

Exact Low Tubal Rank Tensor Recovery from Gaussian Measurements

Canyi Lu¹, Jiashi Feng², Zhouchen Lin^{3,4*}, Shuicheng Yan^{5,2}

¹ Department of Electrical and Computer Engineering, Carnegie Mellon University

² Department of Electrical and Computer Engineering, National University of Singapore

³ Key Laboratory of Machine Perception (MOE), School of EECS, Peking University

⁴ Cooperative Medianet Innovation Center, Shanghai Jiao Tong University

⁵ 360 AI Institute

canyilu@gmail.com, elefjia@nus.edu.sg, zlin@pku.edu.cn, eleyans@nus.edu.sg

Abstract

The recent proposed Tensor Nuclear Norm (TNN) [Lu *et al.*, 2016; 2018a] is an interesting convex penalty induced by the tensor SVD [Kilmer and Martin, 2011]. It plays a similar role as the matrix nuclear norm which is the convex surrogate of the matrix rank. Considering that the TNN based Tensor Robust PCA [Lu *et al.*, 2018a] is an elegant extension of Robust PCA with a similar tight recovery bound, it is natural to solve other low rank tensor recovery problems extended from the matrix cases. However, the extensions and proofs are generally tedious. The general atomic norm provides a unified view of low-complexity structures induced norms, e.g., the ℓ_1 -norm and nuclear norm. The sharp estimates of the required number of generic measurements for exact recovery based on the atomic norm are known in the literature. In this work, with a careful choice of the atomic set, we prove that TNN is a special atomic norm. Then by computing the Gaussian width of certain cone which is necessary for the sharp estimate, we achieve a simple bound for guaranteed low tubal rank tensor recovery from Gaussian measurements. Specifically, we show that by solving a TNN minimization problem, the underlying tensor of size $n_1 \times n_2 \times n_3$ with tubal rank r can be exactly recovered when the given number of Gaussian measurements is $O(r(n_1 + n_2 - r)n_3)$. It is order optimal when comparing with the degrees of freedom $r(n_1 + n_2 - r)n_3$. Beyond the Gaussian mapping, we also give the recovery guarantee of tensor completion based on the uniform random mapping by TNN minimization. Numerical experiments verify our theoretical results.

1 Introduction

Many engineering problems look for solutions to underdetermined systems of linear equations: a system is considered underdetermined if there are fewer equations than unknowns.

*Corresponding author.

Suppose we are given information about an object $\mathbf{x}_0 \in \mathbb{R}^d$ of the form $\Phi \mathbf{x}_0 \in \mathbb{R}^m$ where Φ is an $m \times d$ matrix. We want the bound on the number of rows m to ensure that \mathbf{x}_0 is the unique minimizer to the problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_A, \text{ s.t. } \Phi \mathbf{x}_0 = \Phi \mathbf{x}. \quad (1)$$

Here $\|\cdot\|_A$ is a norm with some suitable properties which encourage solutions to conform to some notion of simplicity. For example, the compressed sensing problem aims to recover a sparse vector \mathbf{x}_0 from (1) by taking $\|\cdot\|_A$ as the ℓ_1 -norm $\|\mathbf{x}\|_1$. We would like to know that how many measurements are required to recover an s -sparse \mathbf{x}_0 . This of course depends on the kind of measurements. For instance, it is shown in [Candès *et al.*, 2006] that $20s \log d$ randomly selected Fourier coefficients are sufficient. If the Gaussian measurement map (Φ has entries i.i.d. sampled from a Gaussian distribution with mean 0 and variance $\frac{1}{m}$) is used, $2s \log \frac{d}{s} + \frac{5}{4}s$ measurements are needed [Donoho and Tanner, 2009; Chandrasekaran *et al.*, 2012]. Another interesting structured object is the low-rank matrix $\mathbf{X}_0 \in \mathbb{R}^{n_1 \times n_2}$. In this case, the i th component of a linear operator is given by $[\Phi(\mathbf{X}_0)]_i = \langle \Phi_i, \mathbf{X}_0 \rangle$, where $\Phi_i \in \mathbb{R}^{n_1 \times n_2}$. This includes the matrix completion problem [Candès and Recht, 2009] as a special case based on a proper choice of Φ_i . By taking $\|\cdot\|_A$ as the matrix nuclear norm $\|\mathbf{X}\|_*$, the convex program (1) recovers \mathbf{X}_0 provided that the number of measurements is of the order $\mu(\mathbf{X}_0)r(n_1 + n_2 - r) \log^2(n_1 + n_2)$, where r is the rank of \mathbf{X}_0 and $\mu(\mathbf{X}_0)$ is the incoherence parameter [Candès and Recht, 2009; Chen, 2015]. Compared with the degrees of freedom $r(n_1 + n_2 - r)$ of the rank- r matrix, such a rate is optimal (up to a logarithmic factor). If the Gaussian measurement map is used, about $3r(n_1 + n_2 - r)$ samples are sufficient for exact recovery [Recht *et al.*, 2010].

Beyond the sparse vector and low-rank matrix, there have some other structured signals which can be recovered by (1). The work [Chandrasekaran *et al.*, 2012] gives some more examples, presents a unified view of the convex programming to inverse problems and provides a simple framework to derive exact recovery bounds for a variety of simple models. Their considered models are formed as the sum of a few atoms from some elementary atomic sets. The convex programming formulation is based on minimizing the norm induced by the

convex hull of the atomic set; this norm is referred to as the atomic norm (the ℓ_1 -norm and nuclear norm are special cases). By using the properties of the atomic norm, an analysis of the underlying convex geometry provides sharp estimates of the number of generic measurements required for exact recovery of models from partial information. A key step to estimate the required number of measurements is to compute the Gaussian width of the tangent cone associated with the atomic norm ball.

