Worst-Case-Optimal Estimation of a Quadratic Form by Quadratic Functionals of its Linear Observations

Simon Foucart*— Texas A&M University

Abstract

In this article, it is established that, in order for a quadratic form of an unknown object to be estimated optimally, in a worst-case sense, from the *a priori* knowledge that the object belongs to an hyperellipsoid and the *a posteriori* availability of linear observations, a quadratic functional of these observations can be used. The proof of existence of such an optimal quadratic functional relies on a slightly refined minimax theorem. In addition, when the underlying Hilbert space is finite-dimensional, a computational construction of the optimal quadratic functional is uncovered. The results are then extended to take into account some observation errors modeled deterministically.

Key words and phrases: optimal recovery, minimax theorems, S-lemma, S-procedure.

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

1 Framing of the Problem

The adage that "optimal estimation of linear functionals can be achieved with linear functionals" is a cornerstone of Optimal Recovery [10], and in turn of Information-Based Complexity [15]. The result is credited to Smolyak [13], although it was published later by his doctoral advisor Bakvalov [1]. The purpose of this article is to establish a similar result where "quadratic" replaces "linear". One starts by formalizing precisely what is meant.

The goal here is to estimate a quantity $\Gamma(f)$ depending on an element f from a vector space F—often, but not necessarily, a space of functions. This element f is only partially known through:

- some *a priori* information, expressed as

$$f \in \mathcal{K} \qquad \text{for some model set } \mathcal{K} \subseteq F;$$

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

- some *a posteriori* information resulting from linear observations $y_i = \lambda_i(f)$, $i = 1, \ldots, m$, summarized as
$$y = \Lambda f \qquad \text{for some linear map } \Lambda : F \to \mathbb{R}^m.$$

The estimation of the quantity of interest $\Gamma : F \to G$ is encapsulated by a map $\Delta : \mathbb{R}^m \to G$, whose performance is measured in a worst-case framework

- either via $\mathrm{lwce}_y(\Delta(y))$, when $y \in \mathbb{R}^m$ is considered fixed and
$$\mathrm{lwce}_y(g) := \sup\{\|\Gamma(f) - g\| : f \in \mathcal{K}, \Lambda f = y\};$$

- or, when $y \in \mathbb{R}^m$ is unfixed, via
$$\mathrm{gwce}(\Delta) := \sup\{\|\Gamma(f) - \Delta(\Lambda f)\| : f \in \mathcal{K}\}.$$

The grail of Optimal Recovery is to find estimation maps Δ^{opti} that minimize these notions of worst-case error, either locally, i.e., $\mathrm{lwce}_y(\Delta^{\mathrm{opti}}(y)) \le \mathrm{lwce}_y(g)$ for all $g \in G$, or globally, i.e., $\mathrm{gwce}(\Delta^{\mathrm{opti}}) \le \mathrm{gwce}(\Delta)$ for all $\Delta : \mathbb{R}^m \to G$. The adage from the beginning refers to this global setting. It conveys that, when the model set $\mathcal{K}$ is convex and symmetric about the origin, if $\Gamma = \gamma : F \to \mathbb{R}$ is a linear functional—lowercase Greek letters are used for functionals—then it is possible to choose an optimal estimation map $\delta^{\mathrm{opti}} : \mathbb{R}^m \to \mathbb{R}$ as a linear functional. In fact, the requirement that $\mathcal{K}$ is symmetric about the origin can be dropped if one allows for estimation functionals that are affine instead of linear, as originally shown by Sukharev [14].

Much less is known when the functional γ to be estimated is not linear. Assuming again that $\mathcal{K}$ is a convex set, it was recently established in [4] that, if $\gamma(f)$ is the largest of several linear functionals of f (or even the kth largest), then it is possible to choose a globally optimal estimation functional δ^{opti} as a maximum of linear functionals (or as the maximum of minima of linear functionals). This article focuses on the case where γ is a quadratic functional. This case had been considered in [2], where it was shown in a specific situation that quadratic estimation functionals can provide near-optimality with respect to a different performance measure, more statistical in nature to account for stochastic observation errors. A similar view was taken in [8], where the authors further advocate an "operational" approach: enable more general situations by sacrificing closed-form solutions for computed ones.

This is the approach adopted here, too. For an arbitrary quadratic form $\gamma(f) = \langle Af, f\rangle$ defined by a self-adjoint operator A on the space F—now considered a Hilbert space and thus denoted by H—and for general ellispoidal model set $\mathcal{K}$, it will be shown in Section 3 that it is possible to find a globally optimal recovery functional $\delta^{\mathrm{opti}} : \mathbb{R}^m \to \mathbb{R}$ expressed as $\delta^{\mathrm{opti}}(y) = \langle By, y\rangle + \beta$ for some symmetric matrix $B \in \mathbb{R}^{m \times m}$ and some scalar $\beta \in \mathbb{R}$. Such an existence result (Theorem 2), based on a slight refinement of Sion minimax theorem, will be accompanied in Section 4 by a computational construction of the optimal quadratic estimation functional (Theorem 6), based on Yakubovich S-lemma. The survey [11] discusses this lemma, as well as related results which will be introduced in due time. Namely, in Section 5, Polyak S-procedure will allow for the computational construction accompanying an extension of the main result (Theorem 7) covering the presence of deterministic observation errors. In Section 6, Finsler lemma will also be used to highlight the computation, redundant in retrospect, of a key lower bound (Proposition 8). The content just described shall be preceded by Section 2, which quickly settles the local setting, and succeeded by an appendix, which contains the proof of the necessary adjustment of the minimax theorem.

2 Locally Optimal Estimation Functionals

In the scenario considered here, the quantity of interest γ can be a quadratic functional on a Hilbert space H, instead of merely a quadratic form, i.e.,

$$\gamma(f) = \langle Af, f\rangle + 2\langle a, f\rangle + \alpha, \qquad f \in H, \tag{1}$$

for some self-adjoint operator A on H, some vector $a \in H$, and some scalar $\alpha \in \mathbb{R}$. The *a posteriori* information, considered fixed in this section, is given by $y = \Lambda(f)$ for some unrestricted linear map $\Lambda : H \to \mathbb{R}^m$, while the model set $\mathcal{K}$ describing the *a priori* information is specifically given as

$$\mathcal{K} = \{f \in H : \|Tf\| \leq 1\} \tag{2}$$

for some linear operator $T : H \to H$. The goal is to determine the so-called Chebyshev center of $\gamma(\mathcal{K}_y)$, where the set $\mathcal{K}_y := \{f \in \mathcal{K} : \Lambda(f) = y\}$ collects all the model- and data-consistent elements f. In other words, one aims at determining the minimizer $g^{\rm cheb}$ of the local worst-case error $\mathrm{lwce}_y(g) = \sup\{|\gamma(f) - g| : \|Tf\| \leq 1, \gamma(f) = y\}$. Since the set $\mathcal{K}_y$ is convex, the set $\gamma(\mathcal{K}_y)$ is a convex subset of the real-line, i.e, an interval. It is then clear that the Chebyshev center and Chebyshev radius are obtained as

$$g^{\rm cheb} = \frac{\sigma + \iota}{2}, \qquad \mathrm{lwce}_y(g^{\rm cheb}) = \frac{\sigma - \iota}{2}, \tag{3}$$

where

$$\sigma := \sup_{f \in H} \{\gamma(f) : \|Tf\| \leq 1, \Lambda f = y\}, \tag{4}$$

$$\iota := \inf_{f \in H} \{\gamma(f) : \|Tf\| \leq 1, \Lambda f = y\}. \tag{5}$$

For the problem to make sense, i.e., for $\sigma < +\infty$ and $\iota > -\infty$ independently of γ, it is assumed

$$\ker(\Lambda) \cap \ker(T) = \{0\}, \tag{6}$$

otherwise a nonzero $h \in \ker(\Lambda) \cap \ker(T)$ and a fixed $f_0 \in \mathcal{K}_y$ would generate the unbounded family $f_t = f_0 + th$, $t \in \mathbb{R}$, of elements belonging to $\mathcal{K}_y$. Elementary as the above considerations are, computing the quantities (4) and (5) is not completely straightforward. This is where the S-lemma come into play. Here is its statement for a finite-dimensional Hilbert space H—note that, since an optimization program eventually needs to be solved, assuming finite-dimensionality for the rest of this section is not considered a restriction.

