

Recovery of Corrupted Low-Rank Tensors

Haiyan Fan, Gangyao Kuang

Department of Electronic Science and Engineering, National University of Defense Technology, Changsha, China

Email: hy_fan@yeah.net

How to cite this paper: Fan, H.Y. and Kuang, G.Y. (2017) Recovery of Corrupted Low-Rank Tensors. *Applied Mathematics*, 8, 229-244.

<https://doi.org/10.4236/am.2017.82019>

Received: January 30, 2017

Accepted: February 25, 2017

Published: February 28, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper studies the problem of recovering low-rank tensors, and the tensors are corrupted by both impulse and Gaussian noise. The problem is well accomplished by integrating the tensor nuclear norm and the l_1 -norm in a unified convex relaxation framework. The nuclear norm is adopted to explore the low-rank components and the l_1 -norm is used to exploit the impulse noise. Then, this optimization problem is solved by some augmented-Lagrangian-based algorithms. Some preliminary numerical experiments verify that the proposed method can well recover the corrupted low-rank tensors.

Keywords

Low-Rank Tensor, Tensor Recovery, Augmented Lagrangian Method, Impulsive Noise, Mixed Noise

1. Introduction

The problem of exploiting low-dimensional structures in high-dimensional data is taking on increasing importance in image, text and video processing, and web search, where the observed data lie in very high dimensional spaces. The principal component analysis (PCA) proposed in [1] is the most widely used tool for low-rank component extraction and dimensionality reduction. It is easily implementable and efficient for data mildly corrupted by small noise. However, this PCA method is sensitive to data corrupted by heavy impulse noise or outlying observations. Then, a number of robust PCA methods were proposed. But none of these approaches yield a polynomial-time algorithm with strong performance guarantees under broad conditions. The proposed Robust PCA [2] is one among the earliest polynomial-time algorithms. Assume that a data matrix $X \in \mathbb{R}^{m \times n_2}$ consists of a low-rank matrix A_0 and a sparse matrix E_0 . Then, A_0 and E_0 can be recovered with a high probability by solving the following convex relaxation problem (if A_0 is low-rank and satisfies some incoherent conditions, E_0 is sufficiently sparse):

$$\begin{aligned} & \min_{A,E} \|A\|_* + \lambda \|E\|_1 \\ & \text{s.t. } X = A + E \end{aligned} \tag{1}$$

where $\|A\|_*$ denotes the nuclear norm of A and $\|E\|_1$ denotes the l_1 -norm of E . Nuclear norm and l_1 -norm are used to induce low rank and sparsity, specifically. $\lambda > 0$ is a parameter balancing the low rank and sparsity. Candes *et al.* [2] prove that $\lambda = 1/\sqrt{\max(n_1, n_2)}$ works with a high probability for recovering any low-rank, incoherent matrix. Notably, Chandrasekaran *et al.* [3] also consider the problem of decomposing a given data matrix into sparse and low-rank components, and give sufficient conditions for convex programming to succeed. Their work was motivated by applications in system identification and learning of graphical models. In contrast, Candes *et al.* [2] were motivated by robust principal component computations in high dimensional settings when there were erroneous and missing entries; missing entries were not considered in [3]. In [3], the parameter λ is data-dependent, and may have to be selected by solving a number of convex programs, while Candes *et al.* [2] provide a universal value of λ , namely, $\lambda = 1/\sqrt{\max(n_1, n_2)}$. The distinction between the two results is a consequence of different assumptions about the origin of the data matrix X .

In many real world applications, we need to consider the model defined in Equation (1) under more complicated circumstance [4] [5]. First, only a fraction of entries of X can be observed. This is the well-known matrix completion problem. If the unknown matrix is known to have low rank or approximately low rank, then accurate and even exact recovery is possible by nuclear norm minimization [6] [7]. Second, the observed data are corrupted by both impulse noise (sparse but large) and Gaussian noise (small but dense). Let X be the superposition of low-rank matrix A , the impulse noise matrix E and the Gaussian noise F , *i.e.*, $X = A + E + F$. The Gaussian noise of the observed entries is assumed to be small in the sense that $\|P_\Omega(F)\|_F \leq \delta$, where δ is the Gaussian noise level and $\|\cdot\|_F$ is the Frobenius norm. Then, to be broadly applicable, we consider the following extension of model defined in Equation (1):

$$\begin{aligned} & \min_{A,E} \|A\|_* + \lambda \|E\|_1 \\ & \text{s.t. } \|P_\Omega(X - A - E)\|_F \leq \delta \end{aligned} \tag{2}$$