This work focuses the study on the low-rank tensor which is an interesting object structured that has many applications in signal processing. Recovering low-rank tensor is not easy since the tensor rank is not well defined. There have several tensor rank definitions, but each has its limitation. For example, the CP rank, defined as the smallest number of rank one tensor decomposition, is generally NP hard to compute. Also, its convex envelope is in general intractable. The tractable Tucker rank is more widely used. However, considering the low Tucker rank tensor recovery problem, the required number of measurements of existing convex model is much higher than the degrees of freedom [Mu *et al.*, 2014]. This is different from the nuclear norm minimization for low-rank matrix recovery which has order optimal rate [Chen, 2015].

In this work, we first study the low tubal rank tensor recovery from Gaussian measurements. Tensor RPCA [Lu *et al.*, 2016; 2018a] studies the low tubal rank tensor recovery from sparse corruptions by Tensor Nuclear Norm (TNN) minimization. We show that TNN is a new instance of the atomic norm based on a proper choice of the atomic set. From the perspective of atomic norm minimization, we give the low tubal rank recovery guarantee from Gaussian measurements. Specifically, to recover a tensor of size $n_1 \times n_2 \times n_3$ with tubal rank r from Gaussian measurement by TNN minimization, the required number of measurements is $O(r(n_1 + n_2 - r)n_3)$. It is order optimal when comparing with the degrees of freedom $r(n_1 + n_2 - r)n_3$. Second, we study the tensor completion problem from uniform random sampling. We show that, to recover a tensor of tubal rank r , the sampling complexity is $O(r \min(n_1, n_2)n_3 \log^2(\min(n_1, n_2)n_3))$, which is order optimal (up to a log factor). The same problem has been studied in [Zhang and Aeron, 2017] but its proofs have several errors.

2 Notations and Preliminaries

We introduce some notations used in this paper. We denote tensors by boldface Euler script letters, e.g., \mathcal{A} , matrices by boldface capital letters, e.g., \mathbf{A} , vectors by boldface lower-case letters, e.g., \mathbf{a} , and scalars by lowercase letters, e.g., a . We denote \mathbf{I}_n as the $n \times n$ sized identity matrix. The field of real number and complex number are denoted as \mathbb{R} and \mathbb{C} , respectively. For a 3-way tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, we denote its (i, j, k) -th entry as \mathcal{A}_{ijk} or a_{ijk} and use the Matlab notation $\mathcal{A}(i, :, :)$, $\mathcal{A}(:, i, :)$ and $\mathcal{A}(:, :, i)$ to respectively denote the i -th horizontal, lateral and frontal slice. More often, the frontal slice $\mathcal{A}(:, :, i)$ is denoted compactly as $\mathbf{A}^{(i)}$. The tube is denoted as $\mathcal{A}(i, j, :)$. The inner product of \mathbf{A} and \mathbf{B} in $\mathbb{C}^{n_1 \times n_2}$ is defined as $\langle \mathbf{A}, \mathbf{B} \rangle = \text{Tr}(\mathbf{A}^* \mathbf{B})$, where \mathbf{A}^* denotes the conjugate transpose of \mathbf{A} and $\text{Tr}(\cdot)$ denotes the matrix trace. The inner product of \mathcal{A} and \mathcal{B} in

$\mathbb{C}^{n_1 \times n_2 \times n_3}$ is defined as $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i=1}^{n_3} \langle \mathbf{A}^{(i)}, \mathbf{B}^{(i)} \rangle$. 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 $\lfloor t \rfloor$ as the nearest integer less than or equal to t and $\lceil t \rceil$ as the one greater than or equal to t . We denote the ℓ_1 -norm as $\|\mathcal{A}\|_1 = \sum_{ijk} |a_{ijk}|$, the infinity norm as $\|\mathcal{A}\|_\infty = \max_{ijk} |a_{ijk}|$ and the Frobenius norm as $\|\mathcal{A}\|_F = \sqrt{\sum_{ijk} |a_{ijk}|^2}$. The same norms are used for matrices and vectors. The spectral norm of a matrix \mathbf{A} 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 $\|\mathbf{A}\|_* = \sum_i \sigma_i(\mathbf{A})$.

For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, by 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 3-rd 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., $\mathcal{A} = \text{ifft}(\bar{\mathcal{A}}, [], 3)$. In particular, we denote $\bar{\mathbf{A}}$ as a block diagonal matrix with i -th block on the diagonal as the frontal slice $\bar{\mathbf{A}}^{(i)}$ of $\bar{\mathcal{A}}$, i.e.,

$$\bar{\mathbf{A}} = \text{bdiag}(\bar{\mathcal{A}}) = \begin{bmatrix} \bar{\mathbf{A}}^{(1)} & & & \\ & \bar{\mathbf{A}}^{(2)} & & \\ & & \ddots & \\ & & & \bar{\mathbf{A}}^{(n_3)} \end{bmatrix}.$$

The block circulant matrix of \mathcal{A} is defined as

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

The block circulant matrix can be block diagonalized, i.e.,

$$(\mathbf{F}_{n_3} \otimes \mathbf{I}_{n_1}) \cdot \text{bcirc}(\mathcal{A}) \cdot (\mathbf{F}_{n_3}^{-1} \otimes \mathbf{I}_{n_2}) = \bar{\mathbf{A}},$$

where $\mathbf{F}_{n_3} \in \mathbb{C}^{n_3 \times n_3}$ is the discrete Fourier transformation matrix, \otimes denotes the Kronecker product. Note that $(\mathbf{F}_{n_3} \otimes \mathbf{I}_{n_1}) / \sqrt{n_3}$ is orthogonal. We define the following operators

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

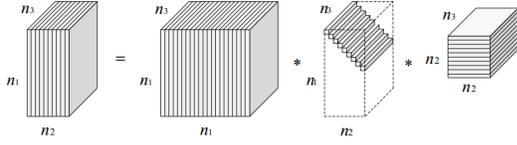
Definition 1. (t-product) [Kilmer and Martin, 2011] Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times l \times n_3}$. Then the t-product $\mathcal{A} * \mathcal{B}$ is defined to be a tensor $\mathcal{C} \in \mathbb{R}^{n_1 \times l \times n_3}$,

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

The frontal slices of $\bar{\mathcal{A}}$ has the following property

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

Using the above property, the work [Lu *et al.*, 2018a] proposes a more efficient way for computing t-product than the method in [Kilmer and Martin, 2011].


