Beyond the Signs: Nonparametric Tensor Completion via Sign Series

Anonymous Authors[1]

Abstract

We consider the problem of tensor estimation from noisy observations with possibly missing entries. A nonparametric approach to tensor completion is developed based on a new model which we coin as sign representable tensors. The model represents the signal tensor of interest using a series of structured sign tensors. Unlike earlier methods, the sign series representation effectively addresses both low- and high-rank signals, while encompassing many existing tensor models—including CP models, Tucker models, single index models, certain hypergraphon models—as special cases. We show that the sign tensor series are theoretically characterized, and computationally estimable, by classification tasks with carefully-specified weights. Excess risk bounds, estimation error rates, and sample complexities are established. We demonstrate the outperformance of our approach over previous methods on two datasets, one on human brain connectivity networks and the other on author-topic mining.

1. Introduction

Higher-order tensors have recently received much attention in enormous fields including social networks (Anandkumar et al., 2014), neuroscience (Wang et al., 2017), and genomics (Wang et al., 2019). Tensor methods provide effective representation of the hidden structure in multiway data. In this paper we consider the signal plus noise model,

$$\mathcal{Y} = \Theta + \mathcal{E}, \tag{1}$$

where $\mathcal{Y} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is an order-K data tensor, Θ is an unknown signal tensor of interest, and $\mathcal{E}$ is the noise tensor. Our goal is to accurately estimate Θ from the incomplete, noisy observation of $\mathcal{Y}$. In particular, we focus on the following two problems:

- Q1 [Nonparametric tensor estimation]. How to flexibly estimate Θ under a wide range of structures, including both low-rankness and high-rankness?
- Q2 [Tensor completion]. How many observed tensor entries do we need to consistently estimate the signal Θ?

Inadequacies of common low-rank models. The signal plus noise model (3) is popular in tensor literature. Existing methods perform estimation based on low-rankness of Θ (Montanari and Sun, 2018; Cai et al., 2019). Common low-rank models include CP tensors (Anandkumar et al., 2014), Tucker tensors (De Lathauwer et al., 2000), and block tensors (Wang and Zeng, 2019). While these methods have shown great success in signal recovery, tensors sought in application often violate the low-rankness. Here we provide two examples to illustrate the limitation of classical models.

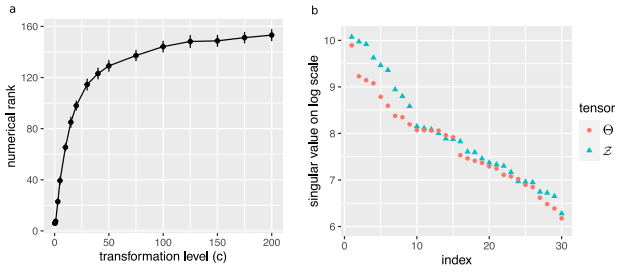


Figure 1. (a) Numerical rank of Θ versus c in the first example. (b) Top $d = 30$ tensor singular values in the second example.

The first example reveals the sensitivity of tensor rank to order-preserving transformations. Let $\mathcal{Z} \in \mathbb{R}^{30\times30\times30}$ be an order-3 tensor with CP rank$(\mathcal{Z}) = 3$ (formal definition is deferred to end of this section). Suppose a monotonic transformation $f(z) = (1 + \exp(-cz))^{-1}$ is applied to $\mathcal{Z}$ entrywise, and we let the signal Θ in model (1) be the tensor after transformation. Figure 1(a) plots the numerical rank of Θ versus c. As we see, the rank increases rapidly with c, rending traditional low-rank tensor methods ineffective even in the presence of mild order-preserving nonlinearities. In applications of digital processing (Karbasi and Oh, 2012) and genomics analysis (Wang et al., 2019), the tensor of interest often undergoes unknown transformation prior to measurements. The sensitivity to transformation makes the low-rank model less desirable in practice.

The second example demonstrates the inadequacy of classical low-rankness in representing important special struc-

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

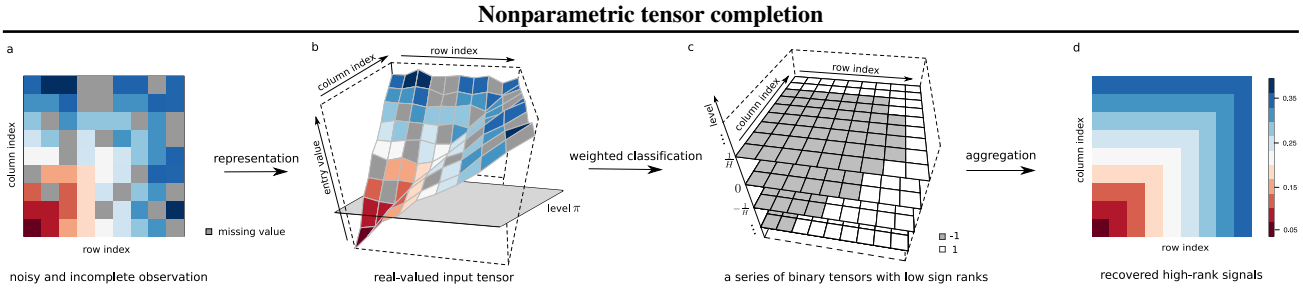


Figure 2. Illustration of our method. For visualization purpose, we plot an order-2 tensor (a.k.a. matrix); similar procedure applies to higher-order tensors. (a): noisy and incomplete tensor input. (b) and (c): main steps of estimating sign tensor series $\mathrm{sgn}(\Theta - \pi)$ for all $\pi \in \{-1, \ldots, -\frac{1}{H}, 0, \frac{1}{H}, \ldots, 1\}$. (d) estimated signal $\hat{\Theta}$. The depicted example is a full-rank signal matrix based on Example 5 in Section 3.

tures. Here we consider the signal tensor of the form $\Theta = \log(1 + \mathcal{Z})$, where $\mathcal{Z} \in \mathbb{R}^{d\times d\times d}$ is an order-3 tensor with entries $\mathcal{Z}(i, j, k) = \frac{1}{d}\max(i, j, k)$ for $i, j, k \in \{1, \ldots, d\}$. The matrix analogy of Θ was studied in Chan and Airoldi (2014) as graph functions. In this case neither Θ nor $\mathcal{Z}$ is low-rank; in fact, the rank is no smaller than the dimension d as illustrated in Figure 1(b). Again, classical low-rank models fail to address this type of tensor structure.

In the above and many other examples, the signal tensors Θ of interest have high rank. Classical low-rank models will miss these important structures. New methods that allow flexible tensor modeling have yet to be developed.

Our contributions. We develop a new model called sign representable tensors to address the aforementioned challenges. Figure 2 illustrates our main idea. Our approach is built on the nonparametric sign representation of signal tensors, and we propose to estimate the sign tensors through a series of weighted classifications. In contrast to existing methods, our method is guaranteed to recover a wide range of low- and high-rank signals. We highlight two main contributions that set our work apart from earlier literature.

