

---

# Beyond the Signs: Nonparametric Tensor Completion via Learning Reduction

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

1 We consider the problem of tensor estimation from noisy observations with possibly  
2 missing entries. A nonparametric approach to tensor completion is developed based  
3 on a new model which we coin as sign representable tensors. The model represents  
4 the signal tensor of interest using a series of structured sign tensors. Unlike earlier  
5 methods, the sign series representation effectively addresses both low- and high-  
6 rank signals, while encompassing many existing tensor models—including CP  
7 models, Tucker models, single index models, structured tensors with repeating  
8 entries—as special cases. We show tensor estimation problem can be reduced to a  
9 series of classification tasks with carefully-specified weights. Excess risk bounds,  
10 estimation error rates, and sample complexities are established. We demonstrate  
11 the outperformance of our approach over previous methods on two datasets, one  
12 on human brain connectivity networks and the other on topic data mining.

## 13 1 Introduction

14 Higher-order tensors have recently received much attention in enormous fields including social  
15 networks [2], neuroscience [31], and genomics [23]. Tensor methods provide effective representation  
16 of the hidden structure in multiway data. In this paper we consider the signal plus noise model,

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

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

20 Q1 [Nonparametric tensor estimation]. How to flexibly estimate  $\Theta$  under a wide range of structures,  
21 including both low-rankness and high-rankness?

22 Q2 [Complexity of tensor completion]. How many observed tensor entries do we need to consistently  
23 estimate the signal  $\Theta$ ?

24 **Inadequacies of low-rank models.** The signal plus noise model (2) is popular in tensor literature.  
25 Existing methods estimate the signal tensor based on low-rankness of  $\Theta$  [24, 28]. Common low-rank  
26 models include Canonical Polyadic (CP) tensors [21], Tucker tensors [10], and block tensors [33].  
27 While these methods have shown great success in theory, tensors in applications often violate the  
28 low-rankness. Here we provide two examples to illustrate the limitation of classical models.

29 The first example reveals the sensitivity of tensor rank to order-preserving transformations. Let  
30  $\mathcal{Z} \in \mathbb{R}^{30 \times 30 \times 30}$  be an order-3 tensor with CP rank( $\mathcal{Z}$ ) = 3 (formal definition is deferred to end  
31 of this section). Suppose a monotonic transformation  $f(z) = (1 + \exp(-cz))^{-1}$  is applied to  $\mathcal{Z}$   
32 entrywise, and we let the signal  $\Theta$  in model (1) be the tensor after transformation. Figure 1a plots the  
33 numerical rank (see Section ?? in Appendix) of  $\Theta$  versus  $c$ . As we see, the rank increases rapidly with  
34  $c$ , rendering traditional low-rank tensor methods ineffective in the presence of mild order-preserving

nonlinearities. In digital processing [15] and genomics analysis [23], the tensor of interest often undergoes unknown transformation prior to measurements. The sensitivity to transformation makes the low-rank model less desirable in practice. **larger spacing between two figures.**

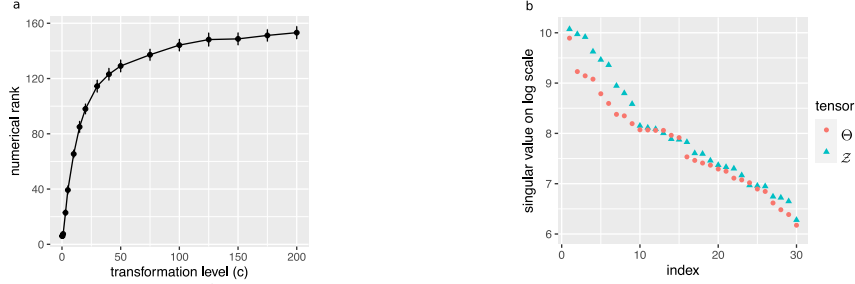


Figure 1: (a) Numerical rank of  $\Theta$  versus  $c$  in the first example. (b) Top  $d = 30$  tensor singular values in the second example.

The second example demonstrates the inadequacy of classical low-rankness in representing special structures. 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  $i, j, k \in \{1, \dots, d\}$ . The matrix analogy of  $\Theta$  was studied by [7] in graphon analysis. In this case neither  $\Theta$  nor  $\mathcal{Z}$  is low-rank; in fact, the rank is no smaller than the dimension  $d$  as illustrated in Figure 1b. Again, classical low-rank models fail to address this type of tensor structure.

In the above and many other examples, the signal tensors  $\Theta$  is of high rank. Classical low-rank models will miss these important structures. The observations have motivated us to develop more flexible tensor modeling.

**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 sign series representation of the signal tensor, 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.

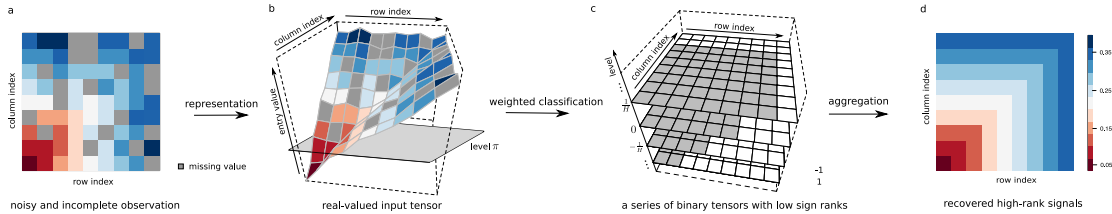


Figure 2: Illustration of our method. For visualization purpose, we plot an order-2 tensor (a.k.a. matrix); similar procedure applies to higher-order tensors. (a): a noisy and incomplete tensor input. (b) and (c): main steps of estimating sign tensor series  $\text{sgn}(\Theta - \pi)$  for  $\pi \in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

Statistically, the problem of high-rank tensor estimation is challenging. Existing estimation theory [2, 28, 6] 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 model literature.