 Figure 1: Illustration of the t-SVD of an $n_1 \times n_2 \times n_3$ tensor.

Definition 2. (Conjugate transpose) [Lu et al., 2016; 2018a] The conjugate transpose of a tensor \mathcal{A} of size $n_1 \times n_2 \times n_3$ is the $n_2 \times n_1 \times n_3$ tensor \mathcal{A}^* 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 the tensor whose first frontal slice is the $n \times n$ identity matrix, and other frontal slices are all zeros.

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

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

Definition 5. (F-diagonal Tensor) [Kilmer and Martin, 2011] A tensor is called f-diagonal if each of its frontal slices is a diagonal matrix.

Theorem 1. (T-SVD) [Lu et al., 2018a; Kilmer and Martin, 2011] Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Then it can be factored as

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

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

Theorem 1 gives the t-SVD based on t-product. See Figure 1 for an illustration. Theorem 1 appears first in [Kilmer and Martin, 2011] but their proof is not rigorous since it cannot guarantee that \mathcal{U} and \mathcal{V} are real tensors. The work [Lu et al., 2018a] fixes this issue by using property (2), and further gives a more efficient way for computing t-SVD (see Algorithm 1). Algorithm 1 only needs to compute $\lceil \frac{n_3+1}{2} \rceil$ matrix SVDs, while this number is n_3 by the method in [Kilmer and Martin, 2011]. The entries of the first frontal slice $\mathcal{S}(:, :, 1)$ are called as the singular values of the tensor \mathcal{A} . The number of nonzero singular values is equivalent to the tensor tubal rank.

Definition 6. (Tensor tubal rank) [Lu et al., 2018a] 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 $c\mathcal{S}$, where \mathcal{S} is from the t-SVD of $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$. We can write

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

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., $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$, 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 $\mathcal{V} \in \mathbb{R}^{n_2 \times r \times n_3}$, in which $\mathcal{U}^* * \mathcal{U} = \mathcal{I}$ and $\mathcal{V}^* * \mathcal{V} = \mathcal{I}$. We use the skinny t-SVD throughout this paper.

Definition 7. (Tensor nuclear norm) [Lu et al., 2018a] Let $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$ 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 $r = \text{rank}_t(\mathcal{A})$.

Algorithm 1 T-SVD

Input: $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$.

Output: T-SVD components \mathcal{U} , \mathcal{S} and \mathcal{V} of \mathcal{A} .

1. Compute $\bar{\mathcal{A}} = \text{fft}(\mathcal{A}, [], 3)$.
 2. Compute each frontal slice of $\bar{\mathcal{U}}$, $\bar{\mathcal{S}}$ and $\bar{\mathcal{V}}$ from $\bar{\mathcal{A}}$ by **for** $i = 1, \dots, \lceil \frac{n_3+1}{2} \rceil$ **do**

$$[\bar{\mathcal{U}}^{(i)}, \bar{\mathcal{S}}^{(i)}, \bar{\mathcal{V}}^{(i)}] = \text{SVD}(\bar{\mathcal{A}}^{(i)});$$
end for
for $i = \lceil \frac{n_3+1}{2} \rceil + 1, \dots, n_3$ **do**

$$\bar{\mathcal{U}}^{(i)} = \text{conj}(\bar{\mathcal{U}}^{(n_3-i+2)});$$

$$\bar{\mathcal{S}}^{(i)} = \bar{\mathcal{S}}^{(n_3-i+2)};$$

$$\bar{\mathcal{V}}^{(i)} = \text{conj}(\bar{\mathcal{V}}^{(n_3-i+2)});$$
end for
 3. Compute $\mathcal{U} = \text{ifft}(\bar{\mathcal{U}}, [], 3)$, $\mathcal{S} = \text{ifft}(\bar{\mathcal{S}}, [], 3)$, and $\mathcal{V} = \text{ifft}(\bar{\mathcal{V}}, [], 3)$.
-

The above definition of TNN is defined based on t-SVD. It is equivalent to $\frac{1}{n_3} \|\bar{\mathcal{A}}\|_*$ as given in [Lu et al., 2016]. Indeed,

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

Above the factor $\frac{1}{n_3}$ is from the property $\|\mathbf{F}_{n_3}\|_F^2 = n_3$, where \mathbf{F}_{n_3} is the discrete Fourier transformation matrix.

Definition 8. (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 $\|\mathcal{A}\| = \|\text{bcirc}(\mathcal{A})\|$.

TNN is the dual norm of the tensor spectral norm, and vice versa. Define the tensor average rank as $\text{rank}_a(\mathcal{A}) = \frac{1}{n_3} \text{bcirc}(\mathcal{A})$. Then the convex envelope of the tensor average rank is the tensor nuclear within the set $\{\mathcal{A} \mid \|\mathcal{A}\| \leq 1\}$. It is worth mentioning that the above definition of tensor nuclear norm is different from the one in [Zhang and Aeron, 2017] due to the factor $\frac{1}{n_3}$. This factor is crucial in theoretical analysis. Intuitively, it makes the model, theoretical proof and the way for optimization consistent with the matrix cases.

