

Robust Low-Tubal-Rank Tensor Completion via Convex Optimization*

Qiang Jiang and Michael Ng[†]

Department of Mathematics, The University of Hong Kong, Hong Kong
 michaelkwokpong@gmail.com

Abstract

This paper considers the problem of recovering multidimensional array, in particular third-order tensor, from a random subset of its arbitrarily corrupted entries. Our study is based on a recently proposed algebraic framework in which the tensor-SVD is introduced to capture the low-tubal-rank structure in tensor. We analyze the performance of a convex program, which minimizes a weighted combination of the tensor nuclear norm, a convex surrogate for the tensor tubal rank, and the tensor ℓ_1 norm. We prove that under certain incoherence conditions, this program can recover the tensor exactly with overwhelming probability, provided that its tubal rank is not too large and that the corruptions are reasonably sparse. The number of required observations is order optimal (up to a logarithm factor) when comparing with the degrees of freedom of the low-tubal-rank tensor. Numerical experiments verify our theoretical results and real-world applications demonstrate the effectiveness of our algorithm.

1 Introduction

Real data, such as electroencephalography (EEG), video, and color image, is usually multidimensional in nature. Tensor-based modeling is a proper choice for such data because it is capable of taking full advantage of the underlying multilinear structures to provide better understanding and higher accuracy. In many applications, multidimensional data is probably incomplete or corrupted owing to various unpredictable or unavoidable factors. The tremendous demand for clean data leads to increasing interest in tensor sampling, estimation and recovery during the last decade.

The strategies and their performance for tensor sampling and recovery rely heavily on the low-rank decomposition used to reveal the intrinsic structure in the multidimensional data. For matrix (second-order tensor) data, its rank defined

by means of the smallest number of rank-one matrix decomposition, can be easily obtained via singular value decomposition (SVD). The nuclear norm has been shown to be the convex envelope of matrix rank over a unit ball of spectral norm. Under some mild assumptions, efficient techniques based on convex programming to minimize the nuclear norm can recover the low-rank matrices accurately (or perhaps exactly) from severely compressive [Wright *et al.*, 2013], incomplete [Candés and Recht, 2009; Chen, 2013], or corrupted observations [Candés *et al.*, 2011; Li, 2013].

Unlike the matrix case, several definitions of tensor rank and their convex relaxations can be found in literature. However, each has its own drawbacks. The CANDECOMP/PARAFAC (CP) decomposition [Kolda and Bader, 2009] compresses a tensor as the sum of rank-one outer products. The minimal number of such decomposition is defined as the CP rank, which is NP-hard to compute in general [Kolda and Bader, 2009]. Although efforts [Jain and Oh, 2014; Shah *et al.*, 2015; Karlsson *et al.*, 2016] have been made to recover low-CP-rank tensor in some special cases, it still remains computationally intractable to determine the CP-rank or its best convex approximation. The Tucker rank of a tensor is a vector whose entries are the ranks of unfolding matrices along each mode respectively [Kolda and Bader, 2009]. Inspired by the relation between matrix rank and nuclear norm, Liu *et al.* proposed the Sum of Nuclear Norms (SNN), which is the (weighted) sum of the nuclear norms of unfolding matrices, as a convex relaxation for Tucker rank. The effectiveness of this tractable measure has been successfully validated in both theoretical analyses [Tomioka *et al.*, 2011; Gu *et al.*, 2014] and practical settings (see, *e.g.*, [Tomioka *et al.*, 2010; Gandy *et al.*, 2011] and reference therein). Considering that the approximation may be substantially sub-optimal in certain situations, Mu *et al.* further designed a better (but still sub-optimal) convex surrogate for tensor rank through a more balanced matricization.

The tensor-product and associated algebraic constructs introduced for third [Kilmer and Martin, 2011] and higher order tensors [Martin *et al.*, 2013], provide a new framework in which we can obtain a SVD-like factorization named the tensor-SVD (t-SVD) [Kilmer and Martin, 2011; Martin *et al.*, 2013], and derive a notion of tensor rank referred to as the tubal rank [Kilmer *et al.*, 2013]. Using this new algebraic framework, Zhang and Aeron and Lu *et al.* gave sufficient

*This work was conducted when the authors were with the Department of Mathematics, Hong Kong Baptist University.

[†]Corresponding Author

conditions for convex programming to succeed in exact recovery of low-tubal-rank tensors from incomplete and grossly corrupted observations respectively.

This paper considers a more challenging problem, namely, recovering a low-rank tensor from a small fraction of its entries, part of which are even arbitrarily corrupted. In essence, this problem is tensor-based generalization of robust matrix completion (RMC) [Li, 2013; Shang *et al.*, 2014]. This is why we call it robust tensor completion (RTC) hereafter. Leveraging on the t-SVD algebraic framework, we study the performance of a convex optimization model and provide recovery guarantee. Specifically, we show that when certain tensor incoherence conditions are satisfied, the convex program can exactly recover the low-tubal-rank tensor with high probability, provided that the tubal rank is not too large and that the corruptions are reasonably sparse. The sampling complexity is order optimal (up to a logarithm factor) when comparing with the degrees of freedom of the tensor.

We are aware that the RTC problem has been theoretically examined in [Huang *et al.*, 2014]. Our study departs from that work on several fronts. First, the t-SVD algebraic framework, in which third-order tensors are treated as linear operators over matrices oriented laterally [Kilmer *et al.*, 2013; Braman, 2010], is quite different from the multilinear algebraic setup for Tucker decomposition. Besides, the tubal rank and tensor nuclear norm differ seriously from the Tucker rank and its convex relaxation SNN. Consequently, the recovery theory established in [Huang *et al.*, 2014] is not directly comparable to our result.

2 Notations and Preliminaries

We first go over some notations. In this paper, tensors are denoted by boldface Euler letters, *e.g.*, \mathcal{A} , matrices by boldface capital letters, *e.g.*, \mathbf{A} , vectors by boldface lowercase letters, *e.g.*, \mathbf{a} , and scalars by lowercase letters, *e.g.*, a . The field of real number and complex number are denoted as \mathbb{R} and \mathbb{C} , respectively. We denote $\lfloor n \rfloor$ as the nearest integer less than or equal to n and $\lceil n \rceil$ as the one greater than or equal to n .