where Ω is a subset of the index set of entries $\{1, 2, \dots, n_1\} \times \{1, 2, \dots, n_2\}$. It's assumed that only these entries $\{X - ij, (i, j) \in \Omega\}$ can be observed. The operator $P_\Omega : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^{n_1 \times n_2}$ is a orthogonal projection onto the span of matrices vanishing outside of Ω so that the ij -th entry of $P_\Omega(X)$ is X_{ij} if $(i, j) \in \Omega$ and zero otherwise. The problem defined in Equation (2) can be solved by the classical Augmented Lagrangian Method (ALM). The separable structure emerging in the objective function and the constraints entails the idea of splitting the corresponding augmented Lagrangian function to derive more efficient numerical algorithms. Tao *et al.* [5] developed the Alternating Splitting Augmented Lagrangian Method (ASALM) and its variant (VASALM), and the Pa-

parallel Splitting Augmented Lagrangian method (PSALM) for solving Equation (2).

One shortcoming of model defined in Equation (2) is that it can only handle matrix (two-way) data. However, the real-world data are ubiquitously in multi-way, also referred to as tensor. For example, a color image is a 3-way object with column, row and color modes; a greyscale video is indexed by two spatial variables and one temporal variable. If we use the model defined in Equation (2) to process the tensor data, we have to unfold the multi-way data into a matrix. Such a preprocessing usually leads to the loss of the inherent structure high-dimensional information in the original observations. To avoid this negative factor, a common approach is to manipulate the tensor data by taking the advantage of its multi-dimensional structure. Tensor analysis have many applications in computer vision [8], diffusion Magnetic Resonance Imaging (MRI) [9] [10] [11], quantum entanglement problem [12], spectral hypergraph theory [13] and higher-order Markov chains [14].

The goal of this paper is to study the Tensor Robust PCA which aims to accurately recover a low-rank tensor from impulse and Gaussian noise. The observations can also be incomplete. Tensors of low rank appear in a variety of applications such as video processing ($d = 3$) [15], time-dependent 3D imaging ($d = 4$), ray tracing where the material dependent bidirectional reflection function is an order four tensor that has to be determined from measurements [15], numerical solution of the electronic Schrödinger equation ($d = 3N$, where N is the number of particles) [16] [17] [18], machine learning [19] and more. More specifically, suppose we are given a data tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ (d is the number of ways), and it can be decomposed as

$$\mathcal{X} = \mathcal{A}_0 + \mathcal{E}_0 + \mathcal{F}_0 \quad (3)$$

where \mathcal{A}_0 is low-rank and \mathcal{E}_0 is sparse. \mathcal{F}_0 is Gaussian noise with the noise level being δ . Then, we try to recover the low-rank \mathcal{A}_0 through the following convex relaxation problem:

$$\begin{aligned} \min_{\mathcal{A}, \mathcal{E}} & \|\mathcal{A}\|_* + \lambda \|\mathcal{E}\|_1 \\ \text{s.t.} & \|\mathcal{P}_\Omega(\mathcal{X} - \mathcal{A} - \mathcal{E})\|_F \leq \delta \end{aligned} \quad (4)$$

1.1. Related Work

Although the recovery of low-rank matrix has been well studied, the research of low-rank tensor recovery is still lacking. This is mainly because it's difficult to define a satisfactory tensor rank which enjoys similar good properties as the matrix case. Several different definitions of tensor rank have been proposed but each has its limitation. For example, the CP rank [20] is defined as the smallest number of rank one tensors that sum up to \mathcal{X} , where a rank one tensor is of the form $\mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \dots \otimes \mathbf{u}_d$. Expectedly, the CP rank is NP-hard to compute. Its convex relaxation is also intractable. Another more popular direction is to use the tractable Tucker rank [20] and its convex relaxation. For a d -way tensor \mathcal{X} , the Tucker rank is a vector defined as

$$\text{rank}_{tc}(\mathcal{X}) := \left(\text{rank}(\mathbf{X}^{(1)}), \text{rank}(\mathbf{X}^{(2)}), \dots, \text{rank}(\mathbf{X}^{(d)}) \right),$$

where $\mathbf{X}^{(i)}$ is the mode- i matricization of \mathcal{X} . Motivated by the fact that the nuclear norm is the convex envelop of the matrix rank within the unit ball of the spectral norm. The Sum of Nuclear Norms (SNN), defined as $\sum_i \|\mathbf{X}^{(i)}\|_*$, is used as a convex surrogate of the Tucker rank. This approach is effective, but SNN is not a tight convex relaxation of Tucker rank.

