

# Beyond the Signs: Nonparametric Tensor Completion via Sign Series

Anonymous Authors<sup>1</sup>

## Abstract

We consider the problem of tensor estimation from noisy observations with 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 using a series of low sign-rank tensor. Unlike earlier methods, the sign series representation effectively addresses both low- and high-rank signal tensors, 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 estimatable, by classification tasks with carefully-specified weights. The excess risk rate, estimation error bound, and sample complexity are established. The results uncover the joint contribution of statistical bias-variance errors and discretization errors. Numerical results demonstrate the robustness of our proposal over previous tensor methods.

## 1. Introduction

High-order tensors have received increasing recent attention in enormous fields including social networks (Anandkumar et al., 2014), neuroscience (Wang et al., 2017) and genomics (Wang et al., 2019). In this paper we focus on the popular signal plus noise model,

$$\mathcal{Y} = \Theta + \mathcal{E}, \quad (1)$$

where  $\mathcal{Y}$  is an order- $K$  data tensor,  $\Theta$  is a signal tensor of interest, and  $\mathcal{E}$  is the noise tensor. Our goal is to accurately estimate  $\Theta$  from the incomplete, noisy observation. In particular, we focus on the following two problems:

- Q1 [Nonparametric tensor estimation]. How to flexibly

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

estimate  $\Theta$  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  $\Theta$ ?

### 1.1. Inadequacies of common low-rank models

The signal plus noise model (3) is common in tensor literature. Most existing methods perform estimation based on the low-rankness of  $\Theta$  (Anandkumar et al., 2014; Montanari and Sun, 2018; Kadmon and Ganguli, 2018; Cai et al., 2019). While these methods have shown great success in low-rank recovery, practical signals might not be truly low rank. Here we provide two examples to illustrate the limitation of classical low-rank models.

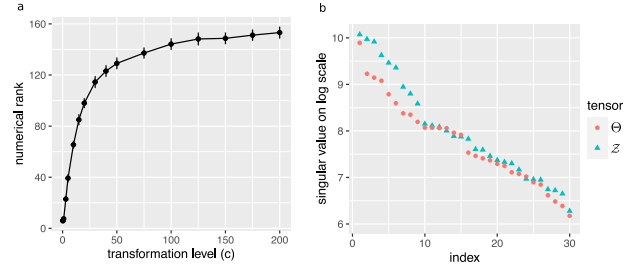


Figure 1. (a) Numerical rank of  $\Theta = f(\mathcal{Z})$  versus  $c$  in Example 1. (b) Top  $d = 30$  tensor singular values in Example 2.

The first example reveals the sensitivity of tensor rank to order-preserving transformations. Let  $\mathcal{Z} \in \mathbb{R}^{30 \times 30 \times 30}$  be an order-3 tensor with  $\text{rank}(\mathcal{Z}) = 3$ . Suppose a monotonic transformation  $f(z) = (1 + \exp(-cz))^{-1}$  is applied to  $\mathcal{Z}$  entrywise, and we observe data from model (3) with the signal tensor  $\Theta = f(\mathcal{Z})$ . Figure 1(a) plots the numerical rank of  $\Theta$  versus  $c$ . As we see, the rank increases rapidly with  $c$ , rendering 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 some unknown transformation prior to measurements. The sensitivity to transformation therefore makes the low-rank model less desirable in practice.

The second example shows that the classical low-rankness may exclude important specially-structured tensors. Here we consider the signal tensor of the form  $\Theta = \log(1 +$