For a third-order tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we denote its (i, j, k) -th entry as \mathcal{A}_{ijk} and denote its i -th horizontal, lateral and frontal slice using the MATLAB notation $\mathcal{A}(i, :, :)$, $\mathcal{A}(:, i, :)$ and $\mathcal{A}(:, :, i)$, respectively. For simplicity, the front slice $\mathcal{A}(:, :, i)$ is denoted compactly as $\mathcal{A}^{(i)}$. $\mathcal{A}(i, j, :)$ denotes a tubal fiber oriented into the board obtained by fixing the first two indices and varying the third. Moreover, a tensor tube of size $1 \times 1 \times n_3$ is denoted as $\hat{\mathbf{a}}$ and a tensor column of size $n_1 \times 1 \times n_3$ is denoted as $\hat{\mathbf{b}}$.

For a vector $\mathbf{v} \in \mathbb{C}^n$, the ℓ_2 -norm is $\|\mathbf{v}\|_2 = \sqrt{\sum_i |v_i|^2}$. The inner product of \mathbf{A} and \mathbf{B} in $\mathbb{C}^{n_1 \times n_2}$ is given by $\langle \mathbf{A}, \mathbf{B} \rangle = \text{Tr}(\mathbf{A}^H \mathbf{B})$, where \mathbf{A}^H denotes the conjugate transpose of \mathbf{A} and $\text{Tr}(\cdot)$ denotes the matrix trace. The spectral norm of a matrix $\mathbf{A} \in \mathbb{C}^{n_1 \times n_2}$ is denoted as $\|\mathbf{A}\| = \max_i \sigma_i(\mathbf{A})$, where $\sigma_i(\mathbf{A})$'s are the singular values of \mathbf{A} . The matrix nuclear norm is $\|\mathbf{A}\|_* = \sum_i \sigma_i(\mathbf{A})$. The inner product of \mathcal{A} and \mathcal{B} in $\mathbb{C}^{n_1 \times n_2 \times n_3}$ is defined as $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i=1}^{n_3} \langle \mathcal{A}^{(i)}, \mathcal{B}^{(i)} \rangle$. For any $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, the complex conjugate of \mathcal{A} is denoted as $\text{conj}(\mathcal{A})$ which takes the complex conjugate of all entries of \mathcal{A} . We denote

the ℓ_1 -norm as $\|\mathcal{A}\|_1 = \sum_{ijk} |\mathcal{A}_{ijk}|$, the infinity norm as $\|\mathcal{A}\|_\infty = \max_{ijk} |\mathcal{A}_{ijk}|$ and the Frobenius norm as $\|\mathcal{A}\|_F = \sqrt{\sum_{ijk} |\mathcal{A}_{ijk}|^2}$. These norms reduce to the corresponding vector or matrix norms if \mathcal{A} is a vector or a matrix.

For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, using the MATLAB command `fft`, we denote $\bar{\mathcal{A}} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ as the result of Fast Fourier Transformation (FFT) of \mathcal{A} along the third dimension, *i.e.*, $\bar{\mathcal{A}} = \text{fft}(\mathcal{A}, [], 3)$. In the same fashion, we can compute \mathcal{A} from $\bar{\mathcal{A}}$ using the inverse FFT, *i.e.*, $\mathcal{A} = \text{ifft}(\bar{\mathcal{A}}, [], 3)$. Furthermore, we define the block circulant matrix of \mathcal{A} as

$$\text{bcirc}(\mathcal{A}) = \begin{bmatrix} \mathbf{A}^{(1)} & \mathbf{A}^{(n_3)} & \dots & \mathbf{A}^{(2)} \\ \mathbf{A}^{(2)} & \mathbf{A}^{(1)} & \dots & \mathbf{A}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}^{(n_3)} & \mathbf{A}^{(n_3-1)} & \dots & \mathbf{A}^{(1)} \end{bmatrix},$$

and the following operators

$$\text{unfold}(\mathcal{A}) = \begin{bmatrix} \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \\ \vdots \\ \mathbf{A}^{(n_3)} \end{bmatrix}, \quad \text{fold}(\text{unfold}(\mathcal{A})) = \mathcal{A}.$$

Definition 1 (t-product) [Kilmer and Martin, 2011] The *t-product* $\mathcal{A} * \mathcal{B}$ of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_4 \times n_2 \times n_3}$ is a tensor $\mathcal{C} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$,

$$\mathcal{C} = \mathcal{A} * \mathcal{B} = \text{fold}(\text{bcirc}(\mathcal{A}) \cdot \text{unfold}(\mathcal{B})).$$

For the frontal slices of $\bar{\mathcal{A}}$, we have

$$\begin{cases} \bar{\mathcal{A}}^{(1)} \in \mathbb{R}^{n_1 \times n_2}, \\ \text{conj}(\bar{\mathcal{A}}^{(i)}) = \bar{\mathcal{A}}^{(n_3-i+2)}, \quad i = 2, \dots, \lfloor \frac{n_3+1}{2} \rfloor. \end{cases} \quad (1)$$

Using the above property, we can compute t-product by a more efficient way (see [Lu *et al.*, 2018, Algorithm 1]).

Definition 2 (Conjugate transpose) [Kilmer and Martin, 2011] The conjugate transpose of a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the tensor $\mathcal{A}^H \in \mathbb{R}^{n_2 \times n_1 \times n_3}$ obtained by conjugate transposing each of the frontal slice and then reversing the order of transposed frontal slices 2 through n_3 .

Definition 3 (Identity tensor) [Kilmer and Martin, 2011] The identity tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is defined to be a tensor whose first frontal slice is the $n \times n$ identity matrix and whose other frontal slices are zero matrices.

