

Worst-Case Learning under a Multi-fidelity Model

Simon Foucart* and Nicolas Hengartner†

Abstract

Inspired by multi-fidelity methods in computer simulations, this article introduces procedures to design surrogates for the input/output relationship of a high-fidelity code. These surrogates should be learned from runs of both the high-fidelity and low-fidelity codes and be accompanied by error guarantees that are deterministic rather than stochastic. For this purpose, the article advocates a framework tied to a theory focusing on worst-case guarantees, namely Optimal Recovery. The multi-fidelity considerations triggered new theoretical results in three scenarios: the globally optimal estimation of linear functionals, the globally optimal approximation of arbitrary quantities of interest in Hilbert spaces, and their locally optimal approximation, still within Hilbert spaces. The latter scenario boils down to the determination of the Chebyshev center for the intersection of two hyperellipsoids. It is worth noting that the mathematical framework presented here, together with its possible extension, seems to be relevant in several other contexts briefly discussed.

Key words and phrases: Multi-fidelity methods, Surrogate models, Optimal Recovery, Minimax problems, Chebyshev centers.

AMS classification: 65D15, 68Q99, 90C22, 90C47.

1 Introduction

High-fidelity computer models of complex physical systems provide some unprecedentedly detailed insights into processes ranging from climate response to anthropogenic greenhouse gas emissions [1], to the dynamics of turbulence and combustion of scramjets [2] and rocket motors [3], to modeling the dynamic responses of materials to shocks [4]. These models couple elementary physical, chemical, and biological processes at scale, leading to nonlinear dynamics that often reveal emergent phenomena [5].

*Texas A&M University

†Los Alamos National Laboratory

S. F. is partially supported by grants from the NSF (DMS-2053172) and from the ONR (N00014-20-1-2787). S. F. is also grateful for a Development Fellowship from the Texas A&M University System National Laboratories Office, which made visits to Los Alamos National Laboratory possible during 2023–24.

The computational cost of running high-fidelity models can be prohibitively high, however, which may limit their usefulness for some applications. Low-fidelity models of the high-fidelity codes rely on streamlined assumptions such as linearization, lower-dimensional approximations, simplified physics, complexity management limiting the number and order of interactions, and domain coarsening, to significantly reduce the computational burden. *Multi-fidelity* models [6, 7] combine the results from both low- and high-fidelity models, using the low-fidelity ones to extensively explore an approximation of the relationship between inputs and outputs and capitalizing on the higher fidelity ones to finetune this relationship. We note that the notions of high-fidelity and low-fidelity are only defined relative to one another.

Surrogate models are data-driven approximations of the relationship between inputs and outputs. They are essential ingredients in optimization [8] and in estimation and uncertainty quantification of computer models [9, 10]. Developing surrogate models for multi-fidelity computer models is challenging and remains an active field of research [11, 12]. While modern machine learning methods [13, 14, 15] have successfully combined low- and high-fidelity data, they are not well equipped to bound the pointwise prediction error. Multi-fidelity models based on Gaussian processes [16, 17] can provide stochastic bounds on the error. Yet, for some applications, deterministic bounds are desirable. The theory of Optimal Recovery—the centerpiece of this article, going back to the seventies [18]—is tailored to produce such deterministic bounds, e.g. for surrogates of computer simulations. These bounds offer a complementary view of the uncertainty associated with the surrogate, with a different interpretation but often similar ranges. See [19] for an example of comparison between stochastic and deterministic uncertainty quantification.

This article contributes to the existing body of research on surrogate modeling by extending the theory of Optimal Recovery to the design of surrogates for multi-fidelity computer simulations. The benefit of this approach is the explicit determination of upper bounds for the error of prediction, or even of full approximation, of the true input/output relationship by a surrogate constructed precisely to make this error small under some prior assumption about the relationship. Thus, it provides novel deterministic uncertainty quantification tools for multi-fidelity computer models.

In mathematical terms, let the input/output relationship for the high-fidelity code be represented by a function f_0 and let a function f_1 represent this relationship for the low-fidelity code. The prior assumption takes the form of a condition $f_0 \in \mathcal{K}_0$, together with a condition $f_0 - f_1 \in \mathcal{K}_1$ expressing the belief that, in a multi-fidelity setting, the bias $f_0 - f_1$ should be well-behaved, hopefully of lower complexity than f_0 . Executing the high- and low-fidelity codes provides the data

$$y_{0,i} = \lambda_{0,i}(f_0), \quad i = 1, \dots, m_0, \quad \text{and} \quad y_{1,i} = \lambda_{1,i}(f_1), \quad i = 1, \dots, m_1,$$

where the linear functionals $\lambda_{0,i}, \lambda_{1,i}$ are often (but not always) point evaluations. These data serve as constraints on the unknown functions, expressed via the sets $\mathcal{D}_0 = \{f : \lambda_{0,i}(f) = y_{0,i} \text{ for all } i\}$ and $\mathcal{D}_1 = \{f : \lambda_{1,i}(f) = y_{1,i} \text{ for all } i\}$. The error of full approximation, say, is obtained as the

optimal value of the min-max program

$$(1) \quad \inf_h \sup\{\|f_0 - h\| : f_0 \in \mathcal{K}_0 \cap \mathcal{D}_0, f_1 - f_0 \in \mathcal{K}_1, f_1 \in \mathcal{D}_1\},$$

while the minimizer h amounts to the optimal surrogate.

This formalism extends to more than two fidelity levels. Although conceptually straightforward, such an extension introduces theoretical challenges for solving suitable generalizations of (1), so our results primarily focus on two levels. Nonetheless, we want to make the point that the core mathematical problem arises in contexts beyond the multi-fidelity setting considered in this paper, and as such is of interest to a broad audience. Below is a non-exhaustive list of four examples that can benefit from our Optimal Recovery perspective.

Accelerated life testing. Accelerated life testing subjects items to conditions that enhances the physical and chemical processes associated with aging. For example, increasing the temperature of explosives promotes chemical degradation. We can quantify this effect in many ways, e.g. gas chromatography to characterize volatile and semi-volatile mixtures [20], possibly complemented with time-of-flight mass spectroscopy [21], and ‘dynamic testing’ to measure the energetic release. These destructive experiments capture physical and chemical characteristics that are differentially impacted by aging. For each experiment, we are interested in learning the expected response $f_t(x)$ to the t -th experiment, where x represents aging. The framework for Optimal Recovery in multi-fidelity experiments applies here, allowing us to combine and learn from multiple related tests.

Multi-physics sensing. Nuclear particles, such as electrons, muons, neutrons and protons penetrate and interact with matter in different but related ways. For example, muons and electrons interact with matter through electro-magnetic forces (Coulomb scattering) [22, 23], whereas neutrons scatter when colliding with the nucleus of atoms. Both of these processes are enhanced in high- Z materials and there is interest in combining the probing ability of multiple types of particles. The presented Optimal Recovery framework applies here, too, as we seek to learn the function $f_t(x, z)$, where x is the spatial position in a probed object and z are characteristics of the individual particles of type t .

Panel data. Panel data consists of short time series observed on multiple individuals, with common dynamic for its temporal evolution [24], possibly depending on covariates x . For times $t = 0, 1, \dots, T$, we denote by $f_t(x)$ the response at time t as a function of covariate x . The temporal evolution of the process makes it plausible that measurements at time t may inform, or at least bound, the measurements at time t' . That is, we can again leverage the theory in this paper to combine data from multiple time points to estimate the dependence of the response at time t on the covariate x , even if the measurements are from different items.

Theoretical tool for transfer learning. Transfer learning has emerged as a powerful machine-learning paradigm to combine related yet distinct datasets, see the recent review articles [25, 26]. The formalism in this paper provides a way to quantify the benefit of combining datasets by computing uncertainty bound of surrogates with and without the addition of ‘related datasets’.

2 Preliminaries, Problem Formulation, and Summary of Results

In this section, we fix the notation, introduce the concepts, and provide a self-contained overview of *Optimal Recovery*, specialized to the problem of building surrogates in our multi-fidelity context. In addition to [18], the books [27, 28] can provide the interested reader with a more modern discussion about the theory and applications of Optimal Recovery.

In this field, the goal is to exploit the assumption that an unknown target function f_0 belongs to a given subset $\mathcal{K} \subseteq F$ of a Banach space F , together with observed constraints $y_{0,1}, \dots, y_{0,m_0}$, which we call data, to bound a quantify of interest $Q(f_0)$. For simplicity, the latter is taken to be a linear map $Q : F \rightarrow Z$ into another Banach space Z . This framework includes the problem of full approximation of the function f_0 by setting $Q = \text{Id}$, the identify map. The Banach space F already encapsulates some *a priori* assumptions about f_0 and the so-called model set \mathcal{K}_0 further reflects our educated knowledge about realistic objects.

Throughout this paper, we will suppose that the data constraints

$$y_{0,i} = \lambda_{0,i}(f_0), \quad i = 1, \dots, m_0,$$

are obtained by applying linear functionals $\lambda_{0,1}, \dots, \lambda_{0,m_0}$ to f_0 . It is convenient to vectorize these data and use the notation $y_0 = \Lambda_0 f_0 = [\lambda_{0,1}(f_0); \dots; \lambda_{0,m_0}(f_0)] \in \mathbb{R}^{m_0}$ and refer to Λ_0 as the observation map from F to \mathbb{R}^{m_0} . A special case of particular interest is when $\lambda_{0,i}(f_0) = f_0(x^{(0,i)})$ is a point evaluation functional. This evokes similarities with Statistical Learning Theory. However, the main difference is that the $x^{(i)}$'s are regarded as fixed here and not as independent realizations of a random variable. As a result, the performance of a learning/recovery procedure cannot be assessed in an average case. Thus, one opts for an assessment focusing on the worst case, see (2) and (3) below, which is a distinctive feature of the Optimal Recovery framework.

In our specific setting, let us recall that multi-fidelity supplies an extra dimension via some side information on an object f_1 related to f_0 . Indeed, the latter is available through its own *a priori* assumption on the bias $f_0 - f_1$ expressed via

$$f_0 - f_1 \in \mathcal{K}_1.$$

For instance, if we have knowledge that f_1 is ε -close to f_0 , then \mathcal{K}_1 would simply be a ball of radius ε . As for the data constraints, they are observations made directly on f_1 , i.e.,

$$y_{1,i} = \lambda_{1,i}(f_1), \quad i = 1, \dots, m_1.$$

These data are again vectorized as $y_1 = \Lambda_1 f_1$, where $\Lambda_1 : F \rightarrow \mathbb{R}^{m_1}$ is a linear map.

To estimate $Q(f_0) \in Z$, we want to produce an approximant \hat{z} built from the combined data $y = [y_0; y_1] \in \mathbb{R}^{m_0+m_1}$. In other words, we want to construct a mapping Δ from $\mathbb{R}^{m_0+m_1}$ to Z ,

which we call a recovery map. Note that it may be cognizant of the model sets \mathcal{K}_0 and \mathcal{K}_1 . We are interested in *optimal* recovery maps, the meaning of which being tied to how the performance of a generic Δ is assessed. The value of $\|Q(f_0) - \Delta(y)\|_Z$ evidently quantifies the recovery error for fixed f_0 and f_1 , but since f_0 and f_1 are unknown, we settle on a worst-case viewpoint to quantify performance by taking the supremum over all f_0 and f_1 that are consistent with model and data. This leads to the following two notions of worst-case recovery error:

- the local worst-case error, at a fixed $y = [y_0; y_1] \in \mathbb{R}^{m_0+m_1}$, is defined as

$$(2) \quad \text{lwce}_y(z) = \sup_{\substack{f_0 \in \mathcal{K}_0, f_0 - f_1 \in \mathcal{K}_1 \\ \Lambda_0 f_0 = y_0, \Lambda_1 f_1 = y_1}} \|Q(f_0) - z\|_Z;$$

- the global worst-case error is defined as

$$(3) \quad \text{gwce}(\Delta) = \sup_{f_0 \in \mathcal{K}_0, f_0 - f_1 \in \mathcal{K}_1} \|Q(f_0) - \Delta([\Lambda_0 f_0; \Lambda_1 f_1])\|_Z.$$

A globally, resp. locally, optimal recovery map $\Delta^{\text{opt}} : \mathbb{R}^{m_0+m_1} \rightarrow Z$ is a map that minimizes $\text{gwce}(\Delta)$ over all $\Delta : \mathbb{R}^{m_0+m_1} \rightarrow Z$, resp. such that $\Delta^{\text{opt}}(y)$ minimizes $\text{lwce}_y(z)$ over all $z \in Z$ at every $y \in \mathbb{R}^{m_0+m_1}$. It is straightforward to see that a locally optimal recovery map is automatically globally optimal, implying that globally optimal recovery maps are somewhat easier to come by than locally optimal recovery maps. This will also be exemplified by the representative cases investigated in the remainder of this article. Treated by increasing order of difficulty, they leverage some techniques that have recently proved useful in Optimal Recovery. Below is an summary of some results for the three scenarios we considered.

