

Guarantees of Augmented Trace Norm Models in Tensor Recovery

Ziqiang Shi^{1*}, Jiqing Han¹, Tieran Zheng¹, Ji Li²

¹Harbin Institute of Technology, Harbin, China

²Beijing GuoDianTong Network Technology, Beijing, China

shiziqiang7@gmail.com; shiziqiang@cn.fujitsu.com

Abstract

This paper studies the recovery guarantees of the models of minimizing $\|\mathcal{X}\|_* + \frac{1}{2\alpha}\|\mathcal{X}\|_F^2$ where \mathcal{X} is a tensor and $\|\mathcal{X}\|_*$ and $\|\mathcal{X}\|_F$ are the trace and Frobenius norm of respectively. We show that they can efficiently recover low-rank tensors. In particular, they enjoy exact guarantees similar to those known for minimizing $\|\mathcal{X}\|_*$ under the conditions on the sensing operator such as its null-space property, restricted isometry property, or spherical section property. To recover a low-rank tensor \mathcal{X}^0 , minimizing $\|\mathcal{X}\|_* + \frac{1}{2\alpha}\|\mathcal{X}\|_F^2$ returns the same solution as minimizing $\|\mathcal{X}\|_*$ almost whenever $\alpha \geq 10 \max_i \|X_{(i)}^0\|_2$.

1 Introduction

Low-rank tensor recovering problem is the generalization of sparse vector recovery and low-rank matrix recover to tensor data [Gandy *et al.*, 2011; Tomioka *et al.*, 2010; Liu *et al.*, 2009]. It has drawn lots of attention from researchers in different fields in the past several years. It is not only a simple mathematical generalization of sparse vector recovery and low-rank matrix recover. They have wide applications in science and engineering, e.g. multimedia signal processing (compression, image in-painting, BRDF data estimation), machine learning (collaborative filtering), and bioinformatics (EEG signal analysis). The fundamental problem of low-rank tensor recovery is to find a tensor of (nearly) lowest rank from an underdetermined $\mathfrak{F}(\mathcal{X}) = b$, where \mathfrak{F} is a linear operator and $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is a N -way tensor. Although researchers have proposed several algorithms to solve the problem, there're only a few results about the theory part, that is under what conditions the underline tensor data will be recovered exactly.

To recover a low-rank tensor $\mathcal{X}^0 \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ from linear measurements $b = \mathfrak{F}(\mathcal{X}^0)$, a powerful approach is the convex model [Gandy *et al.*, 2011; Liu *et al.*, 2009]

$$\min_{\mathcal{X}} \{\|\mathcal{X}\|_* : \mathfrak{F}(\mathcal{X}) = b\}, \quad (1)$$

*Ziqiang Shi is, now, with the Fujitsu Research & Development Center, Beijing, China.

where

$$\|\mathcal{X}\|_* := \frac{1}{N} \sum_{i=1}^N \|X_{(i)}\|_* \quad (2)$$

and $X_{(i)}$ is the mode- i unfolding of \mathcal{X} , $\|X_{(i)}\|_*$ is the trace norm of the matrix $X_{(i)}$, i.e. the sum of the singular values of $X_{(i)}$. For vector b with noise or generated by an approximately low-rank tensor, a variant of (1) is [Liu *et al.*, 2009]

$$\min_{\mathcal{X}} \{\|\mathcal{X}\|_* : \|\mathfrak{F}(\mathcal{X}) - b\|_2 \leq \sigma\}. \quad (3)$$

Despite empirical success, the recovery guarantees of tensor recovery algorithms has not been fully elucidated. Recently, several authors [Lai and Yin, 2012; Recht *et al.*, 2007; Mo and Li, 2011] have got excellent results in the guarantees of sparse vector and low-rank matrix recovery. In this paper, we try to generalize these results to low-rank tensor recovery.

This paper mainly studies the guarantees of minimization of the augmented objective $\|\mathcal{X}\|_* + \frac{1}{2\alpha}\|\mathcal{X}\|_F^2$. The augmented model for (1) and (3) are

$$\min_{\mathcal{X}} \{\|\mathcal{X}\|_* + \frac{1}{2\alpha}\|\mathcal{X}\|_F^2 : \mathfrak{F}(\mathcal{X}) = b\}. \quad (4)$$

and

$$\min_{\mathcal{X}} \{\|\mathcal{X}\|_* + \frac{1}{2\alpha}\|\mathcal{X}\|_F^2 : \|\mathfrak{F}(\mathcal{X}) - b\|_2 \leq \sigma\}. \quad (5)$$

respectively. These are natural generalizations of the augmented model for vector and matrix data [Lai and Yin, 2012] to tensor case.

Tomioka *et al.* [Tomioka *et al.*, 2010] analyzed the statistical performance of a tensor decomposition algorithm in the setting of random Gaussian design. But they didn't give results about general linear operator \mathfrak{F} . Our work gives results about that the general linear operator \mathfrak{F} , that is under what kind of the null-space property, the restricted isometry property (RIP), and the spherical section property, the tensor can be exactly recovered. And furthermore we give results both for the original and augmented model. To our best knowledge, this is the first work to extend the RIP, the null-space property, and the spherical section property based results from the sparse vector recovery to tensor case. At last we should say this is not a trivial extension, because of the dependency among multiple constraints in different dimensions of tensor data, and we only obtain sufficient conditions

for the exact tensor recovery, while for the matrix case, the conditions are also necessary. The sufficient and necessary condition is a challenge.

2 Notations

We adopt the nomenclature mainly used by Kolda and Bader on tensor decompositions and applications [Kolda and Bader, 2009], and also a few symbols of [De Lathauwer *et al.*, 2000; Tomioka *et al.*, 2011].