Statistically, the problem of high-rank estimation is fundamentally challenging. Existing estimation theory (Montanari and Sun, 2018; Cai et al., 2019; Ghadermarzy et al., 2019) exclusively focuses on the regime of fixed r growing d. However, such premise fails in high-rank tensors, where the rank may grow with, or even exceed, the dimension. A proper notion of nonparametric complexity is crucial. We show that, somewhat surprisingly, the sign tensor series not only preserves all information in the original signals, but also brings the benefits of flexibility and accuracy over classical low-rank models. The results fill the gap between parametric (low-rank) and nonparametric (high-rank) tensors, thereby greatly enriching the tensor literature.

From computational perspective, optimizations regarding tensors are in general NP-hard. Fortunately, tensors sought in application are often specially-structured, for which a number of efficient algorithms are available (Ghadermarzy et al., 2018; Wang and Li, 2020; Han et al., 2020). Our high-rank tensor estimate is provably reductable to a series of classifications, and it enjoys the divide-and-conquer nature that facilitates computation. The ability to import and adapt existing tensor algorithms is one advantage of our method.

We also highlight the challenges associated with tensors compared to matrices. High-rank matrix estimation is recently studied under nonlinear models (Ganti et al., 2015) and subspace clustering (Ongie et al., 2017; Fan and Udell, 2019). However, the problem for high-rank tensors is more prevalent, because the tensor rank often exceeds the dimension when order $K \geq 3$ (Anandkumar et al., 2017). This is in sharp contrast to matrices. We show that, applying matrix methods to higher-order tensors results in suboptimal estimates. A full exploitation of the higher-order structure is needed; this is another challenge we address in this paper.

Notation. We use $\mathrm{sgn}(\cdot)\colon \mathbb{R} \to \{-1, 1\}$ to denote the sign function, where $\mathrm{sgn}(y) = 1$ if $y \geq 0$ and -1 otherwise. We allow univariate functions, such as $\mathrm{sgn}(\cdot)$ or general $f\colon \mathbb{R} \to \mathbb{R}$, to be applied to tensors in an element-wise manner. We denote $a_n \lesssim b_n$ if $\limsup_n b_n/a_n \to 0$. We use the shorthand $[n]$ to denote the n-set $\{1, \ldots, n\}$ for $n \in \mathbb{N}_+$. Let $\mathcal{Y} \in \mathbb{R}^{d_1\times\cdots\times d_K}$ denote an order-K $(d_1, \ldots, d_k)$-dimensional tensor, and $\mathcal{Y}(\omega) \in \mathbb{R}$ denote the tensor entry indexed by $\omega \in [d_1] \times \cdots \times [d_K]$. An event E is said to occur "with high probability" if $\mathbb{P}(E)$ tends to 1 as the tensor dimension $d_{\min} = \min_k d_k \to \infty$. The Canonical Polyadic (CP) decomposition (Hitchcock, 1927) for a tensor $\Theta \in \mathbb{R}^{d_1\times\cdots\times d_K}$ is defined as

$$\Theta = \sum_{s=1}^{r} \lambda_s \boldsymbol{a}_s^{(1)} \otimes \cdots \otimes \boldsymbol{a}_s^{(K)}, \tag{2}$$

where $\lambda_1 \geq \cdots \geq \lambda_r > 0$ are called tensor singular values, and $\boldsymbol{a}_s^{(k)}$ are norm-1 tensor singular vectors, for all $s \in [r]$, $k \in [K]$, and $\otimes$ denote the outer product of vectors. The minimal r for which the decomposition (2) holds is called the tensor rank, denoted as $\mathrm{rank}(\Theta)$.

2. Model and proposal overview

Let $\mathcal{Y}$ be an order-K $(d_1, \ldots, d_K)$-dimensional data tensor generated from the following model,

$$\mathcal{Y} = \Theta + \mathcal{E}, \tag{3}$$

where $\Theta \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is the unknown signal tensor of interest, and $\mathcal{E}$ is a noise tensor consisting of mean-zero, independent but not necessarily identically distributed entries. We allow heterogenous noise such that the marginal distribution of noise entry $\mathcal{E}(\omega)$ may depend on ω. In general, assume that $\mathcal{Y}(\omega)$ take values in a bounded interval $[-A, A]$. We assume $A = 1$ throughout the paper.

Our observation is an incomplete data tensor from (3), denoted $\mathcal{Y}_\Omega$, where $\Omega \subset [d_1] \times \cdots \times [d_K]$ is the index set of observed entries. We consider a general model on Ω that allows both uniform and non-uniform samplings. Specifically, let $\Pi = \{\pi_\omega\}$ be an arbitrarily predefined probability distribution over the full index set with $\sum_{\omega \in [d_1] \times \cdots \times [d_K]} \pi_\omega = 1$. Assume that the entries ω in Ω are i.i.d. draws with replacement from the full index set using distribution Π. The sampling rule will be denoted as $\omega \sim \Pi$.

Before describing our main results, we provide the intuition behind our method. In the two examples in Section 1, the high-rankness in the signal Θ makes the estimation challenging. However, if we examine the sign of the π-shifted signal $\mathrm{sgn}(\Theta - \pi)$ for any given π, then, these sign tensors share the same sign patterns as low-rank tensors. Indeed, the signal tensor in example 1 has the same sign pattern as a rank-4 tensor, since $\mathrm{sgn}(\Theta - \pi) = \mathrm{sgn}(\mathcal{Z} - f^{-1}(\pi))$. The signal tensor in example 2 has the same sign pattern as a rank-2 tensor, since $\mathrm{sgn}(\Theta - \pi) = \mathrm{sgn}(\max(i, j, k) - d(e^\pi - 1))$.

The above observation suggests a general framework to estimate both low- and high-rank signal tensors. Figure 2 illustrates the main crux of our method. We dichotomize the data tensor into a series of sign tensors, $\mathrm{sgn}(\mathcal{Y}_\Omega - \pi)$, for all $\pi \in \mathcal{H} = \{-1, \ldots, -\frac{1}{H}, 0, \frac{1}{H}, \ldots, 1\}$, and then estimate the sign signals, $\mathrm{sgn}(\Theta - \pi)$, by performing classification

$$\hat{\mathcal{Z}}_\pi = \underset{\text{low rank tensor } \mathcal{Z}}{\arg\min} L(\mathrm{sgn}(\mathcal{Z}),\ \mathrm{sgn}(\mathcal{Y}_\Omega - \pi)).$$

Here, $L(\cdot, \cdot)$ denotes a carefully-designed classification objective function which will be described in later sections. Our final proposed tensor estimate takes the form

$$\hat{\Theta} = \frac{1}{2H+1} \sum_{\pi \in \mathcal{H}} \mathrm{sgn}(\hat{\mathcal{Z}}_\pi).$$

Our approach is built on the nonparametric sign representation of signal tensors, and the estimate $\hat{\Theta}$ is essentially learned from dichotomized tensor series $\{\mathrm{sgn}(\mathcal{Y}_\Omega - \pi) : \pi \in \mathcal{H}\}$ with certain weights. Interestingly, we show that a proper analysis of dichotomized data not only preserves all information in the original signals, but also brings benefits of accuracy and flexibility over classical low-rank models.

