Worst-Case Completion of Tensors with Approximately Few ANOVA Terms

Simon Foucart* and Jingchun Shao†— Texas A&M University

Abstract

In this article, the problem of completing a tensor from some incomplete knowledge of its entries is treated by adopting a worst-case perspective, given the realistic assumption that the tensor's low-order ANOVA terms are dominant. We survey and leverage some recent all-purpose results from the field of Optimal Recovery to provide solutions on a theoretical level. But the accompanying constructions of optimal completion procedures, which often feature semidefinite programs, are not directly applicable in the tensor case due to the huge dimensions involved. To resolve the issue, we put forward a storage-friendly way to produce low-order ANOVA projections based on the fast Fourier transform (FFT), while exploiting the specificities of the completion problem to efficiently compute regularizers and extremal eigenvalues. Numerical experiments on synthetic tensors and real-world datasets demonstrate the accuracy and scalability of our FFT-based method.

Key words and phrases: Optimal recovery, ANOVA decomposition, tensor computations

AMS classification: 15A69, 41A27, 65F99.

1 Introduction

In this article, we are seeking optimal algorithms for the recovery of a d-way tensor $X \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ known *a priori* to have approximate low complexity and observed *a posteriori* via the direct entries

$$y_{\mathbf{i}} = X[\mathbf{i}] = X[i_1, \ldots, i_d], \qquad \mathbf{i} \in \mathcal{O},$$

where $\mathcal{O}$ represents the multi-index set of observed entries. Without being too formal for now, approximate low complexity means that the low-order terms dominate in the ANOVA decomposition $X = \sum_{S \subseteq [1:d]} \Pi_S(X)$, about which details will appear in Section 2. Our recovery task consists in completing the tensor X based on this knowledge and on the observations $y_{\mathbf{i}} = X[\mathbf{i}]$, $\mathbf{i} \in \mathcal{O}$, with a

*S. F. is partially funded by the NSF (DMS-2505204).

†S. F. and J. S. also acknowledge support from Texas A&M University through the ASCEND Initiative.

focus on the worst-case setting. Thus, the task shall be carried out in the framework of Optimal Recovery.

This problem is closely related to tensor completion, where one aims at recovering a tensor from a limited number of entries by imposing an additional low-complexity structure. A common structural choice is low tensor rank, which leads to tensor analogues of matrix completion and compressive sensing. For example, low-rank tensor completion can be formulated as optimization programs over tensors of a fixed rank—such as Canonical Polyadic (CP) rank [Jain and Oh, 2014] and Tensor Train (TT) rank [Steinlechner, 2016]—or relaxed as convex optimization programs involving the tensor nuclear norm [Liu et al., 2013]. To solve these programs efficiently, Riemannian optimization methods have been widely developed [Kressner et al., 2014]. Other works explore structured sampling schemes, e.g. Zhang [2019] demonstrates that, in the third-order Tucker setting, a tensor can be exactly recovered from noiseless cross-tensor measurements whose count matches the intrinsic degrees of freedom. However, unlike in the matrix case, there is no universal preference among tensor norms and computations featuring these norms are notoriously difficult. For instance, computing the tensor rank and the tensor nuclear norm are generally NP-hard problems [Hillar and Lim, 2013]. In contrast to this traditional low-rank viewpoint, the present work assumes that the underlying tensor can be well approximated by the low-order terms of its ANOVA decomposition.

In the general framework of Optimal Recovery, tensors X give way to abstract objects f living in a Banach space F, the *a priori* information takes the form $f \in \mathcal{C}$ for a so-called model set $\mathcal{C} \subseteq F$, and the *a posteriori* information occurs as $y = \Lambda(f)$ for a linear map $\Lambda : F \to \mathbb{R}^m$ termed observation map. Towards the objective of optimally recovering a quantity of interest $\Gamma(f)$ for some $\Gamma : F \to G$, a recovery process is simply viewed as a map $\Delta : y \in \mathbb{R}^m \to \Delta(y) \in G$. To assess its performance, there are two important notions of worst-case errors:

- the global worst-case error of Δ, as defined by

$$\mathrm{gwce}(\Delta) = \sup_{f \in \mathcal{C}} \|\Gamma(f) - \Delta(\Lambda f)\|_G;$$

- the local worst-case error of $\Delta(y) =: g$ at the specific $y \in \Lambda(\mathcal{C})$, as defined by

$$\mathrm{lwce}_y(g) = \sup_{\substack{f \in \mathcal{C} \\ \Lambda(f) = y}} \|\Gamma(f) - g\|.$$

Our goal shall be to construct optimal recovery maps $\Delta^{\mathrm{opti}} : \mathbb{R}^m \to G$, be they globally optimal, i.e., $\mathrm{gcwe}(\Delta^{\mathrm{opti}}) \le \mathrm{gcwe}(\Delta)$ for any $\Delta : \mathbb{R}^m \to G$, or locally optimal, i.e., $\mathrm{lcwe}_y(\Delta^{\mathrm{opti}}(y)) \le \mathrm{lcwe}_y(g)$ for any $g \in G$. Of note, a locally optimal recovery map is automatically globally optimal.

A fertile specification of the general framework, extended to incorporate observation errors, has been studied by the first author in the series of works [Foucart et al., 2022, Foucart and Liao, 2023, Foucart et al., 2023, Foucart and Liao, 2024a,b, Foucart and Hengartner, 2025]. In this

specification, the spaces F and G are Hilbert spaces, which we assume from now on, and the model set $\mathcal{C}$ is an hyperellipsoid

$$\mathcal{C}_T := \{f \in F : \|T(f)\| \leq 1\}$$

relative to an operator $T : F \to F$. Of particular interest is the case of an approximability set

$$\mathcal{C}_{V,\varepsilon} := \{f \in F : \operatorname{dist}(f, V) \leq \varepsilon\} = \{f \in F : \|P_{V^\perp}(f)\| \leq \varepsilon\}$$

relative to a linear subspace V of F and to a parameter $\varepsilon > 0$. This approximability set coincides with the hyperellipsoid by taking $T = \varepsilon^{-1} P_{V^\perp}$ to be the suitably scaled orthogonal projector onto the orthogonal complement of V. Of further particular interest is the case of an observation map $\Lambda : F \to \mathbb{R}^m$ that satisfies $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$.

These two cases prevail in our tensor scenario. Indeed, the low-complexity assumption translates into X being well-approximated (up to some accuracy ε) by its orthogonal projection onto the ANOVA space of level s (for a small integer $s \geq 0$). Moreover, the assumption $\Lambda\Lambda^* = \mathrm{Id}_{\mathbb{R}^m}$ clearly holds when $\Lambda(X) = X_{|\mathcal{O}}$ consists of entries of X indexed by a set $\mathcal{O}$. Seemingly, then, there is not much to do, as the above-mentioned series of works gave an (almost) complete theoretical solution of the Optimal Recovery problem at hand. There is a caveat, though: because of the huge nominal dimension of the space $\mathbb{R}^{n_1 \times \cdots \times n_d}$, the computational recipes uncovered earlier cannot be executed in a straightforward manner. In fact, storing tensors—in a naive way—might not even be an option. Thus, the contribution of the present work is twofold. Firstly, it serves as a survey of disparate theoretical results—and even complements them with two new contributions. Secondly, and more importantly, it puts the theoretical results to the strenuous test of numerical realization with tensors, where the naive computational recipes must be adjusted.

The organization of the rest of the article is as follows. In Section 2, in order to formally introduce our approximability set, we recall the main features of the ANOVA decomposition, favoring an alternative view over the traditional one. Importantly, this view will lead to more efficient computations involving the fast Fourier transform (FFT) and to guarantees that our worst-case completion problem is well-posed. In Section 3, we study the optimal completion problem within the scenario of accurate observations, starting with an exposition of existing Optimal Recovery results and continuing with their computational adaptations to the high-dimensional tensor framework. Section 4 considers the optimal completion problem in the more realistic—and more difficult—scenario of inaccurate observations. Starting again with an exposition of existing results, it also contains new results, notably one improving the practical construction of Chebyshev centers in a situation that comprises the tensor completion task. It closes again with computational adaptations to the tensor framework. Finally, in Section 5, our numerical solutions are validated on a brief selection of relevant examples. Their MATLAB implementations can be found in the dedicated GitHub repository `https://github.com/Jingchun-Shao/OR_ANOVA`, which also contains a file allowing to reproduce the tests presented here.

2 The ANOVA Decomposition

In this section, we formalize our model assumption based on the ANOVA decomposition of a tensor. As a matter of fact, the ANOVA decomposition more generally applies to real-valued functions x depending on d variables $t_1 \in \Omega_1, \dots, t_d \in \Omega_d$, each Ω_j being equipped with a probability measure μ_j. Tensors in $\mathbb{R}^{n_1 \times \cdots \times n_d}$ are just the special case of $\Omega_j = [1 : n_j]$ equipped with the normalized counting measure. We will therefore now use the notation $x \in \mathbb{R}^{\Omega_1 \times \cdots \times \Omega_d}$ as a compromise between generic objects $f \in F$ and proper tensors $X \in \mathbb{R}^{n_1 \times \cdots \times n_d}$.

2.1 Traditional and alternative views

The ANOVA decomposition of a real-valued function $x \in \mathbb{R}^{\Omega_1 \times \cdots \times \Omega_d}$ of d variables is often encountered as the $L_2(\mu_1 \otimes \cdots \otimes \mu_d)$-orthogonal sum

$$x(t) = \sum_{S \subseteq [1:d]} x_S(t). \tag{1}$$

The functions x_S depend only on the variables t_j for $j \in S$ and are defined recursively via

$$x_S = P_S(x) - \sum_{\substack{R \subseteq S \\ R \neq S}} x_R, \tag{2}$$

where the operator P_S averages out all the variables outside of S. Thus, with $P_{-\ell}$ as a shorthand for $P_{[1:d]\setminus\{\ell\}}$, we have

$$P_S = \prod_{\ell \notin S} P_{-\ell}, \qquad \text{with} \quad (P_{-\ell}x)(t_1, \dots, t_d) := \int_{\Omega_\ell} x(t_1, \dots, t_{\ell-1}, \tau, t_{\ell+1}, \dots, t_d)\, d\mu_\ell(\tau).$$

Besides the traditional view (1)-(2) of the ANOVA decomposition, there is an alternative one found e.g. in [Takemura, 1983]. It will soon be exploited for efficient computing and is arguably quite simple to establish. We include our explanation for completeness.

Proposition 1. The ANOVA decomposition of a d-variate function x is

$$x = \sum_{S \subseteq [1:d]} \Pi_S(x), \tag{3}$$

where the linear operators Π_S are mutually orthogonal projectors given by

$$\Pi_S = \prod_{j \in S} (\mathrm{Id} - P_{-j})\ P_S = \sum_{R \subseteq S} (-1)^{|S|-|R|} P_R. \tag{4}$$

Proof. The arguments only use the fact that the operators P_{-j}, which average out the jth variable, are orthogonal projectors which commute with one another. It starts with the following expansion:

$$\mathrm{Id} = \prod_{i=1}^{d} (\mathrm{Id} - P_{-i} + P_{-i}) = \sum_{S \subseteq [1:d]} \prod_{j \in S} (\mathrm{Id} - P_{-j}) \prod_{\ell \notin S} P_{-\ell} = \sum_{S \subseteq [1:d]} \Pi_S.$$

This is the decomposition (3) in which we have defined

$$\Pi_S := \prod_{j \in S} (\mathrm{Id} - P_{-j}) \prod_{\ell \notin S} P_{-\ell} = \prod_{j \in S} (\mathrm{Id} - P_{-j}) \, P_S.$$

The commutativity of the P_{-i} readily implies that $\Pi_S^* = \Pi_S$ and $\Pi_S \Pi_S = \Pi_S$, i.e., that the Π_S are orthogonal projectors. Moreover, they are mutually orthogonal, i.e., $\Pi_S \Pi_{S'} = 0$ for $S \neq S'$—indeed, picking e.g. $i \in S' \setminus S$, say, one sees that $\Pi_S \Pi_{S'}$ contains the product $P_{-i}(\mathrm{Id} - P_{-i}) = 0$. Notice that the expansion of $\prod_{j \in S} (\mathrm{Id} - P_{-j})$ in the expression for Π_S yields

$$\Pi_S = \prod_{j \in S} (\mathrm{Id} - P_{-j}) \, P_S = \sum_{R \subseteq S} \prod_{i \in S \setminus R} (-P_{-i}) \, P_S = \sum_{R \subseteq S} (-1)^{|S| - |R|} P_R,$$

which is the second identity in (4). Finally, the recursive definition (2) is retrieved by way of yet another expansion, namely

$$P_S = \prod_{i \in S} (\mathrm{Id} - P_{-i} + P_{-i}) \, P_S = \sum_{R \subseteq S} \prod_{j \in R} (\mathrm{Id} - P_{-j}) \prod_{\ell \in S \setminus R} P_{-\ell} \, P_S = \sum_{R \subseteq S} \prod_{j \in R} (\mathrm{Id} - P_{-j}) \, P_R = \sum_{R \subseteq S} \Pi_R,$$

which yield $\Pi_S = P_S - \sum_{\substack{R \subseteq S \\ R \neq S}} \Pi_R$ after a rearrangement. □

2.2 The ANOVA-based approximability set

In order to sidestep the curse of dimensionality for the approximation of multivariate functions x, it is common to make some variable-diminution assumptions, e.g. (i) x belong to a weighted function space, where the importance of each variable t_j is quantified by a weight $\gamma_j > 0$ (see [Sloan and Woźniakowski, 1998]); (ii) x depend only on a few variables, either coordinate variables t_j with indices j belonging to an unknown set S (see [DeVore et al., 2011]) or reduced variables $\langle a_k, t \rangle$ with unknown vectors a_k (see [Cohen et al., 2012]); (iii) x is partially separable, i.e., the sum of functions of few variables. Assumption (iii) is evidently the one we shall focus on, specifically postulating that only the first few terms of the ANOVA decomposition of x matter, say those indexed by subsets S of $[1:d]$ with $|S| \leq s$ for some small integer $s \geq 0$.