More recently, the work [21] proposed the tensor tubal rank based on a new tensor decomposition scheme denoted as t-SVD. The t-SVD is based on a new tensor-tensor product which enjoys many similar properties as the matrix case. Based on the computable t-SVD, the tensor nuclear norm [22] is used to replace the tubal rank for low-rank tensor recovery (from incomplete/corrupted tensors). The problem of recovering the unknown low-rank tensor \mathcal{F} from incomplete samples \mathcal{F}_0 can be solved by the following convex program [21],

$$\begin{aligned} \min_{\mathcal{A}} \|\mathcal{A}\|_* \\ \text{s.t. } \mathcal{P}_{\Omega}(\mathcal{A}) = \mathcal{P}_{\Omega}(\mathcal{A}_0) \end{aligned} \quad (5)$$

where $\|\cdot\|_*$ is the nuclear norm of \mathcal{A} , Ω is the index set of known elements in the original tensor, and \mathcal{P}_{Ω} is the projector onto the span of tensors. Lu *et al.* [23] extended the work and studied the Tensor Robust Principal Component (TRPCA) problem, as defined in the following equation:

$$\begin{aligned} \min_{\mathcal{A}, \mathcal{E}} \|\mathcal{A}\|_* + \lambda \|\mathcal{E}\|_1 \\ \text{s.t. } \mathcal{X} = \mathcal{A} + \mathcal{E} \end{aligned} \quad (6)$$

In this work, we go one step further, and consider recovering low-rank and sparse components of tensors from incomplete and noisy observations as defined in Equation (4).

1.2. Paper Contribution

The contributions of this work are two-fold:

- A unified convex relaxation framework is proposed for the problem of recovering low-rank and sparse components of tensors from incomplete and noisy observations. Three augmented-Lagrangian-based algorithms are developed for the optimization problem.
- Numerical experiments on synthetical data validate the efficacy of our proposed denoising approach.

1.3. Paper Organization

The rest of the paper is organized as follows. In Section 2, some preliminaries that are useful for the subsequent analysis are provided. In Section 3, three augmented-Lagrangian-based methods are developed for the problem defined in Equation (4). In Section 4, some numerical experiments verify the justification of the model defined in Equation (4) and the efficiency of the proposed algorithms. Finally, in Section 5, we make some conclusions and discuss some topics

for future work.

2. Preliminaries

In this section, we list some lemmas concerning the shrinkage operators, which will be used at each iteration of the proposed augmented Lagrangian type methods to solve the generated subproblems.

Lemma 1. For $\tau > 0$, and $T \in \mathbb{R}^{n_1 \times n_2}$, the solution of the following problem (7) obeys

$$\arg \min_S \left\{ \frac{1}{2} \|S - T\|_F^2 + \tau \|S\|_1 \right\} \tag{7}$$

is given by $\text{shrink}(T, \tau)$. $\text{shrink}(\cdot, \cdot)$ is a soft shrinkage operator and defined as:

$$\text{shrink}(a, \kappa) \begin{cases} a - \kappa & a > \kappa \\ 0 & |a| \leq \kappa \\ a + \kappa & a < -\kappa \end{cases} \tag{8}$$

Lemma 2. Consider the singular value decomposition (SVD) of a matrix $A \in \mathbb{R}^{n_1 \times n_2}$ of rank r .

$$A = Q^* S^* V, \quad S = \text{diag}(\{\sigma_i\}_{1 \leq i \leq r}) \tag{9}$$

where $Q \in \mathbb{R}^{n_1 \times r}$ and $V \in \mathbb{R}^{n_2 \times r}$ are orthogonal, and the singular values σ_i are real and positive. Then, for all $\tau > 0$, define the soft-thresholding operator \mathcal{D} ,

$$\mathcal{D}_\tau(A) := Q^* \mathcal{D}_\tau(S) V, \quad \mathcal{D}_\tau(S) = \text{diag}(\{(\sigma_i - \tau)_+\}_{1 \leq i \leq r}) \tag{10}$$

where x_+ is the operator that $x_+ = \max(0, x)$. Then, for each $\tau > 0$ and $B \in \mathbb{R}^{n_1 \times n_2}$, the singular value shrinkage operator (10) obeys

$$\mathcal{D}_\tau(A) = \arg \min_B \left\{ \frac{1}{2} \|B - A\|_F^2 + \tau \|B\|_* \right\} \tag{11}$$

3. Algorithm

An alternative model to study the problem defined in Equation (4) is the following nuclear-norm- and l_1 -norm- normalized least squares problem:

$$\min_{A, \mathcal{E}} \|A\|_* + \lambda_1 \|\mathcal{E}\|_1 + \lambda_2 \|\mathcal{P}_\Omega(\mathcal{X} - A - \mathcal{E})\|_F^2 \tag{12}$$