In the first scenario (Section 3), the quantity of interest is required to be a linear functional. For instance, in the space $F = C(\mathcal{X})$ of continuous functions on a compact set \mathcal{X} , suppose that the observations $\lambda_{0,i}(f_0) = f_0(x^{(0,i)})$ and $\lambda_{1,i}(f_1) = f_1(x^{(1,i)})$ are point evaluations and that the quantity of interest $Q(f_0) = f_0(x)$ is the evaluation at another point $x \in \mathcal{X}$, i.e., we are trying to predict f_0 at new point. Suppose also that the model sets are based on approximation capabilities by linear subspaces $\mathcal{V}_0, \mathcal{V}_1$ of $C(\mathcal{X})$ with parameters $\varepsilon_0, \varepsilon_1 > 0$, i.e.,

$$\mathcal{K}_0 = \{g \in C(\mathcal{X}) : \text{dist}(g, \mathcal{V}_0) \leq \varepsilon_0\} \quad \text{and} \quad \mathcal{K}_1 = \{g \in C(\mathcal{X}) : \text{dist}(g, \mathcal{V}_1) \leq \varepsilon_1\}.$$

Then, a *globally* optimal estimation map takes the form

$$y = [y_0; y_1] \in \mathbb{R}^{m_0+m_1} \mapsto \sum_{i=1}^{m_0} a_{0,i}^{\text{opt}} y_{0,i} + \sum_{i=1}^{m_1} a_{1,i}^{\text{opt}} y_{1,i} \in \mathbb{R},$$

where the vector $a^{\text{opt}} = [a_0^{\text{opt}}; a_1^{\text{opt}}] \in \mathbb{R}^{m_0} \times \mathbb{R}^{m_1}$ is solution to the ℓ_1 -optimization program

$$\underset{[a_0; a_1] \in \mathbb{R}^{m_0+m_1}}{\text{minimize}} \quad [\varepsilon_0 \|a_0\|_1 + (\varepsilon_0 + \varepsilon_1) \|a_1\|_1] \quad \text{s.to } M^{(0)}[a_0; a_1] = b \text{ and } M^{(1)}a_1 = 0$$

for some explicit matrices $M^{(0)}, M^{(1)}$, and vector b . This result is in fact an instantiation of a more general result (Corollary 3). Indeed, in the case of estimation of a linear functional Q , the result can be extended to any number of model sets $\mathcal{K}_0, \mathcal{K}_1, \dots, \mathcal{K}_T$ —which do not need to be approximability sets either—and therefore apply to the four examples mentioned in the introduction. This extension can also be viewed as a first step towards a ‘dynamical Optimal Recovery’, where the object of interest is some time-varying f_t initially modeled by $f_0 \in \mathcal{K}_0$ and whose evolution is modeled by $f_{t-1} - f_t \in \mathcal{K}_t$, $t \geq 1$. As a stylized example, we are thinking of the estimation of average temperatures at a past time $t = 0$ given observations at this time $t = 0$ and at subsequent times $t \geq 1$. For this climate application, the use of Optimal Recovery in the ‘static’ framework was proposed in [29].

In the second scenario (Section 4), there is no restriction on the quantity of interest $Q : F \rightarrow Z$ (except that it is a linear map), but there are restrictions on the spaces F and Z , namely, they should be Hilbert spaces. Thus, when the objects f_0 and f_1 are functions observed via point evaluations, reproducing kernel Hilbert spaces provide the right framework. For arbitrary linear observation maps $\Lambda_0 : F \rightarrow \mathbb{R}^{m_0}$ and $\Lambda_1 : F \rightarrow \mathbb{R}^{m_1}$, suppose that the model sets are given by

$$\mathcal{K}_0 = \{g \in F : \|P_0 g\| \leq \varepsilon_0\} \quad \text{and} \quad \mathcal{K}_1 = \{g \in F : \|P_1 g\| \leq \varepsilon_1\},$$

where the linear operators P_0, P_1 map into Hilbert spaces, too. Then, a *globally* optimal recovery map is produced by constrained regularization with an explicitly determined parameter. Precisely, it is obtained as $Q \circ \Delta_{\tau^\sharp} : \mathbb{R}^{m_0+m_1} \rightarrow Z$, where Δ_τ is defined for $\tau \in [0, 1]$ by $\Delta_\tau([y_0; y_1]) = f_0^\tau \in F$ and

$$[f_0^\tau; f_1^\tau] = \underset{f=[f_0; f_1] \in F \times F}{\operatorname{argmin}} \left[(1-\tau)\|P_0 f_0\|^2 + \tau\|P_1(f_0 - f_1)\|^2 \right] \quad \text{s.to } \Lambda_0 f_0 = y_0, \Lambda_1 f_1 = y_1.$$

As for the parameter $\tau^\sharp \in [0, 1]$, it is selected as $\tau^\sharp = c_1^\sharp / (c_0^\sharp + c_1^\sharp)$, where $c_0^\sharp, c_1^\sharp \geq 0$ are solutions to the semidefinite program

$$\begin{aligned} \underset{c_0, c_1 \geq 0}{\operatorname{minimize}} \quad & c_0 \varepsilon_0^2 + c_1 \varepsilon_1^2 \\ \text{s.to} \quad & c_0 \|P_0 f_0\|^2 + c_1 \|P_1(f_0 - f_1)\|^2 \geq \|Q f_0\|^2 \quad \text{for all } f_0 \in \ker(\Lambda_0), f_1 \in \ker(\Lambda_1). \end{aligned}$$

This statement, exploiting a recent result from [30], does not extend to any number of model sets. This is because the underlying tool—Polyak’s S-procedure, see [31]—is similarly limited in the number of quadratic constraints it can handle. Beyond the multi-fidelity setting, situations that seem pertinent to this scenario involve destructive observations. As a stylized example, we can consider f_0 as a graph signal, i.e., a function defined on the finitely many vertices of a graph, whence it is relevant to take the operator P_0 describing \mathcal{K}_0 as the square-root of the graph Laplacian. We can also imagine that the very fact of observing f_0 is destructive, in the sense that it alters f_0 into a closeby f_1 , whence it is relevant to take the operator P_1 as the identity. In a non-destructive framework, the graph-signal application of Optimal Recovery was proposed in [32].

In the third scenario (Section 6), the setting is similar to the one from the second scenario, except that a *locally* optimal recovery map is targeted instead of a globally optimal one. In the spirit of [33], which extended the result from [34] to an arbitrary linear quantity of interest $Q \neq \text{Id}$, we first show that the minimal local worst-case error (aka local radius of information or Chebyshev radius) is upper bounded by the square root of the optimal value of the semidefinite program

$$(4) \quad \begin{aligned} & \underset{b, c_0, c_1 \geq 0}{\text{minimize}} && c_0(\varepsilon_0^2 - \|P_0 \Lambda_0^\dagger y_0\|^2) + c_1(\varepsilon_1^2 - \|P_1(\Lambda_0^\dagger y_0 - \Lambda_1^\dagger y_1)\|^2) + b \\ & \text{s.to} && \begin{bmatrix} c_0 M_0 + c_1 M_1 & & -c_1 M_1 \\ & & c_1 M_1 \end{bmatrix} \succeq \begin{bmatrix} N_0 & & 0 \\ & & 0 \end{bmatrix} \\ & \text{and} && \begin{bmatrix} c_0 M_0 + c_1 M_1 & & -c_1 M_1 & & c_0 L_0 y_0 + c_1 L_1 y \\ & & c_1 M_1 & & -c_1 L_1 y \\ (c_0 L_0 y_0 + c_1 L_1 y)^* & & -(c_1 L_1 y)^* & & b \end{bmatrix} \succeq 0, \end{aligned}$$

where L_0, L_1, M_0, M_1 are expressed in terms of $P_0, P_1, \Lambda_0, \Lambda_1$, while N_0 is expressed in terms of Q, Λ_0 . Next, we show that this upper bound can actually be equal to the genuine minimal local worst-case error under some additional assumptions, e.g. $\ker(\Lambda_0) \subseteq \ker(P_1)$ or $\ker(\Lambda_0) \subseteq \ker(\Lambda_1)$. The locally optimal recovery map is still produced by constrained regularization: it takes the form $Q \circ \Delta_{\tau^y}$, where the parameter $\tau^y \in [0, 1]$ now depends on y via $\tau^y = c_1^y / (c_0^y + c_1^y)$, with $b^y, c_0^y, c_1^y \geq 0$ being solutions to (4). Beyond the multi-fidelity setting, we can envisage a stylized example in which f_0 represents a physical entity and f_1 represents its digital twin. The observation functionals $\lambda_{1,i}$'s are chosen among the $\lambda_{0,i}$'s—hence ensuring that $\ker(\Lambda_0) \subseteq \ker(\Lambda_1)$ —for the purpose of confirming that f_1 is a convincing twin of f_0 .

The third scenario directly relies on more general results about Chebyshev radii and centers for a model set given by the intersection of two hyperellipsoids. We isolate these seemingly novel results in Section 5. Not only do they supply alternative—and arguably simpler (compare to [35])—arguments for the exact determination of other Chebyshev centers, but they also provide a *necessary and sufficient* orthogonality condition for a semidefinite-relaxation upper bound to agree with the true Chebyshev radius. We refer to Theorem 5 for the precise statement.

3 Globally Optimal Estimation of Linear Functionals

This section deals with the first scenario, where the quantity of interest Q is a linear functional defined on an arbitrary Banach space F . Exploiting the classical theory of Optimal Recovery, we can swiftly guarantee that, under reasonable assumptions, there is a globally optimal recovery map which is linear. Our main focus will be on the efficient construction of such a map. We shall do so in an extended framework where more than two model sets can be involved. Namely, the object f_0 comes with related objects f_1, \dots, f_T , which altogether are available through the model

assumptions

$$f_0 \in \mathcal{K}_0, \quad f_{t-1} - f_t \in \mathcal{K}_t, \quad t = 1, \dots, T,$$

as well as the observed data

$$y_0 = \Lambda_0 f_0 \in \mathbb{R}^{m_0}, \quad y_t = \Lambda_t f_t \in \mathbb{R}^{m_t}, \quad t = 1, \dots, T.$$

To simplify our analysis, it is useful to reduce this extended framework to a more familiar one by considering the compound vector

$$f = \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_T \end{bmatrix} \in F^{T+1}.$$

The model assumptions on this vector simply become $f \in \mathcal{K}$, where

$$\mathcal{K} = \left\{ \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_T \end{bmatrix} \in F^{T+1} : f_0 \in \mathcal{K}_0, f_{t-1} - f_t \in \mathcal{K}_t, t = 1, \dots, T \right\},$$

while the observed data takes the condensed form, with $m := m_0 + m_1 + \dots + m_T$,

$$y = \Lambda f \in \mathbb{R}^m, \quad \text{where } y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_T \end{bmatrix} \quad \text{and} \quad \Lambda = \begin{bmatrix} \Lambda_0 & 0 & \cdots & 0 \\ 0 & \Lambda_1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \Lambda_T \end{bmatrix}.$$

This notation allows us to write the worst-case errors of a recovery map $\Delta : \mathbb{R}^m \rightarrow \mathbb{R}$ in the simpler form

$$(5) \quad \text{gwce}(\Delta) = \sup_{f \in \mathcal{K}} |\tilde{Q}(f) - \Delta(\Lambda f)|,$$

where $\tilde{Q} : F^{T+1} \rightarrow \mathbb{R}$ is the linear functional defined by $\tilde{Q}([f_0; f_1; \dots; f_T]) = Q(f_0)$. This reduction facilitates the derivation of the following abstract statement, soon to be specialized in more tangible situations.

Theorem 1. Suppose that the model sets $\mathcal{K}_0, \mathcal{K}_1, \dots, \mathcal{K}_T$ are symmetric and convex. If the quantity of interest Q is a linear functional, then a globally optimal recovery map is given by

$$\Delta^{\text{opt}} : y = [y_0; y_1; \dots; y_T] \in \mathbb{R}^m \mapsto \sum_{i=1}^{m_0} a_{0,i}^{\text{opt}} y_{0,i} + \sum_{t=1}^T \sum_{i=1}^{m_t} a_{t,i}^{\text{opt}} y_{t,i} \in \mathbb{R},$$

where the vector $a^{\text{opt}} = [a_0^{\text{opt}}; a_1^{\text{opt}}; \dots; a_T^{\text{opt}}] \in \mathbb{R}^m$ is a solution to

$$(6) \quad \text{minimize}_{a=[a_0; a_1; \dots; a_T] \in \mathbb{R}^m} \sup_{g_0 \in \mathcal{K}_0} \left| \left(Q - \sum_{t=0}^T \sum_{i=1}^{m_t} a_{t,i} \lambda_{t,i} \right) (g_0) \right| + \sum_{s=1}^T \sup_{g_s \in \mathcal{K}_s} \left| \left(\sum_{t=s}^T \sum_{i=1}^{m_t} a_{t,i} \lambda_{t,i} \right) (g_s) \right|.$$

