Beyond the Signs: Nonparametric Tensor Completion via Learning Reduction to Sign Series

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
3   based on a new model which we coin as sign representable tensors. The model
4   represents the signal tensor of interest using a series of structured sign tensors.
5   Unlike earlier methods, the sign series representation effectively addresses both
6   low- and high-rank signals, while encompassing many existing tensor models—
7   including CP models, Tucker models, single index models, structured tensors with
8   repeating entries—as special cases. We show the reduction of nonparametric tensor
9   estimation to a series of structured classifications with carefully-specified weights,
10 and we develop a learning reduction machinery to empower existing low-rank
11 tensor algorithms for more challenging high-rank estimation. Excess risk bounds,
12 estimation errors, and sample complexities are established. We demonstrate the
13 outperformance of our approach over previous methods on two datasets, one on
14 human brain connectivity networks and the other on topic data mining.

1   Introduction

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

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

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

22 Q1 [Nonparametric tensor estimation]. How to flexibly estimate Θ under a wide range of structures,
23 including both low-rankness and high-rankness?

24 Q2 [Complexity of tensor completion]. How many observed tensor entries do we need to consistently
25 estimate the signal Θ?

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

31 The first example reveals the sensitivity of tensor rank to order-preserving transformations. Let
32 $\mathcal{Z} \in \mathbb{R}^{30 \times 30 \times 30}$ be an order-3 tensor with CP rank$(\mathcal{Z}) = 3$ (formal definition is deferred to end
33 of this section). Suppose a monotonic transformation $f(z) = (1 + \exp(-cz))^{-1}$ is applied to $\mathcal{Z}$

entrywise, and we let the signal Θ in model (1) be the tensor after transformation. Figure 1a plots the numerical rank (see Section **??** in Appendix) of Θ versus c. As we see, the rank increases rapidly with 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.
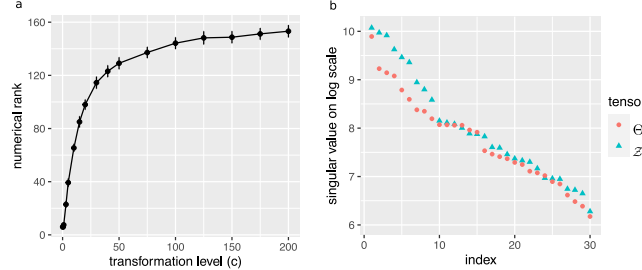

Figure 1: (a) Numerical rank of Θ versus c in the first example. (b) Top $d = 30$ tensor singular values in the second example.

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 Θ was studied by [7] in graphon analysis. In this case neither Θ nor $\mathcal{Z}$ is low-rank; in fact, the rank is no smaller than the dimension d as illustrated in Figure 1b. Again, classical low-rank models fail to address this type of tensor structure.

In the above and many other examples, the signal tensors Θ 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.
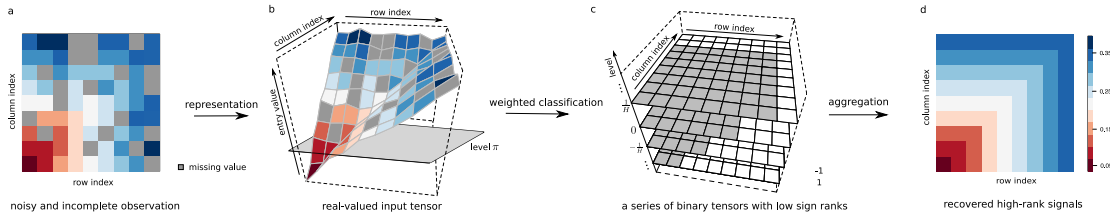

Figure 2: Illustration of our method. For visualization purpose, we plot an order-2 tensor (a.k.a. matrix); similar procedure applies to higher-order tensors. (a): a noisy and incomplete tensor input. (b) and (c): main steps of estimating sign tensor series $\text{sgn}(\Theta - \pi)$ for $\pi \in \{-1, \dots, -1/H, 0, 1/H, \dots, 1\}$. (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,

69 the problem for high-rank tensors is more challenging, because the tensor rank often exceeds the
70 dimension when order $K \geq 3$ [3]. This is in sharp contrast to matrices. We show that, applying
71 matrix methods to higher-order tensors results in suboptimal estimates. A full exploitation of the
72 higher-order structure is needed; this is another challenge we address in this paper.