In many practical applications, high-dimensional functions are indeed largely governed by individual variables and low-order interactions between them, i.e., somewhat equivalently, by their low-order

ANOVA terms. This observation motivates our use of ANOVA truncation as a low-complexity model. It also helps explain the empirical success of quasi-Monte Carlo methods: as shown by Owen [2003], ANOVA decompositions often reveal that standard quadrature test functions have an effective dimension much smaller than the ambient dimension. A similar principle underlies the High-Dimensional Model Representation (HDMR) framework of [Li, Wang, Rosenthal, and Rabitz, 2001], where complex chemical and molecular systems are approximated by sums of low-order component functions. From the approximation-theoretic side, Griebel [2006] emphasizes that ANOVA-type decompositions reveal the relative importance of variables and their interactions and discusses practical examples where low-order or rapidly decaying ANOVA structure appears, including molecular dynamics, Markov-process representations, and high-dimensional problems in mathematical finance.

To make things more precise, we now introduce the space

$$V_s := \bigoplus_{|S|\le s}^{\perp} \operatorname{range}(\Pi_S) = \operatorname{range}(\Pi_{\le s}), \qquad \text{where} \quad \Pi_{\le s} := \sum_{|S|\le s} \Pi_S. \tag{5}$$

Rather than assuming that our functions $x \in \mathbb{R}^{\Omega_1\times\cdots\times\Omega_d}$—or rather our tensors $X \in \mathbb{R}^{n_1\times\cdots\times n_d}$—belong exactly to V_s, we stipulate that they are well approximated by elements of V_s, say up to accuracy $\varepsilon > 0$. This leads us to use as model set the following approximability set:

$$\mathcal{C}_{V_s,\varepsilon} := \{X \in \mathbb{R}^{n_1\times\cdots\times n_d} : \|X - \Pi_{\le s}(X)\| \le \varepsilon\},$$

which appears as an hyperellipsoidal model set $\mathcal{C}_T$ with $T = \varepsilon^{-1}(\mathrm{Id}-\Pi_{\le s})$. For such hyperellipsoidal model sets, a necessary condition for the Optimal Recovery problem to even make sense is easily[1] seen to be $\ker(T) \cap \ker(\Lambda) = \{0\}$. In our situation, this means that

$$V_s \cap \ker(\Lambda) = \{0\}. \tag{6}$$

Viewed as the injectivity of $\Lambda_{|V_s}$, this condition implies $\dim(V_s) \le |\mathcal{O}|$. To close this subsection, we estimate the dimension of V_s before uncovering a specific set $\mathcal{O}$ with size $|\mathcal{O}| = \dim(V_s)$ for which Condition (6) is fulfilled.

Lemma 2. The dimension of the space V_s defined in (5) is

$$k_s := \dim(V_s) = \sum_{|S|\le s} \prod_{j\in S} (n_j - 1). \tag{7}$$

In the particular case $n_1 = \cdots = n_d =: n$, it can be lower- and upper-estimated as

$$\left(\frac{d\,(n-1)}{s}\right)^s \le k_s \le \left(\frac{e\,d\,(n-1)}{s}\right)^s.$$

[1] If there was a nonzero $v \in \ker(T)$ such that $\Lambda(v) = 0$, then, picking some $f_0 \in \mathcal{C}_T$ with $\Lambda(f_0) = y$, any $f_t := f_0 + tv$ would belong to $\mathcal{C}_T$ and satisfy $\Lambda(f_t) = 0$, while $\|f_t - g\| \ge t\|v\| - \|f_0 - g\| \to +\infty$ as $t \to +\infty$, so the local worst-case error $\mathrm{lwce}_y(g)$ for the recovery of $\Gamma = \mathrm{Id}$ would be infinite for any y and g. Likewise for the global worst-case error of any Δ.

Proof. Since rank and trace coincide for (not necessarily orthogonal) projectors and in view of the second identity of (4), namely $\Pi_S = \sum_{R\subseteq S}(-1)^{|S|-|R|}P_R$, we observe that, for a fixed $S \subseteq [1:d]$,

$$\begin{aligned}\mathrm{rk}(\Pi_S) = \mathrm{tr}(\Pi_S) &= \sum_{R\subseteq S}(-1)^{|S|-|R|}\mathrm{tr}(P_R) = \sum_{R\subseteq S}(-1)^{|S|-|R|}\mathrm{rk}(P_R) = \sum_{R\subseteq S}(-1)^{|S|-|R|}\prod_{i\in R} n_i \\ &= \prod_{j\in S}(n_j - 1).\end{aligned}$$

Now, since the space V_s is the orthogonal sum of the spaces range(Π_S) for $|S| \le s$, its dimension is equal to $\sum_{|S|\le s}\mathrm{rk}(\Pi_S)$, which justifies (7). When $n_1 = \cdots = n_d = n$, this yields

$$\dim(V_s) = \sum_{r=0}^{s}\binom{d}{r}(n-1)^r.$$

For the lower bound, we only consider the summand corresponding to $r = s$ and use $\binom{d}{r} \ge \left(\frac{d}{s}\right)^s$. For the upper bound, we use $(n-1)^r \le (n-1)^s$ and the inequality $\sum_{r=0}^{s}\binom{d}{r} \le \left(\frac{ed}{s}\right)^s$, see e.g. [Foucart, 2022, Lemma 2.6]. □

The exact statement for the fulfillment of Condition 6 is included below, but its proof is postponed until the next subsection, as it uses an ingredient introduced there.

Proposition 3. With V_s being the subspace of $\mathbb{R}^{n_1\times\cdots\times n_d}$ defined in (5), Condition (6) holds for the observation map $\Lambda : X \in \mathbb{R}^{n_1\times\cdots\times n_d} \mapsto \big(X[\mathbf{i}], i \in \mathcal{O}\big) \in \mathbb{R}^{|\mathcal{O}|}$ associated any multi-index set $\mathcal{O}$ containing the set $\mathcal{I}_{d,s} = \big\{\mathbf{i} = (i_1,\ldots,i_d) : |\{j \in [1:d] : i_j \ne 1\}| \le s\big\}$.

2.3 Efficient computations of the ANOVA projections

Assuming for the sake of discussion that $n_1 = \cdots = n_d = n$, we consider a tensor $X \in \mathbb{R}^{n\times\cdots\times n}$ and a subset S of $[1:d]$ with size $|S| = s$. The cost of naively computing the ANOVA projection $X_S = \Pi_S(X)$ via the recursive definition (2), namely via $X_S = P_S(X) - \sum_{\substack{R\subseteq S\\ R\ne S}} X_R$, is at least $2\,s!\,n^d$. Indeed, evaluating $P_S(X)$ involves n^{d-s} additions to average out the variables outside of S, to be done at each of the n^s tuples $(t_i, i \in S)$, for a total cost of n^d. Thus, with c_r denoting the cost of computing x_R when $|R| = r$, we have $c_0 = n^d$, while the recursive definition (2) gives

$$c_s = n^d + \sum_{r=0}^{s-1}\binom{s}{r}c_r.$$

The claimed lower bound uses $\sum_{k\ge 0}\binom{k}{s}z^{k-s} = (1-z)^{-s-1}$ in the explicit solution of this recursion, namely

$$c_s = n^d\sum_{k=0}^{\infty}\frac{k^s}{2^k}, \qquad \text{so that} \qquad \frac{c_s}{s!\,n^d} \ge \sum_{k=0}^{\infty}\binom{k}{s}\frac{1}{2^k} = 2.$$

The cost of naively computing the ANOVA projection $\Pi_{\leq s}(X) = \sum_{|S|\leq s} \Pi_S(X)$ will be even higher. We are going to show that this cost can actually be reduced to $n^d \ln(n^d)$ through a Fourier-based approach exploiting the identity (4) representing Π_S as a product of orthogonal projectors. The approach, conceptually identical to the cluster expansion techniques widely used in computational materials science (see e.g. [Drautz, 2019]), is implemented in the dedicated repository accompanying this article. Starting with a fixed index set S, it can be formulated as follows by naturally extending the considerations to the complex setting.

Theorem 4. The ANOVA component of $X \in \mathbb{C}^{n_1 \times \cdots \times n_d}$ indexed by $S \subseteq [1:d]$ can be expressed as

$$\Pi_S(X) = \mathcal{F}^{-1}\big(M_S \odot \mathcal{F}(X)\big),$$

where $\mathcal{F}$ is the d-dimensional Fourier transform on $\mathbb{C}^{n_1 \times \cdots \times n_d}$ producing discrete Fourier coefficients and where the mask $M_S \in \{0,1\}^{n_1 \times \cdots \times n_d}$ is given for $\mathbf{i} = (i_1, \ldots, i_d) \in [1:n_1] \times \cdots \times [1, n_d]$ by

$$(M_S)_{\mathbf{i}} = \begin{cases} 1 & \text{if } i_j \neq 1 \text{ for } j \in S \text{ and } i_\ell = 1 \text{ for } \ell \notin S, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. For each $j \in [1:d]$, let $(\varphi_1^{(j)}, \varphi_2^{(j)}, \ldots, \varphi_{n_j}^{(j)})$ denote an unnormalized Fourier basis of $\mathbb{C}^{n_j}$ with $\varphi_1^{(j)}$ being equal to the all-one vector. Then, for $\mathbf{i} = (i_1, \ldots, i_d) \in [1:n_1] \times \cdots \times [1, n_d]$, let $\phi_{\mathbf{i}} := \varphi_{i_1}^{(1)} \otimes \cdots \otimes \varphi_{i_d}^{(d)}$. It is easily seen—recalling that $P_{-\ell}$ averages out ℓ-th variable—that

$$P_{-\ell}(\phi_{\mathbf{i}}) = \begin{cases} \phi_{\mathbf{i}} & \text{if } i_\ell = 1, \\ 0 & \text{if } i_\ell \neq 1. \end{cases}$$

Therefore, in view of the leftmost identity of (4) affirming that $\Pi_S = \prod_{j \in S}(\mathrm{Id} - P_{-j}) \prod_{\ell \notin S} P_{-\ell}$, we obtain

$$\Pi_S(\phi_{\mathbf{i}}) = \left.\begin{cases} \phi_{\mathbf{i}} & \text{if } i_j \neq 1 \text{ for all } j \in S \text{ and } i_\ell = 1 \text{ for all } \ell \notin S, \\ 0 & \text{otherwise.} \end{cases}\right\} = (M_S)_{\mathbf{i}}\, \phi_{\mathbf{i}}.$$

Finally, for any $X \in \mathbb{C}^{n_1 \times \cdots \times n_d}$ written as $X = \mathcal{F}^{-1}\big(\mathcal{F}(X)\big) = \sum_{\mathbf{i}} \mathcal{F}(X)_{\mathbf{i}}\, \phi_{\mathbf{i}}$, we conclude that

$$\Pi_S(X) = \sum_{\mathbf{i}} \mathcal{F}(X)_{\mathbf{i}}\, (M_S)_{\mathbf{i}}\, \phi_{\mathbf{i}} = \sum_{\mathbf{i}} (M_S \odot \mathcal{F}(X))_{\mathbf{i}}\, \phi_{\mathbf{i}} = \mathcal{F}^{-1}\big(M_S \odot \mathcal{F}(X)\big),$$

as announced. □

Our claim about the cost of the above approach stems from neglecting the cost of constructing M_S, which is precomputed offline, and considering only the costs of constructing the d-dimensional Fourier transform on $\mathbb{C}^{n^d}$ and its inverse, each being known (see e.g. [Dudgeon and Mersereau, 1984, p.76]) to be of order $n^d \ln(n^d)$. In addition, recall that we are more interested in computing the

orthogonal projection of X onto the approximation space V_s, i.e., in $\Pi_{\le s}(X) = \sum_{|S|\le s} \Pi_S(X)$, than in computing an individual $\Pi_S(X)$. The computational cost remains of order $n^d \ln(n^d)$, though. This is because the following corollary indicates that the majority of the cost still stems from the Fourier transform and its inverse.

Corollary 5. The order-s ANOVA projection of $X \in \mathbb{C}^{n_1 \times \cdots \times n_d}$ can be expressed as

$$\Pi_{\le s}(X) = \mathcal{F}^{-1}\big(M_{\le s} \odot \mathcal{F}(X)\big), \tag{8}$$

where the binary mask $M_{\le s} \in \{0,1\}^{n_1 \times \cdots \times n_d}$ is precomputed offline through its entries given for $\mathbf{i} = (i_1, \ldots, i_d) \in [1:n_1] \times \cdots \times [1:n_d]$ by

$$(M_{\le s})_{\mathbf{i}} = \begin{cases} 1 & \text{if } |\{j \in [1:d] : i_j \neq 1\}| \le s, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. In view of $\Pi_{\le s}(X) = \sum_{|S|\le s} \Pi_S(X)$, the identity (8) follows with $M_{\le s} := \sum_{|S|\le s} M_S$. Note that $M_{\le s}$ takes values in $\{0,1\}$ because the summands $M_S \in \{0,1\}^{n_1 \times \cdots \times n_d}$ are disjointly supported, as Theorem 4 says that $(M_S)_{\mathbf{i}} = 1$ if and only if $S = \{j \in [1:d] : i_j \neq 1\}$. Moreover, the entry $(M_{\le s})_{\mathbf{i}}$ equals 1 if and only if there exists $S \subseteq [1:d]$ with $|S| \le s$ such that $(M_S)_{\mathbf{i}} = 1$, i.e., such that $S = \{j \in [1:d] : i_j \neq 1\}$. In other words, one has $(M_{\le s})_{\mathbf{i}} = 1$ if and only if $|\{j \in [1:d] : i_j \neq 1\}| \le s$, as desired. □

As a theoretical consequence of the above interpretation of the order-s ANOVA projection, we now prove that Condition 6 can be fulfilled with an observation set of minimal size $k_s = \dim(V_s)$. This set is the support of $M_{\le s}$, denoted by $\mathcal{I}_{d,s}$ in Proposition 3. However, for notational convenience, the proof below shall feature a version of the Fourier transform where the all-one vector appear in the last position. As a result, the necessary cyclic shifts transform this set into

$$\mathcal{I}_{d,s} = \big\{\mathbf{i} = (i_1, \ldots, i_d) : |\{j \in [1:d] : i_j \neq n_j\}| \le s\big\}.$$

