Incoherent Tensor Norms and Their Applications in Higher Order Tensor Completion

Ming Yuan* and Cun-Hui Zhang†

University of Wisconsin-Madison and Rutgers University

(June 14, 2016)

Abstract

In this paper, we investigate the sample size requirement for a general class of nuclear norm minimization methods for higher order tensor completion. We introduce a class of tensor norms by allowing for different levels of coherence, which allows us to leverage the incoherence of a tensor. In particular, we show that a kth order tensor of rank r and dimension $d \times \cdots \times d$ can be recovered perfectly from as few as $O((r^{(k-1)/2}d^{3/2} + r^{k-1}d)(\log(d))^2)$ uniformly sampled entries through an appropriate incoherent nuclear norm minimization. Our results demonstrate some key differences between completing a matrix and a higher order tensor: They not only point to potential room for improvement over the usual nuclear norm minimization but also highlight the importance of explicitly accounting for incoherence, when dealing with higher order tensors.

*Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706. The research of Ming Yuan was supported in part by NSF FRG Grant DMS-1265202, and NIH Grant 1U54AI117924-01.

†Department of Statistics and Biostatistics, Rutgers University, Piscataway, New Jersey 08854. The research of Cun-Hui Zhang was supported in part by NSF Grants DMS-1129626 and DMS-1209014

1 Introduction

Data in the format of tensors, or multilinear arrays, arise naturally in many modern applications. A kth order hypercubic tensor of dimension $d \times \cdots \times d$ has d^k entries so that these datasets typically are of fairly large size even for moderate d and small k. Therefore, it is oftentimes impractical to observe or store the entire tensor, which naturally brings about the question of tensor completion: How to reconstruct a kth order tensor $\boldsymbol{T} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ from observations $\{\boldsymbol{T}(\omega) : \omega \in \Omega\}$ where Ω is a uniformly sampled subset from $[d_1] \times \cdots \times [d_k]$? Here $[d] = \{1, \ldots, d\}$. The goal of this paper is to study in its full generality a class of tensor completion methods via nuclear norm minimization focusing on higher order tensors ($k \geq 3$).

1.1 Tensor completion

Obviously, for reconstructing $\boldsymbol{T}$ from a subset of its entries to be possible at all, $\boldsymbol{T}$ needs to have some sort of low dimensional structure which is often characterized by certain notion of low-rankness. In particular, let $\mathcal{L}_j(\boldsymbol{X})$ be the linear subspace of $\mathbb{R}^{d_j}$ spanned by the mode-j fibers:

$$\left\{\boldsymbol{X}(a_1, \ldots, a_{j-1}, \cdot, a_{j+1}, \ldots, a_k) \in \mathbb{R}^{d_j} : a_1 \in [d_1], \ldots, a_k \in [d_k]\right\}.$$

Denote by $r_j(\boldsymbol{X})$ the dimension of $\mathcal{L}_j(\boldsymbol{X})$. The tuplet $\{r_1(\boldsymbol{X}), \ldots, r_k(\boldsymbol{X})\}$ is the so-called Tucker ranks of $\boldsymbol{X}$. It is not hard to see that there are a total of $O(r^{k-1}d)$ free parameters in specifying a kth order hypercubic tensor of dimension $d \times \cdots \times d$ whose Tucker ranks are upper bounded by r, which suggests the possibility of recovering a large tensor of low rank from a fairly small fraction of the entries.

In addition to low-rankness, it is also essential to tensor completion that every entry of $\boldsymbol{T}$ contains similar amount of information about the entire tensor so that missing any of them would not stop us from being able to reconstruct it – a property that can be formally characterized through the *coherence* of the linear subspace $\mathcal{L}_j(\boldsymbol{T})$. See, e.g, Candès and Recht (2008). More specifically, the coherence of an r dimensional linear subspace U of $\mathbb{R}^d$

is defined as

$$\mu(U) = \frac{d}{r} \max_{1\le i\le d} \|\boldsymbol{P}_U \boldsymbol{e}_i\|_{\ell_2}^2 = \frac{\max_{1\le i\le d} \|\boldsymbol{P}_U \boldsymbol{e}_i\|_{\ell_2}^2}{d^{-1}\sum_{i=1}^d \|\boldsymbol{P}_U \boldsymbol{e}_i\|_{\ell_2}^2},$$

where $\boldsymbol{P}_U$ is the orthogonal projection onto U and $\boldsymbol{e}_i$'s are the canonical basis for $\mathbb{R}^d$. We call a tensor $\boldsymbol{X}$ μ_*-incoherent if

$$\mu_j(\boldsymbol{X}) := \mu(\mathcal{L}_j(\boldsymbol{X})) \le \mu_*.$$

An especially popular class of techniques to tensor completion is based on nuclear norm minimization where we seek among all tensors that agree with $\boldsymbol{T}$ on all observed entries the one with the smallest nuclear norm.

1.2 Nuclear norm minimization

Recall that the spectral and nuclear norms of a tensor $\boldsymbol{X} \in \mathbb{R}^{d_1\times\cdots\times d_k}$ are defined as

$$\|\boldsymbol{X}\| = \sup_{\boldsymbol{u}_j \in \mathbb{R}^{d_k}:\|\boldsymbol{u}_j\|_{\ell_2}\le 1} \langle \boldsymbol{X}, \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k\rangle$$

and

$$\|\boldsymbol{X}\|_* = \sup_{\boldsymbol{Y}\in\mathbb{R}^{d_1\times\cdots\times d_k}:\|\boldsymbol{Y}\|\le 1} \langle \boldsymbol{X}, \boldsymbol{Y}\rangle,$$

respectively, where $\langle\cdot,\cdot\rangle$ is the usual vectorized inner product, and $\|\cdot\|_{\ell_p}$ stands for the usual ℓ_p norm in a vector space. The usual nuclear norm minimization proceeds by solving the following convex optimization problem:

$$\min_{\boldsymbol{X}\in\mathbb{R}^{d_1\times\cdots\times d_k}} \|\boldsymbol{X}\|_* \qquad \text{subject to } \mathcal{P}_\Omega \boldsymbol{X} = \mathcal{P}_\Omega \boldsymbol{T}, \tag{1}$$

where $\mathcal{P}_\Omega : \mathbb{R}^{d_1\times\cdots\times d_k} \to \mathbb{R}^{d_1\times\cdots\times d_k}$ is a linear operator such that

$$\mathcal{P}_\Omega \boldsymbol{X}(\omega) = \begin{cases} \boldsymbol{X}(\omega) & \text{if } \omega \in \Omega \\ 0 & \text{otherwise} \end{cases}.$$

The solution to (1) is our reconstruction of $\boldsymbol{T}$. This approach was first introduced for matrices, that is $k = 2$, by Candès and Recht (2008) and Candès and Tao (2009). Similar approaches have also been adopted later for higher order tensors. See, e.g., Liu et al. (2009), Signoretto, Lathauwer and Suykens (2010), Gandy et al. (2011), Tomioka, Hayashi and

Kashima (2010), Tomioka et al. (2011), Mu et al. (2013), Jain and Oh (2014), and Yuan and Zhang (2014), among many others.

Of particular interest here is the requirement on the cardinality $|\Omega|$, which we shall refer to as the sample size, to ensure that $\boldsymbol{T}$ can be reconstructed perfectly (with high probability) via nuclear norm minimization (1). It is now well understood that in the case of matrices ($k = 2$), a $d \times d$ incoherent matrix of rank r can be recovered with high probability if $|\Omega| \gtrsim rd \cdot \text{polylog}(d)$ under suitable conditions, where $a \gtrsim b$ means that $a > Cb$ for some constant $C > 0$ independent of r and d, and polylog(d) stands for a certain polynomial of $\log(d)$. See, e.g., Recht (2010), and Gross (2011) among many others. It is clear that this sample size requirement is nearly optimal since the number of free parameters needed to specify a $d \times d$ rank r matrix is of the order $O(rd)$.

The situation for higher order tensors is more complicated as there are multiple ways to generalize the matrix style nuclear norm. A common practice is to first reshape a high order tensor to a matrix and then apply the techniques such as (1) to the unfolded matrix. In doing so, one recasts the problem of completing a kth order tensor, say of dimension $d \times \cdots \times d$, as a problem of completing a $d^{\lfloor k/2 \rfloor} \times d^{\lceil k/2 \rceil}$ matrix. Following the results for matrices, it can be shown that the sample size requirement for recovering a kth order hypercubic tensor of dimension $d \times \cdots \times d$ and whose Tucker ranks are bounded by r in this fashion is

$$|\Omega| \gtrsim r^{\lfloor k/2 \rfloor} d^{\lceil k/2 \rceil} \text{polylog}(d).$$

However, as Yuan and Zhang (2014) recently pointed out, this strategy is often suboptimal and direct minimization of the tensor nuclear norm yields a tighter sample size requirement at least when $k = 3$. In particular they show that, under suitable conditions, a $d \times d \times d$ tensor whose Tucker ranks are bounded by r can be recovered perfectly with high probability if

$$|\Omega| \gtrsim (r^{1/2} d^{3/2} + r^2 d) \text{polylog}(d).$$

Following their argument, it is also possible to show that, when $k > 3$, the sample size required for exact recovery via tensor nuclear norm minimization is

$$|\Omega| \gtrsim d^{k/2} \text{poly}(r, \log(d)),$$

where $\text{poly}(\cdot,\cdot)$ is a certain polynomial in both arguments. However, it remains unknown to what extent such a sample size requirement is tight for nuclear norm minimization based approaches. The main goal of this paper is to address this question. Indeed, we show that this sample size condition for higher order tensor can be much improved.

1.3 Incoherent nuclear norm minimization

The key ingredient of our approach is to define a new class of tensor nuclear norms that explicitly account for the incoherence of the linear subspaces spanned by the fibers of a tensor in defining its nuclear norm. More specifically, for a $\boldsymbol{\delta} = (\delta_1, \dots, \delta_k) \in (0,1]^k$, let

$$\mathscr{U}_{j_1 j_2}(\boldsymbol{\delta}) = \{\boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k : \|\boldsymbol{u}_j\|_{\ell_2} \le 1,\ \forall j; \|\boldsymbol{u}_j\|_{\ell_\infty} \le \delta_j, \forall j \ne j_1, j_2\}$$

be the set of all rank-one tensors satisfying incoherent conditions in "directions" other than j_1 and j_2. Then

$$\mathscr{U}(\boldsymbol{\delta}) = \bigcup_{1\le j_1 < j_2 \le k} \mathscr{U}_{j_1 j_2}(\boldsymbol{\delta})$$

is the collection of all rank-one tensors satisfying certain incoherence conditions in all but two directions. For a kth order tensor $\boldsymbol{X} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$, define a norm

$$\|\boldsymbol{X}\|_{\circ,\boldsymbol{\delta}} = \sup_{\boldsymbol{Y} \in \mathscr{U}(\boldsymbol{\delta})} \langle \boldsymbol{Y}, \boldsymbol{X} \rangle.$$

Note that when $\boldsymbol{\delta} = \mathbf{1} := (1, \dots, 1)^\top$, the ℓ_∞ constraint in defining $\|\boldsymbol{X}\|_{\circ,\boldsymbol{\delta}}$ becomes inactive so that $\|\boldsymbol{X}\|_{\circ,\mathbf{1}} = \|\boldsymbol{X}\|$, the usual tensor spectral norm. We can view $\|\cdot\|_{\circ,\boldsymbol{\delta}}$ as a *incoherent* spectral norm. We can also define the incoherence nuclear norm as the dual of the incoherence spectral norm:

$$\|\boldsymbol{X}\|_{\star,\boldsymbol{\delta}} = \sup_{\|\boldsymbol{Y}\|_{\circ,\boldsymbol{\delta}} \le 1} \langle \boldsymbol{Y}, \boldsymbol{X} \rangle,$$

so that $\|\boldsymbol{X}\|_{\star,\mathbf{1}}$ reduces to the usual tensor nuclear norm.

Instead of minimizing the usual tensor nuclear norm, we now consider recovering $\boldsymbol{T}$ via the following nuclear norm minimization problem:

$$\min_{\boldsymbol{X} \in \mathbb{R}^{d_1 \times \cdots \times d_k}} \|\boldsymbol{X}\|_{\star,\boldsymbol{\delta}} \qquad \text{subject to } \mathcal{P}_\Omega \boldsymbol{X} = \mathcal{P}_\Omega \boldsymbol{T}. \tag{2}$$

It is clear that (2) reduces to the usual nuclear norm minimization (1) if $\boldsymbol{\delta} = \mathbf{1}$. But as we shall see later, it could be extremely beneficial to take smaller values for δ_js. Our goal is to investigate the appropriate choices of $\boldsymbol{\delta}$, and when $\boldsymbol{T}$ can be recovered through the incoherent nuclear norm minimization (2).