From computational perspective, optimizations regarding tensors are in general NP-hard. Fortunately, tensors sought in applications are specially-structured, for which a number of efficient algorithms are available [15, 32, 18]. Our high-rank tensor estimate is provably reducible to a series of classifications, and its divide-and-conquer nature facilitates efficient 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 [14] and subspace clustering [29, 12]. However, the problem for high-rank tensors is more challenging, because the tensor rank often exceeds the dimension when order  $K \geq 3$  [3]. 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  $\text{sgn}(\cdot): \mathbb{R} \rightarrow \{-1, 1\}$  to denote the sign function, where  $\text{sgn}(y) = 1$  if  $y \geq 0$  and  $-1$  otherwise. We allow univariate functions, such as  $\text{sgn}(\cdot)$  and general  $f: \mathbb{R} \rightarrow \mathbb{R}$ , to be applied to tensors in an element-wise manner. We denote  $a_n \lesssim b_n$  if  $\lim_{n \rightarrow \infty} a_n/b_n \leq c$  for some constant  $c \geq 0$ . We use the shorthand  $[n]$  to denote the  $n$ -set  $\{1, \dots, n\}$  for  $n \in \mathbb{N}_+$ . Let  $\Theta \in \mathbb{R}^{d_1 \times \dots \times d_K}$  denote an order- $K$   $(d_1, \dots, d_K)$ -dimensional tensor, and  $\Theta(\omega) \in \mathbb{R}$  denote the tensor entry indexed by  $\omega \in [d_1] \times \dots \times [d_K]$ . An event  $E$  is said to occur “with very high probability” if  $\mathbb{P}(E)$  tends to 1 faster than any polynomial of tensor dimension  $d := \min_k d_k \rightarrow \infty$ . The CP decomposition [21] is defined by  $\Theta = \sum_{s=1}^r \lambda_s \mathbf{a}_s^{(1)} \otimes \dots \otimes \mathbf{a}_s^{(K)}$ , where  $\lambda_1 \geq \dots \geq \lambda_r > 0$  are tensor singular values,  $\mathbf{a}_s^{(k)} \in \mathbb{R}^{d_k}$  are norm-1 tensor singular vectors, and  $\otimes$  denotes the outer product of vectors. The minimal  $r \in \mathbb{N}_+$  in the CP decomposition is called the tensor rank, denoted  $\text{rank}(\Theta)$ .

## 2 Model and proposal overview

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

$$\mathcal{Y} = \Theta + \mathcal{E}, \quad (2)$$

where  $\Theta \in \mathbb{R}^{d_1 \times \dots \times d_K}$  is an unknown signal tensor of interest, and  $\mathcal{E}$  is a noise tensor consisting of zero-mean, 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  $\omega$ . For simplicity, we assume the noise is bounded; the extension to a sub-Gaussian noise is provided in the Appendix. Here, we assume the range of  $\mathcal{Y}$  is the bounded interval  $[-1, 1]$  for cleaner exposition. Our observation is an incomplete data tensor from (2), 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  $\Pi = \{p_\omega\}$  be an arbitrarily predefined probability distribution over the full index set with  $\sum_{\omega \in [d_1] \times \dots \times [d_K]} p_\omega = 1$ . Assume that the entries  $\omega$  in  $\Omega$  are i.i.d. draws with replacement from distribution  $\Pi$ . The sampling rule is denoted as  $\omega \sim \Pi$ .

**Initiation.** 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  $\Theta$  makes the estimation challenging. Now let us examine the sign of the  $\pi$ -shifted signal  $\text{sgn}(\Theta - \pi)$  for any given  $\pi \in [-1, 1]$ . It turns out that, these sign tensors share the same sign patterns as low-rank tensors. Indeed, the signal tensor in the first example 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 the second example has the same sign pattern as a rank-2 tensor, since  $\text{sgn}(\Theta(i, j, k) - \pi) = \text{sgn}(\max(i, j, k) - d(e^\pi - 1))$  (see Example 5 in Section 3).

**Overview.** 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 propose to estimate the signal tensor  $\Theta$  by taking the average over sign tensors

$$\hat{\Theta} = \frac{1}{2H+1} \sum_{\pi \in \mathcal{H}} \text{sgn}(\hat{\mathcal{Z}}_\pi), \quad \text{where} \quad \hat{\mathcal{Z}}_\pi = \arg \min_{\text{low rank tensor } \mathcal{Z}} \text{Weighted-Loss}(\text{sgn}(\mathcal{Z}), \text{sgn}(\mathcal{Y}_\Omega - \pi)),$$

where  $\text{sgn}(\hat{\mathcal{Z}}_\pi) \in \{-1, 1\}^{d_1 \times \dots \times d_K}$  is a sign tensor that estimates sign signal,  $\text{sgn}(\Theta - \pi)$ , at a series of  $\pi \in \mathcal{H} = \{-1, \dots, -1/H, 0, 1/H, \dots, 1\}$ , and  $\text{Weighted-Loss}(\cdot, \cdot)$  denotes a carefully-designed classification objective function which will be described in later sections. To obtain  $\text{sgn}(\hat{\mathcal{Z}}_\pi)$  for a given  $\pi$ , we propose to dichotomize the data tensor into sign tensor,  $\text{sgn}(\mathcal{Y}_\Omega - \pi)$ , and estimate the sign signal by performing weighted classification based on objective  $\text{Weighted-Loss}(\cdot, \cdot)$ . Our approach is built on the nonparametric sign representation of signal tensors. We show that a careful aggregation 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. Unlike traditional methods, the sign representation is guaranteed to recover both low- and high-rank signals. The method enjoys statistical effectiveness and computational efficiency.