73 **Notation.** We use $\text{sgn}(\cdot)\colon \mathbb{R} \to \{-1, 1\}$ to denote the sign function, where $\text{sgn}(y) = 1$ if $y \geq 0$ and
74 -1 otherwise. We allow univariate functions, such as $\text{sgn}(\cdot)$ and general $f\colon \mathbb{R} \to \mathbb{R}$, to be applied to
75 tensors in an element-wise manner. We denote $a_n \lesssim b_n$ if $\lim_{n \to \infty} a_n/b_n \leq c$ for some constant
76 $c \geq 0$. We use the shorthand $[n]$ to denote the n-set $\{1, \ldots, n\}$ for $n \in \mathbb{N}_+$. Let $\Theta \in \mathbb{R}^{d_1 \times \cdots \times d_K}$
77 denote an order-K $(d_1, \ldots, d_K)$-dimensional tensor, and $\Theta(\omega) \in \mathbb{R}$ denote the tensor entry indexed
78 by $\omega \in [d_1] \times \cdots \times [d_K]$. An event E is said to occur "with very high probability" if $\mathbb{P}(E)$ tends to 1
79 faster than any polynomial of tensor dimension $d := \min_k d_k \to \infty$. The CP decomposition [21] is
80 defined by $\Theta = \sum_{s=1}^{r} \lambda_s \boldsymbol{a}_s^{(1)} \otimes \cdots \otimes \boldsymbol{a}_s^{(K)}$, where $\lambda_1 \geq \cdots \geq \lambda_r > 0$ are tensor singular values,
81 $\boldsymbol{a}_s^{(k)} \in \mathbb{R}^{d_k}$ are norm-1 tensor singular vectors, and $\otimes$ denotes the outer product of vectors. The
82 minimal $r \in \mathbb{N}_+$ in the CP decomposition is called the tensor rank, denoted $\text{rank}(\Theta)$.

2 Model and proposal overview

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

$$\mathcal{Y} = \Theta + \mathcal{E}, \tag{2}$$

85 where $\Theta \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is an unknown signal tensor of interest, and $\mathcal{E}$ is a noise tensor consisting of
86 zero-mean, independent but not necessarily identically distributed entries. We allow heterogenous
87 noise, in that the marginal distribution of noise entry $\mathcal{E}(\omega)$ may depend on ω. For simplicity, we
88 assume the noise is bounded; the extension to a sub-Gaussian noise is provided in the Appendix. Here,
89 we assume the range of $\mathcal{Y}$ is the bounded interval $[-1, 1]$ for cleaner exposition. Our observation
90 is an incomplete data tensor from (2), denoted $\mathcal{Y}_\Omega$, where $\Omega \subset [d_1] \times \cdots \times [d_K]$ is the index set
91 of observed entries. We consider a general model on Ω that allows both uniform and non-uniform
92 samplings. Specifically, let $\Pi = \{p_\omega\}$ be an arbitrarily predefined probability distribution over the
93 full index set with $\sum_{\omega \in [d_1] \times \cdots \times [d_K]} p_\omega = 1$. Assume that the entries ω in Ω are i.i.d. draws with
94 replacement from distribution Π. The sampling rule is denoted as $\omega \sim \Pi$.

95 **Intuition.** Before describing our main results, we provide the intuition behind our method. In the
96 two examples in Section 1, the high-rankness in the signal Θ makes the estimation challenging.
97 Now let us examine the sign of the π-shifted signal $\text{sgn}(\Theta - \pi)$ for any given $\pi \in [-1, 1]$. It
98 turns out that, these sign tensors share the same sign patterns as low-rank tensors. Indeed, the
99 signal tensor in the first example has the same sign pattern as a rank-4 tensor, since $\text{sgn}(\Theta - \pi) =$
100 $\text{sgn}(\mathcal{Z} - f^{-1}(\pi))$. The signal tensor in the second example has the same sign pattern as a rank-2
101 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).