1.4 Outline

Our main result provides a sample size requirement for recovering an incoherent and low rank tensor $\boldsymbol{T} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ via (2). In particular, our result implies that a kth order hypercubic tensor of dimension $d \times \cdots \times d$ whose Tucker ranks are bounded by r can be reconstructed perfectly by the solution of (2) with appropriate choices of $\boldsymbol{\delta}$, as long as

$$|\Omega| \gtrsim (r^{(k-1)/2} d^{3/2} + r^{k-1} d)(\log(d))^2.$$

This represents a drastic improvement over the requirement for the usual nuclear norm minimization. It is especially worth noting that, perhaps somewhat surprisingly, the sample size given above depends on the order k only through the rank r which, in most situations of interest, is small. It is also instructive to look at the case when a tensor is of finite rank, that is $r = O(1)$. The sample size requirement in such cases becomes $O(d^{3/2}(\log(d))^2)$ for any fixed order k, which suggests the possibility of a tremendous amount of data reduction even for moderate ks.

In establishing the sample size requirement for the proposed incoherent nuclear norm minimization approach, we developed various algebraic properties of incoherent tensor norms including a characterization of the subdifferential of the incoherent tensor nuclear norm which generalizes earlier results for matrices (Watson, 1992) and for the usual nuclear norm with third order tensors (Yuan and Zhang, 2014).

Also essential to our analysis are large deviation bounds under the incoherent spectral norm we derived for randomly sampled tensors, which may be of independent interest. These probabilistic bounds show a tighter concentration behavior of random tensors under incoherent norm than under the usual spectral norm, an observation we exploited to establish tighter sample size requirement for tensor completion. We note that concentration inequalities such as the ones presented here are the basis for many problems beyond tensor completion. For examples, it is plausible that these bounds could prove useful in developing improved sampling schemes for higher order tensor sparsification. See, e.g., Nguyen, Drineas and Tran (2015). These applications are beyond the scope of the current paper and we shall leave them for future studies.

The rest of the paper is organized as follows. In the next section, we introduce the notion of incoherent tensor norms and establish some algebraic properties of these norms useful for our analysis. In Section 3, we derive large deviation bounds for randomly sampled tensors. Building on the tool developed in Sections 2 and 3, we provide the sample size requirement for the incoherent nuclear norm minimization in Section 4. We conclude with some discussions and remarks in Section 5

2 Subdifferential of Incoherent Tensor Nuclear Norm

Note that the optimization problem (2) is convex. In order to show that $\boldsymbol{T}$ can be recovered via (2), it suffices to find a member from the subdifferential of $\|\cdot\|_{\star,\boldsymbol{\delta}}$ at $\boldsymbol{T}$ that can certify it as the unique solution to (2). To this end, we need to characterize the subdifferential of $\|\cdot\|_{\star,\boldsymbol{\delta}}$, which we shall do in this section.

We first note several immediate yet useful observations of the incoherent spectral and nuclear norms. We shall make repeated use of these simple properties without mentioning in the rest of paper.

Proposition 1. *For any tensor* $\boldsymbol{X}\in\mathbb{R}^{d_1\times\cdots\times d_k}$ *and* $\boldsymbol{\delta}\in(0,1]^k$*,*

$$\|\boldsymbol{X}\|_{\rm HS}^2 := \langle \boldsymbol{X},\boldsymbol{X}\rangle \le \|\boldsymbol{X}\|_{\circ,\boldsymbol{\delta}}\|\boldsymbol{X}\|_{\star,\boldsymbol{\delta}},$$

and

$$\|\boldsymbol{X}\|_{\circ,\boldsymbol{\delta}} \le \|\boldsymbol{X}\| \le \|\boldsymbol{X}\|_{\rm HS} \le \|\boldsymbol{X}\|_{*} \le \|\boldsymbol{X}\|_{\star,\boldsymbol{\delta}}.$$

Recall that, for a tensor $\boldsymbol{X}$, $\mathcal{L}_j(\boldsymbol{X})$ is the linear subspace of $\mathbb{R}^{d_j}$ spanned by the mode-j fibers of $\boldsymbol{X}$. Denote by $\boldsymbol{P}_j(\boldsymbol{X})$ the orthogonal projection to $\mathcal{L}_j(\boldsymbol{X})$. For brevity, we omit the dependence of $\boldsymbol{P}_j$ and $\mathcal{L}_j$ on $\boldsymbol{X}$ hereafter when no confusion occurs. Write

$$\mathcal{Q}^0_{\boldsymbol{X}} = \boldsymbol{P}_1 \otimes \cdots \otimes \boldsymbol{P}_k.$$

It is clear that for any $\boldsymbol{u}_j \in \mathbb{R}^{d_j}$, we have

$$\langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k, \boldsymbol{X} \rangle = \langle \boldsymbol{P}_1 \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{P}_k \boldsymbol{u}_k, \boldsymbol{X} \rangle,$$

This immediately implies that

Proposition 2. *Let* $\delta_j \geq \max_{\|\boldsymbol{u}\|_{\ell_2} \leq 1} \|\boldsymbol{P}_j(\boldsymbol{X})\boldsymbol{u}\|_{\ell_\infty}$*, for* $j = 1, \ldots, k$*. Then, for any tensor* $\boldsymbol{W} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$*,*

$$\|\mathcal{Q}^0_{\boldsymbol{X}} \boldsymbol{W}\|_{\circ,\boldsymbol{\delta}} = \|\mathcal{Q}^0_{\boldsymbol{X}} \boldsymbol{W}\| \leq \|\boldsymbol{W}\|_{\circ,\boldsymbol{\delta}}.$$

Consequently, $\|\boldsymbol{X}\|_{\star,\boldsymbol{\delta}} = \|\boldsymbol{X}\|_*$.

Propositions 1 indicates that the incoherent nuclear norm is greater than the usual nuclear norm in general. But Proposition 2 shows that the two norms are equal if a tensor is indeed incoherent. This gives some intuition on the potential benefits of minimizing the incoherent instead of the usual nuclear norm. Because more penalty is levied on tensors that are not incoherent, compared with the usual nuclear norm minimization (1), it is more plausible that the solution of (2) is incoherent. Given that the truth is known apriori to be incoherent, it is more likely that incoherent tensor nuclear norm minimization produces exact recovery. This advantage will be more precisely quantified by the much refined sample size requirement we shall establish later.

We are now in position to describe a characterization of the subdifferential of $\|\cdot\|_{\star,\boldsymbol{\delta}}$. Let $\boldsymbol{P}_j^\perp = \boldsymbol{I} - \boldsymbol{P}_j$ be the projection to the orthogonal complement $\mathcal{L}_j^\perp$ of $\mathcal{L}_j$ in $\mathbb{R}^{d_j}$. Write

$$\mathcal{Q}_{\boldsymbol{X}} = \mathcal{Q}^0_{\boldsymbol{X}} + \sum_{j=1}^{k} \boldsymbol{P}_1 \otimes \cdots \otimes \boldsymbol{P}_{j-1} \otimes \boldsymbol{P}_j^\perp \otimes \boldsymbol{P}_{j+1} \otimes \cdots \otimes \boldsymbol{P}_k.$$

It is easy to see that

$$\mathcal{Q}^\perp_{\boldsymbol{X}} := \mathcal{I} - \mathcal{Q}_{\boldsymbol{X}} = \sum_{1 \leq j_1 < j_2 \leq k} \mathcal{Q}^\perp_{\boldsymbol{X}, j_1, j_2},$$

where $\mathcal{I}$ is the identity operator on the appropriate space, and

$$\mathcal{Q}^{\perp}_{\boldsymbol{X},j_1,j_2} = \boldsymbol{P}_1 \otimes \cdots \otimes \boldsymbol{P}_{j_1-1} \otimes \boldsymbol{P}^{\perp}_{j_1} \otimes \boldsymbol{P}_{j_1+1} \otimes \cdots \otimes \boldsymbol{P}_{j_2-1} \otimes \boldsymbol{P}^{\perp}_{j_2} \otimes \mathcal{I} \otimes \cdots \otimes \mathcal{I}.$$

We note that $\mathcal{Q}^{\perp}_{j_1,j_2}$ is the orthogonal projection to the linear space of all $\boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k$ such that $\boldsymbol{u}_j$ is in either $\mathcal{L}_j$ or $\mathcal{L}^{\perp}_j$ for $1 \le j \le j_2$ and that j_1 and j_2 are the only indices with $\boldsymbol{u}_j \in \mathcal{L}^{\perp}_j$.

Theorem 1. *Let $\delta_j \ge \max_{\|\boldsymbol{u}\|_{\ell_2} \le 1} \|\boldsymbol{P}_j(\boldsymbol{X})\boldsymbol{u}\|_{\ell_\infty}$, for $j = 1, \ldots, k$. Then there exists an $\boldsymbol{W}_0 \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ such that*

$$\mathcal{Q}^0_{\boldsymbol{X}} \boldsymbol{W}_0 = \boldsymbol{W}_0, \qquad \|\boldsymbol{W}_0\|_{\circ,\boldsymbol{\delta}} = 1, \qquad \text{and} \qquad \|\boldsymbol{X}\|_{\star,\boldsymbol{\delta}} = \langle \boldsymbol{W}_0, \boldsymbol{X} \rangle.$$

Moreover, for any $\boldsymbol{Y} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$

$$\|\boldsymbol{Y}\|_{\star,\boldsymbol{\delta}} \ge \|\boldsymbol{X}\|_{\star,\boldsymbol{\delta}} + \frac{2}{k(k-1)} \|\mathcal{Q}^{\perp}_{\boldsymbol{X}} \boldsymbol{Y}\|_{\star,\boldsymbol{\delta}} + \langle \boldsymbol{W}_0, \boldsymbol{Y} - \boldsymbol{X} \rangle.$$

Proof of Theorem 1. Let $\widetilde{\boldsymbol{W}}_0$ be the dual of $\boldsymbol{X}$ satisfying $\|\widetilde{\boldsymbol{W}}_0\|_{\circ,\boldsymbol{\delta}} = 1$ and $\langle \widetilde{\boldsymbol{W}}_0, \boldsymbol{X} \rangle = \|\boldsymbol{X}\|_{\star,\boldsymbol{\delta}}$. Set $\boldsymbol{W}_0 = \mathcal{Q}^0_{\boldsymbol{X}} \widetilde{\boldsymbol{W}}_0$. Since $\boldsymbol{X} = \mathcal{Q}^0_{\boldsymbol{X}} \boldsymbol{X}$ and $\mathcal{Q}^0_{\boldsymbol{X}}$ is an orthogonal projection, we have $\mathcal{Q}^0_{\boldsymbol{X}} \boldsymbol{W}_0 = \boldsymbol{W}_0$, $\mathcal{Q}^{\perp}_{\boldsymbol{X}} \boldsymbol{X} = 0$ and $\|\boldsymbol{X}\|_{\star,\boldsymbol{\delta}} = \langle \boldsymbol{W}_0, \boldsymbol{X} \rangle \le \|\boldsymbol{W}_0\|_{\circ,\boldsymbol{\delta}} \|\boldsymbol{X}\|_{\star,\boldsymbol{\delta}}$. This, along with Proposition 2, proves the first statement.

To prove the second statement, we first show that for any $\boldsymbol{W}_1 \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ such that $\|\boldsymbol{W}_1\|_{\circ,\boldsymbol{\delta}} \le 2/\{k(k-1)\}$, we have

$$\|\boldsymbol{W}_0 + \mathcal{Q}^{\perp}_{\boldsymbol{X}} \boldsymbol{W}_1\|_{\circ,\boldsymbol{\delta}} \le 1. \tag{3}$$

To this end, note first that

$$\begin{aligned} \|\boldsymbol{W}_0 + \mathcal{Q}^{\perp}_{\boldsymbol{X}} \boldsymbol{W}_1\|_{\circ,\boldsymbol{\delta}} &= \sup_{\boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k \in \mathscr{U}(\boldsymbol{\delta})} \langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k, \boldsymbol{W}_0 + \mathcal{Q}^{\perp}_{\boldsymbol{X}} \boldsymbol{W}_1 \rangle \\ &= \max_{1 \le j_1 < j_2 \le k} \left\{ \sup_{\boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k \in \mathscr{U}_{j_1 j_2}(\boldsymbol{\delta})} \langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k, \boldsymbol{W}_0 + \mathcal{Q}^{\perp}_{\boldsymbol{X}} \boldsymbol{W}_1 \rangle \right\}. \end{aligned}$$

It then suffices to show that for any $1 \le j_1 < j_2 \le k$, and $\boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k \in \mathscr{U}_{j_1,j_2}(\boldsymbol{\delta})$,

$$\langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k, \boldsymbol{W}_0 + \mathcal{Q}^{\perp}_{\boldsymbol{X}} \boldsymbol{W}_1 \rangle \le 1.$$

As the statement is not specific to the index label, we assume without loss of generality that $j_1 = 1$ and $j_2 = 2$; Otherwise, a different decomposition of $\mathcal{Q}_{\boldsymbol{X}}^{\perp}$ is needed beginning with the projection $\mathcal{I} \otimes \cdots \otimes \mathcal{I} \otimes \boldsymbol{P}_{j_1}^{\perp} \otimes \mathcal{I} \otimes \cdots \otimes \mathcal{I} \otimes \boldsymbol{P}_{j_2}^{\perp} \otimes \mathcal{I} \otimes \cdots \otimes \mathcal{I}$. Recall that

$$\begin{aligned}\langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k, \mathcal{Q}_{\boldsymbol{T}}^{\perp} \boldsymbol{W}_1 \rangle &\leq \sum_{1 \leq j_3 < j_4 \leq k} \langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k, \mathcal{Q}_{j_3,j_4}^{\perp} \boldsymbol{W}_1 \rangle \\ &\leq \frac{1}{2} k(k-1) \max_{1 \leq j_3 < j_4 \leq k} \langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k, \mathcal{Q}_{j_3,j_4}^{\perp} \boldsymbol{W}_1 \rangle.\end{aligned}$$