3. Sign representable tensors

This section develops sign representable tensor models for Θ in (3). We characterize the algebraic and statistical properties of sign tensor series, which serves the theoretical foundation for our method.

3.1. Sign-rank and sign tensor series

Let Θ be a tensor of interest and $\mathrm{sgn}(\Theta)$ the corresponding sign patten. The sign pattern induces an equivalence relationship between tensors. Two tensors are called sign equivalent, denoted $\simeq$, if they share the same sign patterns.

Definition 1 (Sign-rank)**.** The sign-rank of a tensor $\Theta \in \mathbb{R}^{d_1 \times \cdots \times d_L}$ is the minimal rank among all tensors that share the same sign patterns as Θ; i.e.,

$$\mathrm{srank}(\Theta) = \min\{\mathrm{rank}(\Theta') : \Theta' \simeq \Theta,\ \Theta' \in \mathbb{R}^{d_1 \times \cdots \times d_K}\}.$$

The sign-rank is also called *support rank* (Cohn and Umans, 2013), *minimal rank* (Alon et al., 2016), and *nondeterministic rank* (De Wolf, 2003). Earlier work defines sign-rank for binary tensors/matrices only; we extend the notion to continuous-valued tensors. Note that the sign-rank concerns only the sign pattern but discards the magnitudes information of Θ. In particular, $\mathrm{srank}(\Theta) = \mathrm{srank}(\mathrm{sgn}\Theta)$.

Like most tensor problems (Hillar and Lim, 2013), determining the sign-rank for a general tensor is NP hard (Alon et al., 2016). Fortunately, tensors arisen in application often possess special structures that facilitate analysis. We show that the low sign-rank family is strictly broader than usual low-rank family. By definition, the sign-rank is upper bounded by the usual rank. More generally,

Proposition 1 (Upper bounds of sign-rank)**.** *For any strictly monotonic function $g : \mathbb{R} \to \mathbb{R}$ with $g(0) = 0$,*

$$\mathrm{srank}(\Theta) \le \mathrm{rank}(g(\Theta)).$$

Conversely, the sign-rank is often dramatically smaller than the usual rank, as we have shown in Section 1.

Proposition 2 (Broadness)**.** *For every order $K \ge 2$ and dimension d, there exist order-K $(d, \ldots, d)$-dimensional tensors Θ such that* $\mathrm{rank}(\Theta) = d$ *and* $\mathrm{srank}(\Theta) = 2$.

We provide several examples in Appendix, in which the tensor rank grows with dimension d but the sign-rank remains a constant. The results highlight the advantages of using sign-rank in the high-dimensional tensor analysis. Corollary 1 and Proposition 2 together demonstrate the strict broadness of low sign-rank family over the usual low-rank family.

We now introduce a tensor family, which we coin as "sign representable tensors", for the signal tensors in model (3).

Definition 2 (Sign representable tensors)**.** Fix a level $\pi \in [-1, 1]$. A tensor Θ is called (r, π)-sign representable, if the tensor $(\Theta - \pi)$ has sign-rank bounded by r. A tensor Θ is called r-sign (globally) representable, if Θ is (r, π)-sign

representable for all $\pi \in [-1,1]$. The collection $\{\mathrm{sgn}(\Theta - \pi)\colon \pi \in [-1,1]\}$ is called the sign tensor series. We use $\mathscr{P}_{\mathrm{sgn}}(r) = \{\Theta\colon \mathrm{srank}(\Theta - \pi) \le r \text{ for all } \pi \in [-1,1]\}$ to denote the r-sign representable tensor family.

We show that the r-sign representable tensor family is a general model that incorporates most existing tensor models, including low-rank tensors, single index models, GLM models, and certain hypergraphon models.

Example 1 (CP/Tucker low-rank models)**.** The CP and Tucker low-rank tensors are the two most popular tensor models (Kolda and Bader, 2009). Let Θ be a low-rank tensor with CP rank r. We see that Θ belongs to the sign representable family; i.e., $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r+1)$ (the constant 1 is due to $\mathrm{rank}(\Theta - \pi) \le r+1$). Similar results hold for Tucker low-rank tensors $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r+1)$, where $r = \prod_k r_k$ with r_k being the k-th mode Tucker rank of Θ.

Example 2 (Tensor block models (TBMs))**.** Tensor block model (Wang and Zeng, 2019; Chi et al., 2020) assumes a checkerbord structure among tensor entries under marginal index permutation. The signal tensor Θ takes at most r distinct values, where r is the total number of multiway blocks. Our model incorporates TBM because $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r)$.

Example 3 (Generalized linear models (GLMs))**.** Let $\mathcal{Y}$ be a binary tensor from a logistic model (Wang and Li, 2020) with mean $\Theta = \mathrm{logit}(\mathcal{Z})$, where $\mathcal{Z}$ is a latent low-rank tensor. By definition, Θ is a low-rank sign representable tensor; note that Θ itself is a high-rank tensor (see Section 1). Same conclusion holds for general exponential-family tensor mean with a (known) link function (Hong et al., 2020).

Example 4 (Single index models (SIMs))**.** Single index model is a flexible semiparametric model proposed in economics (Robinson, 1988) and high-dimensional statistics (Balabdaoui et al., 2019; Ganti et al., 2017). We here extend the model to higher-order tensors Θ. The SIM assumes the existence of a (unknown) monotonic function $g\colon \mathbb{R} \to \mathbb{R}$ such that $g(\Theta)$ has rank r. We see that Θ belongs to the sign representable family; i.e., $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r+1)$.

Example 5 (Min/Max hypergraphon)**.** Graphon is a popular nonparametric model for networks (Chan and Airoldi, 2014; Xu, 2018). Here we revisit the model introduced in Section 1 for generality. Let Θ be an order-K tensor generated from the hypergraphon $\Theta(i_1,\dots,i_K) = \log(1 + \max_k x_{i_k}^{(k)})$, where $x_{i_k}^{(k)} \sim \mathrm{Unif}[0,1]$ i.i.d. for all $i_k \in [d_k]$ and $k \in [K]$. We conclude that $\Theta \in \mathscr{P}_{\mathrm{sgn}}(2)$, because every sign tensor $\mathrm{sgn}(\Theta - \pi)$ in the series of $\pi \in [0,\ \log 2]$ is a block tensor with at most two blocks.

The results extend to general min/max hypergraphons. Let $g(\cdot)$ be a continuous univariate function with at most $r \ge 1$ real-valued roots in the equation $g(z) = \pi$; this property holds, e.g., when $g(z)$ is a polynomial of degree r. Then, the tensor Θ generated from $\Theta(i_1,\dots,i_K) = g(\max_k x_{i_k}^{(k)})$ belongs to $\mathscr{P}_{\mathrm{sgn}}(r+1)$. Same conclusion applies if the maximum is replaced by minimum.