Equation (12) can be reformulated into the following favourable form:

$$\begin{aligned} \min_{A, \mathcal{E}, \mathcal{F}} \|A\|_* + \lambda_1 \|\mathcal{E}\|_1 + \lambda_2 \|\mathcal{P}_\Omega(\mathcal{F})\|_F^2 \\ \text{s.t. } \mathcal{X} = A + \mathcal{E} + \mathcal{F} \end{aligned} \tag{13}$$

Alternating Direction Method of Multiplier (ADMM), which is an extension of ALM algorithm, can be used to solve the tensor recovery problem defined in (13). With given $(\mathcal{A}^k, \mathcal{E}^k, \mathcal{F}^k)$, the ADMM generate the new iterates via the following scheme:

$$\begin{cases} \mathcal{A}^{k+1} = \arg \min_{\mathcal{A}} \left(\|\mathcal{A}\|_* + \frac{\beta}{2} \left\| \mathcal{X} - (\mathcal{A} + \mathcal{E}^k + \mathcal{F}^k) + \frac{\Lambda_1^k}{\beta} \right\| \right) \\ \mathcal{E}^{k+1} = \arg \min_{\mathcal{E}} \left(\lambda_1 \|\mathcal{E}\|_1 + \frac{\beta}{2} \left\| \mathcal{X} - (\mathcal{A}^{k+1} + \mathcal{E} + \mathcal{F}^k) + \frac{\Lambda_1^k}{\beta} \right\| \right) \\ \mathcal{F}^{k+1} = \arg \min_{\mathcal{F} \in \mathcal{B}} \left(\lambda_2 \|\mathcal{F}\|_F^2 + \frac{\beta}{2} \left\| \mathcal{X} - (\mathcal{A}^{k+1} + \mathcal{E}^{k+1} + \mathcal{F}) + \frac{\Lambda_1^k}{\beta} \right\| \right) \\ \Lambda_1^{k+1} = \Lambda_1^k + \beta \left(\mathcal{X} - (\mathcal{A}^{k+1} + \mathcal{E}^{k+1} + \mathcal{F}^{k+1}) \right) \end{cases} \quad (14)$$

See **Algorithm 1** for the optimization details.

3.1. Stopping Criterion

It can be easily verified that the iterates generated by the proposed ADMM algorithm can be characterized by

$$\begin{cases} \mathbf{0} \in \partial \|\mathcal{A}^{k+1}\|_* - \left[\Lambda_1^k - \beta (\mathcal{A}^{k+1} + \mathcal{E}^k + \mathcal{F}^k - \mathcal{X}) \right] \\ \mathbf{0} \in \partial \left(\lambda_1 \|\mathcal{E}^{k+1}\|_1 \right) - \left[\Lambda_1^k - \beta (\mathcal{A}^{k+1} + \mathcal{E}^{k+1} + \mathcal{F}^k - \mathcal{X}) \right] \\ \mathbf{0} \in \partial \left(\lambda_2 \|\mathcal{F}^{k+1}\|_F^2 \right) - \left[\Lambda_1^k - \beta (\mathcal{A}^{k+1} + \mathcal{E}^{k+1} + \mathcal{F}^{k+1} - \mathcal{X}) \right] \\ \Lambda_1^{k+1} = \Lambda_1^k - \left[(\mathcal{A}^{k+1} + \mathcal{E}^{k+1} + \mathcal{F}^{k+1}) - \mathcal{X} \right] \end{cases} \quad (15)$$

which is equivalent to

$$\begin{cases} \mathbf{0} \in \partial \|\mathcal{A}^{k+1}\|_* - \Lambda_1^{k+1} + \beta (\mathcal{E}^k - \mathcal{E}^{k+1}) + \beta (\mathcal{F}^k - \mathcal{F}^{k+1}) \\ \mathbf{0} \in \partial \left(\lambda_1 \|\mathcal{E}^{k+1}\|_1 \right) - \Lambda_1^{k+1} + \beta (\mathcal{F}^k - \mathcal{F}^{k+1}) \\ \mathbf{0} \in \partial \left(\lambda_2 \|\mathcal{F}^{k+1}\|_F^2 \right) - \Lambda_1^{k+1} \\ \Lambda_1^{k+1} = \Lambda_1^k - \beta \left[(\mathcal{A}^{k+1} + \mathcal{E}^{k+1} + \mathcal{F}^{k+1}) - \mathcal{X} \right] \end{cases} \quad (16)$$

Algorithm 1. Optimization framework for problem defined in Equation (13).