3 Tensor Nuclear Norm Is an Atomic Norm

Based on the above tensor tubal rank, this work considers the following problem. Suppose that we have a linear map $\Phi : \mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \mathbb{R}^m$ and the observations $\mathbf{y} = \Phi(\mathcal{M})$ for $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ which has tubal rank r . Our goal is to recover the underlying \mathcal{M} from the observations \mathbf{y} . This can be achieved by solving the following convex program

$$\hat{\mathcal{X}} = \arg \min_{\mathcal{X}} \|\mathcal{X}\|_*, \text{ s.t. } \mathbf{y} = \Phi(\mathcal{X}). \quad (3)$$

Now, how many measurements are required to guarantee the exact recovery (i.e., $\hat{\mathcal{X}} = \mathcal{M}$)? This problem is an extension of the low-rank matrix recovery problem [Recht et al., 2010].

To answer the above question, we will use the unified theory in [Chandrasekaran *et al.*, 2012] which provides sharp estimates of the number of measurements required for exact and robust recovery of models from Gaussian measurements. The key challenge is to reformulate TNN as a special case of the atomic norm and compute the Gaussian width. In this section, we will show that TNN is a special case of the atomic norm.

Let A be a collection of atoms that is a compact subset of \mathbb{R}^p and $\text{conv}(A)$ be its convex hull. The atomic norm induced by A is defined as [Chandrasekaran *et al.*, 2012]

$$\|\mathbf{x}\|_A = \inf \left\{ \sum_{\mathbf{a} \in A} c_{\mathbf{a}} : \mathbf{x} = \sum_{\mathbf{a} \in A} c_{\mathbf{a}} \mathbf{a}, c_{\mathbf{a}} \geq 0, \forall \mathbf{a} \in A \right\}.$$

We also need some other notations which will be used in the analysis. The support function of A is given as

$$\|\mathbf{x}\|_A^* = \sup \{ \langle \mathbf{x}, \mathbf{a} \rangle : \mathbf{a} \in A \}.$$

If $\|\cdot\|_A$ is a norm, the support function $\|\cdot\|_A^*$ is the dual norm of this atomic norm.

A convex set C is a cone if it is closed under positive linear combinations. The polar C^* of a cone C is the cone

$$C^* = \{ \mathbf{x} \in \mathbb{R}^p : \langle \mathbf{x}, \mathbf{z} \rangle \leq 0, \forall \mathbf{z} \in C \}.$$

The tangent cone at nonzero \mathbf{x} is defined as

$$T_A(\mathbf{x}) = \text{cone} \{ \mathbf{z} - \mathbf{x} : \|\mathbf{z}\|_A \leq \|\mathbf{x}\|_A \}.$$

The normal cone $N_A(\mathbf{x})$ at \mathbf{x} is defined as

$$N_A(\mathbf{x}) = \{ \mathbf{s} : \langle \mathbf{s}, \mathbf{z} - \mathbf{s} \rangle \leq 0, \forall \mathbf{z} \text{ s.t. } \|\mathbf{z}\|_A \leq \|\mathbf{x}\|_A \}.$$

Note that the normal cone $N_A(\mathbf{x})$ is the conic hull of the subdifferential of the atomic norm at \mathbf{x} .

By a proper choice of the set A , the atomic norm reduces to several well-known norms. For example, let $A \subset \mathbb{R}^p$ be the set of unit-norm one-sparse vectors $\{\pm \mathbf{e}_i\}_{i=1}^p$. Then k -sparse vectors in \mathbb{R}^p can be constructed using a linear combination of k elements of the atomic set and the atomic norm $\|\mathbf{x}\|_A$ reduces to the ℓ_1 -norm. Let A be the set of rank-one matrices of unit-Euclidean-norm. Then the rank- k matrices can be constructed using a linear combination of k elements of the atomic set and the atomic norm reduces to the matrix nuclear norm. Some other examples of atomic norms can be found in [Chandrasekaran *et al.*, 2012]. At the following, we define a new atomic set A , and show that TNN is also an atomic norm induced by such an atomic set.

Let D be a set of the following matrices, i.e., $\mathbf{D} \in D$ where

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & & & \\ & \mathbf{D}_2 & & \\ & & \ddots & \\ & & & \mathbf{D}_{n_3} \end{bmatrix} \in \mathbb{C}^{n_1 n_3 \times n_2 n_3},$$

where $\mathbf{D}_i \in \mathbb{C}^{n_1 \times n_2}$ and there exists k such that $\mathbf{D}_k \neq 0$, $\text{rank}(\mathbf{D}_k) = 1$, $\|\mathbf{D}_k\|_F = 1$, and $\mathbf{D}_j = 0$, for all $j \neq k$. Then, for any $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we have

$$\|\bar{\mathcal{A}}\|_* = \inf \left\{ \sum_{\bar{\mathbf{D}} \in D} c_{\bar{\mathbf{D}}} : \bar{\mathcal{A}} = \sum_{\bar{\mathbf{D}} \in D} c_{\bar{\mathbf{D}}} \bar{\mathbf{D}}, c_{\bar{\mathbf{D}}} \geq 0, \forall \bar{\mathbf{D}} \in D \right\}.$$