3.2. Statistical characterization of sign tensors via weighted classification

Accurate estimation of a sign representable tensor depends on the behavior of sign tensor series, $\mathrm{sgn}(\Theta - \pi)$. In this section, we show that sign tensors are completely characterizes by weighted classification. The results bridge the algebraic and statistical properties of sign representable tensors.

For a given $\pi \in [-1,1]$, define a π-shifted data tensor $\bar{\mathcal{Y}}_\Omega$ with entries $\bar{\mathcal{Y}}(\omega) = (\mathcal{Y}(\omega) - \pi)$ for $\omega \in \Omega$. We propose a weighted classification objective

$$L(\mathcal{Z}, \bar{\mathcal{Y}}_\Omega) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \underbrace{|\bar{\mathcal{Y}}(\omega)|}_{\text{entry-specific weight}} \times \underbrace{|\mathrm{sgn}\mathcal{Z}(\omega) - \mathrm{sgn}\bar{\mathcal{Y}}(\omega)|}_{\text{classification loss}}, \tag{4}$$

where $\mathcal{Z} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is the decision variable to be optimized, $|\bar{\mathcal{Y}}(\omega)|$ is the entry-specific weight equal to the distance from the tensor entry to the target level π. The weight incorporates the magnitude information in the classification, where entries far away from the target level are penalized more heavily in the objective.

Our proposed entry-specific weight is important for characterizing $\mathrm{sgn}(\Theta - \pi)$, as we show now. Define the weighted classification risk

$$\mathrm{Risk}(\mathcal{Z}) = \mathbb{E}_{\mathcal{Y}_\Omega} L(\mathcal{Z}, \bar{\mathcal{Y}}_\Omega), \tag{5}$$

where the expectation is taken with respect to $\mathcal{Y}_\Omega$ under model (3) and the sampling distribution $\omega \sim \Pi$. Note that the form of $\mathrm{Risk}(\cdot)$ implicitly depends on π; we suppress π when no confusion arises.

Proposition 3 (Global optimum of weighted risk)**.** *Suppose the data $\mathcal{Y}_\Omega$ is generated from model (3) with $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r)$. Then, for all $\bar{\Theta}$ that are sign equivalent to $\mathrm{sgn}(\Theta - \pi)$,*

$$\begin{aligned}\mathrm{Risk}(\bar{\Theta}) &= \inf\{\mathrm{Risk}(\mathcal{Z})\colon \mathcal{Z} \in \mathbb{R}^{d_1 \times \cdots \times d_K}\},\\ &= \inf\{\mathrm{Risk}(\mathcal{Z})\colon \mathrm{rank}\mathcal{Z} \le r\}.\end{aligned}$$

The results show that the sign tensor $\mathrm{sgn}(\Theta - \pi)$ optimizes the weighted classification risk. This fact suggests a practical procedure to estimate $\mathrm{sgn}(\Theta - \pi)$ via empirical risk optimization of $L(\mathcal{Z}, \bar{\mathcal{Y}}_\Omega)$. In order to establish the recovery guarantee, we need address the uniqueness (up to sign equivalence) of the optimizer for $\mathrm{Risk}(\cdot)$. The local behavior of Θ around π turns out to play a key role in the accuracy.

Some additional notation is needed. We use $\mathcal{N} = \{\pi\colon \mathbb{P}_{\omega \sim \Pi}(\Theta(\omega) = \pi) \ne 0\}$ to denote the set of mass points of Θ under Π. Assume there exists a constant $C > 0$, independent of tensor dimension, such that $|\mathcal{N}| \le C$. Note that both Π and Θ implicitly depend on the tensor dimension. Our

assumptions are imposed to $\Pi = \Pi(d)$ and $\Theta = \Theta(d)$ in the high-dimensional regime uniformly as $d := \min_k d_k \to \infty$.

Assumption 1 (α-smoothness)**.** Fix $\pi \notin \mathcal{N}$. Assume there exist constants $\alpha = \alpha(\pi) \geq 0, c = c(\pi) > 0$, independent of tensor dimension, such that,

$$\sup_{0 \leq t < \rho(\pi, \mathcal{N})} \frac{\mathbb{P}_{\omega \sim \Pi}[\omega \colon |\Theta(\omega) - \pi| \leq t]}{t^{\alpha}} \leq c, \qquad (6)$$

where $\rho(\pi, \mathcal{N}) := \min_{\pi' \in \mathcal{N}} |\pi - \pi'|$ denotes the distance from π to the nearest point in $\mathcal{N}$. The largest possible $\alpha = \alpha(\pi)$ is called the smoothness index at level π. We make the convention that $\alpha = \infty$ if the set $\{\omega \colon |\Theta(\omega) - \pi| \leq t\}$ has zero measure, implying few entries of Θ around the level π. We call the tensor Θ is α-globally smooth, if (6) holds with a global constant $c > 0$ for all $\pi \in [-1, 1]$ except for a finite number of levels.

The smoothness index α quantifies the intrinsic hardness of recovering $\mathrm{sgn}(\Theta - \pi)$ from $\mathrm{Risk}(\cdot)$. The recovery is more difficult at levels where the point mass concentrates (small α). A large value of $\alpha > 1$ corresponds a plateau-type zero density of $\Theta(\omega)$ around π, or equivalently, when the cumulative distribution function (CDF) $F(\pi) := \mathbb{P}_{\omega \sim \Pi}[\Theta(\omega) \leq \pi]$ remains flat around π. A small value of $\alpha < 1$ indicates the nonexistent (infinite) density at level π, or equivalently, when the CDF jumps at π. A typical case is $\alpha = 1$ when the CDF has finite non-zero derivative in the vicinity of π. Table 3 illustrates the CDFs for various common models; the details will be described in Simulation section.

We now reach the main theorem in this section. For two tensors Θ_1, Θ_2, define the mean absolute error (MAE)

$$\mathrm{MAE}(\Theta_1, \Theta_2) \stackrel{\text{def}}{=} \mathbb{E}_{\omega \sim \Pi} \|\Theta_1 - \Theta_2\|_1.$$

Theorem 1 (Identifiability)**.** *Under Assumption 1, for all tensors* $\bar{\Theta} \simeq \mathrm{sgn}(\Theta - \pi)$ *and tensors* $\mathcal{Z} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$,

$$\mathrm{MAE}(\mathrm{sgn}\mathcal{Z}, \mathrm{sgn}\bar{\Theta}) \leq C(\pi) \left[\mathrm{Risk}(\mathcal{Z}) - \mathrm{Risk}(\bar{\Theta})\right]^{\alpha/(\alpha+1)},$$

where $C(\pi) > 0$ *is independent of* $\mathcal{Z}$.