## 3 Oracle properties of sign representable tensors

This section develops sign representable tensor models for  $\Theta$  in (2). We characterize the algebraic and statistical properties of sign tensor series, which serves the foundation for our method.

### 3.1 Sign-rank and sign tensor series

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

**Definition 1** (Sign-rank). The sign-rank of a tensor  $\Theta \in \mathbb{R}^{d_1 \times \dots \times d_K}$  is defined by the minimal rank among all tensors that share the same sign pattern 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 [9], minimal rank [1], and nondeterministic rank [11]. Earlier work defines sign-rank for binary-valued tensors; we extend the notion to continuous-valued tensors. Note that the sign-rank concerns only the sign pattern but discards the magnitude information of  $\Theta$ . In particular,  $\text{srnk}(\Theta) = \text{srnk}(\text{sgn}(\Theta))$ .

Like most tensor problems [20], determining the sign-rank for a general tensor is NP hard [1]. Fortunately, tensors arisen in applications often possess special structures that facilitate analysis. By definition, the sign-rank is upper bounded by the tensor rank. More generally, we have the following upper bounds.

**Proposition 1** (Upper bounds of the sign-rank). *For any strictly monotonic function  $g: \mathbb{R} \rightarrow \mathbb{R}$  with  $g(0) = 0$ , the following holds true,  $\text{srnk}(\Theta) \leq \text{rank}(g(\Theta))$ .*

The sign-rank can be much smaller than the tensor rank, as we have shown in Section 1.

**Proposition 2** (Broadness). *For every order  $K \geq 2$  and dimension  $d$ , there exist tensors  $\Theta \in \mathbb{R}^{d \times \dots \times d}$  such that  $\text{rank}(\Theta) \geq d$  but  $\text{srnk}(\Theta - \pi) \leq 2$  for all  $\pi \in \mathbb{R}$ .*

We provide additional examples in the 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. Propositions 1 and 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”.

**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  $\mathcal{P}_{\text{sgn}}(r) = \{\Theta : \max_{\pi \in [-1, 1]} \text{srnk}(\Theta - \pi) \leq r\}$  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 structured tensors with repeating entries.

**Example 1** (CP/Tucker low-rank models). The CP and Tucker low-rank tensors are the two most popular tensor models [25]. 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  $\text{rank}(\Theta - \pi) \leq r + 1$ ). Similar results hold for 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 [33, 8] 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 logistic model [32] with mean  $\Theta = \text{logit}(\mathcal{Z})$ , where  $\mathcal{Z}$  is a latent low-rank tensor. Notice that  $\Theta$  itself may be high-rank (see Section 1). By definition,  $\Theta$  is a low-rank sign representable tensor. Same conclusion holds for general exponential-family models with a (known) link function [22].

**Example 4** (Single index models (SIMs)). Single index model is a flexible semiparametric model proposed in economics [30] and high-dimensional statistics [4, 13]. 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** (Structured tensors with repeating entries). Here we revisit the model introduced in Section 1. Let  $\Theta$  be an order- $K$  tensor with entries  $\Theta(i_1, \dots, i_K) = \log(1 + \max_k x_{i_k}^{(k)})$ , where  $x_{i_k}^{(k)}$  are given number in  $[0, 1]$  for all  $i_k \in [d_k], k \in [K]$ . We conclude that  $\Theta \in \mathcal{P}_{\text{sgn}}(2)$ , because the sign tensor  $\text{sgn}(\Theta - \pi)$  with an arbitrary  $\pi \in (0, \log 2)$  is a block tensor with at most two blocks (see Figure 2c). The results extend to general structured tensors with repeating entries. Let  $g(\cdot)$  be a polynomial of degree  $r$  and the associated tensor  $\Theta$  with entries  $\Theta(i_1, \dots, i_K) = g(\max_k x_{i_k}^{(k)})$ . We have  $\Theta \in \mathcal{P}_{\text{sgn}}(2r)$  (see proofs in the Appendix).

### 3.2 Statistical characterization of sign tensors via weighted classification

In this section, we show that sign tensors are characterized by weighted classification. The results bridge the algebraic and statistical properties of sign representable tensors.

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  $\omega \in \Omega$ . We propose a weighted classification objective function

$$L(\mathcal{Z}, \bar{\mathcal{Y}}_\Omega) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \underbrace{|\bar{\mathcal{Y}}(\omega)|}_{\text{weight}} \times \underbrace{|\text{sgn} \mathcal{Z}(\omega) - \text{sgn} \bar{\mathcal{Y}}(\omega)|}_{\text{classification loss}}, \quad (3)$$

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 entry-specific weights incorporate the magnitude information into classification, where entries far away from the target level are penalized more heavily in the objective. In the special case of binary tensor  $\mathcal{Y} \in \{-1, 1\}^{d_1 \times \dots \times d_K}$  and target level  $\pi = 0$ , the loss (3) reduces to usual classification loss.