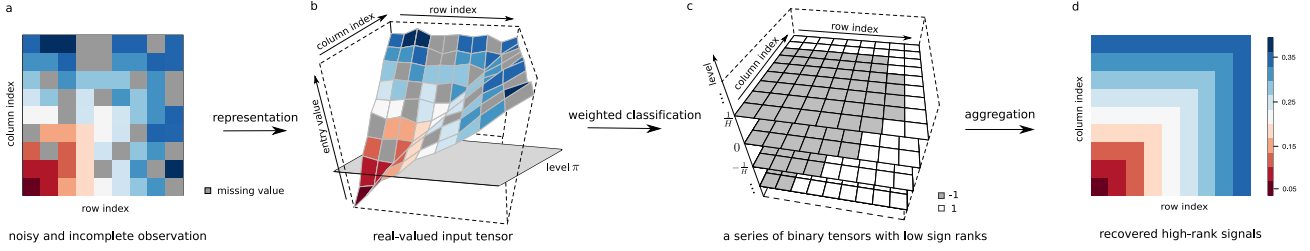


Figure 2. Illustration of our proposed method. For visualization purpose, we plot an order-2 tensor (a.k.a. matrix) in the figure; similar procedure applies to higher-order tensors. (a): input tensor  $\mathcal{Y}_\Omega$  with noisy and incomplete entries. (b) and (c): our method uses weighted classification to estimate sign tensors  $\text{sgn}(\Theta - \pi)$  for a sequence of levels  $\pi \in \{-1, \dots, -\frac{1}{H}, 0, \frac{1}{H}, \dots, 1\}$ . (d) output tensor  $\hat{\Theta}$  with denoised and imputed entries. The depicted example is based on Example 5, where the true signal matrix has full rank.

$\mathcal{Z}$ ), where  $\mathcal{Z}$  is an order-3 tensor with entries  $\mathcal{Z}(i, j, k) = \frac{1}{d} \max(i, j, k)$  for  $(i, j, k) \in [d]^3$ . The matrix analogy of this  $\Theta$  was studied in Chan and Airolidi (2014) in the context of graphon analysis. In this case neither  $\Theta$  nor  $\mathcal{Z}$  is low-rank; indeed, both tensors have rank lower bounded by dimension  $d$  as illustrated in Figure 1(b). However, classical low-rank models fail to address these types of structures.

In the above and many other examples, the signal tensors  $\Theta$  of interest are of high rank. These structures are hardly detectable using classical low-rank models. New methods that allow flexible tensor modeling have yet to be developed.

## 1.2. Our proposal and contribution

We briefly provide the intuition of our method and summarize the contributions. In the earlier two high-rank examples, if we examine the sign of the  $\pi$ -shifted signal,  $\text{sgn}(\Theta - \pi)$  for any given  $\pi$ , then these sign tensors exhibit easily detectable low-rankness. In particular, the signal tensor in example 1 has the same sign pattern as a rank-4 tensor, since  $\text{sgn}(\Theta - \pi) = \text{sgn}(\mathcal{Z} - f^{-1}(\pi))$ . The signal tensor in example 2 has the same sign pattern as a rank-2 tensor, since  $\text{sgn}(\Theta - \pi) = \text{sgn}(\max(i, j, k) - d(e^\pi - 1))$ .

The above observation suggests a general framework to analyze 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,  $\text{sgn}(\mathcal{Y}_\Omega - \pi)$ , for  $\pi \in \mathcal{H} = \{-1, \dots, -\frac{1}{H}, 0, \frac{1}{H}, \dots, 1\}$ , and then estimate the sign signals,  $\text{sgn}(\Theta - \pi)$ , by performing classification

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

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

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

Our approach is built on the nonparametric sign representation of signal tensors, and the estimate  $\hat{\Theta}$  is essentially esti-

mated from dichotomized tensor series  $\{\text{sgn}(\mathcal{Y}_\Omega - \pi) : \pi \in \mathcal{H}\}$ . Surprisingly, we show that proper analysis based on dichotomized data not only preserves all information in the original signals, but also brings benefits of accuracy and flexibility over classical low-rank models. The method enjoys both statistical effectiveness and computational efficiency.

Our method features several advantages. In contrast to earlier methods (Hong et al., 2020; Filipović and Jukić, 2015), our method recovers the signal tensor under a wide range of low- and high-rank tensor models. Our method brings the nonparametric advantages of flexibility into tensor estimation and completion. The algorithm achieves computational efficiency by leveraging classification and divide-and-conquer algorithms.