Yakubovich S-lemma: given two quadratic functionals $q_i(h) = \langle R_i h, h\rangle + 2\langle r_i, f\rangle + \rho_i$, $i = 0, 1$, defined on a finite-dimensional Hilbert space H, the equivalence

$$[q_0(h) \le 0 \;\; \textit{whenever } q_1(h) \le 0] \iff [\exists c \ge 0 : q_0 \le c\, q_1]$$

holds provided $q_1(h) < 0$ for some $h \in H$.

In the present scenario, this tool allows one to uncover a computational recipe for the construction of the locally optimal estimation functional. The construction is implemented in the reproducible MATLAB file accompanying this article[1]. There, it is exploited to show that this locally optimal functional is not quadratic.

Theorem 1. For a finite-dimensional Hilbert space H, a linear observation map $\Lambda : H \to \mathbb{R}^m$, and the model set $\mathcal{K}$ of (2) obeying the assumption (6), the locally optimal recovery map for the estimation of the quadratic functional (1) is given by $\delta^{\rm cheb} : y \in \mathbb{R}^m \mapsto (d' + d'')/2 \in \mathbb{R}$, where d' and d'' are, respectively, the optimal values of the following semidefinite programs, in which $P = \mathrm{Id}_H - \Lambda^*(\Lambda\Lambda^*)^{-1}\Lambda$:

$$\underset{\substack{c\ge0\\d}}{\text{minimize}}\; d \quad \text{subject to} \quad \left[\begin{array}{c|c} P(cT^*T - A)P & -P(Af_y + a) \\ \hline -(Af_y+a)^*P & d - c(1 - \|Tf_y\|^2) - \langle Af_y, f_y\rangle - 2\langle a, f_y\rangle - \alpha \end{array}\right] \succeq 0,$$

$$\underset{\substack{c\ge0\\d}}{\text{maximize}}\; d \quad \text{subject to} \quad \left[\begin{array}{c|c} P(cT^*T + A)P & P(Af_y + a) \\ \hline (Af_y+a)^*P & -d - c(1 - \|Tf_y\|^2) + \langle Af_y, f_y\rangle + 2\langle a, f_y\rangle + \alpha \end{array}\right] \succeq 0.$$

Proof. Recall that the locally optimal estimation functional $\delta^{\rm opti}$ associates to each $y \in \mathbb{R}^m$ the Chebyshev center $g^{\rm cheb}$ of $\gamma(\mathcal{K}_y)$ as given in (3). Accordingly, the task at hand consists in showing that σ coincides with d' and ι with d''. Focusing first on $\sigma = \sup\{\gamma(f) : \|Tf\| \le 1, \Lambda(f) = y\}$, note that it can be expressed as

$$\sigma = \inf_d\; d \quad \text{subject to } \langle Af, f\rangle + 2\langle a, f\rangle + \alpha \le d \text{ whenever } \|Tf\|^2 \le 1,\ \Lambda f = y.$$

By writing $f = f_y + h$ for a fixed $f_y \in H$ with $\|Tf_y\|^2 \le 1$ and $\Lambda f_y = y$ and for $h \in \ker(\Lambda)$, one views the above constraint as the nonpositivity of a quadratic functional of h occurring as a consequence of the nonpositivity of another quadratic functional of h. Thus, the S-lemma will apply provided the latter quadratic functional can take a strictly negative value. Here, it is convenient to choose $f_y = \operatorname{argmin}\{\|Tf\|^2 : \Lambda(f) = y\}$, which is characterized by the orthogonality condition

[1] Available on the author's webpage and the repository `https://github.com/foucart/COR`.

$\langle Tf_y, Th\rangle = 0$ for all $h \in \ker(\Lambda)$, together with the equality $\Lambda f_y = y$—as a matter of fact, one has $f_y = T^{-1}T^{-*}\Lambda^*(\Lambda T^{-1}T^{-*}\Lambda^*)^{-1}y$. Indeed, the strict inequality $\|Tf_y\|^2 < 1$ holds: select some $f \neq f_y$ with $\|Tf\|^2 \leq 1$ and $\Lambda f = y$, so that $f - f_y \in \ker(\Lambda)$, hence $T(f - f_y) \neq 0$ by (6), and then the orthogonality of $T(f_y)$ to $T(f - f_y)$ yields $\|Tf_y\|^2 = \|Tf\|^2 - \|T(f - f_y)\|^2 < 1$. Therefore, the S-lemma ensures that the above constraint is equivalent to the existence of $c \geq 0$ such that

$$\langle A(f_y + h), f_y + h\rangle + 2\langle a, f_y + h\rangle + \alpha - d \leq c(\|T(f_y + h)\|^2 - 1) \qquad \text{for all } h \in \ker(\Lambda).$$

Rearranging and also making use of the orthogonality condition $\langle Tf_y, Th\rangle = 0$, this reads

$$\langle (cT^*T - A)h, h\rangle - 2\langle Af_y + a, h\rangle + d - c(1 - \|Tf_y\|^2) - \langle Af_y, f_y\rangle - 2\langle a, f_y\rangle - \alpha \geq 0 \qquad \text{for all } h \in \ker(\Lambda).$$

More succinctly, with P representing the orthogonal projection onto $\ker(\Lambda)$ (so that $\mathrm{Id}_H - P$ is the orthogonal projection onto $\ker(\Lambda)^\perp = \mathrm{ran}(\Lambda^*)$, easily leading to the announced expression of P), this $c \geq 0$ must satisfy

$$\mathsf{constraint}_{A,a,\alpha}(c, d): \qquad \left[\begin{array}{c|c} P(cT^*T - A)P & -P(Af_y + a) \\ \hline -(Af_y + a)^*P & d - c(1 - \|Tf_y\|^2) - \langle Af_y, f_y\rangle - 2\langle a, f_y\rangle - \alpha \end{array}\right] \succeq 0.$$

All in all, incorporating $c \geq 0$ as an optimization variable, one arrives at

$$\sigma = \inf_{c \geq 0, d} d \quad \text{subject to } \mathsf{constraint}_{A,a,\alpha}(c, d),$$

so that σ is indeed equal to the optimal value d' of the first semidefinite program. As for ι, notice that $-\iota$ can be written as $\sup\{(-\gamma)(f) : \|Tf\| \leq 1, \Lambda(f) = y\}$, so it admits a similar expression, up to the change $(A, a, \alpha) \leftrightarrow (-A, -a, -\alpha)$, i.e., $-\iota = \inf\{d : c \geq 0, d \text{ satisfy } \mathsf{constraint}_{-A,-a,-\alpha}(c, d)\}$. As a result, one obtains $\iota = \sup\{-d : c \geq 0, d \text{ satisfy } \mathsf{constraint}_{-A,-a,-\alpha}(c, d)\}$. Changing d to $-d$ shows that ι is equal to the optimal value d'' of the second semidefinite program. $\square$

3 Globally Optimal Estimation Functionals: Existence

Generally speaking, a drawback of the local setting is that an expensive optimization program needs to be solved anew for each data vector $y \in \mathbb{R}^m$—except in favorable circumstances, e.g. when Γ is a linear map between Hilbert spaces and $\mathcal{K}$ is an ellipsoid, in which case the Chebyshev center depends linearly on y. This aspect partly explains the predominance of the global setting, considered in this section. Here, the quantity of interest γ is restricted to be a quadratic form, i.e.,

$$(7) \qquad \gamma(f) = \langle Af, f\rangle, \qquad f \in H,$$

for a self-adjoint operator A on a Hilbert space H. The *a posteriori* information represented by the linear map $\Lambda : H \to \mathbb{R}^m$ does not change, and there is almost no change either to the *a priori* information, represented by the model set

$$(8) \qquad \mathcal{K} = \{f \in H : \|Tf\| \leq 1\}, \qquad T : H \to H \text{ being an isomorphism.}$$

Note that the added requirement on T automatically guarantees that the assumption (6) holds. The main result of this section, so far nonconstructive, reads as follows.