Proof of Proposition 3. Showing that $V_s \cap \ker(\Lambda) = \{0\}$ consists in proving that, if $X \in \mathrm{range}(\Pi_{\le s})$ satisfies $X[\mathbf{i}] = 0$ for all $\mathbf{i} \in \mathcal{O}$, then $X = 0$. In view of (8), such an X can be written as $X = \mathcal{F}^{-1}(Z)$, where $Z := M_{\le s} \odot \mathcal{F}(X)$ is supported on $\mathcal{I}_{d,s}$. Taking $\mathcal{O} \supseteq \mathcal{I}_{d,s}$ into account, it is therefore enough to prove that

$$\text{if } Z \text{ is supported on } \mathcal{I}_{d,s} \text{ and satisfies } \mathcal{F}(Z)[\mathbf{i}] = 0 \text{ for all } i \in \mathcal{I}_{d,s}, \text{ then } Z = 0. \tag{9}$$

We shall prove by induction on $d \ge 1$ that (9) holds for all $s = 1, \ldots, d$. The base case $d = 1$ is nothing but the invertibility of the one-dimensional discrete Fourier transform. Now, assuming that (9) holds up to $d-1$ for all relevant s, our goal is to prove that it holds for d and all relevant s. As

the case $s = d$ is again nothing but the invertibility of the d-dimensional discrete Fourier transform, we suppose that $s \leq d-1$. At this point, it is useful to remark that

$$\mathcal{I}_{d,s} = (\mathcal{I}_{d-1,s} \times \{n_d\}) \sqcup (\mathcal{I}_{d-1,s-1} \times [1 : n_d - 1]) \qquad \text{and} \qquad \mathcal{I}_{d-1,s-1} \subseteq \mathcal{I}_{d-1,s}.$$

For instance, the former is used to explicit the identities $0 = \mathcal{F}(Z)[\mathbf{i}]$ for any $\mathbf{i} \in \mathcal{I}_{d,s}$ as

$$\begin{aligned}
0 &= \sum_{\mathbf{k} \in \mathcal{I}_{d,s}} Z[k_1, \ldots, k_{d-1}, k_d] \, e^{\mathrm{i}2\pi \left(\frac{i_1 k_1}{n_1} + \cdots + \frac{i_{d-1} k_{d-1}}{n_{d-1}} + \frac{i_d k_d}{n_d} \right)} \\
&= \sum_{(k_1, \ldots, k_{d-1}) \in \mathcal{I}_{d-1,s}} Z[k_1, \ldots, k_{d-1}, n_d] \, e^{\mathrm{i}2\pi \left(\frac{i_1 k_1}{n_1} + \cdots + \frac{i_{d-1} k_{d-1}}{n_{d-1}} \right)} \\
&\quad + \sum_{(k_1, \ldots, k_{d-1}) \in \mathcal{I}_{d-1,s-1}} \left(\sum_{k_d=1}^{n_d - 1} Z[k_1, \ldots, k_{d-1}, k_d] e^{\mathrm{i}2\pi \frac{i_d k_d}{n_d}} \right) e^{\mathrm{i}2\pi \left(\frac{i_1 k_1}{n_1} + \cdots + \frac{i_{d-1} k_{d-1}}{n_{d-1}} \right)}.
\end{aligned}$$

At this point, there are two cases to separate, namely $\mathbf{i} \in \mathcal{I}_{d-1,s} \times \{n_d\}$ and $\mathbf{i} \in \mathcal{I}_{d-1,s-1} \times [1 : n_d - 1]$. In the fist case, one obtains that, for all $(i_1, \ldots, i_{d-1}) \in \mathcal{I}_{d-1,s}$,

$$\begin{aligned}
0 &= \sum_{(k_1, \ldots, k_{d-1}) \in \mathcal{I}_{d-1,s} \setminus \mathcal{I}_{d-1,s-1}} Z[k_1, \ldots, k_{d-1}, n_d] \, e^{\mathrm{i}2\pi \left(\frac{i_1 k_1}{n_1} + \cdots + \frac{i_{d-1} k_{d-1}}{n_{d-1}} \right)} \\
&\quad + \sum_{(k_1, \ldots, k_{d-1}) \in \mathcal{I}_{d-1,s-1}} \left(\sum_{k_d=1}^{n_d} Z[k_1, \ldots, k_{d-1}, k_d] \right) e^{\mathrm{i}2\pi \left(\frac{i_1 k_1}{n_1} + \cdots + \frac{i_{d-1} k_{d-1}}{n_{d-1}} \right)}.
\end{aligned}$$

Invoking the induction hypothesis for $(d-1, s)$, one deduces that

$$\text{for all } (k_1, \ldots, k_{d-1}) \in \mathcal{I}_{d-1,s} \setminus \mathcal{I}_{d-1,s-1}, \qquad Z[k_1, \ldots, k_{d-1}, n_d] = 0, \tag{10}$$

$$\text{for all } (k_1, \ldots, k_{d-1}) \in \mathcal{I}_{d-1,s-1}, \qquad \sum_{k_d=1}^{n_d} Z[k_1, \ldots, k_{d-1}, k_d] = 0. \tag{11}$$

In the second case, making use of (10), one obtains that, for all $(i_1, \ldots, i_{d-1}) \in \mathcal{I}_{d-1,s-1}$ and all $i_d \in [1 : n_d - 1]$,

$$0 = \sum_{(k_1, \ldots, k_{d-1}) \in \mathcal{I}_{d-1,s-1}} \left(\sum_{k_d=1}^{n_d} Z[k_1, \ldots, k_{d-1}, k_d] e^{\mathrm{i}2\pi \frac{i_d k_d}{n_d}} \right) e^{\mathrm{i}2\pi \left(\frac{i_1 k_1}{n_1} + \cdots + \frac{i_{d-1} k_{d-1}}{n_{d-1}} \right)}.$$

Invoking the induction hypothesis for $(d-1, s-1)$, one deduces that, for all $(k_1, \ldots, k_{d-1}) \in \mathcal{I}_{d-1,s-1}$,

$$\sum_{k_d=1}^{n_d} Z[k_1, \ldots, k_{d-1}, k_d] e^{\mathrm{i}2\pi \frac{i_d k_d}{n_d}} = 0 \qquad \text{for all } i_d \in [1 : n_d - 1],$$

which also holds for $i_d = n_d$ by virtue of (11). Collectively interpreted as the one-dimensional discrete Fourier transform of $Z[k_1, \ldots, k_{d-1}, \cdot]$, these equalities imply that $Z[k_1, \ldots, k_{d-1}, k_d] = 0$

for all $(k_1, \ldots, k_{d-1}) \in \mathcal{I}_{d-1,s-1}$ and all $k_d \in [1 : n_d]$. In particular, the latter yields $Z[\mathbf{k}] = 0$ for all $\mathbf{k} \in \mathcal{I}_{d-1,s-1} \times [1 : n_d - 1]$. It also yields $Z[k_1, \ldots, k_{d-1}, n_d] = 0$ for all $(k_1, \ldots, k_{d-1}) \in \mathcal{I}_{d-1,s-1}$, which, combined with (10), implies that $Z[k_1, \ldots, k_{d-1}, n_d] = 0$ for all $(k_1, \ldots, k_{d-1}) \in \mathcal{I}_{d-1,s}$, or equivalently that $Z[\mathbf{k}] = 0$ for all $\mathbf{k} \in \mathcal{I}_{d-1,s} \times \{n_d\}$. Altogether, these two cases ensures that $Z[\mathbf{k}] = 0$ for all $\mathbf{k} \in \mathcal{I}_{d,s}$. Since Z was supported on $\mathcal{I}_{d,s}$, one concludes that $Z = 0$, hence establishing (9) for (d, s). This concludes the inductive proof. □

3 Optimal Completion from Accurate Observations

In this section, we discuss the Optimal Recovery problem, as abstractly described in the introduction but specified to the Hilbert setting and the hyperellipsoidal model set $\mathcal{C}_T = \{f \in F : \|T(f)\| \le 1\}$. In particular, we continue—for just a little longer—to work under the idealized assumption that the vector $y = \Lambda(f)$ is not corrupted by any observation error.

3.1 Summary of known results

Fixed worst-case errors. For a fixed and linear recovery map $\Delta^{\mathrm{lin}} : \mathbb{R}^m \to G$, determining its *global* worst-case error is rather simple. Its square can be expressed as

$$\begin{aligned} \mathrm{gwce}(\Delta^{\mathrm{lin}})^2 &:= \sup_{\|T(f)\|^2 \le 1} \|\Gamma(f) - \Delta^{\mathrm{lin}}(\Lambda f)\|^2 \\ &= \inf_{c \ge 0} c \quad \text{s.to } \|(\Gamma - \Delta^{\mathrm{lin}}\Lambda)(f)\|^2 \le c\,\|T(f)\|^2 \ \ \forall f \in F. \end{aligned}$$

When the Hilbert space F is of finite dimension, say N, the above translates into an expression computable by semidefinite programming[2], namely

$$\mathrm{gwce}(\Delta^{\mathrm{lin}})^2 = \inf_{c \ge 0} c \quad \text{s.to } c\,T^*T - (\Gamma - \Delta^{\mathrm{lin}}\Lambda)^*(\Gamma - \Delta^{\mathrm{lin}}\Lambda) \succeq 0. \tag{12}$$

In the more delicate *local* setting, for fixed $y \in \mathbb{R}^m$ and $g \in G$, with

$$f_y := \operatorname{argmin} \{\|T(f)\| : \ \Lambda(f) = y\},$$

it was shown in Theorem 1 of [Foucart, Liao, Shahrampour, and Wang, 2022] (albeit formulated slightly differently and only for $\Gamma = \mathrm{Id}$) that the squared local worst-case error can be expressed as

$$\begin{aligned} \mathrm{lwce}_y(g)^2 &:= \sup_{\substack{\|T(f)\|^2 \le 1 \\ \Lambda(f) = y}} \|\Gamma(f) - g\|^2 \\ &= \inf_{c \ge 0, d \in \mathbb{R}} \|\Gamma(f_y) - g\|^2 + d \\ &\qquad \text{s.to } \langle (cT^*T - \Gamma^*\Gamma)h, h\rangle + 2\langle \Gamma^*(g - \Gamma(f_y)), h\rangle + d - c(1 - \|T(f_y)\|^2) \ge 0 \ \forall h \in \ker(\Lambda). \end{aligned}$$

[2] If T was invertible (which is not the case for an approximability set), then solving a semidefinite program is an overkill: an eigenvalue computation is enough, as $\mathrm{gwce}(\Delta^{\mathrm{lin}}) = \mathrm{eig}_{\max}((\Gamma - \Delta^{\mathrm{lin}}\Lambda)T^{-1})$.

The main tool for the derivation of this expression was the S-lemma. When the Hilbert space F is of finite dimension N, considering a matrix $M \in \mathbb{R}^{N \times (N-m)}$ whose columns span $\ker(\Lambda)$, the above again translates into an expression computable by semidefinite programming, namely

$$\text{(13)} \quad \operatorname{lwce}_y(g)^2 = \|\Gamma(f_y) - g\|^2 + \inf_{c \geq 0, d \in \mathbb{R}} d : \quad \left[\begin{array}{c|c} M^*(cT^*T - \Gamma^*\Gamma)M & M^*\Gamma^*(g - \Gamma f_y) \\ \hline (g - \Gamma f_y)^*\Gamma M & d - c(1 - \|Tf_y\|^2) \end{array}\right] \succeq 0.$$

Minimized worst-case errors. In both the global and local settings, computing an optimal recovery map becomes possible by further minimization. In the global setting, admitting the theoretical fact that there are linear recovery maps among the optimal ones, we can minimize the expression (12) over all linear maps Δ^{lin}. We indeed end up with a computationally feasible program after noticing that the constraint $cT^*T - (\Gamma - \Delta^{\text{lin}}\Lambda)^*(\Gamma - \Delta^{\text{lin}}\Lambda) \succeq 0$ can be reformulated as

$$\left[\begin{array}{c|c} \mathrm{Id} & \Gamma - \Delta^{\text{lin}}\Lambda \\ \hline (\Gamma - \Delta^{\text{lin}}\Lambda)^* & cT^*T \end{array}\right] \succeq 0.$$

In the local setting, for each $y \in \Lambda(\mathcal{C}_T)$, finding the minimizer[3] over all g of the expression (13) is again computationally feasible, thanks to the affine dependence on (c, d, g) of the constraint and to the remark that

$$\|\Gamma(f_y) - g\|^2 = \inf_{t \geq 0} t : \quad \left[\begin{array}{c|c} \mathrm{Id} & \Gamma(f_y) - g \\ \hline (\Gamma(f_y) - g)^* & t \end{array}\right] \succeq 0,$$

where the latter constraint depends affinely on (g, t). All in all, one arrives at a semidefinite program with variables c, d, g, and t.

We shall not discuss efficient ways to solve such semidefinite programs. In fact, we would rather bypass them, since they are quite inappropriate to deal with tensors. Indeed, current semidefinite solvers struggle when the involved matrices have size in the thousands—not that they take a long time to run, but simply that they do not run because of memory issues. To the point, in the present situation, we are faced with matrices of size equal to the dimension of the tensor space, i.e., $n_1 \times \cdots \times n_d$, so already ten thousands for four-way tensors with each mode dimension equal to ten. Thus, an alternative to this semidefinite programming approach would be welcome, and fortuitously there is one. It covers both the global and local settings because one explicitly constructs the locally optimal recovery map outputting the Chebyshev center—usually called central algorithm in Information-Based Complexity—which is then automatically globally optimal. For $\Gamma = \mathrm{Id}$, it is known (see [Binev et al., 2017] for the case $T = \varepsilon^{-1}P_{V^\perp}$ and even [Kowalski et al., 1995, Theorem 299.1] for an arbitrary T) to be given via the minimal-norm interpolant (aka spline algorithm) as

$$\text{(14)} \qquad \Delta^{\text{cheb}} : y \in \mathbb{R}^m \mapsto f_y = \Big[\operatorname*{argmin}_{f \in F} \|T(f)\| : \ \Lambda(f) = y\Big] \in F.$$

[3] This minimizer is known as a Chebyshev center (of $\{f \in F : \|T(f)\| \leq 1, \Lambda(f) = y\}$) and is unique in our situation.