Input: Observation \mathcal{X} , parameters λ_1, λ_2

Output: Useful signal \mathcal{A}

Initialization: Choose parameters $\eta > 1$ and $\beta_{\min}, \beta_{\max}$ with $0 < \beta_{\min} < \beta_{\max} < +\infty$, Initialize an iteration counter $k \leftarrow 0$ and a bounded starting point, Initialize $\beta^0 \in [\beta_{\min}, \beta_{\max}]$;

repeat

Update \mathcal{A}^{k+1} ,

Update \mathcal{E}^{k+1} ,

Update \mathcal{F}^{k+1} ,

Update penalty parameter $\beta^{k+1} \leftarrow \min(\eta\beta^k, \beta_{\max})$

Update Lagrangian multiplier $\Lambda_1^{k+1} \leftarrow \Lambda_1^k + \beta^k (\mathcal{X} - (\mathcal{A}^{k+1} + \mathcal{E}^{k+1} + \mathcal{F}^{k+1}))$,

if some stopping criterion is satisfied; **then**

Break;

else

$k \leftarrow k + 1$;

end if

until exceed the maximum number of outer loop.

Equation (16) shows that the distance of the iterates $(\mathcal{A}^{k+1}, \mathcal{E}^{k+1}, \mathcal{F}^{k+1})$ to the solution $(\mathcal{A}^*, \mathcal{E}^*, \mathcal{F}^*, \Lambda^*)$ can be characterized by $\beta(\|\mathcal{E}^k - \mathcal{E}^{k+1}\| + \|\mathcal{F}^k - \mathcal{F}^{k+1}\|)$ and $\frac{1}{\beta}\|\Lambda_1^k - \Lambda_1^{k+1}\|$. Thus, a straightforward stopping criterion for **Algorithm 1** is:

$$\min \left\{ \beta(\|\mathcal{E}^k - \mathcal{E}^{k+1}\| + \|\mathcal{F}^k - \mathcal{F}^{k+1}\|), \frac{1}{\beta}\|\Lambda_1^k - \Lambda_1^{k+1}\| \right\} \leq \epsilon \tag{17}$$

Here ϵ is an infinitesimal number, e.g., 10^{-6} .

3.2. Convergence Analysis

In this subsection, we mainly analyze the convergence of ADMM for solving problem defined in Equation (13). We denote $f_1(\cdot) = \|\cdot\|_*$, $f_2(\cdot) = \|\cdot\|_F^2$, and $f_3(\cdot) = \|\cdot\|_1$. $f_2(\cdot)$ is strongly convex, while $f_1(\cdot)$ and $f_3(\cdot)$ are convex terms, but may not be strongly convex. The problem defined in Equation (13) can be reformulated as

$$\begin{aligned} \min_{\mathcal{A}, \mathcal{E}, \mathcal{F}} & f_1(\mathcal{A}) + \lambda_2 f_2(\mathcal{F}) + \lambda_1 f_3(\mathcal{E}) \\ \text{s.t.} & \chi = \mathcal{A} + \mathcal{E} + \mathcal{F} \end{aligned} \tag{18}$$

Definition 1. (Convex and Strongly Convex) Let $f : \mathbb{R}^n \rightarrow [-\infty, +\infty]$, if the domain of f denoted by $f := \{\mathbf{x} \in \mathbb{R}^n, f(\mathbf{x}) < +\infty\}$ is not empty, f is considered to be proper. If for any $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$, we always have $f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y}), \forall t \in [0, 1]$, then it is considered that f is convex. Furthermore, f is considered to be strongly convex with the modulus $\mu > 0$, if

$$\begin{aligned} f(t\mathbf{x} + (1-t)\mathbf{y}) & \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y}) \\ & - \frac{1}{2}\mu t(1-t)\|\mathbf{x} - \mathbf{y}\|^2, \forall t \in [0, 1] \end{aligned} \tag{19}$$

Cai *et al.* [24] and Li *et al.* [25] have proved the convergence of Extended Alternating Direction Method of Multipliers (e-ADMM) with only one strongly convex function for the case $m = 3$.

Assumption 1. In Equation (18), f_1 and f_3 are convex, and f_2 is strongly convex with modulus $\mu_2 > 0$.