Theorem 2. For a Hilbert space H, a linear observation map $\Lambda : H \to \mathbb{R}^m$, and the model set $\mathcal{K}$ of (8), one can find a quadratic functional which is a globally optimal functional for the estimation of the quadratic form (7), i.e., a map $\delta^{\rm opti} : \mathbb{R}^m \to \mathbb{R}$ written as

$$\delta^{\rm opti}(y) = \langle By, y\rangle + \beta, \qquad y \in \mathbb{R}^m,$$

for some symmetric matrix $B \in \mathbb{R}^{m\times m}$ and some scalar $\beta \in \mathbb{R}$ that satisfies

$$\mathrm{gwce}(\delta^{\rm opti}) \le \mathrm{gwce}(\delta) \qquad \text{for any } \delta : \mathbb{R}^m \to \mathbb{R}.$$

The argument will be a slick one, based on the following series three lemmas and on a refinement of Sion minimax theorem (see Theorem 9 in the appendix). The first lemma is a well known lower bound for the global worst-case error of any estimation functional.

Lemma 3. With $\Lambda : F \to \mathbb{R}^m$, $\mathcal{K} \subseteq F$, and $\gamma : F \to \mathbb{R}$ all being arbitrary, any estimation functional $\delta : \mathbb{R}^m \to \mathbb{R}$ satisfies

$$\mathrm{gwce}(\delta) \ge \mathrm{lb} := \sup_{\substack{f',f''\in\mathcal{K}\\ \Lambda f' = \Lambda f''}} \frac{\gamma(f') - \gamma(f'')}{2}.$$

The second lemma in the series concerns the global worst-case error of estimation functionals that are specifically quadratic.

Lemma 4. With an arbitrary linear map Λ from a Hilbert space H to $\mathbb{R}^m$, if $\mathcal{K} \subseteq H$ is convex and symmetric about the origin and if the functional $\gamma : H \to \mathbb{R}$ obeys $\gamma(-f) = \gamma(f)$ for all $f \in H$, then any quadratic functional of the form $\delta_{B,b,\beta}(y) = \langle By, y\rangle + 2\langle b, y\rangle + \beta$, $y \in \mathbb{R}^m$, satisfies

$$\mathrm{gwce}(\delta_{B,b,\beta}) \ge \mathrm{gwce}(\delta_B) = \frac{\sigma_B - \iota_B}{2},$$

where $\delta_B : \mathbb{R}^m \to \mathbb{R}$ is defined $y \in \mathbb{R}^m$ by $\delta_B(y) = \langle By, y\rangle + \dfrac{\sigma_B + \iota_B}{2}$ and where

$$\sigma_B := \sup_{f\in\mathcal{K}} \big(\gamma(f) - \langle B(\Lambda f), \Lambda f\rangle\big), \qquad \iota_B := \inf_{f\in\mathcal{K}} \big(\gamma(f) - \langle B(\Lambda f), \Lambda f\rangle\big).$$

The third lemma in the series is a simple but crucial observation for the application of the refined minimax theorem.

Lemma 5. If S is a self-adjoint operator on a Hilbert space H and if $|\cdot|$ is a seminorm on H, then the sets

$$\big\{f' f'^* - f'' f''^* : \ f', f'' \in H, |f'| \le 1, |f''| \le 1, \langle Sf', f'\rangle - \langle Sf'', f''\rangle \le c\big\}$$

are path-connected (hence connected) for all $c \in \mathbb{R}$.

The justification of this series of lemma is postponed for the short time it will take to deduce the main theorem from them.

Proof of Theorem 2. Lemma 4 indicates that the minimal global worst-case error of a quadratic estimation functional is achieved by a functional of the form $\delta_B(y) = \langle By, y\rangle + \beta$, $y \in \mathbb{R}^m$, for some symmetric matrix $B \in \mathbb{R}^{m\times m}$ and some scalar β depending on B. It also gives the value of the global worst-case error of such a δ_B: with $S_B := A - \Lambda^* B \Lambda$, it is

$$(9)\qquad \mathrm{gwce}(\delta_B) = \frac{1}{2}\left(\sup_{f'\in\mathcal{K}} \langle S_B f', f'\rangle - \inf_{f''\in\mathcal{K}} \langle S_B f'', f''\rangle \right) = \frac{1}{2} \sup_{f',f''\in\mathcal{K}} \big(\langle S_B f', f'\rangle - \langle S_B f'', f''\rangle \big)$$
$$= \frac{1}{2} \sup_{f',f''\in\mathcal{K}} \mathrm{tr}\big(S_B (f' f'^* - f'' f''^*)\big).$$

Consequently, one can assert that

$$\inf_{\delta \text{ quadratic}} \mathrm{gwce}(\delta) = \frac{1}{2} \inf_{B\in\mathcal{Y}} \max_{F\in\mathcal{X}} \mathrm{tr}(S_B F),$$

where $\mathcal{Y} := \{B \in \mathbb{R}^{m\times m} : B^* = B\}$ and $\mathcal{X} = \{f' f'^* - f'' f''^* : f', f'' \in H, \|Tf'\| \le 1, \|Tf''\| \le 1\}$ are respectively convex and compact sets—the compacteness is relative to the weak*-topology, thanks to Banach–Alaoglu theorem, since $\mathcal{X}$ is closed and bounded (due to T being an isomorphism). Moreover, for any $F \in \mathcal{X}$, the map $B \in \mathcal{Y} \mapsto -\mathrm{tr}(S_B F) = -\mathrm{tr}(AF) + \mathrm{tr}(\Lambda^* B \Lambda F)$ is affine, hence is upper semicontinuous and has convex strict superlevel sets, while for any $B \in \mathcal{Y}$, the map $F \in \mathcal{X} \mapsto -\mathrm{tr}(S_B F)$ is continuous, hence is lower semicontinuous, and has connected sublevel sets by Lemma 5. Consequently, Sion minimax theorem refined as Theorem 9 applies to give

$$\sup_{B\in\mathcal{Y}} \min_{F\in\mathcal{X}} (-\mathrm{tr}(S_B F)) = \min_{F\in\mathcal{X}} \sup_{B\in\mathcal{Y}} (-\mathrm{tr}(S_B F)).$$

Since switching the sign has the effect of replacing the sup's by inf's and the min's by max's, one arrives at

$$\inf_{\delta \text{ quadratic}} \mathrm{gwce}(\delta) = \frac{1}{2} \max_{F\in\mathcal{X}} \inf_{B\in\mathcal{Y}} \mathrm{tr}(S_B F) = \frac{1}{2} \max_{F\in\mathcal{X}} \inf_{B\in\mathcal{Y}} \big(\mathrm{tr}(AF) - \mathrm{tr}(B\Lambda F \Lambda^*)\big)$$
$$= \frac{1}{2} \max_{F\in\mathcal{X}} \begin{cases} \mathrm{tr}(AF) & \text{if } \Lambda F \Lambda^* = 0 \\ -\infty & \text{otherwise} \end{cases}$$
$$= \frac{1}{2} \max_{F\in\mathcal{X}} \Big\{ \mathrm{tr}(AF) : \Lambda F \Lambda^* = 0 \Big\}.$$

Notice that the constraint $\Lambda F \Lambda^* = 0$ for $F = f' f'^* - f'' f''^*$ reads $\Lambda f' f'^* \Lambda^* = \Lambda f'' f''^* \Lambda^*$, which in turn is equivalent to $\Lambda f' = \pm \Lambda f''$: indeed, $\Lambda f'$ and $\Lambda f''$ are colinear because both span the range of $\Lambda f' f'^* \Lambda^* = \Lambda f'' f''^* \Lambda^*$ and their norms are seen to be equal by taking the trace. Therefore, replacing f'' by $-f''$ is necessary, it follows that

$$\inf_{\delta \text{ quadratic}} \mathrm{gwce}(\delta) = \frac{1}{2} \sup_{\substack{f',f''\in\mathcal{K} \\ \Lambda f' = \Lambda f''}} \mathrm{tr}\big(A(f' f'^* - f'' f''^*)\big) = \frac{1}{2} \sup_{\substack{f',f''\in\mathcal{K} \\ \Lambda f' = \Lambda f''}} \big(\langle Af', f'\rangle - \langle Af'', f''\rangle\big).$$

Recognizing the lower bound lb from Lemma 3, one concludes that $\inf\{\mathrm{gwce}(\delta) : \delta \text{ quadratic}\}$ is as small as the smallest global worst-case error of any estimation functional $\delta : \mathbb{R}^m \to \mathbb{R}^m$, which is the desired result. □

To make the argument complete, one now turns to the justification of the series of lemma.