The result establishes the stability of recovering sign tensors $\mathrm{sgn}(\Theta - \pi)$ from optimizing the population risk (5). Furthermore, the bound immediately suggests the uniqueness of the optimizer for $\mathrm{Risk}(\cdot)$ up to a zero-measure set under Π, as long as $\alpha \neq 0$. We find that a higher value of α implies more stable recovery. Similar results hold for the empirical risk minimization (4) (see Section 4 for details).

We conclude this section by relating Assumption 1 to the examples described in Section 3.1. For simplicity, suppose Π is the uniform sampling for now. Based on definition, the tensor block model is ∞-globally smooth. This is because there are finitely many elements in $\mathcal{N}$, i.e, the set of distinct block means in Θ. Furthermore, we have $\alpha = \infty$ for all $\pi \notin \mathcal{N}$, since the numerator in (6) is zero for all such π. The min/max hypergaphon with a r-degree polynomial function is 1-globally smooth, because $\alpha = 1$ for all π in the function range except at most $(r-1)$ many (stationary) levels.

4. Nonparametric tensor completion via sign series

In previous sections we have established the sign representation and sign series recovery for the signal tensor. In this section, we present our algorithm proposed in Section 2 (Figure 2) in details. We provide the estimation error bound and address the empirical implementation of the algorithm.

4.1. Estimation error and sample complexity

Given a noisy incomplete tensor observation $\mathcal{Y}_\Omega$ from model (3), we cast the problem of estimating Θ into a series of weighted classifications. Specifically we propose a tensor estimator using the sign representation,

$$\hat{\Theta} = \frac{1}{2H+1} \sum_{\pi \in \mathcal{H}} \mathrm{sgn}\hat{\mathcal{Z}}_\pi, \qquad (7)$$

where $\hat{\mathcal{Z}}_\pi \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is the π-weighted classifier estimated at levels $\pi \in \mathcal{H} = \{-1, \ldots, -\frac{1}{H}, 0, \frac{1}{H}, \ldots, 1\}$,

$$\hat{\mathcal{Z}}_\pi = \underset{\mathcal{Z} \colon \mathrm{rank}\mathcal{Z} \leq r}{\arg\min} \; L(\mathcal{Z}, \mathrm{sgn}(\mathcal{Y}_\Omega - \pi)). \qquad (8)$$

Here $L(\cdot, \cdot)$ denotes the weighted classification objective defined in (4), where we have plugged $\bar{\mathcal{Y}}_\Omega = (\mathcal{Y}_\Omega - \pi)$ in the expression, and the rank constraint follows from Proposition 3. For the theory, we assume the true r is known; in practice, r could be chosen in a data adaptive fashion via cross-validation or elbow method (Hastie et al., 2009). Figure 2 illustrates the main steps in our proposed estimation.

The next theorem establishes the statistical convergence for the sign tensor estimation (8), which is an important ingredient for the signal tensor estimator $\hat{\Theta}$ in (7).

Theorem 2 (Sign tensor errors)**.** *Suppose* $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r)$ *and* $\Theta(\omega)$ *is* α*-globally smooth under* $\omega \sim \Pi$*, where* $\alpha \in (0, 1]$*. Let* $\hat{\mathcal{Z}}_\pi$ *be the estimator in* (8)*, and denote* $d_{\max} = \max_{k \in [K]} d_k$*. Then, for all* $\pi \in [-1, 1]$ *except for a finite number of levels, with very high probability over* $\mathcal{Y}_\Omega$*,*

$$\mathrm{MAE}(\mathrm{sgn}\hat{\mathcal{Z}}_\pi, \mathrm{sgn}(\Theta - \pi)) \lesssim \left(\frac{d_{\max} r}{|\Omega|}\right)^{\frac{\alpha}{\alpha+2}} + \frac{1}{\rho^2(\pi, \mathcal{N})} \frac{d_{\max} r}{|\Omega|}. \qquad (9)$$

Theorem 2 provides the error bound for the estimated sign tensors. Compared to the population version results in Theorem 1, we here explicitly reveal the dependence of accuracy

Algorithm 1 Nonparametric tensor completion

Input: Noisy and incomplete data tensor $\mathcal{Y}_\Omega$, rank r, resolution parameter H

for $\pi \in \mathcal{H} = \{-1, \dots, -\frac{1}{H}, 0, \frac{1}{H}, \dots, 1\}$ **do**
  Random initialization of tensor factors $\boldsymbol{A}_k = [\boldsymbol{a}_1^{(k)}, \dots, \boldsymbol{a}_r^{(k)}] \in \mathbb{R}^{d_k \times r}$ for all $k \in [K]$.
  while not convergence **do**
    for $k = 1, \dots, K$ **do**
      Update $\boldsymbol{A}_k$ while holding others fixed: $\boldsymbol{A}_k \leftarrow \arg\min_{\boldsymbol{A}_k \in \mathbb{R}^{d_k \times r}} \sum_{\omega \in \Omega} |\mathcal{Y}(\omega) - \pi| F(\mathcal{Z}(\omega)\mathrm{sgn}(\mathcal{Y}(\omega) - \pi))$, where $F(\cdot)$ is the large-margin loss, and $\mathcal{Z} = \sum_{s \in [r]} \boldsymbol{a}_s^{(1)} \otimes \cdots \otimes \boldsymbol{a}_s^{(K)}$ is a rank-r tensor.
    end for
  end while
  Return $\mathcal{Z}_\pi \leftarrow \sum_{s \in [r]} \boldsymbol{a}_s^{(1)} \otimes \cdots \otimes \boldsymbol{a}_s^{(K)}$.
end for

Output: Estimated signal tensor $\hat{\Theta} = \frac{1}{2H+1} \sum_{\pi \in \mathcal{H}} \mathrm{sgn}(\mathcal{Z}_\pi)$.

on the sample complexity and on the level π. The result (10) shows the sign error decreases with $|\Omega|$, provided that $\alpha \neq 0$. In particular, our sign estimate achieves consistent recovery using as few as $\tilde{O}(d_{\max} r)$ noisy observations.

Combining the sign representability of the signal tensor and the sign estimation accuracy, we obtain the main results on our nonparametric tensor estimation method.

Theorem 3 (Tensor estimation error)**.** *Consider the same conditions of Theorem 2. Let $\hat{\Theta}$ be the estimate in (7). With very high probability over $\mathcal{Y}_\Omega$,*

$$\mathrm{MAE}(\hat{\Theta}, \Theta) \lesssim \left(\frac{d_{\max} r}{|\Omega|}\right)^{\alpha/(\alpha+2)} + \frac{1}{H} + H\frac{d_{\max} r}{|\Omega|}. \quad (10)$$

In particular, setting $H \asymp \left(\frac{|\Omega|}{d_{\max} r}\right)^{1/2}$ yields the error bound

$$\mathrm{MAE}(\hat{\Theta}, \Theta) \lesssim \left(\frac{d_{\max} r}{|\Omega|}\right)^{\frac{\alpha}{\alpha+2} \vee \frac{1}{2}}. \quad (11)$$

Theorem 3 demonstrates the convergence rate of our tensor estimation. The bound (10) reveals three sources of errors: the estimation error inherited from sign tensors, the bias from sign series representations, and the variance thereof. The resolution parameter H controls the bias-variance tradeoff. We remark that the signal estimation error (11) is generally no better than the corresponding sign error (10). This is to be expected, since magnitude estimation is a harder problem than sign estimation.

