The Annals of Statistics
2019, Vol. 47, No. 2, 936–964
https://doi.org/10.1214/18-AOS1694

CROSS: EFFICIENT LOW-RANK TENSOR COMPLETION

BY ANRU ZHANG[1]

University of Wisconsin-Madison

The completion of tensors, or high-order arrays, attracts significant attention in recent research. Current literature on tensor completion primarily focuses on recovery from a set of uniformly randomly measured entries, and the required number of measurements to achieve recovery is not guaranteed to be optimal. In addition, the implementation of some previous methods are NP-hard. In this article, we propose a framework for low-rank tensor completion via a novel tensor measurement scheme that we name Cross. The proposed procedure is efficient and easy to implement. In particular, we show that a third-order tensor of Tucker rank-(r_1, r_2, r_3) in p_1-by-p_2-by-p_3 dimensional space can be recovered from as few as $r_1r_2r_3 + r_1(p_1 - r_1) + r_2(p_2 - r_2) + r_3(p_3 - r_3)$ noiseless measurements, which matches the sample complexity lower bound. In the case of noisy measurements, we also develop a theoretical upper bound and the matching minimax lower bound for recovery error over certain classes of low-rank tensors for the proposed procedure. The results can be further extended to fourth or higher-order tensors. Simulation studies show that the method performs well under a variety of settings. Finally, the procedure is illustrated through a real dataset in neuroimaging.

1. Introduction. Tensors, or high-order arrays, commonly arise in a wide range of applications, including neuroimaging [Guhaniyogi, Qamar and Dunson (2017), Li and Zhang (2017), Li, Zhou and Li (2013), Sun and Li (2016), Zhou, Li and Zhu (2013)], recommender systems [Karatzoglou et al. (2010), Rendle and Schmidt-Thieme (2010), Sun et al. (2017)], hyperspectral image compression [Li and Li (2010)], multienergy computed tomography [Li et al. (2014), Semerci et al. (2014)], computer vision [Liu et al. (2013)], 3D light field displays [Wetzstein et al. (2012)] and scientific computation [Oseledets and Tyrtyshnikov (2009)]. With the revolutionary development of modern technologies, the rapid increase in data dimension, memory and time expenses outgrows the power of computing devices, which makes it difficult to work directly on the complete datasets and models. For example, a tensor of dimension 10^4-by-10^4-by-10^4 would be difficult to upload into the Random Access Memory (RAM) of a typical computer, making it hard to directly perform operations that involves all entries of the tensor. In order to

Received November 2016; revised November 2017.

[1]Supported in part by NSF Grant DMS-1811868 and NIH Grant R01-GM131399-01.

MSC2010 subject classifications. Primary 62H12; secondary 62C20.

Key words and phrases. Cross tensor measurement, denoising, minimax rate-optimal, neuroimaging, tensor completion.

conduct various statistical tensor data analyses, such as SVD or PCA [Richard and Montanari (2014), Zhang and Xia (2017)] and Monte Carlo algorithms for computations on large tensors [Guhaniyogi, Qamar and Dunson (2017), Johndrow, Bhattacharya and Dunson (2017)] when limited computation power is available, a fast and sufficient tensor compression is essential. To this end, a natural idea is to sample a small portion of entries from the original tensor dataset that preserves the important structural information and allows efficient recovery. By storing these entries to RAM, the follow-up tensor data analysis can be highly facilitated.

Tensor completion, whose central goal is to recover low-rank tensors based on limited numbers of measurable entries, is a plausible idea for compression and decompression of high-dimensional low-rank tensors. Such problems have been central and well studied for order-2 tensors (i.e., matrices) in the fields of high-dimensional statistics and machine learning for the last decade. A large body of matrix completion literatures focused on the scenario of uniformly randomly sampled observations [Agarwal, Negahban and Wainwright (2012), Candès and Tao (2010), Keshavan, Montanari and Oh (2010), Koltchinskii, Lounici and Tsybakov (2011), Negahban and Wainwright (2011), Rohde and Tsybakov (2011)], but there exists another line of works where the observations are collected by other means, such as deterministically sampling patterns [Pimentel-Alarcón, Boston and Nowak (2016)], column-subset-selection [Cai, Cai and Zhang (2016), Krishnamurthy and Singh (2013), Rudelson and Vershynin (2007), Wang and Singh (2015)] and general sampling distributions [Klopp (2014)]. There are efficient procedures for matrix completion with strong theoretical guarantees. For example, for a p_1-by-p_2 matrix of rank-r, whenever roughly $O(r(p_1 + p_2)\operatorname{polylog}(p_1 + p_2))$ uniformly randomly selected entries are observed, one can achieve nice recovery with high probability using convex algorithms such as matrix nuclear norm minimization [Candès and Tao (2010), Recht (2011)] and max-norm minimization [Cai and Zhou (2016), Srebro and Shraibman (2005)]. For matrix completion, the required number of measurements nearly matches the degrees of freedom, $O((p_1 + p_2)r)$, for p_1-by-p_2 matrices of rank-r.

Although significant progress has been made for matrix completion, similar problems for order-3 or higher tensors are far more difficult. There have been some recent literature, including Barak and Moitra (2016), Bhojanapalli and Sanghavi (2015), Gandy, Recht and Yamada (2011), Kressner, Steinlechner and Vandereycken (2014), Mu et al. (2014), Shah, Rao and Tang (2015), Yuan and Zhang (2016, 2017) that studied tensor completion based on similar formulations. To be specific, let $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ be an order-3 low-rank tensor, and $\mathbf{\Omega}$ be a subset of $[1 : p_1] \times [1 : p_2] \times [1 : p_3]$. The goal of tensor completion is to recover $\mathbf{X}$ based on the observable entries indexed by $\mathbf{\Omega}$. Most of the previous literature focuses on the setting where the indices of the observable entries are uniformly randomly selected. For example, Gandy, Recht and Yamada (2011), Liu et al. (2013) proposed the matricization nuclear norm minimization, which requires $O(rp^2 \operatorname{polylog}(p))$ observations to recover order-3 tensors of dimension p-by-p-by-p and Tucker

rank-(r, r, r). Later, Bhojanapalli and Sanghavi (2015), Jain and Oh (2014) considered an alternative minimization method for completion of low-rank tensors with CP decomposition and orthogonal factors. Yuan and Zhang (2016, 2017) proposed the tensor nuclear norm minimization algorithm for tensor completion with noiseless observations and further proved that their proposed method has guaranteed performance for p-by-p-by-p tensors of Tucker rank-(r, r, r) with high probability when $|\mathbf{\Omega}| \geq O((r^{1/2}p^{3/2} + r^2 p)\,\mathrm{polylog}(p))$. However, it is unclear whether the required number of measurements in this literature could be further improved or not. In addition, some of these proposed procedures, such as tensor matrix nuclear norm minimization, are proved to be computationally NP-hard, making them very difficult to apply in real problems. Recently, Barak and Moitra (2016) further showed that the completion of p-by-p-by-p low-rank tensors is computationally infeasible when only $O(p^{3/2})$ uniform random entries are observable, unless a more efficient algorithm exists for boolean satisfiability problem.

The central goal of this paper is to address the following question: is it possible to perform efficient low-rank tensor completion with a small number of observable entries? If so, what is the sample complexity, that is, the minimal number of entries one needs to observe, so that there exist fast algorithms for tensor completion with guaranteed performance? This problem is important to statistical learning theory and is inevitable in many high-dimensional tensor data analyses. Given the previous discussions, to sample entries uniformly at random may not be an optimal strategy to achieve the central goal. Instead, we propose a novel tensor measurement scheme and the corresponding efficient low-rank tensor completion algorithm. We name our methods *Cross Tensor Measurement Scheme* because the measurement set is in the shape of a high-dimensional cross contained in the tensor. We show that one can recover an unknown, Tucker rank-(r_1, r_2, r_3), and p_1-by-p_2-by-p_3 tensor $\mathbf{X}$ with

$$|\mathbf{\Omega}| = r_1 r_2 r_3 + r_1(p_1 - r_1) + r_2(p_2 - r_2) + r_3(p_3 - r_3)$$

noiseless Cross tensor measurements. This outperforms the previous methods in literature, and matches the degrees of freedom for all rank-(r_1, r_2, r_3) tensors of dimensions p_1-by-p_2-by-p_3. To the best of our knowledge, we are among the first to achieve this optimal rate. We also develop the corresponding recovery method for more general cases where measurements are taken with noise. The central idea is to transform the observable matricizations by singular value decomposition and perform the adaptive trimming scheme to denoise each block.

To illustrate the properties of the proposed procedure, both theoretical analyses and simulation studies are provided. We derive upper- and lower-bound results to show that the proposed recovery procedure can accommodate different levels of noise and achieve the optimal rate of convergence for a large class of low-rank tensors. Although the exact low-rank assumption is used in the theoretical analysis, some simulation settings show that such an assumption is not really necessary in practice, as long as the singular values of each matricization of the original tensor decays sufficiently.

It is worth emphasizing that because the proposed algorithms only involve basic matrix operations such as matrix multiplication and singular value decomposition, it is tuning-free in many general situations and can be implemented efficiently to handle large-scale problems. In fact, our simulation study shows that the recovery of a 500-by-500-by-500 tensor can be done stably within, on average, 10 seconds.

We also apply the proposed procedure to a 3-d MRI imaging dataset that comes from a study on Attention-deficit/hyperactivity disorder (ADHD). We show that with a limited number of Cross tensor measurements and the corresponding tensor completion algorithm, one can estimate the underlying low-rank structure of 3-d images as well as if one observes all entries of the image.

This work also relates to some previous results other than tensor completion in the literature. Mahoney, Maggioni and Drineas (2008) considered the tensor CUR decomposition, which aims to represent the tensor as the product of a subtensor and two matrices. However, simply applying their work cannot lead to optimal results in tensor completion since treating tensors as matrix slices would lose useful structures of tensors. Krishnamurthy and Singh (2013) proposed a sequential tensor completion algorithm under adaptive samplings. Their result requires $O(pr^{2.5}\log(r))$ number of entries for p-by-p-by-p order-3 tensors under the more restrictive CP rank-r condition, which is much larger than that of our method. Rauhut, Schneider and Stojanac (2017) considered a tensor recovery setting where each observation is a general linear projections of the original tensor. However, their theoretical analysis heavily relies on a conjecture that is difficult to check. Oseledets, Savostianov and Tyrtyshnikov (2008) provided an existence proof for rank-r Tucker-like approximations for p-by-p-by-p tensors with $O(r^3 + pr)$ parameters. Caiafa and Cichocki (2010) introduced representations for p_1-by-p_2-by-p_3 Tucker rank-(r, r, r) tensors based on $r^3 + r(p_1 + p_2 + p_3)$ selected entries. In Caiafa and Cichocki (2015), they further introduced a multiway projection scheme for stable, robust, and fast low-rank tensor reconstruction, which requires $O(pr^2)$ measurements and some tuning parameters, such as the rank of the tensor, for implementation. To the extent of our knowledge, we are among the first to develop the tensor completion scheme that is efficient, easy to implement, tuning-free, and allows exact tensor completion in the noiseless setting and achieves optimal estimation error in the noisy setting under the minimal sample size.

The rest of the paper is organized as follows. After an introduction to the notation and preliminaries in Section 2.1, we present the Cross tensor measurement scheme in Section 2.2. Based on the proposed measurement scheme, the tensor completion algorithms for both noiseless and noisy case are introduced in Sections 2.3 and 2.4, respectively. We further analyze the theoretical performance of the proposed algorithms in Section 3. The numerical performance of algorithms are investigated in a variety of simulation studies in Section 4. We then apply the proposed procedure to a real dataset of brain MRI imaging in Section 5. In Section 6, we briefly discuss the extensions of main results. The proofs of the main results are finally collected in the Supplementary Material [Zhang (2019)].

2. Cross tensor measurements and completion: Methodology.

2.1. *Basic notation and preliminaries*. We start with basic notation and results that will be used throughout the paper. The upper case letters, for example, X, Y, Z, are generally used to represent matrices. For $X \in \mathbb{R}^{p_1 \times p_2}$, the singular value decomposition can be written as $X = U\Sigma V^\top$. Suppose $\operatorname{diag}(\Sigma) = (\sigma_1(X), \ldots, \sigma_{\min\{p_1,p_2\}}(X))$, then $\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_{\min\{p_1,p_2\}}(X) \geq 0$ are the singular values of X. Especially, we note $\sigma_{\min}(X) = \sigma_{\min\{p_1,p_2\}}(X)$ and $\sigma_{\max}(X) = \sigma_1(X)$ as the smallest and largest singular value of X. Additionally, the matrix spectral norm and Frobenius norm are denoted as $\|X\| = \max_{u \in \mathbb{R}^{p_2}} \frac{\|Xu\|_2}{\|u\|_2}$ and $\|X\|_F = \sqrt{\sum_{i=1}^{p_1}\sum_{j=1}^{p_2} X_{ij}^2} = \sqrt{\sum_{i=1}^{\min\{p_1,p_2\}} \sigma_i^2(X)}$, respectively. We denote $\mathbb{P}_X \in \mathbb{R}^{p_1 \times p_1}$ as the projection operator onto the column space of X. Specifically, $\mathbb{P}_X = X(X^\top X)^\dagger X^\top = XX^\dagger$. Here, $(\cdot)^\dagger$ is the Moore–Penrose pseudo-inverse. Let $\mathbb{O}_{p,r}$ be the set of all p-by-r orthogonal columns, that is, $\mathbb{O}_{p,r} = \{V \in \mathbb{R}^{p \times r} : V^\top V = I_r\}$, where I_r represents the identity matrix of dimension r.