This turns out to be a linear map, with expression $\Delta^{\mathrm{cheb}} = (T^*T)^{-1}\Lambda^*\big(\Lambda(T^*T)^{-1}\Lambda^*\big)^{-1}$ when T is invertible. For an arbitrary linear Γ, it can be said (see Lemma 6 in [Foucart and Hengartner, 2025]) that $\Gamma \circ \Delta^{\mathrm{cheb}}$ is locally, hence globally, optimal and that the minimal squared local worst-case error equals

$$\min_g \mathrm{lwce}_y(g)^2 = \mathrm{eig}_{\max}\big([T_{|W}^* T_{|W}]^{-1/2}\Gamma_{|W}^*\Gamma_{|W}[T_{|W}^* T_{|W}]^{-1/2}\big) \times \big(1 - \|T(f_y)\|^2\big), \tag{15}$$

where $\mathrm{eig}_{\max}$ denotes the largest eigenvalue and where W stands for $\ker(\Lambda)$. Note that the condition $\ker(T) \cap \ker(\Lambda) = \{0\}$ is implicitly used here.

We will discuss below efficient ways to compute (14) and (15) for our tensorial problem—certainly, the explicit expression of Δ^{cheb} is not suitable, since T^*T is too large for direct inversion, having size $N \times N$ with $N = n_1 \cdots n_d$. But before specifying to tensors, we point out another expression, which is valid in a general setting. It involves the Riesz representers $(u_1, \ldots, u_m)$ of the coordinate functionals of $\Lambda : F \to \mathbb{R}^m$, so that $\Lambda(f) = [\langle u_1, f\rangle; \ldots; \langle u_m, f\rangle]$ for all $f \in F$, as well as a basis $(v_1, \ldots, v_k)$ of V. These appear though the Gramian $G \in \mathbb{R}^{m\times m}$ of $(u_1, \ldots, u_m)$ and the cross Gramian $C \in \mathbb{R}^{m\times k}$ of $(u_1, \ldots, u_m)$ with $(v_1, \ldots, v_k)$, which have entries

$$G_{i,i'} = \langle u_i, u_{i'}\rangle, \qquad C_{i,j} = \langle u_i, v_j\rangle, \qquad i, i' = 1, \ldots, m,\ j = 1, \ldots, k.$$

The desired expression was obtained in [Foucart, Liao, Shahrampour, and Wang, 2022, Theorem 2] (see also [Foucart, 2022, Proposition 10.3]) and reads

$$\Delta^{\mathrm{cheb}}(y) = \sum_{i=1}^m a_i u_i + \sum_{j=1}^k b_j v_j, \tag{16}$$

where

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

Notice that the sizes of the matrices to invert have gone down from the dimension N of the ambient space to the more manageable dimensions m of $\mathrm{range}(\Lambda)$ and k of V. Furthermore, in our tensor completion problem—and any problem where $\Lambda\Lambda^* = \mathrm{Id}$—the system $(u_1, \ldots, u_m)$ is an orthonormal basis of $\mathrm{range}(\Lambda^*)$, so no inversion of G is required since then $G = I_m$.

3.2 Tensor computations

At first sight, the summary we just presented seems to solve our abstract Optimal Recovery problem in the accurate scenario, at least in theory. In practice, constructing the optimal recovery map of (14) and evaluating the minimal worst-case error of (15) can run into computational challenges, especially when the dimension N of the ambient space is large. This is definitely the case for our tensor completion problem, where the dimension $N = n_1 \cdots n_d$ grows exponentially with d. Another issue is storage: when N is that large, storing the operator T in matrix form is unsuitable and we should instead exploit affordable ways to apply T. Thankfully, in case $T = \varepsilon^{-1} P_{V_s^\perp}$, which is a multiple of $\mathrm{Id} - \Pi_{\le s}$, the recipe (8) provides just that. The discussion below focuses on this case.

Optimal recovery map. As already mentioned, even if T was invertible, it is unrealistic to use direct methods for matrix inversions in the construction of the optimal recovery map via the formula $\Delta^{\mathrm{cheb}} = (T^*T)^{-1}\Lambda^*\big(\Lambda(T^*T)^{-1}\Lambda^*\big)^{-1}$. Indirect methods are more appropriate, particularly the conjugate gradient method, which is the method of choice when the operator is positive semidefinite and accessible only through its action. The method essentially aims at solving the minimization problem (14) using iterative solvers. As these solvers handle unconstrained problems more naturally than constrained one, we approximate the exact solution by way of a regularized problem that adds a large penalty enforcing the constraint to the objective. More details on this regularization will be given in Subsection 4.3, which will reveal that, provided $\Lambda\Lambda^* = \mathrm{Id}$, the exact—not approximate—solution to (14) can be obtained by solving two regularized problems.

Our preferred strategy is to exploit the formula (16). In our tensor completion problem, the basis $(u_1, \dots, u_m)$ of range(Λ^*) is given by the system $(E_{\mathbf{i}}, \mathbf{i} \in \mathcal{O})$, where $E_{\mathbf{i}} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is the canonical tensor equal to one on the $\mathbf{i}$th entry and to zero on all other entries, and so $G = \big[\langle u_i, u_{i'}\rangle\big] = I_m$. As for the basis $(v_1, \dots, v_{k_s})$ of V_s, Corollary 5 implies that it can be taken as the orthonormal system $(\mathcal{F}^{-1}(E_{\mathbf{j}}), \mathbf{j} \in \mathcal{I}_{d,s})$. The computational cost of producing this system is $k_s\, n^d \ln(n^d)$. Ignoring the cost of forming $\sum_i a_i u_i + \sum_j b_j v_j$ when a and b are available, the remaining cost for realizing the formula (16) then divides as:

(i) computing the cross Gramian $C \in \mathbb{R}^{m \times k_s}$: this is negligible, because the $\mathcal{F}^{-1}(E_{\mathbf{j}})$ have been produced above, so that $C = \big[\langle u_i, v_j\rangle\big] = \big[\mathcal{F}^{-1}(E_{\mathbf{j}})[\mathbf{i}]\big]$ is already accessible;

(ii) computing the vector $b \in \mathbb{R}^{k_s}$: forming $C^\top y$ requires $\mathcal{O}(k_s m)$ operations, forming $C^\top C$ requires $\mathcal{O}(k_s^2 m)$, and computing $b = (C^\top C)^{-1} C^\top y$ by solving the linear system $(C^\top C)b = C^\top y$ requires $\mathcal{O}(k_s^3)$, contributing to a total number of $\mathcal{O}(k_s m + k_s^2 m + k_s^3) = \mathcal{O}(k_s^2 m)$ operations, in view of $k_s \le m$;

(iii) computing the vector $a \in \mathbb{R}^m$: forming $a = y - Cb$ only requires a total number of $\mathcal{O}(k_s m)$ operations, which is dominated by the cost of (ii).

Putting everything together, the total cost of computing $\Delta^{\mathrm{cheb}}(y)$ is

$$\mathcal{O}(k_s(n^d \ln(n^d) + k_s m)) \quad \text{operations.}$$

For instance, thinking of s as a fixed small integer, so that $k_s \asymp d^s n^s$ by Lemma 2, and in the realistic situation where $m \asymp k_s$, we have $k_s m \lesssim k_s^2 \lesssim d^{2s} n^{2s} \lesssim n^d$, resulting in a cost of $\mathcal{O}(k_s n^d \ln(n^d))$. This is not much larger than the cost $\mathcal{O}(n^d \ln(n^d))$ of computing one ANOVA projection.

In passing, we mention another strategy for the computation of $\Delta^{\mathrm{cheb}}(y) = f_y$. The spirit is that of the representer theorem: since the solution has the form (16), we can exploit this knowledge when performing the minimization of (14), leading to the optimization program

$$\underset{a \in \mathbb{R}^m, b \in \mathbb{R}^k}{\text{minimize}} \left\| \sum_{i=1}^m a_i P_{V^\perp} u_i \right\| \qquad \text{s.to} \;\; a_i + \sum_{j=1}^k b_j \langle u_i, v_j\rangle = y_i \quad \text{for all } i \in [1:m].$$

We did not explore if there is a computational advantage in solving the latter via iterative methods, say via conjugate gradient.

Minimal worst-case error. Once we have computed the Chebyshev center f_y—that is, the locally optimal recovery map evaluated at y—determining the Chebyshev radius—that is, the minimal local worst-case error of f_y at y—is not necessarily straightforward. Resorting to (13) with $g = f_y$ is natural, of course, but this involves a semidefinite program whose matrices are too large to handle in a tensor setting. Resorting to (15) is more promising: as f_y has just been obtained, computing the term $(1 - \|T(f_y)\|^2)$ is not a problem, so it remains to compute the largest eigenvalue of the positive semidefinite operator $B^{-1/2}AB^{-1/2}$, where $A := \Gamma_{|W}^* \Gamma_{|W}$ and $B := T_{|W}^* T_{|W}$. Again, it is unrealistic to store these operators in matrix form, but their actions can be computed relatively cheaply, say when $T = \varepsilon^{-2} P_{V_s^\perp} = \varepsilon^{-2}(\mathrm{Id} - \Pi_{\le s})$. In this case, the power method is appropriate. It consists (usually with $B = I$) of iterating the scheme $Bx^{(t)} = Ax^{(t-1)}$, so that $\langle x^{(t)}, x^{(0)} \rangle / \langle x^{(0)}, x^{(0)} \rangle \to \mathrm{eig}_{\max}(B^{-1}A) = \mathrm{eig}_{\max}(B^{-1/2}AB^{-1/2})$ as $t \to \infty$. A more in-depth discussion will appear in Subsection 4.3.

4 Optimal Completion from Inaccurate Observations

In this section, we discuss the Optimal Recovery problem in the more realistic situation of an imperfect observation process. In an abstract formulation, the *a posteriori* information is now given as $y = \Lambda(f) + e$ for some unknown error vector $e \in \mathbb{R}^m$. As for the *a priori* information, it still reads $f \in \mathcal{C}$ for some model set $\mathcal{C} \subseteq F$, but is also augmented with $e \in \mathcal{E}$ for some uncertainty set $\mathcal{E} \subseteq \mathbb{R}^m$. For the recovery of a quantity of interest $\Gamma : F \to G$, the two notions of worst-case error are adjusted to

$$\mathrm{gwce}(\Delta) = \sup_{\substack{f \in C \\ e \in \mathcal{E}}} \|\Gamma(f) - \Delta(\Lambda f + e)\|,$$

$$\mathrm{lwce}_y(g) = \sup_{\substack{f \in \mathcal{C} \\ y - \Lambda f \in \mathcal{E}}} \|\Gamma(f) - g\|.$$

We note that this inaccurate scenario can be reduced to the accurate scenario by considering the compound object $\widetilde{f} := (f, e)$, with *a priori information* $\widetilde{f} \in \widetilde{\mathcal{C}} := \mathcal{C} \times \mathcal{E}$ and *a posteriori information* $y = \widetilde{\Lambda}(\widetilde{f}) := \Lambda f + e$. However, this formal reduction is of limited use for the construction of optimal recovery maps. As a case in point, assuming both model set $\mathcal{C}$ and uncertainty set $\mathcal{E}$ to be hyperellipsoids, the compound model set $\widetilde{\mathcal{C}}$ is not an hyperellipsoid anymore, so the results from the previous section are inapplicable. Below, we (almost) untangle this case, with specific model

and uncertainty sets given as

$$\mathcal{C}_T = \{f \in F : \|T(f)\| \le 1\},$$
$$\mathcal{E}_U = \{e \in \ell_2^m : \|U(e)\| \le 1\},$$

for some operators $T : F \to F$ and $U : \ell_2^m \to \ell_2^m$. Of particular interest are $T = \varepsilon^{-1} P_{V^\perp}$ and $U = \eta^{-1}\mathrm{Id}$, so that the *a priori* information reads $\mathrm{dist}(f, V) \le \varepsilon$ and $\|e\| \le \eta$.

4.1 Summary of known results

Fixed worst-case errors. For a fixed linear recovery map $\Delta^{\mathrm{lin}} : \mathbb{R}^m \to G$, the squared *global* worst-case error is

$$\begin{aligned} \mathrm{gwce}(\Delta^{\mathrm{lin}})^2 &:= \sup_{\substack{\|Tf\|^2 \le 1 \\ \|Ue\|^2 \le 1}} \|\Gamma(f) - \Delta^{\mathrm{lin}}(\Lambda f + e)\|^2 \\ &= \inf_{a,b \ge 0} a + b \quad \text{s.to } a\|Tf\|^2 + b\|Ue\|^2 \ge \|(\Gamma - \Delta^{\mathrm{lin}}\Lambda)f - \Delta^{\mathrm{lin}} e\|^2 \quad \forall f \in F, e \in \mathbb{R}^m. \end{aligned} \tag{17}$$

Although this has not been stated explicitly before (except in [Foucart and Liao, 2023, Lemma 12] for $\Gamma = \mathrm{Id}$, $T = \varepsilon^{-1} P_{V^\perp}$, and $U = \eta^{-1}\mathrm{Id}$), the argument simply relies on (the infinite-dimensional version of) Polyak S-procedure involving three quadratic forms. When the Hilbert space F is of finite dimension, the above translates into an expression computable by semidefinite programming, namely

$$\mathrm{gwce}(\Delta^{\mathrm{lin}})^2 = \inf_{a,b \ge 0} a+b \text{ s.to } \left[\begin{array}{c|c} aT^*T & 0 \\ \hline 0 & bU^*U \end{array}\right] \succeq \left[\begin{array}{c|c} (\Gamma - \Delta^{\mathrm{lin}}\Lambda)^*(\Gamma - \Delta^{\mathrm{lin}}\Lambda) & (\Gamma - \Delta^{\mathrm{lin}}\Lambda)^*\Delta^{\mathrm{lin}} \\ \hline (\Delta^{\mathrm{lin}})^*(\Gamma - \Delta^{\mathrm{lin}}\Lambda) & (\Delta^{\mathrm{lin}})^*\Delta^{\mathrm{lin}}. \end{array}\right].$$

In the more delicate *local* setting, because of the presence of linear terms in the quadratic functionals (e.g. $\|\Gamma(f) - g\|^2$ contains $\langle \Gamma(f), g\rangle$), the S-procedure is not exact, so a correct expression for the worst-case error is not immediately available—only an upper bound is.