Above we use the property of the rank one matrix decomposition of a matrix. This is equivalent to

$$\begin{aligned} \|\bar{\mathcal{A}}\|_* &= \inf \left\{ \sum_{\bar{\mathbf{D}} \in D} c_{\bar{\mathbf{D}}} : \bar{\mathcal{A}} = \sum_{\bar{\mathbf{D}} \in D} c_{\bar{\mathbf{D}}} \bar{\mathbf{D}}, c_{\bar{\mathbf{D}}} \geq 0, \forall \bar{\mathbf{D}} \in D \right\} \\ &= \inf \left\{ \sum_{\bar{\mathbf{D}} \in D} c_{\bar{\mathbf{D}}} : \mathcal{A} = \sum_{\bar{\mathbf{D}} \in D} c_{\bar{\mathbf{D}}} \bar{\mathbf{D}}, c_{\bar{\mathbf{D}}} \geq 0, \forall \bar{\mathbf{D}} \in D \right\}, \end{aligned} \quad (4)$$

where (4) uses the linear property of the inverse discrete Fourier transformation along the third dimension of a three way tensor. Motivated by (4), we define the atomic set A as

$$A = \{ \mathcal{W} \in \mathbb{C}^{n_1 \times n_2 \times n_3} : \mathcal{W} = n_3 \mathcal{D}, \bar{\mathbf{D}} \in D \}. \quad (5)$$

By $\|\mathcal{A}\|_* = \frac{1}{n_3} \|\bar{\mathcal{A}}\|_*$, we have the following result.

Theorem 2. *Let A be the set defined as in (5). The atomic norm $\|\mathcal{A}\|_A$ is TNN, i.e.,*

$$\begin{aligned} \|\mathcal{A}\|_* &= \|\mathcal{A}\|_A \\ &= \inf \left\{ \sum_{\mathcal{W} \in A} c_{\mathcal{W}} : \mathcal{A} = \sum_{\mathcal{W} \in A} c_{\mathcal{W}} \mathcal{W}, c_{\mathcal{W}} \geq 0, \forall \mathcal{W} \in A \right\}. \end{aligned}$$

For any $\mathcal{W} \in A$, we have $\|\mathcal{W}\|_* = n_3 \|\mathcal{D}\|_* = \|\bar{\mathcal{D}}\|_* = 1$. So the convex hull $\text{conv}(A)$ is the TNN ball in which TNN is less than or equal to one. Interpreting TNN as a special atomic norm by choosing a proper atomic set is crucial for the low-rank tensor recovery guarantee.

4 Low-rank Tensor Recovery from Gaussian Measurements

The Corollary 3.3 in [Chandrasekaran *et al.*, 2012] shows that \mathbf{x}_0 is the unique solution to problem (1) with high probability provided $m \geq \omega^2(T_A(\mathbf{x}_0) \cap \mathbb{S}^{p-1}) + 1$. Here, $T_A(\mathbf{x}_0)$ is the tangent cone at $\mathbf{x}_0 \in \mathbb{R}^p$, \mathbb{S}^{p-1} is the unit sphere, and $\omega(S)$ is the Gaussian width of a set S , defined as

$$\omega(S) = \mathbb{E}_{\mathbf{g}} \left[\sup_{\mathbf{z} \in S} \mathbf{g}^T \mathbf{z} \right],$$

where \mathbf{g} is a vector of independent zero-mean unit-variance Gaussians. To apply such a result for our low tubal rank recovery, we need to estimate the Gaussian width of our atomic set A defined in (5).

Theorem 3. *Let $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a tubal rank r tensor and A in (5). We have that*

$$\omega(T_A(\mathcal{M}) \cap \mathbb{S}^{n_1 n_2 n_3 - 1}) \leq \sqrt{3r(n_1 + n_2 - r)n_3}. \quad (6)$$

Now, by using (6) and the Corollary 3.3 in [Chandrasekaran *et al.*, 2012], we have the following main result.

Theorem 4. *Let $\Phi : \mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \mathbb{R}^n$ be a random map with i.i.d. zero-mean Gaussian entries having variance $\frac{1}{m}$ and $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a tensor of tubal rank r . Then, with high probability, we have:*

(1) *exact recovery: $\hat{\mathcal{X}} = \mathcal{M}$, where $\hat{\mathcal{X}}$ is the unique optimum of (3), provided that $m \geq 3r(n_1 + n_2 - r)n_3 + 1$;*

(2) **robust recovery:** $\|\hat{\mathcal{X}} - \mathcal{M}\|_F \leq \frac{2\delta}{\epsilon}$, where $\hat{\mathcal{X}}$ is optimal to

$$\hat{\mathcal{X}} = \arg \min_{\mathcal{X}} \|\mathcal{X}\|_*, \text{ s.t. } \|\mathbf{y} - \Phi(\mathcal{X})\|_2 \leq \delta, \quad (7)$$

$$\text{provided that } m \geq \frac{3r(n_1+n_2-r)n_3+3/2}{(1-\epsilon)^2}.$$

The above theorem shows that the tensor with tubal rank r can be recovered exactly by solving the convex program (3) or approximately by (7) when the required number of measurements is of the order $O(r(n_1 + n_2 - r)n_3)$. Note that such a rate is optimal compared with the degrees of freedom of a tensor with tubal rank r .

Theorem 5. A $n_1 \times n_2 \times n_3$ sized tensor with tubal rank r has at most $r(n_1 + n_2 - r)n_3$ degrees of freedom.

It is worth mentioning that the guarantee for low tubal rank tensor recovery in Theorem 4 is an extension of the low matrix guarantee in [Recht *et al.*, 2010; Chandrasekaran *et al.*, 2012]. If $n_3 = 1$, the tensor \mathcal{X} reduces to a matrix, the tensor tubal rank reduces to the matrix rank, and TNN reduces to the matrix nuclear norm. Thus the convex program (3) and the theoretical guarantee in Theorem 4 include the low rank matrix recovery model and guarantee as special cases, respectively. Compared with the existing low rank tensor recovery guarantees (based on different tensor ranks, e.g., [Mu *et al.*, 2014]) which are not order optimal, our guarantee enjoys the same optimal rate as the matrix case and our model (3) is computable.