In the special case of full observation with equal dimension $d_1 = \cdots = d_K = d$, our signal estimate achieves convergence rate

$$\mathrm{MAE}(\hat{\Theta}, \Theta) \leq \left(\frac{r}{d^{K-1}}\right)^{\frac{\alpha}{\alpha+2} \vee \frac{1}{2}}.$$

In contrast to earlier methods, our estimation accuracy applies to both low- and high-rank signal tensors Θ. The rate depends on the sign complexity $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r)$, and this r is often much smaller than the usual tensor rank (see Section 3.1). Therefore, our method not only incorporates the earlier work on low-rank tensor estimation but also addresses a broad family that was previously impossible.

We apply our method to selected examples in Section 3.1, and compare the results with existing literature. The numerical comparison is provided in Section 5.

Example 2 (Tensor block models)**.** Consider a tensor block model with r multiway blocks in total. Our result implies a rate $\mathcal{O}(d^{-(K-1)/2})$ by taking $\alpha = \infty$. This rate agrees with the previous root-mean-square error (RMSE) for block tensor estimation (Wang and Zeng, 2019).

Example 3 (CP/Tucker and GLMs)**.** Consider a GLM tensor $\Theta = g(\mathcal{Z})$, where g is a known link function and $\mathcal{Z}$ is a latent low-rank tensor. Suppose the marginal density of $\Theta(\omega)$ is bounded from above. Applying our results with $\alpha = 1$ yields $\mathcal{O}(d^{-(K-1)/3})$. Note that this rate is slightly slower than the parametric RMSE rate (Zhang and Xia, 2018; Wang and Li, 2020). One possible reason is that our result remains valid for unknown g and more general high-rank tensors with $\alpha = 1$. The tensor family involved in our upper bound is broader than parametric models.

The following sample complexity for nonparamtric tensor completion is a direct consequence of Theorem 3.

Corollary 1 (Sample complexity for nonparametric completion)**.** *Under the same conditions of Theorem 3, with very high probability over $\mathcal{Y}_\Omega$,*

$$\mathrm{MAE}(\hat{\Theta}, \Theta) \to 0, \quad \textit{as} \quad \frac{|\Omega|}{d_{\max} r} \to \infty.$$

Our result improves earlier work (Ghadermarzy et al., 2019) by allowing both low- and high-rank signals. Note that $\tilde{\mathcal{O}}(d_{\max} r)$ roughly matches the degree of freedom in sign tensors, suggesting the optimality of our sample complexity.

4.2. Numerical implementation

This section addresses the practical implementation of our estimation (7) illustrated in Figure 2. Our sign representation of tensor estimate Θ is a simple average of H sign

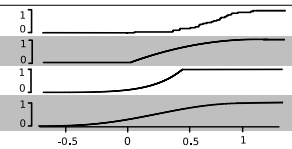
Simulation	Signal Tensor Θ	Rank	Sign Rank	Global α	CDF of Tensor Entries	Noise
1	$\mathcal{C} \times \boldsymbol{M}_1 \times \boldsymbol{M}_2 \times \boldsymbol{M}_3$	3^3	2	∞		Uniform $[-0.3, 0.3]$
2	$\lvert \boldsymbol{a} \otimes \mathbf{1} \otimes \mathbf{1} - \mathbf{1} \otimes \boldsymbol{a} \otimes \mathbf{1} \rvert$	d	≤ 3	1		Normal $\mathcal{N}(0, 0.15)$
3	$\log(0.5 + Z_{\max})$	$\geq d$	2	1		Uniform $[-0.1, 0.1]$
4	$2.5 - \exp(\mathcal{Z}_{\min}^{1/3})$	$\geq d$	2	1		Normal $\mathcal{N}(0, 0.15)$

Figure 3. Simulation models used for comparison. We use $\boldsymbol{M}_k \in \{0,1\}^{d\times 3}$ to denote membership matrices, $\mathcal{C} \in \mathbb{R}^{3\times3\times3}$ the block means, $\boldsymbol{a} = \frac{1}{d}(1, 2, \ldots, d)^T \in \mathbb{R}^d$, $\mathcal{Z}_{\max}$ and $\mathcal{Z}_{\min}$ are order-3 tensors with entries $\frac{1}{d}\max(i, j, k)$ and $\frac{1}{d}\min(i, j, k)$ respectively.

tensors, which can be solved in a divide-and-conquer fashion. Briefly, we estimate the sign tensors $\mathcal{Z}_\pi$ (which are detailed in the next paragraph) for the series $\pi \in \mathcal{H}$ through parallel implementation, and then aggregate the results for the final output. This approach maintains low computational cost similar to a single sign tensor estimation.

For the sign tensor estimation (8), the problem reduces to binary tensor decomposition with a weighted classification loss. A number of algorithms have been developed for this problem (Ghadermarzy et al., 2018; Wang and Li, 2020; Hong et al., 2020). We adopt similar ideas by tailoring the algorithms to our contexts. Following the common practice in classification, we replace the binary loss $\ell(z, y) = |\text{sgn}z - \text{sgn}y|$ with a surrogate loss $F(m)$ as a continuous function of margin $m \overset{\text{def}}{=} z\text{sgn}(y)$. Examples of large-margin loss are hinge loss $F(m) = (1 - m)_+$, logistic loss $F(m) = \log(1 + e^{-m})$, and nonconvex ψ-loss $F(m) = 2\min(1, (1 - m)_+)$ with $m_+ = \max(m, 0)$. We implement the hinge loss and logistic loss in our algorithm, although our framework is applicable to general large-margin losses (Bartlett et al., 2006).

The rank constraints in the optimization (7) have been extensively studied in literature. Recent developments involve convex norm relaxation (Ghadermarzy et al., 2018) and nonconvex optimization (Wang and Li, 2020; Han et al., 2020). Unlike matrices, computing the tensor convex norm is NP hard, so we choose (non-convex) alternating optimization due to its numerical efficiency. Briefly, we use the rank decomposition (2) of $\mathcal{Z} = \mathcal{Z}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_K)$ where $\boldsymbol{A}_k = [\boldsymbol{a}_1^{(k)}, \ldots, \boldsymbol{a}_r^{(k)}] \in \mathbb{R}^{d_k \times r}$ are unknown factor matrices to optimize. We numerically solve (7) by optimizing one factor $\boldsymbol{A}_k$ at a time while holding others fixed. Each sub-optimization reduces to a convex optimization with a low-dimensional decision variable. Following common practice in tensor optimization (Anandkumar et al., 2014; Hong et al., 2020), we run the optimization from multiple initializations to locate a final estimate with the lowest objective value. The full procedure is described in Algorithm 1.