Assumption 2. The optimal solution set for the problem defined in Equation (18) is nonempty, i.e., there exist $(\mathcal{A}^*, \mathcal{E}^*, \mathcal{F}^*, \Lambda_1^*) \in \Omega^*$ such that the following requirements can be satisfied:

$$\nabla f_1(\mathcal{A}^*) - \Lambda^* = \mathbf{0}, \tag{20}$$

$$\lambda_2 \nabla f_2(\mathcal{F}^*) - \Lambda_1^* = \mathbf{0}, \tag{21}$$

$$\lambda_1 \nabla f_3(\mathcal{E}^*) - \Lambda_1^* = \mathbf{0}, \tag{22}$$

$$\mathcal{A}^* + \mathcal{E}^* + \mathcal{F}^* - \chi = \mathbf{0}, \tag{23}$$

Theorem 1. Assume that Assumption 1 and Assumption 2 hold. Let $(\mathcal{A}^k, \mathcal{E}^k, \mathcal{F}^k, \Lambda_1^k)$ be the sequence generated by Algorithm 1 for solving the problem

defined in Equation (18). If $\beta \in \left(0, \frac{6\mu_2}{13}\right)$, the limit point of $(\mathcal{A}^k, \mathcal{F}^k, \mathcal{E}^k, \Lambda_1^k)$ is an optimal solution to Equation (18). Moreover, the objective function converges to the optimal value and the constraint violation converges to zero, i.e.,

$$\lim_{k \rightarrow \infty} \left\| f_1(\mathcal{A}^*) + \lambda_2 f_2(\mathcal{F}^*) + \lambda_1 f_3(\mathcal{E}^*) - f^* \right\| = 0 \tag{24}$$

and

$$\lim_{k \rightarrow \infty} \left\| \chi - (\mathcal{A} + \mathcal{E} + \mathcal{F}) \right\| = 0 \tag{25}$$

where f^* denotes the optimal objective value for the problem defined in Equation (18). In our specific application, $\beta \in \left(0, \frac{6 * 2\lambda_2}{13}\right)$ can sufficiently ensure the convergence [24].

3.3. Parameter Choice

In our optimization framework given in Equation (13), there are three parameters β , λ_1 and λ_2 . As mentioned in Lu [23], λ_1 does not need any tuning and can be set to $1/\sqrt{n_{(1)}n_3}$, where $n_{(1)} = \max\{n_1, n_2\}$. Besides, the value of β is limited to the range $\beta \in \left(0, \frac{6 * 2\lambda_2}{13}\right)$ to ensure the convergence of our algorithm (based on the analysis in Theorem 1). Thus, the value of λ_2 is important for the performance of our algorithm. For simplicity, we consider the case when \mathcal{A} is only degraded by Gaussian noise \mathcal{F} without sparse noise \mathcal{E} , that is:

$$\min \frac{1}{2} \|\mathcal{F}\|_F^2 + \frac{1}{2\lambda_2} \|\mathcal{A}\|_* \quad \text{s.t. } \chi = \mathcal{A} + \mathcal{F} \tag{26}$$

The solution for Equation (26) is equal to χ but with singular values being shifted towards zero by soft thresholding. λ_2 should be set large enough to remove noise (i.e., to keep the variance low), and small to avoid over-shrinking of the original tensor \mathcal{A} (i.e., to keep the bias low). For the matrix case (i.e., $n_3 = 1$), Candes *et al.* [26] have deduced the proper value of λ_2 , as shown in the following theorem.

Theorem 2. *Supposing that the Gaussian noise term $\mathbf{F} \in \mathbb{R}^{n \times n}$, and each entry $n_{i,j}$ is iid normally distributed, we can have that for $\mathcal{N}(0, \sigma^2)$, $\|\mathbf{F}\|_F^2 \leq (n + \sqrt{8n})\sigma^2$ with high probability. Then, $\frac{1}{2\lambda_1} = \sqrt{(n + \sqrt{8n})}\sigma$. That is,*

$$\lambda_2 = \frac{1}{2\sqrt{(n + \sqrt{8n})}\sigma}.$$

Based on this conclusion, we derive the required conditions for convex program defined in Equation (13) to accurately recover the low-rank component \mathcal{A} from corrupted observations. Our derivations are given in the following main result.

Main Result 1. *Assume that the low-rank tensor $\mathcal{A}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ obeys the incoherence conditions [27]. The support set Ω of \mathcal{E}_0 is uniformly distributed*

among all sets of cardinality m . Then, there exist universal constants $c_1, c_2 > 0$ such that with probability at least $1 - c_1 n^{-c_2}$, $\mathcal{A}_0, \mathcal{E}_0$ is the unique minimizer to problem defined in Equation (13). The values of λ_1 and λ_2 can be determined as $\lambda_2 = \frac{1}{2\sqrt{n_{(1)}n_3} + \sqrt{8n_{(1)}n_3}\sigma}$ and $\lambda_1 = \frac{1}{\sqrt{n_{(1)}n_3}}$. In the same time, the rank of