The *order* N of a tensor is the number of dimensions, also known as ways or modes. Matrices (tensor of order two) are denoted by upper case letters, e.g. X , and lower case letters for the elements, e.g. x_{ij} . Higher-order tensors (order three or higher) are denoted by Euler script letters, e.g. \mathcal{X} , and element (i_1, i_2, \dots, i_N) of a N -order tensor \mathcal{X} is denoted by $x_{i_1 i_2 \dots i_N}$. *Fibers* are the higher-order analogue of matrix rows and columns. A fiber is defined by fixing every index but one. The mode- n fibers are all vectors $x_{i_1 \dots i_{n-1} i_{n+1} \dots i_N}$ that obtained by fixing the values of $\{i_1, i_2, \dots, i_N\} \setminus i_n$. The mode- n *unfolding*, also known as *matricization*, of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is denoted by $X_{(n)}$ and arranges the mode- n fibers to be the columns of the resulting matrix. The unfolding operator is denoted as $\text{unfold}(\cdot)$. The opposite operation is $\text{refold}(\cdot)$, denotes the refolding of the matrix into a tensor. The tensor element (i_1, i_2, \dots, i_N) is mapped to the matrix element (i_n, j) , where

$$j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^N (i_k - 1) J_k \quad \text{with} \quad J_k = \prod_{\substack{m=1 \\ m \neq n}}^{k-1} I_m$$

Therefore, $X_{(n)} \in \mathbb{R}^{I_n \times I_1 \dots I_{n-1} I_{n+1} \dots I_N}$. The n -rank of a N -dimensional tensor \mathcal{X} , denoted as $\text{rank}_n(\mathcal{X})$ is the column rank of $X_{(n)}$, i.e. the dimension of the vector space spanned by the mode- n fibers. We say a tensor \mathcal{X} is rank (r_1, \dots, r_N) when $r_k = \text{rank}_k(\mathcal{X})$ for $k = 1, \dots, N$, and denoted as $\text{rank}(\mathcal{X})$. We introduce an ordering among tensors by the rank: $\text{rank}(\mathcal{X}) \leq \text{rank}(\mathcal{Y}) \Leftrightarrow (\text{rank}_1(\mathcal{X}), \dots, \text{rank}_N(\mathcal{X})) \preceq (\text{rank}_1(\mathcal{Y}), \dots, \text{rank}_N(\mathcal{Y})) \Leftrightarrow \text{rank}_i(\mathcal{X}) \leq \text{rank}_i(\mathcal{Y}), i = 1, \dots, N$. The inner product of two same-size tensors $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is defined as $\text{vec}(\mathcal{X})^\top \cdot \text{vec}(\mathcal{Y})$, where vec is a vectorization. The corresponding norm is $\|\mathcal{X}\|_F = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$, which is often called the Frobenius norm.

The n -th mode product of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ with a matrix $U \in \mathbb{R}^{J \times I_n}$ is denoted by $\mathcal{X} \times_n U$ and is of size $I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N$. Elementwise, we have

$$(\mathcal{X} \times_n U)_{i_1 \dots i_{n-1} j i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \dots i_N} U_{j i_n}. \quad (6)$$

Every tensor can be written as the product [De Lathauwer *et al.*, 2000]

$$\mathcal{A} = \mathcal{S} \times_1 U^{(1)} \times_2 U^{(2)} \dots \times_N U^{(N)}, \quad (7)$$

in which:

- $U^{(n)}$ is a unitary $I_n \times I_n$ matrix,

- \mathcal{S} is a $I_1 \times I_2 \times \dots \times I_N$ -tensor of which the subtensors $\mathcal{S}_{i_n=\alpha}$, obtained by fixing the n -th index to α , have the properties of:

1) all-orthogonality: two subtensors $\mathcal{S}_{i_n=\alpha}$ and $\mathcal{S}_{i_n=\beta}$ are orthogonal for all possible values of n, α and β subject to $\alpha \neq \beta$: $\langle \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} \rangle = 0$,

2) ordering: for all possible values of n , one has $\|\mathcal{S}_{i_n=1}\|_F \geq \|\mathcal{S}_{i_n=2}\|_F \geq \dots \geq \|\mathcal{S}_{i_n=I_n}\|_F \geq 0$.

The Frobenius norms $\|\mathcal{S}_{i_n=i}\|_F$, symbolized by $\sigma_i^{(n)}$, are mode- n singular values of \mathcal{A} , that means the singular values of $A_{(n)}$.

This is called the higher-order singular value decomposition (HOSVD) of a tensor \mathcal{A} in [De Lathauwer *et al.*, 2000]. Some properties of this HOSVD which will be used in this paper are listed below as lemmas:

Lemma 1. ([De Lathauwer *et al.*, 2000] Property 6). *Let the HOSVD of \mathcal{A} be given as in (7), and let r_n be equal to the highest index for which $\|\mathcal{S}_{i_n=r_n}\|_F > 0$; then one has $\text{rank}(A_{(n)}) = r_n$.*

Lemma 2. ([De Lathauwer *et al.*, 2000] Property 8). *Let the HOSVD of \mathcal{A} be given as in (7); due to the unitarily invariant of the Frobenius norm, one has $\|\mathcal{A}\|_F^2 = \sum_{i=1}^{I_1} (\sigma_i^{(1)})^2 = \dots = \sum_{i=1}^{I_N} (\sigma_i^{(N)})^2 = \|\mathcal{S}\|_F^2$.*

3 Motivations and contributions

To explain why model (4) is interesting, we conducted following tensor completion simulations

$$\min_{\mathcal{X}} \left\{ \|\mathcal{X}\|_* + \frac{1}{2\alpha} \|\mathcal{X}\|_F^2 : x_{i_1 i_2 \dots i_N} = m_{i_1 i_2 \dots i_N}, i_1 i_2 \dots i_N \in \Omega \right\}. \quad (8)$$

to compare it with model (1) based tensor completion

$$\min_{\mathcal{X}} \left\{ \|\mathcal{X}\|_* : x_{i_1 i_2 \dots i_N} = m_{i_1 i_2 \dots i_N}, i_1 i_2 \dots i_N \in \Omega \right\}. \quad (9)$$

The facade image data of [Liu *et al.*, 2009] was used here to be an example. For color image has RGB values in each pixel, so it is indeed a 3-way tensor. Models (8) and (9) were solved to high accuracy by the solver LRTC [Liu *et al.*, 2009]. For each model, we measured and recorded

$$\text{recovery relative error} : \|\mathcal{X}^* - \mathcal{X}_0\|_F / \|\mathcal{X}_0\|_F. \quad (10)$$

The relative errors are depicted as functions of the number of iterations in Figure 1(a).