## 1.3. Preliminaries

We use the shorthand  $[n]$  to denote the  $n$ -set  $\{1, \dots, n\}$  for  $n \in \mathbb{N}_+$ . We use  $\otimes$  to denote the outer product of vectors,  $\|\mathbf{x}\|_2$  to denote the vector 2-norm, and  $\mathbf{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 = 1\}$  to denote the  $(d-1)$ -dimensional unit sphere. Let  $\mathcal{Y} \in \mathbb{R}^{d_1 \times \dots \times d_K}$  denote an order- $K$   $(d_1, \dots, d_K)$ -dimensional tensor, and  $\mathcal{Y}(\omega) \in \mathbb{R}$  denote the tensor entry indexed by  $\omega \in [d_1] \times \dots \times [d_K]$ . The Frobenius norm of  $\mathcal{Y}$  is defined as  $\|\mathcal{Y}\|_F = \sqrt{\sum_{\omega} \mathcal{Y}^2(\omega)}$ . Unlike matrices, various notions of decomposition have been developed for tensors of order  $K \geq 3$ . The Canonical Polyadic (CP) tensor decomposition (Hitchcock, 1927) for a tensor  $\Theta \in \mathbb{R}^{d_1 \times \dots \times d_K}$  is defined as

$$\Theta = \sum_{s=1}^r \lambda_s \mathbf{a}_s^{(1)} \otimes \dots \otimes \mathbf{a}_s^{(K)}, \quad (2)$$

where  $\lambda_1 \geq \dots \geq \lambda_r > 0$  are called tensor singular values, and  $\mathbf{a}_s^{(k)} \in \mathbf{S}^{d_k-1}$  are called tensor singular vectors, for all  $s \in [r]$ ,  $k \in [K]$ . The minimal  $r$  for which the decomposition (2) holds is called the tensor rank, denoted as  $\text{rank}(\Theta)$ .

We use  $\text{sgn}(\cdot) : \mathbb{R} \rightarrow \{-1, 1\}$  to denote the sign function, where  $\text{sgn}(y) = 1$  if  $y \geq 0$  and  $-1$  otherwise. We allow

univariate functions, such as  $\text{sgn}(\cdot)$  or general  $f: \mathbb{R} \rightarrow \mathbb{R}$ , to be applied to tensors in an element-wise manner.

## 2. Observation model

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

$$\mathcal{Y} = \Theta + \mathcal{E}, \quad (3)$$

where  $\Theta \in \mathbb{R}^{d_1 \times \dots \times d_K}$  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 in that the marginal distribution of noise entry  $\mathcal{E}(\omega)$  may depend on  $\omega$ . In general, we allow  $\mathcal{Y}(\omega)$  to take either continuous or discrete value in a bounded interval  $[-A, A]$ , where  $A > 0$ . Unless otherwise specified, we assume  $A = 1$  throughout the paper.

Our observation is an incomplete data tensor from (3), denoted  $\mathcal{Y}_\Omega$ , where  $\Omega \subset [d_1] \times \dots \times [d_K]$  is the index set of observed entries. We consider a general model on  $\Omega$  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  $\sum_{\omega \in [d_1] \times \dots \times [d_K]} \pi_\omega = 1$ . We assume the entries  $\omega$  in  $\Omega$  are i.i.d. draws with replacement from the full index set using distribution  $\Pi$ . The sampling rule will be denoted as  $\omega \sim \Pi$ .

## 3. Sign representable tensors

In this section, we develop sign representable tensor models for  $\Theta$  in (3). The algebraic and statistical characterization of sign tensor series provides the accuracy guarantee for our method.

### 3.1. Sign-rank and sign tensor series

Let  $\Theta$  be a continuous-valued tensor, and  $\text{sgn}(\Theta)$  be the corresponding sign pattern. 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  $\Theta \in \mathbb{R}^{d_1 \times \dots \times d_K}$  is the minimal rank among all tensors that share the same sign patterns as  $\Theta$ ; i.e.,

$$\text{srnk}(\Theta) = \min\{\text{rank}(\Theta'): \Theta' \simeq \Theta, \Theta' \in \mathbb{R}^{d_1 \times \dots \times d_K}\}.$$

The sign-rank is also called *support rank* (Cohn and Umans, 2013), *minimal rank* (Alon et al., 2016), and *nondeterministic rank* (De Wolf, 2003) in the field of combinatorics and information theory. 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  $\Theta$ . In particular,  $\text{srnk}(\Theta) = \text{srnk}(\text{sgn}(\Theta))$ .