Proof of Lemma 3. For $f', f'' \in \mathcal{K}$ with $\Lambda f' = \Lambda f''$, let y denote this common value. One has

$$\begin{aligned}\frac{\gamma(f') - \gamma(f'')}{2} &= \frac{1}{2} \quad (\gamma(f') - \delta(y)) \ + \frac{1}{2} \quad (\delta(y) - \gamma(f''))\\ &\leq \frac{1}{2} \sup_{f \in \mathcal{K}} |\gamma(f) - \delta(\Lambda f)| + \frac{1}{2} \sup_{f \in \mathcal{K}} |\gamma(f) - \delta(\Lambda f)| = \mathrm{gwce}(\delta).\end{aligned}$$

The announced result follows by taking the supremum over these f' and f''. □

Proof of Lemma 4. As a first step, one observes that the global worst-case error of $\delta_{B,b,\beta}$ is minimized when $b = 0$. To see this, notice that $\delta_{B,0,\beta}(y) = \big(\delta_{B,b,\beta}(y) + \delta_{B,b,\beta}(-y)\big)/2$ for any $y \in \mathbb{R}^m$ and that, for $f \in \mathcal{K}$, and hence $-f \in \mathcal{K}$, one has

$$\begin{aligned}|\gamma(f) - \delta_{B,0,\beta}(\Lambda f)| &= \frac{1}{2} \left| \big[\gamma(f) - \delta_{B,b,\beta}(\Lambda(f))\big] + \big[\gamma(-f) - \delta_{B,b,\beta}(\Lambda(-f))\big] \right|\\ &\leq \frac{1}{2} \left|\gamma(f) - \delta_{B,b,\beta}(\Lambda(f))\right| + \frac{1}{2} \left|\gamma(-f) - \delta_{B,b,\beta}(\Lambda(-f))\right|\\ &\leq \frac{1}{2} \mathrm{gwce}(\delta_{B,b,\beta}) + \frac{1}{2} \mathrm{gwce}(\delta_{B,b,\beta}) = \mathrm{gwce}(\delta_{B,b,\beta}).\end{aligned}$$

Taking the supremum over $f \in \mathcal{K}$ yields $\mathrm{gwce}(\delta_{B,0,\beta}) \leq \mathrm{gwce}(\delta_{B,b,\beta})$, as stated.

As a second step, one establishes that $(\sigma_B + \iota_B)/2$ is the optimal value of the parameter β. Indeed, considering f^+, f^- achieving (or coming arbitrarily close to) the supremum and infimum defining σ_B and ι_B, one has

$$\mathrm{gwce}(\delta_{B,0,\beta}) = \sup_{f \in \mathcal{K}} |\gamma(f) - \langle B(\Lambda f), \Lambda f\rangle - \beta| \geq \begin{cases} \gamma(f^+) - \langle B(\Lambda f^+), \Lambda f^+\rangle - \beta &= \sigma_B - \beta, \\ -\gamma(f^-) + \langle B(\Lambda f^-), \Lambda f^-\rangle + \beta &= -\iota_B + \beta, \end{cases}$$

so the lower estimate $\mathrm{gwce}(\delta_{B,0,\beta}) \geq (\sigma_B - \iota_B)/2$ follows by averaging these two values. This lower estimate is indeed achieved at δ_B, i.e., with $\beta = \beta^{\mathrm{opti}} := (\sigma_B + \iota_B)/2$, since then one has, for any $f \in \mathcal{K}$,

$$\gamma(f) - \langle B(\Lambda f), \Lambda f\rangle - \beta^{\mathrm{opti}} \begin{cases} \leq \sigma_B - \dfrac{\sigma_B + \iota_B}{2} &= +\dfrac{\sigma_B - \iota_B}{2}, \\ \geq \iota_B - \dfrac{\sigma_B + \iota_B}{2} &= -\dfrac{\sigma_B - \iota_B}{2}. \end{cases}$$

This means that $\mathrm{gwce}(\delta_B) = \sup_{f \in \mathcal{K}} |\gamma(f) - \langle B(\Lambda f), \Lambda f\rangle - \beta^{\mathrm{opti}}| \leq (\sigma_B - \iota_B)/2$, as desired. □

Proof of Lemma 5. Let $\{f'f'^* - f''f''^* : f', f'' \in H, |f'| \le 1, |f''| \le 1, \langle Sf', f'\rangle - \langle Sf'', f''\rangle \le c\}$ be denoted by $\mathcal{X}_{S,c}$ and let $F_0, F_1 \in \mathcal{X}_{S,c}$ be written as $F_{0/1} = f'_{0/1}{f'_{0/1}}^* - f''_{0/1}{f''_{0/1}}^*$ with $|f'_{0/1}| \le 1$, $|f''_{0/1}| \le 1$, and $\langle Sf'_{0/1}, f'_{0/1}\rangle - \langle Sf''_{0/1}, f''_{0/1}\rangle \le c$. Let also signs $\varepsilon', \varepsilon'' \in \{\pm 1\}$ be specified as $\varepsilon' = -\mathrm{sign}\langle Sf'_0, f'_1\rangle$ and $\varepsilon'' = +\mathrm{sign}\langle Sf''_0, f''_1\rangle$. For $\tau \in [0,1]$, one defines $f'(\tau) := (1-\tau)f'_0 + \tau\varepsilon' f'_1$ and $f''(\tau) := (1-\tau)f''_0 + \tau\varepsilon'' f''_1$. The map $\tau \in [0,1] \mapsto f'(\tau)f'(\tau)^* - f''(\tau)f''(\tau)^*$ is a continuous path joining $F_0 = f'_0{f'_0}^* - f''_0{f''_0}^*$ (since $f'(0) = f'_0$ and $f''(0) = f''_0$) to $F_1 = f'_1{f'_1}^* - f''_1{f''_1}^*$ (since $f'(1) = \varepsilon' f'_1$ and $f''(1) = \varepsilon'' f''_1$). To make sure that this map does not leave the set $\mathcal{X}_{S,c}$, it remains to verify that $|f'(\tau)| \le 1$, $|f''(\tau)| \le 1$, and $\langle Sf'(\tau), f'(\tau)\rangle - \langle Sf''(\tau), f''(\tau)\rangle \le c$ for all $\tau \in [0,1]$. The former two simply result from a triangle inequality, while the latter one follows from

$$\begin{aligned}
\langle Sf'(\tau),&f'(\tau)\rangle - \langle Sf''(\tau), f''(\tau)\rangle \\
&= (1-\tau)^2 \times \begin{bmatrix} \langle Sf'_0, f'_0\rangle \\ -\langle Sf''_0, f''_0\rangle \end{bmatrix} + \tau^2 \times \begin{bmatrix} \langle Sf'_1, f'_1\rangle \\ -\langle Sf''_1, f''_1\rangle \end{bmatrix} + 2(1-\tau)\tau \times \begin{bmatrix} \epsilon'\langle Sf'_0, f'_1\rangle \\ -\epsilon''\langle Sf''_0, f''_1\rangle \end{bmatrix} \\
&\le (1-\tau)^2 \times c + \tau^2 \times c + 2(1-\tau)\tau \times 0 = \big((1-\tau)^2 + \tau^2\big) \times c \le c.
\end{aligned}$$

□

Remark. Although not needed here, it is worth pointing out that in the expression (see (9))

$$\mathrm{gwce}(\delta_{\mathrm{B}}) = \frac{1}{2} \sup_{\substack{\|Tf'\| \le 1 \\ \|Tf''\| \le 1}} \mathrm{tr}\big((A - \Lambda^* B \Lambda)(f'f'^* - f''f''^*)\big)$$

of the global worst-case error of a quadratic functional relative to a symmetric matrix $B \in \mathbb{R}^{m \times m}$, it can be assumed that the maximizers $\widehat{f}'$ and $\widehat{f}''$ satisfy the orthogonality condition $T\widehat{f}' \perp T\widehat{f}''$. To see this, consider the eigendecomposition of the symmetric rank-2 matrix $M = T(\widehat{f}'\widehat{f}'^* - \widehat{f}''\widehat{f}''^*)T^*$ written as

$$T(\widehat{f}'\widehat{f}'^* - \widehat{f}''\widehat{f}''^*)T^* = \lambda g' g'^* + \mu g'' g''^* \qquad \text{with} \qquad \lambda \ge \mu,\ \|g'\| = 1,\ \|g''\| = 1,\ g' \perp g''.$$

