. The identities (32) now read $\|P_{\mathcal{V}^\perp} h_\sharp\|^2 = \varepsilon^2$ and $\|\Lambda^* \Lambda h_\sharp\|^2 = \eta^2$, i.e., $\|\Lambda h_\sharp\|^2 = \eta^2$. Setting $f = h_\sharp$

and $e = -\Lambda h_\sharp$, which satisfy $\|P_{\mathcal{V}^\perp} f\| = \varepsilon$ and $\|e\| = \eta$, the very definition of the global worst-case error yields

$$\begin{aligned} \mathrm{gwce}(\Delta_\tau)^2 &\geq \|f - \Delta_\tau(\Lambda f + e)\|^2 = \|h_\sharp\|^2 \\ &= \frac{1}{\lambda_\sharp}\big\langle \big((1-\tau_\sharp) P_{\mathcal{V}^\perp} + \tau_\sharp \Lambda^* \Lambda\big) h_\sharp, h_\sharp \big\rangle \\ &= \frac{1-\tau_\sharp}{\lambda_\sharp} \|P_{\mathcal{V}^\perp} h_\sharp\|^2 + \frac{\tau_\sharp}{\lambda_\sharp} \|\Lambda^* \Lambda h_\sharp\|^2 \\ &= \frac{1-\tau_\sharp}{\lambda_\sharp} \varepsilon^2 + \frac{\tau_\sharp}{\lambda_\sharp} \eta^2. \end{aligned} \tag{47}$$

Together, the inequalities (46) and (47) justify our claim about the value of the global worst-case error. In passing, it is worth noticing that the above argument reveals that $f = h_\sharp$ and $e = -\Lambda h_\sharp$ are extremal in the defining expression for the global worst-case error of the regularization map Δ_τ independently of the parameter $\tau \in [0,1]$.

References

A. Beck and Y. C. Eldar. Regularization in regression with bounded noise: a Chebyshev center approach. *SIAM Journal on Matrix Analysis and Applications*, 29(2):606–625, 2007.

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.

S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Z. Chen and S. Haykin. On different facets of regularization theory. *Neural Computation*, 14(12): 2791–2846, 2002.

A. Cohen, W. Dahmen, O. Mula, and J. Nichols. Nonlinear reduced models for state and parameter estimation. *arXiv preprint arXiv:2009.02687*, 2020.

R. DeVore, G. Petrova, and P. Wojtaszczyk. Data assimilation and sampling in Banach spaces. *Calcolo*, 54(3):963–1007, 2017.

S. Diamond and S. Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

M. Ettehad 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. *Mathematical Pictures at a Data Science Exhibition*. Cambridge University Press, To appear.

S. Foucart, C. Liao, S. Shahrampour, and Y. Wang. Learning from non-random data in Hilbert spaces: an optimal recovery perspective. *arXiv preprint arXiv:2006.03706*, 2020.

A. L. Garkavi. On the optimal net and best cross-section of a set in a normed space. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 26(1):87–106, 1962.

M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, second edition, 2009.

Y. Maday, A. T. Patera, J. D. Penn, and M. Yano. A parameterized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics. *International Journal for Numerical Methods in Engineering*, 102(5):933–965, 2015.

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.

L. Plaskota. *Noisy Information and Computational Complexity*. Cambridge University Press, 1996.

I. Pólik and T. Terlaky. A survey of the S-lemma. *SIAM Review*, 49(3):371–418, 2007.

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.

B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2002.

Appendix

This additional section collects justifications for a few facts that were mentioned but not explained in the main text. These facts are: the uniqueness of a Chebyshev center for the model- and data-consistent set (see page 6), the efficient computation of the solution to (7) when $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$ (see page 11), the form of Newton method when solving equation (28) (see page 15), and the reason why the constraint of (39) always implies the constraint of (38) (see pages 21 and 23).

Uniqueness of the Chebyshev center. Let $\widehat{f}_1, \widehat{f}_2$ be two Chebyshev centers, i.e., minimizers of $\max\{\|f-g\| : \|P_{\mathcal{V}^\perp} g\| \le \varepsilon, \|\Lambda g - y\| \le \eta\}$ and let μ be the value of the minimum. Consider $\overline{g} \in H$ such that $\|(\widehat{f}_1 + \widehat{f}_2)/2 - \overline{g}\| = \max\{\|(\widehat{f}_1 + \widehat{f}_2)/2 - g\| : \|P_{\mathcal{V}^\perp} g\| \le \varepsilon, \|\Lambda g - y\| \le \eta\}$. Then

$$\begin{aligned}\mu &\le \|(\widehat{f}_1 + \widehat{f}_2)/2 - \overline{g}\| \le \frac{1}{2}\|\widehat{f}_1 - \overline{g}\| + \frac{1}{2}\|\widehat{f}_2 - \overline{g}\| \\ &\le \frac{1}{2}\max\{\|\widehat{f}_1 - g\| : \|P_{\mathcal{V}^\perp} g\| \le \varepsilon, \|\Lambda g - y\| \le \eta\} + \frac{1}{2}\max\{\|\widehat{f}_2 - g\| : \|P_{\mathcal{V}^\perp} g\| \le \varepsilon, \|\Lambda g - y\| \le \eta\} \\ &= \frac{1}{2}\mu + \frac{1}{2}\mu = \mu.\end{aligned}$$

Thus, equality must hold all the way through. This implies that $\widehat{f}_1 - \overline{g} = \widehat{f}_2 - \overline{g}$, i.e., that $\widehat{f}_1 = \widehat{f}_2$, as expected.