Motivated by the above example, we show in this paper that any $\alpha \geq 10 \max_i \|X_{(i)}^0\|_2$ guarantees that problem (4) either recovers \mathcal{X}^0 exactly or returns an approximate of it nearly as good as the solution of problem (1). Specifically, we show that several properties of \mathfrak{F} , such as the null-space property (a simple condition used in, e.g., [Liu *et al.*, 2009; Donoho and Huo, 2001; Gribonval and Nielsen, 2003; Zhang, 2005; Oymak *et al.*, 2011]), the restricted isometry principle [Candes and Tao, 2005], and the spherical section property [Zhang, 2008], which have been used in the recovery

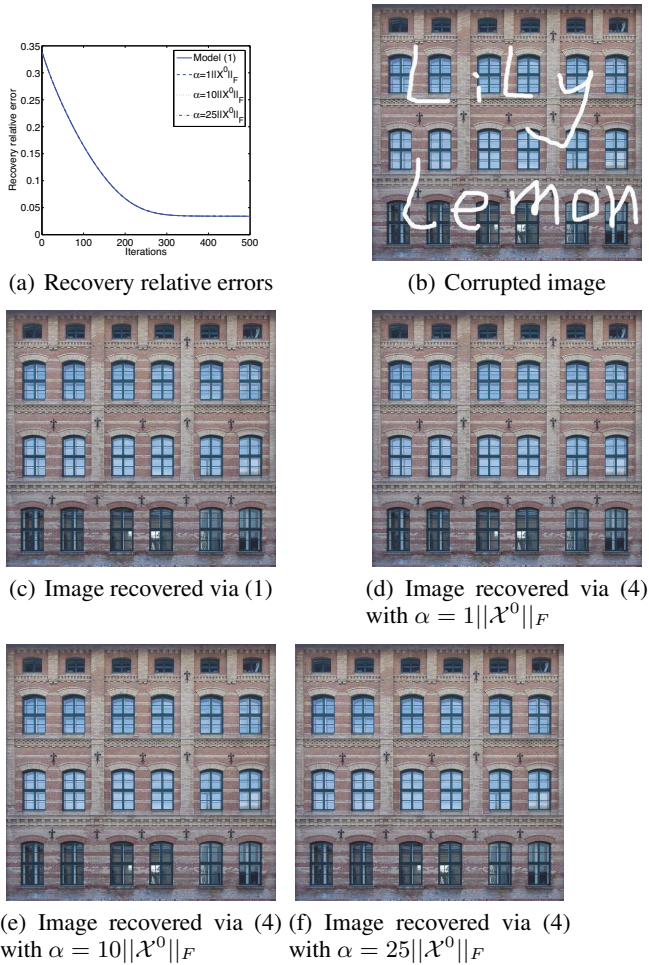


Figure 1: Facade in-painting.

guarantees for vectors and matrices, can also guarantee the tensor recovery by model (1) and (4).

Even though \mathcal{X}^0 not known when α is set, $\max_i \|X_{(i)}^0\|_2$ is often easy to estimate. When $\max_i \|X_{(i)}^0\|_2$ is not available, using inequalities $\|\mathcal{X}^0\|_F \geq \max_i \|X_{(i)}^0\|_2$, one gets the more conservative formulae $\alpha \geq 10\|\mathcal{X}^0\|_F$. Furthermore, when \mathfrak{F} satisfies the RIP, one has $\|\mathcal{X}^0\|_F \leq \beta\|b\|_2$ for some $\beta > 0$; hence, one has the option to use the even more conservative formula $\alpha \geq 10\beta\|b\|_2$.

4 Tensor recovery guarantees

This section establishes recovery guarantees for the original and augmented trace norm models (1) and (4). The results are given based on the properties of \mathfrak{F} including the null-space property (NSP) in Theorem 5 and 6, the restricted isometry principle (RIP) [Candes and Tao, 2005] in Theorem 8 and 9, the spherical section property (SSP) [Zhang, 2008] in Theorem 10. These results adapt and generalize of the work in [Lai and Yin, 2012].

4.1 Null space property

The wide use of NSP for recovering sparse vector and low-rank matrices can be found in e.g. [Donoho and Huo, 2001; Gribonval and Nielsen, 2003; Zhang, 2005; Oymak *et al.*, 2011]. In this subsection, we extend the NSP conditions on \mathfrak{F} for tensor recovery. Throughout this subsection, we let $\sigma_i(X), i = 1, \dots, m$ denote the i -th largest singular value of matrix X of rank m or less, and $\Sigma(X) = \text{diag}(\sigma_1(X), \dots, \sigma_m(X))$ denote the diagonal matrix of singular values and $s(X) = (\sigma_1(X), \dots, \sigma_m(X))$. $\|X\|_2 = \sigma_1(X)$ denotes the spectral norms of X .

We will need the following two technical lemmas for the introduction of the tensor NSP conditions.

Lemma 3. ([Horn and Johnson, 1990] Theorem 7.4.51). *Let X and Y be two matrices of the same size. Then we have*

$$\sum_{i=1}^m |\sigma_i(X) - \sigma_i(Y)| \leq \|X - Y\|_* \quad (11)$$

and

$$\sum_{i=1}^m (\sigma_i(X) - \sigma_i(Y))^2 \leq \|X - Y\|_F^2. \quad (12)$$

Lemma 4. ([Lai and Yin, 2012] Equation (19)). *Let x and h be two vectors of the same size, $\mathcal{S} := \text{supp}(x)$ and $\mathcal{Z} = \mathcal{S}^c$. Here \mathcal{S} is the index set of the nonzero support of x , and \mathcal{S}^c is the complement of \mathcal{S} . Then we have*

$$\|x + h\|_1 \geq \|x\|_1 + \|h_{\mathcal{Z}}\|_1 - \|h_{\mathcal{S}}\|_1 \quad (13)$$

and

$$\begin{aligned} \|x + h\|_1 + \frac{1}{2\alpha} \|x + h\|_2^2 &\geq \left[\|x\|_1 + \frac{1}{2\alpha} \|x\|_2^2 \right] \\ &+ \left[\|h_{\mathcal{Z}}\|_1 - \left(1 + \frac{\|x_{\mathcal{S}}\|_{\infty}}{\alpha}\right) \|h_{\mathcal{S}}\|_1 \right] + \frac{1}{2\alpha} \|h\|_2^2. \end{aligned} \quad (14)$$

Now we give an NSP type sufficient condition for problem (1).