102 The above observation suggests a general framework to estimate both low- and high-rank signal
103 tensors. Figure 2 illustrates the main crux of our method. We propose to estimate the signal tensor Θ
104 by taking the average over sign tensors

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

105 where $\text{sgn}(\hat{\mathcal{Z}}_\pi) \in \{-1, 1\}^{d_1 \times \cdots \times d_K}$ is a sign tensor that estimates sign signal, $\text{sgn}(\Theta - \pi)$, at a series
106 of $\pi \in \mathcal{H} = \{-1, \ldots, -1/H, 0, 1/H, \ldots, 1\}$, and Weighted-Loss$(\cdot, \cdot)$ denotes a carefully-designed
107 classification objective function which will be described in Section 3.2. To obtain $\text{sgn}(\hat{\mathcal{Z}}_\pi)$ for a
108 given π, we propose to dichotomize the data tensor into sign tensor, $\text{sgn}(\mathcal{Y}_\Omega - \pi)$, and estimate
109 the sign signal by performing weighted classification based on objective Weighted-Loss$(\cdot, \cdot)$. Our
110 approach is built on the nonparametric sign representation of signal tensors. We show that a careful
111 aggregation of dichotomized data not only preserves all information in the original signals, but also
112 brings benefits of accuracy and flexibility over classical low-rank models. Unlike traditional methods,
113 the sign representation is guaranteed to recover both low- and high-rank signals. The method enjoys
114 statistical effectiveness and computational efficiency.

3

3 Oracle properties of sign representable tensors

This section develops sign representable tensor models for Θ 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 Θ be the tensor of interest, and $\mathrm{sgn}(\Theta)$ 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 $\Theta \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is defined by the minimal rank among all tensors that share the same sign pattern as Θ; i.e.,

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

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 Θ. In particular, $\mathrm{srank}(\Theta) = \mathrm{srank}(\mathrm{sgn}\Theta)$.

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 $g\colon \mathbb{R} \to \mathbb{R}$ with $g(0) = 0$, the following holds true,* $\mathrm{srank}(\Theta) \leq \mathrm{rank}(g(\Theta))$.

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 $\Theta \in \mathbb{R}^{d \times \cdots \times d}$ such that* $\mathrm{rank}(\Theta) \geq d$ *but* $\mathrm{srank}(\Theta - \pi) \leq 2$ *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 Θ is called (r, π)-sign representable, if the tensor $(\Theta - \pi)$ has sign-rank bounded by r. A tensor Θ is called r-sign (globally) representable, if Θ is (r, π)-sign representable for all $\pi \in [-1, 1]$. The collection $\{\mathrm{sgn}(\Theta - \pi)\colon \pi \in [-1, 1]\}$ is called the sign tensor series. We use $\mathscr{P}_{\mathrm{sgn}}(r) = \{\Theta\colon \max_{\pi \in [-1,1]} \mathrm{srank}(\Theta - \pi) \leq r\}$ 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 Θ be a low-rank tensor with CP rank r. We see that Θ belongs to the sign representable family; i.e., $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r + 1)$ (the constant 1 is due to $\mathrm{rank}(\Theta - \pi) \leq r + 1$). Similar results hold for Tucker low-rank tensors $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r + 1)$, where $r = \prod_k r_k$ with r_k being the k-th mode Tucker rank of Θ.

Example 2 (Tensor block models (TBMs))**.** Tensor block model [33, 8] assumes a checkerboard structure among tensor entries under marginal index permutation. The signal tensor Θ takes at most r distinct values, where r is the total number of multiway blocks. Our model incorporates TBM because $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r)$.

Example 3 (Generalized linear models (GLMs))**.** Let $\mathcal{Y}$ be a binary tensor from a logistic model [32] with mean $\Theta = \mathrm{logit}(\mathcal{Z})$, where $\mathcal{Z}$ is a latent low-rank tensor. Notice that Θ itself may be high-rank (see Section 1). By definition, Θ 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 $g\colon \mathbb{R} \to \mathbb{R}$ such that $g(\Theta)$ has rank r. We see that Θ belongs to the sign representable family; i.e., $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r + 1)$.