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 family of low sign-rank tensors is strictly broader than usual low-rank tensors. This is because sign-rank is upper bounded by the usual rank. More generally,

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

$$\text{srnk}(\Theta) \leq \text{rank}(g(\Theta)).$$

Conversely, the sign-rank can be dramatically smaller than the usual rank, as we have shown in Section 1.1.

**Proposition 2** (Broadness). *For every order  $K \geq 2$  and dimension  $d$ , there exist order- $K$   $(d, \dots, d)$ -dimensional tensors  $\Theta$  such that  $\text{rank}(\Theta) = d$  and  $\text{srnk}(\Theta) = 2$ .*

Several common examples are provided in Appendix, in which we show 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 tensor family over the usual low-rank tensor 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  $\Theta$  is called  $(r, \pi)$ -sign representable, if the tensor  $(\Theta - \pi)$  has sign-rank bounded by  $r$ . A tensor  $\Theta$  is called  $r$ -sign (globally) representable, if  $\Theta$  is  $(r, \pi)$ -sign representable for all  $\pi \in [-1, 1]$ . The collection  $\{\text{sgn}(\Theta - \pi): \pi \in [-1, 1]\}$  is called the sign tensor series. We use  $\mathcal{P}_{\text{sgn}}(r) = \{\Theta: \text{srnk}(\Theta - \pi) \leq r \text{ for all } \pi \in [-1, 1]\}$  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 models are the two most popular tensor models (Anandkumar et al., 2014; Montanari and Sun, 2018; Kadmon and Ganguli, 2018; Cai et al., 2019). Let  $\Theta$  be a low-rank tensor with CP rank  $r$ . We see that  $\Theta$  belongs to the sign representable family, i.e.,  $\Theta \in \mathcal{P}_{\text{sgn}}(r + 1)$  (the constant 1 is due to  $\text{rank}(\Theta - \pi) \leq r + 1$ ). Similarly, Tucker low-rank tensors  $\Theta \in \mathcal{P}_{\text{sgn}}(r + 1)$ , where  $r = \prod_k r_k$  with  $r_k$  being the  $k$ -th mode Tucker rank of  $\Theta$ .

**Example 2** (Tensor block models (TBMs)). Tensor block model (Wang and Zeng, 2019; Chi et al., 2020) assumes a

checkerboard structure among tensor entries under marginal index permutation. The signal tensor  $\Theta$  takes at most  $r$  distinct values, where  $r$  is the total number of multiway blocks. Our model incorporates TBM because  $\Theta \in \mathcal{P}_{\text{sgn}}(r)$ .

**Example 3** (Generalized linear models (GLMs)). Let  $\mathcal{Y}$  be a binary tensor from a tensor logistic model (Wang and Li, 2020) with mean  $\Theta = \mathbb{E}(\mathcal{Y}) = \text{logit}(\mathcal{Z})$ , where  $\mathcal{Z}$  is a latent low-rank tensor. By definition,  $\Theta$  is a low-rank sign representable tensor, although the usual rank of  $\Theta$  is often high (see Section 1.1). Same conclusion holds for general exponential-family tensors with a (known) link function (Hong et al., 2020).

**Example 4** (Single index models (SIMs)). Single index model is a flexible semiparametric model initially proposed in economics (Robinson, 1988) and has recently been popular in high-dimensional statistics (Balabdaoui et al., 2019; Ganti et al., 2017; Alquier and Biau, 2013). We here extend the model to high-dimensional tensors  $\Theta$ . The SIM assumes the existence of a (unknown) monotonic function  $g: \mathbb{R} \rightarrow \mathbb{R}$  such that  $g(\Theta)$  has rank  $r$ . We see that  $\Theta$  belongs to the sign representable family; i.e.,  $\Theta \in \mathcal{P}_{\text{sgn}}(r + 1)$ .