Minimized global worst-case error. In the less challenging *global* setting, there is now a complete solution for the construction of a recovery map with minimal worst-case error. It is given by regularization via the map

$$\Delta^\sigma : y \in \mathbb{R}^m \mapsto f_y^\sigma := \left[\operatorname*{argmin}_{f \in F} \ (1-\sigma)\|T(f)\|^2 + \sigma\|U(y - \Lambda f)\|^2\right] \in F, \tag{18}$$

which is incidentally a linear map that can be expressed as $\Delta^\sigma = [(1-\sigma)T^*T + \sigma\Lambda^*U^*U\Lambda]^{-1}\sigma\Lambda^*U^*U$. Melkman and Micchelli [1979] were the first to show that a globally optimal recovery map occurs as Δ^σ for *some* parameter $\sigma \in [0, 1]$. Much later, in the case $\Gamma = \mathrm{Id}$, $T = \varepsilon^{-1} P_{V^\perp}$, and $U = \eta^{-1}\mathrm{Id}$, Foucart and Liao [2023] exhibited a way to compute such an optimal regularization parameter by

solving a semidefinite program. This was extended to arbitrary (linear) Γ and T in [Foucart et al., 2023] and further simplified in [Foucart and Liao, 2024a], which also covered the case of an arbitrary (linear) U. In short, a globally optimal recovery map is given by $\Gamma \circ \Delta^{\sigma^\sharp}$, where the optimal parameter is $\sigma^\sharp = b^\sharp/(a^\sharp + b^\sharp)$ with

$$(a^\sharp, b^\sharp) = \underset{a,b \geq 0}{\operatorname{argmin}} \ a + b \quad \text{s.to} \ \ a\|Tf\|^2 + b\|U\Lambda f\|^2 \geq \|\Gamma f\|^2 \ \ \forall f \in F. \tag{19}$$

When the Hilbert space F is of finite dimension, the above can be solved as a semidefinite program, since the constraint reformulates as

$$a\, T^*T + b\, \Lambda^* U^* U \Lambda \succeq \Gamma^* \Gamma.$$

Note that this semidefinite program holds the value of the minimal global worst-case error, namely

$$\min_{\Delta} \operatorname{gwce}(\Delta)^2 = \min_{a,b \geq 0} \ a + b \quad \text{s.to} \ a\, T^*T + b\, \Lambda^* U^* U \Lambda \succeq \Gamma^* \Gamma. \tag{20}$$

In the particular situation where $T = \varepsilon^{-1} P_{V^\perp}$, $U = \eta^{-1}\mathrm{Id}$, and $\Lambda\Lambda^* = \mathrm{Id}$—which occurs in our tensorial situation—any $\sigma \in [0,1]$ actually leads to optimality when $\Gamma = \mathrm{Id}$, i.e., for full recovery, the global worst-case error of Δ^σ is independent of σ, as shown in [Foucart and Liao, 2023], so there is no need to solve any semidefinite program. Although this fact does not generalize to arbitrary quantities of interest, the next subsection shall uncover other natural Γ displaying this peculiarity.

Another bonus of the particular situation where $T = \varepsilon^{-1} P_{V^\perp}$, $U = \eta^{-1}\mathrm{Id}$, and $\Lambda\Lambda^* = \mathrm{Id}$ is a nice expression of the regularizer f_y^σ. Namely, with the notation already used in (16), we have

$$f_y^\sigma = \tau \sum_{i=1}^m a_i u_i + \sum_{j=1}^k b_j v_j, \tag{21}$$

where $\tau \in [0,1]$ depends on $\sigma \in [0,1]$ according to the correspondence

$$\tau = \frac{\sigma\varepsilon^2}{(1-\sigma)\eta^2 + \sigma\varepsilon^2} \quad \longleftrightarrow \quad \sigma = \frac{\tau\eta^2}{(1-\tau)\varepsilon^2 + \tau\eta^2}. \tag{22}$$

As a matter of fact, it was established in [Foucart and Liao, 2023, Proposition 3 and Appendix] that the right-hand side of (21) equals the minimizer of $(1-\tau)\|P_{V^\perp} f\|^2 + \tau\|y - \Lambda f\|^2$, which coincides with f_y^σ under the correspondence (22) when $T = \varepsilon^{-1} P_{V^\perp}$ and $U = \eta^{-1}\mathrm{Id}$.

Minimized local worst-case error. The *local* setting is the most delicate one, only solved for the particular situation $\Gamma = \mathrm{Id}$, $T = \varepsilon^{-1} P_{V^\perp}$, $U = \eta^{-1}\mathrm{Id}$, and $\Lambda\Lambda^* = \mathrm{Id}$ in [Foucart and Liao, 2023]—fortuitously, these restrictions hold in our tensor completion problem. Fixing y, Beck and

Eldar [2007] exhibited, when[4] $\Gamma = \mathrm{Id}$, a 'candidate' Chebyshev center by uncovering an upper bound for $\mathrm{lwce}_y(g)^2$ and then minimizing over all g. This is in the spirit of (13), except that the S-lemma is replaced by the S-procedure, which is not guaranteed to be exact, so the relaxation only provides an upper bound. Such a process was somewhat duplicated in Theorem 2 of [Foucart and Liao, 2024b] for an arbitrary (linear) Γ, leading to

$$\min_g \mathrm{lwce}_y(g)^2 \le \min_{\substack{a,b\ge 0\\ t\in\mathbb{R}}} a + b - b\|Uy\|^2 + t \qquad \text{s.to } aT^*T + b\Lambda^*U^*U\Lambda \succeq \Gamma^*\Gamma$$
$$\text{and } \left[\begin{array}{c|c} aT^*T + b\Lambda^*U^*U\Lambda & b\Lambda^*U^*Uy \\ \hline b(\Lambda^*U^*Uy)^* & t \end{array}\right] \succeq 0,$$

while the minimizers $a^\sharp, b^\sharp \ge 0$ yield the candidate Chebsyshev center $g^{\mathrm{cand}} = \Gamma(f_y^{\sigma^\sharp})$, with $f_y^{\sigma^\sharp}$ still denoting the solution to the regularization problem from (18) with parameter $\sigma^\sharp = b^\sharp/(a^\sharp + b^\sharp)$. The article [Foucart and Liao, 2024b] also indicated that this candidate Chebyshev center agrees with the true Chebyshev center determined in [Foucart and Liao, 2023] when $\Gamma = \mathrm{Id}$, $T = \varepsilon^{-1}P_V^\perp$, $U = \eta^{-1}\mathrm{Id}$, and $\Lambda\Lambda^* = \mathrm{Id}$. For arbitrary Γ, T, U, and Λ, Foucart and Hengartner [2025] later showed (see Theorem 5) that the candidate Chebyshev center coincides with the true Chebyshev center if and only if the orthogonality condition $\langle Tf_y^{\sigma^\sharp}, Th_{\sigma^\sharp}\rangle = 0$ hold, where h_σ generically denotes a leading eigenvector of $[(1-\sigma)T^*T+\sigma\Lambda^*U^*U\Lambda]^{-1/2}\Gamma^*\Gamma[(1-\sigma)T^*T+\sigma\Lambda^*U^*U\Lambda]^{-1/2}$. Incidentally, these conditions were precisely the ones used as sufficient conditions in [Foucart and Liao, 2023] (see the proof of Theorem 8 there) to determine the true Chebyshev center in the specific situation where $\Gamma = \mathrm{Id}$, $T = \varepsilon^{-1}P_{V^\perp}$, $U = \eta^{-1}\mathrm{Id}$, and $\Lambda\Lambda^* = \mathrm{Id}$.

All in all, in this specific situation, the true Chebyshev center can be computed as a regularizer $f_y^{\sigma^\sharp}$ for some parameter $\sigma^\sharp \in [0,1]$ than can be determined by solving a semidefinite program. But resorting to semidefinite programming is unnecessary: working with the alternative parametrization (22), it was established in [Foucart and Liao, 2023] that the optimal $\tau^\sharp \in [0,1]$ is the unique solution between $1/2$ and $\varepsilon/(\varepsilon+\eta)$ to the equation

$$\mathrm{eig}_{\min}((1-\tau)P_{V^\perp} + \tau\Lambda^*\Lambda) = \frac{(1-\tau)^2\varepsilon^2 - \tau^2\eta^2}{(1-\tau)\varepsilon^2 - \tau\eta^2 + (1-\tau)\tau(1-2\tau)\delta^2}, \tag{23}$$

where δ is precomputed as the common (when $\Lambda\Lambda^* = \mathrm{Id}$) value of $\min\{\|P_{V^\perp}f\| : \Lambda f = y\}$ and of $\min\{\|\Lambda f - y\| : f \in V\}$. This equation can be—and was—solved efficiently via Newton method, since the derivatives of both sides are available. It was indeed uncovered in the appendix of [Foucart and Liao, 2023] that $e(\tau) := \mathrm{eig}_{\min}((1-\tau)P_{V^\perp} + \tau\Lambda^*\Lambda)$ satisfies

$$\frac{de(\tau)}{d\tau} = \frac{1-2\tau}{\tau(1-\tau)}\frac{e(\tau)(1-e(\tau))}{1-2e(\tau)}.$$

As such, a possibly expensive eigencomputation is, at first sight, needed at each Newton iteration. This downside shall be removed in the next subsection.

[4]Technically speaking, the restriction $U = \eta^{-1}\mathrm{Id}$ was also in place, but the case of an arbitrary U can always be reduced to the case $U = \mathrm{Id}$ in the local setting, since the set $\mathcal{S}_U = \{f \in F : \|Tf\| \le 1, \|U(y - \Lambda f)\| \le 1\}$ of model- and data-consistent objects can be interpreted as $\mathcal{S}_{\mathrm{Id}}$ by replacing y by Uy and Λ by $U\Lambda$.

After this step, the value of the (squared) Chebyshev center becomes available, too, in the particular situation. It was indeed revealed in [Foucart and Liao, 2023, last remark before Section 4], and more generally in [Foucart and Hengartner, 2025], that

$$\min_g \mathrm{lwce}_y(g)^2 = \frac{(1-\tau^\sharp)\varepsilon^2 + \tau^\sharp\eta^2 - (1-\tau^\sharp)\tau^\sharp\delta^2}{e(\tau^\sharp)} = \frac{(1-\tau^\sharp)(\varepsilon^2 - \|P_{V^\perp} f^\sharp\|^2) + \tau^\sharp(\eta^2 - \|y - \Lambda f^\sharp\|^2)}{e(\tau^\sharp)}, \tag{24}$$

where the second expression is closer in spirit to (15).

4.2 Two novel results

The results presented in this section are both proved under the assumptions that T is a multiple of an orthogonal projector, that U is a multiple of the identity, and that Λ satisfies $\Lambda\Lambda^* = \mathrm{Id}$, which means that $\Lambda^*\Lambda$ is an orthogonal projector from F onto $\mathrm{range}(\Lambda^*)$. These assumptions are met in our tensor completion problem, where $T = \varepsilon^{-1} P_{V_s^\perp}$ and $U = \eta^{-1}\mathrm{Id}$. Moreover, it is readily checked that $\Lambda\Lambda^* = \mathrm{Id}$ because the observation map $\Lambda : \mathbb{R}^{n_1\times\cdots\times n_d} \to \mathbb{R}^{\mathcal{O}}$ for which $\Lambda(X) = X_{|\mathcal{O}}$ has adjoint $\Lambda^* : \mathbb{R}^{\mathcal{O}} \to \mathbb{R}^{n_1\times\cdots\times n_d}$ given by $\Lambda^*(y) = \sum_{\mathbf{i}\in\mathcal{O}} y_{\mathbf{i}} E_{\mathbf{i}}$, where $E_{\mathbf{i}} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ is the canonical tensor equal to one on the $\mathbf{i}$th entry and to zero on all other entries. As mentioned in the previous subsection, under these assumptions, every regularization map Δ^σ, $\sigma \in [0,1]$, is globally optimal for the recovery of $\Gamma = \mathrm{Id}$. We provide below a much simpler proof than the original one from [Foucart and Liao, 2023] and, in the process, we uncover a similar result for further quantities of interest.

Theorem 6. Given the model and uncertainty sets $\{f \in F : \|P_{V^\perp} f\| \le \varepsilon\}$ and $\{e \in \ell_2^m : \|e\| \le \eta\}$, if $\Lambda\Lambda^* = \mathrm{Id}$, then, when $\mathrm{range}(\Gamma^*) \subseteq \ker(\Lambda)$ or when $\Gamma = \mathrm{Id}$, one has

$$\mathrm{gwce}(\Gamma \circ \Delta^\sigma) = \inf_\Delta \mathrm{gwce}(\Delta) \qquad \text{for all } \sigma \in (0,1),$$

i.e., all regularization maps (18), when composed with Γ, produce globally optimal recovery maps.