Definition 4 (Orthogonal tensor) [Kilmer and Martin, 2011] A tensor $\mathcal{Q} \in \mathbb{R}^{n \times n \times n_3}$ is orthogonal if it satisfies

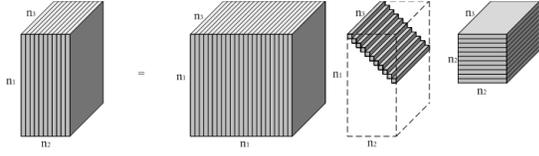
$$\mathcal{Q}^H * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^H = \mathcal{I}.$$

Definition 5 (f-diagonal tensor) [Kilmer and Martin, 2011] A tensor \mathcal{A} is called *f-diagonal* if each frontal slice $\mathcal{A}^{(i)}$ is a diagonal matrix.

Definition 6 (t-SVD) [Kilmer and Martin, 2011; Lu *et al.*, 2018] For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the t-SVD of \mathcal{A} is given by

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^H,$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a *f-diagonal* tensor.


 Figure 1: Sketch map of the t-SVD for a $n_1 \times n_2 \times n_3$ tensor.

The entries of the first frontal slice $\mathcal{S}(:,:,1)$ are called the singular values of the tensor \mathcal{A} [Lu *et al.*, 2018]. Figure 1 illustrates the t-SVD for a $n_1 \times n_2 \times n_3$ tensor. This factorization can be obtained by computing matrix SVDs in the Fourier domain. This idea was proposed first in [Kilmer and Martin, 2011] and later some related works [Kilmer *et al.*, 2013; Zhang and Aeron, 2017], but the way for computing \mathcal{U} and \mathcal{V} cannot guarantee that they are real tensors. The study [Lu *et al.*, 2018] uses property (1) to construct \mathcal{U} and \mathcal{V} , which leads to a more efficient way for computing t-SVD (see [Lu *et al.*, 2018, Algorithm 2]), and thus fixes this issue. It is only required to run $\lceil \frac{n_3+1}{2} \rceil$ matrix SVDs while the number is n_3 for the original method [Kilmer and Martin, 2011].

Definition 7 (Tensor tubal rank) [Lu *et al.*, 2018] For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor tubal rank, denoted as $\text{rank}_t(\mathcal{A})$, is defined as the number of nonzero singular values of \mathcal{S} , where \mathcal{S} is from the t-SVD of $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^H$. We can write

$$\text{rank}_t(\mathcal{A}) = \#\{i, \mathcal{S}(i, i, 1) \neq 0\} = \#\{i, \mathcal{S}(i, i, :) \neq \mathbf{0}\}.$$

For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with tubal rank r , it has the skinny t-SVD, i.e., $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{U}^H$ where $\mathcal{U} \in \mathbb{R}^{n_1 \times r \times n_3}$, $\mathcal{S} \in \mathbb{R}^{r \times r \times n_3}$, and $\mathcal{V} \in \mathbb{R}^{n_2 \times r \times n_3}$ such that $\mathcal{U}^H * \mathcal{U} = \mathcal{I}$ and $\mathcal{V}^H * \mathcal{V} = \mathcal{I}$. The skinny t-SVD will be used in this paper.

There are some connections between tensor tubal rank and other tensor ranks. First, suppose the Tucker rank of \mathcal{A} be $(\text{rank}(\mathcal{A}_{(1)}), \text{rank}(\mathcal{A}_{(2)}), \text{rank}(\mathcal{A}_{(3)}))$, where $\mathcal{A}_{(i)}$ is the i th mode matricization of \mathcal{A} , and we have $\text{rank}_t(\mathcal{A}) \leq \min(\text{rank}(\mathcal{A}_{(1)}), \text{rank}(\mathcal{A}_{(2)}))$. This indicates that a tensor with low Tucker rank also has low tubal rank. Moreover, let the CP decomposition of \mathcal{A} be $\mathcal{A} = \sum_{i=1}^r \mathbf{a}_i^{(1)} \circ \mathbf{a}_i^{(2)} \circ \mathbf{a}_i^{(3)}$, where \circ denotes the outer product. Then $\bar{\mathcal{A}} = \sum_{i=1}^r \mathbf{a}_i^{(1)} \circ \mathbf{a}_i^{(2)} \circ \bar{\mathbf{a}}_i^{(3)}$, where $\bar{\mathbf{a}}_i^{(3)} = \text{fft}(\mathbf{a}_i^{(3)})$. This implies that each frontal slice of $\bar{\mathcal{A}}$ is the sum of r rank-one matrices and $\bar{\mathcal{A}}$ has CP rank at most r . Therefore, the tubal rank of \mathcal{A} is not larger than r . In a word, the low-tubal-rank assumption is weaker than the low-Tucker-rank and low-CP-rank assumptions.

Definition 8 (Tensor nuclear norm) [Lu *et al.*, 2018] Let $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^H$ be the t-SVD of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. The tensor nuclear norm of \mathcal{A} is defined as the sum of the tensor singular values, i.e., $\|\mathcal{A}\|_* = \sum_{i=1}^r \mathcal{S}(i, i, 1)$, where $r = \text{rank}_t(\mathcal{A})$.

The above definition of tensor nuclear norm (TNN) is equivalent to that given in [Lu *et al.*, 2016]. Indeed,

$$\begin{aligned} \|\mathcal{A}\|_* &= \sum_{i=1}^r \mathcal{S}(i, i, 1) = \langle \mathcal{S}, \mathcal{I} \rangle = \frac{1}{n_3} \langle \bar{\mathcal{S}}, \bar{\mathcal{I}} \rangle \\ &= \frac{1}{n_3} \sum_{i=1}^{n_3} \langle \bar{\mathcal{S}}^{(i)}, \bar{\mathcal{I}}^{(i)} \rangle = \frac{1}{n_3} \sum_{i=1}^{n_3} \|\bar{\mathcal{A}}^{(i)}\|_* \end{aligned}$$

Definition 9 (Tensor spectral norm) [Lu *et al.*, 2016] The tensor spectral norm of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, denoted as $\|\mathcal{A}\|$, is defined as $\|\mathcal{A}\| = \|\text{bcirc}(\mathcal{A})\| = \max_i \|\bar{\mathcal{A}}^{(i)}\|$.