**Example 5** (Min/Max hypergraphon). Graphon is a popular nonparametric model for networks (Chan and Airoldi, 2014; Xu, 2018), and we have extended the model for tensors in Section 1.1. Here we revisit the model for generality. Let  $\Theta$  be an order- $K$  tensor generated from the hypergraphon  $\Theta(i_1, \dots, i_K) = \log(1 + \max_k x_{i_k}^{(k)})$ , where  $x_{i_k}^{(k)} \sim \text{Unif}[0, 1]$  i.i.d. for all  $i_k \in [d_k]$  and  $k \in [K]$ . Every sign tensor  $\text{sgn}(\Theta - \pi)$  in the series of  $\pi \in [0, \log 2]$  is a block tensor with at most two blocks, so  $\Theta \in \mathcal{P}_{\text{sgn}}(2)$ .

The results extend to general min/max hypergraphons. Let  $g(\cdot)$  be a continuous univariate function with at most  $r \geq 1$  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  $\Theta$  generated from  $\Theta(i_1, \dots, i_K) = g(\max_k x_{i_k}^{(k)})$  belongs to  $\mathcal{P}_{\text{sgn}}(r + 1)$ . 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 crucially depends on the behavior of sign tensor series  $\text{sgn}(\Theta - \pi)$ . In this section, we show that weighted classification completely characterizes the sign tensors. The results bridge the algebraic and statistical properties of sign representable tensors, thereby providing the theoretical guarantee for our nonparametric algorithm (Figure 2).

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

a weighted classification objective

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

where  $\mathcal{Z} \in \mathbb{R}^{d_1 \times \dots \times d_K}$  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  $\pi$ . The objective reduces to usual classification loss in the special case when the data tensor is binary and the target level  $\pi = 0$ .

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

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

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  $\text{Risk}(\cdot)$  implicitly depends on  $\pi$ ; we suppress  $\pi$  when no confusion arises.

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

$$\begin{aligned} \text{Risk}(\bar{\Theta}) &= \inf\{\text{Risk}(\mathcal{Z}): \mathcal{Z} \in \mathbb{R}^{d_1 \times \dots \times d_K}\}, \\ &= \inf\{\text{Risk}(\mathcal{Z}): \text{rank}\mathcal{Z} \leq r\}. \end{aligned}$$

The results show that the sign tensor  $\text{sgn}(\Theta - \pi)$  optimizes the weighted classification risk. This fact suggests a practical procedure to estimate  $\text{sgn}(\Theta - \pi)$  via optimizing the empirical risk  $L(\mathcal{Z}, \bar{\mathcal{Y}}_\Omega)$ . The entry-specific weight incorporates the magnitude information in the classification, where entries far away from the target level are penalized more heavily in the objective.

In order to establish the recovery guarantee, we address the uniqueness (up to sign equivalence) of the optimizer for  $\text{Risk}(\cdot)$ . Essentially, this is a converse problem of Proposition 3 asking whether optimizing  $\text{Risk}(\cdot)$  is sufficient for recovering  $\text{sgn}(\Theta - \pi)$ . The local behavior of  $\Theta$  around  $\pi$  turn out to play a key role in the accuracy guarantee.

Some additional notation is needed. We use  $\mathcal{N} = \{\pi: \mathbb{P}_{\omega \sim \Pi}(\Theta(\omega) = \pi) \neq 0\}$  to denote the set of mass points of  $\Theta$  under  $\Pi$ . Assume there exists a constant  $C > 0$ , independent of tensor dimension, such that  $|\mathcal{N}| \leq C$ . Note that both  $\Pi$  and  $\Theta$  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  $d := \min_k d_k \rightarrow \infty$ .