Proof. For now, let Γ be an arbitrary (linear) quantity of interest. According to (17) and (20), with the change $c = a\varepsilon^{-2}$ and $d = b\eta^{-2}$, both $\mathrm{gwce}(\Gamma \circ \Delta^\sigma)^2$ and $\inf_\Delta \mathrm{gwce}(\Delta)^2$ are mininizers of $c\varepsilon^2 + d\eta^2$ subject to, respectively,

$$c\|P_{V^\perp} f\|^2 + d\|e\|^2 \ge \|\Gamma(f - \Delta^\sigma(\Lambda f + e))\|^2 \qquad \forall f \in F, e \in \mathbb{R}^m, \tag{25}$$

$$c\|P_{V^\perp} f\|^2 + d\|\Lambda f\|^2 \ge \|\Gamma f\|^2 \qquad \forall f \in F. \tag{26}$$

Thus, it is enough to show that the constraints (25) and (26) are equivalent. Since it is clear that (25) implies (26) by taking $e = -\Lambda f$, it remains to show that (26) implies (25). So let us assume that (26) holds in the form

$$c\|P_{V^\perp} f\|^2 + d\|P_{W^\perp} f\|^2 \ge \|\Gamma f\|^2 \qquad \forall f \in F, \tag{27}$$

where $P_{W^\perp} = \Lambda^*\Lambda$ is the orthogonal projector onto $\mathrm{range}(\Lambda^*) = W^\perp$, $W := \ker(\Lambda)$. Let us now consider $f \in F$ and $e \in \mathbb{R}^m$. Introducing $g := f - \Delta^\sigma(\Lambda f + e)$, we need to prove that

$$c\|P_{V^\perp} f\|^2 + d\|e\|^2 \overset{?}{\geq} \|\Gamma g\|^2. \tag{28}$$

Note that, in view of $\|e\|^2 = \|\Lambda^* e\|^2$, the left-hand side is

$$c\|P_{V^\perp} f\|^2 + d\|e\|^2 = \alpha + \beta, \qquad \text{where } \begin{cases} \alpha = c\|P_{V+W} \quad P_{V^\perp} f\|^2 + d\|P_{V+W} \quad \Lambda^* e\|^2, \\ \beta = c\|P_{(V+W)^\perp} P_{V^\perp} f\|^2 + d\|P_{(V+W)^\perp} \Lambda^* e\|^2. \end{cases}$$

Because we can express the regularization map as[5] $\Delta^\sigma = [(1-\sigma)P_{V^\perp} + \sigma P_{W^\perp}]^{-1}\sigma\Lambda^*$, we have

$$\begin{aligned} g &= [(1-\sigma)P_{V^\perp} + \sigma P_{W^\perp}]^{-1}((1-\sigma)P_{V^\perp} f + \sigma P_{W^\perp} f - \sigma\Lambda^*(\Lambda f + e)) \\ &= [(1-\sigma)P_{V^\perp} + \sigma P_{W^\perp}]^{-1}((1-\sigma)P_{V^\perp} f - \sigma\Lambda^* e). \end{aligned}$$

This implies that $(1-\sigma)P_{V^\perp} g + \sigma P_{W^\perp} g = (1-\sigma)P_{V^\perp} f - \sigma\Lambda^* e$, and hence

$$(1-\sigma)(P_{V^\perp} g - P_{V^\perp} f) = -\sigma(\Lambda^* e + P_{W^\perp} g) \quad \text{belongs to } V^\perp \cap W^\perp = (V+W)^\perp. \tag{29}$$

Applying P_{V+W} to (29) leads to $P_{V+W}(P_{V^\perp} g - P_{V^\perp} f) = 0$ and $P_{V+W}(\Lambda^* e + P_{W^\perp} g) = 0$, so that

$$P_{V+W} P_{V^\perp} f = P_{V+W} P_{V^\perp} g \qquad \text{and} \qquad P_{V+W}\Lambda^* e = -P_{V+W} P_{W^\perp} g.$$

As a result, we obtain

$$\alpha = c\|P_{V+W} P_{V^\perp} g\|^2 + d\|P_{V+W} P_{W^\perp} g\|^2 = c\|P_{V^\perp} P_{V+W} g\|^2 + d\|P_{W^\perp} P_{V+W} g\|^2,$$

where the last step used the fact that orthogonal projectors $P_{Z'}$ and $P_{Z^\perp}$ commute for subspaces $Z \subseteq Z'$ (by virtue of $P_{Z^\perp} = \mathrm{Id} - P_Z$ and of $P_{Z'} P_Z = P_Z P_{Z'} = P_Z$). Therefore, as a consequence of the constraint (26) expressed as in (27), we arrive at

$$\alpha \geq \|\Gamma P_{V+W} g\|^2.$$

This is enough to brush aside the case $\mathrm{range}(\Gamma^*) \subseteq \ker(\Lambda)$, as then $\mathrm{range}(\Gamma^*) \subseteq V + W$ yields $P_{V+W}\Gamma^* = \Gamma^*$, or equivalently $\Gamma P_{V+W} = \Gamma$, so that the previous inequality reads $\alpha \geq \|\Gamma g\|^2$ and the inequality $\alpha + \beta \geq \|\Gamma g\|^2$ requested in (28) immediately follows.

For the quantity of interest $\Gamma = \mathrm{Id}$, in addition to $\alpha \geq \|P_{V+W} g\|^2$, we shall also establish that $\beta \geq \|P_{(V+W)^\perp} g\|^2$, yielding $\alpha + \beta \geq \|g\|^2$ to validate (28). To this end, we apply $P_{(V+W)^\perp}$ to (29), which leads to, in view of $(V+W)^\perp \subseteq V^\perp$ and $(V+W)^\perp \subseteq W^\perp$ and after some rearrangement,

$$P_{(V+W)^\perp} g = (1-\sigma)P_{(V+W)^\perp} f - \sigma P_{(V+W)^\perp}\Lambda^* e.$$

[5] The existence of the inverse owes to the running assumption (6) that $V \cap W = \{0\}$, which, in infinite dimensions, must be strengthened to the existence of some $\delta > 0$ such that $\max\{\|P_{V^\perp} f\|, \|P_{W^\perp} f\|\} \geq \delta\|f\|$ for all $f \in F$.

By convexity, we deduce that

$$\|P_{(V+W)^\perp} g\|^2 \le (1-\sigma)\|P_{(V+W)^\perp} f\|^2 + \sigma \|P_{(V+W)^\perp}\Lambda^* e\|^2.$$

Notice that the constraint (27) with $\Gamma = \mathrm{Id}$ forces $c, d \ge 1$ by taking some nonzero $f \in W$ and $f \in V$. Consequently, we arrive at

$$\|P_{(V+W)^\perp} g\|^2 \le c\|P_{(V+W)^\perp} f\|^2 + d\|P_{(V+W)^\perp}\Lambda^* e\|^2 = c\|P_{(V+W)^\perp} P_{V^\perp} f\|^2 + d\|P_{(V+W)^\perp}\Lambda^* e\|^2 = \beta.$$

As this was the last inequality to establish, the proof is now complete. □

In our tensor completion problem, suitable quantities of interest include restrictions to unobserved entries, i.e., cases for which $\Gamma(X) = X_{|\mathcal{U}}$ with $\mathcal{U} \subseteq \mathcal{O}^c$, as then $\mathrm{range}(\Gamma^*) = \mathrm{span}\{E_{\mathbf{i}}, \mathbf{i} \in \mathcal{U}\}$ is clearly included in $\ker(\Lambda)$. In these cases, there is even a more direct argument exploiting (21). When written as $\Delta^\sigma(y) = \tau \sum_{\mathbf{i}\in\mathcal{O}} a_i E_{\mathbf{i}} + \sum_{j=1}^{k_s} b_j v_j$, this identity yields $\Gamma(\Delta^\sigma(y)) = \sum_{j=1}^{k_s} b_j v_j$ independently of σ, and hence $\mathrm{gwce}(\Gamma \circ \Delta^\sigma) = \sup\{\|\Gamma(f) - \Gamma(\Delta^\sigma(\Lambda f + e))\| : \|P_{V_s^\perp} f\| \le \varepsilon,\ \|e\| \le \eta\}$ is independent of σ, too.

For the second novel result, we transition from the global setting to the local setting and provide a more explicit solution to the full recovery problem in the situation where $T = \varepsilon^{-1} P_{V^\perp}$, $U = \eta^{-1}\mathrm{Id}$, and $\Lambda\Lambda^* = \mathrm{Id}$.

Theorem 7. Given the model and uncertainty sets $\{f \in F : \|P_{V^\perp} f\| \le \varepsilon\}$ and $\{e \in \ell_2^m : \|e\| \le \eta\}$, if $\Lambda\Lambda^* = \mathrm{Id}$, then the locally optimal recovery map for $\Gamma = \mathrm{Id}$ is given by

$$y \in \mathbb{R}^m \mapsto \left[\operatorname*{argmin}_{f\in F} \ (1-\tau^\sharp)\|P_{V^\perp} f\|^2 + \tau^\sharp \|y - \Lambda f\|^2\right] \in F,$$

where the regularization parameter $\tau^\sharp$, depending on y, is the unique solution between $1/2$ and $\varepsilon/(\varepsilon+\eta)$ to the sextic equation

$$c\left((1-\tau)\varepsilon^2 - \tau\eta^2 + (1-\tau)\tau(1-2\tau)\delta^2\right)^2 - (\varepsilon^2 - \eta^2 + (1-2\tau)\delta^2)((1-\tau)^2\varepsilon^2 - \tau^2\eta^2) = 0, \tag{30}$$

with $c := \mathrm{eig}_{\min}(P_{V^\perp} + \Lambda^*\Lambda)(2 - \mathrm{eig}_{\min}(P_{V^\perp} + \Lambda^*\Lambda)$ and $\delta := \min_{\Lambda f = y} \|P_{V^\perp} f\| = \min_{f\in V} \|\Lambda f - y\|$.

Proof. As recalled in the previous subsection, the locally optimal recovery map is $y \in \mathbb{R}^m \mapsto \Delta^{\sigma^\sharp} y$, where the regularization parameter $\sigma^\sharp$ corresponds via (22) to the unique solution $\tau^\sharp$ between $1/2$ and $\varepsilon/(\varepsilon+\eta)$ to

$$e(\tau) = r(\tau), \qquad \text{where} \quad r(\tau) := \frac{(1-\tau)^2\varepsilon^2 - \tau^2\eta^2}{(1-\tau)\varepsilon^2 - \tau\eta^2 + (1-\tau)\tau(1-2\tau)\delta^2}.$$

Here $e(\tau) := \mathrm{eig}_{\min}((1-\tau)P_{V^\perp} + \tau\Lambda^*\Lambda)$ satisfies the differential equation

$$\frac{de(\tau)}{d\tau} = \frac{1-2\tau}{\tau(1-\tau)} \frac{e(\tau)(1-e(\tau))}{1-2e(\tau)}. \tag{31}$$

The original article [Foucart and Liao, 2023] failed to notice, however, that this differential equation could be solved somewhat explicitly. To this end, consider the function of $\tau \in [0,1]$ defined by

$$C(\tau) = \frac{e(\tau)(1-e(\tau))}{\tau(1-\tau)} = \frac{e(\tau)-e(\tau)^2}{\tau-\tau^2},$$

whose derivative, thanks to (31), becomes

$$\frac{dC(\tau)}{d\tau} = \frac{1-2e(\tau)}{\tau(1-\tau)}\frac{de(\tau)}{d\tau} - \frac{e(\tau)(1-e(\tau))}{(\tau(1-\tau))^2}(1-2\tau) = 0.$$

This implies that the function C is constant, equal to $c := C(1/2)$, say, and we deduce that

$$e(\tau)(1-e(\tau)) = c\,\tau(1-\tau) \qquad \text{for all } \tau \in [0,1].$$

On the one hand, since the optimal parameter $\tau^\sharp$, lying between $1/2$ and $\varepsilon/(\varepsilon+\eta)$, must satisfy $e(\tau) = r(\tau)$ and in turn $r(\tau)(1-r(\tau)) = c\,\tau(1-\tau)$, we derive, after expressing $1-r(\tau)$ as a ratio, that $\tau^\sharp$ must be a solution to

$$\frac{\tau(1-\tau)(\varepsilon^2-\eta^2+(1-2\tau)\delta^2)((1-\tau)^2\varepsilon^2-\tau^2\eta^2)}{\big((1-\tau)\varepsilon^2-\tau\eta^2+(1-\tau)\tau(1-2\tau)\delta^2\big)^2} = c\,\tau(1-\tau),$$

which is readily equivalent to the announced sextic equation.

On the other hand, let τ be a solution to this sextic equation lying between $1/2$ and $\varepsilon/(\varepsilon+\eta)$. Reverting the above, we obtain that $r(\tau)(1-r(\tau)) = c\,\tau(1-\tau)$, and since $e(\tau)(1-e(\tau)) = c\,\tau(1-\tau)$, we deduce that $e(\tau)(1-e(\tau)) = r(\tau)(1-r(\tau))$. This implies that $e(\tau) = r(\tau)$—which is what we need for $\tau = \tau^\sharp$ to follow—or $e(\tau) = 1-r(\tau)$—which we are aiming to invalidate. To this end, it is useful to recall from [Foucart and Liao, 2023, Proof of Theorem 8] that the denominator $d(\tau)$ of $1-r(\tau)$ does not vanish between $1/2$ and $\varepsilon/(\varepsilon+\eta)$, that $\delta \le \varepsilon+\eta$, and that $e(\tau) \in [0,1/2]$. By contradiction, assume now that $e(\tau) = 1-r(\tau) \neq r(\tau)$, so that $1-r(\tau) \in [0,1/2)$, and consider e.g. the case $\varepsilon \ge \eta$. Since then $1/2 \le \tau \le \varepsilon/(\varepsilon+\eta)$, the numerator $n(\tau)$ of $1-r(\tau)$ satisfies

$$\frac{n(\tau)}{\tau(1-\tau)} = \varepsilon^2-\eta^2+(1-2\tau)\delta^2 \ge \varepsilon^2-\eta^2+(1-2\tau)(\varepsilon+\eta)^2 \ge \varepsilon^2-\eta^2+\Big(1-\frac{2\varepsilon}{\varepsilon+\eta}\Big)(\varepsilon+\eta)^2 = 0,$$

i.e., its is nonnegative, hence the denominator $d(\tau)$ must be positive. Therefore, the inequality $1-r(\tau) < 1/2$ yields $2n(\tau)-d(\tau) < 0$. However, after a simplification step, we have

$$\begin{aligned}
2n(\tau)-d(\tau) &= (2\tau-1)[(1-\tau)\varepsilon^2+\tau\eta^2-\tau(1-\tau)\delta^2] \\
&\ge (2\tau-1)[(1-\tau)\varepsilon^2+\tau\eta^2-\tau(1-\tau)(\varepsilon^2+\eta^2+2\varepsilon\eta)] \\
&= (2\tau-1)[(1-\tau)^2\varepsilon^2+\tau^2\eta^2-2\tau(1-\tau)\varepsilon\eta] = (2\tau-1)[(1-\tau)\varepsilon-\tau\eta]^2 \\
&\ge 0.
\end{aligned}$$

This provides the contradiction desired to finish the proof. □

4.3 Tensor computations

As in the accurate scenario, there are satisfying solutions to the abstract Optimal Recovery problem in the inaccurate scenario, but further challenges occur for their computational implementations due to the high dimensionality of the tensor completion problem. We discuss our resolutions below.