Theorem 5. (Tensor NSP condition for (1)). *Assume \mathcal{X}^0 is fixed, problem (1) uniquely recovers all tensors \mathcal{X}^0 of rank (r_1, \dots, r_N) or less from the measurements $\mathfrak{F}(\mathcal{X}^0) = b$, if all $\mathcal{H} \in \text{Null}(\mathfrak{F}) \setminus \{0\}$ satisfy*

$$\sum_{i=1}^N \sum_{j=r_i+1}^{I_i} \sigma_j(H_{(i)}) > \sum_{i=1}^N \sum_{j=1}^{r_i} \sigma_j(H_{(i)}). \quad (15)$$

We can extend this result to problem (4) as follows.

Theorem 6. (Tensor NSP condition for (4)). *Assume \mathcal{X}^0 is fixed, problem (4) uniquely recovers all tensors \mathcal{X}^0 of rank (r_1, \dots, r_N) or less from the measurements $\mathfrak{F}(\mathcal{X}^0) = b$, if all $\mathcal{H} \in \text{Null}(\mathfrak{F}) \setminus \{0\}$ satisfy*

$$\sum_{i=1}^N \sum_{j=r_i+1}^{I_i} \sigma_j(H_{(i)}) \geq \sum_{i=1}^N \left(1 + \frac{\|X_{(i)}^0\|_2}{\alpha}\right) \sum_{j=1}^{r_i} \sigma_j(H_{(i)}). \quad (16)$$

Remark 1. For any finite $\alpha > 0$, (16) is stronger than (15) due to the extra term $\frac{\|X_{(i)}^0\|_2}{\alpha}$. Since various uniform recovery results establish conditions that guarantee (15), one can tighten these conditions so that they guarantee (16) and thus the uniform recovery by problem (4). How much tighter these conditions have to be depends on the value $\frac{\|X_{(i)}^0\|_2}{\alpha}$.

4.2 Tensor restricted isometry principle

In this subsection, we generalize the RIP-based guarantees to tensor case and show that any $\alpha \geq 10 \max_i \|X_{(i)}^0\|_2$ guarantees exact recovery by (4).

Definition 1. (Tensor RIP). Let $\mathfrak{M}_{(r_1, r_2, \dots, r_N)} := \{\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N} : \text{rank}(X_{(n)}) \leq r_n, n = 1, \dots, N\}$. The RIP constant $\delta_{(r_1, r_2, \dots, r_N)}$ of linear operator \mathfrak{F} is the smallest value such that

$$\begin{aligned} (1 - \delta_{(r_1, r_2, \dots, r_N)}) \|\mathcal{X}\|_F^2 &\leq \|\mathfrak{F}(\mathcal{X})\|_2^2 \\ &\leq (1 + \delta_{(r_1, r_2, \dots, r_N)}) \|\mathcal{X}\|_F^2 \end{aligned} \quad (17)$$

holds for all $\mathcal{X} \in \mathfrak{M}_{(r_1, r_2, \dots, r_N)}$.

The following recovery theorems will characterize the power of the tensor restricted isometry constants. The first theorem generalizes Lemma 1.3 in [Candes and Tao, 2005] and Theorem 3.2 in [Recht et al., 2007] to low-rank tensor recovery.

Theorem 7. Suppose $\delta_{(2r_1, 2r_2, \dots, 2r_N)} < 1$ for some $(r_1, r_2, \dots, r_N) \succeq (1, 1, \dots, 1)$. Then \mathcal{X}^0 is the only tensor of rank at most (r_1, r_2, \dots, r_N) satisfying $\mathfrak{F}(\mathcal{X}) = b$.

The proof of the preceding theorem is identical to the argument given by Candes and Tao and is an immediate consequence of our definition of the constant $\delta_{(r_1, r_2, \dots, r_N)}$. No adjustment is necessary in the transition from sparse vectors and low-rank matrices to low-rank tensors. The key property used is the sub-additivity of the rank. Adapting results in [Lai and Yin, 2012; Mo and Li, 2011], we give the uniform recovery conditions for (1) below.

Theorem 8. (RIP condition for exact recovery by (1)). Let \mathcal{X}^0 be a tensor with rank (r_1, r_2, \dots, r_N) or less. Problem (1) exactly recovers \mathcal{X}^0 from measurements $b = \mathfrak{F}(\mathcal{X}^0)$ if \mathfrak{F} satisfies the RIP with $\delta_{(I_1, \dots, 2r_n, \dots, I_N)} < 0.4931$, for $n = 1, \dots, N$.

Next we carry out a similar study for the augmented model (4).

Theorem 9. (RIP condition for exact recovery by (4)). Let \mathcal{X}^0 be a tensor with rank (r_1, r_2, \dots, r_N) or less. The augmented model (3) exactly recovers \mathcal{X}^0 from measurements $b = \mathfrak{F}(\mathcal{X}^0)$ if \mathfrak{F} satisfies the RIP with $\delta_{(I_1, \dots, 2r_n, \dots, I_N)} < 0.4404$, $n = 1, \dots, N$ and $\alpha \geq 10 \max_i \|X_{(i)}^0\|_2$.