**Assumption 1** ( $\alpha$ -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,

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



where  $\rho(\pi, \mathcal{N}) := \min_{\pi' \in \mathcal{N}} |\pi - \pi'|$  denotes the distance from  $\pi$  to the nearest point in  $\mathcal{N}$ . The largest possible  $\alpha = \alpha(\pi)$  is called the smoothness index at level  $\pi$ . We make the convention that  $\alpha = \infty$  if the set  $\{\omega: |\Theta(\omega) - \pi| \leq t\}$  has zero measure, implying few entries of  $\Theta$  around the level  $\pi$ . We call the tensor  $\Theta$  is  $\alpha$ -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  $\alpha$  quantifies the intrinsic hardness of recovering  $\text{sgn}(\Theta - \pi)$  from  $\text{Risk}(\cdot)$ . The value of  $\alpha$  depends on both the sampling distribution  $\omega \sim \Pi$  and the behavior of  $\Theta(\omega)$ . The recovery is more difficult at levels where the point mass concentrates (small  $\alpha$ ). A large value of  $\alpha > 1$  corresponds a plateau-type zero density of  $\Theta(\omega)$  around  $\pi$ , or equivalently, when the cumulative distribution function (CDF)  $F(\pi) := \mathbb{P}_{\omega \sim \Pi}[\Theta(\omega) \leq \pi]$  remains flat around  $\pi$ . A small value of  $\alpha < 1$  indicates the nonexistent (infinite) density at level  $\pi$ , or equivalently, when the CDF jumps at  $\pi$ . A typical case is  $\alpha = 1$  when the CDF has finite non-zero derivative in the vicinity of  $\pi$ . 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  $\Theta_1, \Theta_2$ , define the mean absolute error (MAE)

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

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

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

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

The result establishes the stability of recovering sign tensors  $\text{sgn}(\Theta - \pi)$  from optimizing the population risk (5). Furthermore, the bound immediately suggests the uniqueness of the optimizer for  $\text{Risk}(\cdot)$  up to a zero-measure set under  $\Pi$ , as long as  $\alpha \neq 0$ . We find that a higher value of  $\alpha$  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  $\Pi$  is the uniform sampling for now. Based on definition, the tensor block model is  $\infty$ -globally smooth. This is because there are finitely many elements in  $\mathcal{N}$ , i.e., the set of distinct block means in  $\Theta$ . Furthermore, we have  $\alpha = \infty$  for all  $\pi \notin \mathcal{N}$ , since the numerator in (6) is zero for all such  $\pi$ . The min/max hypergraphon with a  $r$ -degree polynomial function is 1-globally smooth, because  $\alpha = 1$  for all  $\pi$  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 1.2 (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  $\Theta$  into a series of weighted classifications. Specifically we propose a tensor estimator using the sign representation,

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

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

$$\hat{\mathcal{Z}}_\pi = \arg \min_{\mathcal{Z}: \text{rank} \mathcal{Z} \leq r} L(\mathcal{Z}, \text{sgn}(\mathcal{Y}_\Omega - \pi)). \quad (8)$$

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  $\Theta \in \mathcal{P}_{\text{sgn}}(r)$  and  $\Theta(\omega)$  is  $\alpha$ -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  $\mathcal{Y}_\Omega$ ,*

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

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 on the sample complexity and on the level  $\pi$ . 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.

Simulation	Signal Tensor $\Theta$	Rank	Sign Rank	Global $\alpha$	CDF of Tensor Entries	Noise
1	$\mathcal{C} \times \mathbf{M}_1 \times \mathbf{M}_2 \times \mathbf{M}_3$	$3^3$	2	$\infty$		Uniform $[-0.3, 0.3]$
2	$ \mathbf{a} \otimes \mathbf{1} \otimes \mathbf{1} - \mathbf{1} \otimes \mathbf{a} \otimes \mathbf{1} $	$d$	unknown	1		Normal $\mathcal{N}(0, 0.15)$
3	$\log(0.5 + \mathcal{Z}_{\max})$	$\geq d$	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  $\mathbf{M}_k \in \{0, 1\}^{d \times 3}$  to denote membership matrices,  $\mathcal{C} \in \mathbb{R}^{3 \times 3 \times 3}$  the block means,  $\mathbf{a} = \frac{1}{d}(1, 2, \dots, d)^T \in \mathbb{R}^d$ ,  $\mathcal{Z}_{\max}$  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.

**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  $\mathcal{Y}_{\Omega}$ ,*

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

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

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

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 trade-off. 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  $d_1 = \dots = d_K = d$ , our signal estimate achieves convergence rate

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

In contrast to earlier methods, our estimation accuracy applies to both low- and high-rank signal tensors  $\Theta$ . The rate depends on the sign complexity  $\Theta \in \mathcal{P}_{\text{sgn}}(r)$ , 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 nonparametric tensor completion is an immediate corollary of Theorem 3.