We use bold upper case letters, for example, $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ to denote tensors. If $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$, $E_t \in \mathbb{R}^{m_t \times p_t}$, $t = 1, 2, 3$. The *mode products* (tensor-matrix product) is defined as

$$\mathbf{X} \times_1 E_1 \in \mathbb{R}^{m_1 \times p_2 \times p_3}, \qquad (\mathbf{X} \times_1 E_1)_{ijk} = \sum_{s=1}^{p_1} E_{1,is}\mathbf{X}_{sjk},$$

where $i \in [1 : m_1]$, $j \in [1 : p_2]$, $k \in [1 : p_3]$. The mode-2 product $\mathbf{X} \times_2 E_2$ and mode-3 product $\mathbf{X} \times_3 E_3$ can be defined similarly. Interestingly, the products along different modes satisfy the commutative law, for example, $\mathbf{X} \times_t E_t \times_s E_s = \mathbf{X} \times_s E_s \times_t E_t$ if $s \neq t$. The *matricization* (or unfolding, flattening in literature), $\mathcal{M}_t(\mathbf{X})$, maps a tensor $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ into a matrix $\mathcal{M}_t(\mathbf{X}) \in \mathbb{R}^{p_t \times \prod_{s \neq t} p_s}$, so that for any $i \in \{1, \ldots, p_1\}$, $j \in \{1, \ldots, p_2\}$, $k \in \{1, \ldots, p_3\}$,

$$\begin{aligned}\mathbf{X}_{ijk} &= \big(\mathcal{M}_1(\mathbf{X})\big)_{[i,(j+p_2(k-1))]} \\ &= \big(\mathcal{M}_2(\mathbf{X})\big)_{[j,(k+p_3(i-1))]} = \big(\mathcal{M}_3(\mathbf{X})\big)_{[k,(i+p_1(j-1))]}.\end{aligned}$$

The tensor Hilbert–Schmidt norm and tensor spectral norm, which are defined as

$$\|\mathbf{X}\|_{\mathrm{HS}} = \sqrt{\sum_{i=1}^{p_1}\sum_{j=1}^{p_2}\sum_{k=1}^{p_3}\mathbf{X}_{ijk}^2},$$

$$\|\mathbf{X}\|_{\mathrm{op}} = \max_{u \in \mathbb{R}^{p_1}, v \in \mathbb{R}^{p_2}, w \in \mathbb{R}^{p_3}} \frac{\mathbf{X} \times_1 u \times_2 v \times_3 w}{\|u\|_2\|v\|_2\|w\|_2},$$

will be intensively used in this paper. It is also noteworthy that the general calculation of the tensor operator norm is NP-hard [Hillar and Lim (2013)]. Unlike matrices, there is no universal definition of rank for third or higher-order tensors. Standing out from various definitions, the *Tucker rank* [Tucker (1966)] has been

widely utilized in literature, and its definition is closely associated with the following *Tucker decomposition*: for $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$,

$$\begin{aligned} &\mathbf{X} = \mathbf{S} \times_1 U_1 \times_2 U_2 \times_3 U_3, \qquad \text{or equivalently} \\ &\mathbf{X}_{ijk} = \sum_{i'j'k'} s_{i'j'k'} U_{1,ii'} U_{2,jj'} U_{3,kk'}. \end{aligned} \tag{2.1}$$

Here, $\mathbf{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ is referred to as the *core tensor*, $U_k \in \mathbb{O}_{p_k, r_k}$. The minimum number of triplets (r_1, r_2, r_3) are defined as the *Tucker rank* of $\mathbf{X}$ which we denote as $\text{rank}(\mathbf{X}) = (r_1, r_2, r_3)$. The Tucker rank can be calculated easily by the rank of each matricization: $r_t = \text{rank}(\mathcal{M}_t(\mathbf{X}))$. It is also easy to prove that the triplet (r_1, r_2, r_3) satisfies $r_t \le p_t$, $\max^2\{r_1, r_2, r_3\} \le r_1 r_2 r_3$. For a more detailed survey of tensor decomposition, readers are referred to Kolda and Bader (2009).

We also use the following symbols to represent subarrays. For any subsets Ω_1, Ω_2, etc., we use $X_{[\Omega_1, \Omega_2]}$ to represent the submatrix of X with row indices Ω_1 and column indices Ω_2. The subtensors are denoted similarly: $\mathbf{X}_{[\Omega_1, \Omega_2, \Omega_3]}$ represents the tensors with mode-t indices in Ω_t for $t = 1, 2, 3$. For better presentation, we use bracket to represent index sets. Particularly for any integers $a \le b$, let $[a : b] = \{a, \ldots, b\}$ and let ":" alone represent the whole index set. Thus, $U_{[:, 1:r]}$ represents the first r columns of U; $\mathbf{X}_{[\Omega_1, \Omega_2, :]}$ represents the subtensor of $\mathbf{X}$ with mode-1 indices Ω_1, mode-2 indices Ω_2 and all mode-3 indices.

Now we establish the lower bound for the minimum number of measurements for Tucker low-rank tensor completion based on counting the degrees of freedom.

PROPOSITION 1 [Degrees of freedom for rank-(r_1, r_2, r_3) tensors in $\mathbb{R}^{p_1 \times p_2 \times p_3}$]. *Assume that* $r_1 \le p_1, r_2 \le p_2, r_3 \le p_3$, $\max^2\{r_1, r_2, r_3\} \le r_1 r_2 r_3$, *then the degrees of freedom of all rank-*(r_1, r_2, r_3) *tensors in* $\mathbb{R}^{p_1 \times p_2 \times p_3}$ *is*

$$r_1 r_2 r_3 + (p_1 - r_1) r_1 + (p_2 - r_2) r_2 + (p_3 - r_3) r_3.$$

REMARK 1. Beyond order-3 tensors, we can show the degrees of freedom for rank-$(r_1, \ldots, r_d)$ order-d tensors in $\mathbb{R}^{p_1 \times \cdots \times p_d}$ is $\prod_{t=1}^d r_t + \sum_{t=1}^d r_t (p_t - r_t)$ similarly.

Proposition 1 provides a lower bound and the benchmark for the number of measurements to guarantee low-rank tensor completion, that is, $r_1 r_2 r_3 + \sum_{t=1}^3 r_t \times (p_t - r_t)$. Since the previous methods are not guaranteed to achieve this lower bound, we focus on developing the first measurement scheme that can both work efficiently and reach this benchmark.

2.2. *Cross tensor measurements.* In this section, we propose a novel Cross tensor measurement scheme. Suppose the targeting unknown tensor $\mathbf{X}$ is of p_1-by-

p_2-by-p_3, we let

$$\begin{aligned} &\Omega_1 \subseteq [1:p_1], \qquad \Omega_2 \subseteq [1:p_2], \qquad \Omega_3 \subseteq [1:p_3], \\ &|\Omega_t| = m_t, \qquad t = 1,2,3; \\ &\Xi_1 \subseteq \Omega_2 \times \Omega_3, \qquad \Xi_2 \subseteq \Omega_3 \times \Omega_1, \qquad \Xi_3 \subseteq \Omega_1 \times \Omega_2, \\ &|\Xi_t| = g_t, \qquad t = 1,2,3. \end{aligned} \tag{2.2}$$

Then we measure the entries of $\mathbf{X}$ using the following indices set:

$$\begin{aligned} \boldsymbol{\Omega} = (\Omega_1 \times \Omega_2 \times \Omega_3) \cup \big([1:p_1] \times \Xi_1\big) \\ \cup \big([1:p_2] \times \Xi_2\big) \cup \big([1:p_3] \times \Xi_3\big), \end{aligned} \tag{2.3}$$

where

$$\Omega_1 \times \Omega_2 \times \Omega_3 = \big\{(i,j,k) : i \in \Omega_1, j \in \Omega_2, k \in \Omega_3\big\}$$

are *body measurements*;

$$\left.\begin{aligned} [1:p_1] \times \Xi_1 &= \big\{(i,j,k) : i \in [1:p_1], (j,k) \in \Xi_1\big\} \\ [1:p_2] \times \Xi_2 &= \big\{(i,j,k) : j \in [1:p_2], (k,i) \in \Xi_2\big\} \\ [1:p_3] \times \Xi_3 &= \big\{(i,j,k) : k \in [1:p_3], (i,j) \in \Xi_3\big\} \end{aligned}\right\} \tag{2.4}$$

are *arm measurements*.

Meanwhile, the intersections among body and arm measurements, which we refer to as *joint measurements*, also play important roles in our analysis:

$$\begin{aligned} \Omega_1 \times \Xi_1 &= (\Omega_1 \times \Omega_2 \times \Omega_3) \cap \big([1:p_1] \times \Xi_1\big) \\ &= \big\{(i,j,k) : i \in \Omega_1, (j,k) \in \Xi_1\big\}, \\ \Omega_2 \times \Xi_2 &= (\Omega_1 \times \Omega_2 \times \Omega_3) \cap \big([1:p_2] \times \Xi_2\big) \\ &= \big\{(i,j,k) : j \in \Omega_2, (k,i) \in \Xi_2\big\}, \\ \Omega_3 \times \Xi_3 &= (\Omega_1 \times \Omega_2 \times \Omega_3) \cap \big([1:p_3] \times \Xi_3\big) \\ &= \big\{(i,j,k) : k \in \Omega_3, (i,j) \in \Xi_3\big\}. \end{aligned} \tag{2.5}$$

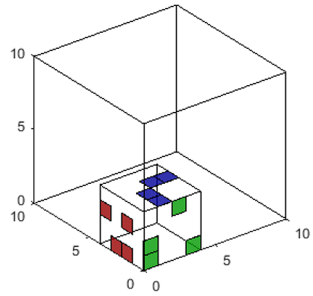
A pictorial illustration of the body, arm and joint measurements is provided in Figure 1. Since the measurements are generally cross-shaped, we refer to $\boldsymbol{\Omega}$ as the *Cross Tensor Measurement Scheme*. It is easy to see that the total number of measurements for the proposed scheme is $m_1m_2m_3 + g_1(p_1 - m_1) + g_2(p_2 - m_2) + g_3(p_3 - m_3)$ and the sampling ratio is

$$\frac{\#\text{Observable samples}}{\#\text{All parameters}} = \frac{m_1m_2m_3 + \sum_{t=1}^{3} g_t(p_t - m_t)}{p_1p_2p_3}. \tag{2.6}$$

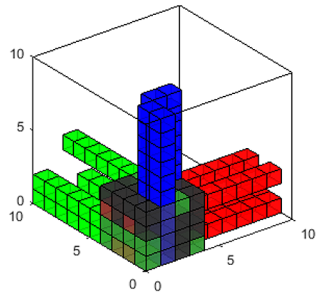
Based on these measurements, we focus on the following model:

$$\mathbf{Y}_{\boldsymbol{\Omega}} = \mathbf{X}_{\boldsymbol{\Omega}} + \mathbf{Z}_{\boldsymbol{\Omega}}, \qquad \text{that is, } Y_{ijk} = X_{ijk} + Z_{ijk}, (i,j,k) \in \boldsymbol{\Omega}, \tag{2.7}$$

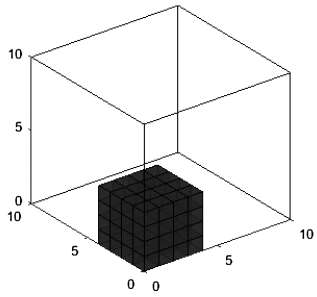
where $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ correspond to the original tensor, observed values and unknown noise term, respectively.



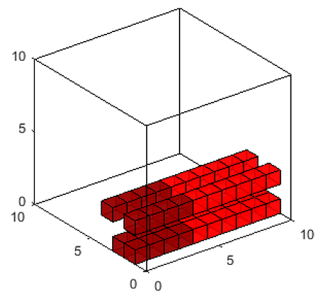
(a) Ω_t, Ξ_t illustration



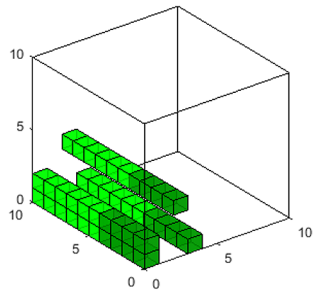
(b) All measurements



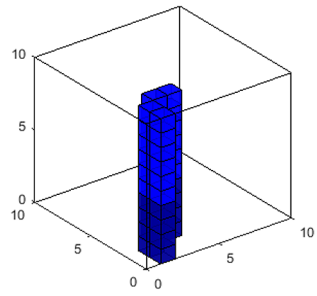
(c) Body measurements $\mathbf{Y}_{[\Omega_1,\Omega_2,\Omega_3]}$



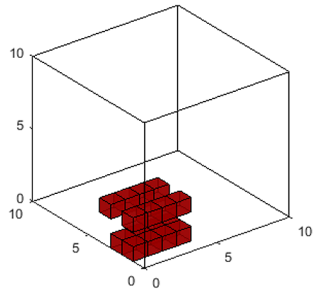
(d) Arm measurements $\mathbf{Y}_{\Xi_1}$



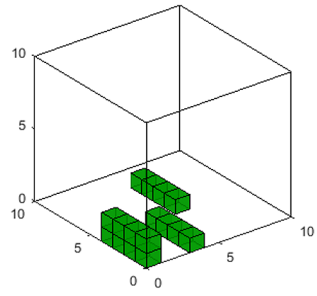
(e) Arm measurements $\mathbf{Y}_{\Xi_2}$



(f) Arm measurements $\mathbf{Y}_{\Xi_3}$

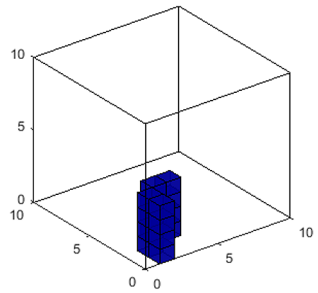


(g) Joint measurements $\mathbf{Y}_{\Omega_1\times\Xi_1}$



(h) Joint measurements $\mathbf{Y}_{\Omega_2\times\Xi_2}$

FIG. 1. *Illustrative example for the Cross Tensor Measurements Scheme. For a better illustration, here we assume* $\Omega_t = [1 : m_t]$, $p_1 = p_2 = p_3 = 10$, $m_1 = m_2 = m_3 = g_1 = g_2 = g_3 = 4$.