Our proposed weighted classification function (3) is important for characterizing  $\text{sgn}(\Theta - \pi)$ . Define the weighted classification risk

$$\text{Risk}(\mathcal{Z}) = \mathbb{E}_{\mathcal{Y}_\Omega} L(\mathcal{Z}, \bar{\mathcal{Y}}_\Omega), \quad (4)$$

where the expectation is taken with respect to  $\mathcal{Y}_\Omega$  under model (2) and the sampling distribution  $\omega \sim \Pi$ . 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 (2) with  $\Theta \in \mathcal{P}_{\text{sgn}}(r)$ . Then, for all  $\Theta$  that are sign equivalent to  $\text{sgn}(\Theta - \pi)$ ,*

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

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 empirical risk optimization of  $L(\mathcal{Z}, \bar{\mathcal{Y}}_\Omega)$ . In order to establish the recovery guarantee, we shall address the uniqueness (up to sign equivalence) for the optimizer of  $\text{Risk}(\cdot)$ . The local behavior of  $\Theta$  around  $\pi$  plays a key role in the accuracy.

Some additional notation is needed for stating the results in full generality. Let  $d_t = \prod_{k=1}^K d_k$  denote the total number of tensor entries, and  $\Delta s = 1/d_t$  a small tolerance. We quantify distribution of entries in tensor  $\Theta$  using a pseudo density, i.e., histogram with bin width  $2\Delta s$ . Specifically, let  $G(\pi) := \mathbb{P}_{\omega \sim \Pi}[\Theta(\omega) \leq \pi]$  denote the cumulative distribution function (CDF) of  $\Theta(\omega)$  under  $\omega \sim \Pi$ . We partition  $[-1, 1] = \mathcal{N} \cup \mathcal{N}^c$ , where  $\mathcal{N}$  consists of levels whose pseudo density based on  $2\Delta s$ -bin is asymptotically unbounded; i.e.,

$$\mathcal{N} = \left\{ \pi \in [-1, 1] : \frac{G(\pi + \Delta s) - G(\pi - \Delta s)}{\Delta s} \geq C \right\}, \text{ for some universal constant } C > 0,$$

and  $\mathcal{N}^c$  otherwise. Both  $\Theta$  and its induced CDF  $G$  implicitly depend on the tensor dimension.

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

$$\sup_{\Delta s \leq t < \rho(\pi, \mathcal{N})} \frac{G(\pi + t) - G(\pi - t)}{t^\alpha} \leq c, \quad (5)$$

where  $\rho(\pi, \mathcal{N}) := \min_{\pi' \in \mathcal{N}} |\pi - \pi'| + \Delta s$  denotes the adjusted distance from  $\pi$  to the nearest point in  $\mathcal{N}$ . The largest possible  $\alpha = \alpha(\pi)$  in (5) is called the smoothness index at level  $\pi$ . We make the convention that  $\alpha = \infty$  if the numerator in (5) is zero. A tensor  $\Theta$  is called  $\alpha$ -globally smooth, if (5) holds with global constants  $\alpha > 0, c > 0$  for all  $\pi \notin \mathcal{N}$ .

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 easier at levels where points are less concentrated around  $\pi$  with a large value of  $\alpha > 1$ , or equivalently, when  $G(\pi)$  remains almost flat around  $\pi$ . A small value of  $\alpha < 1$  indicates the nonexistent (infinite) density at level  $\pi$ , or equivalently, when the  $G(\pi)$  jumps by greater than the tolerance  $\Delta s$  at  $\pi$ . Table 1 illustrates the  $G(\pi)$  for various models of  $\Theta$  (see Section 5).

We now reach the main theorem in this section. For two tensors  $\Theta_1, \Theta_2$ , define the mean absolute error (MAE) as  $\text{MAE}(\Theta_1, \Theta_2) \stackrel{\text{def}}{=} \mathbb{E}_{\omega \sim \Pi} |\Theta_1(\omega) - \Theta_2(\omega)|$ .

**Theorem 1** (Identifiability). *Assume  $\Theta \in \mathcal{P}_{\text{sgn}}(r)$  is  $\alpha$ -globally smooth. Then, for all  $\pi \notin \mathcal{N}$  and tensors  $\bar{\Theta} \simeq \text{sgn}(\Theta - \pi)$ , we have*

$$\text{MAE}(\text{sgn}\mathcal{Z}, \text{sgn}\bar{\Theta}) \lesssim C(\pi) [\text{Risk}(\mathcal{Z}) - \text{Risk}(\bar{\Theta})]^{\alpha/(\alpha+1)} + \Delta s, \quad \text{for all } \mathcal{Z} \in \mathbb{R}^{d_1 \times \dots \times d_K},$$

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

The result establishes the recovery stability of sign tensors  $\text{sgn}(\Theta - \pi)$  using optimization with population risk (4). The bound immediately shows the uniqueness of the optimizer for  $\text{Risk}(\cdot)$  up to a  $\Delta s$ -measure set under  $\Pi$ . We find that a higher value of  $\alpha$  implies more stable recovery, as intuition would suggest. Similar results hold for optimization with sample risk (3) (see Section 4).

We conclude this section by applying Assumption 1 to the examples described in Section 3.1. For simplicity, suppose  $\Pi$  is the uniform sampling. The tensor block model is  $\infty$ -globally smooth. This is because the set  $\mathcal{N}$  consists of finite  $2\Delta s$ -bin's covering the distinct block means in  $\Theta$ . Furthermore, we have  $\alpha = \infty$  for all  $\pi \notin \mathcal{N}$ , since the numerator in (5) is zero. Similarly, the high-rank tensor  $\Theta(i_1, \dots, i_K) = \log(1 + \max_{\ell=1, \dots, K} i_\ell/d)$  is  $\infty$ -globally smooth because  $\alpha = \infty$  for all  $\pi$  except those in  $\mathcal{N}$ , where  $\mathcal{N}$  collects  $d$  many  $2\Delta s$ -bin's covering  $\log(1 + i/d)$  for  $i = 1, \dots, d$ .