By definition,

$$\begin{aligned}&\langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k, \mathcal{Q}_{j_3,j_4}^{\perp} \boldsymbol{W}_1 \rangle \\ =\ &\langle \boldsymbol{P}_1 \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{P}_{j_3-1} \boldsymbol{u}_{j_3-1} \otimes \boldsymbol{P}_{j_3}^{\perp} \boldsymbol{u}_{j_3} \otimes \cdots \otimes \boldsymbol{P}_{j_4-1} \boldsymbol{u}_{j_4-1} \otimes \boldsymbol{P}_{j_4}^{\perp} \boldsymbol{u}_{j_4} \otimes \cdots \otimes \boldsymbol{u}_k, \boldsymbol{W}_1 \rangle.\end{aligned}$$

Because $\|\boldsymbol{u}\|_{\ell_\infty} \leq \delta_j$ for all $j \geq 2$ and $\|\boldsymbol{P}_j \boldsymbol{u}\|_{\ell_\infty} \leq \delta_j \leq \delta_j$ for all $\boldsymbol{u} \in \mathbb{R}^{d_k}$ with $\|\boldsymbol{u}\|_{\ell_2} \leq 1$, we have

$$\begin{aligned}&\langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k, \mathcal{Q}_{j_3,j_4}^{\perp} \boldsymbol{W}_1 \rangle \\ \leq\ &\|\boldsymbol{P}_{j_3}^{\perp} \boldsymbol{u}_{j_3}\|_{\ell_2} \|\boldsymbol{P}_{j_4}^{\perp} \boldsymbol{u}_{j_4}\|_{\ell_2} \sup_{\boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k \in \mathscr{U}_{j_3 j_4}(\boldsymbol{\delta})} \langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k, \boldsymbol{W}_1 \rangle \\ \leq\ &\|\boldsymbol{P}_{j_3}^{\perp} \boldsymbol{u}_{j_3}\|_{\ell_2} \|\boldsymbol{P}_{j_4}^{\perp} \boldsymbol{u}_{j_4}\|_{\ell_2} \|\boldsymbol{W}_1\|_{\circ,\boldsymbol{\delta}} \\ \leq\ &\frac{2}{k(k-1)} \|\boldsymbol{P}_{j_3}^{\perp} \boldsymbol{u}_{j_3}\|_{\ell_2} \|\boldsymbol{P}_{j_4}^{\perp} \boldsymbol{u}_{j_4}\|_{\ell_2}.\end{aligned}$$

Together with the fact that

$$\begin{aligned}\langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k, \mathcal{Q}_{\boldsymbol{X}}^{0} \boldsymbol{W}_0 \rangle &= \langle \boldsymbol{P}_1 \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{P}_k \boldsymbol{u}_k, \boldsymbol{W}_0 \rangle \\ &\leq \|\boldsymbol{W}_0\|_{\circ,\mathbf{1}} \prod_{j=1}^{k} \|\boldsymbol{P}_j \boldsymbol{u}_j\|_{\ell_2},\end{aligned}$$

we get, for any $\boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k \in \mathscr{U}_{j_1 j_2}(\boldsymbol{\delta})$,

$$\begin{aligned}&\langle \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k, \mathcal{Q}_{\boldsymbol{X}}^{0} \boldsymbol{W}_0 + \mathcal{Q}_{\boldsymbol{X}}^{\perp} \boldsymbol{W}_1 \rangle \\ \leq\ &\prod_{j=1}^{k} \|\boldsymbol{P}_j \boldsymbol{u}_j\|_{\ell_2} + \max_{1 \leq j_3 < j_4 \leq k} \|\boldsymbol{P}_{j_3}^{\perp} \boldsymbol{u}_{j_3}\|_{\ell_2} \|\boldsymbol{P}_{j_4}^{\perp} \boldsymbol{u}_{j_4}\|_{\ell_2} \\ \leq\ &\max_{1 \leq j_3 < j_4 \leq k} \left\{ \|\boldsymbol{P}_{j_3} \boldsymbol{u}_{j_3}\|_{\ell_2} \|\boldsymbol{P}_{j_4} \boldsymbol{u}_{j_4}\|_{\ell_2} + \|\boldsymbol{P}_{j_3}^{\perp} \boldsymbol{u}_{j_3}\|_{\ell_2} \|\boldsymbol{P}_{j_4}^{\perp} \boldsymbol{u}_{j_4}\|_{\ell_2} \right\} \\ \leq\ &\max_{1 \leq j_3 < j_4 \leq k} \left\{ \left(\|\boldsymbol{P}_{j_3} \boldsymbol{u}_{j_3}\|_{\ell_2}^2 + \|\boldsymbol{P}_{j_3}^{\perp} \boldsymbol{u}_{j_3}\|_{\ell_2}^2 \right)^{1/2} \left(\|\boldsymbol{P}_{j_4} \boldsymbol{u}_{j_4}\|_{\ell_2}^2 + \|\boldsymbol{P}_{j_4}^{\perp} \boldsymbol{u}_{j_4}\|_{\ell_2}^2 \right)^{1/2} \right\} \\ =\ &1.\end{aligned}$$

It then follows that

$$\begin{aligned}\|\boldsymbol{Y}\|_{\star,\boldsymbol{\delta}} - \|\boldsymbol{X}\|_{\star,\boldsymbol{\delta}} &\geq \max_{\|\boldsymbol{W}_1\|_{\circ,\boldsymbol{\delta}}\leq 2/\{k(k-1)\}} \langle \boldsymbol{W}_0 + \mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{W}_1, \boldsymbol{Y}-\boldsymbol{X}\rangle \\ &= \frac{\|\mathcal{Q}_{\boldsymbol{X}}^{\perp}\boldsymbol{Y}\|_{\star,\boldsymbol{\delta}}}{k(k-1)/2} + \langle \boldsymbol{W}_0, \boldsymbol{Y}-\boldsymbol{X}\rangle.\end{aligned}$$

This completes the proof. □

Theorem 1 provides a sufficient condition for a tensor to be in the subdifferential $\partial\|\boldsymbol{X}\|_{\star,\boldsymbol{\delta}}$. More specifically, it states that there exists a $\boldsymbol{W}_0$ so that for any $\boldsymbol{W}_1$ such that $\boldsymbol{W}_1 = \mathcal{Q}_{\boldsymbol{X}}^{\perp}\boldsymbol{W}_1$ and $\|\boldsymbol{W}_1\|_{\circ,\boldsymbol{\delta}} \leq 2/\{k(k-1)\}$,

$$\boldsymbol{W}_0 + \boldsymbol{W}_1 \in \partial\|\boldsymbol{X}\|_{\star,\boldsymbol{\delta}}.$$

This characterization generalizes the earlier result by Yuan and Zhang (2014) for the special case when $k=3$ and $\boldsymbol{\delta}=\mathbf{1}$.

3 Concentration under Incoherent Spectral Norm

A main technical tool for many tensor related problems is the large deviation bounds for the spectral norm of a random tensor. We shall use such bounds, in particular, to construct a dual certificate for (2) later on.

Let $\boldsymbol{A} \in \mathbb{R}^{d_1\times\cdots\times d_k}$ be an arbitrary but fixed tensor. We are interested in the behavior of randomly sampled tensors

$$\boldsymbol{X}_i = (d_1\cdots d_k)\mathcal{P}_{\omega_i}\boldsymbol{A}, \qquad i=1,\ldots,n,$$

where ω_is are iid uniform random variables on $[d_1]\times\cdots\times[d_k]$. Write

$$\bar{\boldsymbol{X}} = \frac{1}{n}\left(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n\right).$$

It is clear that $\mathbb{E}\bar{\boldsymbol{X}} = \boldsymbol{A}$. We are interested in bounding the incoherent spectral norm of its deviation from the mean $\|\bar{\boldsymbol{X}} - \boldsymbol{A}\|_{\circ,\boldsymbol{\delta}}$.

Denote by

$$\|\boldsymbol{A}\|_{\max} = \max_{\omega\in[d_1]\times\cdots\times[d_k]} |\boldsymbol{A}(\omega)|.$$

For brevity, write

$$d = \frac{1}{k}\sum_{1\le j\le k} d_j, \qquad \text{and} \qquad d_* = (d_1\cdots d_k)^{1/k},$$

and

$$\delta_* = (\delta_1\cdots\delta_k)^{1/k}, \qquad \text{and} \qquad \delta_{**} = \min_{1\le j_1<j_2\le k}\sqrt{\delta_{j_1}\delta_{j_2}}.$$

We first give a general concentration bound.

Theorem 2. *Suppose that d is sufficiently large such that*

$$\frac{8e}{9\log 2}k^2(\log d)^3 \le d.$$

For any $\alpha > 0$ and

$$t \ge 160(3\alpha+7)\frac{k}{n}\sqrt{d\log d_*}(2\delta_* d_*)^k\|\boldsymbol{A}\|_{\max}\max_{1\le j_1<j_2\le k}\left\{\left(\frac{n}{\delta_{j_1}^2 d_{j_1}\delta_{j_2}^2 d_{j_2}} + \frac{\log d}{\delta_{j_1}^2\delta_{j_2}^2}\right)\right\}^{1/2},$$

then

$$\mathbb{P}\left\{\left\|\bar{\boldsymbol{X}} - \boldsymbol{A}\right\|_{\circ,\boldsymbol{\delta}} \ge t\right\} \le \frac{1}{2}k^2 d^{-\alpha} + \frac{1}{4(\log 2)^2}k^2(\log d)^2 \times$$
$$\times\left\{\exp\left(-\frac{9nt^2}{64kd_*^k\|\boldsymbol{A}\|_{\max}^2\log d_*}\right) + \exp\left(-\frac{9nt}{32k\delta_*^k\delta_{**}^{-2}d_*^k\|\boldsymbol{A}\|_{\max}\log d_*}\right)\right\}.$$

The proof relies on the following result which is an extension of Lemma 9 of Yuan and Zhang (2015) to accommodate an ℓ_∞ bound.

Lemma 1. *Let $\delta \in [1/\sqrt{d}, 1]$ and m be an integer with $2^{m/2} < \delta\sqrt{d} \le 2^{(m+1)/2}$. Then,*

$$\max_{\|\boldsymbol{u}\|_{\ell_2}\le 1, \|\boldsymbol{u}\|_{\ell_\infty}\le\delta} \boldsymbol{u}^\top\boldsymbol{a} \le (2/c)\max\left\{\boldsymbol{w}^\top\boldsymbol{a} : \|\boldsymbol{w}\|_{\ell_2}\le c, \boldsymbol{w}\in\{\pm c2^{j/2}/\sqrt{2d}, j=0,\ldots,m\}^d\right\}$$

for all $0 < c \le 1$. Moreover,

$$\left|\left\{\boldsymbol{w} : \|\boldsymbol{w}\|_{\ell_2}\le c, \boldsymbol{w}\in\{\pm c2^{j/2}/\sqrt{2d}, j=0,\ldots,m\}^d\right\}\right| \le \exp\big(1.344 + 3.082\times d\big).$$

For brevity, the proof of Lemma 1 is deferred to the Appendix. We now present the proof of Theorem 2.

Proof of Theorem 2. The standard symmetrization argument gives

$$\begin{aligned}\mathbb{P}\left\{\left\|\bar{\boldsymbol{X}}-\boldsymbol{A}\right\|_{\circ,\boldsymbol{\delta}}\geq 3t\right\} &\leq \max_{\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\in\mathscr{U}(\boldsymbol{\delta})}\mathbb{P}\left\{\left\langle\bar{\boldsymbol{X}}-\boldsymbol{A},\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\right\rangle\geq t\right\}\\ &\quad +4\,\mathbb{P}\left\{\left\|\frac{1}{n}\sum_{i=1}^n\epsilon_i\boldsymbol{X}_i\right\|_{\circ,\boldsymbol{\delta}}\geq t\right\}.\end{aligned}$$

See, e.g., Giné and Zinn (1984). For any fixed $\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\in\mathscr{U}(\boldsymbol{\delta})$, we have

$$\begin{aligned}\mathbb{E}\left\langle\boldsymbol{X}_i,\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\right\rangle &= \left\langle\boldsymbol{A},\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\right\rangle,\\ \left|\left\langle\boldsymbol{X}_i,\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\right\rangle\right| &\leq (d_1\cdots d_k)\left(\|\boldsymbol{u}_1\|_{\ell_\infty}\cdots\|\boldsymbol{u}_k\|_{\ell_\infty}\right)\|\boldsymbol{A}\|_{\max}\\ &\leq (d_1\cdots d_k)(\delta_1\cdots\delta_k)\|\boldsymbol{A}\|_{\max}/\delta_{**}^2,\end{aligned}$$

and

$$\mathrm{var}(\langle\boldsymbol{X}_i,\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\rangle)\leq\mathbb{E}\left\langle\boldsymbol{X}_i,\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\right\rangle^2\leq(d_1\cdots d_k)\|\boldsymbol{A}\|_{\max}^2.$$

Therefore, by the Bernstein inequality,

$$\begin{aligned}\mathbb{P}\left\{\left\|\bar{\boldsymbol{X}}-\boldsymbol{A}\right\|_{\circ,\boldsymbol{\delta}}\geq 3t\right\} &\leq \exp\left(-\frac{nt^2}{4d_*^k\|\boldsymbol{A}\|_{\max}^2}\right)+\exp\left(-\frac{(3/4)\delta_{**}^2nt}{d_*^k\delta_*^k\|\boldsymbol{A}\|_{\max}}\right)\\ &\quad +4\mathbb{P}\left\{\left\|\frac{1}{n}\sum_{i=1}^n\epsilon_i\boldsymbol{X}_i\right\|_{\circ,\boldsymbol{\delta}}\geq t\right\}\end{aligned}$$

We now proceed to bound the last term on the right hand side.