(i) Joint measurements $\mathbf{Y}_{\Omega_3\times\Xi_3}$

FIG. 1. (*Continued*).

Algorithm 1 Cross: Tensor Completion with Noiseless Observations

1: Input: noiseless observations $\mathbf{Y}_{ijk}$, $(i,j,k)\in\boldsymbol{\Omega}$ from (2.3).
2: Construct Y_{Ξ_1}, Y_{Ξ_2}, X_{Ξ_3}, $Y_{\Omega_1\times\Xi_1}$, $Y_{\Omega_2\times\Xi_2}$, $Y_{\Omega_3\times\Xi_3}$ as (2.8).
3: Calculate

$$R_1 = Y_{\Xi_1}Y^{\dagger}_{\Omega_1\times\Xi_1}\in\mathbb{R}^{p_1\times m_1},\qquad R_2 = Y_{\Xi_2}Y^{\dagger}_{\Omega_2\times\Xi_2}\in\mathbb{R}^{p_2\times m_2},$$

$$R_3 = Y_{\Xi_3}Y^{\dagger}_{\Omega_3\times\Xi_3}\in\mathbb{R}^{p_3\times m_3}.$$

4: Calculate the final estimator

$$\hat{\mathbf{X}} = \mathbf{Y}_{[\Omega_1,\Omega_2,\Omega_3]}\times_1 R_1\times_2 R_2\times_3 R_3.$$

2.3. *Recovery algorithm—noiseless case.* When $\mathbf{X}$ is exactly low-rank and the observations are noiseless, that is, $\mathbf{Y}_{ijk}=\mathbf{X}_{ijk}$, we can recover $\mathbf{X}$ with the following algorithm. We first construct the *arm matricizations*, *joint matricizations* and *body matricizations* based on (2.4) and (2.5),

$$Y_{\Xi_t} = \mathcal{M}_t(\mathbf{Y}_{[1:p_t]\times\Xi_t})\in\mathbb{R}^{p_t\times g_t},\qquad\text{(Arm matricizations)} \tag{2.8}$$

$$Y_{t,\Omega} = \mathcal{M}_t(\mathbf{Y}_{[\Omega_1,\Omega_2,\Omega_3]})\in\mathbb{R}^{m_t\times\prod_{s\neq t}m_s},\qquad\text{(Body matricizations)} \tag{2.9}$$

$$Y_{\Omega_t\times\Xi_t} = \mathcal{M}_t(\mathbf{Y}_{\Omega_t\times\Xi_t})\in\mathbb{R}^{m_t\times g_t}.\qquad\text{(Joint matricizations)} \tag{2.10}$$

In the noiseless setting, we propose the following formula to complete $\mathbf{X}$:

$$\hat{\mathbf{X}} = \mathbf{Y}_{[\Omega_1,\Omega_2,\Omega_3]}\times_1 R_1\times_2 R_2\times_3 R_3, \tag{2.11}$$

$$\text{where } R_t = Y_{\Xi_t}Y^{\dagger}_{\Omega_t\times\Xi_t}\in\mathbb{R}^{p_t\times m_t},\ t=1,2,3. \tag{2.12}$$

The procedure is summarized in Algorithm 1. The theoretical guarantee for this proposed algorithm is provided in Theorem 1.

THEOREM 1 (Exact recovery in noiseless setting). *Suppose* $\mathbf{X}\in\mathbb{R}^{p_1\times p_2\times p_3}$, $\operatorname{rank}(\mathbf{X})=(r_1,r_2,r_3)$. *Assume all Cross tensor measurements are noiseless, that is,* $\mathbf{Y}_{\boldsymbol{\Omega}}=\mathbf{X}_{\boldsymbol{\Omega}}$. *If* $\operatorname{rank}(Y_{\Omega_t\times\Xi_t})=r_t$ *and* $\min\{m_t,g_t\}\geq r_t$ *for* $t=1,2,3$ [*so that*

$|\mathbf{\Omega}| \geq r_1 r_2 r_3 + r_1(p_1 - r_1) + r_2(p_2 - r_2) + r_3(p_3 - r_3)]$, *then*

$$\begin{aligned} \mathbf{X} &= \mathbf{Y}_{[\Omega_1,\Omega_2,\Omega_3]} \times_1 R_1 \times_2 R_2 \times_3 R_3, \\ R_t &= Y_{\Xi_t} Y^{\dagger}_{\Omega_t \times \Xi_t}, \qquad t = 1, 2, 3. \end{aligned} \tag{2.13}$$

Moreover, if there are $\tilde{M}_t \in \mathbb{R}^{m_t \times r_t}$, $\tilde{N}_t \in \mathbb{R}^{g_t \times r_t}$ *such that* $\tilde{M}_t^\top X_{\Omega_t \times \Xi_t} \tilde{N}_t \in \mathbb{R}^{r_t \times r_t}$ *is nonsingular for* $t = 1, 2, 3$, *then we further have*

$$\mathbf{X} = \mathbf{Y}_{[\Omega_1,\Omega_2,\Omega_3]} \times_1 \tilde{R}_1 \times_2 \tilde{R}_2 \times_3 \tilde{R}_3, \qquad \tilde{R}_t = Y_{\Xi_t} \tilde{N}_t^\top \big(\tilde{M}^\top Y_{\Omega_t \times \Xi_t} \tilde{N}\big)^{-1} \tilde{M}_t^\top.$$

Theorem 1 shows that, in the noiseless setting, as long as $\min\{m_t, g_t\} \geq r_t$, both $\mathcal{M}_t(\mathbf{Y})$ and its m_t-by-g_t submatrix $Y_{\Omega_t \times \Xi_t}$ are of rank r_t, exact recovery by Algorithm 1 can be guaranteed. Therefore, the minimum required number of measurements for the proposed Cross tensor measurement scheme is $r_1 r_2 r_3 + r_1(p_1 - r_1) + r_2(p_2 - r_2) + r_3(p_3 - r_3)$ when we set $m_t = g_t = r_t$, which exactly matches the lower bound established in Proposition 1 and outperforms the previous methods in the literature.

On the other hand, Algorithm 1 heavily relies on the noiseless assumption. In fact, calculating $Y^{\dagger}_{\Omega_t \times \Xi_t} = (X_{\Omega_t \times \Xi_t} + Z_{\Omega_t \times \Xi_t})^{\dagger}$ is unstable even with low levels of noise, which ruins the performance of Algorithm 1. Since we rarely have noiseless observations in practice, we focus on the setting with nonzero noise for the rest of the paper.

2.4. *Recovery algorithm—noisy case.* In this section, we propose the following procedure for recovery in the noisy setting. The proposed algorithm is divided into four steps and an illustrative example is provided in Figure 2 for readers' better understanding.

- (*Step* 1: *Construction of matricizations*.) Same as Algorithm 1, Construct the *arm, body and joint matricizations* as (2.8) and (2.9) [see Figure 2(a)],

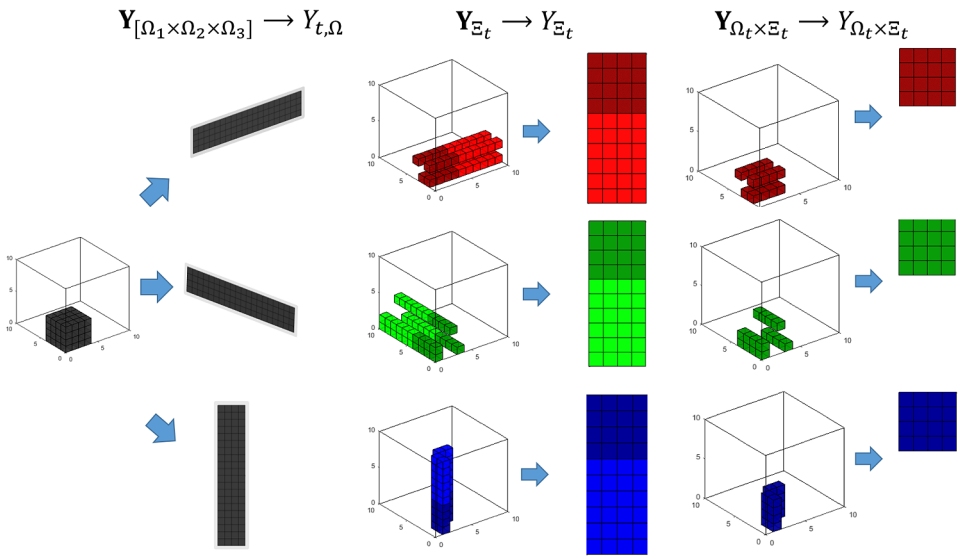
$$\begin{aligned} &\text{Arm} : Y_{\Xi_1}, Y_{\Xi_2}, Y_{\Xi_3}; \qquad \text{Body} : Y_{1,\Omega}, Y_{2,\Omega}, Y_{3,\Omega}; \\ &\text{Joint} : Y_{\Omega_1 \times \Xi_1}, Y_{\Omega_2 \times \Xi_2}, Y_{\Omega_3 \times \Xi_3}. \end{aligned}$$

- (*Step* 2: *Rotation*.) For $t = 1, 2, 3$, we calculate the singular value decompositions of Y_{Ξ_t} and $Y_{t,\Omega}$, then store

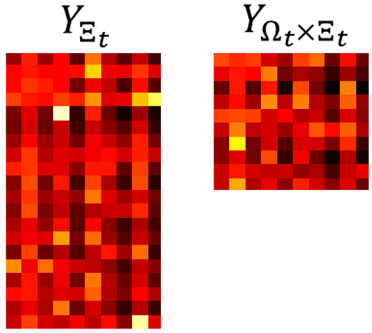
$$\begin{aligned} V_t^{(A)} &\in \mathbb{O}_{g_t} \qquad \text{as the right singular vectors of } Y_{\Xi_t}, \\ U_t^{(B)} &\in \mathbb{O}_{m_t} \qquad \text{as the left singular vectors of } Y_{t,\Omega}. \end{aligned} \tag{2.14}$$

  Here, the superscripts "(A), (B)" represent arm and body, respectively. We calculate the following rotation for arm and joint matricizations based on SVDs [see Figure 2(b), (c)]:

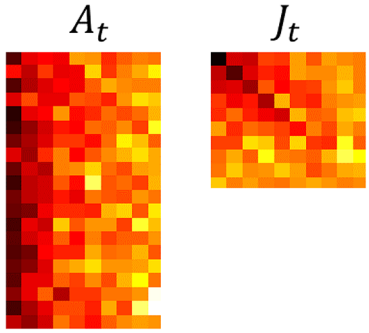
$$\begin{aligned} A_t &= Y_{\Xi_t} \cdot V_t^{(A)} \in \mathbb{R}^{p_t \times g_t}, \\ J_t &= \big(U_t^{(B)}\big)^\top \cdot Y_{\Omega_t \times \Xi_t} \cdot V_t^{(A)} \in \mathbb{R}^{m_t \times g_t}. \end{aligned} \tag{2.15}$$



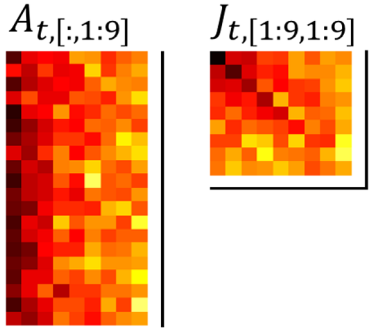
(a) Step 1. All matricizations: $Y_{t,\Omega}$, Y_{Ξ_t} and $Y_{\Omega_t\times\Xi_t}$



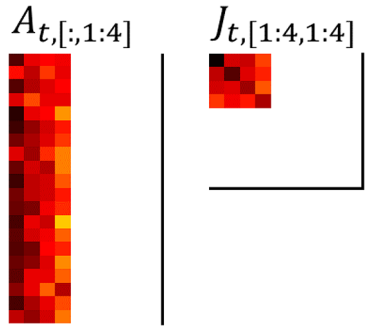
(b) Heatmap illustration of Y_{Ξ_t}, $Y_{\Omega_t\times\Xi_t}$ (darker blocks mean larger absolute values)



(c) Step 2. We obtain A_t, J_t after rotation



(d) Step 3. Intermediate process of trimming



(e) Step 3. Eventually located at $\hat{r}_t = 4$

FIG. 2. *Illustration of the proposed procedure in noisy setting.*

As we can see from Figure 2(c), the magnitude of A_t's columns and J_t's both columns and rows decreases front to back. Therefore, the important factors of Y_{Ξ_t} and $Y_{\Omega_t \times \Xi_t}$ are moved to front rows and columns in this step.