167 **Example 5** (Structured tensors with repeating entries). Here we revisit the model introduced in
168 Section 1. Let Θ be an order-K tensor with entries $\Theta(i_1, \ldots, i_K) = \log(1 + \max_k x_{i_k}^{(k)})$, where $x_{i_k}^{(k)}$
169 are given number in $[0, 1]$ for all $i_k \in [d_k], k \in [K]$. We conclude that $\Theta \in \mathscr{P}_{\mathrm{sgn}}(2)$, because the
170 sign tensor $\mathrm{sgn}(\Theta - \pi)$ with an arbitrary $\pi \in (0, \log 2)$ is a block tensor with at most two blocks
171 (see Figure 2c). The results extend to general structured tensors with repeating entries. Let $g(\cdot)$ be a
172 polynomial of degree r and the associated tensor Θ with entries $\Theta(i_1, \ldots, i_K) = g(\max_k x_{i_k}^{(k)})$. We
173 have $\Theta \in \mathscr{P}_{\mathrm{sgn}}(2r)$ (see proofs in the Appendix).

3.2 Statistical characterization of sign tensors via weighted classification

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

177 For a given $\pi \in [-1, 1]$, define a π-shifted data tensor $\bar{\mathcal{Y}}_\Omega$ with entries $\bar{\mathcal{Y}}(\omega) = (\mathcal{Y}(\omega) - \pi)$ for
178 $\omega \in \Omega$. We propose a weighted classification objective function

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

179 where $\mathcal{Z} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is the decision variable to be optimized, $|\bar{\mathcal{Y}}(\omega)|$ is the entry-specific weight
180 equal to the distance from the tensor entry to the target level π. The entry-specific weights incorporate
181 the magnitude information into classification, where entries far away from the target level are
182 penalized more heavily in the objective. In the special case of binary tensor $\mathcal{Y} \in \{-1, 1\}^{d_1 \times \cdots \times d_K}$
183 and target level $\pi = 0$, the loss (3) reduces to usual classification loss.

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

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

186 where the expectation is taken with respect to $\mathcal{Y}_\Omega$ under model (2) and the sampling distribution
187 $\omega \sim \Pi$. The form of $\mathrm{Risk}(\cdot)$ implicitly depends on π; we suppress π when no confusion arises.

188 **Proposition 3** (Global optimum of weighted risk). *Suppose the data $\mathcal{Y}_\Omega$ is generated from model* (2)
189 *with $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r)$. Then, for all $\bar{\Theta}$ that are sign equivalent to $\mathrm{sgn}(\Theta - \pi)$,*

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

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

194 Some additional notation is needed for stating the results in full generality. Let $d_t = \prod_{k=1}^{K} d_k$
195 denote the total number of tensor entries, and $\Delta s = 1/d_t$ a small tolerance. We quantify distribution
196 of entries in tensor Θ using a pseudo density, i.e., histogram with bin width $2\Delta s$. Let $G(\pi) :=$
197 $\mathbb{P}_{\omega \sim \Pi}[\Theta(\omega) \leq \pi]$ denote the cumulative distribution function (CDF) of $\Theta(\omega)$ under $\omega \sim \Pi$. We
198 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
199 asymptotically unbounded; i.e,

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

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

201 **Assumption 1** (α-smoothness). Fix $\pi \notin \mathcal{N}$. Assume there exist constants $\alpha = \alpha(\pi) > 0, c =$
202 $c(\pi) > 0$, independent of tensor dimension, such that,

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

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

5

207 The smoothness index α quantifies the intrinsic hardness of recovering $\text{sgn}(\Theta - \pi)$ from $\text{Risk}(\cdot)$.
208 The value of α depends on both the sampling distribution $\omega \sim \Pi$ and the behavior of $\Theta(\omega)$. The
209 recovery is easier at levels where points are less concentrated around π with a large value of $\alpha > 1$,
210 or equivalently, when $G(\pi)$ remains almost flat around π. A small value of $\alpha < 1$ indicates the
211 nonexistent (infinite) density at level π, or equivalently, when the $G(\pi)$ jumps by greater than the
212 tolerance Δs at π. Table 1 illustrates the $G(\pi)$ for various models of Θ (see Section 5).