## 4 Nonparametric tensor completion via sign series

In previous sections we have established the sign series representation and its relationship to classification. In this section, we present our learning reduction proposal (Figure 2) in details. We provide the estimation error bound and address the empirical implementation of the method.

### 4.1 Statistical error and sample complexity

Given a noisy incomplete tensor observation  $\mathcal{Y}_\Omega$  from model (2), we cast the problem of estimating  $\Theta$  into a series of weighted classifications. Specifically we propose the tensor estimate using the sign representation,

$$\hat{\Theta} = \frac{1}{2H+1} \sum_{\pi \in \mathcal{H}} \text{sgn}\hat{\mathcal{Z}}_\pi, \quad \text{where } \hat{\mathcal{Z}}_\pi = \arg \min_{\mathcal{Z}: \text{rank}\mathcal{Z} \leq r} L(\mathcal{Z}, \mathcal{Y}_\Omega - \pi), \quad (6)$$

where  $\mathcal{H} = \{-1, \dots, -1/H, 0, 1/H, \dots, 1\}$  is a series of level to aggregate,  $L(\cdot, \cdot)$  denotes the weighted classification objective defined in (3), and the rank constraint on  $\mathcal{Z}$  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 [19].

The next theorem establishes the statistical convergence for the sign tensor estimate (6), which is an important ingredient for the final signal tensor estimate  $\hat{\Theta}$  in (6).

**Theorem 2** (Sign tensor estimation). *Suppose  $\Theta \in \mathcal{P}_{\text{sgn}}(r)$  and  $\Theta(\omega)$  is  $\alpha$ -globally smooth under  $\omega \sim \Pi$ . Let  $\hat{\mathcal{Z}}_\pi$  be the estimate in (6),  $d_{\max} = \max_{k \in [K]} d_k$ , and  $t_d = d_{\max} r / |\Omega| \lesssim 1$ . Then, for all  $\pi \notin \mathcal{N}$ , with very high probability over  $\mathcal{Y}_\Omega$ ,*

$$\text{MAE}(\text{sgn}\hat{\mathcal{Z}}_\pi, \text{sgn}(\Theta - \pi)) \lesssim t_d^{\alpha/(\alpha+2)} + \frac{1}{\rho^2(\pi, \mathcal{N})} t_d. \quad (7)$$

Theorem 2 provides the error bound for the sign tensor estimation. Compared to the population results in Theorem 1, we explicitly reveal the dependence of accuracy on the sample complexity and the level  $\pi$ . The result demonstrates the polynomial decay of sign errors with  $|\Omega|$ . Our sign estimate achieves consistent recovery using as few as  $\tilde{O}(d_{\max} r)$  noisy entries.

Recall that  $\mathcal{N}$  collects the levels for which the sign tensor is possibly nonrecoverable. Let  $|\mathcal{N}|$  be the covering number of  $\mathcal{N}$  with  $2\Delta s$ -bin's, i.e.,  $|\mathcal{N}| = \lceil \mu(\mathcal{N})/2\Delta s \rceil$ , where  $\mu$  is the Lebesgue measure.

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 (6) and  $t_d = d_{\max} r / |\Omega|$ . With very high probability over  $\mathcal{Y}_\Omega$ ,*

$$\text{MAE}(\hat{\Theta}, \Theta) \lesssim t_d^{\alpha/(\alpha+2)} + \frac{1 + |\mathcal{N}|}{H} + H t_d. \quad (8)$$

*In particular, setting  $H \asymp (1 + |\mathcal{N}|)^{1/2} t_d^{-1/2}$  yields the error bound  $\text{MAE}(\hat{\Theta}, \Theta) \lesssim t_d^{\alpha/(\alpha+2)} + t_d^{1/2} (1 + |\mathcal{N}|)^{1/2}$ .*

Theorem 3 demonstrates the convergence rate of our tensor estimation. The bound (8) reveals three sources of errors: the estimation error for 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 (8) is generally no better than the corresponding sign error (7). 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

$$\text{MAE}(\hat{\Theta}, \Theta) \lesssim \sqrt{r} d^{-\alpha(K-1)/(\alpha+2)} + \sqrt{r} d^{-(K-1)/2} (1 + |\mathcal{N}|)^{1/2}.$$

Compared to earlier methods, our estimation accuracy applies to both low- and high-rank signal tensors. 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). Our result also reveals that the convergence becomes favorable as the order of data tensor increases.

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

**Example 2** (TBMs). Consider a tensor block model with  $r$  multiway blocks. Our result implies a rate  $\mathcal{O}(d^{-(K-1)/2})$  by taking  $\alpha = \infty$  and  $|\mathcal{N}| \leq r^K$ . This rate agrees with the previous root-mean-square error (RMSE) for block tensor estimation [33].

**Example 3** (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 uniformly bounded as  $d \rightarrow \infty$ . Applying our results with  $\alpha = 1$  and finite  $|\mathcal{N}|$  yields  $\mathcal{O}(d^{-(K-1)/3})$ . This rate is slightly slower than the parametric RMSE rate [36, 32], as expected. The reason is that our estimate remains valid for unknown  $g$  and general high-rank tensors. The nonparametric rate is the price one has to pay for not knowing the form  $\Theta = g(\mathcal{Z})$  as a priori.

**Example 4** (SIMs). The earlier example has shown the nonparametric rate  $\mathcal{O}(d^{-(K-1)/3})$  when applying our method to single index tensor model. In the matrix case with  $K = 2$ , our result yields error rate  $\mathcal{O}(d^{-1/3})$ , which is faster compared to the RMSE rate  $\mathcal{O}(d^{-1/4})$  obtained by [14].