- (*Step* 3: *Adaptive trimming*.) Since A_t and J_t are contaminated with noise, in this step we denoise them by trimming the lower ranking columns of A_t and both lower ranking columns and rows of J_t. To decide the number of rows and columns to trim, it will be good to have an estimate for (r_1, r_2, r_3), say $(\hat{r}_1, \hat{r}_2, \hat{r}_3)$. We will show later in theoretical analysis that a good choice of $\hat{r}_t$ should satisfy

$$(J_t)_{[1:\hat{r}_t, 1:\hat{r}_t]} \text{ is nonsingular} \quad \text{and} \quad \big\|(A_t)_{[:,1:\hat{r}_t]}(J_t)^{-1}_{[1:\hat{r}_t, 1:\hat{r}_t]}\big\| \le \lambda_t \tag{2.16}$$

for $t = 1, 2, 3$. $\lambda_t = c_t\sqrt{p_t/m_t}$ is the tuning parameter here, and the discussion of selection method is provided a little while later. Our final estimator for r_t is the largest $\hat{r}_t$ that satisfies Condition (2.16), and can be found by verifying (2.16) for all possible r_t's. It is worth mentioning that this step shares similar ideas with structured matrix completion in Cai, Cai and Zhang (2016) [see Figure 2(d) and (e)].

- (*Step* 4: *Assembling*.) Finally, given $\hat{r}_1, \hat{r}_2, \hat{r}_3$ obtained from Step 3, we calculate

$$\bar{R}_t = (A_t)_{[:,1:\hat{r}_t]}(J_t)^{-1}_{[1:\hat{r}_t, 1:\hat{r}_t]}\big(V^{(A)}_{t,[1:\hat{r}_t,:]}\big)^{\top} \in \mathbb{R}^{p_t \times m_t}, \qquad t = 1, 2, 3, \tag{2.17}$$

and recover the original low-rank tensor $\mathbf{X}$ by

$$\hat{\mathbf{X}} = \mathbf{Y}_{[\Omega_1, \Omega_2, \Omega_3]} \times_1 \bar{R}_1 \times_2 \bar{R}_2 \times_3 \bar{R}_3. \tag{2.18}$$

The procedure is summarized as Algorithm 2. It is worth mentioning that both Algorithms 1 and 2 can be easily extended to fourth and higher-order tensors.

Selection of tuning parameter. The tuning parameter λ_t is a key factor to the performance of final estimation. Intuitively speaking, a larger value of λ_t yields a higher trimming level and a lower-rank estimation. As we will illustrate in the theoretical and numerical analyses, one can simply choose $\lambda_t = 3\sqrt{p_t/m_t}$ in a variety of situations. When more computing power is available, a practical data-driven approach for selecting c_t via K-fold subsampling cross-validation can be applied instead. The procedure is described as below, and the detailed numerical analyses for tuning parameter selection is provided in Section 4.

Suppose the body and arms of $\mathbf{X}$ are observed as (2.2)–(2.7) and let $T \subseteq [1.5, 4]$ be a grid of candidate values of c_t. For $l = 1, \ldots, L$, we randomly select subset $\Omega_t^{(\text{train},l)} \subseteq \Omega_t$ with cardinality $|\Omega_t^{(\text{train},l)}| \approx |\Omega_t| \cdot (K-1)/K$. Recall $\Xi_t \subseteq \Omega_{t+1} \times \Omega_{t+2}$, we further denote $\Xi_t^{(\text{train},l)} = \Xi_t \cap (\Omega_{t+1}^{(\text{train},l)} \times \Omega_{t+2}^{(\text{train},l)})$. Apply our proposed procedure based on the Cross measurements

$$\begin{aligned}\boldsymbol{\Omega}^{(\text{train},l)} &= \Omega_1^{(\text{train},l)} \times \Omega_2^{(\text{train},l)} \times \Omega_3^{(\text{train},l)} \\ &\quad \cup \big([1:p_1] \times \Xi_1^{(\text{train},l)}\big) \cup \big([1:p_2] \times \Xi_2^{(\text{train},l)}\big) \cup \big([1:p_3] \times \Xi_3^{(\text{train},l)}\big)\end{aligned}$$

Algorithm 2 Noisy Tensor Completion with Cross Measurements

Input: entries $\mathbf{Y}_{ijk}$, $(i,j,k) \in \boldsymbol{\Omega}$ from (2.3), $\lambda_1, \lambda_2, \lambda_3$.

Construct arm, body and joint matricizations as (2.8) and (2.9),

$$Y_{\Xi_t} \in \mathbb{R}^{p_t \times g_t}, \qquad Y_{\Omega_t \times \Xi_t} \in \mathbb{R}^{m_t \times g_t},$$
$$Y_{t,\Omega} \in \mathbb{R}^{m_t \times (\prod_{s \neq t} m_s)}, \qquad t = 1, 2, 3.$$

Calculate $U_t^{(B)}$ and $V_t^{(A)}$ via SVDs:

$$U_t^{(B)} \in \mathbb{O}_{m_t} \qquad \text{as the left singular vectors of } Y_{t,\Omega};$$
$$V_t^{(A)} \in \mathbb{O}_{g_t} \qquad \text{as the right singular vectors of } Y_{\Xi_t}.$$

Rotate the arm and joint measurements as

$$A_t = Y_{\Xi_t} \cdot V^{(A)} \in \mathbb{R}^{p_t \times g_t}, \qquad J_t = \big(U^{(B)}\big)^\top \cdot Y_{\Omega_t \times \Xi_t} \cdot V^{(A)} \in \mathbb{R}^{m_t \times g_t}.$$

for t = 1, 2, 3 **do**
  for $s = \min\{g_t, m_t\} : -1 : 1$ **do**
    if $J_{t,[1:s,1:s]}$ is not singular and $\|A_{t,[:,1:s]} J_{t,[1:s,1:s]}^{-1}\| \le \lambda_t$ **then**
      $\hat{r}_t = s$; **break** from the loop;
    end if
  end for
  If $\hat{r}_t$ is still unassigned **then** $\hat{r}_t = 0$.
end for

Calculate

$$\bar{R}_t = A_{t,[:,1:s]} J_{t,[1:s,1:s]}^{-1} \big(V_{t,[1:\hat{r}_t,:]}^{(A)}\big)^\top \in \mathbb{R}^{p_t \times m_t}, \qquad t = 1, 2, 3.$$

Compute the final estimator

$$\hat{\mathbf{X}} = \mathbf{Y}_{[\Omega_1,\Omega_2,\Omega_3]} \times_1 \bar{R}_1 \times_2 \bar{R}_2 \times_3 \bar{R}_3.$$

with each $c_t \in T$, then denote the resulting estimation as $\hat{\mathbf{X}}^l(c_t)$ for $l = 1, \ldots, N$. Next, the prediction error is evaluated on the observations outside the training set

$$\hat{R}(c_t) = \sum_{l=1}^{L} \sum_{(i,j,k) \in \boldsymbol{\Omega} \setminus \boldsymbol{\Omega}^{(\text{train},l)}} \big|\big(\hat{\mathbf{X}}^l(c_t)\big)_{ijk} - \mathbf{Y}_{ijk}\big|^2, \qquad c_t \in T,$$

where $\boldsymbol{\Omega}$ is defined (2.3). Finally, we select $c_t^* = \arg\min_{c_t \in T} \hat{R}(c_t)$, and apply the proposed procedure again with tuning parameter $\lambda_t = c_t^* \sqrt{p_t / m_t}$.

3. Theoretical analysis. In this section, we investigate the theoretical performance for the proposed procedure in the last section. Recall that our goal is to recover $\mathbf{X}$ from $\mathbf{Y}_{\boldsymbol{\Omega}}$ based on (2.3). Similarly, one can further define the arm, joint

and body matricizations for $\mathbf{X}$, $\mathbf{Z}$, that is, X_{Ξ_t}, $X_{\Omega_t\times\Xi_t}$, $X_{t,\Omega}$, Z_{Ξ_t}, $Z_{\Omega_t\times\Xi_t}$ and $Z_{t,\Omega}$ for $t=1,2,3$ in the same fashion as Y_{Ξ_t}, $Y_{\Omega_t\times\Xi_t}$ and $Y_{t,\Omega}$ in (2.8), (2.9) and (2.10). We first present the following theoretical guarantees for low-rank tensor completion based on noisy observations via Algorithm 2.

THEOREM 2. *Suppose* $\mathbf{X}\in\mathbb{R}^{p_1\times p_2\times p_3}$, *rank*$(\mathbf{X})=(r_1,r_2,r_3)$. *Assume we observe* $\mathbf{Y}_{\boldsymbol{\Omega}}$ *based on Cross tensor measurement scheme* (2.4), *where* $\mathbf{X}_\Omega$ *satisfies* $\operatorname{rank}(X_{\Omega_t\times\Xi_t})=r_t$ *and*

$$\begin{aligned}&\sigma_{r_t}(X_{\Omega_t\times\Xi_t})>5\|Z_{\Omega_t\times\Xi_t}\|,\qquad \sigma_{r_t}(X_{\Xi_t})>5\|Z_{\Xi_t}\|,\\ &\sigma_{r_t}(X_{t,\Omega})>5\|Z_{t,\Omega}\|.\end{aligned}\tag{3.1}$$

We further define

$$\xi_t=\|X^{\dagger}_{\Omega_t\times\Xi_t}X_{t,\Omega}\|,\qquad t=1,2,3.\tag{3.2}$$

Applying Algorithm 2 with $\lambda_1,\lambda_2,\lambda_3$ *satisfying*

$$\lambda_t\ge 2\|X_{t,\Xi_t}X^{\dagger}_{\Omega_t\times\Xi_t}\|,\qquad t=1,2,3,\tag{3.3}$$

we have the following upper bound results for some uniform constant C:

$$\begin{aligned}\|\hat{\mathbf{X}}-\mathbf{X}\|_{\mathrm{HS}}&\le C\lambda_1\lambda_2\lambda_3\|\mathbf{Z}_{[\Omega_1,\Omega_2,\Omega_3]}\|_{\mathrm{HS}}\\&\quad+C\lambda_1\lambda_2\lambda_3\left(\sum_{t=1}^{3}\xi_t\|Z_{\Omega_t\times\Xi_t}\|_F+C\sum_{t=1}^{3}\frac{\xi_t}{\lambda_t}\|Z_{\Xi_t}\|_F\right),\\ \|\hat{\mathbf{X}}-\mathbf{X}\|_{\mathrm{op}}&\le C\lambda_1\lambda_2\lambda_3\|\mathbf{Z}_{[\Omega_1,\Omega_2,\Omega_3]}\|_{\mathrm{op}}\\&\quad+C\lambda_1\lambda_2\lambda_3\left(\sum_{t=1}^{3}\xi_t\|Z_{\Omega_t\times\Xi_t}\|+C\sum_{t=1}^{3}\frac{\xi_t}{\lambda_t}\|Z_{\Xi_t}\|\right).\end{aligned}$$

It is helpful to explain the meanings of the conditions used in Theorem 2. The singular value gap condition (3.1) is assumed in order to guarantee that signal dominates the noise in the observed blocks. λ_t and ξ_t are important factors in our analysis which represent "arm-joint" and "joint-body" ratio, respectively. These factors roughly indicate how much information is contained in the body and arm measurements and how much impact the noisy terms have on the upper bound, all of which implicitly indicate the difficulty of the problem. Based on λ_t, ξ_t, we consider the following classes of low-rank tensors, the perturbation $\mathbf{Z}$ and indices of observations:

$$\begin{aligned}\mathcal{F}&=\mathcal{F}_{\{\lambda_t\},\{\xi_t\}}\\&=\left\{(\mathbf{X},\mathbf{Z},\Omega_t,\Xi_t):\begin{array}{l}\mathbf{X}\in\mathbb{R}^{p_1\times p_2\times p_3},\operatorname{rank}(\mathbf{X})\le(r_1,r_2,r_3);\\ \|X_{\Xi_t}X^{\dagger}_{\Omega_t\times\Xi_t}\|\le\lambda_t,\|X^{\dagger}_{\Omega_t\times\Xi_t}X_{t,\Omega}\|\le\xi_t.\\ \sigma_{r_t}(X_{\Omega_t\times\Xi_t})\ge 5\|Z_{\Omega_t\times\Xi_t}\|,\sigma_{r_t}(X_{\Xi_t})\ge 5\|Z_{\Xi_t}\|,\\ \sigma_{r_t}(X_{t,\Omega})\ge 5\|Z_{t,\Omega}\|;\end{array}\right\}\end{aligned}\tag{3.4}$$

and provide the following lower bound result over $\mathcal{F}_{\{\lambda_t\},\{\xi_t\}}$.