213 We now reach the main theorem in this section. For two tensors Θ_1, Θ_2, define the mean absolute
214 error (MAE) as $\text{MAE}(\Theta_1, \Theta_2) \stackrel{\text{def}}{=} \mathbb{E}_{\omega \sim \Pi}|\Theta_1(\omega) - \Theta_2(\omega)|$.
215 **Theorem 1** (Identifiability)**.** *Assume $\Theta \in \mathscr{P}_{\text{sgn}}(r)$ is α-globally smooth. Then, for all $\pi \notin \mathcal{N}$ and*
216 *tensors $\bar{\Theta} \simeq \text{sgn}(\Theta - \pi)$, we have*

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

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

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

222 We conclude this section by applying Assumption 1 to the examples described in Section 3.1. For
223 simplicity, suppose Π is the uniform sampling. The tensor block model is ∞-globally smooth. This
224 is because the set $\mathcal{N}$ consists of finite $2\Delta s$-bin's covering the distinct block means in Θ. Furthermore,
225 we have $\alpha = \infty$ for all $\pi \notin \mathcal{N}$, since the numerator in (5) is zero. Similarly, the high-rank tensor
226 $\Theta(i_1, \ldots, i_K) = \log(1 + \max_{\ell=1,\ldots K} i_\ell/d)$ is ∞-globally smooth because $\alpha = \infty$ for all π except
227 those in $\mathcal{N}$, where $\mathcal{N}$ collects d many $2\Delta s$-bin's covering $\log(1 + i/d)$ for $i = 1, \ldots, d$.

4 Nonparametric tensor completion via sign series

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

4.1 Statistical error and sample complexity

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

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

236 where $\mathcal{H} = \{-1, \ldots, -1/H, 0, 1/H, \ldots, 1\}$ is a series of level to aggregate, $L(\cdot, \cdot)$ denotes the
237 weighted classification objective defined in (3), and the rank constraint on $\mathcal{Z}$ follows from Proposi-
238 tion 3. For the theory, we assume the true r is known; in practice, r could be chosen in a data adaptive
239 fashion via cross-validation or elbow method [19].

240 The next theorem establishes the statistical convergence for the sign tensor estimate (6), which is an
241 important ingredient for the final signal tensor estimate $\hat{\Theta}$ in (6).
242 **Theorem 2** (Sign tensor estimation)**.** *Suppose $\Theta \in \mathscr{P}_{\text{sgn}}(r)$ and $\Theta(\omega)$ is α-globally smooth under*
243 *$\omega \sim \Pi$. 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*
244 *$\pi \notin \mathcal{N}$, with very high probability over $\mathcal{Y}_\Omega$,*

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

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

249 Recall that $\mathcal{N}$ collects the levels for which the sign tensor is possibly nonrecoverable. Let $|\mathcal{N}|$ be the
250 covering number of $\mathcal{N}$ with $2\Delta s$-bin's, i.e, $|\mathcal{N}| = \lceil \mu(\mathcal{N})/2\Delta s \rceil$, where μ is the Lebesgue measure.
251 Combining the sign representability of the signal tensor and the sign estimation accuracy, we obtain
252 the main results on our nonparametric tensor estimation method.

253 **Theorem 3** (Tensor estimation error). *Consider the same conditions of Theorem 2. Let $\hat{\Theta}$ be the*
254 *estimate in* (6) *and* $t_d = d_{\max}r/|\Omega|$. *With very high probability over* $\mathcal{Y}_\Omega$,

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

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

256 Theorem 3 demonstrates the convergence rate of our tensor estimation. The bound (8) reveals three
257 sources of errors: the estimation error for sign tensors, the bias from sign series representations, and
258 the variance thereof. The resolution parameter H controls the bias-variance tradeoff. We remark that
259 the signal estimation error (8) is generally no better than the corresponding sign error (7). This is to
260 be expected, since magnitude estimation is a harder problem than sign estimation.

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

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

263 Compared to earlier methods, our estimation accuracy applies to both low- and high-rank signal
264 tensors. The rate depends on the sign complexity $\Theta \in \mathscr{P}_{\text{sgn}}(r)$, and this r is often much smaller
265 than the usual tensor rank (see Section 3.1). Our result also reveals that the convergence becomes
266 favorable as the order of data tensor increases.