Optimal recovery maps. In both the global and local settings, the optimal recovery maps are given by regularization. When $T = \varepsilon^{-1} P_{V^\perp}$ and $U = \eta^{-1}\mathrm{Id}$, they take the form

$$\Delta_\tau(y) = f_{y,\tau} = \operatorname*{argmin}_{f \in F} \ (1-\tau)\|P_{V^\perp} f\|^2 + \tau \|y - \Lambda f\|^2.$$

Beware of the difference with f_y^σ defined in (18)—in fact, we have $f_{y,\tau} = f_y^\sigma$ under the correspondence (22). Either way, an unconstrained least-squares problem needs to be solved. Using τ, the normal equations read $A_\tau f_{y,\tau} = \tau \Lambda^* y$, where $A_\tau := (1-\tau) P_{V^\perp} + \tau \Lambda^* \Lambda$ is a positive semidefinite operator—actually, positive definite thanks to our standing assumption (6). This framework is favorable for conjugate gradient, especially since the action of A_τ on a tensor X splits onto two computationally cheap, matrix-free actions, namely: $P_{V^\perp} X = (\mathrm{Id} - \Pi_{\le s}) X$ is computed via (8) in $\mathcal{O}(n^d \ln(n^d))$ operations and $\Lambda^* \Lambda X$, obtained from X by zeroing out the entries outside of $\mathcal{O}$, is computed in $\mathcal{O}(m)$ operations. We do not spell out the conjugate gradient iterations adapted to our case, but we mention that each iteration involves two applications of A_τ and that the solution $f_{y,\tau}$ is achieved in a finite number $t_{\max}$ of iterations. It is guaranteed that $t_{\max} \le N = n_1 \cdots n_d$, but $t_{\max}$ was typically much smaller in our experiments which imposed a stopping criterion based on residual. All in all, the total operation cost for computing the regularizer $f_{y,\tau}$ via conjugate gradient is $\mathcal{O}(t_{\max} n^d \ln(n^d))$.

We note that $f_{y,\tau}$, as the minimizer of $\|P_{V^\perp} f\|^2 + (\tau/(1-\tau))\|y - \Lambda f\|^2$, intuitively tend to the minimizer f_y of $\|P_{V^\perp} f\|^2$ subject to $\Lambda f = y$, see [Foucart and Liao, 2024a, Appendix] for a proper justification of this statement. The point here is that the Chebyshev center in the accurate scenario, i.e. f_y, can be approximately computed via conjugate gradient, too. In the particular case where $T = \varepsilon^{-1} P_{V^\perp}$, $U = \eta^{-1}\mathrm{Id}$, and $\Lambda\Lambda^* = \mathrm{Id}$, it even occurs that $f_y = f_{y,0}$ can be computed *exactly* via two conjugate gradient calls. Indeed, recalling the affine dependence (21) of $f_{y,\tau}$, written as $f_{y,\tau} = f_{y,0} + \tau(f_{y,1} - f_{y,0})$, we can deduce $f_y = f_{y,0}$ from $f_{y,\tau}$ computed at two values of $\tau \in (0,1)$, e.g. $f_y = 2 f_{y,1/3} - f_{y,2/3}$.

To finally compute the optimal recovery maps, now that we are able to compute each regularizer $f_{y,\tau}$, we need to determine the optimal parameter $\tau^\sharp$. In the *global* setting, this is generally done by solving the semidefinite program (20), yet solving it naively will be inefficient due to the high dimensionality of our tensor completion problem. Fortuitously, the assumptions $T = \varepsilon^{-1} P_{V^\perp}$, $U = \eta^{-1}\mathrm{Id}$, and $\Lambda\Lambda^* = \mathrm{Id}$ hold for this problem, so any $\tau \in [0,1]$ leads to a globally optimal recovery map when $\Gamma(X) = X_{|\mathcal{U}}, \mathcal{U} \subseteq \mathcal{O}^c$, and when $\Gamma(X) = X$, as shown in Theorem 6. In the *local* setting,

again thanks to the assumptions $T = \varepsilon^{-1} P_{V^\perp}$, $U = \eta^{-1}\mathrm{Id}$, and $\Lambda\Lambda^* = \mathrm{Id}$ being met, Theorem 7 establishes that the optimal parameter is obtained by solving the sextic equation (30). To do so, we first need to compute δ and c. The quantity $\delta = \min\{\|P_{V^\perp} f\| : \Lambda f = y\} = \min\{\|\Lambda f - y\| : f \in V\}$ is available after the computation of any $f_{y,\tau}$, according to $\delta = \|P_{V^\perp} f_{y,\tau}\|/\tau = \|\Lambda f_{y,\tau} - y\|/(1-\tau)$, see [Foucart and Liao, 2023, Proposition 3]. The quantity $c = \mathrm{eig}_{\min}(P_{V^\perp} + \Lambda^*\Lambda)(2 - \mathrm{eig}_{\min}(P_{V^\perp} + \Lambda^*\Lambda)$ shall by determined by performing a power method only once. For comparison, before Theorem 7, the determination of the optimal parameter $\tau^\sharp$ necessitated one eigenanalysis (power method or else) at each Newton iteration. To compute c, it is enough to compute

$$\mathrm{eig}_{\min}(P_{V^\perp} + \Lambda^*\Lambda) = 2 - \mathrm{eig}_{\max}(B), \qquad \text{where} \quad B := 2\mathrm{Id} - P_{V^\perp} - \Lambda^*\Lambda \succeq 0.$$

The standard power method to determine the largest absolute eigenvalue of B, also applicable to matrices that are not positive semidefinite, would produce the 'powers' $x^{(t)} = B^t x^{(0)}$ of an arbitrary initial vector $x^{(0)}$ and approximate $\mathrm{eig}_{\max}(B)$ by $|\langle Ax^{(t)}, x^{(t)}\rangle|/\|x^{(t)}\|^2$. We modify this strategy slightly by averaging over several random initial vectors in order to provide a PAC (short for probably approximately correct) guarantee of accuracy with a number of iterations known beforehand. We emphasize that each iteration $x^{(t)} = Bx^{(t-1)}$ still exploits the fact that the action of B again splits as two computationally cheap, matrix-free actions. The result we rely on is stated and proved below, although it is likely to have been expressed in a resembling form elsewhere before.

Proposition 8. Let $B \in \mathbb{R}^{N\times N}$ be positive semidefinite matrix with largest eigenvalue $\lambda_{\max}$. Given integers $k, t \geq 0$, consider independent standard gaussian vectors $g^{(1)}, \dots, g^{(k)} \in \mathbb{R}^N$ and form the computable randomized estimator

$$\widehat{\lambda}_{k,t} := \left[\frac{5}{k}\sum_{j=1}^{k} \left\langle B^t g^{(j)}, g^{(j)}\right\rangle\right]^{1/t}.$$

Given two small parameters $\iota, \theta > 0$, as soon as $k \geq 25 \ln\left(\frac{2}{\theta}\right)$ and $t \geq \dfrac{\ln(9N)}{\ln(1+\iota)}$, it holds with probability at least $1 - \theta$ that

$$\lambda_{\max} \leq \widehat{\lambda}_{k,t} \leq (1+\iota)\lambda_{\max}.$$

Proof. Write the eigendecomposition of B as

$$B = \sum_{i=1}^{N} \lambda_i v_i v_i^*, \quad \text{with } \lambda_{\max} = \lambda_1 \geq \cdots \geq \lambda_N \geq 0 \text{ and } (v_1, \dots, v_N) \text{ an orthonormal basis of } \mathbb{R}^N.$$

For each standard gaussian vector $g^{(j)}$, we have $g^{(j)} = \sum_{i=1}^N g_i^{(j)} v_i$, where all the $g_i^{(j)} = \langle g^{(j)}, v_i\rangle$ are independent standard gaussian variables. The random variable $\xi_{j,t} := \langle B^t g^{(j)}, g^{(j)}\rangle = \sum_{i=1}^N \lambda_i^t (g_i^{(j)})^2$ therefore has expectation

$$\mathbb{E}_t := \mathbb{E}\big[\langle B^t g^{(j)}, g^{(j)}\rangle\big] = \sum_{i=1}^{N} \lambda_i^t \begin{cases} \leq & N\lambda_{\max}^t, \\ \geq & \lambda_{\max}^t. \end{cases}$$

Based on the moment generating function of the square of a standard gaussian variable (see e.g. [Foucart, 2022, (B.5)]), the moment generating function of $\xi_{j,t}$ at $u > 0$ is

$$\begin{aligned}\mathbb{E}[\exp(u\xi_{j,t})] &= \prod_{i=1}^N \mathbb{E}[\exp(u\lambda_i^t (g_i^{(j)})^2)] = \prod_{i=1}^N \frac{1}{\sqrt{1-2u\lambda_i^t}} \\ &\le \prod_{i=1}^N \exp(u\lambda_i^t + 4u^2\lambda_i^{2t}) = \exp(u\mathbb{E}_t + 4u^2\mathbb{E}_{2t}) \\ &\le \exp(u\mathbb{E}_t + 4u^2\mathbb{E}_t^2).\end{aligned}$$

As a result, for any $\omega \in (0,1)$ and making the choice $u = \omega/(8\mathbb{E}_t)$, an application of Markov inequality yields

$$\begin{aligned}\mathbb{P}\Big[\frac{1}{k}\sum_{j=1}^k \big\langle B^t g^{(j)}, g^{(j)}\big\rangle > (1+\omega)\mathbb{E}_t\Big] &= \mathbb{P}\Big[\exp\Big(u\sum_{j=1}^k \xi_{j,t}\Big) > \exp(uk(1+\omega)\mathbb{E}_t)\Big] \\ &\le \frac{\mathbb{E}\big[\exp\big(u\sum_{j=1}^k \xi_{j,t}\big)\big]}{\exp(uk(1+\omega)\mathbb{E}_t)} = \prod_{j=1}^k \frac{\mathbb{E}[\exp(u\xi_{j,t})]}{\exp(u(1+\omega)\mathbb{E}_t)} \le \prod_{j=1}^k \frac{\exp(u\mathbb{E}_t + 4u^2\mathbb{E}_t^2)}{\exp(u(1+\omega)\mathbb{E}_t)} \\ &= \prod_{j=1}^k \exp(-u\omega\mathbb{E}_t + 4u^2\mathbb{E}_t^2) = \prod_{j=1}^k \exp\Big(-\frac{\omega^2}{16}\Big) = \exp\Big(-\frac{\omega^2 k}{16}\Big).\end{aligned}$$

In a similar fashion and with details left to the reader, we also derive

$$\mathbb{P}\Big[\frac{1}{k}\sum_{j=1}^k \big\langle B^t g^{(j)}, g^{(j)}\big\rangle < (1-\omega)\mathbb{E}_t\Big] \le \exp\Big(-\frac{\omega^2 k}{16}\Big).$$

This means that, with failure probability at most $2\exp(-\omega^2 k/16)$, we have

$$\frac{1}{k}\sum_{j=1}^k \big\langle B^t g^{(j)}, g^{(j)}\big\rangle \begin{cases} \le & (1+\omega)\mathbb{E}_t \le (1+\omega)N\lambda_{\max}^t, \\ \ge & (1-\omega)\mathbb{E}_t \ge (1-\omega)\lambda_{\max}^t, \end{cases}$$

which implies that

$$\lambda_{\max} \le \Big[\frac{1}{(1-\omega)k}\sum_{j=1}^k \big\langle B^t g^{(j)}, g^{(j)}\big\rangle\Big]^{1/t} \le \Big(\frac{1+\omega}{1-\omega}N\Big)^{1/t}\lambda_{\max}.$$

By choosing $\omega = 4/5$, we conclude that $\lambda_{\max} \le \widehat{\lambda}_{k,t} \le (9N)^{1/t}\lambda_{\max}$ with failure probability at most $2\exp(-k/25)$. We arrive at $\lambda_{\max} \le \widehat{\lambda}_{k,t} \le (1+\iota)\lambda_{\max}$ with failure probability at most θ when $k \ge 25\ln(2/\theta)$ and $t \ge \ln(9N)/\ln(1+\iota)$, as announced in Proposition 8. □

Minimal worst-case errors. In the general situation, computing the minimal *global* worst-case error via the semidefinite program (20) seems to be the only reliable grip to put our hands on.

But this approach is not suitable in the high-dimensional setting of tensors, unless we probe the inner workings of semidefinite solvers in search for saving opportunities—a task which we have not attempted. In the particular situation where $T = \varepsilon^{-1}P_{V^\perp}$, $U = \eta^{-1}\mathrm{Id}$, $\Lambda\Lambda^* = \mathrm{Id}$, and $\Gamma = \mathrm{Id}$, although we know that the minimal global worst-case error is achieved as $\mathrm{gwce}(\Delta^\sigma)$ for any the regularization parameter $\sigma \in [0, 1]$, the computation of each of these still materializes, at first sight, as the semidefinite program (17). Fortunately, it is also known that the minimal global worst-case error agrees with the minimal local worst-case error at $y = 0$ in the general situation, and since the latter can be efficiently computed at any y in the particular situation—see below—the minimal global worst-case error will ensue (see [Foucart and Liao, 2023, last remark of Section 4] for details).