THEOREM 3 (Lower bound). *Suppose positive integers r_t, p_t satisfy $4 \le r_t \le p_t/2$. The arm, body and joint measurement errors are bounded as*

$$\begin{aligned} \|\mathbf{Z}_{[\Omega_1,\Omega_2,\Omega_3]}\|_{\mathrm{HS}} &\le C^{(B)}, \\ \|Z_{t,\Xi_t}\|_F &\le C_t^{(A)}, \\ \|Z_{t,\Omega_t\times\Xi_t}\|_F &\le C_t^{(J)}. \end{aligned} \tag{3.5}$$

If $C_t^{(J)} \le \min\{C_t^{(A)}, C^{(B)}\}$, $\xi_t \ge 3$, $\lambda_t > 1$, then there exists uniform constant $c > 0$ such that

$$\begin{aligned} &\inf_{\hat{\mathbf{X}}} \sup_{\substack{(\mathbf{X},\mathbf{Z},\Omega_t,\Xi_t)\in\mathcal{F} \\ \mathbf{Z}\ \text{satisfies (3.5)}}} \|\hat{\mathbf{X}} - \mathbf{X}\|_{\mathrm{HS}} \\ &\quad \ge c\lambda_1\lambda_2\lambda_3 C^{(B)} + c\lambda_1\lambda_2\lambda_3 \sum_{t=1}^{3}\left(\xi_t C_t^{(J)} + \frac{\xi_t}{\lambda_t} C_t^{(A)}\right). \end{aligned} \tag{3.6}$$

Similarly, suppose $C^{(B)}$, $C_t^{(A)}$, $C_t^{(J)}$ are the upper bound for arm, body and joint measurement errors in tensor and matrix operator norms, respectively, that is,

$$\begin{aligned} \|\mathbf{Z}_{[\Omega_1,\Omega_2,\Omega_3]}\|_{\mathrm{op}} &\le C^{(B)}, \\ \|Z_{t,\Xi_t}\| &\le C_t^{(A)}, \\ \|Z_{t,\Omega_t\times\Xi_t}\| &\le C_t^{(J)}. \end{aligned} \tag{3.7}$$

Suppose $C_t^{(J)} \le \min\{C_t^{(A)}, C^{(B)}\}$, $\xi_t \ge 3$, $\lambda_t > 1$, then

$$\begin{aligned} &\inf_{\hat{\mathbf{X}}} \sup_{\substack{(\mathbf{X},\mathbf{Z},\Omega_t,\Xi_t)\in\mathcal{F} \\ \mathbf{Z}\ \text{satisfies (3.7)}}} \|\hat{\mathbf{X}} - \mathbf{X}\|_{\mathrm{op}} \\ &\quad \ge c\lambda_1\lambda_2\lambda_3 C^{(B)} + c\lambda_1\lambda_2\lambda_3 \sum_{t=1}^{3}\left(\xi_t C_t^{(J)} + \frac{\xi_t}{\lambda_t} C_t^{(A)}\right). \end{aligned} \tag{3.8}$$

REMARK 2. Theorems 2 and 3 together yield the optimal rate of recovery in $\mathcal{F}$ in both Hilbert–Schmidt and operator norms:

$$\inf_{\hat{\mathbf{X}}} \sup_{\substack{(\mathbf{X},\mathbf{Z},\Omega_t,\Xi_t)\in\mathcal{F} \\ \mathbf{Z}\ \text{satisfies (3.5)}}} \|\hat{\mathbf{X}} - \mathbf{X}\|_{\mathrm{HS}} \asymp \lambda_1\lambda_2\lambda_3 \left\{ C^{(B)} + \sum_{t=1}^{3}\left(\xi_t C_t^{(J)} + \frac{\xi_t}{\lambda_t} C_t^{(A)}\right)\right\},$$

$$\inf_{\hat{\mathbf{X}}} \sup_{\substack{(\mathbf{X},\mathbf{Z},\Omega_t,\Xi_t)\in\mathcal{F} \\ \mathbf{Z}\ \text{satisfies (3.7)}}} \|\hat{\mathbf{X}} - \mathbf{X}\|_{\mathrm{op}} \asymp \lambda_1\lambda_2\lambda_3 \left\{ C^{(B)} + \sum_{t=1}^{3}\left(\xi_t C_t^{(J)} + \frac{\xi_t}{\lambda_t} C_t^{(A)}\right)\right\}.$$

REMARK 3. There have been a number of existing lower bound results on the estimation error in matrix/tensor estimation literatures. For example, Negahban and Wainwright (2012) considered the setting that one observes uniformly randomly selected entries with noise; Candès and Plan (2011) developed a sharp oracle lower bound when the measurement matrices satisfies restrict isometry property (RIP), Koltchinskii, Lounici and Tsybakov (2011), Rohde and Tsybakov (2011) considered the setting that the measurement matrices/tensors are i.i.d. randomly generated. Raskutti, Yuan and Chen (2017) considered multiresponse regularized tensor regression, autoregressive regression and interaction model with random Gaussian measurements. As the proposed Cross tensor measurement scheme does not satisfy the assumptions of these existing settings, these previous results cannot be directly applied.

As we can see from the theoretical analyses, the choice of λ_t is crucial toward the recovery performance of Algorithm 2. Theorem 2 provides a guideline for such a choice depending on the unknown parameter $\|X_{\Xi_t} X^{\dagger}_{\Omega_t \times \Xi_t}\|$, which is hard to obtain in practice. However, we can choose $\lambda_t = 3\sqrt{p_t/m_t}$ in a variety of settings. Specifically, in the analysis below, we show under random sampling scheme that Ω_1, Ω_2, Ω_3, Ξ_1, Ξ_2, Ξ_3 are uniformly randomly selected from $[1 : p_1]$, $[1 : p_2]$, $[1 : p_3]$, $\Omega_2 \times \Omega_3$, $\Omega_3 \times \Omega_1$, $\Omega_1 \times \Omega_2$, Algorithm 2 with $\lambda_t = 3\sqrt{p_t/m_t}$ will have guaranteed performance. The choice of $\lambda_t = 3\sqrt{p_t/m_t}$ and the one by cross-validation will be further examined in simulation studies later.

THEOREM 4. *Suppose* $\mathbf{X}$ *is with Tucker decomposition* $\mathbf{X} = \mathbf{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$, *where* $\mathbf{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$, $U_1 \in \mathbb{O}_{p_1,r_1}$, $U_2 \in \mathbb{O}_{p_2,r_2}$, $U_3 \in \mathbb{O}_{p_3,r_3}$ *and* U_1, U_2, U_3, $\mathcal{M}_1(\mathbf{S} \times_2 U_2 \times_3 U_3)$, $\mathcal{M}_2(\mathbf{S} \times_1 U_1 \times_3 U_3)$, $\mathcal{M}_3(\mathbf{S} \times_1 U_1 \times_2 U_2)$ *all satisfy the matrix incoherence conditions*:

$$\begin{aligned} &\frac{p_t}{r_t} \max_{1 \le j \le p_t} \big\|\mathbb{P}_{U_t} e_j^{(p_t)}\big\|_2^2 \le \rho, \\ &\frac{\prod_{s \ne t} p_s}{r_t} \max_{1 \le j \le \prod_{s \ne t} p_s} \big\|\mathbb{P}_{\mathcal{M}_t(\mathbf{S} \times_{(t+1)} U_{t+1} \times_{(t+2)} U_{t+2})^{\top}} \cdot e_j^{(\prod_{s \ne t} p_s)}\big\|_2^2 \le \rho, \end{aligned} \tag{3.9}$$

where $e_j^{(p)}$ *is the* j*th canonical basis in* $\mathbb{R}^p$. *Suppose we are given random Cross tensor measurements that* Ω_t *and* Ξ_t *are uniformly randomly chosen* m_t *and* g_t *values from* $\{1, \dots, p_t\}$ *and* $\prod_{s \ne t} \Omega_s$, *respectively. If for* $t = 1, 2, 3$,

$$\begin{aligned} &\sigma_{\min}\big(\mathcal{M}_t(\mathbf{S})\big) \\ &\quad \ge \max\Bigg\{10\sqrt{\frac{p_1 p_2 p_3}{m_t g_t}}\|Z_{\Omega_t \times \Xi_t}\|, 10\sqrt{\frac{p_1 p_2 p_3}{m_1 m_2 m_3}}\|Z_{t,\Omega}\|, \\ &\qquad 19\sqrt{\frac{p_1 p_2 p_3}{p_t g_t}}\|Z_{\Xi_t}\|\Bigg\}, \end{aligned} \tag{3.10}$$

Algorithm 2 with $\lambda_t = 3\sqrt{p_t/m_t}$ *yields*

$$\|\hat{\mathbf{X}} - \mathbf{X}\|_{\mathrm{HS}} \le C\sqrt{\frac{p_1p_2p_3}{m_1m_2m_3}}\|\mathbf{Z}_{[\Omega_1,\Omega_2,\Omega_3]}\|_{\mathrm{HS}} + C\sqrt{p_1p_2p_3}\sum_{t=1}^{3}\Big(\frac{\|Z_{\Omega_t\times\Xi_t}\|_F}{\sqrt{g_tm_t}} + \frac{\|Z_{\Xi_t}\|_F}{\sqrt{g_tp_t}}\Big),$$

$$\|\hat{\mathbf{X}} - \mathbf{X}\|_{\mathrm{op}} \le C\sqrt{\frac{p_1p_2p_3}{m_1m_2m_3}}\|\mathbf{Z}_{[\Omega_1,\Omega_2,\Omega_3]}\|_{\mathrm{op}} + C\sqrt{p_1p_2p_3}\sum_{t=1}^{3}\Big(\frac{\|Z_{\Omega_t\times\Xi_t}\|}{\sqrt{g_tm_t}} + \frac{\|Z_{\Xi_t}\|}{\sqrt{g_tp_t}}\Big),$$

with probability at least $1 - 2\sum_{t=1}^{3} r_t\{\exp(-m_t/(16r_t\rho)) + \exp(-g_t/(64r_t\rho))\}$.

REMARK 4. The incoherence conditions (3.9) are widely used in matrix and tensor completion literature [see, e.g., Candès and Tao (2010), Recht (2011), Yuan and Zhang (2017)]. Their conditions basically characterize every entry of X as containing a similar level of information for the whole tensor. Therefore, we should have enough knowledge to recover the original tensor based on the observable entries.

REMARK 5. For a better illustration of the proposed procedure, it is helpful to briefly discuss the matrix counterpart of Cross tensor measurement scheme and recovery algorithm here. Suppose X is a p_1-by-p_2 unknown low-rank matrix, a row index subset $\Omega_1 \subseteq [1 : p_1]$ and a column index subset $\Omega_2 \subseteq [1 : p_2]$ are randomly generated, and one observe the rows $X_{[\Omega_1,:]}$ and columns $X_{[:,\Omega_2]}$. We aim to recover the original low-rank matrix X from observations of $X_{[\Omega_1,\Omega_2]}$, $X_{[\Omega_1,\Omega_2^c]}$ and $X_{[\Omega_1^c,\Omega_2]}$. This problem, which has been studied recently in Wagner and Zuk (2015) and Cai, Cai and Zhang (2016) in the context of *row and column matrix completion* and *structured matrix completion*, would be a matrix analogy to tensor completion via Cross tensor measurements. In the noiseless setting, it was shown that the low-rank matrix X can be recovered by the well-regarded Schur complement,

$$\hat{X} = X \qquad \text{where } \hat{X} = X_{[:,\Omega_2]}X^{\dagger}_{[\Omega_1,\Omega_2]}X_{[\Omega_1,:]}. \tag{3.11}$$

In the noisy setting, the estimation scheme based on a sequential truncation and MLE-based approach were proposed and analyzed in Wagner and Zuk (2015) and Cai, Cai and Zhang (2016), respectively.

Although the Cross tensor measurement scheme shares similarities with the above matrix completion setting, the proposed tensor recovery procedure shows distinct aspects and is much more difficult to analyze in various ways. First,

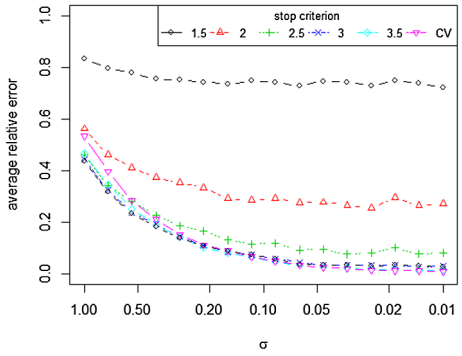
in matrix settings one fully observes an "L" shaped region including $\mathbf{X}_{[\Omega_1,\Omega_2]}$, $\mathbf{X}_{[\Omega_1^c,\Omega_2]}$ and $\mathbf{X}_{[\Omega_1,\Omega_2^c]}$ [Cai, Cai and Zhang (2016), Wagner and Zuk (2015)]. However, the analog of the "L" shape in tensor settings, $\mathbf{X}_{[\Omega_1^c,\Omega_2,\Omega_3]}$, $\mathbf{X}_{[\Omega_1,\Omega_2^c,\Omega_3]}$ and $\mathbf{X}_{[\Omega_1,\Omega_2,\Omega_3^c]}$, include $O(p_1m_2m_3 + m_1p_2m_3 + m_1m_2p_3)$ entries in total, which is far more than the level achieved by Cross. Such difference also makes it difficult to directly apply the original analysis in matrix setting to Cross. Second, the Cross tensor measurement scheme involves more complicated tensor operations than its matrix counterpart. In particular, the tensor recovery formula (2.13) involves seven terms with three inverses including body, arms and joints, making its analysis far more demanding than that of (3.11) where only three submatrices and one inverse are involved. Third, the analysis of the proposed tensor completion algorithm relies on tensor terminology and algebra, which are much more complicated than the matrix ones. For example, the ℓ_2 and Frobenius norms of a matrix can be well characterized by its singular values. However, there is no such correspondence for tensors.

4. Simulation study. In this section, we investigate the numerical performance of the proposed procedure in a variety of settings. We repeat each setting 1000 times and record the average relative loss in Hilbert–Schmidt norm, that is, $\|\hat{\mathbf{X}} - \mathbf{X}\|_{\mathrm{HS}}/\|\mathbf{X}\|_{\mathrm{HS}}$.