Proof. Given that $\mathcal{K}_0, \mathcal{K}_1, \dots, \mathcal{K}_T \subseteq F$ are all symmetric and convex, so is $\mathcal{K} \subseteq F^{T+1}$. Moreover, Q being a linear functional, \tilde{Q} is also a linear functional. Thus, according to the reformulation (5) of the global worst-case error, the foundational result of Smolyak (see e.g. [27, Theorem 4.7] or [28, Theorem 9.3]) ensures that, among the recovery maps minimizing $\text{gwce}(\Delta)$, there exists one which is linear, say of the form $\Delta_a = \langle a, \cdot \rangle$, i.e., $\Delta_a([y_0; y_1; \dots; y_T]) = \sum_{i=1}^{m_0} a_{0,i} y_{0,i} + \sum_{t=1}^T \sum_{i=1}^{m_t} a_{t,i} y_{t,i}$. To find a globally optimal recovery map, it is therefore enough to minimize $\text{gwce}(\Delta_a)$ over all $a = [a_0; a_1; \dots; a_T] \in \mathbb{R}^m$. Coming back to the original formulation of the global worst-case error and substituting the specific form of Δ_a , we obtain

$$\begin{aligned} \text{gwce}(\Delta_a) &= \sup_{\substack{f_0 \in \mathcal{K}_0 \\ f_{t-1} - f_t \in \mathcal{K}_t, t=1, \dots, T}} \left| Q(f_0) - \left(\sum_{i=1}^{m_0} a_{0,i} \lambda_{0,i}(f_0) + \sum_{t=1}^T \sum_{i=1}^{m_t} a_{t,i} \lambda_{t,i}(f_t) \right) \right| \\ &= \sup_{\substack{f_0 \in \mathcal{K}_0 \\ f_{t-1} - f_t \in \mathcal{K}_t, t=1, \dots, T}} \left| \left(Q - \sum_{t=0}^T \sum_{i=1}^{m_t} a_{t,i} \lambda_{t,i} \right) (f_0) + \left(\sum_{t=1}^T \sum_{i=1}^{m_t} a_{t,i} \lambda_{t,i} \right) (f_0 - f_t) \right|. \end{aligned}$$

Introducing temporarily $\mu := Q - \sum_{t=0}^T \sum_{i=1}^{m_t} a_{t,i} \lambda_{t,i}$ and $\nu_t := \sum_{i=1}^{m_t} a_{t,i} \lambda_{t,i}$ for $t = 1, \dots, T$, while writing $g_0 = f_0$ and $g_s = f_{s-1} - f_s$ for $s = 1, \dots, T$, the quantity inside the absolute values is

$$\mu(g_0) + \sum_{t=1}^T \nu_t \left(\sum_{s=1}^t g_s \right) = \mu(g_0) + \sum_{s=1}^T \left(\sum_{t=s}^T \nu_t \right) (g_s).$$

Therefore, we arrive at

$$\text{gwce}(\Delta_a) = \sup_{\substack{g_0 \in \mathcal{K}_0 \\ g_s \in \mathcal{K}_s, s=1, \dots, T}} \left| \mu(g_0) + \sum_{s=1}^T \left(\sum_{t=s}^T \nu_t \right) (g_s) \right| = \sup_{g_0 \in \mathcal{K}_0} |\mu(g_0)| + \sum_{s=1}^T \sup_{g_s \in \mathcal{K}_s} \left| \left(\sum_{t=s}^T \nu_t \right) (g_s) \right|,$$

where the last step relied on the symmetry of $\mathcal{K}_0, \mathcal{K}_1, \dots, \mathcal{K}_T$ to transform the absolute value of the sum into the sum of absolute values before decoupling the suprema. Taking the defining expression of μ, ν_1, \dots, ν_T into account finally leads to $\text{gwce}(\Delta_a)$ being equal to the objective function of (6). The announced result is consequently justified. \square

To make things less abstract, we now introduce some particular model sets $\mathcal{K}_0, \mathcal{K}_1, \dots, \mathcal{K}_T$ based on approximation capabilities by linear subspaces $\mathcal{V}_0, \mathcal{V}_1, \dots, \mathcal{V}_T$ of F with parameters $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_T > 0$. As such, let us consider

$$(7) \quad \mathcal{K}_t = \{g \in F : \text{dist}(g, \mathcal{V}_t) \leq \varepsilon_t\}, \quad t = 0, 1, \dots, T.$$

For instance, if $\mathcal{V}_t = \{0\}$ for $t = 1, \dots, T$, then the conditions $f_{t-1} - f_t \in \mathcal{K}_t$ simply mean that $\|f_{t-1} - f_t\| \leq \varepsilon_t$, i.e., that f_{t-1} and f_t are ε_t -close. Moreover, the condition $f_0 \in \mathcal{K}_0$ means that f_0 is ε_0 -close to being in the space \mathcal{V}_0 . Such an assumption is made implicitly in many numerical procedures, e.g. when devising quadrature formulas that are exact on a space of polynomials and whose accuracy depends on how close the function to integrate is from a polynomial. Note that, as

in the classical case where $\mathcal{K}_1, \dots, \mathcal{K}_T$ are absent, to avoid infinite worst-case errors, it is assumed from the outset that

$$\mathcal{V}_0 \cap \ker(\Lambda_0) = \{0\}, \quad \text{so that} \quad n_0 := \dim(\mathcal{V}_0) \leq m_0.$$

With the above approximability sets $\mathcal{K}_0, \mathcal{K}_1, \dots, \mathcal{K}_T$, Theorem 1 becomes the following statement.

Theorem 2. Suppose that $\mathcal{K}_0, \mathcal{K}_1, \dots, \mathcal{K}_T$ are the approximability sets defined in (7). If the quantity of interest Q is a linear functional, then a globally optimal recovery map is given by

$$\Delta^{\text{opt}} : y = [y_0; y_1; \dots; y_T] \in \mathbb{R}^m \mapsto \sum_{i=1}^{m_0} a_{0,i}^{\text{opt}} y_{0,i} + \sum_{t=1}^T \sum_{i=1}^{m_t} a_{t,i}^{\text{opt}} y_{t,i} \in \mathbb{R},$$

where the vector $a^{\text{opt}} = [a_0^{\text{opt}}; a_1^{\text{opt}}; \dots; a_T^{\text{opt}}] \in \mathbb{R}^m$ is a solution to

(8)

$$\begin{aligned} & \underset{a=[a_0; a_1; \dots; a_T] \in \mathbb{R}^m}{\text{minimize}} \left[\left\| Q - \sum_{t=0}^T \sum_{i=1}^{m_t} a_{t,i} \lambda_{t,i} \right\|_* \times \varepsilon_0 + \sum_{s=1}^T \left\| \sum_{t=s}^T \sum_{i=1}^{m_t} a_{t,i} \lambda_{t,i} \right\|_* \times \varepsilon_s \right] \\ (9) \quad & \text{s.to} \quad \left(\sum_{t=0}^T \sum_{i=1}^{m_t} a_{t,i} \lambda_{t,i} - Q \right)_{|\mathcal{V}_0} = 0 \quad \text{and} \quad \left(\sum_{t=s}^T \sum_{i=1}^{m_t} a_{t,i} \lambda_{t,i} \right)_{|\mathcal{V}_s} = 0 \quad \text{for all } s = 1, \dots, T. \end{aligned}$$

Proof. The quantity to minimize in (6) is made of several suprema, each supremum taking the form $S = \sup_{g \in F} \{|\eta(g)| : \text{dist}(g, \mathcal{V}) \leq \varepsilon\}$ for some linear functional η , linear subspace $\mathcal{V} \subseteq F$, and parameter $\varepsilon > 0$. Here, the constraint $\text{dist}(g, \mathcal{V}) \leq \varepsilon$ is equivalent to the existence of $v \in \mathcal{V}$ such that $h := g - v \in F$ satisfies $\|h\| \leq \varepsilon$. It follows that the supremum S can be written as

$$\begin{aligned} S &= \sup_{\substack{h \in F \\ v \in \mathcal{V}}} \{|\eta(h + v)| : \|h\| \leq \varepsilon\} = \sup_{h \in F} \{|\eta(h)| : \|h\| \leq \varepsilon\} + \sup_{v \in \mathcal{V}} \{|\eta(v)|\} \\ &= \begin{cases} +\infty & \text{in case } \eta(v) \neq 0 \text{ for some } v \in \mathcal{V}, \\ \|\eta\|_* \times \varepsilon & \text{in case } \eta(v) = 0 \text{ for all } v \in \mathcal{V}. \end{cases} \end{aligned}$$

Since we are interested in minimizing S , we must place ourselves in the latter case and impose the constraint $\eta|_{\mathcal{V}} = 0$. Substituting the detailed expression of η for each supremum yields the announced result. \square

Still, it is not apparent that Theorem 2 leads to an efficient construction of the vector a^{opt} because the dual norm on F may not be fitted for computations. For this to happen, we need to deal with spaces F on a case-to-case basis. For illustration, we shall take $F = C(\mathcal{X})$, the space of continuous functions on a compact set \mathcal{X} , equipped with the norm $\|g\| = \max\{|g(x)|, x \in \mathcal{X}\}$. As observation functionals, we select point evaluations, i.e., $\lambda_{t,i} = \delta_{x^{(t,i)}}$, at points $x^{(t,i)}$ that are all distinct. Thus,

if $Q(f_0) = f_0(x)$ for some $x \notin \{x^{(t,i)}, t = 0, 1, \dots, T, i = 1, \dots, m_t\}$ or $Q(f_0) = \text{vol}(\mathcal{X})^{-1} \int_{\mathcal{X}} f_0(x) dx$, we can easily see (using Tietze extension theorem) that

$$\left\| Q - \sum_{t=0}^T \sum_{i=1}^{m_t} a_{t,i} \lambda_{t,i} \right\|_* = 1 + \sum_{t=0}^T \sum_{i=1}^{m_t} |a_{t,i}|, \quad \left\| \sum_{t=s}^T \sum_{i=1}^{m_t} a_{t,i} \lambda_{t,i} \right\|_* = \sum_{t=s}^T \sum_{i=1}^{m_t} |a_{t,i}|, \quad s = 1, \dots, T.$$

With $(v_{t,1}, \dots, v_{t,m_t})$ denoting bases for \mathcal{V}_t , $t = 0, 1, \dots, T$, we introduce the vector $b \in \mathbb{R}^{n_0}$ and the matrices $M^{(s)} \in \mathbb{R}^{n_s \times (m_s + \dots + m_T)}$, $s = 0, 1, \dots, T$, with entries

$$b_j = v_{0,j}(x) \quad \text{and} \quad M_{j,(t,i)}^{(s)} = v_{s,j}(x^{(t,i)}).$$

With this notation in place, we can now state the following instantiation of Theorem 2, whose careful verification is left to the reader.

Corollary 3. Suppose that $\mathcal{K}_0, \mathcal{K}_1, \dots, \mathcal{K}_T \subseteq C(\mathcal{X})$ are the approximability sets defined in (7). If the observation functionals $\lambda_{t,i}$ are evaluations at points $x^{(t,i)} \in \mathcal{X}$ that are all distinct, then a globally optimal recovery map for the evaluation of f_0 at new $x \in \mathcal{X}$ is given by

$$\Delta^{\text{opt}} : y = [y_0; y_1; \dots; y_T] \in \mathbb{R}^m \mapsto \sum_{i=1}^{m_0} a_{0,i}^{\text{opt}} y_{0,i} + \sum_{t=1}^T \sum_{i=1}^{m_t} a_{t,i}^{\text{opt}} y_{t,i} \in \mathbb{R},$$

where $a^{\text{opt}} = [a_0^{\text{opt}}; a_1^{\text{opt}}; \dots; a_T^{\text{opt}}] \in \mathbb{R}^m$ is a solution to

$$\underset{[a_0; a_1; \dots; a_T] \in \mathbb{R}^m}{\text{minimize}} \sum_{t=0}^T (\varepsilon_0 + \dots + \varepsilon_t) \|a_t\|_1 \quad \text{s.to } M^{(0)}[a_0; \dots; a_T] = b, \quad M^{(s)}[a_s; \dots; a_T] = 0, \quad s = 1, \dots, T.$$

It is worth mentioning that the above optimization program—a weighted ℓ_1 -minimization—can be recast as a standard-form linear program. Furthermore, by freezing a_1, \dots, a_T to their optimal values, we obtain a standard-form linear program with n_0 equality constraints in the variable $a_0 \in \mathbb{R}^{m_0}$, so one can output (e.g. by the simplex algorithm) a solution $a_0^{\text{opt}} \in \mathbb{R}^{m_0}$ which is n_0 -sparse. Since the latter can be precomputed offline before the data $y \in \mathbb{R}^m$ are collected, it becomes unnecessary to collect the $y_{0,i}$'s for $i \notin \text{supp}(a_0^{\text{opt}})$. This is potentially significant because the high-fidelity functionals $\lambda_{0,i}$ are typically expensive to evaluate.