It has been shown that TNN is the dual norm of the tensor spectral norm, and vice versa [Zhang *et al.*, 2014]. If we define the tensor average rank as $\text{rank}_a(\mathcal{A}) = \frac{1}{n_3} \text{rank}(\text{bcirc}(\mathcal{A}))$, then the convex envelope of the tensor average rank is the tensor nuclear norm within the set $\{\mathcal{A} \mid \|\mathcal{A}\| \leq 1\}$ [Lu *et al.*, 2018]. We would like to emphasize that the above definition of TNN is different from the one in [Zhang and Aeron, 2017] due to the factor $\frac{1}{n_3}$ and this factor is crucial in our theoretical analysis.

3 Exact Recovery Guarantee

Let us consider the RTC problem formally. Suppose we are given a third-order tensor \mathcal{L}_0 having low tubal rank and corrupted by a sparse term \mathcal{E}_0 . Here, both \mathcal{L}_0 and \mathcal{E}_0 are of arbitrary magnitude. We do not know the tubal rank of \mathcal{L}_0 . Also we have no idea about the locations of the nonzero entries of \mathcal{E}_0 , not even how many there are. Can we recover \mathcal{L}_0 accurately (perhaps even exactly) and efficiently from an observed subset of the noisy data $\mathcal{X} = \mathcal{L}_0 + \mathcal{E}_0$?

Mathematically, the problem can be represented by

$$\min_{\mathcal{L}, \mathcal{E}} \text{rank}_t(\mathcal{L}) + \lambda \|\mathcal{E}\|_0, \quad \text{s.t.}, \mathcal{P}_\Omega(\mathcal{L} + \mathcal{E}) = \mathcal{P}_\Omega(\mathcal{X}), \quad (2)$$

where λ is a penalty parameter and \mathcal{P}_Ω is a linear projection such that the entries in the set Ω are given while the remaining entries are missing. The optimization problem (2) is generally NP-hard due to the discrete nature of the tubal rank and the ℓ_0 pseudo-norm. Replacing these two terms by their convex surrogates, namely, tensor nuclear norm and ℓ_1 -norm respectively, leads to the following convex optimization problem

$$\min_{\mathcal{L}, \mathcal{E}} \|\mathcal{L}\|_* + \lambda \|\mathcal{E}\|_1, \quad \text{s.t.}, \mathcal{P}_\Omega(\mathcal{L} + \mathcal{E}) = \mathcal{P}_\Omega(\mathcal{X}). \quad (3)$$

Our model (3) is equivalent to the following tensor completion (TC) problem when there is no corruption, i.e., $\mathcal{E} = \mathbf{0}$,

$$\min_{\mathcal{L}} \|\mathcal{L}\|_* \quad \text{s.t.}, \mathcal{P}_\Omega(\mathcal{L}) = \mathcal{P}_\Omega(\mathcal{X}), \quad (4)$$

and reduces to the tensor robust principal component analysis (TRPCA) problem when Ω is the entire set of indices,

$$\min_{\mathcal{L}, \mathcal{E}} \|\mathcal{L}\|_* + \lambda \|\mathcal{E}\|_1, \quad \text{s.t.}, \mathcal{L} + \mathcal{E} = \mathcal{X}. \quad (5)$$

3.1 Tensor Incoherence Conditions

As discussed in [Huang *et al.*, 2014; Lu *et al.*, 2016; Zhang and Aeron, 2017; Lu *et al.*, 2018], exact recovery is hopeless if most entries of \mathcal{X} are equal to zero. When \mathcal{X} is both low-rank and sparse (e.g., $\mathcal{X}_{111} = 1$ and zeros everywhere else), we are not able to identify the low-rank tensor \mathcal{L}_0 . To make the problem meaningful, we need some tensor incoherence conditions on \mathcal{L}_0 . We denote $\bar{\mathbf{e}}_i$ as the tensor column basis, which is a tensor of size $n_1 \times 1 \times n_3$ with its $(i, 1, 1)$ -th entry

¹In this situation, it is impossible to exactly recover \mathcal{E}_0 (some of its entries are simply not observed!), unless the observed set is identical to the support of \mathcal{E}_0 .

equaling 1 and the rest equaling 0. We also define the tensor tube basis \mathbf{e}_k , which is a tensor of size $1 \times 1 \times n_3$ with its $(1, 1, k)$ -th entry equaling 1 and the rest equaling 0.

Definition 10 (Tensor Incoherence Conditions) Assume that $\text{rank}_t(\mathcal{L}_0) = r$ and its skinny t-SVD is $\mathcal{L}_0 = \mathcal{U} * \mathcal{S} * \mathcal{V}^H$. \mathcal{L}_0 is said to satisfy the tensor incoherence conditions with parameter $\mu_0 > 0$ if

$$\max_{i=1, \dots, n_1} \|\mathcal{U}^H * \mathbf{e}_i\|_F \leq \sqrt{\frac{\mu_0 r}{n_1}}, \quad (6)$$

$$\max_{j=1, \dots, n_2} \|\mathcal{V}^H * \mathbf{e}_j\|_F \leq \sqrt{\frac{\mu_0 r}{n_2}}, \quad (7)$$

and

$$\|\mathcal{U} * \mathcal{V}^H\|_\infty \leq \sqrt{\frac{\mu_0 r}{n_1 n_2 n_3}}. \quad (8)$$

The incoherence conditions (6)-(8) reduce to the regular matrix incoherence conditions [Candés and Recht, 2009; Candés *et al.*, 2011; Li, 2013] when $n_3 = 1$. According to [Chen, 2013], we name (6) and (7) tensor standard incoherence conditions, and (8) tensor joint incoherence condition. Zhang and Aeron indicate that the joint incoherence condition (8) is unnecessary for tensor completion, while we get the same conclusion in an alternative way. However, this incoherence condition is indispensable for obtaining exact solution to the TRPCA and RTC problems, as shown in our analysis². Another concern arises if the corruption term \mathcal{E}_0 has low tubal rank. This can be avoided by assuming that the support of \mathcal{E}_0 is distributed uniformly at random.