267 We apply our method to the main examples in Section 3.1, and compare the results with existing
268 literature. The numerical comparison is provided in Section 5.
269 **Example 2** (TBMs). Consider a tensor block model with r multiway blocks. Our result implies a rate
270 $\mathcal{O}(d^{-(K-1)/2})$ by taking $\alpha = \infty$ and $|\mathcal{N}| \le r^K$. This rate agrees with the previous root-mean-square
271 error (RMSE) for block tensor estimation [33].
272 **Example 3** (GLMs). Consider a GLM tensor $\Theta = g(\mathcal{Z})$, where g is a known link function and $\mathcal{Z}$
273 is a latent low-rank tensor. Suppose the marginal density of $\Theta(\omega)$ is uniformly bounded as $d \to \infty$.
274 Applying our results with $\alpha = 1$ and finite $|\mathcal{N}|$ yields $\mathcal{O}(d^{-(K-1)/3})$. This rate is slightly slower
275 than the parametric RMSE rate [35, 32], as expected. The reason is that our estimate remains valid
276 for unknown g and general high-rank tensors. The nonparametric rate is the price one has to pay for
277 not knowing the form $\Theta = g(\mathcal{Z})$ as a priori.
278 **Example 4** (SIMs). The earlier example has shown the nonparametric rate $\mathcal{O}(d^{-(K-1)/3})$ when
279 applying our method to single index tensor model. In the matrix case with $K = 2$, our result yields
280 error rate $\mathcal{O}(d^{-1/3})$, which is faster compared to the RMSE rate $\mathcal{O}(d^{-1/4})$ obtained by [14].
281 **Example 5** (Structured tensors with repeating entries). We consider a more general model than that
282 in Section 1. Consider a r-sign representable tensor $\Theta \in \mathscr{P}_{\text{sgn}}(r)$ with at most d distinct entries with
283 repetition pattern. Applying our results with $\alpha = \infty$ and $|\mathcal{N}| = d$ yields the rate $\mathcal{O}(d^{-(K-2)/2})$.

284 The following corollary reveal the sample complexity for nonparamtric tensor completion.
285 **Corollary 1** (Sample complexity for nonparametric completion). *Under the same conditions of*
286 *Theorem 3 with* $\alpha \neq 0$ *and bounded* $|\mathcal{N}|$, *with high probability over* $\mathcal{Y}_\Omega$,

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

287 Our result improves earlier work [34, 16, 26] by allowing both low- and high-rank signals. Interest-
288 ingly, the sample requirements depend only on the sign complexity $d_{\max}r$ but not the nonparametric
289 complexity α. Note that $\tilde{\mathcal{O}}(d_{\max}r)$ roughly matches the degree of freedom of sign tensors, suggesting
290 the optimality of our sample requirements.

4.2 Numerical implementation via learning reduction

292 Cut details. Emphasize learning reduction. Meta algorithm is a simple average of $2H + 1 \asymp \text{poly}(d)$
293 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 ψ-loss $F(m) = 2\min(1, (1 - m)_+)$ with $m_+ = \max(m, 0)$. We implement the hinge loss and logistic loss in our algorithm, although our framework is applicable to general large-margin losses [5]. The full procedure is described in Algorithm 4.2.

Algorithm 1 Nonparametric tensor completion

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

Add guarantee from 0-1 loss to hinge loss. Users can choose their favorite (polynomial) 1-bit tensor algorithms, as long as these algorithms are sample efficient. Revise the pseudo algorithm to reflect the meta vs. base algorithms. Skip the details of base algorithm.

5 Experiments

5.1 Finite-sample accuracy

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**). The performance is assessed under both complete and incomplete observations. We generate signal tensors 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 $d \in \{15, 20, \ldots, 55, 60\}$, $r = 2$, $H = 10 + (d - 15)/5$ 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 Θ	Rank	Sign Rank	α	$|\mathcal{N}|$	CDF of Tensor Entries	Noise
1	$\mathcal{C} \times \boldsymbol{M}_1 \times \boldsymbol{M}_2 \times \boldsymbol{M}_3$	3^3	$\leq 3^3$	∞	$\leq 3^3$		Uniform $[-0.3, 0.3]$
2	$|\boldsymbol{a} \otimes \mathbf{1} \otimes \mathbf{1} - \mathbf{1} \otimes \boldsymbol{a} \otimes \mathbf{1}|$	d	≤ 3	1	0		Normal $\mathcal{N}(0, 0.15)$
3	$\log(0.5 + \mathcal{Z}_{\max})$	$\geq d$	2	∞	d		Uniform $[-0.1, 0.1]$
4	$2.5 - \exp(\mathcal{Z}_{\min}^{1/3})$	$\geq d$	2	∞	d		Normal $\mathcal{N}(0, 0.15)$