Computation of the regularized solution. Let $(v_1, \dots, v_n)$ be a basis for $\mathcal{V}$ and let $u_1, \dots, u_m$ denote the Riesz representers of the observation functionals $\lambda_1, \dots, \lambda_m$, which form an orthonormal basis for $\mathrm{ran}(\Lambda^*)$ under the assumption that $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$. With $C \in \mathbb{R}^{m \times n}$ representing the cross-gramian with entries $\langle u_i, v_j \rangle = \lambda_i(v_j)$, the solution to the regularization program (7) is given, even when H is infinite dimensional, by

$$f_\tau = \tau \sum_{i=1}^m a_i u_i + \sum_{j=1}^n b_j v_j,$$

where the coefficient vectors $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$ are computed according to

$$b = \left(C^\top C\right)^{-1} C^\top y \qquad \text{and} \qquad a = y - Cb.$$

This is fairly easy to see for $\tau = 0$ and it has been established in [Foucart, Liao, Shahrampour, and Wang, 2020, Theorem 2] for $\tau = 1$, so the general result follows from Proposition 3. Alternatively, it can be obtained by replicating the steps from the proof of the case $\tau = 1$ with minor changes.

Newton method. Equation (28) takes the form $F(\tau) = 0$, where

$$F(\tau) = \lambda_{\min}((1-\tau)R + \tau S) - \frac{(1-\tau)^2\varepsilon^2 - \tau^2\eta^2}{(1-\tau)\varepsilon^2 - \tau\eta^2 + (1-\tau)\tau(1-2\tau)\delta^2}.$$

Newton method produces a sequence $(\tau_k)_{k\geq 0}$ converging to a solution using the recursion

$$\tau_{k+1} = \tau_k - \frac{F(\tau_k)}{F'(\tau_k)}, \qquad k \geq 0. \tag{48}$$

In order to apply this method, we need the ability to compute the derivative of F with respect to τ. Setting $\lambda_{\min} = \lambda_{\min}((1-\tau)R + \tau S)$, this essentially reduces to the computation of $d\lambda_{\min}/d\tau$, which is performed via the argument below. Note that the argument is not rigorous, as we take for granted the differentiability of the eigenvalue $\lambda_{\min}$ and of a normalized eigenvector h associated with it. However, nothing prevents us from applying the scheme (48) using the expression for $d\lambda_{\min}/d\tau$ given in (49) below and agree that a solution has been found if the output τ_K satisfies $F(\tau_K) < \iota$ for some prescribed tolerance $\iota > 0$. Now, the argument starts form the identities

$$((1-\tau)R + \tau S)h = \lambda_{\min} h \qquad \text{and} \qquad \langle h, h\rangle = 1,$$

which we differentiate to obtain

$$(S-R)h + ((1-\tau)R + \tau S)\frac{dh}{d\tau} = \frac{d\lambda_{\min}}{d\tau} h + \lambda_{\min}\frac{dh}{d\tau} \qquad \text{and} \qquad 2\Big\langle h, \frac{dh}{d\tau}\Big\rangle = 0.$$

By taking the inner product with h in the first identity and using the second identity, we derive

$$\langle (S-R)h, h\rangle = \frac{d\lambda_{\min}}{d\tau}, \qquad \text{i.e.,} \qquad \frac{d\lambda_{\min}}{d\tau} = \|Sh\|^2 - \|Rh\|^2.$$

According to Lemma 9, this expression can be transformed, after some work, into

$$\frac{d\lambda_{\min}}{d\tau} = \frac{1-2\tau}{\tau(1-\tau)} \, \frac{\lambda_{\min}(1-\lambda_{\min})}{1-2\lambda_{\min}}. \tag{49}$$

Relation between semidefinite constraints. Suppose that the constraint of (39) holds for a regularization map Δ_τ. In view of the expressions

$$\Delta_\tau = \big((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda\big)^{-1}(\tau\Lambda^*) \quad \text{and} \quad \mathrm{Id} - \Delta_\tau\Lambda = \big((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda\big)^{-1}((1-\tau)P_{\mathcal{V}^\perp}),$$

this constraint also reads

$$\left[\begin{array}{c|c} cP_{\mathcal{V}^\perp} & 0 \\ \hline 0 & d\,\mathrm{Id}_{\mathbb{R}^m} \end{array}\right] \succeq \left[\begin{array}{c} (1-\tau)P_{\mathcal{V}^\perp} \\ \hline \tau\Lambda \end{array}\right] \big((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda\big)^{-2} \Big[(1-\tau)P_{\mathcal{V}^\perp} \mid \tau\Lambda^*\Big].$$

Multiplying on the left by $\Big[P_{\mathcal{V}^\perp} \mid \Lambda^*\Big]$ and on the right by $\left[\begin{array}{c} P_{\mathcal{V}^\perp} \\ \hline \Lambda \end{array}\right]$ yields

$$cP_{\mathcal{V}^\perp} + d\Lambda^*\Lambda \succeq ((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda)\big((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda\big)^{-2}((1-\tau)P_{\mathcal{V}^\perp} + \tau\Lambda^*\Lambda) = \mathrm{Id}.$$

This is the constraint of (38).