**Example 5** (Structured tensors with repeating entries). We consider a more general model than that in Section 1. Consider a  $r$ -sign representable tensor  $\Theta \in \mathcal{P}_{\text{sgn}}(r)$  with at most  $d$  distinct entries with repetition pattern. Applying our results with  $\alpha = \infty$  and  $|\mathcal{N}| = d$  yields the rate  $\mathcal{O}(d^{-(K-2)/2})$ .

The following corollary reveal the sample complexity for nonparametric tensor completion.

**Corollary 1** (Sample complexity for nonparametric completion). *Under the same conditions of Theorem 3 with  $\alpha \neq 0$  and bounded  $|\mathcal{N}|$ , with high probability over  $\mathcal{Y}_\Omega$ ,*

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

Our result improves earlier work [35, 16, 26] by allowing both low- and high-rank signals. Interestingly, the sample requirements depend only on the sign complexity  $d_{\max} r$  but not the nonparametric complexity  $\alpha$ . Note that  $\tilde{\mathcal{O}}(d_{\max} r)$  roughly matches the degree of freedom of sign tensors, suggesting the optimality of our sample requirements.

## 4.2 Numerical implementation via learning reduction

**Cut details. Emphasize learning reduction. Meta algorithm is a simple average of  $2H + 1 \asymp \text{poly}(d)$  terms. Base algorithm can be chosen from a variety of polynomial tensor solvers...Skip the detailed optimization.** We take a learning reduction approach to estimate the signal in a divide-and-conquer fashion (Figure 2). The mega algorithm takes the average of  $2H + 1$  sign tensors based on the sign representations of the signal tensor. The base algorithm estimates sign tensors  $\mathcal{Z}_\pi$  for the series  $\pi \in \mathcal{H}$  through parallel implementation. The estimate enjoys low computational cost similar to a single sign tensor estimation.

The base algorithm reduces to binary tensor decomposition with a weighted classification loss. A number of algorithms have been developed for this problem [15, 32, 22]. 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  $F(m)$  using a continuous function of margin  $m := z\text{sgn}(y)$ . Examples of large-margin loss are hinge loss  $F(m) = (1 - m)_+$ , logistic loss  $F(m) = \log(1 + e^{-m})$ , and nonconvex  $\psi$ -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 [5].