5 Exact Tensor Completion Guarantee

Theorem 4 gives the recovery guarantee of program (3) based on the Gaussian measurements. In this section, we consider the tensor completion problem which is a special case of (3) but based on the uniform random mapping. Suppose that $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\text{rank}_t(\mathcal{M}) = r$. We consider the Bernoulli model in this work: the entries of \mathcal{M} are independently observed with probability p . We denote the set of the indices of the observed entries as Ω . We simply denote $\Omega \sim \text{Ber}(p)$. Then, the tensor completion problem asks for recovering \mathcal{M} from the observations $\{\mathcal{M}_{i,j,k} \mid (i,j,k) \in \Omega\}$. We can solve this problem by solving the following program

$$\min_{\mathcal{X}} \|\mathcal{X}\|_*, \text{ s.t. } \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{M}), \quad (8)$$

where $\mathcal{P}_\Omega(\mathcal{X})$ denotes the projection of \mathcal{X} on the observed set Ω . The above model extends the matrix completion task by convex nuclear norm minimization [Candès and Recht, 2009]. To guarantee the exact recovery, we need the following tensor incoherence conditions [Lu *et al.*, 2018a]

$$\max_{i=1, \dots, n_1} \|\mathcal{U}^* * \hat{\mathbf{e}}_i\|_F \leq \sqrt{\frac{\mu r}{n_1 n_3}}, \quad (9)$$

$$\max_{j=1, \dots, n_2} \|\mathcal{V}^* * \hat{\mathbf{e}}_j\|_F \leq \sqrt{\frac{\mu r}{n_2 n_3}}, \quad (10)$$

where $\hat{\mathbf{e}}_i$ denotes the tensor column basis, which is a tensor of size $n \times 1 \times n_3$ with its $(i, 1, 1)$ -th entry equaling 1 and the rest equaling 0. We also define the tensor tube basis $\hat{\mathbf{e}}_k$, 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. Denote $n_{(1)} = \max(n_1, n_2)$ and $n_{(2)} = \min(n_1, n_2)$.

$r = \text{rank}_t(\mathcal{X}_0) = 0.2n$				
n	$\text{rank}_t(\mathcal{X}_0)$	m	$\text{rank}_t(\hat{\mathcal{X}})$	$\frac{\ \hat{\mathcal{X}} - \mathcal{X}_0\ _F}{\ \mathcal{X}_0\ _F}$
10	2	541	2	1.2e-9
20	4	2161	4	1.6e-9
30	6	4861	6	1.5e-9
$r = \text{rank}_t(\mathcal{X}_0) = 0.3n$				
n	$\text{rank}_t(\mathcal{X}_0)$	m	$\text{rank}_t(\hat{\mathcal{X}})$	$\frac{\ \hat{\mathcal{X}} - \mathcal{X}_0\ _F}{\ \mathcal{X}_0\ _F}$
10	3	766	3	1.6e-9
20	6	3061	6	1.2e-9
30	9	6886	9	1.2e-9

Table 1: Exact low tubal rank tensor recovery from Gaussian measurements with sufficient number of measurements.

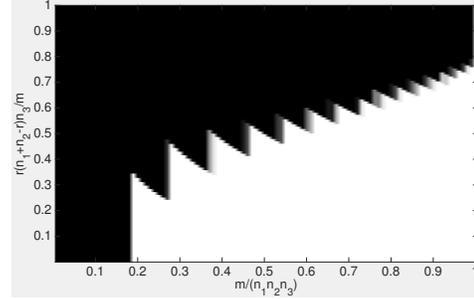


Figure 2: Phase transitions for low tubal rank tensor recovery from Gaussian measurements. Fraction of correct recoveries is across 10 trials, as a function of $\frac{r(n_1+n_2-r)n_3}{m}$ (y-axis) and sampling rate $\frac{m}{n_1 n_2 n_3}$. In this test, $n_1 = n_2 = 30, n_3 = 5$.

Theorem 6. Let $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with $\text{rank}_t(\mathcal{M}) = r$ and the skinny t -SVD be $\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$. Suppose that the indices $\Omega \sim \text{Ber}(p)$ and the tensor incoherence conditions (9)-(10) hold. There exist universal constants $c_0, c_1, c_2 > 0$ such that if

$$p \geq \frac{c_0 \mu r \log^2(n_{(1)} n_3)}{n_{(2)} n_3},$$

then \mathcal{M} is the unique solution to (8) with probability at least $1 - c_1(n_1 + n_2)^{-c_2}$.

Theorem 6 shows that, to recover a $n_1 \times n_2 \times n_3$ sized tensor with tubal rank r , the sampling complexity is $O(rn_{(1)}n_3 \log^2(n_{(1)}n_3))$. Such a bound is tight compared with the degrees of freedom¹.

6 Experiments

In this section, we conduct experiments to first verify the exact recovery guarantee in Theorem 4 for (3) from Gaussian measurements, then to verify the exact recovery guarantee in Theorem 6 for tensor completion (8). Both (3) and (8) can be solved by the standard ADMM [Lu *et al.*, 2018b]².

6.1 Exact Recovery from Gaussian Measurements

To verify Theorem 4, we can reformulate (3) as

$$\hat{\mathcal{X}} = \arg \min_{\mathcal{X}} \|\mathcal{X}\|_*, \text{ s.t. } \mathbf{y} = \text{Avec}(\mathcal{X}), \quad (11)$$

¹The proofs in [Zhang and Aeron, 2017] for tensor completion have several errors. Their used TNN definition is different from ours.