4 Globally Optimal Recovery in Hilbert Spaces

This section deals with the second scenario, where the quantity of interest is an arbitrary linear map $Q : F \rightarrow Z$ —it is not restricted to be a functional anymore—but F and Z are restricted to be Hilbert spaces. In close adherence to our multi-fidelity model, there are only two model sets \mathcal{K}_0 and \mathcal{K}_1 here. The latter are hyperellipsoids, i.e., they have the form

$$(10) \quad \mathcal{K}_0 = \{g \in F : \|P_0 g\| \leq \varepsilon_0\} \quad \text{and} \quad \mathcal{K}_1 = \{g \in F : \|P_1 g\| \leq \varepsilon_1\},$$

where the linear operators P_0 and P_1 also map into Hilbert spaces. Note that the approximability sets introduced in (7) are covered in this setting, as it suffices to choose $P_0 = P_{\mathcal{V}_0^\perp}$ and $P_1 = P_{\mathcal{V}_1^\perp}$, i.e., the orthogonal projectors onto the orthogonal complements of \mathcal{V}_0 and \mathcal{V}_1 , respectively. For the full recovery problem, i.e., when $Q = \text{Id}_F$, it is arguably natural to try and estimate the original f_0 by a constrained regularizer, namely by the element f_0^τ obtained via

$$[f_0^\tau; f_1^\tau] := \underset{[f_0; f_1] \in F \times F}{\text{argmin}} \left[(1 - \tau) \|P_0 f_0\|^2 + \tau \|P_1(f_0 - f_1)\|^2 \right] \quad \text{s.to} \quad \Lambda_0 f_0 = y_0, \Lambda_1 f_1 = y_1$$

for some parameter $\tau \in [0, 1]$. Rather than tuning this parameter by, say, cross-validation, we shall uncover a principled way of selecting the optimal τ . By doing so, we will also guarantee that the associated constrained regularization map is globally optimal, even among all possible recovery maps. The theorem stated below, which extends to the estimation of other quantities of interest besides $Q = \text{Id}_F$, follows from a recent result derived in [30, Section 2] on Optimal Recovery with a two-hyperellipsoid-intersection model set. The key is to reduce our multi-fidelity setting to this related setting, in the spirit of Section 3. Specifically, we reformulate the global worst-case error of any $\Delta : \mathbb{R}^{m_0+m_1} \rightarrow Z$ as

$$(11) \quad \text{gwce}(\Delta) := \sup_{\substack{\|P_0 f_0\| \leq \varepsilon_0 \\ \|P_1(f_0 - f_1)\| \leq \varepsilon_1}} \|Q(f_0) - \Delta([\Lambda_0 f_0; \Lambda_1 f_1])\|_Z = \sup_{\substack{\|\tilde{R}f\| \leq 1 \\ \|Sf\| \leq 1}} \|\tilde{Q}(f) - \Delta(\Lambda f)\|_Z,$$

where $f \in F \times F$ denotes again the compound vector

$$f = \begin{bmatrix} f_0 \\ f_1 \end{bmatrix},$$

and where the Hilbert-valued linear maps R , S , Λ , and \tilde{Q} are defined on $F \times F$ by

$$(12) \quad R(f) = \frac{1}{\varepsilon_0} P_0(f_0), \quad S(f) = \frac{1}{\varepsilon_1} P_1(f_0 - f_1), \quad \Lambda f = \begin{bmatrix} \Lambda_0 f_0 \\ \Lambda_1 f_1 \end{bmatrix}, \quad \tilde{Q}(f) = Q(f_0).$$

To avoid infinite worst-case errors, it is assumed from now on that

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

or, when $\dim(F) = \infty$, that there is $\delta > 0$ such that $\max\{\|Rh\|, \|Sh\|\} \geq \delta \|h\|$ for all $h \in \ker(\Lambda)$. In terms of our multi-fidelity framework, this reads

$$(13) \quad [P_0(f_0) = 0, P_1(f_0 - f_1) = 0, \Lambda_0(f_0) = 0, \Lambda_1(f_1) = 0] \implies [f_0 = 0, f_1 = 0].$$

In the cases $\ker(\Lambda_0) \subseteq \ker(P_1)$ and $\ker(\Lambda_1) \subseteq \ker(P_0)$ featured in Section 6, it can be verified that (13) holds if and only if both $\ker(P_0) \cap \ker(\Lambda_0) = \{0\}$ and $\ker(P_1) \cap \ker(\Lambda_1) = \{0\}$ hold.

Theorem 4. Suppose that F is a Hilbert space and that the model sets $\mathcal{K}_0, \mathcal{K}_1 \subseteq F$ are the hyperellipsoidal sets defined in (10). If the linear quantity of interest $Q : F \rightarrow Z$ maps into a

Hilbert space, then a globally optimal recovery map is given by $\Delta^{\text{opt}} : y \in \mathbb{R}^{m_0+m_1} \mapsto Q(f_0^\#) \in Z$, where

$$[f_0^\#; f_1^\#] = \underset{f=[f_0; f_1] \in F \times F}{\operatorname{argmin}} \left[(1 - \tau^\#) \|P_0 f_0\|^2 + \tau^\# \|P_1(f_0 - f_1)\|^2 \right] \quad \text{s.to} \quad \Lambda_0 f_0 = y_0, \Lambda_1 f_1 = y_1.$$

Here, the regularization parameter is $\tau^\# = c_1^\# / (c_0^\# + c_1^\#)$, with $c_0^\#, c_1^\# \geq 0$ being solutions to

$$(14) \quad \underset{c_0, c_1 \geq 0}{\operatorname{minimize}} [c_0 \varepsilon_0^2 + c_1 \varepsilon_1^2] \\ \text{s.to} \quad c_0 \|P_0 f_0\|^2 + c_1 \|P_1(f_0 - f_1)\|^2 \geq \|Q f_0\|^2 \quad \text{for all } f_0 \in \ker(\Lambda_0), f_1 \in \ker(\Lambda_1).$$

Proof. According to the reformulation (11) of the global worst-case error of $\Delta : \mathbb{R}^{m_0+m_1} \rightarrow Z$ and invoking [30, Theorem 1], we know that an optimal recovery map is given by $\Delta^{\text{opt}} = \tilde{Q} \circ \Delta_{a^\#, b^\#}$, where

$$\Delta_{a^\#, b^\#} : y \in \mathbb{R}^{m_0+m_1} \mapsto \left(\underset{f \in F \times F}{\operatorname{argmin}} a^\# \|Rf\|^2 + b^\# \|Sf\|^2 \quad \text{s.to} \quad \Lambda f = y \right) \in F \times F,$$

and where $a^\#, b^\# \geq 0$ are solutions to

$$\underset{a, b \geq 0}{\operatorname{minimize}} (a + b) \quad \text{s.to} \quad a \|Rf\|^2 + b \|Sf\|^2 \geq \|\tilde{Q}f\|^2 \quad \text{for all } f \in \ker \Lambda.$$

Making the change of optimization variables $c_0 = a/\varepsilon_0^2$ and $c_1 = b/\varepsilon_1^2$ while taking the expressions of R, S, Λ , and \tilde{Q} into account, we immediately see that the latter program reduces to (14). Likewise, we see that $\Delta_{a^\#, b^\#}(y) =: [f_0^\#; f_1^\#]$ reduces to the minimizer of $c_0^\# \|P_0 f_0\|^2 + c_1^\# \|P_1(f_0 - f_1)\|^2$ —or equivalently of $(1 - \tau^\#) \|P_0 f_0\|^2 + \tau^\# \|P_1(f_0 - f_1)\|^2$ —subject to $\Lambda_0 f_0 = y_0$ and $\Lambda_1 f_1 = y_1$, i.e., $[f_0^\#, f_1^\#] = [f_0^{\tau^\#}, f_1^{\tau^\#}]$, so that $\Delta^{\text{opt}}(y) = \tilde{Q}([f_0^{\tau^\#}, f_1^{\tau^\#}]) = Q(f_0^\#)$, as announced. \square

Remark. Here are a few comments to put the theoretical results of Theorem 4 into perspective.

a) Although not obvious at first sight, the optimal recovery map $\Delta^{\text{opt}} = \tilde{Q} \circ \Delta_{a^\#, b^\#}$ is linear. Indeed, the map $\Delta_{a^\#, b^\#}$ can be expressed (see [30, Subsection 2.1]) as

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

where $R_{\mathcal{N}}$ and $S_{\mathcal{N}}$ denote the restrictions of R and S to $\mathcal{N} = \ker(\Lambda)$.

b) Theorem 4 is valid for $\dim(F) = \infty$ (since the underpinning result from [30] is), but it is in finite dimensions that (14) becomes a manageable semidefinite program. Indeed, if $(h_1^0, \dots, h_{k_0}^0)$ and $(h_1^1, \dots, h_{k_1}^1)$ are bases for $\ker(\Lambda_0)$ and $\ker(\Lambda_1)$ and if $H_0 : \mathbb{R}^{k_0} \rightarrow \ker(\Lambda_0)$ and $H_1 : \mathbb{R}^{k_1} \rightarrow \ker(\Lambda_1)$ represent the transformations defined by $H_0(u) = \sum_{i=1}^{k_0} u_i h_i^0$ and $H_1(v) = \sum_{i=1}^{k_1} v_i h_i^1$, then we can reformulate the constraint $c_0 \|P_0 f_0\|^2 + c_1 \|P_1(f_0 - f_1)\|^2 \geq \|Q f_0\|^2$ for all $f_0 \in \ker(\Lambda_0)$ and $f_1 \in \ker(\Lambda_1)$ as

$$\left[\begin{array}{c|c} H_0^*(c_0 P_0^* P_0 + c_1 P_1^* P_1 - Q^* Q) H_0 & H_0^*(c_1 P_1^* P_1) H_1 \\ \hline H_1^*(c_1 P_1^* P_1) H_0 & H_1^*(c_1 P_1^* P_1) H_1 \end{array} \right] \succeq 0.$$

Note that the square matrix above has size $k_0 + k_1 \geq 2 \dim(F) - m_0 - m_1$, independently of the dimension of Z . Intuitively, if Z is low dimensional, e.g. $Z = \mathbb{R}$ when Q is a linear functional, then we should be able to lower the computational cost of producing an optimal recovery map, but this does not follow from Theorem 4.

- c) Theorem 4 substantially relies on a result from [30], itself relying on the so-called S-procedure, and in particular on its ‘no-linear-term’ version. Loosely speaking, given quadratic functions q_0, q_1, \dots, q_K on \mathbb{K}^n , the S-procedure—which is surveyed in [36]—relates the condition $[q_0(x) \leq 0$ whenever $q_1(x) \leq 0, \dots, q_K(x) \leq 0]$ to the condition of existence of $c_1, \dots, c_K \geq 0$ such that $q_0 \leq c_1 q_1 + \dots + c_K q_K$. Clearly, the latter condition always implies the former condition. The S-procedure is said to be exact if, conversely, the former implies the latter. Exactness occurs, under some strict feasibility conditions, when $K = 1$ (this is Yakubovich’s S-Lemma) and when $K = 2$ if $\mathbb{K} = \mathbb{C}$ (see [37]). It also occurs when $K = 2$ if $\mathbb{K} = \mathbb{R}$ provided the quadratic functions are all of the form $q_k(x) = \langle A_k x, x \rangle + \alpha_k$, i.e., they do not feature any linear term $\langle a_k, x \rangle$. Established by Polyak in [31], this is what we called the ‘no-linear-term’ S-procedure. Not only is this version generally invalid for $K > 2$, but the result from [30] is invalid, too, when more than two hyperellipsoidal model sets are involved. Consequently, unless a more restrictive setting is enforced, we cannot extend Theorem 4 to the situation considered in Section 3, where we dealt with model sets $\mathcal{K}_0, \mathcal{K}_1, \dots, \mathcal{K}_T$ with $T > 1$.

5 Chebyshev Center for the Intersection of Two Hyperellipsoids

Still within Hilbert spaces, our next goal is to tackle the more subtle problem of local optimality. As was just made apparent in the global setting, our multi-fidelity model is tied to the two-hyperellipsoid-intersection model, so we start with some investigations about this general model, bearing in mind that the local setting was not considered in [30]. Readers mainly interested in the multi-fidelity aspect can skip to Section 6 and bypass the technicalities of the present section. Here, we deal with Hilbert-valued linear maps R, S, Λ , and \tilde{Q} which are defined on a common Hilbert space F and which do not have to be taken as in (12). Fixing $y \in \text{ran}(\Lambda)$, we aim at minimizing over all z the local worst-case error

$$\text{lwce}_y(z) = \sup_{\substack{\|Rf\| \leq 1, \|Sf\| \leq 1 \\ \Lambda f = y}} \|\tilde{Q}(f) - z\|.$$

In other words, we are looking for a Chebyshev center of the linear image of a sliced intersection of two hyperellipsoids, namely for a center z^{opt} of a ball with smallest radius that contains the set $\tilde{Q}(\{f \in F : \|Rf\| \leq 1, \|Sf\| \leq 1, \Lambda f = y\})$. In this situation, classical results guarantee the existence and uniqueness of such a Chebyshev center. However, determining the Chebyshev center/radius is an arduous task in general. In the closely related problem of Optimal Recovery from inaccurate data, a semidefinite-programming-based upper bound for the Chebyshev radius was

proposed in [34] when $Q = \text{Id}_F$, along with a candidate Chebyshev center. The genuine Chebyshev center was obtained more directly in [35] but under a specific assumption and still when $Q = \text{Id}_F$. It was later observed that the aforementioned candidate and genuine Chebyshev centers actually coincide, see [33]. There, the upper bound was also generalized to linear quantities of interest $Q \neq \text{Id}_F$. All these findings can now be retrieved from the theorem below. It consists of three parts: the first part provides a computable upper bound for the Chebyshev radius (incidentally involving constrained regularizers); the second part uncovers an orthogonality condition under which this upper bound agrees with the true Chebyshev radius; the third part shows conversely that the upper bound and the true Chebyshev radius can only be equal under this orthogonality condition. The precise statement of the theorem involves the following notation to be used throughout this section: given $y \in \text{ran}(\Lambda)$ and $\tau \in [0, 1]$,

- $f^\tau \in F$ is the constrained minimizer defined as

$$f^\tau := \underset{f \in F}{\text{argmin}} \left[(1 - \tau) \|Rf\|^2 + \tau \|Sf\|^2 \right] \quad \text{s.to} \quad \Lambda f = y;$$

- $\lambda(\tau)$ denotes the largest eigenvalue of $[(1 - \tau)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau S_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1}\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}$, $\mathcal{N} := \ker(\Lambda)$, i.e.,

$$\lambda(\tau) := \lambda_{\max}([(1 - \tau)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau S_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1}\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}});$$

- $h^\tau \in \mathcal{N}$ is defined implicitly via an eigenequation and a normalization, namely

$$\begin{aligned} \tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}h^\tau &= \lambda(\tau)[(1 - \tau)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau S_{\mathcal{N}}^*S_{\mathcal{N}}]h^\tau, \\ \|\tilde{Q}_{\mathcal{N}}h^\tau\|^2 &= \lambda(\tau) \times (1 - ((1 - \tau)\|Rf^\tau\|^2 + \tau\|Sf^\tau\|^2)). \end{aligned}$$

The three-part theorem, whose upcoming proof is technically involved, reads as follows.