Remark 2. Different values of $\delta_{(I_1, \dots, 2r_n, \dots, I_N)}$, $n = 1, \dots, N$ are associated with different conditions on α . Following (34), if $\delta_{(I_1, \dots, 2r_n, \dots, I_N)} < 0.4715$, $n = 1, \dots, N$, $\alpha \geq 10 \max_i \|X_{(i)}^0\|_2$ guarantees exact recovery. If $\delta_{(I_1, \dots, 2r_n, \dots, I_N)} < 0.1273$, $n = 1, \dots, N$, $\alpha \geq 10 \max_i \|X_{(i)}^0\|_2$ guarantees exact recovery. In general, smaller $\delta_{(I_1, \dots, 2r_n, \dots, I_N)}$, $n = 1, \dots, N$ allows a smaller α .

4.3 Spherical section property

Next, we derive exact conditions based on the spherical section property (SSP) [Lai and Yin, 2012; Zhang, 2008]. There is not much discussion on spherical section property (SSP) for low-rank tensor recovery in the literature, here we present an SSP-based result.

Theorem 10. (SSP condition for exact recovery by (4)). Let $\mathfrak{F} : \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N} \rightarrow \mathbb{R}^m$ be a linear operator. Suppose there exists $\Delta > 0$ such that all nonzero $\mathcal{H} \in \text{Null}(\mathfrak{F})$ satisfy

$$\frac{\|\mathcal{H}\|_*}{\|\mathcal{H}\|_F} \geq \sqrt{\frac{m}{\Delta}}. \quad (18)$$

Assume that $\|X_{(i)}^0\|_2$, $(i = 1, \dots, N)$ and $\alpha > 0$ are fixed. If

$$m \geq (2 + \frac{\|X_{(i)}^0\|_2}{\alpha})^2 r_i \Delta, (i = 1, \dots, N), \quad (19)$$

then the null-space condition holds for all nonzero $\mathcal{H} \in \text{Null}(\mathfrak{F})$. Hence is sufficient for problem to recover any \mathcal{X}^0 with rank (r_1, r_2, \dots, r_N) or less from measurements $b = \mathfrak{F}(\mathcal{X}^0)$.

5 Conclusion

In this work we focussed on the recovery guarantees of tensor recovery via convex optimization. We presented general results stating that the extension of some sufficient conditions for the recovery of low-rank matrices using nuclear norm minimization is also sufficient for the recovery of low-rank tensors using tensor nuclear norm minimization. We extended the null-space property, the restricted isometry principle, and the spherical section property conditions to the augmented tensor recovery problems, and find that any $\alpha \geq 10 \max_i \|X_{(i)}^0\|_2$ guarantees that problem (4) either recovers \mathcal{X}^0 exactly or returns an approximate of it nearly as good as the solution of problem (1).

There are some directions that the current study can be extended. In this paper, we have focused on the recovery guarantees of the exact case; it would be meaningful to also analyze the guarantees for the stable recovery. Second, generalize the linearized Bregman algorithm [Yin et al., 2008] for augmented sparse vector and low-rank matrix recovery [Lai and Yin, 2012] to low-rank tensor case. Moreover, from our results, there is a big ‘‘gap’’ between the recovery conditions for matrices and tensors, and we need to fill this ‘‘gap’’ in future work. In a broader context, we believe that the current paper could serve as a basis for examining the augmented trace norm models in tensor recovery.

Appendix

Proof of Theorem 5.

Proof. Pick any tensor \mathcal{X}^0 of rank (r_1, \dots, r_N) or less and let $b = \mathfrak{F}(\mathcal{X}^0)$. For any $\mathcal{H} \in \text{Null}(\mathfrak{F}) \setminus \{0\}$, we have $\mathfrak{F}(\mathcal{X}^0 +$

$\mathcal{H}) = \mathfrak{F}(\mathcal{X}^0) = b$. By using (11), we have

$$\begin{aligned}
\|\mathcal{X}^0 + \mathcal{H}\|_* &= \frac{1}{N} \sum_{i=1}^N \|X_{(i)}^0 + H_{(i)}\|_* \\
&\geq \frac{1}{N} \sum_{i=1}^N \|s(X_{(i)}^0) - s(H_{(i)})\|_1 \\
&\geq \frac{1}{N} \sum_{i=1}^N (\|X_{(i)}^0\|_* + [\sum_{j=r_i+1}^{I_i} \sigma_j(H_{(i)}) - \sum_{j=1}^{r_i} \sigma_j(H_{(i)})]) \\
&= \|\mathcal{X}^0\|_* + \frac{1}{N} \sum_{i=1}^N (\sum_{j=r_i+1}^{I_i} \sigma_j(H_{(i)}) - \sum_{j=1}^{r_i} \sigma_j(H_{(i)}))
\end{aligned} \tag{20}$$

where the first inequality follows from (11). For any nonzero $\mathcal{H} \in \text{Null}(\mathfrak{F})$, $\|\mathcal{H}\|_F^2 > 0$. Hence from (15) and (20), it follows that \mathcal{X}^0 is a unique minimizer of (1). \square

Proof of Theorem 6.

Proof. Pick any tensor \mathcal{X}^0 of rank (r_1, \dots, r_N) or less and let $b = \mathfrak{F}(\mathcal{X}^0)$. For any nonzero $\mathcal{H} \in \text{Null}(\mathfrak{F})$, we have $\mathfrak{F}(\mathcal{X}^0 + \mathcal{H}) = \mathfrak{F}(\mathcal{X}^0) = b$. Thus