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

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 $|\Omega|/d^3$

8

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
Θ, 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
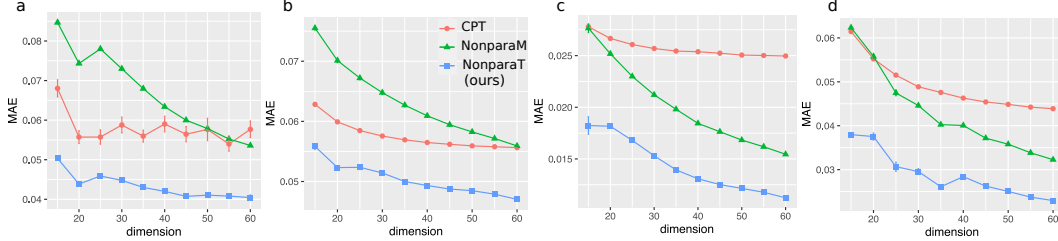underlying signal tensor is often unknown. keep only one figure. panels b and d only. Others in
appendix.

Figure 3: Estimation error versus tensor dimension. need change

5.2 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	$r = 6$	$r = 9$	$r = 12$
NonParaT (Ours)	**0.14**(0.001)	**0.12**(0.001)	**0.12**(0.001)
CPT	0.23(0.006)	0.22(0.004)	0.21(0.006)
NIPS word occurrence dataset			
Method	$r = 6$	$r = 9$	$r = 12$
NonParaT (Ours)	**0.16**(0.002)	**0.15**(0.001)	**0.14**(0.001)
CPT	0.20(0.007)	0.19(0.007)	0.17(0.007)

Table 2: MAE comparison between our method **Non-paraT** ($H = 20$) and CP low-rank method (**CPT**) in the brain and NIPS data analysis. Standard errors are reported in parenthesis.
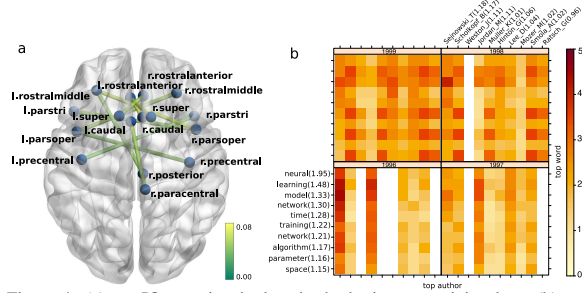

Figure 4: (a) top IQ-associated edges in the brain connectivity data. (b) top authors and words for years 1996-1999 in the NIPS data.

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

6   Conclusion

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

References

[1] Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank versus VC dimension. In *Conference on Learning Theory*, pages 47–80, 2016.

[2] Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15(1):2773–2832, 2014.

[3] Animashree Anandkumar, Rong Ge, and Majid Janzamin. Analyzing tensor power method dynamics in overcomplete regime. *Journal of Machine Learning Research*, 18(1):752–791, 2017.

[4] Fadoua Balabdaoui, Cécile Durot, and Hanna Jankowski. Least squares estimation in the monotone single index model. *Bernoulli*, 25(4B):3276–3310, 2019.

[5] Peter L Bartlett, Michael I Jordan, and Jon D McAuliffe. Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 101(473):138–156, 2006.

[6] Changxiao Cai, Gen Li, H Vincent Poor, and Yuxin Chen. Nonconvex low-rank tensor completion from noisy data. In *Advances in Neural Information Processing Systems*, pages 1863–1874, 2019.

[7] Stanley Chan and Edoardo Airoldi. A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning*, pages 208–216, 2014.

[8] Eric C Chi, Brian J Gaines, Will Wei Sun, Hua Zhou, and Jian Yang. Provable convex co-clustering of tensors. *Journal of Machine Learning Research*, 21(214):1–58, 2020.

[9] Henry Cohn and Christopher Umans. Fast matrix multiplication using coherent configurations. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 1074–1087, 2013.

[10] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

[11] Ronald De Wolf. Nondeterministic quantum query and communication complexities. *SIAM Journal on Computing*, 32(3):681–699, 2003.

[12] Jicong Fan and Madeleine Udell. Online high rank matrix completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8690–8698, 2019.

[13] Ravi Ganti, Nikhil Rao, Laura Balzano, Rebecca Willett, and Robert Nowak. On learning high dimensional structured single index models. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 1898–1904, 2017.