The characterization of extremal eigenvalues yields $\lambda \in [0,1]$ and $\mu \in [-1,0]$, for instance

$$\lambda = \max_{\|g\| \le 1} \langle Mg, g\rangle = \max_{\|g\| \le 1} \big(\langle T\widehat{f}', g\rangle^2 - \langle T\widehat{f}'', g\rangle^2\big) \begin{cases} \le \max_{\|g\| \le 1} \quad \|T\widehat{f}'\|^2\, \|g\|^2 & \le 1, \\ \ge \max_{\|g\| \le 1, g \perp T\widehat{f}''} \langle T\widehat{f}', g\rangle^2 & \ge 0. \end{cases}$$

Then, setting $\bar{f}' := \sqrt{\lambda}\, T^{-1} g'$ and $\bar{f}'' := \sqrt{-\mu}\, T^{-1} g''$, one sees that

$$\begin{aligned}
\mathrm{tr}\big((A - \Lambda^* B\Lambda)(\bar{f}'\bar{f}'^* - \bar{f}''\bar{f}''^*)\big) &= \mathrm{tr}\big((A - \Lambda^* B\Lambda) T^{-1}(\lambda g' g'^* + \mu g'' g''^*) T^{-*}\big) \\
&= \mathrm{tr}\big((A - \Lambda^* B\Lambda)(\widehat{f}'\widehat{f}'^* - \widehat{f}''\widehat{f}''^*)\big) \\
&= \mathrm{gwce}(\delta_{\mathrm{B}}),
\end{aligned}$$

while they clearly satisfy $\|T\bar{f}'\| \le 1$, $\|T\bar{f}''\| \le 1$, as well as the orthogonality condition $T\bar{f}' \perp T\bar{f}''$.

4 Globally Optimal Estimation Functionals: Computation

Under the scenario carried over from the previous section, the focus is now put on the practical construction of a quadratic functional that is globally optimal for the estimation of a quadratic form. Although Theorem 2 guaranteed its existence, it contained no indication on how to produce such an optimal quadratic functional. The result below offers a construction. As in Section 2, since an optimization program eventually needs to be solved, the assumption of finite-dimensionality—allowing linear operators to be identified with matrices—is not seen as restrictive.

Theorem 6. For a finite-dimensional Hilbert space H, a linear observation map $\Lambda : H \to \mathbb{R}^m$, and the model set $\mathcal{K}$ of (8), a globally optimal functional for the estimation of the quadratic form (7) is given as

$$\delta^{\rm opti}(y) = \langle B^{\rm opti} y, y\rangle + \beta^{\rm opti}, \qquad y \in \mathbb{R}^m,$$

where the symmetric matrix $B^{\rm opti} \in \mathbb{R}^{m\times m}$ is obtained as solution to the semidefinite program

$$\underset{\substack{c',c''\geq 0\\ B\,\text{symmetric}}}{\text{minimize}} \; \frac{c'+c''}{2} \quad \text{subject to} \quad \begin{cases} c'T^*T - A + \Lambda^* B \Lambda & \succeq 0,\\ c''T^*T + A - \Lambda^* B\Lambda & \succeq 0. \end{cases} \tag{10}$$

The solutions $\overline{c}', \overline{c}'' \in \mathbb{R}$ yield the optimal parameter $\beta^{\rm opti} = (\overline{c}' - \overline{c}'')/2$, as well as the minimal global worst-case estimation error $(\overline{c}' + \overline{c}'')/2$.

Proof. By Theorem 2, a globally optimal estimation functional can be selected to have the form $\delta(y) = \langle By, y\rangle + \beta$, $y \in \mathbb{R}^m$. For such quadratic functionals, Lemma 4 established that the optimal choice of β as well as the global worst-case error depend on two quantities σ_B and ι_B via

$$\beta^{\rm opti} = \frac{\sigma_B + \iota_B}{2} \qquad \text{and} \qquad \mathrm{gwce}(\delta_B) = \frac{\sigma_B - \iota_B}{2}.$$

The quantity σ_B, for instance, can be converted with the help of the S-lemma as follows (the strict feasibility condition being easily met by $f = 0$):

$$\begin{aligned}
\sigma_B &:= \sup_{\|Tf\|\leq 1} \langle (A - \Lambda^* B\Lambda) f, f\rangle \\
&= \inf_{d} \; d : \; \langle (A - \Lambda^* B\Lambda) f, f\rangle \leq d \text{ whenever } \|Tf\|^2 \leq 1 \\
&= \inf_{d} \; d : \; \exists c \geq 0 \text{ such that } \langle (A - \Lambda^* B\Lambda) f, f\rangle - d \leq c\big(\|Tf\|^2 - 1\big) \; \text{ for all } f \in H \\
&= \inf_{\substack{d\\ c\geq 0}} \; d : \; \langle (cT^*T - (A - \Lambda^* B\Lambda)) f, f\rangle + d - c \geq 0 \; \text{ for all } f \in H.
\end{aligned}$$

The latter constraint decouples as $\langle (cT^*T - A + \Lambda^* B\Lambda) f, f\rangle \geq 0$ for all $f \in H$ and $d - c \geq 0$, i.e., as $cT^*T - A + \Lambda^* B \Lambda \succeq 0$ and $d \geq c$, so the minimal admissible value for d is c, which implies that

$$\sigma_B = \inf_{c \geq 0} c : \; cT^*T - A + \Lambda^* B\Lambda \succeq 0.$$

The quantity $\iota_B = \inf\{\langle (A - \Lambda^* B\Lambda) f, f\rangle : \|Tf\| \leq 1\}$—or rather $-\iota_B$—is likewise converted to

$$-\iota_B = \inf_{c\geq 0} c : cT^*T + A - \Lambda^* B\Lambda \succeq 0.$$

Combining these two expressions with $\mathrm{gwce}(\delta_B) = (\sigma_B - \iota_B)/2$ justifies that, for a fixed B,

$$\mathrm{gwce}(\delta_B) = \inf_{c',c''\geq 0} \frac{c'+c''}{2} \quad \text{subject to} \quad \begin{cases} c'T^*T - A + \Lambda^* B \Lambda & \succeq 0, \\ c''T^*T + A - \Lambda^* B \Lambda & \succeq 0. \end{cases} \tag{11}$$

Further minimizing over symmetric matrices $B \in \mathbb{R}^{m\times m}$—which is computationally feasible thanks to the linear dependence on B of the constraints—leads to the announced semidefinite program. Note that the solutions $\bar{c}', \bar{c}'' \geq 0$ correspond to σ_B and $-\iota_B$ for the optimal B, justifying the final part of the statement. □

The above construction is implemented in the MATLAB file accompanying this article, which can be used to certify that the "plug-in estimate is not optimal". Precisely, by plug-in estimate, one means the functional δ^{plug} equal to $\gamma \circ \Delta^{\mathrm{opti}}$, plus an optimaly chosen scalar β, where Δ^{opti} is the locally (hence globally) optimal estimation map for $\Gamma = \mathrm{Id}_H$. This map associates to $y \in \mathbb{R}^m$ the Chebyshev center of $\mathcal{K}_y := \{f \in H : \|Tf\| \leq 1, \Lambda f = y\}$ and is known (see more in [6, Lemma 6]) to be $y \in \mathbb{R}^m \mapsto \mathrm{argmin}\,\{\|Tf\| : \Lambda f = y\}$, i.e., the linear map $\Delta^{\mathrm{opti}} = T^{-1}T^{-*}\Lambda^*(\Lambda T^{-1}T^{-*}\Lambda^*)^{-1}$. Thus, the plug-in functional δ^{plug} is a quadratic functional whose global-worst case error can be computed according to (11). Numerical tests reveal that in general it is strictly larger that the minimal global worst-case error as computed according to (10)—this conclusion is not apparent for a random T, but it is for $T = \mathrm{Id}_H$.