Theorem 5. Let R, S, \tilde{Q} be Hilbert-valued linear maps defined on a common Hilbert space F . Let also $\Lambda : F \rightarrow \mathbb{R}^m$ be another linear map and let $y \in \mathbb{R}^m$.

(i) The squared Chebyshev radius ρ^2 of the set $\tilde{Q}(\{f \in F : \|Rf\| \leq 1, \|Sf\| \leq 1, \Lambda f = y\})$ is upper-bounded as

$$\rho^2 \leq \inf_{\tau \in [0,1]} \sup_{\substack{\|Rf\| \leq 1, \|Sf\| \leq 1 \\ \Lambda f = y}} \|\tilde{Q}(f) - \tilde{Q}(f^\tau)\|^2 \leq \inf_{\tau \in [0,1]} \|\tilde{Q}h^\tau\|^2.$$

(ii) For the minimizer τ^y of $\|\tilde{Q}h^\tau\|^2$ over $\tau \in [0, 1]$, if the orthogonality condition

$$(15) \quad \langle Rf^{\tau^y}, Rh^{\tau^y} \rangle = 0 \quad \text{and/or} \quad \langle Sf^{\tau^y}, Sh^{\tau^y} \rangle = 0$$

holds, then the squared Chebyshev radius agrees with the upper bound, i.e.,

$$(16) \quad \rho^2 = \|\tilde{Q}h^{\tau^y}\|^2 = \inf_{\tau \in [0,1]} \|\tilde{Q}h^\tau\|^2.$$

(iii) Conversely, if the equality (16) between squared Chebyshev radius and upper bound occurs, then the orthogonality condition (15) holds.

Each part of the theorem will be proved separately. For the first part, we make use of the following lemma, which provides a fresh look at the one-hyperellipsoid model. The result was obtained in [38] when $T = P_{\mathcal{V}^\perp}$ is the orthogonal projector onto the orthogonal complement of a linear space $\mathcal{V} \subseteq F$ and when $Q = \text{Id}_F$, in which case the largest eigenvalue in the statement below reduces to $1/\sigma_{\min}(T_{\mathcal{N}})^2$. When $T = P_{\mathcal{V}^\perp}$ but $Q \neq \text{Id}_F$, it can also be found as [28, Theorem 10.2], albeit with the largest eigenvalue expressed in a different manner.

Lemma 6. Let T, \tilde{Q} be Hilbert-valued linear maps defined on a common Hilbert space F . Let also $\Lambda : F \rightarrow \mathbb{R}^m$ be another linear map whose null space $\mathcal{N} = \ker(\Lambda)$ satisfies $\ker(T) \cap \mathcal{N} = \{0\}$. Given $y \in \mathbb{R}^m$, the Chebyshev center of the set $\tilde{Q}(\{f \in F : \|Tf\| \leq 1, \Lambda f = y\})$ is $\tilde{Q}(\hat{f})$, where

$$\hat{f} := \underset{f \in F}{\text{argmin}} \|Tf\|^2 \quad \text{s.to} \quad \Lambda f = y$$

and its squared Chebyshev radius is equal to

$$(17) \quad \|\tilde{Q}\hat{h}\|^2 = \lambda_{\max}([T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}) \times (1 - \|T\hat{f}\|^2),$$

where $\hat{h} \in \mathcal{N}$ is a leading eigenvector of $[T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}$ normalized so that $\|T\hat{h}\|^2 = 1 - \|T\hat{f}\|^2$.

Proof. Let us first quickly justify the equality in (17), which relies on the properties of \hat{h} via

$$\begin{aligned} \|\tilde{Q}\hat{h}\|^2 &= \langle \tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}\hat{h}, \hat{h} \rangle = \langle [T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}\hat{h}, [T_{\mathcal{N}}^*T_{\mathcal{N}}]\hat{h} \rangle = \langle \lambda_{\max}([T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}})\hat{h}, [T_{\mathcal{N}}^*T_{\mathcal{N}}]\hat{h} \rangle \\ &= \lambda_{\max}([T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}) \times \|T\hat{h}\|^2 = \lambda_{\max}([T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}) \times (1 - \|T\hat{f}\|^2). \end{aligned}$$

The equality in (17) being put aside, our remaining objective is to prove that, for any $z \in \text{ran}(\tilde{Q})$,

$$(18) \quad \sup_{\substack{\|Tf\|^2 \leq 1 \\ \Lambda f = y}} \|\tilde{Q}f - z\|^2 \geq \sup_{\substack{\|Tf\|^2 \leq 1 \\ \Lambda f = y}} \|\tilde{Q}f - \tilde{Q}\hat{f}\|^2 = \|\tilde{Q}\hat{h}\|^2.$$

We start with the equality on the right of (18). To this end, it is crucial to observe that \hat{f} is characterized by $\Lambda\hat{f} = y$ and $\langle T\hat{f}, Th \rangle = 0$ for all $h \in \mathcal{N}$, the latter property stemming from $\|T\hat{f}\|^2 \leq \|T(\hat{f} + h)\|^2$ for all $h \in \mathcal{N}$. Thus, reparametrizing any $f \in F$ such that $\Lambda f = y$ as $f = \hat{f} + h$ with $h \in \mathcal{N}$, we see that the constraint $\|Tf\|^2 \leq 1$, i.e., $\|T(\hat{f} + h)\|^2 \leq 1$, is equivalent to $\|T\hat{f}\|^2 + \|Th\|^2 \leq 1$. Therefore,

$$\sup_{\substack{\|Tf\|^2 \leq 1 \\ \Lambda f = y}} \|\tilde{Q}f - \tilde{Q}\hat{f}\|^2 = \sup_{\substack{\|Th\|^2 \leq 1 - \|T\hat{f}\|^2 \\ h \in \mathcal{N}}} \|\tilde{Q}h\|^2.$$

At this point, we claim that the latter supremum is equal to $\|\tilde{Q}\hat{h}\|^2$. Indeed, for any $h \in \mathcal{N}$ satisfying $\|Th\|^2 \leq 1 - \|T\hat{f}\|^2$, by introducing $u := [T_{\mathcal{N}}^*T_{\mathcal{N}}]^{1/2}h \in \mathcal{N}$ and using the shorthand $\lambda_{\max} := \lambda_{\max}([T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}) = \lambda_{\max}([T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2}\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}[T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2})$, we remark that

$$\begin{aligned} \|\tilde{Q}h\|^2 &= \langle \tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}h, h \rangle = \langle [T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2}\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}[T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2}u, u \rangle \\ &\leq \lambda_{\max} \times \|u\|^2 = \lambda_{\max} \times \|Th\|^2 \leq \lambda_{\max} \times (1 - \|T\hat{f}\|^2), \end{aligned}$$

with equality occurring throughout if and only if $[T_{\mathcal{N}}^* T_{\mathcal{N}}]^{-1/2} \tilde{Q}_{\mathcal{N}}^* \tilde{Q}_{\mathcal{N}} [T_{\mathcal{N}}^* T_{\mathcal{N}}]^{-1/2} u = \lambda_{\max} u$, i.e., $\tilde{Q}_{\mathcal{N}}^* \tilde{Q}_{\mathcal{N}} h = \lambda_{\max} [T_{\mathcal{N}}^* T_{\mathcal{N}}] h$, and $\|Th\|^2 = 1 - \|T\hat{f}\|^2$ —in short, if and only if h equals (one of the) \hat{h} .

Continuing with the inequality on the left of (18), let us consider any $z \in \text{ran}(\tilde{Q})$. By noticing that $\|T(\hat{f} \pm \hat{h})\|^2 = 1$ and that $\Lambda(\hat{f} \pm \hat{h}) = y$, we obtain

$$\begin{aligned} \sup_{\substack{\|Tf\|^2 \leq 1 \\ \Lambda f = y}} \|\tilde{Q}f - z\|^2 &\geq \max_{\pm} \|\tilde{Q}(\hat{f} \pm \hat{h}) - z\|^2 \geq \frac{1}{2} \|\tilde{Q}\hat{f} - z + \tilde{Q}\hat{h}\|^2 + \frac{1}{2} \|\tilde{Q}\hat{f} - z - \tilde{Q}\hat{h}\|^2 \\ &= \|\tilde{Q}\hat{f} - z\|^2 + \|\tilde{Q}\hat{h}\|^2 \geq \|\tilde{Q}\hat{h}\|^2, \end{aligned}$$

which is the desired inequality. The proof is now complete. \square

Proof of Theorem 5, Part (i). For a fixed $\tau \in [0, 1]$, the squared Chebyshev radius ρ^2 , i.e., the squared minimal local worst-case error, is evidently upper-bounded by the squared local worst-case error evaluated at $\tilde{Q}(f^\tau)$. We bound the latter as

$$\sup_{\substack{\|Rf\|^2 \leq 1, \|Sf\|^2 \leq 1 \\ \Lambda f = y}} \|\tilde{Q}(f) - \tilde{Q}(f^\tau)\|^2 \leq \sup_{\substack{(1-\tau)\|Rf\|^2 + \tau\|Sf\|^2 \leq 1 \\ \Lambda f = y}} \|\tilde{Q}(f) - \tilde{Q}(f^\tau)\|^2.$$

If we set $T := [(1 - \tau)R^*R + \tau S^*S]^{1/2}$, so that $\|Tf\|^2 = (1 - \tau)\|Rf\|^2 + \tau\|Sf\|^2$ for all $f \in F$, we can invoke Lemma 6—and its proof, see (18) in particular—to state that the latter supremum is equal to $\|\tilde{Q}\hat{h}\|^2$, where \hat{h} is easily identified in our context with h^τ . We have therefore obtained

$$\rho^2 \leq \sup_{\substack{\|Rf\|^2 \leq 1, \|Sf\|^2 \leq 1 \\ \Lambda f = y}} \|\tilde{Q}(f) - \tilde{Q}(f^\tau)\|^2 \leq \|\tilde{Q}h^\tau\|^2.$$

The required result simply follows by taking the infimum over $\tau \in [0, 1]$. \square

We now turn to the second part of the theorem, for which it is worth isolating another lemma. After uncovering the derivatives of some τ -dependent quantities, the lemma brings forward a crucial piece of information about the minimizer τ^y of $\|Qh^\tau\|^2$.

Lemma 7. The derivatives of the functions G and H defined below are

$$\begin{aligned} \text{if } G(\tau) &:= (1 - \tau)\|Rf^\tau\|^2 + \tau\|Sf^\tau\|^2, & \text{then } G'(\tau) &= \|Sf^\tau\|^2 - \|Rf^\tau\|^2, \\ \text{if } H(\tau) &:= 1/\lambda(\tau), & \text{then } H'(\tau) &= \frac{\|Sh^\tau\|^2 - \|Rh^\tau\|^2}{\|\tilde{Q}h^\tau\|^2}. \end{aligned}$$

As a consequence, the minimizer $\tau^y \in (0, 1)$ of $\|Qh^\tau\|^2$ satisfies

$$\|Rf^{\tau^y}\|^2 + \|Rh^{\tau^y}\|^2 = \|Sf^{\tau^y}\|^2 + \|Sh^{\tau^y}\|^2 = 1.$$

Proof. For $\sigma, \tau \in [0, 1]$, the minimality property of f^σ yields

$$\begin{aligned} (1 - \sigma)\|Rf^\sigma\|^2 + \sigma\|Sf^\sigma\|^2 &\leq (1 - \sigma)\|Rf^\tau\|^2 + \sigma\|Sf^\tau\|^2 \\ &= (1 - \tau)\|Rf^\tau\|^2 + \tau\|Sf^\tau\|^2 + (\sigma - \tau)(\|Sf^\tau\|^2 - \|Rf^\tau\|^2). \end{aligned}$$

In terms of the function G , the latter reads $G(\sigma) \leq G(\tau) + (\sigma - \tau)(\|Sf^\tau\|^2 - \|Rf^\tau\|^2)$. Likewise, by exchanging the roles of τ and σ , we also have $G(\tau) \leq G(\sigma) + (\tau - \sigma)(\|Sf^\sigma\|^2 - \|Rf^\sigma\|^2)$. We combine these two inequalities in the form

$$(\sigma - \tau)(\|Sf^\sigma\|^2 - \|Rf^\sigma\|^2) \leq G(\sigma) - G(\tau) \leq (\sigma - \tau)(\|Sf^\tau\|^2 - \|Rf^\tau\|^2).$$

This immediately implies that the function G is differentiable with

$$G'(\tau) = \|Sf^\tau\|^2 - \|Rf^\tau\|^2.$$

Next, the eigenequation defining h^τ can be written as $[(1 - \tau)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau S_{\mathcal{N}}^*S_{\mathcal{N}}]h^\tau = H(\tau)\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}h^\tau$, and differentiating it with respect to τ , we obtain

$$(-R_{\mathcal{N}}^*R_{\mathcal{N}} + S_{\mathcal{N}}^*S_{\mathcal{N}})h^\tau + [(1 - \tau)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau S_{\mathcal{N}}^*S_{\mathcal{N}}]\frac{dh^\tau}{d\tau} = H'(\tau)\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}h^\tau + H(\tau)\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}\frac{dh^\tau}{d\tau}.$$

Taking the inner product with h^τ , while noticing that

$$\begin{aligned} \left\langle [(1 - \tau)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau S_{\mathcal{N}}^*S_{\mathcal{N}}]\frac{dh^\tau}{d\tau}, h^\tau \right\rangle &= \left\langle \frac{dh^\tau}{d\tau}, [(1 - \tau)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau S_{\mathcal{N}}^*S_{\mathcal{N}}]h^\tau \right\rangle \\ &= \left\langle \frac{dh^\tau}{d\tau}, H(\tau)\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}h^\tau \right\rangle = H(\tau) \left\langle \tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}\frac{dh^\tau}{d\tau}, h^\tau \right\rangle, \end{aligned}$$

we arrive at the identity $\langle (-R_{\mathcal{N}}^*R_{\mathcal{N}} + S_{\mathcal{N}}^*S_{\mathcal{N}})h^\tau, h^\tau \rangle = H'(\tau)\langle \tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}h^\tau, h^\tau \rangle$, which also reads $-\|Rh^\tau\|^2 + \|Sh^\tau\|^2 = H'(\tau)\|\tilde{Q}h^\tau\|^2$. This yields the announced expression for $H'(\tau)$.