²The codes of our methods can be found at <http://sites.google.com/site/canyilu/>

where $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathbf{A} \in \mathbb{R}^{m \times (n_1 n_2 n_3)}$, $\mathbf{y} \in \mathbb{R}^m$ and $\text{vec}(\mathcal{X})$ denotes the vectorization of \mathcal{X} . The elements of \mathbf{A} are with i.i.d. zero-mean Gaussian entries having variance $1/m$. Thus, $\text{Avec}(\mathcal{X})$ gives the linear map $\Phi(\mathcal{X})$.

First, we test on random tensors, provided sufficient number of measurements as suggested in Theorem 4. We generate $\mathcal{X}_0 \in \mathbb{R}^{n \times n \times n}$ of tubal rank r by $\mathcal{X}_0 = \mathcal{P} * \mathcal{Q}$, where $\mathcal{P} \in \mathbb{R}^{n \times r \times n}$ and $\mathcal{Q} \in \mathbb{R}^{r \times n \times n}$ are with i.i.d. standard Gaussian random variables. We generate $\mathbf{A} \in \mathbb{R}^{m \times (n^2 n_3)}$ with its entries being i.i.d., zero-mean, $\frac{1}{m}$ -variance Gaussian variables. Then, let $\mathbf{y} = \text{Avec}(\mathcal{X}_0)$. We choose $n = 10, 20, 30$, $n_3 = 5$, $r = 0.2n$ and $r = 0.3n$. We set the number of measurements $m = 3r(2n - r)n_3 + 1$ as in Theorem 4. The results are given in Table 1, in which $\hat{\mathcal{X}}$ is the solution to (11). It can be seen that the relative errors $\|\hat{\mathcal{X}} - \mathcal{X}_0\|_F / \|\mathcal{X}_0\|_F$ are very small and the tubal ranks of $\hat{\mathcal{X}}$ are correct. Thus, this experiment verifies Theorem 4 for low tubal rank tensor recovery from Gaussian measurements.

Second, we exam the phase transition phenomenon in tubal rank r and the number of measurements m . We set $n_1 = n_2 = 30$ and $n_3 = 5$. We vary m between 1 and $n_1 n_2 n_3$ where the tensor is completely discovered. For a fixed m , we generate all possible tubal ranks such that $r(n_1 + n_2 - r)n_3 \leq m$. For each (m, r) pair, we repeat the following procedure 10 times. We generate \mathcal{X}_0 , \mathbf{A} , \mathbf{y} in the same way as the first experiment above. We declare \mathcal{X}_0 to be recovered if $\|\hat{\mathcal{X}} - \mathcal{X}_0\|_F / \|\mathcal{X}_0\|_F \leq 10^{-3}$. Figure 2 plots the fraction of correct recovery for each pair. The color of the cell in the figure reflects the empirical recovery rate of the 10 runs (scaled between 0 and 1). In all experiments, white denotes perfect recovery, while black denotes failure. It can be seen that there is a large region in which the recovery is correct. When the underlying tubal rank r of \mathcal{X}_0 is relatively larger, the required number of measurements for correct recovery is also larger. Such a result is consistent with our theoretical result. Similar phenomenon can be found in low-rank matrix recovery [Chandrasekaran *et al.*, 2012].

6.2 Exact Tensor Completion

First, we verify the exact tensor completion guarantee in Theorem 6 on random data. We generate $\mathcal{M} \in \mathbb{R}^{n \times n \times n}$ with tubal rank r by $\mathcal{M} = \mathcal{P} * \mathcal{Q}$, where the entries of $\mathcal{P} \in \mathbb{R}^{n \times r \times n}$ and $\mathcal{Q} \in \mathbb{R}^{r \times n \times n}$ are independently sampled from an $\mathcal{N}(0, 1/n)$ distribution. Then we sample $m = pn^3$ elements uniformly from \mathcal{M} to form the known samples. A useful quantity for reference is the number of degrees of freedom $d_r = r(2n - r)n$. The results in Table 1 shows that program (8) gives the correct recovery in the sense that the relative errors are small, less than 10^{-5} and the tubal ranks of the obtained solution are correct. These results well verify the recovery guarantee in Theorem 6.

Second, we examine the recovery phenomenon with varying tubal rank of \mathcal{M} and varying sampling rate p . We consider two sizes of $\mathcal{M} \in \mathbb{R}^{n \times n \times n}$: (1) $n = 40$; (2) $n = 50$. We generate $\mathcal{M} = \mathcal{P} * \mathcal{Q}$, where the entries of $\mathcal{P} \in \mathbb{R}^{n \times r \times n}$ and $\mathcal{Q} \in \mathbb{R}^{r \times n \times n}$ are independently sampled from an $\mathcal{N}(0, 1/n)$ distribution. We set $m = pn^3$. We choose p in $[0.01 : 0.01 : 0.99]$ and $r = 1, 2, \dots, 30$ in the case $n = 40$, and $r = 1, 2, \dots, 35$ in the case $n = 50$.

$$\mathcal{X}_0 \in \mathbb{R}^{n \times n \times n}, r = \text{rank}_t(\mathcal{X}_0), m = pn^3, d_r = r(2n - r)n$$

n	r	$\frac{m}{d_r}$	p	$\text{rank}_t(\hat{\mathcal{X}})$	$\frac{\ \hat{\mathcal{X}} - \mathcal{X}_0\ _F}{\ \mathcal{X}_0\ _F}$
50	3	4	0.47	3	$3.9e-7$
50	5	3	0.57	5	$3.5e-7$
50	10	2	0.72	10	$4.1e-7$
100	5	4	0.39	5	$1.4e-6$
100	10	3	0.57	10	$9.2e-7$
100	15	2	0.56	15	$8.4e-7$
200	5	4	0.20	5	$4.2e-6$
200	10	3	0.29	10	$3.2e-6$
200	20	2	0.38	20	$3.1e-6$
300	10	4	0.26	10	$5.1e-6$
300	20	3	0.39	20	$4.2e-6$
300	30	3	0.57	30	$2.9e-6$

Table 2: Exact tensor completion on random data.