5 Incorporating Observation Errors

In practical situations, the vector $y \in \mathbb{R}^m$ is likely to be contaminated by observation errors, so that $y = \Lambda(f) + e$ for some nonzero $e \in \mathbb{R}^m$. This error vector can be modeled stochastically, as in [2, 8], or deterministically, as in this article, by way of an uncertainty set

$$\mathcal{E} = \{e \in \mathbb{R}^m : \|Ue\|_2 \leq \eta\}, \tag{12}$$

where U is an isomorphism on $\mathbb{R}^m$ and $\eta > 0$ is a bound on the error magnitude. It is well known that this "inaccurate scenario" can be reduced to the "accurate scenario". Precisely, the adjusted global-worst case error takes the form

$$\mathrm{gwce}(\delta) := \sup_{\substack{f\in\mathcal{K}\\ e\in\mathcal{E}}} |\gamma(f) - \delta(\Lambda f + e)| = \sup_{\widetilde{f}\in\widetilde{\mathcal{K}}} |\widetilde{\gamma}(\widetilde{f}) - \delta(\widetilde{\Lambda}\widetilde{f})|,$$

where the "tilde" objects are defined by $\widetilde{\mathcal{K}} := \mathcal{K} \times \mathcal{E} \subseteq \widetilde{H} := H \times \mathbb{R}^m$ and

$$\widetilde{f} := \begin{bmatrix} f \\ \hline e \end{bmatrix} \in \widetilde{H}, \quad \widetilde{\Lambda}\widetilde{f} := \Lambda f + e = \begin{bmatrix} \Lambda & | & \mathrm{Id}_{\mathbb{R}^m} \end{bmatrix} \begin{bmatrix} f \\ \hline e \end{bmatrix}, \quad \widetilde{\gamma}(\widetilde{f}) := \gamma(f) = \left\langle \left[\begin{array}{c|c} A & 0 \\ \hline 0 & 0 \end{array}\right] \begin{bmatrix} f \\ \hline e \end{bmatrix}, \begin{bmatrix} f \\ \hline e \end{bmatrix} \right\rangle.$$

This is a formal reduction and does not automatically allow one to call upon earlier results that relied on the specificity of the model set $\mathcal{K}$, especially when computational issues are concerned. For instance, here the original model set $\mathcal{K}$ was an ellipsoid but the new model set $\widetilde{\mathcal{K}}$ is the intersection of two ellipsoids, so the proof of Theorem 1, say, would not carry over, since the use of the S-lemma was contingent on $\mathcal{K}$ being described by the lone quadratic inequality $\|Tf\| \leq 1$. Fortuitously, there is an analog of the S-lemma covering the case of one quadratic inequality being a consequence of two—not one—quadratic inequalities. Discussed in [11], this result due to B. Polyak has already bear fruits in Optimal Recovery: it was invoked in [5] to show how the optimal parameter of a Tikhonov-style regularization (which is a globally optimal map for the estimation of the identity) can be computed.

Polyak S-procedure: *given three quadratic functionals of the specific form* $q_i(h) = \langle R_i h, h\rangle + \rho_i$, $i = 0, 1, 2$, *defined on a Hilbert space* H *with* $\dim(H) \geq 3$, *the equivalence*

$$[q_0(h) \leq 0 \ \ \textit{whenever}\ q_1(h) \leq 0 \ \textit{and}\ q_2(h) \leq 0] \iff [\exists c_1, c_2 \geq 0 : q_0 \leq c_1\, q_1 + c_2\, q_2]$$

holds provided $q_1(h) < 0$ *and* $q_2(h) < 0$ *for some* $h \in H$ *and* $d_1 R_1 + d_2 R_2 \succ 0$ *for some* $d_1, d_2 \in \mathbb{R}$.

Although of no assistance in the local setting, this S-procedure is quite fitted to the global setting. Indeed, the main result for the "accurate scenario" admits the following generalization.

Theorem 7. Consider a Hilbert space H, a linear observation map $\Lambda : H \to \mathbb{R}^m$, the model set $\mathcal{K}$ of (8), and the uncertainty set $\mathcal{E}$ of (12). For the estimation of the quadratic form (7), there exists a globally optimal functional given as a quadratic functional

$$\delta^{\rm opti}(y) = \langle B^{\rm opti} y, y\rangle + \beta^{\rm opti}, \qquad y \in \mathbb{R}^m.$$

If H is finite-dimensional, then the symmetric matrix $B^{\rm opti} \in \mathbb{R}^{m \times m}$ is obtained by solving the semidefinite program

$$\text{(13)}\quad \underset{\substack{c_1',c_2' \geq 0\\ c_1'',c_2'' \geq 0\\ B\,\text{symmetric}}}{\text{minimize}}\ \frac{c_1' + c_1'' + (c_2' + c_2'')\eta^2}{2} \quad \text{subject to} \quad \begin{cases} \left[\begin{array}{c|c} c_1' T^*T - A + \Lambda^* B \Lambda & \Lambda^* B \\ \hline B\Lambda & c_2' U^* U + B \end{array}\right] \succeq 0, \\ \left[\begin{array}{c|c} c_1'' T^*T + A - \Lambda^* B \Lambda & -\Lambda^* B \\ \hline -B\Lambda & c_2'' U^* U - B \end{array}\right] \succeq 0. \end{cases}$$

The solutions $\overline{c}_1', \overline{c}_2', \overline{c}_1'', \overline{c}_2'' \geq 0$ yield the optimal parameter $\beta^{\rm opti} = \big((\overline{c}_1' - \overline{c}_1'') + (\overline{c}_2' - \overline{c}_2'')\eta^2\big)/2$, as well as the minimal global worst-case estimation error equal to $\big((\overline{c}_1' + \overline{c}_1'') + (\overline{c}_2' + \overline{c}_2'')\eta^2\big)/2$.

Proof. For the existence part, the proof of Theorem 2 carries over, *mutatis mutandis*, using the "tilde" objects. In particular, Lemma 5 shall be applied with a seminorm defined for $\widetilde{f} = [f; e] \in \widetilde{H}$ by $|\widetilde{f}| := \max\{\|Tf\|, \|Ue\|_2/\eta\}$.

Now, for the computation of an optimal quadratic estimation functional δ_B, one needs to minimize, with obvious notation, $(\widetilde{\sigma}_B - \widetilde{\iota}_B)/2$. Concentrating on the first quantity, it is

$$\begin{aligned}\widetilde{\sigma}_B &:= \sup_{\widetilde{f} \in \widetilde{\mathcal{K}}} \langle (\widetilde{A} - \widetilde{\Lambda}^* B \widetilde{\Lambda}) \widetilde{f}, \widetilde{f} \rangle \\ &= \inf_d d: \ \langle (\widetilde{A} - \widetilde{\Lambda}^* B \widetilde{\Lambda}) \widetilde{f}, \widetilde{f} \rangle \le d \ \text{ whenever } \|\widetilde{T}\widetilde{f}\|^2 \le 1 \text{ and } \|\widetilde{U}\widetilde{f}\|^2 \le \eta^2.\end{aligned}$$

From the S-procedure, which applies since the strict feasibility constraint is met with $\widetilde{f} = 0$ and with any $d_1, d_2 > 0$ (recall that T and U are invertible), the above constraint is equivalent to the existence of $c_1, c_2 \ge 0$ such that

$$\langle (\widetilde{A} - \widetilde{\Lambda}^* B \widetilde{\Lambda}) \widetilde{f}, \widetilde{f} \rangle - d \le c_1 (\|\widetilde{T}\widetilde{f}\|^2 - 1) + c_2 (\|\widetilde{U}\widetilde{f}\|^2 - \eta^2) \quad \text{for all } \widetilde{f} \in \widetilde{H},$$

which decouples as $\langle (c_1 \widetilde{T}^*\widetilde{T} + c_2 \widetilde{U}^*\widetilde{U} - \widetilde{A} + \widetilde{\Lambda}^* B \widetilde{\Lambda}) \widetilde{f}, \widetilde{f} \rangle \ge 0$ for all $\widetilde{f} \in \widetilde{H}$ and $d \ge c_1 + c_2 \eta^2$. Incorporating $c_1, c_2 \ge 0$ as optimization variables, taking d as its minimal admissible value, and spelling out the form of $\widetilde{A}$, $\widetilde{\Lambda}$, $\widetilde{T}$, and $\widetilde{U}$ leads to

$$\widetilde{\sigma}_B = \inf_{c_1, c_2 \ge 0} c_1 + c_2 \eta^2 : \ \left[\begin{array}{c|c} c_1 T^* T - A + \Lambda^* B \Lambda & \Lambda^* B \\ \hline B \Lambda & c_2 U^* U + B \end{array} \right] \succeq 0.$$