For brevity, write $\boldsymbol{Y}_i=\epsilon_i\boldsymbol{X}_i$ and

$$\bar{\boldsymbol{Y}}=\frac{1}{n}\sum_{i=1}^n\epsilon_i\boldsymbol{X}_i.$$

Recall that

$$\|\bar{\boldsymbol{Y}}\|_{\circ,\boldsymbol{\delta}}=\max_{1\leq j_1<j_2\leq k}\max_{\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\in\mathscr{U}_{j_1j_2}(\boldsymbol{\delta})}\langle\bar{\boldsymbol{Y}},\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\rangle.$$

Hence,

$$\mathbb{P}\left\{\|\bar{\boldsymbol{Y}}\|_{\circ,\boldsymbol{\delta}}\geq t\right\}\leq\sum_{1\leq j_1<j_2\leq k}\mathbb{P}\left\{\max_{\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\in\mathscr{U}_{j_1j_2}(\boldsymbol{\delta})}\langle\bar{\boldsymbol{Y}},\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\rangle\geq t\right\}.$$

We now bound each of the summands on the right hand side. To fix ideas, we shall treat only the case when $j_1=1$ and $j_2=2$ without loss of generality.

It follows from Lemma 1 that

$$\max_{\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\in\mathscr{U}_{1,2}(\boldsymbol{\delta})}\big\langle \bar{\boldsymbol{Y}}, \boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\big\rangle \le 2^{k+1}\max_{\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\in\mathscr{U}^*_{1,2}(\boldsymbol{\delta})}\langle \bar{\boldsymbol{Y}}, \boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\rangle.$$

where

$$\mathscr{U}^*_{1,2}(\boldsymbol{\delta}) = \Big\{\boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\in\mathscr{U}_{1,2}(\boldsymbol{\delta}) : \|\boldsymbol{u}_j\|_{\ell_2}\le c_j, \boldsymbol{u}_j\in\{\pm 2^{j/2}c_j/\sqrt{2d_j}, j=0,\ldots,m_j\}^{d_j}\Big\}$$

with $m_j = \lceil \log_2(d_j)-1\rceil$ for $j=1,2$, and $m_j = \lceil \log_2(\delta_j^2 d_j)-1\rceil$ for $j>2$. We choose $1/\sqrt{2}\le c_j\le 1$ such that $\{\pm 2^{j/2}c_j/\sqrt{2d_j}, j=0,\ldots,m_j\} = \{\pm 2^{-j/2}, j=2,\ldots,m_j+2\}$ for $j=1,2$, and $c_j=1$ for $j>2$. As $d_1+\cdots+d_k = kd$ and $d\ge 2$,

$$|\mathscr{U}^*_{1,2}(\boldsymbol{\delta})| \le \exp\big(4kd\big).$$

For $\boldsymbol{U} = \boldsymbol{u}_1\otimes\cdots\otimes\boldsymbol{u}_k\in\mathscr{U}^*_{1,2}(\boldsymbol{\delta})$, define

$$\begin{aligned} A_m &= \{(a_1,a_2) : |\boldsymbol{u}_1(a_1)\boldsymbol{u}_2(a_2)| = 2^{-m/2}\},\\ B_m &= \{(a_3,\ldots,a_k) : (a_1,a_2)\in A_m, (a_1,\ldots,a_k)\in\Omega\},\end{aligned}$$

and

$$\boldsymbol{U}_{1,2} = \boldsymbol{u}_1\otimes\boldsymbol{u}_2,\ \boldsymbol{U}_{3,\ldots,k} = \boldsymbol{u}_3\otimes\cdots\otimes\boldsymbol{u}_k.$$

Here and in the sequel, we omit the dependence of $\{A_m, B_m, \boldsymbol{U}_{1,2}, \boldsymbol{U}_{3,\ldots,k}\}$ on $\boldsymbol{U}$ and B_m on Ω when no confusion occurs. For $\boldsymbol{U}\in\mathscr{U}^*_{1,2}(\boldsymbol{\delta})$ and any integer $m_{1,2}\ge 0$,

$$\langle\bar{\boldsymbol{Y}},\boldsymbol{U}\rangle = \langle\bar{\boldsymbol{Y}}, (\mathcal{P}_{C_{1,2}}\boldsymbol{U}_{1,2})\otimes\boldsymbol{U}_{3,\ldots,k}\rangle + \sum_{4\le m\le m_{1,2}}\langle\bar{\boldsymbol{Y}}, (\mathcal{P}_{A_m}\boldsymbol{U}_{1,2})\otimes(\mathcal{P}_{B_m}\boldsymbol{U}_{3,\ldots,k})\rangle,$$

where

$$C_{1,2} = \{(a_1,a_2) : |\boldsymbol{U}_{1,2}(a_1,a_2)|\le 2^{-m_{1,2}/2-1/2}\}.$$

We note that $A_m=\emptyset$ for $m\le 3$.

Write

$$\nu_{1,2}(\bar{\boldsymbol{Y}}) = \max_{a_1\in[d_1],a_2\in[d_2]}\left|\big\{(a_1,\ldots,a_k)\in\mathrm{supp}(\bar{\boldsymbol{Y}}) : a_j\in[d_j], j\ge 3\big\}\right|.$$

We argue that

$$\mathbb{P}\left\{\nu_{1,2}(\bar{\boldsymbol{Y}})\le(3\alpha+7)\left(\frac{n}{d_1d_2}+\log d\right)\right\}\le d^{-\alpha}. \tag{4}$$

When $n/(d_1d_2) \geq \log d$, we can apply Chernoff bound to get, for any fixed $a_1 \in [d_1]$ and $a_2 \in [d_2]$

$$\begin{aligned} &\mathbb{P}\left\{\left|\left\{(a_1,\dots,a_k) \in \mathrm{supp}(\bar{\boldsymbol{Y}}) : a_j \in [d_j], j \geq 3\right\}\right| \geq (3\alpha+7)\frac{n}{d_1d_2}\right\} \\ \leq\ & \exp[-(\alpha+2)n/(d_1d_2)] \leq d^{-(\alpha+2)}. \end{aligned}$$

Similarly, when $n/(d_1d_2) < \log d$, we can also apply Chernoff bound to get

$$\mathbb{P}\left\{\left|\left\{(a_1,\dots,a_k) \in \mathrm{supp}(\bar{\boldsymbol{Y}}) : a_j \in [d_j], j \geq 3\right\}\right| \geq (3\alpha+7)\log d\right\} \leq d^{-(\alpha+2)}.$$

Equation (4) then follows from an application of the union bound.

We shall now proceed conditional on the event that

$$\nu_{1,2}(\bar{\boldsymbol{Y}}) \leq \nu_* := (3\alpha+7)\left(\frac{n}{d_1d_2} + \log d\right).$$

Under this event,

$$|B_m| \leq \nu_*|A_m|.$$

Observe that for any $\boldsymbol{U} = \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_k \in \mathscr{U}^*_{1,2}(\boldsymbol{\delta})$,

$$|A_m| \leq 2^m,\ \|\boldsymbol{U}_{3,\dots,k}\|_{\max} = \|\boldsymbol{u}_3 \otimes \cdots \otimes \boldsymbol{u}_k\|_{\max} \leq \delta_{3,\dots,k}.$$

with $\delta_{3,\dots,k} = \delta_3 \cdots \delta_k$. For integers $0 \leq \ell \leq m \leq m_{1,2}$ define,

$$\begin{aligned} \mathscr{B}_{1,2}(m,\ell) \ = \ & \Big\{\boldsymbol{V} = (\mathcal{P}_{A_m}\boldsymbol{U}_{1,2}) \otimes (\mathcal{P}_B\boldsymbol{U}_{3,\dots,k}) : |A_m| \leq 2^{m-\ell}, \\ & |B| \leq \nu_*|A_m|, \boldsymbol{U}_{1,2} \otimes \boldsymbol{U}_{3,\dots,k} \in \mathscr{U}^*_{1,2}(\boldsymbol{\delta})\Big\}. \end{aligned}$$

It follows that for $\boldsymbol{U} \in \mathscr{U}^*_{1,2}(\boldsymbol{\delta})$ and integers $a_m \geq 0$ with $2^{m-a_m-1} \leq |A_m| \leq 2^{m-a_m}$,

$$(\mathcal{P}_{A_m}\boldsymbol{U}_{1,2}) \otimes (\mathcal{P}_{B_m}\boldsymbol{U}_{3,\dots,k}) \in \mathscr{B}_{1,2}(m,\ell), \quad a_m \leq \ell.$$

As

$$\sum_{m=4}^{m_{1,2}} 2^{-(a_m \wedge (m-3))} \leq 1 + 2\sum_{m=4}^{m_{1,2}} |A_m|/2^m \leq 1 + 2\|\boldsymbol{U}_{1,2}\|_{\mathrm{F}}^2 \leq 3$$

for all $\boldsymbol{U} \in \mathscr{U}^*_{1,2}(\boldsymbol{\delta})$,

$$\sum_{4 \leq m \leq m_{1,2}} \langle \bar{\boldsymbol{Y}}, (\mathcal{P}_{A_m}\boldsymbol{U}_{1,2}) \otimes (\mathcal{P}_{B_m}\boldsymbol{U}_{3,\dots,k})\rangle$$

$$
\begin{aligned}
&\le \sum_{4\le m\le m_{1,2}} 2^{-(a_m\wedge(m-3))/2-\ell_m/2} \max_{\boldsymbol{V}\in\mathscr{B}_{1,2}(m,a_m\wedge(m-3))} 2^{(a_m\wedge(m-3))/2+\ell_m/2}\langle\bar{\boldsymbol{Y}},\boldsymbol{V}\rangle \\
&\le \left(3\sum_{m=4}^{m_{1,2}} 2^{-\ell_m}\right)^{1/2} \max_{4\le m\le m_{1,2}}\max_{0\le\ell\le m-3}\max_{\boldsymbol{V}\in\mathscr{B}_{1,2}(m,\ell)} 2^{\ell/2+\ell_m/2}\langle\bar{\boldsymbol{Y}},\boldsymbol{V}\rangle
\end{aligned}
$$

for any nonnegative integers ℓ_m. Here $a\wedge b=\min\{a,b\}$. It follows that if

$$
\left(3\sum_{m=4}^{m_{1,2}} 2^{-\ell_m}\right)^{1/2} \le 4,
$$

then

$$
\begin{aligned}
\langle\bar{\boldsymbol{Y}},\boldsymbol{U}\rangle &\le \max_{\boldsymbol{U}\in\mathscr{U}^*_{1,2}(\boldsymbol{\delta})}\langle\bar{\boldsymbol{Y}},(\mathcal{P}_{C_{1,2}}\boldsymbol{U}_{1,2})\otimes\boldsymbol{U}_{3,\ldots,k}\rangle \\
&\quad +4\max_{4\le m\le m_{1,2}}\max_{0\le\ell\le m-3}\max_{\boldsymbol{V}\in\mathscr{B}_{1,2}(m,\ell)} 2^{\ell/2+\ell_m/2}\langle\bar{\boldsymbol{Y}},\boldsymbol{V}\rangle. \qquad (5)
\end{aligned}
$$

We note that $\mathcal{P}_{C_{1,2}}=\mathcal{I}$ when $m_{1,2}\le 3$.