Now, in the local setting and with $T = \varepsilon^{-1}P_{V^\perp}$, $U = \eta^{-1}\mathrm{Id}$, $\Lambda\Lambda^* = \mathrm{Id}$, and $\Gamma = \mathrm{Id}$, we have described ways to efficiently compute the regularizers $f_{y,\tau}$, as well as the optimal regularization parameters $\tau^\sharp$ (involving the precomputation of the quantity δ, which depends on y). The minimal eigenvalue $e(\tau^\sharp) = \mathrm{eig}_{\min}((1-\tau^\sharp)P_{V^\perp} + \tau^\sharp\Lambda^*\Lambda)$ follows by (23) and finally the minimal local worst-case error at any y is deduced via (24).

5 Numerical Experiments

In this section, we first verify the accuracy and computational advantage of the FFT-based ANOVA projection. Next, we compare global and local recoveries on synthetic data generated from Sobol's product function. Finally, we evaluate our recovery methods on real-world data: the Energy Efficiency dataset serves as a moderate-scale example, while the Theoretical Multinary Perovskite Oxides dataset provides a large-scale test of computational scalability. MATLAB files allowing one to retrieve our implementations and to reproduce this section's experiments are available from the repository `https://github.com/Jingchun-Shao/OR_ANOVA`.

5.1 Fast ANOVA via the fast Fourier transform

We compare our FFT-based ANOVA construction with the traditional recursive construction on ten random tensors $X \in \mathbb{R}^{n\times\cdots\times n}$. Here, with d denoting the order of the tensor, n the number of grid points in each dimension, and s the maximum ANOVA interaction order, the experiments cover the range $5 \le d \le 10$, $2 \le n \le 4$, and $1 \le s \le 4$.

The relative error between the two implementations varies from 3.09×10^{-16} to 2.19×10^{-15}, showing that they agree to machine precision. We then compare their computational efficiencies. Figure 1 compares running times as the order of the tensor increases from $d = 5$ to $d = 12$, with $n = 4$ and $s = 2$ being fixed. The FFT-based implementation is faster in every tested case. In addition, the

dominant computational scaling of n^d for both methods is reflected by the roughly linear trends seen with the logarithmic y-scale.

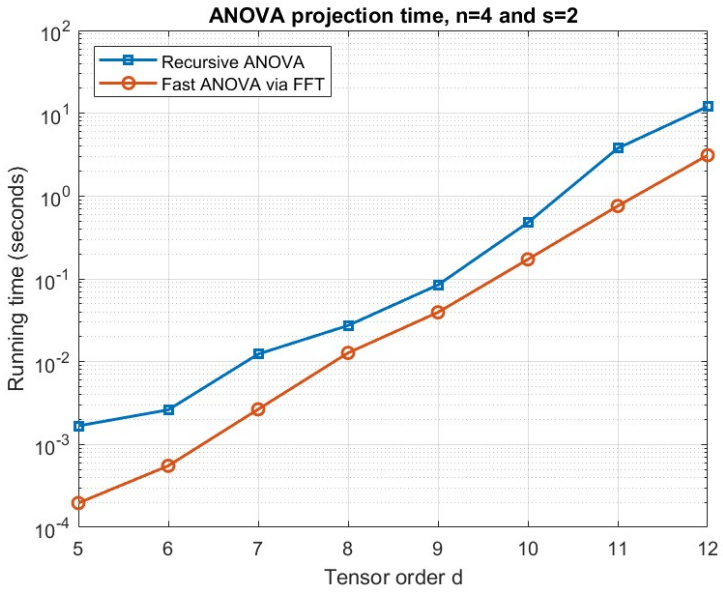

Figure 1: Running times of the FFT-based and recursive ANOVA projections

5.2 Global and local recoveries for Sobol's function

Motivated by Sobol's ANOVA-based sensitivity analysis [Sobol', 1993] and Owen's effective-dimension analysis [Owen, 2003], we consider Sobol's product function defined on $[0,1]^d$ by

$$x(t_1,\dots,t_d) = \prod_{j=1}^{d} \frac{|4t_j - 2| + a_j}{1 + a_j}.$$

We take $d = 8$, with $a_1 = a_2 = 0$ and $a_3 = \cdots = a_8 = 3$, so that the first two variables are more influential than the others six. Discretizing each variable at three points produces a tensor X with $N = 3^8 = 6561$ entries. For $s = 1, 2, 3, 4$, the relative ANOVA approximation errors $\|X - \Pi_{\le s} X\|/\|X\|$ are 33.248%, 10.218%, 2.2759%, and 0.3843%, respectively. We choose $s = 3$, yielding an approximation error of about 2.3% with model dimension $k_s = 577$.

The observations are taken at distinct entries sampled uniformly at random. Their number is $m = 1968$, making up 30% of the grid. The vector of observations is gaussian with $\|e\| = 0.01\|\Lambda X\|$. We choose the parameters ε and η as $\varepsilon = 1.1\|X - \Pi_{\le s} X\|$ and $\eta = 1.1\|e\|$. In view of $\Lambda\Lambda^* = \mathrm{Id}$ in this situation, the globally and locally optimal recovery procedures both return regularizers X_τ: in the global setting, any $\tau \in [0,1]$ leads to the same global worst-case error, so we choose $\tau = 0.5$;

in the local setting, the optimal parameter τ is found by solving the sextic equation. Note that finding it via the semidefinite formulation of Beck and Eldar [2007] was also attempted, but the ambient dimension of $N = 6,561$ was too large for the computation to run given our resources.

Table 1 reports the minimal local worst-case error computed as in (24), as well as the minimal global worst-case error computed as a minimal local worst-case error with $y = 0$, both of them presented in normalized form, i.e., after division by $\|X\| = 166.7529$. This is possible in this synthetic example where the ground truth is available. The local error is smaller than the global error, as should be. Table 1 also reports the pointwise, not worst-case, errors $\|X - X_\tau\|$ (full grid) and $\|X_{\mathcal{O}^c} - (X_\tau)_{\mathcal{O}^c}\|$ (unobserved grid), presented in relative forms, i.e., divided by $\|X\|$ and $\|X_{\mathcal{O}^c}\|$, respectively. We see that the locally optimal procedure slightly reduces the full-grid relative error, while the unobserved-grid relative error remains unchanged, confirming the fact explained after Theorem 6 that changing τ only affects the observed entries in this situation.

Method	Normalized wce	Full-grid relative error	Unobserved relative error
Global	11.999%	2.5439%	2.9301%
Local	11.174%	2.5177%	2.9301%

Table 1: Normalized worst-case errors and pointwise relative errors for Sobol's function with $s = 3$.

5.3 Global and local recoveries on the Energy Efficiency dataset

The Energy Efficiency dataset of [Tsanas and Xifara, 2012] comprises 768 building designs and two responses: the heating load Y_1 and the cooling load Y_2. The inputs $X_1, \ldots, X_5$ describe the geometry, X_6 the orientation, and X_7, X_8 the glazing area and its distribution. In this dataset, each value of X_1 actually determines the whole tuple $(X_1, \ldots, X_5)$, so that X_1 alone indexes geometry. We also treat each pair (X_7, X_8) as one glazing configuration, giving 16 glazing levels. In this way, the axes X_1, X_6, and the configurations (X_7, X_8) give rise to a $12 \times 4 \times 16$ tensor with 768 entries. We randomly split these entries into a training set of 538 observed entries and a testing set of 230 unobserved entries to be recovered.

For $s = 0, 1, 2$, the relative ANOVA approximation errors are computed, respectively, as 41.191%, 4.0922%, 1.2266% for Y_1, and 36.064%, 6.4641%, 4.8717% for Y_2. We retain $s = 1$ and $s = 2$ for our experiments, verifying in passing that the dimensions $k_1 = 30$ and $k_2 = 273$ do not exceed $m = 538$. As for the model parameters ε and η, we choose them in a heuristic way allowing us to deliberately avoid using the ground truth (even though it is available here). We then proceed as in Subsection 5.2 to determine the normalized minimal worst-case errors and the testing (i.e., pointwise) relative errors.

Table 2 once again confirms that the minimal local worst-case errors are smaller than their global counterparts and that the local and global testing errors coincide because changing τ only affects

the observed entries. It also shows that increasing s from 1 to 2 reduces the testing error for Y_1 from 3.98% to 2.48%—both being excellent—but increases it from 6.66% to 8.71% for Y_2. Thus, on this split, second-order recovery predicts Y_1 better, while first-order recovery predicts Y_2 better.

Target	s	Method	Normalized wce	Testing relative error
Y_1	1	Global	9.552%	3.9782%
Y_1	1	Local	8.452%	3.9782%
Y_1	2	Global	15.603%	2.4844%
Y_1	2	Local	15.368%	2.4844%
Y_2	1	Global	13.564%	6.6573%
Y_2	1	Local	11.628%	6.6573%
Y_2	2	Global	29.480%	8.7145%
Y_2	2	Local	26.851%	8.7145%

Table 2: Normalized worst-case errors and testing relative errors for the Energy Efficiency dataset.

5.4 Global recovery on the Multinary Perovskite Oxides dataset

Discovering functional materials is an important but difficult task: the number of candidate compositions grows rapidly, while evaluating each by first-principles calculations is costly. Bare, Morelock, and Musgrave [2023] report formation enthalpies for 66,516 multinary oxide compositions. Their dataset includes compositions with two cations on both sites, $A_{0.5}A'_{0.5}B_{0.5}B'_{0.5}O_3$, allowing $A = A'$ or $B = B'$ when one site contains only a single cation. Motivated by the successful application of HDMR to molecular system in [Li, Wang, Rosenthal, and Rabitz, 2001], we adopt a low-order ANOVA model to capture the effects of individual cation choices and their interactions.

We use the four cation identities A_1, A_2, B_1, B_2 as tensor coordinates, giving a $39 \times 38 \times 39 \times 38$ grid with $N = 2{,}196{,}324$ entries. Of the recorded compositions, 53,213 are used for training and 13,303 for testing. We recover the formation enthalpy via the globally optimal procedure by choosing $\tau = 0.5$ and $s = 1, 2$. The locally optimal procedure, which is still too costly in these dimensions, was not deployed.

Table 3 reports the testing relative errors and shows that including second-order interactions reduces them from 4.33% to 3.38%. Both low-order models achieve test-set relative errors below 5%, highlighting the potential of our method to support materials discovery.

s	Model dimension k_s	Testing relative error
1	151	4.330%
2	8588	3.379%

Table 3: Recovery of formation enthalpy for the Theoretical Multinary Perovskite Oxides dataset.

References

Z. J. L. Bare, R. J. Morelock, and C. B. Musgrave. Dataset of theoretical multinary perovskite oxides. *Scientific Data*, 10:244, 2023.

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

P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Data assimilation in reduced modeling. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):1–29, 2017.

A. Cohen, I. Daubechies, R. DeVore, G. Kerkyacharian, and D. Picard. Capturing ridge functions in high dimensions from point queries. *Constructive Approximation*, 35(2):225–243, 2012.

R. DeVore, G. Petrova, and P. Wojtaszczyk. Approximation of functions of few variables in high dimensions. *Constructive Approximation*, 33(1):125–143, 2011.

R. Drautz. Atomic cluster expansion for accurate and transferable interatomic potentials. *Physical Review B*, 99(1):014104, 2019. doi: 10.1103/PhysRevB.99.014104.

D. E. Dudgeon and R. M. Mersereau. *Multidimensional Digital Signal Processing*. Prentice-Hall Signal Processing Series. Prentice-Hall, Englewood Cliffs, NJ, 1984.

S. Foucart. *Mathematical Pictures at a Data Science Exhibition*. Cambridge University Press, 2022.

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

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

S. Foucart and C. Liao. Radius of information for two intersected centered hyperellipsoids and implications in optimal recovery from inaccurate data. *Journal of Complexity*, 83:101841, 2024a.

S. Foucart and C. Liao. 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*, pages 1–18. Springer, 2024b.

S. Foucart, C. Liao, S. Shahrampour, and Y. Wang. Learning from non-random data in Hilbert spaces: an optimal recovery perspective. *Sampling Theory, Signal Processing, and Data Analysis*, 20(1):5, 2022.

S. Foucart, C. Liao, and N. Veldt. On the optimal recovery of graph signals. In *2023 International Conference on Sampling Theory and Applications (SampTA)*, pages 1–5. IEEE, 2023.

M. Griebel. Sparse grids and related approximation schemes for higher dimensional problems. In *Foundations of Computational Mathematics, Santander 2005*, pages 106–161. Cambridge University Press, 2006.

C. J. Hillar and L.-H. Lim. Most tensor problems are NP-hard. *Journal of the ACM (JACM)*, 60 (6):1–39, 2013.

P. Jain and S. Oh. Provable tensor factorization with missing data. *Advances in Neural Information Processing Systems*, 27, 2014.

M. A. Kowalski, K. A. Sikorski, and F. Stenger. *Selected Topics in Approximation and Computation*. Oxford University Press, 1995.

D. Kressner, M. Steinlechner, and B. Vandereycken. Low-rank tensor completion by Riemannian optimization. *BIT Numerical Mathematics*, 54(2):447–468, 2014.

G. Li, S.-W. Wang, C. Rosenthal, and H. Rabitz. General foundations of high-dimensional model representations. *Journal of Mathematical Chemistry*, 30(1):1–30, 2001.

J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):208–220, 2013.

A. A. Melkman and C. A. Micchelli. Optimal estimation of linear operators in Hilbert spaces from inaccurate data. *SIAM Journal on Numerical Analysis*, 16(1):87–105, 1979.

A. B. Owen. The dimension distribution and quadrature test functions. *Statistica Sinica*, 13(1): 1–17, 2003.

I. H. Sloan and H. Woźniakowski. When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? *Journal of Complexity*, 14(1):1–33, 1998.

I. M. Sobol'. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1(4):407–414, 1993.

M. Steinlechner. Riemannian optimization for high-dimensional tensor completion. *SIAM Journal on Scientific Computing*, 38(5):S461–S484, 2016.

A. Takemura. Tensor analysis of ANOVA decomposition. *Journal of the American Statistical Association*, 78(384):894–900, 1983.

A. Tsanas and A. Xifara. Energy Efficiency [Dataset]. UC Irvine Machine Learning Repository, 2012. URL `https://archive.ics.uci.edu/ml/datasets/Energy+efficiency`.

A. R. Zhang. Cross: Efficient low-rank tensor completion. *The Annals of Statistics*, 47(2):936–964, 2019.