$$\begin{aligned}
&\|\mathcal{X}^0 + \mathcal{H}\|_* + \frac{1}{2\alpha} \|\mathcal{X}^0 + \mathcal{H}\|_F^2 \\
&= \frac{1}{N} \sum_{i=1}^N (\|X_{(i)}^0 + H_{(i)}\|_* + \frac{1}{2\alpha} \|X_{(i)}^0 + H_{(i)}\|_F^2) \\
&\geq \frac{1}{N} \sum_{i=1}^N (\|s(X_{(i)}^0) - s(H_{(i)})\|_1 + \frac{1}{2\alpha} \|s(X_{(i)}^0) - s(H_{(i)})\|_2^2) \\
&\geq \frac{1}{N} \sum_{i=1}^N (\|X_{(i)}^0\|_* + \frac{1}{2\alpha} \|X_{(i)}^0\|_F^2) \\
&+ [\sum_{j=r_i+1}^{I_i} \sigma_j(H_{(i)}) - (1 + \frac{\|X_{(i)}^0\|_2}{\alpha}) \sum_{j=1}^{r_i} \sigma_j(H_{(i)})] + \frac{1}{2\alpha} \|H_{(i)}\|_F^2 \\
&= \|\mathcal{X}^0\|_* + \frac{1}{2\alpha} \|\mathcal{X}^0\|_F^2 + \frac{1}{2\alpha} \|\mathcal{H}\|_F^2 \\
&+ \frac{1}{N} \sum_{i=1}^N (\sum_{j=r_i+1}^{I_i} \sigma_j(H_{(i)}) - (1 + \frac{\|X_{(i)}^0\|_2}{\alpha}) \sum_{j=1}^{r_i} \sigma_j(H_{(i)}))
\end{aligned} \tag{21}$$

where the first inequality follows from (11) and (12), and the second inequality follows from (13) and (14). For any nonzero $\mathcal{H} \in \text{Null}(\mathfrak{F})$, $\|\mathcal{H}\|_F^2 > 0$. Hence, from (21) and (16), it follows that $\mathcal{X}^0 + \mathcal{H}$ leads to a strictly worse objective than \mathcal{X}^0 . That is, \mathcal{X}^0 is the unique solution to problem (4). \square

Proof of Theorem 7.

Proof. Assume, on the contrary, that there exists a tensor with rank (r_1, r_2, \dots, r_N) or less satisfying $\mathfrak{F}(\mathcal{X}) = b$ and $\mathcal{X} \neq \mathcal{X}_0$. Then $\mathcal{Z} := \mathcal{X} - \mathcal{X}_0$ is a nonzero tensor of rank at most

$(2r_1, 2r_2, \dots, 2r_N)$, and $\mathfrak{F}(\mathcal{Z}) = 0$. But then we would have $0 = \|\mathfrak{F}(\mathcal{Z})\|_2^2 \geq (1 - \delta_{(2r_1, 2r_2, \dots, 2r_N)}) \|\mathcal{Z}\|_F^2 > 0$ which is a contradiction. \square

Proof of Theorem 8.

Proof. For any nonzero $\mathcal{H} \in \text{Null}(\mathfrak{F})$, $\|\mathcal{H}\|_F > 0$, let $\mathcal{H} = \mathcal{S} \times_1 U^{(1)} \times_2 U^{(2)} \dots \times_N U^{(N)}$ be the HOSVD of \mathcal{H} . From Proposition 3.7 of [Kolda, 2006] we have $\text{rank}(H_{(n)}) \leq \text{rank}(S_{(n)})$, $(n = 1, \dots, N)$, thus we have $\text{rank}(\mathcal{H}) \preceq \text{rank}(\mathcal{S})$. We decompose $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1 + \dots$, where

$$\begin{aligned}
\mathcal{H}_0 &= \mathcal{S}_{\{i_n:1, \dots, r_n\}} \times_1 U^{(1)} \times_2 U^{(2)} \dots \times_N U^{(N)}, \\
\mathcal{H}_1 &= \mathcal{S}_{\{i_n:r_n+1, \dots, 2r_n\}} \times_1 U^{(1)} \times_2 U^{(2)} \dots \times_N U^{(N)}, \dots
\end{aligned} \tag{22}$$

and $\mathcal{S}_{\{i_n:\alpha\}}$ is the tensor obtained by fixing the n -th index to the index set α , others to zero. Similarly we have $\text{rank}(\mathcal{H})_i \leq \text{rank}(\mathcal{S}_{\{i_n:i r_n+1, \dots, (i+1)r_n\}}) \preceq (I_1, \dots, r_n, \dots, I_n)$, $(i = 0, 1, \dots)$. Due to the unitarily invariant of the Frobenius norm, we have $\|\mathcal{H}_0\|_F = \|\mathcal{S}_{\{i_n:1, \dots, r_n\}}\|_F$, $\|\mathcal{H}_1\|_F = \|\mathcal{S}_{\{i_n:r_n+1, \dots, 2r_n\}}\|_F$, Let $h = (\sigma_1^{(n)}, \sigma_2^{(n)}, \dots, \sigma_{I_n}^{(n)})$ and $h_0 = (\sigma_1^{(n)}, \dots, \sigma_{r_n}^{(n)})$, $h_1 = (\sigma_{r_n+1}^{(n)}, \dots, \sigma_{2r_n}^{(n)})$, $h_2 = (\sigma_{2r_n+1}^{(n)}, \dots, \sigma_{3r_n}^{(n)})$, ..., where $\sigma_i^{(n)}$ is the i -th largest mode- n singular value. From the definition of HOSVD and Lemma 2, we have

$$\begin{aligned}
\|\mathcal{H}_i\|_F^2 &= \|\mathcal{S}_{\{i_n:i r_n+1, \dots, (i+1)r_n\}}\|_F^2 \\
&= \sum_{j=i r_n+1}^{(i+1)r_n} (\sigma_j^{(n)})^2 = \|h_i\|_2^2, (i = 0, 1, \dots).
\end{aligned} \tag{23}$$

Due to the mean-inequation, one has $\|h_i\|_2^2 \geq \|h_i\|_1^2 / r_n$, $(i = 0, 1, \dots)$. Assume that $\|h_1\|_1 = t(\sum_{i \geq 1} \|h_i\|_1)$ with some $t \in [0, 1]$. Then we have $(\sum_{i \geq 2} \|h_i\|_1) = (1 - t)(\sum_{i \geq 1} \|h_i\|_1)$.