---

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

```

---

**Cut** The rank constraints in the optimization (6) have been extensively studied in literature. Recent developments involve convex norm relaxation [15] and nonconvex optimization [32, 18]. 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 (??) of  $\mathcal{Z} = \mathcal{Z}(\mathbf{A}_1, \dots, \mathbf{A}_K)$  to optimize the unknown factor matrices  $\mathbf{A}_k = [\mathbf{a}_1^{(k)}, \dots, \mathbf{a}_r^{(k)}] \in \mathbb{R}^{d_k \times r}$ , where we choose to collect tensor singular values into  $\mathbf{A}_K$ . We numerically solve (6) by optimizing one factor  $\mathbf{A}_k$  at a time while holding others fixed. Each suboptimization reduces to a convex optimization with a low-dimensional decision variable. Following common practice in tensor optimization [2, 22], we run the optimization from multiple initializations to locate a final estimate with the lowest objective value. The full procedure is described in Algorithm 4.2.

## 5 Simulations

In this section, we compare our nonparametric tensor method (**NonParaT**) with two alternative approaches: low-rank tensor CP decomposition (**CPT**), and the matrix version 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 1. The simulation covers a wide range of complexity, including block tensors, transformed low rank tensors, structured tensors with repeating entries. 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 4.2. For **NonParaM**, we apply Algorithm 4.2 to each of the three unfolded matrices and report the average error. All summary statistics are averaged across 30 replicates.



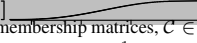
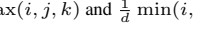
Simulation	Signal Tensor $\Theta$	Rank	Sign Rank	$\alpha$	$ \mathcal{N} $	CDF of Tensor Entries	Noise
1	$\mathcal{C} \times \mathbf{M}_1 \times \mathbf{M}_2 \times \mathbf{M}_3$	$3^3$	$\leq 3^3$	$\infty$	$\leq 3^3$		Uniform $[-0.3, 0.3]$
2	$[\mathbf{a} \otimes \mathbf{1} \otimes \mathbf{1} - \mathbf{1} \otimes \mathbf{a} \otimes \mathbf{1}]$	$d$	$\leq 3$	1	0		Normal $\mathcal{N}(0, 0.15)$
3	$\log(0.5 + \mathcal{Z}_{\max})$	$\geq d$	2	$\infty$	$d$		Uniform $[-0.1, 0.1]$
4	$2.5 - \exp(\mathcal{Z}_{\min}^{1/3})$	$> d$	2	$\infty$	$d$		Normal $\mathcal{N}(0, 0.15)$

Table 1: Simulation models used for comparison. We use  $\mathbf{M}_k \in \{0, 1\}^{d \times d \times d}$  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.



Figure 3 compares the estimation error under full observation. The MAE decreases with tensor dimension for all three methods. We find that 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 therefore, the two nonparametric methods **NonParaT** and **NonparaM** exhibit the greater advantage for signal recovery. Figure ?? 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. We find that **NonParaT** achieves the lowest error among all methods. Our simulation covers a reasonable range of complexities; for example, model 1 has  $3^3$  jumps in the CDF of signal  $\Theta$ , and models 2 and 4 have unbounded noise. Nevertheless, our method shows good performance in spite of model misspecification. This robustness is appealing in practice because the structure of underlying signal tensor is often unknown. **keep only one figure. panels b and d only. Others in appendix.**

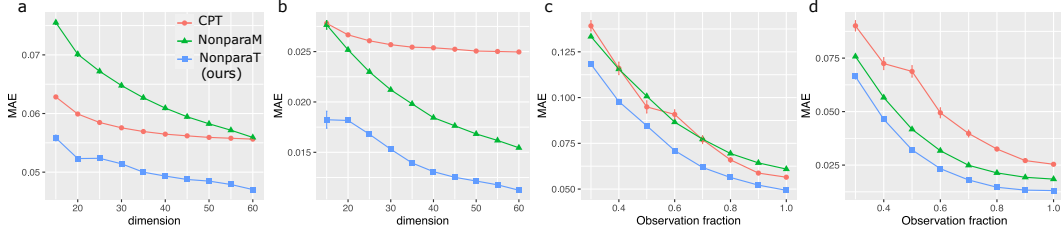


Figure 3: Estimation error versus tensor dimension. **red change**

## 6. Data applications

We apply our method to two tensor datasets, the MRN-114 human brain connectivity data [31], and NIPS data [17]. The brain 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 word occurrence counts in 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 pseudo-count 1. The resulting dataset is an order-3 tensor with entry representing the log counts of words by authors across years.

MRN-114 brain connectivity dataset			
Method	$r = 6$	$r = 9$	$r = 12$
<b>NonparaT (Ours)</b>	<b>0.14(0.001)</b>	<b>0.12(0.001)</b>	<b>0.12(0.001)</b>
CPT	0.23(0.006)	0.22(0.004)	0.21(0.006)
NIPS word occurrence dataset			
Method	$r = 6$	$r = 9$	$r = 12$
<b>NonparaT (Ours)</b>	<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 2: MAE comparison between our method **NonparaT** ( $H = 20$ ) and CP low-rank method (**CPT**) in the brain and NIPS data analysis. Standard errors are reported in parenthesis.

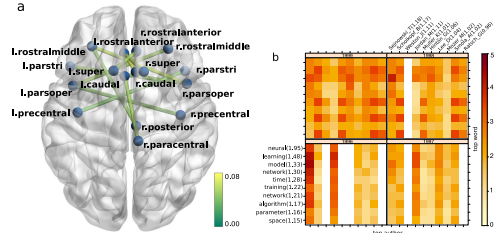


Figure 4: (a) top IQ-associated edges in the brain connectivity data. (b) top authors and words for years 1996-1999 in the NIPS data.

Table 2 compares the prediction accuracy of different methods. Reported MAEs are averaged over five runs of cross-validation, with 20% entries for testing and 80% for training. Our method substantially outperforms the low-rank CP method for every configuration under consideration. 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). 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 4a shows the results from brain connectivity dataset. We plot the top 10 brain edges based on regression analysis of denoised connection strengths against normalized IQ scores. We find that top connections are mostly inter-hemisphere edges, consistent with recent research on brain connectivity with intelligence [27, 31]. Figure 4b illustrates the results from NIPS data, where we plot the entries in  $\hat{\Theta}$  corresponding to top authors and most-frequent words (after excluding generic words such as *figure*, *results*, etc). The identified pattern is consistent with the active topics in the NIPS publication. Among the top words are *neural* (marginal mean = 1.95), *learning* (1.48), and *network* (1.21), whereas top authors are *T. Sejnowski* (1.18), *B. Scholkopf* (1.17), *M. Jordan* (1.11), and *G. Hinton* (1.06). We also find

364 strong heterogeneity among word occurrences across authors and years. For example, *training* and  
 365 *algorithm* are popular words for *B. Scholkopf* and *A. Smola* in 1998-1999, whereas *model* occurs  
 366 more often in *M. Jordan* and in 1996. The detected pattern and achieved accuracy demonstrate the  
 367 applicability of our method.

## 368 7 Conclusion

369 We have developed a tensor completion method that addresses both low- and high-rankness based on  
 370 sign series representation. Our work provides a nonparametric framework for tensor estimation, and  