We have $|\mathscr{U}^*_{1,2}(\boldsymbol{\delta})|\le e^{4kd}$. To bound the cardinality of $\mathscr{B}_{1,2}(m,\ell)$, we pick

$$
m_{1,2}=\max\left\{\lfloor\log_2(4d/(\nu_*\log d_*))\rfloor,0\right\},
$$

so that

$$
\nu_* 2^{m_{1,2}}\log d_*\le 4d\le \nu_* 2^{m_{1,2}+1}\log d_*
$$

if $\nu_*\log d_*\le 4d$ and $m_{1,2}=0$ otherwise. Moreover, for $4\le m\le m_{1,2}$, we pick integers ℓ_m satisfying

$$
\max\left\{2^{m-m_{1,2}},\frac{9}{8k\log d_*}\right\}\le 2^{-\ell_m}<\max\left\{2^{m-m_{1,2}},\frac{9}{4k\log d_*}\right\}.
$$

As

$$
m_{1,2}\le\log_2 d\le k\log(d_*)/\log 2,
$$

we have

$$
\left(3\sum_{m=4}^{m_{1,2}} 2^{-\ell_m}\right)^{1/2}\le\left(\frac{27(1+m_{1,2}-3)}{4k\log d_*}\right)^{1/2}\le\left(\frac{27}{4\log 2}\right)^{1/2}\le 3.121.
$$

We note that $\boldsymbol{U}_{1,2}$ takes value $\pm 2^{-m/2}$ on A_m and $\boldsymbol{U}_{3,\ldots,k}$ takes value in $\pm 2^{j/2}/(\prod_{j=3}^k\sqrt{2d_j})$ for $j=0,\ldots,m_3+\ldots+m_k$. Let $m_{**}=k\log_2(\delta_*^2 d_*)$. As $m_j=\lceil\log_2(\delta_j^2 d_j)-1\rceil$ for $j>2$, each element of $\boldsymbol{U}_{3,\dots,k}$ has at most $2m_{**}+2$ possible values. It follows that

$$\begin{aligned}\log|\mathscr{B}_{1,2}(m,\ell)| &\le \log\left(\sum_{j=1}^{2^{m-\ell}}\binom{d_1d_2}{j}\binom{d_3\cdots d_k}{\lfloor\nu_* j\rfloor}2^j(2m_{**}+2)^{\lfloor\nu_* j\rfloor}\right)\\ &\le \nu_*2^{m-\ell}\left\{\log\left(\frac{ed_3\dots d_k}{\nu_*2^{m-\ell}}\right)+\log(2m_{**}+2)\right\}\\ &\quad +2^{m-\ell}\left\{\log\left(\frac{ed_1d_2}{2^{m-\ell}}\right)+\log 2\right\}+\log 2.\end{aligned}$$

As $x\log(y/x^2)$ is increasing in x for $0<x\le\sqrt{y}/e$ and $4\le m\le m_{1,2}-\ell_m$,

$$\begin{aligned}&2^{-(m-\ell)/2}\log|\mathscr{B}_{1,2}(m,\ell)|\\ \le\ &\nu_*2^{(m_{1,2}-\ell_m)/2}\left\{\log\left(\frac{ed_3\dots d_k}{\nu_*2^{m_{1,2}-\ell_m}}\right)+\log(2m_{**}+2)\right\}\\ &+2^{(m_{1,2}-\ell_m)/2}\left\{\log\left(\frac{ed_1d_2}{2^{m_{1,2}-\ell_m}}\right)+2\log 2\right\}\\ \le\ &\nu_*2^{(m_{1,2}-\ell_m)/2}\left\{\log\left(\frac{e(d_1d_2)^{1/\nu_*}d_3\dots d_k}{\nu_*2^{m_{1,2}-\ell_m}}\right)+\log(2m_{**}+2)\right\}\\ \le\ &\nu_*2^{-\ell_m/2}\left(\frac{4d}{\nu_*\log d_*}\right)^{1/2}\log\left(\frac{d_*^k e(d_1d_2)^{1/\nu_*}2^{\ell_m}(2m_{**}+2)}{d_1d_24d/\log d_*}\right).\end{aligned}$$

Note that

$$\begin{aligned}&e(d_1d_2)^{1/\nu_*}2^{\ell_m}(2m_{**}+2)\log d_*\\ \le\ &(d_1d_2)^{1/\{(1+\alpha)\log d\}}(8e/9)k(\log d_*)^2\{2k\log_2(\delta_*^2d_*)+2\}\\ \le\ &4d_1d_2d,\end{aligned}$$

where the last inequality follows from the fact that $d_*<d$ and the assumption that d is sufficiently large. Thus,

$$2^{-(m-\ell)/2}\log|\mathscr{B}_{1,2}(m,\ell)|\le 2^{-\ell_m/2}k\sqrt{4\nu_*d\log d_*}.$$

It follows that

$$\log|\mathscr{B}_{1,2}(m,\ell)|\le 2^{(m-\ell-\ell_m)/2}k\sqrt{4\nu_*d\log d_*}\le 4kd,\qquad \forall\ 0\le\ell\le m\le m_{1,2}.$$

For any fixed $\boldsymbol{V}\in\mathscr{B}_{1,2}(m,\ell)$, write $Z_i=\langle\boldsymbol{Y}_i,\boldsymbol{V}\rangle$. Then

$$\left\langle\bar{\boldsymbol{Y}},\boldsymbol{V}\right\rangle=\frac{1}{n}(Z_1+\cdots+Z_n).$$

We have

$$\|\boldsymbol{V}\|_{\max} \le 2^{-m/2}\delta_{3,\dots,k} \qquad \text{and} \qquad \|\boldsymbol{V}\|_{\mathrm{HS}}^2 \le 2^{-\ell}.$$

Thus, as $\boldsymbol{Y}_i = \epsilon_i \boldsymbol{X}_i$ and $\boldsymbol{X}_i = (d_1 \cdots d_k)\mathcal{P}_{\omega_i}\boldsymbol{A}$, we have

$$|Z_i| \le d_*^k \|\boldsymbol{A}\|_{\max}\|\boldsymbol{V}\|_{\max} \le 2^{-m/2}\delta_{3,\dots,k} d_*^k \|\boldsymbol{A}\|_{\max}$$

and

$$\mathrm{var}(Z_i) \le \mathbb{E}(Z_i^2) \le d_*^k \|\boldsymbol{A}\|_{\max}^2 \|\boldsymbol{V}\|_{\mathrm{HS}}^2 \le 2^{-\ell} d_*^k \|\boldsymbol{A}\|_{\max}^2.$$

It follows from the Bernstein inequality and the union bound that

$$\begin{aligned}
&\mathbb{P}\left\{\max_{\boldsymbol{V}\in\mathscr{B}_{1,2}(m,\ell)} \left\langle \bar{\boldsymbol{Y}}, \boldsymbol{V}\right\rangle \ge 2^{-(\ell+\ell_m)/2} t\right\} \\
\le\ & |\mathscr{B}_{1,2}(m,\ell)| \exp\left(-\frac{n2^{-\ell-\ell_m}t^2}{2^{1-\ell}d_*^k\|\boldsymbol{A}\|_{\max}^2 + (2/3)2^{-m/2}\delta_{3,\dots,k}d_*^k\|\boldsymbol{A}\|_{\max}2^{-(\ell+\ell_m)/2}t}\right) \\
\le\ & \exp\left(4kd - \frac{n2^{-\ell_m}t^2}{4d_*^k\|\boldsymbol{A}\|_{\max}^2}\right) + \exp\left(2^{(m-\ell-\ell_m)/2}k\sqrt{4\nu_* d\log d_*} - \frac{(3/4)2^{(m-\ell-\ell_m)/2}nt}{\delta_{3,\dots,k}d_*^k\|\boldsymbol{A}\|_{\max}}\right).
\end{aligned}$$

The condition on t implies that

$$t \ge \frac{8}{3n}(\delta_{3,\dots,k}d_*^k\|\boldsymbol{A}\|_{\max})k\sqrt{4\nu_* d\log d_*}.$$

Together with the fact that $2^{-\ell_m} \ge (9/8)/(k\log d_*)$, we get

$$\frac{n2^{-\ell_m}t^2}{4d_*^k\|\boldsymbol{A}\|_{\max}^2} \ge \frac{2\delta_{3,\dots,k}^2 d_*^k k^2(4\nu_* d\log d_*)}{nk\log d_*} \ge (d_1d_2\nu_*/n)8kd \ge 8kd.$$

Therefore,

$$\exp\left(4kd - \frac{n2^{-\ell_m}t^2}{4d_*^k\|\boldsymbol{A}\|_{\max}^2}\right) \le \exp\left(-\frac{n2^{-\ell_m}t^2}{8d_*^k\|\boldsymbol{A}\|_{\max}^2}\right) \le \exp\left(-\frac{9nt^2}{64kd_*^k\|\boldsymbol{A}\|_{\max}^2\log d_*}\right).$$

Similarly, we have

$$\frac{(3/4)2^{(m-\ell-\ell_m)/2}nt}{\delta_{3,\dots,k}d_*^k\|\boldsymbol{A}\|_{\max}} \ge 2\cdot 2^{(m-\ell-\ell_m)/2}k\sqrt{4\nu_* d\log d_*},$$

which implies that

$$\begin{aligned}
&\exp\left(2^{(m-\ell-\ell_m)/2}k\sqrt{4\nu_* d\log d_*} - \frac{(3/4)2^{(m-\ell-\ell_m)/2}nt}{\delta_{3,\dots,k}d_*^k\|\boldsymbol{A}\|_{\max}}\right) \\
\le\ & \exp\left(-\frac{3}{8}\cdot\frac{2^{(m-\ell-\ell_m)/2}nt}{\delta_{3,\dots,k}d_*^k\|\boldsymbol{A}\|_{\max}}\right) \\
\le\ & \exp\left(-\frac{9}{32}\cdot\frac{nt}{(k\log d_*)^{1/2}\delta_{3,\dots,k}d_*^k\|\boldsymbol{A}\|_{\max}}\right).
\end{aligned}$$

For $m_{1,2} \geq 1$, we have

$$2^{-(m_{1,2}+1)/2} \leq \sqrt{(\nu_* \log d_*)/(4d)},$$

so that

$$\frac{3}{4} nt 2^{(m_{1,2}+1)/2} / (\delta_{3,\ldots,k} d_*^k \|\boldsymbol{A}\|_{\max}) \geq 2k\sqrt{4\nu_* d \log d_*}\sqrt{4d/(\nu_* \log d_*)} = 8kd.$$

As

$$|\langle \epsilon_i \boldsymbol{X}_i, (\mathcal{P}_{C_{1,2}} \boldsymbol{U}_{1,2}) \otimes \boldsymbol{U}_{3,\ldots,k} \rangle| \leq 2^{-(m_{1,2}+1)/2} \delta_{3,\ldots,k} d_*^k \|\boldsymbol{A}\|_{\max},$$

we have

$$\begin{aligned}
& \mathbb{P}\left\{ \max_{\boldsymbol{U} \in \mathscr{U}_{1,2}^*(\boldsymbol{\delta})} \langle \bar{\boldsymbol{Y}}, (\mathcal{P}_{C_{1,2}} \boldsymbol{U}_{1,2}) \otimes \boldsymbol{U}_{3,\ldots,k} \rangle \geq t \right\} \\
\leq\ & |\mathscr{U}_{1,2}^*(\boldsymbol{\delta})| \max_{\boldsymbol{U} \in \mathscr{U}_{1,2}^*(\boldsymbol{\delta})} \mathbb{P}\left\{ \langle \bar{\boldsymbol{Y}}, (\mathcal{P}_{C_{1,2}} \boldsymbol{U}_{1,2}) \otimes \boldsymbol{U}_{3,\ldots,k} \rangle \geq t \right\} \\
\leq\ & \exp\left(4kd - \frac{nt^2}{2d_*^k \|\boldsymbol{A}\|_{\max}^2 + 2^{1-(m_{1,2}+1)/2} \delta_{3,\ldots,k} d_*^k \|\boldsymbol{A}\|_{\max} t/3} \right) \\
\leq\ & \exp\left(-\frac{nt^2}{4d_*^k \|\boldsymbol{A}\|_{\max}^2} \right) + \exp\left(-\frac{3d^{1/2} nt}{2\delta_{3,\ldots,k} d_*^k \|\boldsymbol{A}\|_{\max} (\nu_* \log d_*)^{1/2}} \right).
\end{aligned}$$

Finally, for $m_{1,2} = 0$, we have $\nu_* > 4d/\log d_*$, so that the condition on t still implies

$$\frac{(3/4)nt}{\delta_{3,\ldots,k} d_*^k \|\boldsymbol{A}\|_{\max}} \geq 2k\sqrt{4\nu_* d \log d_*} \geq 8kd.$$

Putting the above probability bounds together via (5), we find that

$$\begin{aligned}
& \mathbb{P}\left\{ \max_{\boldsymbol{U} \in \mathscr{U}_{1,2}(\boldsymbol{\delta})} \langle \bar{\boldsymbol{Y}}, \boldsymbol{U} \rangle \geq 2^{k+1} 5t \right\} \\
\leq\ & \mathbb{P}\left\{ \max_{\boldsymbol{U} \in \mathscr{U}_{1,2}^*(\boldsymbol{\delta})} \langle \bar{\boldsymbol{Y}}, \boldsymbol{U} \rangle \geq 5t \right\} \\
\leq\ & \Big(1 + 2 + \ldots + (m_{1,2} - 2) \Big) \times \\
& \times \left\{ \exp\left(-\frac{9nt^2}{64kd_*^k \|\boldsymbol{A}\|_{\max}^2 \log d_*} \right) + \exp\left(-\frac{9}{32} \cdot \frac{nt}{(k \log d_*)^{1/2} \delta_{3,\ldots,k} d_*^k \|\boldsymbol{A}\|_{\max}} \right) \right\}.
\end{aligned}$$

As $m_{1,2} \leq \log_2 d$, the proof is then completed in the light of (4). □

It is instructive to examine the case of hypercubic tensors where $d_1 = \cdots = d_k = d$ and we take $\delta_1 = \cdots = \delta_k = \delta_*$. The following is an immediate consequence of Theorem 2.