3.2 Main Results

Now we present our main results. For convenience, we denote $n_{(1)} = \max(n_1, n_2)$ and $n_{(2)} = \min(n_1, n_2)$.

Theorem 1 Suppose $\mathcal{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ obeys (6)-(8), and the observation set Ω is uniformly distributed among all sets of cardinality $m = \rho n_1 n_2 n_3$. Also suppose that each observed entry is independently corrupted with probability γ . Then, there exist universal constants $c_1, c_2 > 0$ such that with probability at least $1 - c_1(n_{(1)}n_3)^{-c_2}$, the recovery of \mathcal{L}_0 with $\lambda = 1/\sqrt{\rho n_{(1)}n_3}$ is exact, provided that

$$r \leq \frac{c_r n_{(2)}}{\mu_0 (\log(n_{(1)}n_3))^2} \quad \text{and} \quad \gamma \leq c_\gamma \quad (9)$$

where c_r and c_γ are two positive constants.

The theorem tells us that the target tensor \mathcal{L}_0 whose singular vectors $\mathcal{U}(:, i, :)$ and $\mathcal{V}(:, j, :)$ are reasonably spread, can be exactly recovered with probability nearly one from a subset of its entries even if they are arbitrarily corrupted. All we require is that the tubal rank of the tensor \mathcal{L}_0 is not too large, to be exact, on the order of $n_{(2)}/(\mu_0 (\log(n_{(1)}n_3))^2)$ and the corruption term \mathcal{E}_0 is sufficiently sparse. We would like to emphasize that the only ‘‘random distribution’’ in our

²The proof of Theorem 1, the optimization algorithm to solve (3), and additional experimental results are given in the longer version of this paper: <http://www.math.hkbu.edu.hk/~mng/qiang-mng.pdf>

assumptions concerns the locations of the nonzero entries of \mathcal{E}_0 , but not on their magnitudes or signs. Another remarkable fact is that there is no tuning parameter in our model.

The recovery guarantees for problems (4) and (5) are naturally implied by Theorem 1 as in the following corollaries

Corollary 1 Suppose $\mathcal{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ obeys (6) and (7) and m entries of \mathcal{L}_0 are observed with locations sampled uniformly at random, then there exist universal constants $c_0, c_1, c_2 > 0$ such that if

$$m \geq c_0 \mu_0 r n_{(1)} n_3 (\log(n_{(1)}n_3))^2, \quad (10)$$

\mathcal{L}_0 is the unique minimizer to (4) with probability at least $1 - c_1(n_{(1)}n_3)^{-c_2}$.

Corollary 2 Suppose $\mathcal{L}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ obeys (6)-(8) and \mathcal{E}_0 has support uniformly distributed with probability γ . Then, there exist universal constants $c_1, c_2 > 0$ such that with probability at least $1 - c_1(n_{(1)}n_3)^{-c_2}$, $(\mathcal{L}_0, \mathcal{E}_0)$ is the unique minimizer to (5) with $\lambda = 1/\sqrt{\rho n_{(1)}n_3}$, provided that

$$r \leq \frac{c_r n_{(2)}}{\mu_0 (\log(n_{(1)}n_3))^2} \quad \text{and} \quad \gamma \leq c_\gamma \quad (11)$$

where c_r and c_γ are two positive constants.

It is worth pointing out that the theoretical results given here can be easily extended to the case of p th-order tensors with $p \geq 3$, by exploiting the higher-order t-SVD algebraic framework [Martin *et al.*, 2013].

3.3 Connections with Prior Works

Recall that the t-product for third-order tensors and the tubal nuclear norm reduce to the standard matrix multiplication and the matrix nuclear norm respectively when $n_3 = 1$. In this sense, Theorem 1.3 in [Li, 2013] can be viewed as a special case of Theorem 1 in this paper.

The two works [Zhang and Aeron, 2017; Lu *et al.*, 2018] are the most similar to our study. However, they simply focus on the TC and TRPCA problems, which are both special cases of our model. Corollary 2 seems slightly different from Theorem 4.1 in [Lu *et al.*, 2018], but they are actually in good agreement with each other. Note that our incoherence parameter μ_0 and the incoherence parameter μ given in [Lu *et al.*, 2018] have the relationship $\mu_0 = \frac{\mu}{n_3}$. Substituting the equation into (6)-(8) and (11), we obtain the identical results.

The number of observed entries m required by Corollary 1 is $O(r n_{(1)} n_3 (\log(n_{(1)}n_3))^2)$. For a $n_1 \times n_2 \times n_3$ tensor with tubal rank r , its degrees of freedom is $O(r(n_1 + n_2)n_3)$. Therefore, such rate is order optimal (up to a logarithmic factor). The similar results have been observed in the matrix cases [Candés and Recht, 2009; Chen, 2013].

4 Experiments

We conduct a series of experiments to demonstrate the validity of our theorem, and show possible applications of our model and algorithm. As suggested by Theorem 1, the parameter λ is set to be $\lambda = 1/\sqrt{\rho n_{(1)}n_3}$ in all the experiments unless otherwise specified. For practical problems, it is possible to further improve the performance by turning λ cautiously. Nevertheless, the default value is often a good rule of thumb. We employ the standard alternating direction method of multipliers (ADMM) algorithm to solve (3).

$r = 0.05n, \rho = 0.9, \gamma = 0.1$				$r = 0.1n, \rho = 0.8, \gamma = 0.2$			
n	r	$\text{rank}_t(\mathcal{L})$	$\frac{\ \mathcal{L} - \mathcal{L}_0\ _F}{\ \mathcal{L}_0\ _F}$	n	r	$\text{rank}_t(\mathcal{L})$	$\frac{\ \mathcal{L} - \mathcal{L}_0\ _F}{\ \mathcal{L}_0\ _F}$
100	5	5	1.68×10^{-10}	100	10	10	8.53×10^{-10}
200	10	10	5.25×10^{-12}	200	20	20	1.13×10^{-11}
300	15	15	4.80×10^{-12}	300	30	30	2.26×10^{-12}

Table 1: Exact recovery on random data with different sizes. In the two scenarios, synthetic tensors are with different tubal ranks, percentage of missing entries and proportion of grossly corrupted observations.