In a very similar way, one derives

$$\begin{aligned}-\widetilde{\iota}_B &:= \sup_{\widetilde{f} \in \widetilde{\mathcal{K}}} \langle (-\widetilde{A} + \widetilde{\Lambda}^* B \widetilde{\Lambda}) \widetilde{f}, \widetilde{f} \rangle \\ &= \inf_{c_1, c_2 \ge 0} c_1 + c_2 \eta^2 : \ \left[\begin{array}{c|c} c_1 T^* T + A - \Lambda^* B \Lambda & -\Lambda^* B \\ \hline -B \Lambda & c_2 U^* U - B \end{array} \right] \succeq 0.\end{aligned}$$

Putting these two infima together in $(\widetilde{\sigma}_B - \widetilde{\iota}_B)/2$ while further minimizing over symmetric matrices $B \in \mathbb{R}^{m \times m}$ results in the semidefinite program (13). Finally, keeping in mind that $\widetilde{\sigma}_B$ corresponds to $\bar{c}_1' + \bar{c}_2' \eta^2$ and that $-\widetilde{\iota}_B$ corresponds to $\bar{c}_1'' + \bar{c}_2'' \eta^2$, the values of the optimal parameter β^{opti} and of the minimal global worst-case error follow. □

6 Postlude on Computational Discovery

I wish to thank Chunyang Liao for asking about the optimal estimation of quadratic functionals at the end of a presentation of the results from [4]. It prompted me to investigate this question, numerically at first because a way to compute the minimal global worst-case error over quadratic functionals was relatively easy to come by (see Theorem 6), so an answer could be speculated if there was also a way to compute the well-known lower bound on the global worst-case error of any estimation functional (see Lemma 3). And there was a way—it led me to the empirical conviction that "optimal estimation of quadratic functionals can be achieved with quadratic functionals".

This discovery owes a lot to the availability of computational tools such as CVX [7]. Later came the formal justification presented in this article, whose unfortunate consequence is to erase a decisive part of the process. Still, I believe that it is well worth mentioning it here. The main tool, also discussed in [11], is again related to the S-lemma.

<u>Finsler lemma:</u> *given two self-adjoint operators R_0, R_1 defined on a finite-dimensional Hilbert space H, the following equivalence holds:*

$$[\langle R_0 f, f\rangle > 0 \text{ whenever } \langle R_1 f, f\rangle = 0 \text{ and } f \neq 0] \iff [\exists d \in \mathbb{R} : R_0 + d\, R_1 \succ 0].$$

Proposition 8. For a finite-dimensional Hilbert space H, a linear observation map $\Lambda : H \to \mathbb{R}^m$, and the model set $\mathcal{K}$ of (8), the lower bound lb on the global worst-case error of any functional for the estimation of the quadratic form (7), as presented in Lemma 3, can be computed as the optimal value of the semidefinite program

$$\text{(14)} \qquad \underset{\substack{c_1,c_2 \geq 0 \\ d}}{\text{minimize}} \; \frac{c_1+c_2}{2} \quad \text{subject to} \quad \left[\begin{array}{c|c} c_1 T^*T - A + d\Lambda^*\Lambda & -d\Lambda^*\Lambda \\ \hline -d\Lambda^*\Lambda & c_2 T^*T + A + d\Lambda^*\Lambda \end{array}\right] \succeq 0.$$

Proof. Introducing the notation $\widehat{f} := (f', f'') \in \widehat{H} := H \times H$, $\widehat{T}'(\widehat{f}) := T(f')$, and $\widehat{T}''(\widehat{f}) := T(f'')$, as well as $\widehat{\Lambda}(\widehat{f}) = \Lambda(f') - \Lambda(f'')$, the lower bound takes the form

$$\begin{aligned} \mathrm{lb} &= \sup_{f',f''} \left\{ \frac{\langle Af', f'\rangle - \langle Af'', f''\rangle}{2} : \|Tf'\|^2 \leq 1, \|Tf''\|^2 \leq 1, \Lambda(f') = \Lambda(f'') \right\} \\ &= \sup_{\widehat{f}} \left\{ \frac{\langle \widehat{A}\widehat{f}, \widehat{f}\rangle}{2} : \|\widehat{T}'\widehat{f}\|^2 \leq 1, \|\widehat{T}''\widehat{f}\|^2 \leq 1, \widehat{f} \in \ker(\widehat{\Lambda}) \right\} \\ &= \inf_d \; d : \frac{\langle \widehat{A}\widehat{f}, \widehat{f}\rangle}{2} \leq d \text{ whenever } \widehat{f} \in \ker(\widehat{\Lambda}) \text{ satisfies } \|\widehat{T}'\widehat{f}\|^2 \leq 1, \|\widehat{T}''\widehat{f}\|^2 \leq 1. \end{aligned}$$

First, relying on Polyak S-procedure in the space $\ker(\widehat{\Lambda})$, one obtains (the strict feasibility condition being easily verified)

$$\begin{aligned} \mathrm{lb} &= \inf_d \; d : \exists c_1, c_2 \geq 0 \text{ such that } \langle \widehat{A}\widehat{f}, \widehat{f}\rangle - 2d \leq c_1(\|\widehat{T}'\widehat{f}\|^2 - 1) + c_2(\|\widehat{T}''\widehat{f}\|^2 - 1) \text{ for all } \widehat{f} \in \ker(\widehat{\Lambda}) \\ &= \inf_{c_1,c_2 \geq 0} \frac{c_1+c_2}{2} : \langle (c_1 \widehat{T}'^* \widehat{T}' + c_2 \widehat{T}''^* \widehat{T}'' - \widehat{A})\widehat{f}, \widehat{f}\rangle \geq 0 \text{ whenever } \langle \widehat{\Lambda}^* \widehat{\Lambda} \widehat{f}, \widehat{f}\rangle = 0. \end{aligned}$$

Note that this coincides with

$$\mathrm{lb}' := \inf_{c_1,c_2 \geq 0} \frac{c_1+c_2}{2} : \langle (c_1 \widehat{T}'^* \widehat{T}' + c_2 \widehat{T}''^* \widehat{T}'' - \widehat{A})\widehat{f}, \widehat{f}\rangle > 0 \text{ whenever } \langle \widehat{\Lambda}^* \widehat{\Lambda} \widehat{f}, \widehat{f}\rangle = 0 \text{ and } \widehat{f} \neq 0.$$

Indeed, on the one hand, if $c_1, c_2 \geq 0$ satisfy the constraint of lb', then they satisfy the constraint of lb, yielding $(c_1+c_2)/2 \geq \mathrm{lb}$, and in turn $\mathrm{lb}' \geq \mathrm{lb}$; on the other hand, given $\varepsilon > 0$, if $c_1, c_2 \geq 0$ satisfy the constraint of lb, then $c_1 + \varepsilon, c_2 + \varepsilon \geq 0$ satisfy the constraint of lb' (recall that T is invertible), yielding $(c_1+\varepsilon+c_2+\varepsilon)/2 \ge \mathrm{lb}'$, and in turn $\mathrm{lb} \ge \mathrm{lb}' - \varepsilon$. Therefore, Finsler lemma can be utilized to arrive at

$$\begin{aligned}\mathrm{lb} &= \inf_{c_1,c_2\ge 0} \frac{c_1+c_2}{2} \, : \; \exists d \in \mathbb{R} \text{ such that } c_1 \widehat{T}'^* \widehat{T}' + c_2 \widehat{T}''^* \widehat{T}'' - \widehat{A} + d \widehat{\Lambda}^* \widehat{\Lambda} \succ 0 \\ &= \inf_{\substack{c_1,c_2\ge 0\\ d}} \frac{c_1+c_2}{2} \, : \; c_1 \widehat{T}'^* \widehat{T}' + c_2 \widehat{T}''^* \widehat{T}'' - \widehat{A} + d \widehat{\Lambda}^* \widehat{\Lambda} \succ 0.\end{aligned}$$

As before, this infimum can be seen to equal the one with positive semidefiniteness replacing positive definiteness—although the straightforward inequality would be enough, as one seeks a lower bound. It now remains to spell out the form of $\widehat{A}$, $\widehat{\Lambda}$, $\widehat{T}'$, and $\widehat{T}''$ in order to conclude the proof. □

7 Appendix: The Refined Minimax Theorem

A crucial ingredient in the existence results proved in this article is a specific minimax theorem. The original minimax theorem von Neumann already admits several extensions that relax the convexity/concavity assumption, notably [3] where it is replaced by convexity-like/concavity-like (in the terminology of [12]) and [12] where it is replaced by quasiconvexity/quasiconvexity. The latter is almost what is needed in this article, except that asking sublevel sets to be convex is too much, yet asking them to be connected is enough. This was essentially present in the short proof of [9], but since the result was not stated as such there, a full argument is included here.