5. Simulation

In this section, we compare our nonparametric tensor method (**NonParaT**) with two alternative approaches: low-rank tensor CP decomposition (**CPT**), and the matrix analogy of our method applied to tensor unfolding (**NonParaM**). We assess the performance under both complete and incomplete observations. The signal tensors are generated based on four models listed in Table 3. The simulation covers a wide range of complexity, including block tensors, transformed low rank tensors, min/max hypergraphon with log and exponential functions. We consider order-3 tensors of equal dimension $d_1 = d_2 = d_3 = d$, and set $d \in \{15, 20, \ldots, 55, 60\}$, $r = 2$, $H = 10 + (d - 15)/5$ in Algorithm 1. For **NonParaM**, we apply Algorithm 1 to each of the three unfolded matrices and report the average error. All summary statistics are averaged across 30 replicates.

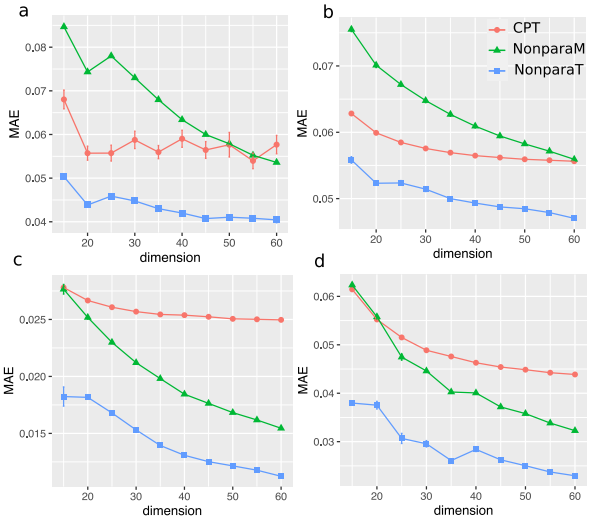


Figure 4. Estimation performance against tensor dimension. Panels (a)-(d) correspond to simulation models 1-4 in Table 3.

Figure 4 compares the estimation error with full observation. We find all three methods show decreased error as the dimension increases. Furthermore, our method **NonParaT** achieves the best performance in all scenarios, whereas the second best method is **CPT** for models 1-2, and **NonParaM** for models 3-4. One possible reason is that models 1-2 have controlled multilinear tensor rank, which makes tensor methods **NonParaT** and **CPT** more accurate than matrix methods. For models 3-4, the rank exceeds the tensor dimension, and the two nonparametric methods **NonParaT** and **NonparaM** are well suited for high rank signals.

Figure 5 shows the completion error against observation fraction. We fix $d = 40$ and gradually increase the observation fraction $\frac{|\Omega|}{d^3}$ from 0.3 to 1. It is seen that **NonParaT**

achieves the lowest error among all methods. Our simulation covers a reasonable range of various complexities; for example, model 1 has 3^3 jumps in the CDF of signal Θ, and models 2 and 4 have unbounded noise. Nevertheless, our method shows good performance in spite of model misspecification. This result is appealing in practice because the structure of underlying signal tensor is often unknown.

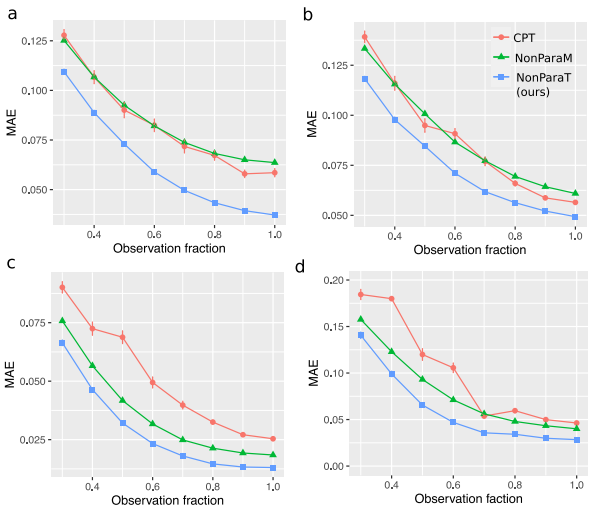


Figure 5. Completion performance against observation fraction. Panels (a)-(d) correspond to simulation models 1-4 in Table 3.

6. Data analysis

We apply our method to two tensor datasets, the MRN-114 human brain connectivity data (Wang et al., 2017), and NIPS data (Globerson et al., 2007). The MRN-144 dataset records the structural connectivity among 68 brain regions for 114 individuals along with their Intelligence Quotient (IQ) scores. We organize the connectivity data into an order-3 tensor, where entries encode the presence or absence of fiber connections between brain regions across individuals. The NIPS dataset consists of counts of words in NIPS papers published from 1987 to 2003. We focus on the top 100 authors, 200 most frequent words, and normalize each word count by log transformation with pseudocount 1. The resulting dataset is an order-3 tensor with entry representing the log counts of words by authors across years.

MRN-144 brain connectivity dataset			
Method	$r = 6$	$r = 9$	$r = 12$
NonparaT (Ours)	**0.14**(0.003)	**0.12**(0.004)	**0.12**(0.007)
CPT	0.23(0.006)	0.22(0.004)	0.21(0.004)
NIPS dataset			
Method	$r = 6$	$r = 9$	$r = 12$
NonparaT (Ours)	**0.16**(0.002)	**0.15**(0.001)	**0.14**(0.001)
CPT	0.20(0.007)	0.19(0.007)	0.17(0.007)

Table 1. Comparison of prediction error in the MRN-114 and NIPS data analysis. For low-rank CPT, we use R function `rTensor` with default hyperparameters, and for our method, we set $H = 20$.

Table 1 compares the prediction accuracy of our method and low-rank CP method. The reported MAE is averaged over five runs of cross-validation, with 20% entries for testing and 80% for training. Our method substantially outperforms the classical low-rank method for every configuration in the considered range. Further increment of rank appears to have little effect on the performance, and we find that increased missingness gives more advantages to our method (see details in Appendix). In contrast, the low-rank CPT exhibits poor prediction even at a relatively high rank. The comparison highlights the advantage of our method in achieving accuracy while maintaining low complexity.

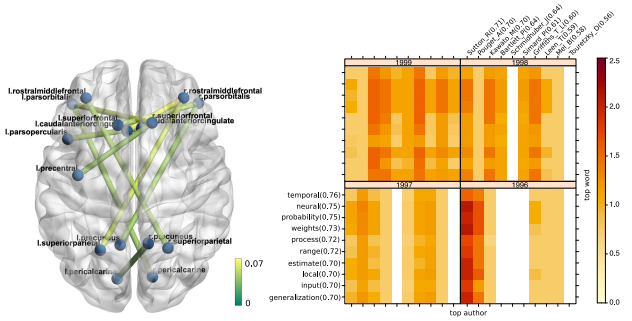


Figure 6. Estimated signal tensor from brain connectivity data (a) and NIPS data (b).