Finally, from the defining property of h^τ , we point out that $\|\tilde{Q}h^\tau\|^2 = (1 - G(\tau))/H(\tau)$ and remark that the derivative of H can also be expressed as $H'(\tau) = (\|Sh^\tau\|^2 - \|Rh^\tau\|^2)H(\tau)/(1 - G(\tau))$. It follows that, at the minimizer $\tau^y \in (0, 1)$ of $\|\tilde{Q}h^\tau\|$, we have

$$\begin{aligned} (19) \quad 0 &= -G'(\tau^y)H(\tau^y) - (1 - G(\tau^y))H'(\tau^y) \\ &= (\|Rf^{\tau^y}\|^2 - \|Sf^{\tau^y}\|^2 + \|Rh^{\tau^y}\|^2 - \|Sh^{\tau^y}\|^2) \times H(\tau^y). \end{aligned}$$

Simplifying by $H(\tau^y) \neq 0$ and rearranging yields $\|Rf^{\tau^y}\|^2 + \|Rh^{\tau^y}\|^2 = \|Sf^{\tau^y}\|^2 + \|Sh^{\tau^y}\|^2$. With κ denoting this common value, we see that

$$\begin{aligned} \kappa &= (1 - \tau^y)(\|Rf^{\tau^y}\|^2 + \|Rh^{\tau^y}\|^2) + \tau^y(\|Sf^{\tau^y}\|^2 + \|Sh^{\tau^y}\|^2) \\ &= ((1 - \tau^y)\|Rf^{\tau^y}\|^2 + \tau^y\|Sf^{\tau^y}\|^2) + ((1 - \tau^y)\|Rh^{\tau^y}\|^2 + \tau^y\|Sh^{\tau^y}\|^2). \end{aligned}$$

Taking into account that

$$\begin{aligned} (1 - \tau^y)\|Rh^{\tau^y}\|^2 + \tau^y\|Sh^{\tau^y}\|^2 &= \langle [(1 - \tau^y)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau^y S_{\mathcal{N}}^*S_{\mathcal{N}}]h^{\tau^y}, h^{\tau^y} \rangle = \langle (1/\lambda(\tau^y))\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}h^{\tau^y}, h^{\tau^y} \rangle \\ &= (1/\lambda(\tau^y))\|\tilde{Q}h^{\tau^y}\|^2 = 1 - ((1 - \tau^y)\|Rf^{\tau^y}\|^2 + \tau^y\|Sf^{\tau^y}\|^2), \end{aligned}$$

we derive that $\kappa = 1$. The desired result is now proved. \square

Proof of Theorem 5, Part (ii). Together with $\|Rf^{\tau^y}\|^2 + \|Rh^{\tau^y}\|^2 = \|Sf^{\tau^y}\|^2 + \|Sh^{\tau^y}\|^2 = 1$, the additional orthogonality conditions $\langle Rf^{\tau^y}, Rh^{\tau^y} \rangle = 0$ and $\langle Sf^{\tau^y}, Sh^{\tau^y} \rangle = 0$ —note that one follows from the other according to the characterization $((1 - \tau^y)R^*R + \tau^y S^*S)f^{\tau^y} \perp \ker \Lambda$ —imply

$$\|R(f^{\tau^y} \pm h^{\tau^y})\|^2 = 1 \quad \text{and} \quad \|S(f^{\tau^y} \pm h^{\tau^y})\|^2 = 1.$$

From here, it follows that, for any $z \in \text{ran}(\tilde{Q})$,

$$\begin{aligned} \text{lwce}_y(z)^2 &= \sup_{\substack{\|Rf\| \leq 1, \|Sf\| \leq 1 \\ \Lambda f = y}} \|\tilde{Q}(f) - z\|^2 \geq \max_{\pm} \|\tilde{Q}(f^{\tau^y} \pm h^{\tau^y}) - z\|^2 \\ &\geq \frac{1}{2} \|\tilde{Q}f^{\tau^y} - z + \tilde{Q}h^{\tau^y}\|^2 + \frac{1}{2} \|\tilde{Q}f^{\tau^y} - z - \tilde{Q}h^{\tau^y}\|^2 = \|\tilde{Q}f^{\tau^y} - z\|^2 + \|\tilde{Q}h^{\tau^y}\|^2 \\ &\geq \|\tilde{Q}h^{\tau^y}\|^2. \end{aligned}$$

This lower bound is the same as the upper bound from Part (i), so that Part (ii) is now proved. \square

Proof of Theorem 5, Part (iii). Let us now assume that the squared Chebyshev radius coincides with the upper bound from Part (i), i.e., that the minimum squared local worst-case error equals $\|\tilde{Q}h^{\tau^y}\|^2$. In this case, we have

$$\|\tilde{Q}h^{\tau^y}\|^2 \leq \text{lwce}_y(\tilde{Q}f^{\tau^y}) = \sup_{\substack{\|Rf\| \leq 1, \|Sf\| \leq 1 \\ \Lambda f = y}} \|\tilde{Q}f - \tilde{Q}f^{\tau^y}\|^2 = \sup_{\substack{\|R(f^{\tau^y} + h)\| \leq 1 \\ \|S(f^{\tau^y} + h)\| \leq 1 \\ h \in \mathcal{N}}} \|\tilde{Q}h\|^2.$$

If the latter supremum is attained at some $\bar{h} \in \mathcal{N}$, we have $\|\tilde{Q}h^{\tau^y}\|^2 \leq \|\tilde{Q}\bar{h}\|^2$. Using the shorthand $T := [(1 - \tau^y)R^*R + \tau^y S^*S]^{1/2}$, the following chain of inequalities makes use of $T^*Tf^{\tau^y} \perp \mathcal{N}$, of $\|R(f^{\tau^y} + \bar{h})\| \leq 1$, $\|S(f^{\tau^y} + \bar{h})\| \leq 1$, and of the normalization of h^{τ^y} :

$$\begin{aligned} \|T\bar{h}\|^2 &= \langle T^*T\bar{h}, \bar{h} \rangle = \langle T^*T(f^{\tau^y} + \bar{h}), (f^{\tau^y} + \bar{h}) \rangle - \langle T^*Tf^{\tau^y}, f^{\tau^y} \rangle \\ &= \langle [(1 - \tau^y)R^*R + \tau^y S^*S](f^{\tau^y} + \bar{h}), (f^{\tau^y} + \bar{h}) \rangle - \langle [(1 - \tau^y)R^*R + \tau^y S^*S]f^{\tau^y}, f^{\tau^y} \rangle \\ &= ((1 - \tau^y)\|R(f^{\tau^y} + \bar{h})\|^2 + \tau^y\|S(f^{\tau^y} + \bar{h})\|^2) - ((1 - \tau^y)\|Rf^{\tau^y}\|^2 + \tau^y\|Sf^{\tau^y}\|^2) \\ &\leq ((1 - \tau^y) \times 1 + \tau^y \times 1) - ((1 - \tau^y)\|Rf^{\tau^y}\|^2 + \tau^y\|Sf^{\tau^y}\|^2) \\ &= 1 - ((1 - \tau^y)\|Rf^{\tau^y}\|^2 + \tau^y\|Sf^{\tau^y}\|^2) = \frac{1}{\lambda(\tau^y)} \|\tilde{Q}h^{\tau^y}\|^2 \\ &\leq \frac{1}{\lambda(\tau^y)} \|\tilde{Q}\bar{h}\|^2 = \frac{1}{\lambda(\tau^y)} \langle [T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2} \tilde{Q}_{\mathcal{N}}^* \tilde{Q}_{\mathcal{N}} [T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2} [T_{\mathcal{N}}^*T_{\mathcal{N}}]^{1/2} \bar{h}, [T_{\mathcal{N}}^*T_{\mathcal{N}}]^{1/2} \bar{h} \rangle \\ &\leq \frac{1}{\lambda(\tau^y)} \lambda_{\max}([T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2} \tilde{Q}_{\mathcal{N}}^* \tilde{Q}_{\mathcal{N}} [T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2}) \|[T_{\mathcal{N}}^*T_{\mathcal{N}}]^{1/2} \bar{h}\|^2 \\ &= \|T_{\mathcal{N}}\bar{h}\|^2. \end{aligned}$$

Since the leftmost term and the rightmost term are the same, equality must hold all the way through, implying that $\|R(f^{\tau^y} + \bar{h})\|^2 = 1$, $\|S(f^{\tau^y} + \bar{h})\|^2 = 1$, and that $[T_{\mathcal{N}}^*T_{\mathcal{N}}]^{1/2}\bar{h}$ is a leading eigenvalue of $[T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2} \tilde{Q}_{\mathcal{N}}^* \tilde{Q}_{\mathcal{N}} [T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2}$, hence $\bar{h} = h^{\tau}$. We thus arrived at $\|R(f^{\tau^y} + h^{\tau^y})\|^2 = 1$. Recalling from Lemma 7 that $\|Rf^{\tau^y}\|^2 + \|Rh^{\tau^y}\|^2 = 1$, we conclude that $\langle Rf^{\tau^y}, Rh^{\tau^y} \rangle = 0$, as desired. \square

Remark. Although Theorem 5 may appear like an abstract statement distant from practical considerations, our candidate locally optimal recovery map, as well as a locally optimal one, can actually be efficiently computed, as revealed in the comments below.

a) The upper bound $\|Qh^{\tau^y}\|^2$ from Theorem 5, Part (i), is the optimal value of the semidefinite program

$$(20) \quad \begin{aligned} & \underset{b, c, d \geq 0}{\text{minimize}} \quad c(1 - \|R\Lambda^\dagger y\|^2) + d(1 - \|S\Lambda^\dagger y\|^2) + b \\ & \text{s.to} \quad cR_{\mathcal{N}}^*R_{\mathcal{N}} + dS_{\mathcal{N}}^*S_{\mathcal{N}} \succeq \tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}} \\ & \text{and} \quad \left[\begin{array}{c|c} cR_{\mathcal{N}}^*R_{\mathcal{N}} + dS_{\mathcal{N}}^*S_{\mathcal{N}} & cR_{\mathcal{N}}^*R\Lambda^\dagger y + dS_{\mathcal{N}}^*S\Lambda^\dagger y \\ \hline (cR_{\mathcal{N}}^*R\Lambda^\dagger y + dS_{\mathcal{N}}^*S\Lambda^\dagger y)^* & b \end{array} \right] \succeq 0. \end{aligned}$$

Moreover, if $b^y, c^y, d^y \geq 0$ are minimizers, then the optimal parameter τ^y is given by $d^y/(c^y + d^y)$.