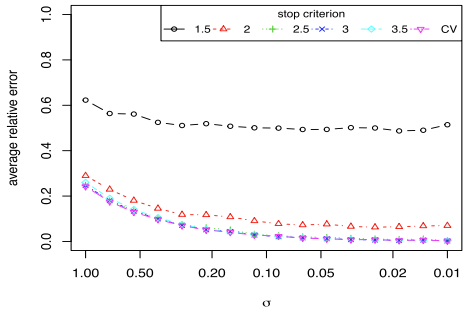
We first focus on the setting with i.i.d. Gaussian noise. To be specific, we randomly generate $\mathbf{X} = \mathbf{S} \times_1 E_1 \times_2 E_2 \times E_3$, where $\mathbf{S} \in \mathbb{R}^{r_1\times r_2\times r_3}$, $E_1 \in \mathbb{R}^{p_1\times r_1}$, $E_2 \in \mathbb{R}^{p_2\times r_2}$, $E_3 \in \mathbb{R}^{p_3\times r_3}$ are all with i.i.d. standard Gaussian entries. We can verify that $\mathbf{X}$ becomes a rank-(r_1, r_2, r_3) tensor with probably 1 whenever r_1, r_2, r_3 satisfy $\max^2(r_1, r_2, r_3) \le r_1r_2r_3$. Then we generate the Cross tensor measurement $\boldsymbol{\Omega}$ as in (2.3) with Ω_t including uniformly randomly selected m_t values from $[1 : p_t]$ and Ξ_t including uniformly randomly selected g_t values from $\prod_{s\neq t}\Omega_s$, and contaminate $\mathbf{X}_{\boldsymbol{\Omega}}$ with i.i.d. Gaussian noise: $\mathbf{Y}_{\boldsymbol{\Omega}} = \mathbf{X}_{\boldsymbol{\Omega}} + \mathbf{Z}_{\boldsymbol{\Omega}}$, where $Z_{ijk} \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$. Under such configuration, we study the influence of different factors, including λ_t, σ, m_t, g_t, p_t to the numerical performance.

Under the Gaussian noise setting, we first compare different choices of tuning parameters λ_t. To be specific, set $p_1 = p_2 = p_3 \in \{50, 80\}$, $m_1 = m_2 = m_3 = g_1 = g_2 = g_3 \in \{10, 15\}$, $r_1 = r_2 = r_3 = 3$ and let σ range from 1 to 0.01. We consider both the fixed tuning parameters: $\lambda_t \in [1.5\sqrt{p_t/m_t}, 3.5\sqrt{p_t/m_t}]$, and the one selected by 5-fold cross-validation. The average relative Hilbert–Schmidt norm loss of $\hat{\mathbf{X}}$ from Algorithm 2 is reported in Figure 3. It can be seen that the average relative loss decays when the noise level is decreasing. After comparing different choices of λ_t, $3\sqrt{p_t/m_t}$ and cross-validation scheme works the best under different σ, which matches our previous suggestions.

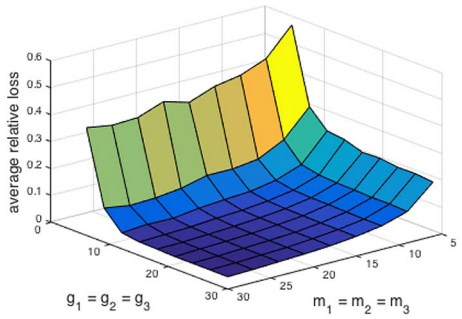
We also compare the effects of $m_t = |\Omega_t|$ and $g_t = |\Xi_t|$ in the numerical performance of Algorithm 2. We set $p_1 = p_2 = p_3 = 50, r_1 = r_2 = r_3 = 3, \sigma = 0.3$,



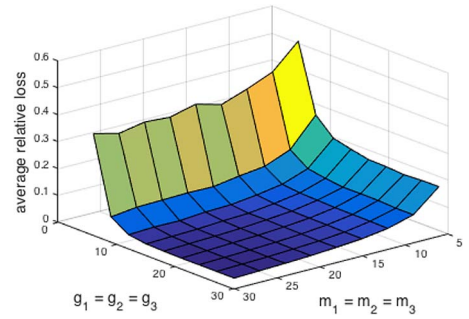
(a) Varying noise level σ and stop criteria ($p_t = 50$, $m_t = 10$)



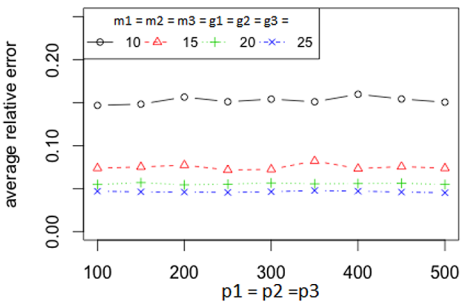
(b) Varying noise level σ and stop criteria ($p_t = 80$, $m_t = 15$)



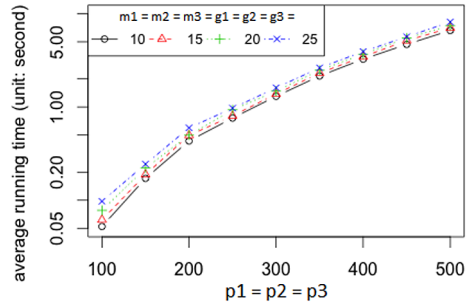
(c) Varying g_t and m_t: $\lambda_t = 3\sqrt{p_t/m_t}$



(d) Varying g_t and m_t: λ_t selected by CV



(e) Average relative loss when p_t, m_t, g_t are varying



(f) Average time cost when p_t, m_t, g_t are varying

FIG. 3. *Numerical performance under Gaussian noise settings.*

let g_t, m_t vary from 6 to 30 and let λ_t be either fixed as $3\sqrt{p_t/m_t}$ or chosen by 5-fold cross-validation. The average relative Hilbert–Schmidt norm loss are plotted in Figure 3(c) and (d). It can be seen that as g_t, m_t grow, namely when more entries

are observable, better recovery performance can be achieved. The performance of $\lambda_t = 3\sqrt{p_t/m_t}$ is still similar to the one by cross-validation.

To further study the impact of high dimensionality to the proposed procedure, we consider the setting where the dimension of $\mathbf{X}$ further grows. Here, $r_1 = r_2 = r_3 = 3, \sigma = 0.3$, $m_1 = m_2 = m_3 = g_1 = g_2 = g_3 \in \{10, 15, 20, 25\}$, p_1, p_2, p_3 grow from 100 to 500 and $\lambda_t = 3\sqrt{p_t/m_t}$. The average relative loss in Hilbert–Schmidt norm and average running time are provided in Figure 3(e) and (f), respectively. Particularly, the recovery of 500-by-500-by-500 tensors involves 125,000,000 variables, but the proposed procedure provides stable recovery within 10 seconds on average by the PC with 3.1 GHz CPU, which demonstrates the efficiency of our proposed algorithm.

The next simulation setting is designed to compare the proposed Algorithm 2 with the Low-rank Tensor Completion (LRTC) proposed by Liu et al. (2013). LRTC is a convexified tensor completion method based on matricization nuclear norm minimization. To avoid nonconvergence runs of LRTC, we set the maximum number of iterations as 500 and all the other tuning parameters as the default values. Let $p_1 = p_2 = p_3 = 50$, $r_1 = r_2 = r_3 = 3$; we consider two settings: (i) $\sigma^2 = 0.3$, m_t, g_t vary from 6 to 20; (ii) $m_t = g_t = 10$, σ^2 varies from 0.01 to 1. We apply both LRTC (with the package downloaded from the authors' website) and our proposed procedure, then present the estimation error in relative Hilbert–Schmidt norm and average running time in Figure 4. It is clear that our proposed procedure achieves significantly smaller estimation error in much shorter running time, which substantially outperforms LRTC.

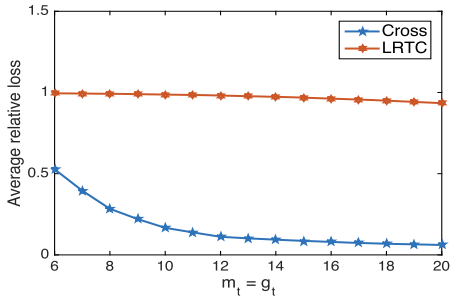
Then we move on to the setting where observations take discrete random values. High-dimensional count data commonly appear in a wide range of applications, including fluorescence microscopy, network flow and microbiome [see, e.g., Cao and Xie (2016), Cao, Zhang and Li (2017), Jiang, Raskutti and Willett (2015), Nowak and Kolaczyk (2000), etc.], where Poisson and multinomial distributions are often used in modeling the counts. In this simulation study, we generate $\mathbf{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$, $E_t \in \mathbb{R}^{p_t \times r_t}$ as absolute values of i.i.d. standard normal random variables, and calculate

$$\mathbf{X} = \frac{\mathbf{S} \times_1 E_1 \times_2 E_2 \times_3 E_3}{\sum_{i=1}^{p_1} \sum_{j=1}^{p_2} \sum_{k=1}^{p_3} (\mathbf{S} \times_1 E_1 \times_2 E_2 \times_3 E_3)_{ijk}}.$$

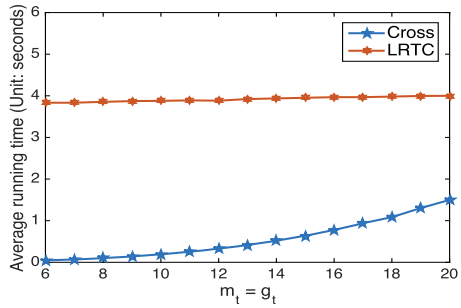
Ω_t, Ξ_t are generated similarly as before, $p_1 = p_2 = p_3 = 50$, $r_1 = r_2 = r_3 = 3$, $m_1 = m_2 = m_3 = g_1 = g_2 = g_3 \in \{10, 15, 20, 25\}$, and $\mathbf{Y} = (\mathbf{Y}_{ijk})_{1 \le i \le p_1, 1 \le j \le p_2, 1 \le k \le p_3}$ are Poisson or multinomial distributed:

$$\mathbf{Y}_{ijk} \sim \text{Poisson}(H\mathbf{X}_{ijk}) \quad \text{or} \quad \mathbf{Y}_{ijk} \sim \text{Multinomial}(N; \mathbf{X}).$$

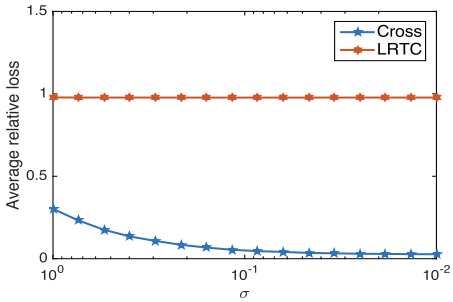
Here, H is a known intensity parameter in Poisson observations and N is the total count parameter in multinomial observations. As shown in Figure 5, the proposed Algorithm 2 performs stably for these two types of noisy structures.



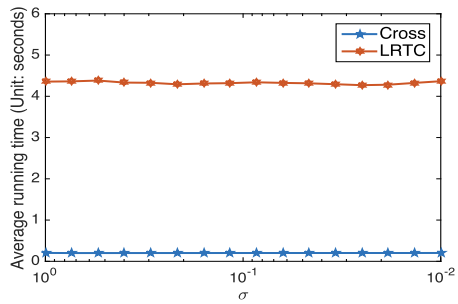
(a) Average relative loss with varying $m_t, g_t \in [6:20]$



(b) Average running time with varying $m_t, g_t \in [6:20]$

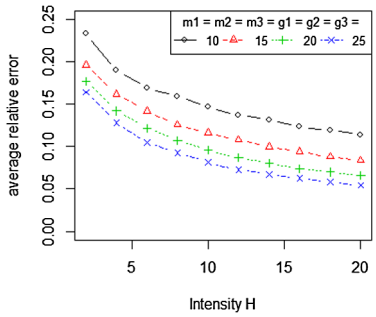


(c) Average relative loss with varying $\sigma^2 \in [0.01, 1]$

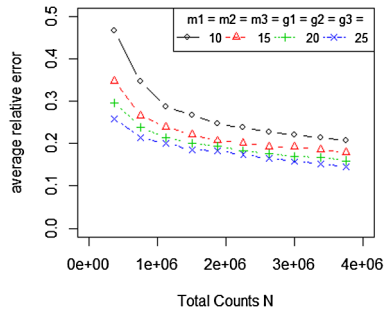


(d) Average running time with varying $\sigma^2 \in [0.01, 1]$

FIG. 4. *Average relative loss and running time for Cross and LRTC.*



(a) Poisson model with varying m_t, g_t and intensity H



(b) Multinomial model with varying m_t, g_t, total count N

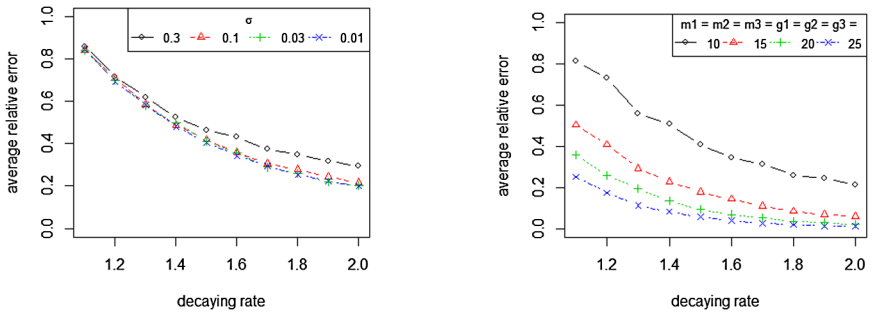
FIG. 5. *Average relative loss in HS norm based on Poisson and multinomial observations. Here,* $p_1 = p_2 = p_3 = 50, r_1 = r_2 = r_3 = 3$.