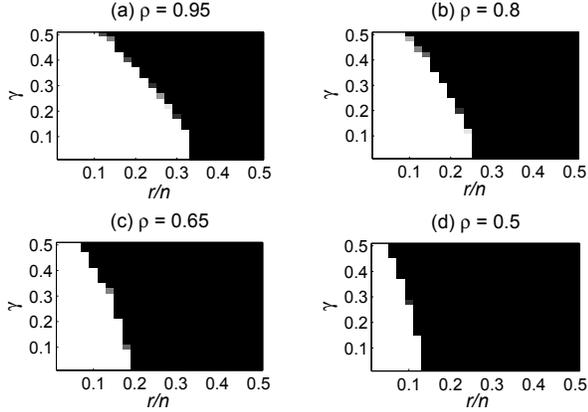


Figure 2: Exact recovery for varying rank and gross corruptions under different proportions of observed entries. Fraction of perfect recoveries across 10 trials, as a function of tubal rank $\text{rank}_t(\mathcal{L}_0)$ (x -axis) and proportion of gross corruptions \mathcal{E}_0 (y -axis).

4.1 Synthetic Tensor Recovery

Validity of Exact Recovery

We first verify the correct recovery phenomenon of Theorem 1 by synthetic problems. For simplicity, we consider the tensors of size $n \times n \times n$ with varying dimension $n = 100, 200$ and 300 . We generate the clean tensor $\mathcal{L}_0 = \mathcal{P} * \mathcal{W}$ with tubal rank $\text{rank}_t(\mathcal{L}_0) = r$, where the entries of $\mathcal{P} \in \mathbb{R}^{n \times r \times n}$ and $\mathcal{W} \in \mathbb{R}^{r \times n \times n}$ are independently sampled from a standard Gaussian distribution $\mathcal{N}(0, 1)$. In addition, a fraction γ of its entries are uniformly corrupted by additive i.i.d. noise from a standard Gaussian distribution $\mathcal{N}(0, 1)$ at random. Finally, we randomly choose a percentage ρ of the noisy tensor entries as our observations.

We test on two cases and summarize the results in Table 1. We set $r = 0.05n$, $\gamma = 0.1$ and $\rho = 0.9$ for the first scenario, and choose a more challenging setting with $r = 0.1n$, $\gamma = 0.2$ and $\rho = 0.8$ for the second scenario. It is clear to see that our method gives the correct rank estimation of \mathcal{L}_0 and the negligible relative error $\|\mathcal{L} - \mathcal{L}_0\|_F / \|\mathcal{L}_0\|_F$ in all cases. These results verify the exact recovery phenomenon as claimed in Theorem 1 pretty well.

Phase Transition

To further corroborate our theoretical results, we check the recovery ability of our algorithm as a function of tubal rank r , fractions of gross corruptions γ and proportion of observed entries ρ . The data are generated by the above-mentioned strategy, where the data size $n = 100$. We set ρ to be different specified values, and vary r and γ to empirically investigate

the probability of recovery success. For each pair (r, γ) , we simulate 10 test instances and declare a trial to be successful if the recovered tensor \mathcal{L} satisfies $\|\mathcal{L} - \mathcal{L}_0\|_F / \|\mathcal{L}_0\|_F \leq 10^{-3}$. Figure 2 reports the fraction of perfect recovery for each pair (black = 0% and white = 100%). We see clearly that there exists a big region in which the recovery is correct for all the cases. Moreover, the larger the percentage of missing values is, the smaller the region of correct recovery becomes.

4.2 Natural Image Restoration

It is well known that the real color images can be well approximated by low-rank matrices on the three channels independently. If a color image is treated as a third-order tensor with each channel corresponding to a frontal slice, then it can be well approximated by a tensor of low tubal rank [Liu *et al.*, 2013; Lu *et al.*, 2016]. In this experiment, we focus on noisy image completion. Unlike the traditional problems of image inpainting and image denoising, this problem aims to simultaneously fill the missing pixels and remove the noise in an image. One typical example is the restoration of archived photographs and films [Kokaram, 2004; Subrahmanyam *et al.*, 2010]. In this case, it is necessary to deal with corruptions and missing values jointly.

We download 50 color images at random from the Berkeley Segmentation Database (BSD) [Martin *et al.*, 2001]. For each image, we randomly set γ pixels corrupted with Gaussian noise $\mathcal{N}(0, \sigma)$, and choose ρ entries to be observed. We compare our algorithm with several approaches for low-rank matrix or tensor recovery, including RPCA [Candés *et al.*, 2011], RMC [Shang *et al.*, 2014], SNN [Huang *et al.*, 2014], and TRPCA [Lu *et al.*, 2016]. For RPCA and RMC, we apply them on each channel independently with $\lambda = 1/\sqrt{n_{(1)}}$. For SNN, we find that its performance is not satisfactory when the parameters λ_i 's are set to the default values [Huang *et al.*, 2014]. As suggested by [Lu *et al.*, 2016], we empirically set $\lambda_1 = \lambda_2 = 15$ and $\lambda_3 = 1.5$ which make SNN work fine for most images. We run TRPCA with $\lambda = 1/\sqrt{n_{(1)}n_3}$. For a comprehensive comparison, we also test BM3D³ [Dabov *et al.*, 2007] on the BSD image set, which is usually referred to as the representative of state-of-the-art algorithms for image restoration. Considering that BM3D is originally proposed for image denoising, we further enhance it with the scheme of ‘‘completion + denoising’’, which means that we first fill the missing pixels without considering the noise and then apply BM3D to the intermediate result. Here, HaLRTC⁴ [Liu *et al.*, 2013] and TNNM [Zhang *et al.*, 2014] are used in the completion step and the corresponding methods are denoted by BM3D+ and BM3D++, respectively.