We next examine the estimated signal tensor $\hat{\Theta}$ form our method. Figure 6a shows the top 10 edges in MSN-144 dataset based on regression analysis of denoised edge connection with IQ scores. We find that top edges are mostly inter-hemisphere, which are consistent with recent research in brain connectivity pattern with intelligence (Li et al., 2009; Wang et al., 2017). Figure 6b illustrates the result from NIPS data, where we plot the top entries in $\hat{\Theta}$ corresponding to top authors and words (after excluding generic words such as *figure*, *current*, etc). The detected patten is consistent with the active topics in the NIPS publication. Among the top words are *temporal* (marginal mean = 0.76), *neural* (0.75), and *generalization* (0.70), whereas top authors are *Richard Sutton* (0.71), *Pouget_A*(0.70), *Kawato_M* (0.70), and *Bartlett Peter* (0.64). We also find strong heterogeneity among word occurrence across authors and years. For example, *neural* and *weights* are popular words for *Tomas Griffiths* in 1998-1999, whereas *temproal* occurs more often in *Richard Sutton et al* in 1996. The achieved denoising accuracy and pattern detection demonstrates the applicability of our method.

7. Conclusion

We have develop a new framework for high-rank tensor estimation based on sign tensor representation. The new method improves the predictive power and enhances interpretability by nonparametric estimation of signal tensors. The application to brain connection dataset and author-topic mining shows the practical utility of the proposed method.

References

Alon, N., S. Moran, and A. Yehudayoff (2016). Sign rank versus vc dimension. In *Conference on Learning Theory*, pp. 47–80.

Alquier, P. and G. Biau (2013). Sparse single-index model. *Journal of Machine Learning Research 14*(Jan), 243–280.

Anandkumar, A., R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky (2014). Tensor decompositions for learning latent variable models. *The Journal of Machine Learning Research 15*(1), 2773–2832.

Anandkumar, A., R. Ge, and M. Janzamin (2017). Analyzing tensor power method dynamics in overcomplete regime. *The Journal of Machine Learning Research 18*(1), 752–791.

Balabdaoui, F., C. Durot, H. Jankowski, et al. (2019). Least squares estimation in the monotone single index model. *Bernoulli 25*(4B), 3276–3310.

Bartlett, P. L., M. I. Jordan, and J. D. McAuliffe (2006). Convexity, classification, and risk bounds. *Journal of the American Statistical Association 101*(473), 138–156.

Cai, C., G. Li, H. V. Poor, and Y. Chen (2019). Nonconvex low-rank tensor completion from noisy data. In *Advances in Neural Information Processing Systems*, pp. 1863–1874.

Chan, S. and E. Airoldi (2014). A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning*, pp. 208–216.

Chi, E. C., B. J. Gaines, W. W. Sun, H. Zhou, and J. Yang (2020). Provable convex co-clustering of tensors. *Journal of Machine Learning Research 21*(214), 1–58.

Cohn, H. and C. Umans (2013). Fast matrix multiplication using coherent configurations. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1074–1087. SIAM.

De Lathauwer, L., B. De Moor, and J. Vandewalle (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications 21*(4), 1253–1278.

De Wolf, R. (2003). Nondeterministic quantum query and communication complexities. *SIAM Journal on Computing 32*(3), 681–699.

Fan, J. and M. Udell (2019). Online high rank matrix completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8690–8698.

Ganti, R., N. Rao, L. Balzano, R. Willett, and R. Nowak (2017). On learning high dimensional structured single index models. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 1898–1904. AAAI Press.

Ganti, R. S., L. Balzano, and R. Willett (2015). Matrix completion under monotonic single index models. *Advances in Neural Information Processing Systems 28*, 1873–1881.

Ghadermarzy, N., Y. Plan, and O. Yilmaz (2018). Learning tensors from partial binary measurements. *IEEE Transactions on Signal Processing 67*(1), 29–40.

Ghadermarzy, N., Y. Plan, and Ö. Yilmaz (2019). Near-optimal sample complexity for convex tensor completion. *Information and Inference: A Journal of the IMA 8*(3), 577–619.

Globerson, A., G. Chechik, F. Pereira, and N. Tishby (2007). Euclidean embedding of co-occurrence data. *Journal of Machine Learning Research 8*(Oct), 2265–2295.

Han, R., R. Willett, and A. Zhang (2020). An optimal statistical and computational framework for generalized tensor estimation. *arXiv preprint arXiv:2002.11255*.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hillar, C. J. and L.-H. Lim (2013). Most tensor problems are NP-hard. *Journal of the ACM (JACM) 60*(6), 45.

Hitchcock, F. L. (1927). The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics 6*(1-4), 164–189.

Hong, D., T. G. Kolda, and J. A. Duersch (2020). Generalized canonical polyadic tensor decomposition. *SIAM Review 62*(1), 133–163.

Karbasi, A. and S. Oh (2012). Robust localization from incomplete local information. *IEEE/ACM Transactions on Networking 21*(4), 1131–1144.

Kolda, T. G. and B. W. Bader (2009). Tensor decompositions and applications. *SIAM Review 51*(3), 455–500.

Li, Y., Y. Liu, J. Li, W. Qin, K. Li, C. Yu, and T. Jiang (2009). Brain anatomical network and intelligence. *PLoS Comput Biol 5*(5), e1000395.

Montanari, A. and N. Sun (2018). Spectral algorithms for tensor completion. *Communications on Pure and Applied Mathematics 71*(11), 2381–2425.

Ongie, G., R. Willett, R. D. Nowak, and L. Balzano (2017, 06–11 Aug). Algebraic variety models for high-rank matrix completion. In D. Precup and Y. W. Teh (Eds.), *Proceedings of the 34th International Conference on Machine Learning*, Volume 70 of *Proceedings of Machine Learning Research*, International Convention Centre, Sydney, Australia, pp. 2691–2700. PMLR.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, 931–954.

Wang, L., D. Durante, R. E. Jung, and D. B. Dunson (2017). Bayesian network–response regression. *Bioinformatics 33*(12), 1859–1866.

Wang, M., J. Fischer, and Y. S. Song (2019). Three-way clustering of multi-tissue multi-individual gene expression data using semi-nonnegative tensor decomposition. *The Annals of Applied Statistics 13*(2), 1103–1127.

Wang, M. and L. Li (2020). Learning from binary multiway data: Probabilistic tensor decomposition and its statistical optimality. *Journal of Machine Learning Research 21*(154), 1–38.

Wang, M. and Y. Zeng (2019). Multiway clustering via tensor block models. In *Advances in Neural Information Processing Systems*, pp. 713–723.

Xu, J. (2018). Rates of convergence of spectral methods for graphon estimation. In *International Conference on Machine Learning*, pp. 5433–5442. PMLR.

Zhang, A. and D. Xia (2018). Tensor SVD: Statistical and computational limits. *IEEE Transactions on Information Theory*.