Although $\mathbf{X}$ is assumed to be exactly low rank in all theoretical studies, it is not necessary in practice. In fact, our simulation study shows that Algorithm 2 performs well when $\mathbf{X}$ is only approximately low rank. Specifically, we fix $p_1 = p_2 = p_3 = 50$, generate $\mathbf{W} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ from i.i.d. standard normal, set $U_1 \in \mathbb{O}_{p_1}, U_2 \in \mathbb{O}_{p_2}, U_3 \in \mathbb{O}_{p_3}$ as uniform random orthogonal matrices, and $E_t = \operatorname{diag}(1, 1, 1^{-\alpha}, \ldots, (p_t - 2)^{-\alpha})$. $\mathbf{X}$ is then constructed as

$$\mathbf{X} = \mathbf{W} \times_1 (E_1 U_1) \times_2 (E_2 U_2) \times_3 (E_3 U_3).$$

Here, α measures the decaying rate of singular values of each matricization of $\mathbf{X}$ and $\mathbf{X}$ becomes exactly rank-(3, 3, 3) when $\alpha = \infty$. We consider different decay rates α, noise levels σ and observation set sizes m_t and g_t. The corresponding average relative Hilbert–Schmidt norm loss is reported in Figure 6. It can be seen that although $\mathbf{X}$ is not exactly low rank, as long as the singular values of each matricization of $\mathbf{X}$ decay sufficiently fast, a desirable completion of $\mathbf{X}$ can still be achieved, which again demonstrates the robustness of the proposed procedure.

5. Real data illustration. In this section, we apply the proposed Cross tensor measurements scheme to a real dataset on attention hyperactivity disorder (ADHD) available from ADHD-200 Sample Initiative (http://fcon_1000.projects.nitrc.org/indi/adhd200/). ADHD is a common disease that affects at least 5–7% of school-age children and may accompany patients throughout their life with direct costs of at least $36 billion per year in the United States. Despite being the most common mental disorder in children and adolescents, the cause of ADHD is largely unclear. To investigate the disease, the ADHD-200 study covered 285 subjects diagnosed with ADHD and 491 control subjects. After data cleaning, the dataset contains 776 tensors of dimension 121-by-145-by-121: $\mathbf{Y}_i$, $i = 1, \ldots, 776$. The storage space for these data through naive format is $121 \times 145 \times 121 \times 776 \times 4\text{B} \approx 6.137$ GB,



(a) Fixed $m_t = g_t = 10$, varying singular value decaying rate α and noise level σ

(b) Fixed $\sigma = 0.3$, varying singular value decaying rate α and m_t, g_t

FIG. 6. *Average relative loss for approximate low-rank tensors.*

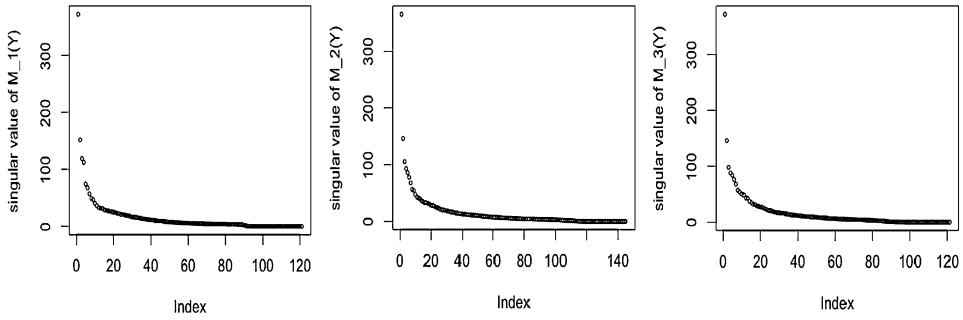


FIG. 7. *Singular value decompositions for each matricization of* $\mathbf{Y}$.

which makes it difficult and costly for sampling, storage and computation. Therefore, we hope to reduce the sampling size for ADHD brain imaging data via the proposed Cross tensor measurement scheme.

Figure 7 shows the singular values of each matricization of a randomly selected $\mathbf{Y}_i$. We can see that $\mathbf{Y}_i$ is approximately Tucker low rank. Similar to previous simulation settings, we uniformly randomly select $\Omega_t \subseteq [1 : p_t]$, $\Xi_t \subseteq \prod_{s\neq t} \Omega_s$ such that $|\Omega_t| = m_t, |\Xi_t| = g_t$. Particularly, we choose $m_t = \text{round}(\rho \cdot p_t)$, $g_t = \text{round}(m_1 m_2 m_3 / p_t)$, where ρ varies from 0.1 to 0.5 and round$(\cdot)$ is the function that rounds its input to the nearest integer. After observing partial entries of each tensor, we apply Algorithm 2 with $\lambda_t = 3\sqrt{p_t/m_t}$ to obtain $\hat{\mathbf{X}}$. Different from some of the previous studies [e.g., Zhou, Li and Zhu (2013)], our algorithm is adaptive and tuning-free so that we do not need to subjectively specify the rank of the target tensors beforehand.

Suppose $\text{rank}(\hat{\mathbf{X}}) = (\hat{r}_1, \hat{r}_2, \hat{r}_3)$, $\hat{U}_1 \in \mathbb{O}_{p_1,r_1}, \hat{U}_2 \in \mathbb{O}_{p_2,r_2}, \hat{U}_3 \in \mathbb{O}_{p_3,r_3}$ are the left singular vectors of $\mathcal{M}_1(\hat{\mathbf{X}})$, $\mathcal{M}_2(\hat{\mathbf{X}})$, $\mathcal{M}_3(\hat{\mathbf{X}})$, respectively. We are interested in investigating the performance of $\hat{\mathbf{X}}$, but the absence of the true tank of the underlying tensor $\mathbf{X}$ makes it difficult to directly compare $\hat{\mathbf{X}}$ and $\mathbf{X}$. Instead, we compare $\hat{\mathbf{X}}$ with $\tilde{\mathbf{X}}$, where $\tilde{\mathbf{X}}$ is the rank-$(\hat{r}_1, \hat{r}_2, \hat{r}_3)$ tensor obtained through the high-order singular value decomposition (HOSVD) [see, e.g., Kolda and Bader (2009)] based on all observations in $\mathbf{Y}$:

$$\tilde{\mathbf{X}} = \mathbf{Y} \times_1 \mathbb{P}_{\tilde{U}_1} \times_2 \mathbb{P}_{\tilde{U}_2} \times_3 \mathbb{P}_{\tilde{U}_3}. \tag{5.1}$$

Here, $\tilde{U}_1 \in \mathbb{O}_{p_1,r_1}, \tilde{U}_2 \in \mathbb{O}_{p_2,r_2}, \tilde{U}_3 \in \mathbb{O}_{p_3,r_3}$ are the first r_1, r_2 and r_3 left singular vectors of $\mathcal{M}_1(\mathbf{Y})$, $\mathcal{M}_2(\mathbf{Y})$ and $\mathcal{M}_3(\mathbf{Y})$, respectively. Particularly, we compare $\|\hat{\mathbf{X}} - \mathbf{Y}\|_{\text{HS}}$ and $\|\tilde{\mathbf{X}} - \mathbf{Y}\|_{\text{HS}}$, that is, the rank-$(r_1, r_2, r_3)$ approximation based on limited number of Cross tensor measurements and the approximation based on all measurements. We also compare $\hat{U}_1$, $\hat{U}_2$, $\hat{U}_3$ and $\tilde{U}_1$, $\tilde{U}_2$, $\tilde{U}_3$ by calculating $\frac{1}{\sqrt{\hat{r}_t}}\|\hat{U}_t^\top \tilde{U}_t\|_F$. The study is performed on 10 randomly selected images and repeated 100 times for each of them. We can immediately see from the result in

TABLE 1
Comparison between $\hat{\mathbf{X}}$ *and* $\tilde{\mathbf{X}}$ *for ADHD brain imaging data*

m_t	Sampling Ratio [See (2.6)]	$\frac{\|\hat{\mathbf{X}}-\mathbf{Y}\|_{\mathrm{HS}}}{\|\tilde{\mathbf{X}}-\mathbf{Y}\|_{\mathrm{HS}}}$	$\frac{1}{\sqrt{\hat{r}_1}}\|\hat{U}_1^\top \tilde{U}_1\|_F$	$\frac{1}{\sqrt{\hat{r}_2}}\|\hat{U}_2^\top \tilde{U}_2\|_F$	$\frac{1}{\sqrt{\hat{r}_3}}\|\hat{U}_3^\top \tilde{U}_3\|_F$
round(.1p_t)	0.0035	1.5086	0.8291	0.8212	0.8318
round(.2p_t)	0.0267	1.2063	0.9352	0.9110	0.9155
round(.3p_t)	0.0832	1.0918	0.9650	0.9571	0.9634
round(.4p_t)	0.1766	1.0506	0.9769	0.9745	0.9828
round(.5p_t)	0.3066	1.0312	0.9832	0.9840	0.9905

Table 1 that on average, $\|\hat{\mathbf{X}} - \mathbf{Y}\|_{\mathrm{HS}}$, that is, rank-$(r_1, r_2, r_3)$ approximation error with limited numbers of Cross tensor measurements, can get very close to $\|\tilde{\mathbf{X}} - \mathbf{Y}\|_{\mathrm{HS}}$, that is, rank-$(r_1, r_2, r_3)$ approximation error with the whole tensor $\mathbf{Y}$. Besides, $\frac{1}{\sqrt{\hat{r}_t}}\|\hat{U}_t^\top \tilde{U}_t\|_F$ is close to 1, which means the singular vectors calculated from limited numbers of Cross tensor measurements are not too far from the ones calculated from the whole tensor.

Therefore, by the proposed Cross Tensor Measurement Scheme and a small fraction of observable entries, we can approximate the leading principle component of the original tensor just as if we have observed all voxels. This illustrates the power of the proposed algorithm.

6. Discussions: Extensions to fourth- and higher-order tensors. In this paper, we propose a novel tensor measurement scheme called Cross and the corresponding low-rank tensor completion algorithm. The theoretical analyses are provided for the proposed procedure to guarantee the optimality in both the sample size requirement and the completion error. The proposed procedure is efficient and easy to implement even for large-scale dataset.

Throughout the paper, we focus our presentations and analyses on order-3 tensors. Moreover, the proposed methods can be easily extended for fourth- or higher-order tensors. Suppose we aim to complete an unknown, order-d and rank-$(r_1, \ldots, r_d)$ tensor: $\mathbf{X} \in \mathbb{R}^{p_1 \times \cdots \times p_d}$. Similarly, we introduce the order-d Cross Tensor Measurement Scheme as

$$\Omega_t \subseteq [1 : p_t], \qquad \text{where } |\Omega_t| = m_t, t = 1, \ldots, d,$$

$$\Xi_t = \prod_{s \neq t} \Omega_s, \qquad \text{where } |\Xi_t| = g_t, t = 1, \ldots, d,$$

$$\boldsymbol{\Omega} = \left(\prod_{t=1}^{d} \Omega_t\right) \bigcup_{t=1}^{d} ([1 : p_t] \times \Xi_t).$$

By observing $\mathbf{Y}_{\boldsymbol{\Omega}} = \mathbf{X}_{\boldsymbol{\Omega}} + \mathbf{Z}_{\boldsymbol{\Omega}}$, we can construct the body, arm and joint matricizations as

$$Y_{t,\Omega} = \mathcal{M}_t(\mathbf{Y}_{[\Omega_1,\ldots,\Omega_d]}), \qquad Y_{\Xi_t} = \mathcal{M}_t(\mathbf{Y}_{[1:p_t]\times\Xi_t}),$$
$$Y_{\Omega_t\times\Xi_t} = \mathcal{M}_t(\mathbf{Y}_{\Omega_t\times\Xi_t}).$$

Similarly as Theorem 1, we can prove that $\mathbf{X}$ can be recovered by

$$\hat{\mathbf{X}} = \mathbf{Y}_{[\Omega_1,\ldots,\Omega_d]} \times_1 R_1 \times_2 \cdots \times_d R_d \qquad \text{where } R_t = Y_{\Xi_t} Y_{\Omega_t\times\Xi_t}^{\dagger}, t = 1,\ldots,d,$$

in the noiseless setting, provided that $\min\{m_t, g_t\} \geq r_t$ [so that $|\boldsymbol{\Omega}| \geq \prod_{t=1}^d r_t + \sum_{t=1}^d r_t(p_t - r_t)$] and some other mild assumptions holds. This result achieves the optimal sampling requirement since the degrees of freedom for rank-$(r_1,\ldots,r_d)$ tensors in $\mathbb{R}^{p_1\times\cdots\times p_d}$ is exactly $\prod_{t=1}^d r_t + \sum_{t=1}^d r_t(p_t - r_t)$ (see Proposition 1 and Remark 1). Additionally, the procedure for order-d tensor completion with noisy Cross measurements essentially follow from the proposed procedure in Algorithm 2, as long as we replace "$t = 1, 2, 3$" by "$t = 1,\ldots,d$". An interesting problem for further exploration is on how to select the tuning parameter λ_t for the general order-d tensor completion.