First from Lemma 2.1, Lemma 2.2 in [Mo and Li, 2011] and (22) we have

$$\sum_{i \geq 2} \|\mathcal{H}_i\|_F^2 \leq \sigma_{2r_n+1}^{(n)} \sum_{i \geq 2} \|h_i\|_1 \leq \frac{t}{r_n} (1 - t) (\sum_{i \geq 1} \|h_i\|_1)^2 \tag{24}$$

and

$$r_n^{\frac{1}{2}} \|\mathcal{H}_i\|_F \leq \|h_i\|_1 + r_n (\sigma_{i r_n+1}^{(n)} - \sigma_{i r_n+1+r_n}^{(n)}) / 4, \tag{25}$$

$i = 2, 3, \dots$

From (25) we have

$$\begin{aligned}
r_n^{\frac{1}{2}} \sum_{i \geq 2} \|\mathcal{H}_i\|_F &\leq \sum_{i \geq 2} \|h_i\|_1 + r_n |d_{2r_n+1}| / 4 \\
&\leq \sum_{i \geq 2} \|h_i\|_1 + \|h_1\|_1 / 4 = (1 - 3t/4) \sum_{i \geq 1} \|h_i\|_1.
\end{aligned} \tag{26}$$

So we have

$$\begin{aligned}
\|\mathfrak{F}(\mathcal{H}_0 + \mathcal{H}_1)\|_2^2 &\geq (1 - \delta_{(I_1, I_2, \dots, 2r_n, \dots, I_N)}) \|\mathcal{H}_0 + \mathcal{H}_1\|_F^2 \\
&= (1 - \delta_{(I_1, I_2, \dots, 2r_n, \dots, I_N)}) (\|\mathcal{H}_0\|_F^2 + \|\mathcal{H}_1\|_F^2) \\
&\geq (1 - \delta_{(I_1, I_2, \dots, 2r_n, \dots, I_N)}) (\|h_0\|_1^2 + \|h_1\|_1^2) / r_n
\end{aligned} \tag{27}$$

and

$$\begin{aligned}
& \|\mathfrak{F}(\mathcal{H}_2 + \mathcal{H}_3 + \dots)\|_2^2 = \sum_{j,k \geq 2} \langle \mathfrak{F}(\mathcal{H}_j), \mathfrak{F}(\mathcal{H}_k) \rangle \\
& = \sum_{j \geq 2} \langle \mathfrak{F}(\mathcal{H}_j), \mathfrak{F}(\mathcal{H}_j) \rangle + 2 \sum_{2 \leq j < k} \langle \mathfrak{F}(\mathcal{H}_j), \mathfrak{F}(\mathcal{H}_k) \rangle \\
& \leq \sum_{j \geq 2} (1 + \delta_{(I_1, I_2, \dots, r_n, \dots, I_N)}) \|\mathcal{H}_j\|_F^2 \\
& + 2\delta_{(I_1, I_2, \dots, 2r_n, \dots, I_N)} \sum_{2 \leq j < k} \|\mathcal{H}_j\|_F \|\mathcal{H}_k\|_F \\
& = \sum_{j \geq 2} \|\mathcal{H}_j\|_F^2 + \delta_{(I_1, I_2, \dots, 2r_n, \dots, I_N)} \left(\sum_{j \geq 2} \|\mathcal{H}_j\|_F \right)^2
\end{aligned} \tag{28}$$

Further more, by (24),(26) we have

$$\begin{aligned}
& \|\mathfrak{F}(\mathcal{H}_2 + \mathcal{H}_3 + \dots)\|_2^2 \leq \sum_{j \geq 2} \|\mathcal{H}_j\|_F^2 \\
& + \delta_{(I_1, I_2, \dots, 2r_n, \dots, I_N)} \left(\sum_{j \geq 2} \|\mathcal{H}_j\|_F \right)^2 \\
& \leq \frac{t}{r_n} (1-t) \left(\sum_{i \geq 1} \|h_i\|_1 \right)^2 + (1-3t/4) \sum_{i \geq 1} \|h_i\|_1 \\
& = \frac{t(1-t) + \delta_{(I_1, I_2, \dots, 2r_n, \dots, I_N)} (1-3t/4)^2}{r_n} \left(\sum_{i \geq 1} \|h_i\|_1 \right)^2.
\end{aligned} \tag{29}$$

Since $\mathfrak{F}(\mathcal{H}_0 + \mathcal{H}_1 + \mathcal{H}_2 \dots) = \mathfrak{F}(\mathcal{H}) = 0$, we have $\|\mathfrak{F}(\mathcal{H}_0 + \mathcal{H}_1)\|_2^2 = \|\mathfrak{F}(\mathcal{H}_2 + \mathcal{H}_3 + \dots)\|_2^2$, by the above equations we have

$$\begin{aligned}
& (1 - \delta_{(I_1, I_2, \dots, 2r_n, \dots, I_N)}) (\|h_0\|_1^2 + \|h_1\|_1^2) / r_n \\
& \leq \frac{t(1-t) + \delta_{(I_1, I_2, \dots, 2r_n, \dots, I_N)} (1-3t/4)^2}{r_n} \left(\sum_{i \geq 1} \|h_i\|_1 \right)^2.
\end{aligned} \tag{30}$$

Hence, let $\delta = \delta_{(I_1, I_2, \dots, 2r_n, \dots, I_N)}$

$$\begin{aligned}
& \|h_0\|_1^2 \\
& \leq \frac{1}{1-\delta} [\delta + (1-3\delta/2)t - (2-25\delta/16)t^2] \left(\sum_{i \geq 1} \|h_i\|_1 \right)^2
\end{aligned} \tag{31}$$

We have a quadratic polynomial of t with $t \in [0, 1]$ in the right-hand side of the above inequality. Hence, by calculus, this quadratic polynomial achieves its maximal value at $t = \frac{1-3\delta/2}{4-25\delta/8} \in [0, 1]$. Therefore we obtain $\|h_0\|_1 \leq \theta \left(\sum_{i \geq 1} \|h_i\|_1 \right)$,

where

$$\theta = \sqrt{\frac{4(1+5\delta-4\delta^2)}{(1-\delta)(32-25\delta)}}. \tag{32}$$

$\delta < (77 - \sqrt{1337})/82 \approx 0.4931$, then $\theta < 1$, we get $\|h_0\|_1 < \left(\sum_{i \geq 1} \|h_i\|_1 \right)$, which is

$$\sum_{j=r_n+1}^{I_n} \sigma_j^{(n)} > \sum_{j=1}^{r_n} \sigma_j^{(n)}. \tag{33}$$

If for all $n = (1, \dots, N)$, we have (33), then we get (15). \square

Proof of Theorem 9.