We consider two different situations in which ρ is set to be 0.9 and 0.7 respectively, and change γ from 0.01 to 0.03 for each case. Table 2 gives the results in terms of average Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity index (SSIM) when $\sigma = 30$. The best results is in bold text and the second one is underlined. Our algorithm yields the best quantitative results for all the cases, significantly better than the runner-up sometimes. The performance of TRPCA

³<http://www.cs.tut.fi/~foi/GCF-BM3D/index.html>

⁴<http://www.cs.rochester.edu/u/jliu/publications.html>

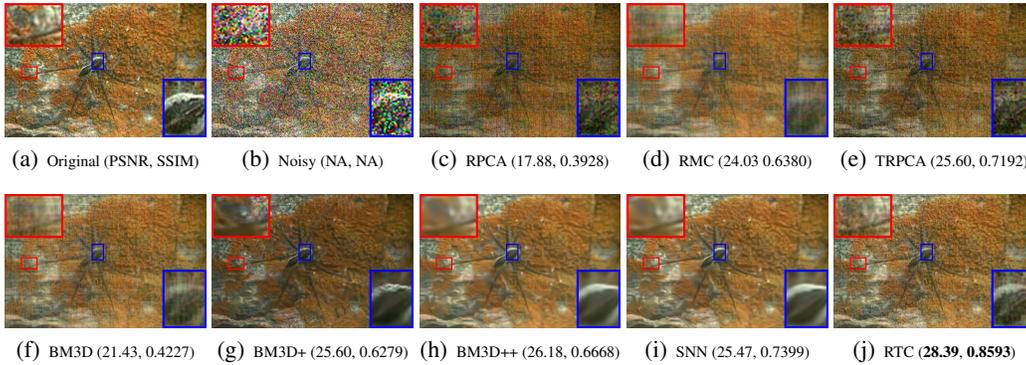


Figure 3: Visual comparison of Image Restoration when $\rho = 0.7$ and $\gamma = 0.3$. Our result contains slightly sharper edges and fewer artifacts.

	$\rho = 0.9$						$\rho = 0.7$					
	$\gamma = 0.1$		$\gamma = 0.2$		$\gamma = 0.3$		$\gamma = 0.1$		$\gamma = 0.2$		$\gamma = 0.3$	
	PSNR	SSIM										
RPCA	27.67	0.8535	27.30	0.8367	26.88	0.8122	21.69	0.5609	20.62	0.4744	19.64	0.4081
RMC	28.11	0.8552	27.82	0.8423	27.53	0.8276	26.33	0.7865	26.09	0.7736	25.84	0.7599
TRPCA	32.31	0.9457	31.59	0.9278	30.91	0.9037	29.25	0.8608	28.69	0.8209	28.06	0.7678
BM3D	28.31	0.8008	28.25	0.8002	28.18	0.7994	20.91	0.3392	20.80	0.3311	20.70	0.3234
BM3D+	30.72	0.8289	30.51	0.8245	30.28	0.8203	29.75	0.8060	29.45	0.7993	29.18	0.7935
BM3D++	30.94	0.8338	30.74	0.8297	30.52	0.8257	30.42	0.8221	30.11	0.8152	29.82	0.8093
SNN	30.14	0.9128	29.60	0.8972	29.11	0.8797	27.75	0.8426	27.35	0.8248	26.97	0.8063
RTC	33.03	0.9566	32.10	0.9400	31.27	0.9185	31.30	0.9296	30.58	0.9091	29.91	0.8831

Table 2: Average PSNR and SSIM obtained by various methods on the BSD image set.

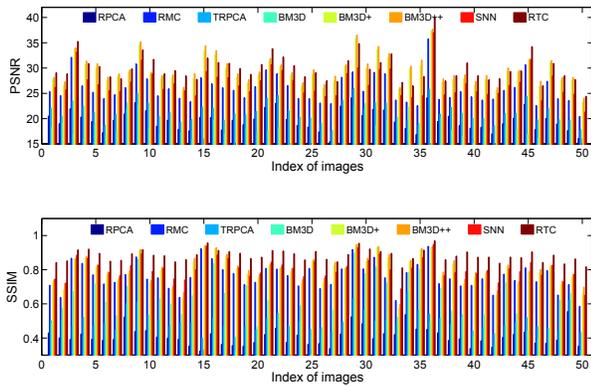


Figure 4: Comparison of the PSNR and SSIM values of various methods for image restoration on 50 images when $\rho = 0.7$ and $\gamma = 0.3$.

is comparable to our algorithm when $\rho = 0.9$, but it deteriorates dramatically as ρ and γ become larger. BM3D is not able to achieve acceptable results, especially when the fraction of missing pixels is relatively high, since it considers removing the noise from the images purely. With an additional completion step, BM3D+ and BM3D++ exhibit remarkably improved performance. Three tensor-based methods, TRPCA, SNN and RTC, perform much better than two matrix-based approaches, RPCA and RMC. The reason is that RPCA and RMC, which conduct the matrix recovery on each channel independently, are incapable of exploiting the information across channels, while the tensor-based methods can take the advantage of the multi-channel structure. We also see that our quantitative results are much better than those obtained

by SNN, which verifies that the t-SVD algebraic framework is more suitable for natural image processing compared with the traditional multilinear algebraic setup.

In Figure 4, we give the results obtained by various methods on all 50 images when $\rho = 0.7$ and $\gamma = 0.3$. Our algorithm outperforms the other methods quantitatively for most images. From the example in Figure 3, we see that our recovered images contain slightly sharper edges and fewer artifacts, exhibited in the enlarged views of the corresponding areas in red and blue boxes.