To justify this claim, we first look at the one-hyperellipsoid situation. By Lemma 6, the squared Chebyshev radius is $\rho^2 = \lambda_{\max}(1 - \|T\hat{f}\|^2)$, where $\lambda_{\max} := \lambda_{\max}([T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2}\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}[T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2})$ can be expressed as

$$\lambda_{\max} = \inf_{a \geq 0} \{a : a\text{Id}_{\mathcal{N}} \succeq [T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2}\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}[T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1/2}\} = \inf_{a \geq 0} \{a : aT_{\mathcal{N}}^*T_{\mathcal{N}} \succeq \tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}\}.$$

Keeping in mind that $T\hat{f} = y$ and that $T^*T\hat{f} \perp \mathcal{N}$ —which yields $\hat{f} = \Lambda^\dagger y - [T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}T_{\mathcal{N}}^*T\Lambda^\dagger y$ —we note that

$$\begin{aligned} \|T\hat{f}\|^2 &= \langle T\hat{f}, T\Lambda^\dagger y + T(\hat{f} - \Lambda^\dagger y) \rangle = \langle T\hat{f}, T\Lambda^\dagger y \rangle = \langle T\Lambda^\dagger y - T[T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}T_{\mathcal{N}}^*T\Lambda^\dagger y, T\Lambda^\dagger y \rangle \\ &= \|T\Lambda^\dagger y\|^2 - \langle [T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}T_{\mathcal{N}}^*T\Lambda^\dagger y, T_{\mathcal{N}}^*T\Lambda^\dagger y \rangle = \|T\Lambda^\dagger y\|^2 - (T_{\mathcal{N}}^*T\Lambda^\dagger y)^*[T_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}(T_{\mathcal{N}}^*T\Lambda^\dagger y). \end{aligned}$$

It follows that the squared Chebyshev center can be written as

$$(21) \quad \rho^2 = \inf_{a \geq 0} \{a(1 - \|T\Lambda^\dagger y\|^2) + (aT_{\mathcal{N}}^*T\Lambda^\dagger y)^*[aT_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}(aT_{\mathcal{N}}^*T\Lambda^\dagger y) : aT_{\mathcal{N}}^*T_{\mathcal{N}} \succeq \tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}\}.$$

Relying on a classical property of Schur complements, we now notice that

$$(22) \quad \begin{aligned} (aT_{\mathcal{N}}^*T\Lambda^\dagger y)^*[aT_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}(aT_{\mathcal{N}}^*T\Lambda^\dagger y) &= \inf_{b \geq 0} \{b : b \geq (aT_{\mathcal{N}}^*T\Lambda^\dagger y)^*[aT_{\mathcal{N}}^*T_{\mathcal{N}}]^{-1}(aT_{\mathcal{N}}^*T\Lambda^\dagger y)\} \\ &= \inf_{b \geq 0} \left\{ b : \left[\begin{array}{c|c} aT_{\mathcal{N}}^*T_{\mathcal{N}} & aT_{\mathcal{N}}^*T\Lambda^\dagger y \\ \hline (aT_{\mathcal{N}}^*T\Lambda^\dagger y)^* & b \end{array} \right] \succeq 0 \right\}. \end{aligned}$$

Putting (21) and (22) together, we now see that ρ^2 has the semidefinite expression

$$\rho^2 = \inf_{a, b \geq 0} \left\{ a(1 - \|T\Lambda^\dagger y\|^2) + b : aT_{\mathcal{N}}^*T_{\mathcal{N}} \succeq \tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}, \left[\begin{array}{c|c} aT_{\mathcal{N}}^*T_{\mathcal{N}} & aT_{\mathcal{N}}^*T\Lambda^\dagger y \\ \hline (aT_{\mathcal{N}}^*T\Lambda^\dagger y)^* & b \end{array} \right] \succeq 0 \right\}.$$

Next, for the two-hyperellipsoid-intersection, our upper bound for the Chebyshev radius is the minimal value of $\lambda(\tau) \times (1 - ((1-\tau)\|Rf^\tau\|^2 + \tau\|Sf^\tau\|^2))$ over $\tau \in [0, 1]$. For a fixed τ , this reduces to the previous situation by considering $T = [(1-\tau)R^*R + \tau S^*S]^{1/2}$, yielding a semidefinite expression with variables $a, b \geq 0$ that ought to be minimized over $\tau \in [0, 1]$ as well. By making the change of variables $c = a(1-\tau)$ and $d = a\tau$, we arrive, after some work, at the semidefinite program announced in (20).

- b) As noted in a), the optimal parameter τ^y can be found by solving the semidefinite program (20) before setting $\tau^y = d^y/(c^y + d^y)$. This parameter can alternatively be obtained by solving the equation $d\|\tilde{Q}h^\tau\|^2/d\tau = 0$ —recall that $\|\tilde{Q}h^\tau\|^2 = (1 - G(\tau))/H(\tau)$ and (see (19)) that $d\|\tilde{Q}h^\tau\|^2/d\tau = F(\tau)/H(\tau)$, where $F(\tau) := (\|Rf^\tau\|^2 + \|Rh^\tau\|^2) - (\|Sf^\tau\|^2 + \|Sh^\tau\|^2)$. Thus, we need to solve the equation $F(\tau) = 0$, which can be done via the Newton iteration

$$\tau \leftarrow \tau - \frac{F(\tau)}{F'(\tau)}$$

provided that $F(\tau)$ and $F'(\tau)$ are accessible throughout the iterations. Note that performing an eigendecomposition at each iteration gives $H(\tau) = 1/\lambda(\tau)$ and h^τ and that f^τ can also be computed, thus making $F(\tau)$ accessible. As for $F'(\tau)$, it can also become accessible in some cases, e.g. when $\tilde{Q} = \text{Id}_F$. To see this, notice that

$$F'(\tau) = 2(\alpha - \beta), \quad \alpha := \left\langle (R^*R - S^*S)f^\tau, \frac{df^\tau}{d\tau} \right\rangle, \quad \beta := \left\langle (R^*R - S^*S)h^\tau, \frac{dh^\tau}{d\tau} \right\rangle.$$

We can access α via the formula $f^\tau = \Lambda^\dagger y - [(1-\tau)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau S_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1}((1-\tau)R_{\mathcal{N}}^*R + \tau S_{\mathcal{N}}^*S)\Lambda^\dagger y$ written as $[(1-\tau)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau S_{\mathcal{N}}^*S_{\mathcal{N}}](f^\tau - \Lambda^\dagger y) = -((1-\tau)R_{\mathcal{N}}^*R + \tau S_{\mathcal{N}}^*S)\Lambda^\dagger y$, which differentiates to give, after some work,

$$\frac{df^\tau}{d\tau} = [(1-\tau)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau S_{\mathcal{N}}^*S_{\mathcal{N}}]^{-1}(R_{\mathcal{N}}^*R - S_{\mathcal{N}}^*S)f^\tau.$$

Turning to β , we notice that the eigenequation $[(1-\tau)R_{\mathcal{N}}^*R_{\mathcal{N}} + \tau S_{\mathcal{N}}^*S_{\mathcal{N}}]h^\tau = H(\tau)\tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}h^\tau$ yields

$$\left\langle h^\tau, \frac{dh^\tau}{d\tau} \right\rangle - \tau \left\langle (R_{\mathcal{N}}^*R_{\mathcal{N}} - S_{\mathcal{N}}^*S_{\mathcal{N}})h^\tau, \frac{dh^\tau}{d\tau} \right\rangle = H(\tau) \left\langle \tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}h^\tau, \frac{dh^\tau}{d\tau} \right\rangle,$$

while

$$\left\langle \tilde{Q}_{\mathcal{N}}^*\tilde{Q}_{\mathcal{N}}h^\tau, \frac{dh^\tau}{d\tau} \right\rangle = \frac{1}{2} \frac{d\|\tilde{Q}h^\tau\|^2}{d\tau} = \frac{F(\tau)}{2H(\tau)},$$

so that β is accessible as soon as $\left\langle h^\tau, \frac{dh^\tau}{d\tau} \right\rangle$ is. But when $\tilde{Q} = \text{Id}_F$, the latter is indeed accessible since

$$\left\langle h^\tau, \frac{dh^\tau}{d\tau} \right\rangle = \frac{1}{2} \frac{d\|h^\tau\|^2}{d\tau} = \frac{1}{2} \frac{d\|\tilde{Q}h^\tau\|^2}{d\tau} = \frac{F(\tau)}{2H(\tau)}.$$

- c) It took considerable effort to determine the candidate locally optimal recovery map and, even then, we need an orthogonality condition to guarantee that we have found the genuine locally optimal recovery map. In contrast, it would take much less effort determine a locally near-optimal recovery map. Indeed, any map assigning to $y \in \mathbb{R}^m$ a model- and data-consistent $f^y \in F$ —i.e., such that $\|Rf^y\| \leq 1$, $\|Sf^y\| \leq 1$, and $\Lambda f^y = y$ —has a local worst-case error that is away from optimality by a factor two. To see this, with \hat{f}^y denoting the genuine Chebyshev

center, we notice that

$$\begin{aligned} \text{lwce}_y(f^y) &= \sup_{\substack{\|Rf\| \leq 1, \|Sf\| \leq 1 \\ \Lambda f = y}} \|Q(f) - Q(f^y)\| \leq \sup_{\substack{\|Rf\| \leq 1, \|Sf\| \leq 1 \\ \Lambda f = y}} \|Q(f) - Q(\hat{f}^y)\| + \|Q(\hat{f}^y) - Q(f^y)\| \\ &\leq 2 \sup_{\substack{\|Rf\| \leq 1, \|Sf\| \leq 1 \\ \Lambda f = y}} \|Q(f) - Q(\hat{f}^y)\| = 2 \text{lwce}_y(\hat{f}^y). \end{aligned}$$

Such a model- and data-consistent f^y can be obtained by solving the quadratic program

$$\underset{f \in F}{\text{minimize}} \max \{ \|Rf\|, \|Sf\| \} \quad \text{s.to} \quad \Lambda f = y.$$

Using arguments similar to the ones from [32, Theorem 4], we can show that the solution f^y to the above program actually agrees with a constrained regularizer f^τ for some $\tau \in [0, 1]$.

6 Locally Optimal Recovery in Hilbert Spaces

This section finally deals with the third multi-fidelity scenario mentioned in Section 2. Still working in Hilbert spaces with f_0 and $f_0 - f_1$ once more modeled as belonging to the hyperellipsoids \mathcal{K}_0 and \mathcal{K}_1 from (10), we now target locally optimal recovery maps, which are more challenging to produce than the globally optimal recovery maps treated in Section 4. More precisely, given a linear quantity of interest $Q : F \rightarrow Z$, we would like to construct, for every $y = [y_0; y_1] \in \mathbb{R}^{m_0+m_1}$, an element $\Delta^{\text{opt}}(y) = z^{\text{opt}}$ that minimizes over all $z \in Z$ the local worst-case error

$$\text{lwce}_y(z) = \sup_{\substack{\|P_0 f_0\| \leq \varepsilon_0, \|P_1(f_0 - f_1)\| \leq \varepsilon_1 \\ \Lambda_0 f_0 = y_0, \Lambda_1 f_1 = y_1}} \|Q(f_0) - z\|.$$

We capitalize on the result from the previous section by reformulating the latter as

$$(23) \quad \text{lwce}_y(z) = \sup_{\substack{\|Rf\| \leq 1, \|Sf\| \leq 1 \\ \Lambda f = y}} \|\tilde{Q}(f) - z\|,$$

where the linear maps R , S , Λ , and \tilde{Q} are the ones introduced in (12). Our main result shows that, under some specific assumptions, constrained regularization yet again provides an optimal recovery map, but this time the optimal parameter $\tau^y \in [0, 1]$ depends on y , making the locally optimal recovery map nonlinear.

Theorem 8. Suppose that F is a Hilbert space and that the model sets $\mathcal{K}_0, \mathcal{K}_1 \subseteq F$ are the hyperellipsoidal sets defined in (10). Suppose also that $\ker(\Lambda_0) \subseteq \ker(P_1)$ or $\ker(\Lambda_0) \subseteq \ker(\Lambda_1)$. If the linear quantity of interest $Q : F \rightarrow Z$ maps into a Hilbert space, then a locally optimal recovery map is given by $\Delta^{\text{opt}} : y \in \mathbb{R}^{m_0+m_1} \mapsto Q(f_0^{\tau^y}) \in Z$, where

$$[f_0^\tau; f_1^\tau] := \underset{f = [f_0; f_1] \in F \times F}{\text{argmin}} [(1 - \tau)\|P_0 f_0\|^2 + \tau\|P_1(f_0 - f_1)\|^2] \quad \text{s.to} \quad \Lambda_0 f_0 = y_0, \Lambda_1 f_1 = y_1.$$

Here, the regularization parameter τ^y is $\tau^y = c_1^y / (c_0^y + c_1^y)$, with $b^y, c_0^y, c_1^y \geq 0$ being solutions to the following semidefinite program featuring $\mathcal{N}_0 := \ker(\Lambda_0)$, $\mathcal{N}_1 := \ker(\Lambda_1)$, $u_0 := P_0 \Lambda_0^\dagger y_0$, and $u_1 := P_1(\Lambda_0^\dagger y_0 - \Lambda_1^\dagger y_1)$:

$$(24) \quad \begin{aligned} & \underset{b, c_0, c_1 \geq 0}{\text{minimize}} && c_0(\varepsilon_0^2 - \|u_0\|^2) + c_1(\varepsilon_1^2 - \|u_1\|^2) + b \\ \text{s.to} &&& \begin{bmatrix} c_0 P_{0, \mathcal{N}_0}^* P_{0, \mathcal{N}_0} + c_1 P_{1, \mathcal{N}_1}^* P_{1, \mathcal{N}_1} & | & -c_1 P_{1, \mathcal{N}_1}^* P_{1, \mathcal{N}_1} \\ -c_1 P_{1, \mathcal{N}_1}^* P_{1, \mathcal{N}_1} & | & c_1 P_{1, \mathcal{N}_1}^* P_{1, \mathcal{N}_1} \end{bmatrix} \succeq \begin{bmatrix} Q_{\mathcal{N}_0}^* Q_{\mathcal{N}_0} & | & 0 \\ 0 & | & 0 \end{bmatrix} \\ \text{and} &&& \begin{bmatrix} c_0 P_{0, \mathcal{N}_0}^* P_{0, \mathcal{N}_0} + c_1 P_{1, \mathcal{N}_1}^* P_{1, \mathcal{N}_1} & | & -c_1 P_{1, \mathcal{N}_1}^* P_{1, \mathcal{N}_1} & | & c_0 P_{0, \mathcal{N}_0}^* u_0 + c_1 P_{1, \mathcal{N}_1}^* u_1 \\ -c_1 P_{1, \mathcal{N}_1}^* P_{1, \mathcal{N}_1} & | & c_1 P_{1, \mathcal{N}_1}^* P_{1, \mathcal{N}_1} & | & -c_1 P_{1, \mathcal{N}_1}^* u_1 \\ \hline (c_0 P_{0, \mathcal{N}_0}^* u_0 + c_1 P_{1, \mathcal{N}_1}^* u_1)^* & | & -(c_1 P_{1, \mathcal{N}_1}^* u_1)^* & | & b \end{bmatrix} \succeq 0. \end{aligned}$$

The key is to leverage the additional assumptions to certify that the necessary (and sufficient) orthogonality condition from Theorem 5 is met, as shown below.