Corollary 1. *Let $\boldsymbol{A} \in \mathbb{R}^{d\times\cdots\times d}$ be a kth order tensor, and $\delta_1 = \cdots = \delta_k = \delta \in (0,1]$, then there exists constant $c_1, c_2 > 0$ depending on k only such that, for any $\beta > 0$,*

$$\left\|\bar{\boldsymbol{X}} - \boldsymbol{A}\right\|_{\circ,\boldsymbol{\delta}} \le c_1(1+\beta)\max\left\{\left(\frac{\log d}{n}\right)^{1/2}\delta^{k-2}d^{k-1/2}, \left(\frac{\log d}{n}\right)\delta^{k-2}d^{k+1/2}\right\}\|\boldsymbol{A}\|_{\max}, \quad (6)$$

with probability at least $1 - c_2 d^{-\beta}$.

Note that the second term on the right hand side of (6) decreases with δ, indicating a tighter concentration bound for $\bar{\boldsymbol{X}} - \boldsymbol{A}$ when it dominates the first term. The bound (6) immediately suggests an effective sampling scheme to approximate incoherent tensors in terms of the usual spectral norm. Suppose that $\boldsymbol{A}$ is μ-incoherent so that

$$\max_{\|\boldsymbol{u}\|_{\ell_2}\le 1}\|\boldsymbol{P}_j(\boldsymbol{A})\boldsymbol{u}\|_{\ell_\infty} \le \sqrt{\mu r_j(\boldsymbol{A})/d}, \qquad j = 1,\ldots,k.$$

Then we can take $\delta = 2\sqrt{\mu r/d}$ where $r = \max_j r_j(\boldsymbol{A})$. Equation (6) now becomes

$$\left\|\bar{\boldsymbol{X}} - \boldsymbol{A}\right\|_{\circ,\boldsymbol{\delta}} \lesssim (\mu r)^{k/2-1}\max\left\{\left(\frac{\log d}{n}\right)^{1/2}d^{(k+1)/2}, \left(\frac{\log d}{n}\right)d^{(k+3)/2}\right\}\|\boldsymbol{A}\|_{\max}.$$

Let $\widehat{\boldsymbol{A}}$ be the projection of $\bar{X}$ onto the space $\mathcal{T}_\mu$ of μ-incoherent tensors:

$$\widehat{\boldsymbol{A}} = \underset{\boldsymbol{Y}\in\mathcal{T}_\mu}{\arg\min}\|\bar{\boldsymbol{X}} - \boldsymbol{Y}\|_{\circ,\boldsymbol{\delta}}.$$

By triangular inequality, $\|\widehat{\boldsymbol{A}} - \boldsymbol{A}\|_{\circ,\boldsymbol{\delta}} \le 2\left\|\bar{\boldsymbol{X}} - \boldsymbol{A}\right\|_{\circ,\boldsymbol{\delta}}$, so that

$$\|\widehat{\boldsymbol{A}} - \boldsymbol{A}\|_{\circ,\boldsymbol{\delta}} \lesssim (\mu r)^{k/2-1}\max\left\{\left(\frac{\log d}{n}\right)^{1/2}d^{(k+1)/2}, \left(\frac{\log d}{n}\right)d^{(k+3)/2}\right\}\|\boldsymbol{A}\|_{\max}.$$

Because both $\widehat{\boldsymbol{A}}$ and $\boldsymbol{A}$ are μ-coherent. Their difference $\widehat{\boldsymbol{A}} - \boldsymbol{A}$ must be $\sqrt{2}\mu$-coherent. In the light of Proposition 2, we know $\|\widehat{\boldsymbol{A}} - \boldsymbol{A}\| = \|\widehat{\boldsymbol{A}} - \boldsymbol{A}\|_{\circ,\boldsymbol{\delta}}$, so that

$$\|\widehat{\boldsymbol{A}} - \boldsymbol{A}\| \lesssim (\mu r)^{k/2-1}\max\left\{\left(\frac{\log d}{n}\right)^{1/2}d^{(k+1)/2}, \left(\frac{\log d}{n}\right)d^{(k+3)/2}\right\}\|\boldsymbol{A}\|_{\max}. \quad (7)$$

In other words, we can approximate $\boldsymbol{A}$ up to the same error bound given by (6), but in terms of the usual spectral norm.

For illustration purposes, consider a more specific case when $\boldsymbol{A}$ admits an orthogonal decomposition

$$\boldsymbol{A} = \sum_{i=1}^{r} \boldsymbol{u}_1^{(i)} \otimes \cdots \otimes \boldsymbol{u}_k^{(i)},$$

for some $\boldsymbol{u}_j^{(i)} \in \mathbb{R}^d$ such that

$$\langle \boldsymbol{u}_j^{(i_1)}, \boldsymbol{u}_j^{(i_2)} \rangle = \begin{cases} 1 & \text{if } i_1 = i_2 \\ 0 & \text{otherwise} \end{cases}.$$

If $\boldsymbol{A}$ is μ-incoherent in that

$$\|\boldsymbol{u}_j^{(i)}\|_{\ell_\infty} \leq \sqrt{\frac{\mu}{d}}, \qquad j = 1, \ldots, k, i = 1, \ldots, r.$$

then

$$\|\boldsymbol{A}\|_{\max} \leq \mu^{k/2} r d^{-k/2}.$$

The approximation error bound given by (7) can now be further simplified as

$$\|\widehat{\boldsymbol{A}} - \boldsymbol{A}\| \lesssim \mu^{k-1} r^{k/2} \max\left\{ \left(\frac{d \log d}{n} \right)^{1/2}, \frac{d^{3/2} \log d}{n} \right\}.$$

In other words, when $\mu^{k-1} = O(1)$, we can approximate $\boldsymbol{A}$ up to an error of ϵ, in terms of the usual spectral norm, based on observations from

$$n \geq C_k \max\left(\frac{r^k d \log d}{\epsilon^2}, \frac{r^{k/2} d^{3/2} \log d}{\epsilon} \right)$$

entries for some constant C_k. If the condition on $\boldsymbol{A}$ is strengthened to $\|\boldsymbol{A}\|_{\max} \lesssim \mu^{k/2} r^{1/2} d^{-k/2}$, then the sample size requirement becomes

$$n \geq C_k \max\left(\frac{r^{k-1} d \log d}{\epsilon^2}, \frac{r^{(k-1)/2} d^{3/2} \log d}{\epsilon} \right).$$

This example shows the importance of leveraging the information that a tensor is incoherent.

4 Tensor Completion

We now turn our attention back to tensor completion through incoherent nuclear norm minimization:

$$\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_{\star, \boldsymbol{\delta}} \text{ subject to } \mathcal{P}_\Omega \boldsymbol{X} = \mathcal{P}_\Omega \boldsymbol{T}. \tag{8}$$

Denote by $\widehat{\boldsymbol{T}}$ the solution to the above convex optimization problem. We shall utilize the results from the previous sections to establish the requirement on the sample size $n := |\Omega|$ so that $\widehat{\boldsymbol{T}} = \boldsymbol{T}$ with high probability when Ω is a uniformly sampled subset of $[d_1] \times \cdots \times [d_k]$.

Recall that $r_j(\boldsymbol{T})$s are the Tucker ranks of $\boldsymbol{T}$. For brevity, we shall omit the dependence of r_js on $\boldsymbol{T}$ for the rest of the section. Denote by

$$r_* = \left[\frac{1}{kd}\sum_{j=1}^{k}\left(\frac{d_j}{r_j}\prod_{\ell=1}^{k} r_\ell\right)\right]^{1/(k-1)},$$

$$\mu_* = \frac{d_*^k}{k r_*^{k-1} d}\max_{i_1,\dots,i_k}\|\mathcal{Q}_{\boldsymbol{T}}(e_{i_1}\otimes\cdots\otimes e_{i_k})\|_{\rm HS}^2, \tag{9}$$

and

$$\alpha_* = (d_*^k/r_*)^{1/2}\|\boldsymbol{W}_0\|_{\max}, \tag{10}$$

where as before, d and d_* are the arithmetic and geometric averages of d_js, and $\boldsymbol{W}_0 \in \mathbb{R}^{d_1\times\cdots\times d_k}$ is the dual of $\boldsymbol{T}$ as specified in Theorem 1. We are now in position to state our main result.

Theorem 3. *Let Ω be a uniformly sampled subset of $[d_1]\times\cdots\times[d_k]$ and $\widehat{\boldsymbol{T}}$ be the solution to (8) with $\delta_j = \sqrt{\lambda_* r_*/d_j}$. There exists a constant $c_k > 0$ depending on k only so that $\mathbb{P}\{\widehat{\boldsymbol{T}} = \boldsymbol{T}\} \ge 1 - d^{-\beta}$ if*

$$\lambda_* \ge \frac{1}{r_*}\max_{1\le j\le k}\{\mu_j(\boldsymbol{T})r_j(\boldsymbol{T})\},$$

and

$$n := |\Omega| \ge c_k(1+\beta)\left((\mu_* + \alpha_*^2\lambda_*^{k-2})r_*^{k-1}d(\log d)^2 + \alpha_*\lambda_*^{k/2-1}r_*^{(k-1)/2}d^{3/2}(\log d)^2\right)$$

Proof of Theorem 3. The main steps of the proof is analogous to those from Yuan and Zhang (2014). We shall outline below these steps while highlighting the key differences moving from third order tensors to higher order tensors, and from usual tensor nuclear norm to incoherent tensor nuclear norm. We begin with a lemma that reduces the problem to finding a dual certificate.

Lemma 2. *Suppose there exists a tensor* $\widetilde{\boldsymbol{G}} \in \mathbb{R}^{d_1\times\cdots\times d_k}$ *such that* $\widetilde{\boldsymbol{G}} = \mathcal{P}_\Omega \widetilde{\boldsymbol{G}}$,

$$\|\mathcal{Q}_{\boldsymbol{T}}\widetilde{\boldsymbol{G}} - \boldsymbol{W}_0\|_{\rm HS} < \frac{\sqrt{n/(2d_*^k)}}{k(k-1)} \tag{11}$$

and

$$\max_{\|\mathcal{Q}_{\boldsymbol{T}}^\perp \boldsymbol{X}\|_{\star,\boldsymbol{\delta}}=1} \langle \widetilde{\boldsymbol{G}}, \mathcal{Q}_{\boldsymbol{T}}^\perp \boldsymbol{X}\rangle < \frac{1}{k(k-1)}. \tag{12}$$

If in addition,

$$\left\|\mathcal{P}_\Omega|_{\text{range}(\mathcal{Q}_{\boldsymbol{T}})}\right\|_{\rm HS\to HS} := \inf\left\{\|\mathcal{P}_\Omega \mathcal{Q}_{\boldsymbol{T}}\boldsymbol{X}\|_{\rm HS} : \|\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{X}\|_{\rm HS}=1\right\} \geq \sqrt{\frac{n}{2d_*^k}}, \tag{13}$$

then $\widehat{\boldsymbol{T}} = \boldsymbol{T}$.

The proof of Lemma 2 is relegated to the proof. In the light of Lemma 2, it now suffices to verify condition (13) and construct a dual certificate $\widetilde{\boldsymbol{G}}$ that satisfies conditions (11) and (12). We first verify condition (13).

Recall that for a linear operator $\mathcal{R} : \mathbb{R}^{d_1\times\cdots\times d_k} \to \mathbb{R}^{d_1\times\cdots\times d_k}$,

$$\|\mathcal{R}\|_{\rm HS\to HS} = \max\left\{\|\mathcal{R}\boldsymbol{X}\|_{\rm HS} : \boldsymbol{X} \in \mathbb{R}^{d_1\times\cdots\times d_k}, \|\boldsymbol{X}\|_{\rm HS} \leq 1\right\}.$$

Here we prove that under the Hilbert-Schmidt norm in the range of $\mathcal{Q}_{\boldsymbol{T}}$,

$$\left\|\mathcal{Q}_{\boldsymbol{T}}\Big((d_*^k/n)\mathcal{P}_\Omega - \mathcal{I}\Big)\mathcal{Q}_{\boldsymbol{T}}\right\|_{\rm HS\to HS} \leq 1/2 \tag{14}$$

with large probability. This implies that as an operator in the range of $\mathcal{Q}_{\boldsymbol{T}}$, the spectrum of $(d_*^k/n)\mathcal{Q}_{\boldsymbol{T}}\mathcal{P}_\Omega\mathcal{Q}_{\boldsymbol{T}}$ is contained in $[1/2, 3/2]$. Consequently, (13) holds via

$$(d_*^k/n)\|\mathcal{P}_\Omega\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{X}\|_{\rm HS}^2 = \left\langle \mathcal{Q}_{\boldsymbol{T}}\boldsymbol{X}, (d_*^k/n)\mathcal{Q}_{\boldsymbol{T}}\mathcal{P}_\Omega\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{X}\right\rangle \geq \frac{1}{2}\|\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{X}\|_{\rm HS}^2.$$

This goal can be achieved by invoking the following result.