The main results on Cross tensor measurements can be further extended from the entrywise observations to the more general projection settings. Suppose $P_t \in \mathbb{O}_{p_t,m_t}$ and $Q_t \in \mathbb{O}_{m_{t+1}m_{t+2},g_t}$ are orthogonal matrices for $t = 1, 2, 3$. We observe the following *body, arm and joint projections* of $\mathbf{X}$:

$$\begin{aligned}
&\mathbf{X}^{(B)} = \mathbf{X} \times_1 P_1^\top \times_2 P_2^\top \times_3 P_3^\top \in \mathbb{R}^{m_1\times m_2\times m_3},\\
&X_t^{(A)} = \mathcal{M}_1\big(\mathbf{X} \times_2 P_2^\top \times_3 P_3^\top\big)\cdot Q_1 \in \mathbb{R}^{p_t\times g_t}, \qquad t = 1,2,3,\\
&X_t^{(J)} = P_1^\top \cdot \mathcal{M}_1\big(\mathbf{X} \times_2 P_2^\top \times_3 P_3^\top\big)\cdot Q_1 \in \mathbb{R}^{m_t\times g_t}, \qquad t = 1,2,3.
\end{aligned} \tag{6.1}$$

The regular Cross tensor measurements discussed in Section 2.2 can be seen as a special case of (6.1), where

$$(P_t)_{ij} = 1_{\{i=\Omega_t(j)\}}, \qquad 1 \leq i \leq p_t, 1 \leq j \leq m_t;$$
$$(Q_t)_{ij} = 1_{\{i=\Xi_t(j)\}}, \qquad 1 \leq i \leq m_{t+1}m_{t+2}, 1 \leq j \leq g_t.$$

When $\mathbf{X}^{(B)}$, $X_t^{(A)}$, $X_t^{(J)}$ are observed without noise, by the similar argument of Theorem 1, one can show that

$$\mathbf{X} = \mathbf{X}^{(B)} \times_1 R_1 \times R_2 \times_3 R_3, \qquad R_t = X_t^{(A)}\big(X_t^{(J)}\big)^{\dagger}, \qquad t = 1,2,3.$$

In the noisy setting, one can similarly apply the proposed Algorithm 2 to recover $\mathbf{X}$. Suppose the additive observation noises are

$$\begin{aligned}
&\mathbf{Z}^{(B)} = \mathbf{Y}^{(B)} - \mathbf{X}^{(B)} \in \mathbb{R}^{m_1\times m_2\times m_3},\\
&Z_t^{(A)} = Y_t^{(A)} - X_t^{(A)} \in \mathbb{R}^{p_t\times g_t}, \qquad t = 1,2,3,\\
&Z_t^{(J)} = Y_t^{(J)} - X_t^{(J)} \in \mathbb{R}^{m_t\times g_t}, \qquad t = 1,2,3.
\end{aligned}$$

If $\hat{\mathbf{X}}$ is the output from Algorithm 2, by similar procedure of Theorem 2, one can show that

$$\begin{aligned}\|\hat{\mathbf{X}}-\mathbf{X}\|_{\mathrm{HS}} &\le C\lambda_1\lambda_2\lambda_3\|\mathbf{Z}^{(B)}\|_{\mathrm{HS}}\\ &\quad + C\lambda_1\lambda_2\lambda_3\Bigg(\sum_{t=1}^{3}\xi_t\|Z_t^{(J)}\|_F + C\sum_{t=1}^{3}\frac{\xi_t}{\lambda_t}\|Z_t^{(A)}\|_F\Bigg),\end{aligned} \tag{6.2}$$

where $\lambda_t \ge 2\|X^{(A)}(X_t^{(J)})^{\dagger}\|$, $\xi_t = \|(X_t^{(J)})^{\dagger}\mathcal{M}_t(\mathbf{X}^{(B)})\|$. This extension yields a possible application of Cross to general tensor estimation problems. Suppose one aims to recover low-rank tensor $\mathbf{X}$ from limited number of (not necessarily entry-wise) measurements. If one can obtain reasonable estimations for the following low-dimensional projections of $\mathbf{X}$,

$$\mathbf{X}^{(B)} = \mathbf{X}\times_1 P_1\times P_2\times P_3 \in \mathbb{R}^{m_1\times m_2\times m_3} \quad \text{and}$$

$$X_t^{(A)} = \mathcal{M}_t\big(\mathbf{X}\times_{t+1} P_{t+1}^{\top}\times_{t+2} P_{t+2}^{\top}\big)Q_t \in \mathbb{R}^{p_t\times g_t},$$

the proposed Algorithm 2 yields an efficient estimation for $\mathbf{X}$ with guaranteed performance (6.2). It would be an interesting future topic to apply Cross tensor measurement scheme to develop efficient algorithm for other low-rank tensor estimation problems, such as tensor completion with uniform random measurements, tensor regression and tensor denoising.

Acknowledgments. The author would like to thank Lexin Li for sharing the ADHD dataset and helpful discussion. The author would also like to thank the Editor, Associate Editor and anonymous referees for valuable suggestions on improving the paper.

SUPPLEMENTARY MATERIAL

Supplement to "Cross: Efficient low-rank tensor completion" (DOI: 10.1214/18-AOS1694SUPP; .pdf). In the supplement, we provide proofs for the main results and technical lemmas. For better presentation for the long proof of Theorem 2, we also provide a table of used notation.

REFERENCES

AGARWAL, A., NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions. *Ann. Statist.* **40** 1171–1197. MR2985947

BARAK, B. and MOITRA, A. (2016). Noisy tensor completion via the sum-of-squares hierarchy. In 29*th Annual Conference on Learning Theory* 417–445.

BHOJANAPALLI, S. and SANGHAVI, S. (2015). A new sampling technique for tensors. Preprint. Available at arXiv:1502.05023.

CAI, T., CAI, T. T. and ZHANG, A. (2016). Structured matrix completion with applications to genomic data integration. *J. Amer. Statist. Assoc.* **111** 621–633. MR3538692

CAI, T. T. and ZHOU, W.-X. (2016). Matrix completion via max-norm constrained optimization. *Electron. J. Stat.* **10** 1493–1525. MR3507371

CAIAFA, C. F. and CICHOCKI, A. (2010). Generalizing the column-row matrix decomposition to multi-way arrays. *Linear Algebra Appl.* **433** 557–573. MR2653820

CAIAFA, C. F. and CICHOCKI, A. (2015). Stable, robust, and super fast reconstruction of tensors using multi-way projections. *IEEE Trans. Signal Process.* **63** 780–793. MR3299461

CANDÈS, E. J. and PLAN, Y. (2011). Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans. Inform. Theory* **57** 2342–2359. MR2809094

CANDÈS, E. J. and TAO, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inform. Theory* **56** 2053–2080. MR2723472

CAO, Y. and XIE, Y. (2016). Poisson matrix recovery and completion. *IEEE Trans. Signal Process.* **64** 1609–1620. MR3548877

CAO, Y., ZHANG, A. and LI, H. (2017). Multi-sample Estimation of Bacterial Composition Matrix in Metagenomics Data. Preprint. Available at arXiv:1706.02380.

GANDY, S., RECHT, B. and YAMADA, I. (2011). Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Probl.* **27** 025010. MR2765628

GUHANIYOGI, R., QAMAR, S. and DUNSON, D. B. (2017). Bayesian tensor regression. *J. Mach. Learn. Res.* **18** 79. MR3714242

HILLAR, C. J. and LIM, L.-H. (2013). Most tensor problems are NP-hard. *J. ACM* **60** 45. MR3144915

JAIN, P. and OH, S. (2014). Provable tensor factorization with missing data. In *Advances in Neural Information Processing Systems* **1** 1431–1439. MIT Press, Cambridge, MA.

JIANG, X., RASKUTTI, G. and WILLETT, R. (2015). Minimax optimal rates for Poisson inverse problems with physical constraints. *IEEE Trans. Inform. Theory* **61** 4458–4474. MR3372365

JOHNDROW, J. E., BHATTACHARYA, A. and DUNSON, D. B. (2017). Tensor decompositions and sparse log-linear models. *Ann. Statist.* **45** 1–38. MR3611485

KARATZOGLOU, A., AMATRIAIN, X., BALTRUNAS, L. and OLIVER, N. (2010). Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems* 79–86. ACM, New York.

KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from a few entries. *IEEE Trans. Inform. Theory* **56** 2980–2998. MR2683452

KLOPP, O. (2014). Noisy low-rank matrix completion with general sampling distribution. *Bernoulli* **20** 282–303. MR3160583

KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. MR2535056

KOLTCHINSKII, V., LOUNICI, K. and TSYBAKOV, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Statist.* **39** 2302–2329. MR2906869

KRESSNER, D., STEINLECHNER, M. and VANDEREYCKEN, B. (2014). Low-rank tensor completion by Riemannian optimization. *BIT* **54** 447–468. MR3223510

KRISHNAMURTHY, A. and SINGH, A. (2013). Low-rank matrix and tensor completion via adaptive sampling. In *Advances in Neural Information Processing Systems* 836–844.

LI, N. and LI, B. (2010). Tensor completion for on-board compression of hyperspectral images. In 2010 *IEEE International Conference on Image Processing* 517–520. IEEE, New York.

LI, L. and ZHANG, X. (2017). Parsimonious tensor response regression. *J. Amer. Statist. Assoc.* **112** 1131–1146. MR3735365

LI, X., ZHOU, H. and LI, L. (2013). Tucker tensor regression and neuroimaging analysis. Preprint. Available at arXiv:1304.5637.

LI, L., CHEN, Z., WANG, G., CHU, J. and GAO, H. (2014). A tensor PRISM algorithm for multi-energy CT reconstruction and comparative studies. *J. X-Ray Sci. Technol.* **22** 147–163.

LIU, J., MUSIALSKI, P., WONKA, P. and YE, J. (2013). Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 208–220.

MAHONEY, M. W., MAGGIONI, M. and DRINEAS, P. (2008). Tensor-CUR decompositions for tensor-based data. *SIAM J. Matrix Anal. Appl.* **30** 957–987. MR2447438

MU, C., HUANG, B., WRIGHT, J. and GOLDFARB, D. (2014). Square deal: Lower bounds and improved relaxations for tensor recovery. In *ICML* 73–81.

NEGAHBAN, S. and WAINWRIGHT, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Statist.* **39** 1069–1097. MR2816348

NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *J. Mach. Learn. Res.* **13** 1665–1697. MR2930649

NOWAK, R. D. and KOLACZYK, E. D. (2000). A statistical multiscale framework for Poisson inverse problems. *IEEE Trans. Inform. Theory* **46** 1811–1825. MR1790322

OSELEDETS, I. V., SAVOSTIANOV, D. V. and TYRTYSHNIKOV, E. E. (2008). Tucker dimensionality reduction of three-dimensional arrays in linear time. *SIAM J. Matrix Anal. Appl.* **30** 939–956. MR2447437

OSELEDETS, I. V. and TYRTYSHNIKOV, E. E. (2009). Breaking the curse of dimensionality, or how to use SVD in many dimensions. *SIAM J. Sci. Comput.* **31** 3744–3759. MR2556560

PIMENTEL-ALARCÓN, D. L., BOSTON, N. and NOWAK, R. D. (2016). A characterization of deterministic sampling patterns for low-rank matrix completion. *IEEE J. Sel. Top. Signal Process.* **10** 623–636.

RASKUTTI, G., YUAN, M. and CHEN, H. (2017). Convex regularization for high-dimensional multiresponse tensor regression. Preprint. Available at arXiv:1512.01215v2.

RAUHUT, H., SCHNEIDER, R. and STOJANAC, Ž. (2017). Low rank tensor recovery via iterative hard thresholding. *Linear Algebra Appl.* **523** 220–262. MR3624675

RECHT, B. (2011). A simpler approach to matrix completion. *J. Mach. Learn. Res.* **12** 3413–3430. MR2877360

RENDLE, S. and SCHMIDT-THIEME, L. (2010). Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining* 81–90. ACM, New York.

RICHARD, E. and MONTANARI, A. (2014). A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems* 2897–2905.

ROHDE, A. and TSYBAKOV, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** 887–930. MR2816342

RUDELSON, M. and VERSHYNIN, R. (2007). Sampling from large matrices: An approach through geometric functional analysis. *J. ACM* **54** 21. MR2351844

SEMERCI, O., HAO, N., KILMER, M. E. and MILLER, E. L. (2014). Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. *IEEE Trans. Image Process.* **23** 1678–1693. MR3191324

SHAH, P., RAO, N. and TANG, G. (2015). Optimal low-rank tensor recovery from separable measurements: Four contractions suffice. Preprint. Available at arXiv:1505.04085.

SREBRO, N. and SHRAIBMAN, A. (2005). Rank, trace-norm and max-norm. In *Learning Theory. Lecture Notes in Computer Science* **3559** 545–560. Springer, Berlin. MR2203286

SUN, W. W. and LI, L. (2016). Sparse low-rank tensor response regression. Preprint. Available at arXiv:1609.04523.

SUN, W. W., LU, J., LIU, H. and CHENG, G. (2017). Provable sparse tensor decomposition. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 899–916. MR3641413

TUCKER, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika* **31** 279–311. MR0205395

WAGNER, A. and ZUK, O. (2015). Low-rank matrix recovery from row-and-column affine measurements. In *Proceedings of the* 32*nd International Conference on Machine Learning* (*ICML*-15) 2012–2020.

WANG, Y. and SINGH, A. (2015). Provably correct algorithms for matrix column subset selection with selectively sampled data. Preprint. Available at arXiv:1505.04343.

WETZSTEIN, G., LANMAN, D., HIRSCH, M. and RASKAR, R. (2012). Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting.

YUAN, M. and ZHANG, C.-H. (2016). On tensor completion via nuclear norm minimization. *Found. Comput. Math.* **16** 1031–1068. MR3529132

YUAN, M. and ZHANG, C.-H. (2017). Incoherent tensor norms and their applications in higher order tensor completion. *IEEE Trans. Inform. Theory* **63** 6753–6766. MR3707566

ZHANG, A. (2019). Supplement to "Cross: Efficient low-rank tensor completion." DOI:10.1214/18-AOS1694SUPP.

ZHANG, A. and XIA, D. (2017). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inform. Theory* **64** 7311–7338.

ZHOU, H., LI, L. and ZHU, H. (2013). Tensor regression with applications in neuroimaging data analysis. *J. Amer. Statist. Assoc.* **108** 540–552. MR3174640

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN-MADISON
MADISON, WISCONSIN 53706
USA
E-MAIL: anruzhang@stat.wisc.edu
URL: http://www.stat.wisc.edu/~anruzhang/