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

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

Our result improves earlier work (Ghadermarzy et al., 2019; Cai 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 tensor series, suggesting the optimality of our sample requirement.

## 4.2. Numerical implementation

This section addresses the practical implementation of our estimation (7) illustrated in Figure 2. Our sign representation of tensor estimate  $\Theta$  is a simple average of  $H$  sign 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 to our contexts. Following the common practice in classification, we replace the binary loss  $\ell(z, y) = |\text{sgnz} - \text{sgny}|$  with a surrogate loss  $F(m)$  as a continuous function of margin  $m \stackrel{\text{def}}{=} z \text{sgn}(y)$ . Examples of large-margin loss are hinge loss  $F(m) = (1 - m)_+$  for support vector machines, logistic loss  $F(m) = \log(1 + e^{-m})$  for important vector machines, and

**Algorithm 1** Nonparametric tensor completion

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

```

1: for  $\pi \in \mathcal{H} = \{-1, \dots, -\frac{1}{H}, 0, \frac{1}{H}, \dots, 1\}$  do
2:   Random initialization of tensor factors  $\mathbf{A}_k = [\mathbf{a}_1^{(k)}, \dots, \mathbf{a}_r^{(k)}] \in \mathbb{R}^{d_k \times r}$  for all  $k \in [K]$ .
3:   while not convergence do
4:     for  $k = 1, \dots, K$  do
5:       Update  $\mathbf{A}_k$  while holding others fixed:  $\mathbf{A}_k \leftarrow \arg \min_{\mathbf{A}_k \in \mathbb{R}^{d_k \times r}} \sum_{\omega \in \Omega} |\mathcal{Y}(\omega) - \pi| F(\mathcal{Z}(\omega) \text{sgn}(\mathcal{Y}(\omega) - \pi))$ ,
       where  $F(\cdot)$  is the large-margin loss, and  $\mathcal{Z} = \sum_{s \in [r]} \mathbf{a}_s^{(1)} \otimes \dots \otimes \mathbf{a}_s^{(K)}$  is a rank- $r$  tensor.
6:     end for
7:   end while
8:   Return  $\mathcal{Z}_\pi \leftarrow \sum_{s \in [r]} \mathbf{a}_s^{(1)} \otimes \dots \otimes \mathbf{a}_s^{(K)}$ .
9: end for
Output: Estimated signal tensor  $\hat{\Theta} = \frac{1}{2H+1} \sum_{\pi \in \mathcal{H}} \text{sgn}(\mathcal{Z}_\pi)$ .
    
```

nonconvex  $\psi$ -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  $\mathcal{Z} = \mathcal{Z}(\mathbf{A}_1, \dots, \mathbf{A}_K)$  where  $\mathbf{A}_k = [\mathbf{a}_1^{(k)}, \dots, \mathbf{a}_r^{(k)}] \in \mathbb{R}^{d_k \times r}$  are unknown factor matrices to optimize. We numerically solve (7) by optimizing one factor  $\mathbf{A}_k$  at a time while holding others fixed, and iterate until convergence. Each suboptimization 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  $d \in \{15, 20, \dots, 55, 60\}$ ,  $r = 2$ ,  $H = 10 + (d - 15)/5$  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.

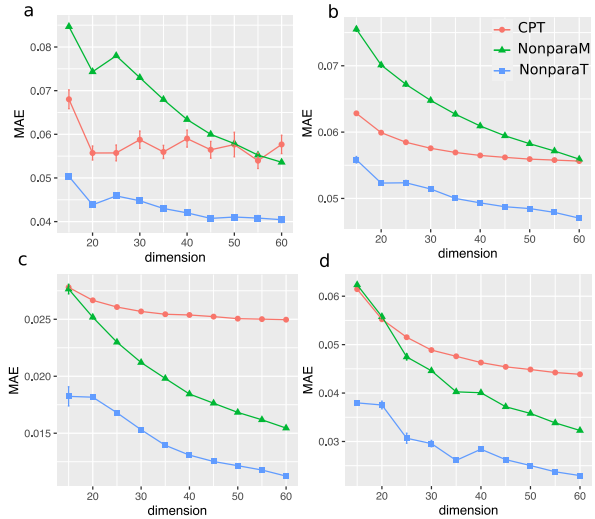


Figure 4. Estimation performance against tensor dimension. Panels (a)-(d) correspond to simulation models 1-4 in Table 3.