5 Conclusions

In this work, we conduct a rigorous study for the RTC problem which aims to learn a low-tubal-rank tensor from partial observations that are arbitrarily corrupted. Our study rests heavily on recently proposed t-SVD algebraic framework, in which we can define the tubal rank and tensor nuclear norm for tensors. Equipped with the new definitions, we show that one can exactly recover a third-order tensor having low tubal rank with high probability and establish a theoretical bound for exact recovery when using a convex optimization algorithm. Numerical experiments verify our theoretical analysis and the real-world applications demonstrate the superiority of our method over other existing approaches. Considering that real data routinely lies in thousands or even billions of dimensions, the computational cost of our method may become expensive. We desire to develop fast algorithms for low-tubal-rank tensor recovery and will explore this important direction in our future work.

Acknowledgements

This work is supported in part by the Hong Kong RGC GRF projects 12306616, 12200317 and 12300218.

References

- [Braman, 2010] Karen Braman. Third-order tensors as linear operators on a space of matrices. *Linear Algebra & Its Applications*, 433(7):1241–1253, 2010.
- [Candés and Recht, 2009] Emmanuel J. Candés and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.
- [Candés *et al.*, 2011] Emmanuel J. Candés, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):1–73, May 2011.
- [Chen, 2013] Yudong Chen. Incoherence-optimal matrix completion. *IEEE TIT*, 61(5):2909–2923, 2013.
- [Dabov *et al.*, 2007] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image restoration by sparse 3D transform-domain collaborative filtering. *IEEE TIP*, 16(8):2080–2095, August 2007.
- [Gandy *et al.*, 2011] Silvia Gandy, Benjamin Recht, and Isao Yamada. Tensor completion and low- n -rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):025010, 2011.
- [Gu *et al.*, 2014] Quanquan Gu, Huan Gui, and Jiawei Han. Robust tensor decomposition with gross corruption. In *Proceedings of the 27th Advances in Neural Information Processing Systems*, pages 1422–1430, 2014.
- [Huang *et al.*, 2014] Bo Huang, Cun Mu, Donald Goldfarb, and John Wright. Provable low-rank tensor recovery. *Optimization-Online*, page 4252, 2014.
- [Jain and Oh, 2014] Prateek Jain and Sewoong Oh. Provable tensor factorization with missing data. In *NIPS*, pages 1431–1439, 2014.
- [Karlsson *et al.*, 2016] Lars Karlsson, Daniel Kressner, and André Uschmajew. Parallel algorithms for tensor completion in the CP format. *Parallel Comput.*, 57:222–234, September 2016.
- [Kilmer and Martin, 2011] Misha E. Kilmer and Carla D. Martin. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3):641–658, 2011.
- [Kilmer *et al.*, 2013] Misha E. Kilmer, Karen Braman, Ning Hao, and Randy C. Hoover. Third-order tensors as operators on matrices: a theoretical and computational framework with applications in imaging. *SIAM Journal on Mat. Anal. & App.*, 34(1):148–172, 2013.
- [Kokaram, 2004] Anil C. Kokaram. On missing data treatment for degraded video and film archives: A survey and a new bayesian approach. *IEEE TIP*, 13(3):397–415, March 2004.
- [Kolda and Bader, 2009] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.
- [Li, 2013] Xiaodong Li. Compressed sensing and matrix completion with constant proportion of corruptions. *Constructive Approximation*, 37(1):73–99, 2013.
- [Liu *et al.*, 2013] Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye. Tensor completion for estimating missing values in visual data. *IEEE T PAMI*, 35(1):208–220, 2013.
- [Lu *et al.*, 2016] Canyi Lu, Jiashi Feng, Yudong Chen, Liu Wei, and Shuicheng Yan. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In *CVPR*, pages 5249–5257, 2016.
- [Lu *et al.*, 2018] Canyi Lu, Jiashi Feng, Yudong Chen, Liu Wei, Zhouchen Lin, and Shuicheng Yan. Tensor robust principal component analysis with a new tensor nuclear norm. *arXiv:1804.03728*, 2018.
- [Martin *et al.*, 2001] David R. Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and application to evaluating segmentation algorithms and measuring ecological statistics. *ICCV*, 2(11):416–423, 2001.
- [Martin *et al.*, 2013] Carla D. Martin, Richard Shafer, and Betsy Larue. An order- p tensor factorization with applications in imaging. *SIAM Journal on Scientific Computing*, 35(1):A474–A490, 2013.
- [Mu *et al.*, 2013] Cun Mu, Huang Bo, John Wright, and Donald Goldfarb. Square deal: Lower bounds and improved relaxations for tensor recovery. In *ICML*, pages 73–81, 2013.
- [Shah *et al.*, 2015] Parikshit Shah, Nikhil Rao, and Gongguo Tang. Sparse and low-rank tensor decomposition. In *NIPS*, pages 2548–2556, 2015.
- [Shang *et al.*, 2014] Fanhua Shang, Yuanyuan Liu, James Cheng, and Hong Cheng. Robust principal component analysis with missing data. In *CIKM*, pages 1149–1158, November 2014.
- [Subrahmanyam *et al.*, 2010] Gorthi R.K.S. Subrahmanyam, Ambasadram N. Rajagopalan, and Rangarajan Aravind. Recursive framework for joint inpainting and denoising of photographic films. *Journal of the Optical Society of America A*, 27(5):1091–1099, 2010.
- [Tomioka *et al.*, 2010] Ryota Tomioka, Kohei Hayashi, and Hisashi Kashima. Estimation of low-rank tensors via convex optimization. *arXiv:1010.0789*, 2010.
- [Tomioka *et al.*, 2011] Ryota Tomioka, Taiji Suzuki, Kohei Hayashi, and Hisashi Kashima. Statistical performance of convex tensor decomposition. In *NIPS*, pages 972–980, 2011.
- [Wright *et al.*, 2013] John Wright, Arvind Ganesh, Kerui Min, and Yi Ma. Compressive principal component pursuit. *Information and Inference*, 2(1):32–68, 2013.
- [Zhang and Aeron, 2017] Zemin Zhang and Shuchin Aeron. Exact tensor completion using t-SVD. *IEEE TSP*, 65(6):1511–1526, 2017.
- [Zhang *et al.*, 2014] Zemin Zhang, Gregory Ely, Shuchin Aeron, Ning Hao, and Misha Kilmer. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In *CVPR*, pages 3842–3849, 2014.