Lemma 3. *Let* Ω *be a uniformly sampled subset from* $[d_1]\times\cdots\times[d_k]$ *without replacement. Then,*

$$\mathbb{P}\left\{\left\|\mathcal{Q}_{\boldsymbol{T}}\left(\frac{d_*^k}{n}\mathcal{P}_\Omega - \mathcal{I}\right)\mathcal{Q}_{\boldsymbol{T}}\right\|_{\rm HS\to HS} \geq \tau\right\} \leq 2kr_*^{k-1}d\,\exp\left(-\frac{\tau^2/2}{1+2\tau/3}\left(\frac{n}{k\mu_* r_*^{k-1}d}\right)\right).$$

Lemma 3 can be proved using the same argument from Yuan and Zhang (2014) in treating low-rank tensors, noting that

$$\mathrm{rank}(\mathcal{Q}_{\boldsymbol{T}}) = \dim\big(\mathrm{range}(\mathcal{Q}_{\boldsymbol{T}})\big) \le \sum_{j=1}^{k} d_j \prod_{\ell\neq j} r_\ell = r_*^{k-1} d.$$

The details are omitted for brevity.

Equation (14) follows immediately from Lemma 3 as soon as

$$n \ge c_k(\beta+1)\mu_* r_*^{k-1} d \log(d).$$

It now remains to show that there exists a dual certificate $\widetilde{\boldsymbol{G}}$ that satisfies conditions (11) and (12). To this end, we apply the now standard "Golfing scheme". See, e.g., Gross (2011) and Recht (2011). As argued by Yuan and Zhang (2014), we can construct a sequence $\{\omega_i : 1 \le i \le n\}$ of iid uniform vectors from $[d_1]\times\cdots\times[d_k]$ such that $\omega_i \in \Omega$ for all $1 \le i \le n$. Let n_1 and n_2 be two natural numbers to be specified later so that $n_1 n_2 \le n$. Write

$$\Omega_j = \{\omega_i : (j-1)n_1 < i \le j n_1\},$$

for $j = 1, 2, \ldots, n_2$. Define

$$\mathcal{R}_j = \mathcal{I} - \frac{1}{n_1}\sum_{i=(j-1)n_1+1}^{j n_1} d_*^k \mathcal{P}_{\omega_i} \tag{15}$$

and

$$\widetilde{\boldsymbol{G}}_j = \sum_{\ell=1}^{j}\big(\mathcal{I}-\mathcal{R}_\ell\big)\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_{\ell-1}\mathcal{Q}_{\boldsymbol{T}}\cdots\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_1\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{W}_0, \quad \widetilde{\boldsymbol{G}} = \widetilde{\boldsymbol{G}}_{n_2}. \tag{16}$$

Since $\omega_i \in \Omega$,

$$\mathcal{P}_\Omega(\mathcal{I}-\mathcal{R}_j) = \mathcal{I}-\mathcal{R}_j,$$

so that $\mathcal{P}_\Omega\widetilde{\boldsymbol{G}} = \widetilde{\boldsymbol{G}}$. It follows from the definition of $\widetilde{\boldsymbol{G}}_j$ that

$$\begin{aligned}\mathcal{Q}_{\boldsymbol{T}}\widetilde{\boldsymbol{G}}_j &= \sum_{\ell=1}^{j}(\mathcal{Q}_{\boldsymbol{T}} - \mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_\ell\mathcal{Q}_{\boldsymbol{T}})(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_{\ell-1}\mathcal{Q}_{\boldsymbol{T}})\cdots(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_1\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{W}_0)\\ &= \boldsymbol{W}_0 - (\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_j\mathcal{Q}_{\boldsymbol{T}})\cdots(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_1\mathcal{Q}_{\boldsymbol{T}})\boldsymbol{W}_0\end{aligned}$$

and for any $\boldsymbol{X} \in \mathbb{R}^{d_1\times\cdots\times d_k}$,

$$\langle \widetilde{\boldsymbol{G}}_j, \mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{X}\rangle = -\Big\langle \sum_{\ell=1}^{j} \mathcal{R}_\ell(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_{\ell-1}\mathcal{Q}_{\boldsymbol{T}})\cdots(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_1\mathcal{Q}_{\boldsymbol{T}})\boldsymbol{W}_0, \mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{X}\Big\rangle.$$

Thus, conditions (11) and (12) hold if

$$\|(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_{n_2})\cdots(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_1)\boldsymbol{W}_0\|_{\rm HS} < \frac{\sqrt{n/(2d_*^k)}}{k(k-1)} \tag{17}$$

and

$$\left\|\sum_{\ell=1}^{n_2} \mathcal{R}_\ell(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_{\ell-1}\mathcal{Q}_{\boldsymbol{T}})\cdots(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_1\mathcal{Q}_{\boldsymbol{T}})\boldsymbol{W}_0\right\|_{\circ,\boldsymbol{\delta}} < \frac{1}{k(k-1)}. \tag{18}$$

We still need to prove that (17) and (18) hold with high probability. For this purpose, we need large deviation bounds for the average of certain iid tensors under the operator, maximum and spectrum norms. The large deviation bounds for the operator and maximum norms are presented in the following lemma.

Lemma 4. *Let ω_i, $i = 1, \ldots, n_1$ be iid uniformly sampled from $[d_1] \times \cdots \times [d_k]$, and*

$$\mathcal{D}_i = \mathcal{Q}_{\boldsymbol{T}}(d_*^k \mathcal{P}_{\omega_i})\mathcal{Q}_{\boldsymbol{T}} - \mathcal{Q}_{\boldsymbol{T}}.$$

Then, for all $\tau > 0$,

$$\mathbb{P}\left\{\left\|\frac{1}{n_1}\sum_{i=1}^{n_1}\mathcal{D}_i\right\|_{\rm HS\to HS} > \tau\right\} \le 2(r_*^{k-1}d)\exp\left(-\frac{\tau^2/2}{1+2\tau/3}\left(\frac{n_1}{\mu_* r_*^{k-1}d}\right)\right). \tag{19}$$

Moreover, for any deterministic $\boldsymbol{X} \in \mathbb{R}^{d_1\times\cdots\times d_k}$ with $\|\boldsymbol{X}\|_{\max} \le 1$,

$$\mathbb{P}\left\{\left\|\frac{1}{n_1}\sum_{i=1}^{n_1}\mathcal{D}_i\boldsymbol{X}\right\|_{\max} \ge \tau\right\} \le 2d_*^k \exp\left(-\frac{\tau^2/2}{1+2\tau/3}\left(\frac{n_1}{\mu_* r_*^{k-1}d}\right)\right). \tag{20}$$

Lemma 4 again follows from identical arguments used by Yuan and Zhang (2014) and the details are omitted for brevity.

Let

$$\boldsymbol{W}_j = (\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_j\mathcal{Q}_{\boldsymbol{T}})\cdots(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_1\mathcal{Q}_{\boldsymbol{T}})\boldsymbol{W}$$

with $\boldsymbol{W}_0 = \boldsymbol{W}$. Since $\mathcal{R}_j$s are iid operators with

$$\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_1\mathcal{Q}_{\boldsymbol{T}} = -(1/n_1)\sum_{i=1}^{n_1}\mathcal{D}_i,$$

Equation (19) yields

$$\begin{aligned}
& \mathbb{P}\left\{\|\boldsymbol{W}_j\|_{\mathrm{HS}} \le \tau_1^j \|\boldsymbol{W}\|_{\mathrm{HS}}, 1 \le j \le n_2\right\} \\
= \;& \mathbb{P}\left\{\|(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_j\mathcal{Q}_{\boldsymbol{T}})\cdots(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_1\mathcal{Q}_{\boldsymbol{T}})\boldsymbol{W}\|_{\mathrm{HS}} \le \tau_1^j \|\boldsymbol{W}\|_{\mathrm{HS}}, 1 \le j \le n_2\right\} \\
\ge \;& 1 - n_2 2(r_*^{k-1}d)\exp\left(-\frac{\tau_1^2/2}{1+2\tau_1/3}\left(\frac{n_1}{\mu_* r_*^{k-1} d}\right)\right).
\end{aligned}$$

This can be used to verify (17) with certain τ_1 satisfying

$$\tau_1^{n_2}\|\boldsymbol{W}\|_{\mathrm{HS}} \le \frac{\sqrt{n/(2d_*^k)}}{k(k-1)},$$

by taking

$$n \ge n_1 n_2 \ge c_k(\beta+1)\mu_* r_*^{k-1} d\log^2(d).$$

Finally, we prove (18). It follows from (20) that

$$\begin{aligned}
& \mathbb{P}\left\{\|\boldsymbol{W}_j\|_{\max} = \|(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_j\mathcal{Q}_{\boldsymbol{T}})\cdots(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_1\mathcal{Q}_{\boldsymbol{T}})\boldsymbol{W}\|_{\max} \le \tau^j\|\boldsymbol{W}\|_{\max},\ 1 \le j \le n_2\right\} \\
\ge \;& 1 - 2n_2 d_*^k \exp\left(-\frac{\tau^2/2}{1+2\tau/3}\left(\frac{n_1}{\mu_* r_*^{k-1} d}\right)\right). \qquad (21)
\end{aligned}$$

It follows from the definition of $\mathcal{R}_j$ in (15) that for any $\boldsymbol{X}$ with $\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{X} = \boldsymbol{X}$,

$$\mathcal{R}_j\boldsymbol{X} = -\frac{1}{n_1}\sum_{i=(j-1)n_1+1}^{jn_1}\left((d_*^k)\mathcal{P}_{\omega_i} - \mathcal{I}\right)\boldsymbol{X}.$$

Recall that

$$\|\boldsymbol{W}\|_{\max} = \alpha_*(r_*/d_*^k)^{1/2}.$$

Note that $\{\omega_i : (j-1)n_1 < i \le jn_1\}$ is independent of $\boldsymbol{W}_{j-1}$ and $\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{W}_{j-1} = \boldsymbol{W}_{j-1}$. By Theorem 2, we have

$$\begin{aligned}
& \mathbb{P}\left\{\|\mathcal{R}_j\boldsymbol{W}_{j-1}\|_{\circ,\boldsymbol{\delta}} > \tau^{j-1}t, \|\boldsymbol{W}_{j-1}\|_{\max}/\tau^{j-1} \le \|\boldsymbol{W}\|_{\max}\right\} \\
\le \;& k^2 d^{-\alpha}/2 + (k^2(\log_2 d)^2/4)\left\{\exp(-4kd) + \exp\left(-\sqrt{4kd(3\alpha+7)\log d}\right)\right\} \\
=: \;& p_{n_1}(t).
\end{aligned}$$

We note that as $\delta_j = \sqrt{\lambda_* r_*/d_j}$ and $\alpha_* = (d_*^k/r_*)^{1/2}\|\boldsymbol{W}_0\|_{\max}$,

$$t \ge \frac{c_k'}{n_1}(3\alpha+7)\sqrt{d\log d}(\lambda_* r_*)^{k/2}\alpha_* r_*^{1/2}\max_{1\le j_1<j_2\le k}\left\{(\lambda_* r_*)^{-2}(\alpha+1)\left(n_1 + d_{j_1}d_{j_2}\log d\right)\right\}^{1/2}$$

$$= \frac{c_k'}{n_1}\sqrt{d\log d_*}(\delta_*^k d_*^k\|\boldsymbol{W}\|_{\max}) \max_{1\le j_1<j_2\le k}\left\{\left(\frac{n_1}{\delta_{j_1}^2 d_{j_1}\delta_{j_2}^2 d_{j_2}}+\frac{\log d}{\delta_{j_1}^2\delta_{j_2}^2}\right)\right\}^{1/2}$$

with $c_k' = 2^k 160$. Together with (21), this yields

$$\begin{aligned}
&\mathbb{P}\left\{\left\|\sum_{j=1}^{n_2}\mathcal{R}_j(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_{j-1}\mathcal{Q}_{\boldsymbol{T}})\cdots(\mathcal{Q}_{\boldsymbol{T}}\mathcal{R}_1\mathcal{Q}_{\boldsymbol{T}})\boldsymbol{W}\right\|_{\circ,\boldsymbol{\delta}} < \frac{1}{k(k-1)}\right\} \\
\ge\ & \mathbb{P}\left\{\|\mathcal{R}_j\boldsymbol{W}_{j-1}\|_{\circ,\boldsymbol{\delta}} < \frac{\tau^{j-1}-\tau^j}{k(k-1)}, \|\boldsymbol{W}_{j-1}\|_{\max}/\tau^{j-1}\le\|\boldsymbol{W}\|_{\max},\ j\le n_2\right\} \\
\ge\ & 1-n_2 p_{n_1}\left(\frac{1-\tau}{k(k-1)}\right) - 2n_2 d_*^k\exp\left(-\frac{\tau^2/2}{1+2\tau/3}\left(\frac{n_1}{\mu_* r_*^{k-1}d}\right)\right),
\end{aligned}$$

which completes the proof. □

5 Concluding Remarks

We introduce a general framework of nuclear norm minimization for tensor completion and investigate the minimum sample size required to ensure prefect recovery. Our work contributes to a fast-growing literature on higher order tensors, beyond matrices. In particular, we argue that incoherence may play a more prominent role in higher order tensor completion. We show that, by appropriately incorporating information about the incoherence of a kth order tensor of rank r and dimension $d\times\cdots\times d$, we can complete it with $O((r^{(k-1)/2}d^{3/2}+r^{k-1}d)(\log(d))^2)$ uniformly sampled entries. This sample size requirement agrees with existing results on recovering a third order tensor (see, e.g., Yuan and Zhang, 2014), and more interestingly, it depends on $k(\ge 3)$ only through the $O(1)$ factor for rank one tensors ($r=1$).