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  $|\Omega|/d^3$  from 0.3 to 1. It is seen that our method achieves the lowest error among all methods. We point out that the simulated models do not necessarily satisfy our method assumption. In particular, models 1 and 3 have unbounded noise, and model 2 has non-monotonic transformation functions. 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.

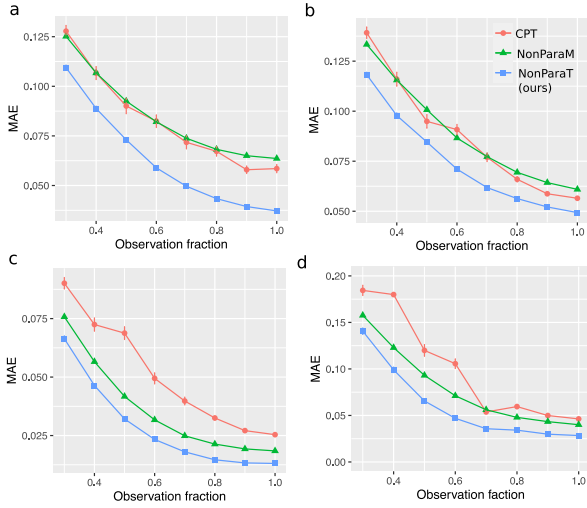


Figure 5. Completion performance against observation fraction. Panels (a)-(d) correspond to simulation models 1-4 in Table 3.

## 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.

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.

We next examine the estimated signal tensor  $\hat{\Theta}$  from 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

MRN-144 brain connectivity dataset			
Method	$r = 6$	$r = 9$	$r = 12$
NonparaT (Ours)	<b>0.14(0.003)</b>	<b>0.12(0.004)</b>	<b>0.12(0.007)</b>
CPT	0.23(0.006)	0.22(0.004)	0.21(0.004)

NIPS dataset			
Method	$r = 6$	$r = 9$	$r = 12$
NonparaT (Ours)	<b>0.16(0.002)</b>	<b>0.15(0.001)</b>	<b>0.14(0.001)</b>
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  $H = 20$ .

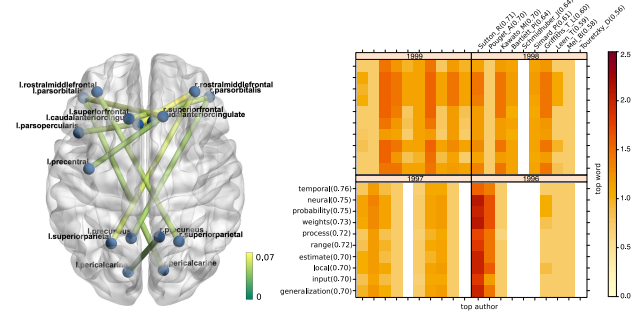


Figure 6. Estimated signal tensor from brain connectivity data (a) and NIPS data (b).

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 pattern 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 *temporal* 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 representations. Statistical accuracies are established, and we demonstrate the outperformance of our approach compared to other methods. The work unlocks several directions of future research. While we have provided numerical evidence for the success of nonconvex approach, the interplay between computational efficiency and statistical accuracy remains an interesting future problem.



## 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.
- 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. Airolidi (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 Wolf, R. (2003). Nondeterministic quantum query and communication complexities. *SIAM Journal on Computing* 32(3), 681–699.
- Filipović, M. and A. Jukić (2015). Tucker factorization with missing data with application to low-rank tensor completion. *Multidimensional systems and signal processing* 26(3), 677–692.
- 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.
- 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.
- Kadmon, J. and S. Ganguli (2018). Statistical mechanics of low-rank tensor decomposition. *Advances in Neural Information Processing Systems* 31, 8201–8212.
- Karbasi, A. and S. Oh (2012). Robust localization from incomplete local information. *IEEE/ACM Transactions on Networking* 21(4), 1131–1144.
- 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.
- 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*.