Lemma 9. If $\ker(\Lambda_0) \subseteq \ker(P_1)$ or $\ker(\Lambda_0) \subseteq \ker(\Lambda_1)$, then, for any $\tau \in (0, 1)$,

$$\langle P_0 f_0^\tau, P_0 h_0 \rangle = 0 \quad \text{and} \quad \langle P_1(f_0^\tau - f_1^\tau), P_1(h_0 - h_1) \rangle = 0 \quad \text{for all } h_0 \in \ker(\Lambda_0), h_1 \in \ker(\Lambda_1).$$

Proof. With the notation of (12), we see that $f^\tau = [f_0^\tau; f_1^\tau]$ is the minimizer of $(1-\tau)\|Rf\|^2 + \tau\|Sf\|^2$ subject to $\Lambda f = y$. As such, it must satisfy $((1-\tau)R^*R + \tau S^*S)f^\tau \perp \ker(\Lambda)$, in other words, $(1-\tau)\langle Rf^\tau, Rh \rangle + \tau\langle Sf^\tau, Sh \rangle = 0$ for all $h = [h_0; h_1] \in \ker \Lambda$. Reinserting the specific expressions of R and S , this reads

$$\frac{1-\tau}{\varepsilon_0^2} \langle P_0(f_0^\tau), P_0(h_0) \rangle + \frac{\tau}{\varepsilon_1^2} \langle P_1(f_0^\tau - f_1^\tau), P_1(h_0 - h_1) \rangle = 0 \quad \text{for all } h_0 \in \ker(\Lambda_0), h_1 \in \ker(\Lambda_1),$$

which itself decouples as the two identities

$$(25) \quad \frac{1-\tau}{\varepsilon_0^2} \langle P_0 f_0^\tau, P_0 h_0 \rangle + \frac{\tau}{\varepsilon_1^2} \langle P_1(f_0^\tau - f_1^\tau), P_1 h_0 \rangle = 0 \quad \text{for all } h_0 \in \ker \Lambda_0,$$

$$(26) \quad \frac{\tau}{\varepsilon_1^2} \langle P_1(f_0^\tau - f_1^\tau), P_1 h_1 \rangle = 0 \quad \text{for all } h_1 \in \ker \Lambda_1.$$

First, let us assume that $\ker(\Lambda_0) \subseteq \ker(P_1)$. Then (25) directly implies that $\langle P_0 f_0^\tau, P_0 h_0 \rangle = 0$ for all $h_0 \in \ker(\Lambda_0)$, while $\langle P_1(f_0^\tau - f_1^\tau), P_1(h_0 - h_1) \rangle = -\langle P_1(f_0^\tau - f_1^\tau), P_1 h_1 \rangle = 0$ for all $h_0 \in \ker(\Lambda_0)$ and $h_1 \in \ker(\Lambda_1)$ according to (26). Second, let us now assume that $\ker(\Lambda_0) \subseteq \ker(\Lambda_1)$. Then (26) implies that $\langle P_1(f_0^\tau - f_1^\tau), P_1 h_0 \rangle = 0$, which, when substituted in (25), yields $\langle P_0 f_0^\tau, P_0 h_0 \rangle = 0$ for all $h_0 \in \ker(\Lambda_0)$ and, when combined with (26), yields $\langle P_1(f_0^\tau - f_1^\tau), P_1(h_0 - h_1) \rangle = 0$ for all $h_0 \in \ker(\Lambda_0)$ and $h_1 \in \ker(\Lambda_1)$. \square

We are now ready to justify the main result of this section.

Proof of Theorem 8. For $y \in \mathbb{R}^{m_0+m_1}$, in view of the reformulation (23) of the local worst-case error, the desired $\Delta^{\text{opt}}(y)$ is the Chebyshev center of the set $\tilde{Q}(\{f \in F : \|Rf\| \leq 1, \|Sf\| \leq 1, \Lambda f = y\})$. By Theorem 5 and using the notation introduced there, the latter is equal to $\tilde{Q}(f^{\tau^y})$, i.e., $Q(f_0^{\tau^y})$, provided that $\langle Rf^{\tau^y}, Rh^{\tau^y} \rangle = 0$, i.e., $\langle P_0 f_0^{\tau^y}, P_0 h_0^{\tau^y} \rangle = 0$. But Lemma 9 guarantees that this orthogonality condition holds for any $h_0 \in \ker(\Lambda_0)$ and in particular for $h_0^{\tau^y}$. Now, according to a) in the remark closing Section 5, the parameter τ^y is obtained as $\tau^y = c_1^y / (c_0^y + c_1^y)$ where $b^y, c_0^y, c_1^y \geq 0$ solve (20). By substituting the expressions for R, S, Λ , and \tilde{Q} from (12), we transform this semidefinite program, after some routine work, into the announced semidefinite program (24). \square

References

- [1] Edwards, P. N.: History of climate modeling. Wiley Interdisciplinary Reviews: Climate Change **2**:1, 128–139 (2011).
- [2] Hoste, J.-J. O., Casseau, V., Fossati, M., Taylor, I. J., Gollan, R.: Numerical modeling and simulation of supersonic flows in propulsion systems by open-source solvers. 21st AIAA International Space Planes and Hypersonics Technologies Conference, 2411 (2017).
- [3] Chen, Y.-S., Chou, T.H., Gu, B.R., Wu, J.-S., Wu, B., Lian, Y.Y., Yang, L.: Multiphysics simulations of rocket engine combustion. Computers & Fluids **45**:1, 29–36 (2011).
- [4] Steinhauser, M. O., Hiermaier, S.: A review of computational methods in materials science: examples from shock-wave and polymer physics. International Journal of Molecular Sciences **10**:12, 5135–5216 (2009).
- [5] Crutchfield, J. P.: The calculi of emergence: computation, dynamics and induction. Physica D: Nonlinear Phenomena **75**:1-3, 11–54 (1994).
- [6] Fernandez-Godino, M. G.: Review of multi-fidelity models. Advances in Computational Science and Engineering **1**:4, 351–400 (2023).
- [7] Peherstorfer, B., Willcox, K., Gunzburger, M.: Survey of multi-fidelity methods in uncertainty propagation, inference, and optimization. SIAM Review **60**:3, 550–591 (2018).
- [8] Koziel, S., Ciaurri, D. E., Leifsson, L.: Surrogate-based methods. In: Computational Optimization, Methods and Algorithms, pp 33–59. Springer (2011).
- [9] Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., Tucker, P. K.: Surrogate-based analysis and optimization. Progress in Aerospace Sciences **41**:1, 1–28 (2005).
- [10] Santner, T. J., Williams, B. J., Notz, W. I.: The Design and Analysis of Computer Experiments (Vol. 1). Springer (2003).
- [11] Forrester, A. I., Sóbester, A., Keane, A. J.: Multi-fidelity optimization via surrogate modelling. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **463**:2088, 3251–3269 (2007).

-
- [12] Park, C., Haftka, R. T., Kim, N. H.: Remarks on multi-fidelity surrogates. *Structural and Multidisciplinary Optimization* **55**, 1029–1050 (2017).
- [13] Zhang, X., Xie, F., Ji, T., Zhu, Z., Zheng, Y.: Multi-fidelity deep neural network surrogate model for aerodynamic shape optimization. *Computer Methods in Applied Mechanics and Engineering* **373**, 113485 (2021).
- [14] Meng, X., Karniadakis, G. E.: A composite neural network that learns from multi-fidelity data: application to function approximation and inverse PDE problems. *Journal of Computational Physics* **401**, 109020 (2020).
- [15] Song, X., Lv, L., Sun, W., Zhang, J.: A radial basis function-based multi-fidelity surrogate model: exploring correlation between high-fidelity and low-fidelity models. *Structural and Multidisciplinary Optimization* **60**, 965–981 (2019).
- [16] Song, J., Chen, Y., Yue, Y.: A general framework for multi-fidelity Bayesian optimization with Gaussian processes. *22nd International Conference on Artificial Intelligence and Statistics*, 3158–3167 (2019).
- [17] Haaland, B., Wang, W., Maheshwari, V.: A framework for controlling sources of inaccuracy in Gaussian process emulation of deterministic computer experiments. *SIAM/ASA Journal on Uncertainty Quantification* **6:2**, 497–521 (2018).
- [18] Micchelli, C. A., Rivlin, T. J.: A survey of optimal recovery. In: *Optimal Estimation in Approximation Theory*. pp 1–54 (1977).
- [19] Frenklach, M., Packard, A., Garcia-Donato, G., Paulo, R., Sacks, J.: Comparison of statistical and deterministic frameworks of uncertainty quantification. *SIAM/ASA Journal on Uncertainty Quantification* **4:1**, 875–901 (2016).
- [20] Corbally, M. A., Hinz, N. S., Freye, C. E.: Comprehensive two-dimensional gas chromatography under low-pressure conditions. *Journal of Chromatography A* **1705**, 464203 (2023).
- [21] Davis, J. T., Beaux, M. F., Freye, C. E.: Evaluation of different getter substrates using two-dimensional gas chromatography with time of flight mass spectrometry. *Journal of Chromatography A* **1689**, 463760 (2023).
- [22] Schultz, L. J., Blanpied, G. S., Borozdin, K. N., Fraser, A. M., Hengartner, N. W., Klimenko, A. V., Morris, C. L., Orum, C., Sossong, M. J.: Statistical reconstruction for cosmic ray muon tomography. *IEEE Transactions on Image Processing* **16:8**, 1985–1993 (2007).
- [23] Cuéllar, L., Borozdin, K. N., Green, J. A., Hengartner, N. W., Morris, C., Schultz, L. J.: Soft cosmic ray tomography for detection of explosives. *IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, 968–970 (2009).
- [24] Hsiao, C.: *Analysis of Panel Data*. Cambridge University Press (2022).
- [25] Weiss, K., Khoshgoftaar, T. M., Wang, D.: A survey of transfer learning. *Journal of Big Data* **3**, 1–40 (2016).

- [26] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q.: A comprehensive survey on transfer learning. *Proceedings of the IEEE* **109**:1, 43–76 (2020).
- [27] Novak, E., Wozniakowski, H.: *Tractability of Multivariate Problems (Vol. 1)*. European Mathematical Society (2010).
- [28] Foucart, S.: *Mathematical Pictures at a Data Science Exhibition*. Cambridge University Press (2022).
- [29] Foucart, S., Hielsberg, M., Mullendore, G. L., Petrova, G., Wojtaszczyk, P.: Optimal algorithms for computing average temperatures. *Mathematics of Climate and Weather Forecasting* **5**, 34–44 (2019).
- [30] Foucart, S., Liao, C.: Radius of information for two intersected centered hyperellipsoids and implications in Optimal Recovery from inaccurate data. *Journal of Complexity* **83**, 101841 (2024).
- [31] Polyak, B. T.: Convexity of quadratic transformations and its use in control and optimization. *Journal of Optimization Theory and Applications* **99**, 553–583 (1998).
- [32] Foucart, S., Liao, C., Veldt, N.: On the optimal recovery of graph signals. *Proceedings of SampTA 2023*, New Haven, pp 1–5, doi: 10.1109/SampTA59647.2023.10301205.
- [33] Foucart, S., Liao, C.: S-procedure relaxation: a case of exactness involving Chebyshev centers. In: *Explorations in the Mathematics of Data Science —The Inaugural Volume of the Center for Approximation and Mathematical Data Analytics*. Birkhäuser (to appear).
- [34] Beck, A., Eldar, Y. C.: Regularization in regression with bounded noise: a Chebyshev center approach. *SIAM Journal on Matrix Analysis and Applications* **29**, 606–625 (2007).
- [35] Foucart, S., Liao, C.: Optimal recovery from inaccurate data in Hilbert spaces: regularize, but what of the parameter? *Constructive Approximation* **57**, 489–520 (2023).
- [36] Pólik, I., Terlaky, T.: A survey of the S-lemma. *SIAM Review* **49**, 371–418 (2007).
- [37] Beck, A., Eldar, Y. C.: Strong duality in nonconvex quadratic optimization with two quadratic constraints. *SIAM Journal on Optimization* **17**, 844–860 (2006).
- [38] Binev, P., Cohen, A., Dahmen, W., DeVore, R., Petrova, G., Wojtaszczyk, P.: Data assimilation in reduced modeling. *SIAM/ASA Journal on Uncertainty Quantification* **5**, 1–29 (2017).