One of the chief challenges when dealing with higher order tensors is computation. Although convex, nuclear norm minimization for higher order tensors is computationally expensive in the worst case. See, e.g., Hillar and Lim (2013). Various relaxations and approximate algorithms have been introduced in recent years to alleviate the computational burden associated with evaluating tensor norms. See, e.g., Nie and Wang (2014), Jiang, Ma and Zhang (2015) and references therein. It is of great interest to study how these techniques can be adopted in the context of tensor completion in general, and nuclear norm minimization in particular. More detailed investigation along this direction is beyond the scope of the current work and we hope to report our progress elsewhere in the near future. Nevertheless, our results here may provide valuable guidance along this direction. For example, our analysis suggests that when developing effective approximation algorithms for higher order tensor completion, it could tremendously beneficial to explicitly take incoherence into account.

References

[1] Candès, E.J. and Recht, B. (2008), Exact matrix completion via convex optimization, *Foundations of Computational Mathematics*, **9**, 717-772.

[2] Candès, E.J. and Tao, T. (2009), The power of convex relaxation: Near-optimal matrix completion, *IEEE Transactions on Information Theory*, **56(5)**, 2053-2080.

[3] Gandy, S., Recht, B. and Yamada, I. (2011), Tensor completion and low-n-rank tensor recovery via convex optimization, *Inverse Problems*, **27(2)**, 025010.

[4] Giné, E. and Zinn, J. (1984), Some limit theorems for empirical processes, *The Annals of Probability*, **12(4)**, 929-989.

[5] Gross, D. (2011), Recovering low-rank matrices from few coefficients in any basis, *IEEE Transaction on Information Theory*, **57**, 1548-1566.

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

[7] Jiang, B., Ma, S. and Zhang, S. (2015), Tensor principal component analysis via convex optimization, *Mathematical Programming Series A*, **150 (2)**, 423-457.

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

[9] Kruskal, J. B. (1989), Rank, decomposition, and uniqueness for 3-way and N-way arrays, in *"Multiway data analysis"*, North-Holland, Amsterdam, pp. 7-18.

[10] Li, N. and Li, B. (2010), Tensor completion for on-board compression of hyperspectral images, In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 517-520.

[11] Liu, J., Musialski, P., Wonka, P. and Ye, J. (2009), Tensor completion for estimating missing values in visual data, In *ICCV*, 2114-2121.

[12] Mu, C., Huang, B., Wright, J. and Goldfarb, D. (2013), Square deal: lower bounds and improved relaxations for tensor recovery, arXiv: 1307.5870.

[13] Nie, W. and Wang L. (2014), Semidefinite relaxations for best rank-1 tensor approximations, *SIAM Jounral on Matrix Analysis and Applications*, **35(3)**, 1155-1179.

[14] Nguyen, N., Drineas, P. and Tran, T. (2015), Tensor sparsification via a bound on the spectral norm of random tensors, *Information and Inference*, **4**, 195-229.

[15] Recht, B. (2011), A simpler approach to matrix completion, *Journal of Machine Learning Research*, **12**, 3413-3430.

[16] Semerci, O., Hao, N., Kilmer, M. and Miller, E. (2013), Tensor based formulation and nuclear norm regularizatin for multienergy computed tomography, to appear in *IEEE Transactions on Image Processing.*

[17] Sidiropoulos N.D. and Nion, N. (2010), Tensor algebra and multi-dimensional harmonic retrieval in signal processing for mimo radar, *IEEE Trans. on Signal Processing*, **58(11)**, 5693-5705.

[18] Signoretto, M., De Lathauwer, L. and Suykens, J. (2010), Nuclear norms for tensors and their use for convex multilinear estimation.

[19] Signoretto, M., Van de Plas, R., De Moor, B. and Suykens, J. (2011), Tensor versus matrix completion: A comparison with application to spectral data, *IEEE SPL*, **18(7)**, 403-406.

[20] Tomioka, R., Hayashi, K. and Kashima, H. (2010), Estimation of low-rank tensors via convex optimization, arXiv preprint arXiv:1010.0789.

[21] Tomioka, R., Suzuki, T., Hayashi, K. and Kashima, H. (2011), Statistical performance of convex tensor decomposition, *Advances in Neural Information Processing Systems (NIPS)*, 137.

[22] Tropp, J. (2012), User-friendly tail bounds for sums of random matrices, *Foundations of Computational Mathematics*, **12**, 389-434.

[23] Yuan, M. and Zhang, C-H. (2014), On tensor completion via nuclear norm minimization, to appear in *Foundation of Computational Mathematics*.

[24] Watson, G. A. (1992), Characterization of the subdifferential of some matrix norms, *Linear Algebra Appl.*, **170**, 33-45.

A Proof of Lemma 1

It suffices to prove the lemma for $c = 1$. Consider without loss of generality $\boldsymbol{a}$ and $\boldsymbol{u}$ with nonnegative components, $\|\boldsymbol{u}\|_{\ell_2} = 1$ and $\|\boldsymbol{u}\|_{\ell_\infty} \le \delta$. Let

$$\boldsymbol{v} = (u_1 \vee d^{-1/2}, \ldots, u_d \vee d^{-1/2})^\top/\sqrt{2},$$

where $a \vee b = \max\{a, b\}$. We have

$$\|\boldsymbol{v}\|_{\ell_\infty} \le \delta/\sqrt{2}, \qquad \sqrt{2}\boldsymbol{v}^\top\boldsymbol{a} \ge \boldsymbol{u}^\top\boldsymbol{a},$$

and

$$\|\boldsymbol{v}\|_{\ell_2}^2 = 2^{-1}\sum_{i=1}^d \max(u_i^2, 1/d) \le 1.$$

Let

$$\boldsymbol{w} \in \{2^{j/2}/\sqrt{2d}, j = 0, \ldots, m\}^d \qquad \text{with } \; w_i \le v_i \le \sqrt{2}w_i, \; \forall\, i = 1, \ldots, d.$$

This is possible as

$$\|\boldsymbol{v}\|_{\ell_\infty} \le \delta/\sqrt{2} \le \sqrt{2}(2^{m/2}/\sqrt{2d}).$$

We have

$$\|\boldsymbol{w}\|_{\ell_2} \le \|\boldsymbol{v}\|_{\ell_2} \le 1 \qquad \text{and} \qquad 2\boldsymbol{w}^\top\boldsymbol{a} \ge \sqrt{2}\boldsymbol{v}^\top\boldsymbol{a} \ge \boldsymbol{u}^\top\boldsymbol{a}.$$

It remains to count the cardinality. Let $\ell_j = \lfloor d/(2^j - 1) \rfloor$. For $1 \le j \le m$,

$$(2^j/(2d)) \left| \{ i : w_i^2 = 2^j/(2d) \} \right| + (2d)^{-1} \left[d - \left| \{ i : w_i^2 = 2^j/(2d) \} \right| \right] \le 1,$$

so that

$$|\{ i : w_i^2 = 2^j/(2d) \}| \le \ell_j.$$

As a choice of $\boldsymbol{w}$ can be made by first picking the sign of its elements, the cardinality of the $\boldsymbol{w}$-collection is no greater than

$$N = 2^d \prod_{j=1}^{m} \sum_{0 \le \ell \le \ell_j} \binom{d}{\ell}.$$

Moreover, for $j \ge 2$, we have $\ell_j \le d/(2^j - 1)$, so that

$$\sum_{\ell=1}^{\ell_j} \binom{d}{\ell} \le \binom{d}{\ell_j} \sum_{\ell=0}^{\ell_j} \Big(\frac{1/(2^j-1)}{1 - 1/(2^j - 1)} \Big)^{\ell_j - \ell} \le \binom{d}{\ell_j} \Big(1 + \frac{1}{2^j - 3} \Big).$$

It follows with an application of the Stirling formula that

$$N \le 4^d \exp \left\{ \sum_{j=2}^{m} \Big(\ell_j \log(ed/\ell_j) + \frac{1}{2^j - 3} \Big) \right\}.$$

Since $x(1 + \log(d/x))$ is increasing in x for $0 \le x \le d$ and $\ell_j \le d/(2^j - 1)$,

$$\log N \le d \log 4 + d \sum_{j=2}^{\infty} \frac{1 + \log(2^j - 1)}{2^j - 1} + \sum_{j=2}^{\infty} \frac{1}{2^j - 3} \le 3.082 \times d + 1.344.$$

The proof is now completed.

B Proof of Lemma 2

Let $\boldsymbol{\Delta} = \widehat{\boldsymbol{T}} - \boldsymbol{T}$. Then $\mathcal{P}_\Omega \boldsymbol{\Delta} = 0$ and

$$\|\boldsymbol{T} + \boldsymbol{\Delta}\|_{\star,\boldsymbol{\delta}} \le \|\boldsymbol{T}\|_{\star,\boldsymbol{\delta}}.$$

It follows from Theorem 2 that

$$\|\boldsymbol{T} + \boldsymbol{\Delta}\|_{\star,\boldsymbol{\delta}} \ge \|\boldsymbol{T}\|_{\star,\boldsymbol{\delta}} + \frac{\|\mathcal{Q}_{\boldsymbol{T}}^{\perp} \boldsymbol{\Delta}\|_{\star,\boldsymbol{\delta}}}{k(k-1)/2} + \langle \boldsymbol{W}_0, \boldsymbol{\Delta} \rangle.$$

Because $\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{W}_0 = \boldsymbol{W}_0$ and

$$\langle \widetilde{\boldsymbol{G}}, \boldsymbol{\Delta}\rangle = \langle \mathcal{P}_\Omega \widetilde{\boldsymbol{G}}, \boldsymbol{\Delta}\rangle = \langle \widetilde{\boldsymbol{G}}, \mathcal{P}_\Omega \boldsymbol{\Delta}\rangle = 0$$

we get

$$\begin{aligned} -\frac{\|\mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{\Delta}\|_{\star,\boldsymbol{\delta}}}{k(k-1)/2} &\geq \langle \boldsymbol{W}_0 - \widetilde{\boldsymbol{G}}, \boldsymbol{\Delta}\rangle \\ &= \langle \mathcal{Q}_{\boldsymbol{T}}(\boldsymbol{W}_0 - \widetilde{\boldsymbol{G}}), \boldsymbol{\Delta}\rangle - \langle \widetilde{\boldsymbol{G}}, \mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{\Delta}\rangle \\ &\geq -\|\boldsymbol{W}_0 - \mathcal{Q}_{\boldsymbol{T}}\widetilde{\boldsymbol{G}}\|_{\mathrm{HS}}\|\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{\Delta}\|_{\mathrm{HS}} - \|\mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{\Delta}\|_{\star,\boldsymbol{\delta}}/\{k(k-1)\}. \end{aligned}$$

It follows that

$$\|\mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{\Delta}\|_{\star,\boldsymbol{\delta}}/\{k(k-1)\} \leq \|\boldsymbol{W}_0 - \mathcal{Q}_{\boldsymbol{T}}\widetilde{\boldsymbol{G}}\|_{\mathrm{HS}}\|\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{\Delta}\|_{\mathrm{HS}}.$$

Recall that

$$\mathcal{P}_\Omega\boldsymbol{\Delta} = \mathcal{P}_\Omega\mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{\Delta} + \mathcal{P}_\Omega\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{\Delta} = 0.$$

Thus, in view of (13) and Proposition 1

$$\frac{\|\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{\Delta}\|_{\mathrm{HS}}}{\sqrt{2d_*^k/n}} \leq \|\mathcal{P}_\Omega\mathcal{Q}_{\boldsymbol{T}}\boldsymbol{\Delta}\|_{\mathrm{HS}} = \|\mathcal{P}_\Omega\mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{\Delta}\|_{\mathrm{HS}} \leq \|\mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{\Delta}\|_{\mathrm{HS}} \leq \|\mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{\Delta}\|_{\star,\boldsymbol{\delta}}. \tag{22}$$

Consequently,

$$\frac{\|\mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{\Delta}\|_{\star,\boldsymbol{\delta}}}{k(k-1)} \leq \sqrt{2d_*^k/n}\|\boldsymbol{W}_0 - \mathcal{Q}_{\boldsymbol{T}}\widetilde{\boldsymbol{G}}\|_{\mathrm{HS}}\|\mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{\Delta}\|_{\star,\boldsymbol{\delta}}.$$

Because of (11), we have $\|\mathcal{Q}_{\boldsymbol{T}}^{\perp}\boldsymbol{\Delta}\|_{\star,\boldsymbol{\delta}} = 0$. Together with (22), we conclude that $\boldsymbol{\Delta} = 0$, or equivalently $\widehat{\boldsymbol{T}} = \boldsymbol{T}$.