\mathcal{A}_0 and the number of non-zero entries of \mathcal{E}_0 should satisfy that

$$\text{rank}_t(\mathcal{A}) \leq \frac{\rho_r n_{(2)}}{\mu_0 \left(\log(n_{(1)}n_3)\right)^2} \quad \text{and} \quad m \leq \rho_s n_1 n_2 n_3^2$$

where $n_{(1)} = \max\{n_1, n_2\}$ and $n_{(2)} = \min\{n_1, n_2\}$. ρ_r and ρ_s are positive constants. The value of penalty parameter β should be within the range of $\left(0, \frac{6 * 2\lambda_2}{13}\right)$ to ensure the convergence.

4. Experiments on Synthetic Data

In this section, we conduct synthetic data and real data experiments to corroborate our algorithm. We investigate the ability of our proposed Robust Low Rank Tensor Approximation (denoted as RLRTA) algorithm for recovering low-rank tensors of various tubal rank from noises of various sparsity and random Gaussian noise of different intensity.

4.1. Exact Recovery from Varying Sparsity of \mathcal{E}

We first verify the correct recovery performance of our algorithm for different sparsity of \mathcal{E} . Be similar to [23], we consider the tensors of size $n \times n \times n$, with varying dimensions $n = 100, 200$. We generate a $\text{rank}_t - r$ tensor $\mathcal{A} = \mathcal{Q} * \mathcal{V}$, where the entries of $\mathcal{Q} \in \mathbb{R}^{n \times r \times n}$ and $\mathcal{V} \in \mathbb{R}^{r \times n \times n}$ are independently sampled from a uniform distribution in interval $(0, 1/n)$. The support set Ω , with size $m = \rho_s n^3$ of sparse component \mathcal{E} is chosen uniformly at random. For all $(i, j, k) \in \Omega$, let $\mathcal{E}_{ijk} = \mathcal{M}_{ijk}$, where \mathcal{M} is a tensor with independent Bernoulli ± 1 entries. For \mathcal{E} , it can be mathematically expressed as

$$\mathcal{E}_{ijk} = \begin{cases} 1, & \text{w.p. } \rho_s/2 \\ 0 & \text{w.p. } 1 - \rho_s \\ -1 & \text{w.p. } \rho_s/2 \end{cases} \quad (27)$$

where w.p. is the abbreviation of “with probability”. We test on two settings: the first scenario with setting $r = \text{rank}_t(\mathcal{A}) = 0.1n$ and $\rho_s = 0.1$. The second scenario with setting $r = \text{rank}_t(\mathcal{A}) = 0.1n$ and $\rho_s = 0.2$.

The Gaussian noise \mathcal{F} in each frontal slice is generated independently with each other, i.e.

$$\mathcal{F}(:, :, i_3) \sim \mathcal{N}\left(0, \sigma_{i_3}^2\right), \quad 1 \leq i_3 \leq n \quad (28)$$

The variance values $\sigma_{i_3}^2$ in each frontal slice are randomly selected from 0 to 0.1. In this sub-subsection 1, our task is to recovery \mathcal{A} from noisy observation $\chi = \mathcal{A} + \mathcal{E} + \mathcal{F}$ with \mathcal{E} of varying sparsity. **Table 1** and **Table 2** show the

Table 1. Correct recovery for random problems of varying sparsity using RLRTA.

$r = \text{rank}_r(\mathcal{F}) = 0.1n, m = \ \mathcal{S}\ _0 = 0.1n^3$					
n	r	m	$\ \tilde{\mathcal{S}}\ _0$	$\frac{\ \mathcal{F} - \tilde{\mathcal{F}}\ _F}{\ \mathcal{F}\ _F}$	$\frac{\ \mathcal{S} - \tilde{\mathcal{S}}\ _F}{\ \mathcal{S}\ _F}$
100	10	1e5	132,399	1.1838e-04	0.3040
200	20	8e5	1,046,860	2.8331e-05	0.3026
$r = \text{rank}_r(\mathcal{F}) = 0.1n, m = \ \mathcal{S}\ _0 = 0.2n^3$					
n	r	m	$\ \tilde{\mathcal{S}}\ _0$	$\frac{\ \mathcal{F} - \tilde{\mathcal{F}}\ _F}{\ \mathcal{F}\ _F}$	$\frac{\ \mathcal{S} - \tilde{\mathcal{S}}\ _F}{\ \mathcal{S}\ _F}$
100	10	2e5	222,128	1.5001e-04	0.3072
200	20	16e5	1,797,586	3.8035e-05	0.3118

Table 2. Correct recovery for random problems of varying sparsity using TRPCA [23].