 371 we obtain results previously impossible. We hope the work opens up new inquiry that allows more  
 372 researchers to contribute to this field.

## 373 References

- 374 [1] Alon, N., S. Moran, and A. Yehudayoff (2016). Sign rank versus VC dimension. In *Conference*  
 375 *on Learning Theory*, pp. 47–80.
- 376 [2] Anandkumar, A., R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky (2014). Tensor decompositions  
 377 for learning latent variable models. *Journal of Machine Learning Research* 15(1), 2773–2832.
- 378 [3] Anandkumar, A., R. Ge, and M. Janzamin (2017). Analyzing tensor power method dynamics in  
 379 overcomplete regime. *Journal of Machine Learning Research* 18(1), 752–791.
- 380 [4] Balabdaoui, F., C. Durot, and H. Jankowski (2019). Least squares estimation in the monotone  
 381 single index model. *Bernoulli* 25(4B), 3276–3310.
- 382 [5] Bartlett, P. L., M. I. Jordan, and J. D. McAuliffe (2006). Convexity, classification, and risk  
 383 bounds. *Journal of the American Statistical Association* 101(473), 138–156.
- 384 [6] Cai, C., G. Li, H. V. Poor, and Y. Chen (2019). Nonconvex low-rank tensor completion from  
 385 noisy data. In *Advances in Neural Information Processing Systems*, pp. 1863–1874.
- 386 [7] Chan, S. and E. Airoldi (2014). A consistent histogram estimator for exchangeable graph models.  
 387 In *International Conference on Machine Learning*, pp. 208–216.
- 388 [8] Chi, E. C., B. J. Gaines, W. W. Sun, H. Zhou, and J. Yang (2020). Provable convex co-clustering  
 389 of tensors. *Journal of Machine Learning Research* 21(214), 1–58.
- 390 [9] Cohn, H. and C. Umans (2013). Fast matrix multiplication using coherent configurations. In  
 391 *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pp.  
 392 1074–1087.
- 393 [10] De Lathauwer, L., B. De Moor, and J. Vandewalle (2000). A multilinear singular value  
 394 decomposition. *SIAM Journal on Matrix Analysis and Applications* 21(4), 1253–1278.
- 395 [11] De Wolf, R. (2003). Nondeterministic quantum query and communication complexities. *SIAM*  
 396 *Journal on Computing* 32(3), 681–699.
- 397 [12] Fan, J. and M. Udell (2019). Online high rank matrix completion. In *Proceedings of the IEEE*  
 398 *Conference on Computer Vision and Pattern Recognition*, pp. 8690–8698.
- 399 [13] Ganti, R., N. Rao, L. Balzano, R. Willett, and R. Nowak (2017). On learning high dimensional  
 400 structured single index models. In *Proceedings of the Thirty-First AAAI Conference on Artificial*  
 401 *Intelligence*, pp. 1898–1904.
- 402 [14] Ganti, R. S., L. Balzano, and R. Willett (2015). Matrix completion under monotonic single  
 403 index models. In *Advances in Neural Information Processing Systems*, pp. 1873–1881.
- 404 [15] Ghadermarzy, N., Y. Plan, and O. Yilmaz (2018). Learning tensors from partial binary measure-  
 405 ments. *IEEE Transactions on Signal Processing* 67(1), 29–40.
- 406 [16] Ghadermarzy, N., Y. Plan, and Ö. Yilmaz (2019). Near-optimal sample complexity for convex  
 407 tensor completion. *Information and Inference: A Journal of the IMA* 8(3), 577–619.

408 [17] Globerson, A., G. Chechik, F. Pereira, and N. Tishby (2007). Euclidean embedding of co-  
409 occurrence data. *Journal of Machine Learning Research* 8, 2265–2295.

410 [18] Han, R., R. Willett, and A. Zhang (2020). An optimal statistical and computational framework  
411 for generalized tensor estimation. *arXiv preprint arXiv:2002.11255*.

412 [19] Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning: data*  
413 *mining, inference, and prediction*. Springer Science & Business Media.

414 [20] Hillar, C. J. and L.-H. Lim (2013). Most tensor problems are NP-hard. *Journal of the ACM*  
415 *(JACM)* 60(6), 45.

416 [21] Hitchcock, F. L. (1927). The expression of a tensor or a polyadic as a sum of products. *Journal*  
417 *of Mathematics and Physics* 6(1-4), 164–189.

418 [22] Hong, D., T. G. Kolda, and J. A. Dueresch (2020). Generalized canonical polyadic tensor  
419 decomposition. *SIAM Review* 62(1), 133–163.

420 [23] Hore, V., A. Viñuela, A. Buil, J. Knight, M. I. McCarthy, K. Small, and J. Marchini (2016).  
421 Tensor decomposition for multiple-tissue gene expression experiments. *Nature genetics* 48(9),  
422 1094.

423 [24] Jain, P. and S. Oh (2014). Provable tensor factorization with missing data. In *Advances in*  
424 *Neural Information Processing Systems*, Volume 27, pp. 1431–1439.

425 [25] Kolda, T. G. and B. W. Bader (2009). Tensor decompositions and applications. *SIAM Re-*  
426 *view* 51(3), 455–500.

427 [26] Lee, C. and M. Wang (2020). Tensor denoising and completion based on ordinal observations.  
428 In *International Conference on Machine Learning*, pp. 5778–5788.

429 [27] Li, Y., Y. Liu, J. Li, W. Qin, K. Li, C. Yu, and T. Jiang (2009). Brain anatomical network and  
430 intelligence. *PLoS Comput Biol* 5(5), e1000395.

431 [28] Montanari, A. and N. Sun (2018). Spectral algorithms for tensor completion. *Communications*  
432 *on Pure and Applied Mathematics* 71(11), 2381–2425.

433 [29] Ongie, G., R. Willett, R. D. Nowak, and L. Balzano (2017). Algebraic variety models for  
434 high-rank matrix completion. In *International Conference on Machine Learning*, pp. 2691–2700.

435 [30] Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: Journal*  
436 *of the Econometric Society* 56(4), 931–954.

437 [31] Wang, L., D. Durante, R. E. Jung, and D. B. Dunson (2017). Bayesian network–response  
438 regression. *Bioinformatics* 33(12), 1859–1866.

439 [32] Wang, M. and L. Li (2020). Learning from binary multiway data: Probabilistic tensor decomp-  
440 osition and its statistical optimality. *Journal of Machine Learning Research* 21(154), 1–38.

441 [33] Wang, M. and Y. Zeng (2019). Multiway clustering via tensor block models. In *Advances in*  
442 *Neural Information Processing Systems*, pp. 713–723.

443 [34] Xu, J. (2018). Rates of convergence of spectral methods for graphon estimation. In *International*  
444 *Conference on Machine Learning*, pp. 5433–5442.

445 [35] Yuan, M. and C.-H. Zhang (2016). On tensor completion via nuclear norm minimization.  
446 *Foundations of Computational Mathematics* 16(4), 1031–1068.

447 [36] Zhang, A. and D. Xia (2018). Tensor SVD: Statistical and computational limits. *IEEE*  
448 *Transactions on Information Theory* 64(11), 7311 – 7338.