[14] Ravi Sastry Ganti, Laura Balzano, and Rebecca Willett. Matrix completion under monotonic single index models. In *Advances in Neural Information Processing Systems*, pages 1873–1881, 2015.

[15] Navid Ghadermarzy, Yaniv Plan, and Ozgur Yilmaz. Learning tensors from partial binary measurements. *IEEE Transactions on Signal Processing*, 67(1):29–40, 2018.

[16] Navid Ghadermarzy, Yaniv Plan, and Özgür Yilmaz. Near-optimal sample complexity for convex tensor completion. *Information and Inference: A Journal of the IMA*, 8(3):577–619, 2019.

[17] Amir Globerson, Gal Chechik, Fernando Pereira, and Naftali Tishby. Euclidean embedding of co-occurrence data. *Journal of Machine Learning Research*, 8:2265–2295, 2007.

[18] Rungang Han, Rebecca Willett, and Anru Zhang. An optimal statistical and computational framework for generalized tensor estimation. *arXiv preprint arXiv:2002.11255*, 2020.

[19] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

[20] Christopher J Hillar and Lek-Heng Lim. Most tensor problems are NP-hard. *Journal of the ACM (JACM)*, 60(6):45, 2013.

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

[22] David Hong, Tamara G Kolda, and Jed A Duersch. Generalized canonical polyadic tensor decomposition. *SIAM Review*, 62(1):133–163, 2020.

[23] Victoria Hore, Ana Viñuela, Alfonso Buil, Julian Knight, Mark I McCarthy, Kerrin Small, and Jonathan Marchini. Tensor decomposition for multiple-tissue gene expression experiments. *Nature genetics*, 48(9):1094, 2016.

[24] Prateek Jain and Sewoong Oh. Provable tensor factorization with missing data. In *Advances in Neural Information Processing Systems*, volume 27, pages 1431–1439, 2014.

[25] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

[26] Chanwoo Lee and Miaoyan Wang. Tensor denoising and completion based on ordinal observations. In *International Conference on Machine Learning*, pages 5778–5788, 2020.

[27] Yonghui Li, Yong Liu, Jun Li, Wen Qin, Kuncheng Li, Chunshui Yu, and Tianzi Jiang. Brain anatomical network and intelligence. *PLoS Comput Biol*, 5(5):e1000395, 2009.

[28] Andrea Montanari and Nike Sun. Spectral algorithms for tensor completion. *Communications on Pure and Applied Mathematics*, 71(11):2381–2425, 2018.

[29] Greg Ongie, Rebecca Willett, Robert D. Nowak, and Laura Balzano. Algebraic variety models for high-rank matrix completion. In *International Conference on Machine Learning*, pages 2691–2700, 2017.

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

[31] Lu Wang, Daniele Durante, Rex E Jung, and David B Dunson. Bayesian network–response regression. *Bioinformatics*, 33(12):1859–1866, 2017.

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

[33] Miaoyan Wang and Yuchen Zeng. Multiway clustering via tensor block models. In *Advances in Neural Information Processing Systems*, pages 713–723, 2019.

[34] Ming Yuan and Cun-Hui Zhang. On tensor completion via nuclear norm minimization. *Foundations of Computational Mathematics*, 16(4):1031–1068, 2016.

[35] Anru Zhang and Dong Xia. Tensor SVD: Statistical and computational limits. *IEEE Transactions on Information Theory*, 64(11):7311 – 7338, 2018.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default [TODO] to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section **??**.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...

    (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [TODO]

    (b) Did you describe the limitations of your work? [TODO]

    (c) Did you discuss any potential negative societal impacts of your work? [TODO]

    (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [TODO]

2. If you are including theoretical results...

    (a) Did you state the full set of assumptions of all theoretical results? [TODO]

    (b) Did you include complete proofs of all theoretical results? [TODO]

3. If you ran experiments...

    (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [TODO]

    (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [TODO]

    (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [TODO]

    (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [TODO]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

    (a) If your work uses existing assets, did you cite the creators? [TODO]

    (b) Did you mention the license of the assets? [TODO]

    (c) Did you include any new assets either in the supplemental material or as a URL? [TODO]

    (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [TODO]

    (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [TODO]

5. If you used crowdsourcing or conducted research with human subjects...

    (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [TODO]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? **[TODO]**

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? **[TODO]**