Theorem 9. Let $\mathcal{X}$ be a compact subset of a topological vector space and $\mathcal{Y}$ be a convex subset of a topological vector space. Assume that a function $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ satisfies

(i) for any $x \in \mathcal{X}$, the function $f(x,\cdot)$ is upper semicontinuous and has convex strict superlevel sets $\mathcal{Y}_{x,c} := \{y \in \mathcal{Y} : f(x,y) > c\}$;

(ii) for any $y \in \mathcal{Y}$, the function $f(\cdot,y)$ is lower semicontinuous and has connected sublevel sets $\mathcal{X}_{y,c} := \{x \in \mathcal{X} : f(x,y) \le c\}$.

Then

$$\sup_{y\in\mathcal{Y}} \min_{x\in\mathcal{X}} f(x,y) = \min_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} f(x,y).$$

Proof. Writing a for the left-hand side and b for the right-hand side, notice that $b \ge a$ is always valid. By contradiction, suppose $a < b$ and consider some $c \in (a,b)$. By (ii), each set $\mathcal{X}_y := \mathcal{X}_{y,c}$ is closed and connected. It is also nonempty, as even each $\mathcal{X}'_y := \{x \in \mathcal{X} : f(x,y) < c\} \subseteq \mathcal{X}_y$ is nonempty—otherwise $\min_{x\in\mathcal{X}} f(x,y) \ge c$, hence $a \ge c$. Notice that $\cap_{y\in\mathcal{Y}} \mathcal{X}_y = \emptyset$—otherwise, there would exist some $x \in \mathcal{X}$ such that $x \in \mathcal{X}_y$, i.e., $f(x,y) \le c$, for all $y \in \mathcal{Y}$, hence $b = \min_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} f(x,y) \le c$. By compactness of $\mathcal{X}$, it follows that there exists finitely many $y_1,\ldots,y_n \in \mathcal{Y}$ such that

$$\bigcap_{i=1}^{n} \mathcal{X}_{y_i} = \emptyset. \tag{15}$$

Moreover, with $\mathcal{C}$ denoting the convex hull of $y_1, \ldots, y_n$, if $y \in \mathcal{C}$, then any $x \in \mathcal{X}_y$ must belong to at least one $\mathcal{X}_{y_i}$—otherwise $f(x, y_i) > c$ for all i, so (i) would imply $f(x, y) > c$, i.e. $x \notin \mathcal{X}_y$. In other words, for any $y \in \mathcal{C} = \mathrm{conv}\{y_1, \ldots, y_n\}$, one has $\mathcal{X}_y \subseteq \bigcup_{i=1}^n \mathcal{X}_{y_i}$, or equivalently $\mathcal{X}_y = \bigcup_{i=1}^n \widetilde{\mathcal{X}}_{y_i}$, where $\widetilde{\mathcal{X}}_{y_i} := \mathcal{X}_{y_i} \cap \mathcal{X}_y$. This yields

$$\text{for any } y \in \mathcal{C}, \text{ there exists } i \in \{1, \ldots, n\} \text{ such that } \mathcal{X}_y \subseteq \mathcal{X}_{y_i}. \tag{16}$$

Indeed, because each $\mathcal{X}_{y_i}$ is closed, each set $\widetilde{\mathcal{X}}_{y_i}$ is closed in $\mathcal{X}_y$ and it is also open in $\mathcal{X}_y$, in view of $\widetilde{\mathcal{X}}_{y_i} = \mathcal{X}_y \setminus \cup_{j=1, j\neq i}^n \widetilde{\mathcal{X}}_{y_j}$. Therefore, selecting i such that $\widetilde{\mathcal{X}}_{y_i} \neq \emptyset$, the connectedness of $\mathcal{X}_y$ ensures that $\widetilde{\mathcal{X}}_{y_i} = \mathcal{X}_y$, or equivalently that $\mathcal{X}_y \subseteq \mathcal{X}_{y_i}$. Now, for each $i = 1, \ldots, n$, defining a set $\mathcal{C}_i := \{y \in \mathcal{C} : \mathcal{X}'_y \subseteq \mathcal{X}_{y_i}\}$, which is nonempty because $y_i \in \mathcal{C}_i$, the statement (16) implies that $\mathcal{C} = \cup_{i=1}^n \mathcal{C}_i$. Momentarily taking for granted that each $\mathcal{C}_i$ is closed in $\mathcal{C}$, and in turn also open in $\mathcal{C}$ in view of $\mathcal{C}_i = \mathcal{C} \setminus \cup_{j=1, j\neq i}^n \mathcal{C}_j$, the convexity, hence connectedness, of $\mathcal{C}$ ensures that $\mathcal{C}_i = \mathcal{C}$ for all $i = 1, \ldots, n$. Consequently, for any $y \in \mathcal{C}$, one has $\mathcal{X}'_y \subseteq \cap_{i=1}^n \mathcal{X}_{y_i}$. This emphatically contradicts (15) and invalidates the assumption that $a < b$.

It remains to justify that each $\mathcal{C}_i$ is closed in $\mathcal{C}$. To this end, let $(z_k)_{k \geq 1}$ be a sequence on $\mathcal{C}_i$ converging to some $z \in \mathcal{C}$, the goal being to show that $z \in \mathcal{C}_i$. Consider $x \in \mathcal{X}'_z$, meaning that $f(x, z) < c$. By the upper semicontinuity of $f(x, \cdot)$, one has $f(x, z_k) < c$ for some $k \geq 1$, i.e., $x \in \mathcal{X}'_{z_k}$. From $z_k \in \mathcal{C}_i$, i.e., $\mathcal{X}'_{z_k} \subseteq \mathcal{X}_{y_i}$, it follows that $x \in \mathcal{X}_{y_i}$. This shows that $\mathcal{X}'_z \subseteq \mathcal{X}_{y_i}$, i.e., $z \in \mathcal{C}_i$, as desired. □

References

[1] Bakhvalov, N. S. (1971). *On the optimality of linear methods for operator approximation in convex classes of functions.* USSR Computational Mathematics and Mathematical Physics, 11(4), 244–249.

[2] Donoho, D. L. and Nussbaum, M. (1990). *Minimax quadratic estimation of a quadratic functional.* Journal of Complexity, 6(3), 290–323.

[3] Fan, K (1953). *Minimax theorems.* Proceedings of the National Academy of Sciences, 39(1), 42–47.

[4] Foucart, S. (2025). *Optimal algorithms for nonlinear estimation with convex models.* arXiv preprint arXiv:2512.20826.

[5] Foucart, S. and Liao, C. (2023). *Optimal recovery from inaccurate data in Hilbert spaces: regularize, but what of the parameter?* Constructive Approximation, 57(2), 489–520.

[6] Foucart, S. and Hengartner, N. (2025). *Worst-case learning under a multifidelity model.* SIAM/ASA Journal on Uncertainty Quantification, 13(1), 171–194.

[7] Grant, M. and Boyd, S. (2013). CVX: Matlab software for disciplined convex programming, version 2.0 beta. `https://cvxr.com/cvx`

[8] Juditsky, A. and Nemirovski, A. (2020). *Estimating linear and quadratic forms via indirect observations*. Bernoulli, 26(4), 2639–2669.

[9] Komiya, H. (1988). *Elementary proof for Sion's minimax theorem*. Kodai Mathematical Journal, 11(1), 5–7.

[10] Micchelli, C. A. and Rivlin, T. J. (1985). *Lectures on optimal recovery*. In: Numerical Analysis Lancaster 1984, 21–93. Springer.

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

[12] Sion, M. (1958). *On general minimax theorems*. Pacific Journal of Mathematics, 8, 171–176.

[13] Smolyak, S. A. (1965). On Optimal Restoration of Functions and Functionals of Them. Candidate dissertation, Moscow State University.

[14] Sukharev, A. G. (1986). *On the existence of optimal affine methods for approximating linear functionals*. Journal of Complexity, 2(4), 317–322.

[15] Traub, J. F., Wasilkowski, G. W., and Wozniakowski, H. (1988). Information-Based Complexity. Academic Press.