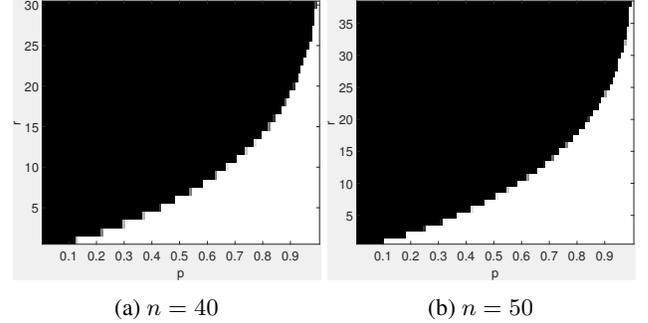


Figure 3: Phase transitions for tensor completion. Fraction of correct recoveries is across 10 trials, as a function of tubal rank r (y-axis) and sampling rate p (x-axis). The results are shown for different sizes of $\mathcal{M} \in \mathbb{R}^{n \times n \times n}$: (a) $n = 40$; (b) $n = 50$.

For each (r, p) triple, we simulate 10 test instances and declare a trial to be successful if the recovered $\hat{\mathcal{X}}$ satisfies $\|\hat{\mathcal{X}} - \mathcal{M}\|_F / \|\mathcal{M}\|_F \leq 10^{-3}$. Figure 3 plots the fraction of correct recovery for each triple (black = 0% and white = 100%). It can be seen that there is a large region in which the recovery is correct. Interestingly, the experiments reveal very similar plots for different n , suggesting that our asymptotic conditions for recovery may be conservative. Such a phenomenon is also consistent with the result in Theorem 6 which shows that the recovery is correct when the sampling rate p is not small and the tubal rank r is relatively low.

7 Conclusion

This paper first considers the exact guarantee of TNN minimization for low tubal rank tensor recovery from Gaussian measurements. We prove that TNN is a new instance of the atomic norm associated with certain atomic set. From the perspective of atomic norm minimization, we give the optimal estimation of the required measurements for the exact low tubal rank tensor recovery. Second, we give the exact recovery guarantee of TNN minimization for tensor completion. This result fixes the errors in the proofs of [Zhang and Aeron, 2017]. Numerical experiments verify our theoretical results.

By treating TNN as an instance of the atomic norm, we can get more results of low tubal rank recovery by using existing results, e.g., [Foygel and Mackey, 2014; Amelunxen *et al.*, 2014]. Beyond the study on the convex TNN, it is also interesting to study the nonconvex models [Lu *et al.*, 2015].

Acknowledgements

J. Feng is partially supported by National University of Singapore startup grant R-263-000-C08-133 and Ministry of Education of Singapore AcRF Tier One grant R-263-000-C21-112. Z. Lin was supported by National Basic Research Program of China (973 Program) (grant no. 2015CB352502), National Natural Science Foundation (NSF) of China (grant nos. 61625301 and 61731018), Qualcomm, and Microsoft Research Asia.

References

- [Amelunxen *et al.*, 2014] Dennis Amelunxen, Martin Lotz, Michael B McCoy, and Joel A Tropp. Living on the edge: Phase transitions in convex programs with random data. *Information and Inference*, pages 224–294, 2014.
- [Candès and Recht, 2009] E.J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.
- [Candès *et al.*, 2006] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *TIT*, 52(2):489–509, 2006.
- [Chandrasekaran *et al.*, 2012] Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky. The convex geometry of linear inverse problems. *Foundations of Computational mathematics*, 12(6):805–849, 2012.
- [Chen, 2015] Yudong Chen. Incoherence-optimal matrix completion. *TIT*, 61(5):2909–2923, May 2015.
- [Donoho and Tanner, 2009] David Donoho and Jared Tanner. Counting faces of randomly projected polytopes when the projection radically lowers dimension. *Journal of the American Mathematical Society*, 22(1):1–53, 2009.
- [Foygel and Mackey, 2014] Rina Foygel and Lester Mackey. Corrupted sensing: Novel guarantees for separating structured signals. *TIT*, 60(2):1223–1247, 2014.
- [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.
- [Lu *et al.*, 2015] Canyi Lu, Changbo Zhu, Chunyan Xu, Shuicheng Yan, and Zhouchen Lin. Generalized singular value thresholding. In *AAAI*, 2015.
- [Lu *et al.*, 2016] Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In *CVPR*. IEEE, 2016.
- [Lu *et al.*, 2018a] Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. Tensor robust principal component analysis with a new tensor nuclear norm. *arXiv preprint arXiv:1804.03728*, 2018.
- [Lu *et al.*, 2018b] Canyi Lu, Jiashi Feng, Shuicheng Yan, and Zhouchen Lin. A unified alternating direction method of multipliers by majorization minimization. *TPAMI*, 40(3):527–541, 2018.
- [Mu *et al.*, 2014] Cun Mu, Bo Huang, John Wright, and Donald Goldfarb. Square deal: Lower bounds and improved relaxations for tensor recovery. In *ICML*, pages 73–81, 2014.
- [Recht *et al.*, 2010] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.
- [Zhang and Aeron, 2017] Zemin Zhang and Shuchin Aeron. Exact tensor completion using t-SVD. *TSP*, 65(6):1511–1526, 2017.