Proof. The proof of Theorem 8 establishes that any nonzero $\mathcal{H} \in \text{Null}(\mathfrak{F})$ satisfies $\|h_0\|_1 \leq \theta \left(\sum_{i \geq 1} \|h_i\|_1 \right)$. Hence, if $(1 +$

$\frac{\|X_{(i)}^0\|_2}{\alpha})\theta \leq 1$, notice $\theta < 1$, we have

$$\begin{aligned}
& \alpha \geq (\theta^{-1} - 1)^{-1} \|X_{(i)}^0\|_2 \\
& = \frac{\|X_{(i)}^0\|_2 \sqrt{4(1+5\delta-4\delta^2)}}{\sqrt{(1-\delta)(32-25\delta)} - \sqrt{4(1+5\delta-4\delta^2)}}.
\end{aligned} \tag{34}$$

For $\delta = 0.04404$, we obtain $(\theta^{-1} - 1)^{-1} \|X_{(i)}^0\|_2 \approx 9.9849 \|X_{(i)}^0\|_2 \leq \alpha$, which proves the theorem. \square

Proof of Theorem 10.

Proof. Condition (16) is equivalent to

$$\sum_{i=1}^N \sum_{j=1}^{I_i} \sigma_j(H_{(i)}) \geq \sum_{i=1}^N \left(2 + \frac{\|X_{(i)}^0\|_2}{\alpha} \right) \sum_{j=1}^{r_i} \sigma_j(H_{(i)}). \tag{35}$$

Since $\sum_{j=1}^{r_i} \sigma_j(H_{(i)}) \leq \sqrt{r_i} \sqrt{\sum_{j=1}^{r_i} \sigma_j(H_{(i)})^2} \leq \sqrt{r_i} \|H_{(i)}\|_F$, (35) holds provide that

$$\sum_{i=1}^N \sum_{j=1}^{I_i} \sigma_j(H_{(i)}) \geq \sum_{i=1}^N \left(2 + \frac{\|X_{(i)}^0\|_2}{\alpha} \right) \sqrt{r_i} \|H_{(i)}\|_F. \tag{36}$$

Now from (18) and (19), one has

$$\|\mathcal{H}\|_* \geq \max_i \left(2 + \frac{\|X_{(i)}^0\|_2}{\alpha} \right) \sqrt{r_i} \|\mathcal{H}\|_F, \tag{37}$$

which is equivalent to

$$\begin{aligned}
& \frac{1}{N} \sum_{i=1}^N \|H_{(i)}^0\|_* \geq \max_i \left(2 + \frac{\|X_{(i)}^0\|_2}{\alpha} \right) \sqrt{r_i} \|\mathcal{H}\|_F \\
& \geq \frac{1}{N} \sum_{i=1}^N \left(2 + \frac{\|X_{(i)}^0\|_2}{\alpha} \right) \sqrt{r_i} \|H_{(i)}\|_F.
\end{aligned} \tag{38}$$

Thus the null-space condition holds. \square

Acknowledgments

The authors would like to thank the anonymous reviewers for their comments. This work was partly supported by the National Natural Science Foundation of China under grant No. 91120303 and No. 61071181.

References

- [Candes and Tao, 2005] E.J. Candes and T. Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215, 2005.
- [De Lathauwer *et al.*, 2000] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.
- [Donoho and Huo, 2001] D.L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *Information Theory, IEEE Transactions on*, 47(7):2845–2862, 2001.
- [Gandy *et al.*, 2011] S. Gandy, B. Recht, and I. Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27:025010, 2011.
- [Gribonval and Nielsen, 2003] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *Information Theory, IEEE Transactions on*, 49(12):3320–3325, 2003.
- [Horn and Johnson, 1990] R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge Univ Pr, 1990.
- [Kolda and Bader, 2009] T.G. Kolda and B.W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455, 2009.
- [Kolda, 2006] T.G. Kolda. *Multilinear operators for higher-order decompositions*. United States. Department of Energy, 2006.
- [Lai and Yin, 2012] M.J. Lai and W. Yin. Augmented ℓ_1 and nuclear-norm models with a globally linearly convergent algorithm. *Arxiv preprint arXiv:1201.4615*, 2012.
- [Liu *et al.*, 2009] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2114–2121. IEEE, 2009.
- [Mo and Li, 2011] Q. Mo and S. Li. New bounds on the restricted isometry constant δ_{2k} . *Applied and Computational Harmonic Analysis*, 2011.
- [Oymak *et al.*, 2011] S. Oymak, K. Mohan, M. Fazel, and B. Hassibi. A simplified approach to recovery conditions for low rank matrices. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 2318–2322. IEEE, 2011.
- [Recht *et al.*, 2007] B. Recht, M. Fazel, and P.A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *Arxiv preprint arxiv:0706.4138*, 2007.
- [Tomioka *et al.*, 2010] R. Tomioka, K. Hayashi, and H. Kashima. Estimation of low-rank tensors via convex optimization. *Arxiv preprint arXiv:1010.0789*, 2010.
- [Tomioka *et al.*, 2011] R. Tomioka, T. Suzuki, K. Hayashi, and H. Kashima. Statistical performance of convex tensor decomposition. *Advances in Neural Information Processing Systems (NIPS)*, page 137, 2011.
- [Yin *et al.*, 2008] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing. *SIAM Journal on Imaging Sciences*, 1(1):143–168, 2008.
- [Zhang, 2005] Y. Zhang. A simple proof for recoverability of ℓ_1 -minimization: Go over or under? *Rice University CAAM Technical Report TR05-09*, 2005.
- [Zhang, 2008] Y. Zhang. Theory of compressive sensing via ℓ_1 -minimization: a non-rip analysis and extensions. *Rice University CAAM Technical Report TR08-11*, 2008.