$r = \text{rank}_r(\mathcal{F}) = 0.1n, m = \ \mathcal{S}\ _0 = 0.1n^3$					
n	r	m	$\ \tilde{\mathcal{S}}\ _0$	$\frac{\ \mathcal{F} - \tilde{\mathcal{F}}\ _F}{\ \mathcal{F}\ _F}$	$\frac{\ \mathcal{S} - \tilde{\mathcal{S}}\ _F}{\ \mathcal{S}\ _F}$
100	10	1e5	575,485	0.0021	0.2805
200	20	8e5	4,594,860	5.4577e-04	0.2727
$r = \text{rank}_r(\mathcal{F}) = 0.1n, m = \ \mathcal{S}\ _0 = 0.2n^3$					
n	r	m	$\ \tilde{\mathcal{S}}\ _0$	$\frac{\ \mathcal{F} - \tilde{\mathcal{F}}\ _F}{\ \mathcal{F}\ _F}$	$\frac{\ \mathcal{S} - \tilde{\mathcal{S}}\ _F}{\ \mathcal{S}\ _F}$
100	10	2e5	576,448	0.0030	0.1597
200	20	16e5	4,609,591	8.3233e-04	0.1707

recovery results of algorithm RLRTA and TRPCA. It’s shown that RLRTA can better recover the low-rank component \mathcal{A} under different sparse component \mathcal{E} .

4.2. Exact Recovery with \mathcal{F} of Varying Intensity

Now we exam the recovery phenomenon with Gaussian noise of varying variances. The generation of $\mathcal{A} \in \mathbb{R}^{n \times n \times n}$ is the same as that in sub-subsection 1 and $r = \text{rank}_r(\mathcal{A}) = 0.1n$. The sparse component \mathcal{E} has sparsity $\rho_s = 0.1$. For simplicity, we assume that \mathcal{F} is white Gaussian noise, that is

$$\mathcal{F}(i_1, i_2, i_3) \sim \mathcal{N}(0, \sigma_w^2) \tag{29}$$

where $1 \leq i_1 \leq n, 1 \leq i_2 \leq n, 1 \leq i_3 \leq n$. The noise variance values σ_w^2 are 0.02, 0.04, 0.06, 0.08 and 0.1, respectively **Table 3** and **Table 4** show the recovery results of algorithm RLRTA and TRPCA. It’s shown that RLRTA can better recover the low-rank component \mathcal{A} under different Gaussian noise \mathcal{F} .

4.3. Phase Transition in Rank and Sparsity

Now we exam the recovery phenomenon with varying rank of \mathcal{A} and varying sparsity of \mathcal{E} . Similar to [23], we consider two sizes of $\mathcal{A} \in \mathbb{R}^{n \times n \times n_3}$: (1)

Table 3. Correct recovery for random problems of varying intensity using RLRTA.

$r = \text{rank}_t(\mathcal{F}) = 0.1n, m = \ \mathcal{S}\ _0 = 0.1n^3, n = 100$					
σ_w^2	r	m	$\ \tilde{\mathcal{S}}\ _0$	$\frac{\ \mathcal{F} - \tilde{\mathcal{F}}\ _F}{\ \mathcal{F}\ _F}$	$\frac{\ \mathcal{S} - \tilde{\mathcal{S}}\ _F}{\ \mathcal{S}\ _F}$
0.02	10	1e5	100,438	1.1900e-04	0.2707
0.04	10	1e5	111,249	1.1886e-04	0.2915
0.06	10	1e5	134,908	1.2041e-04	0.3138
0.08	10	1e5	160,814	1.5044e-04	0.3451
0.10	10	1e5	184,006	2.2189e-04	0.3824
$r = \text{rank}_t(\mathcal{F}) = 0.1n, m = \ \mathcal{S}\ _0 = 0.1n^3, n = 200$					
σ_w^2	r	m	$\ \tilde{\mathcal{S}}\ _0$	$\frac{\ \mathcal{F} - \tilde{\mathcal{F}}\ _F}{\ \mathcal{F}\ _F}$	$\frac{\ \mathcal{S} - \tilde{\mathcal{S}}\ _F}{\ \mathcal{S}\ _F}$
0.02	20	8e5	803,564	2.8135e-05	0.2709
0.04	20	8e5	889,969	2.8269e-05	0.2913
0.06	20	8e5	1,080,602	2.9120e-05	0.3145
0.08	20	8e5	1,287,048	3.6441e-05	0.3460
0.10	20	8e5	1,472,423	5.4040e-05	0.3821

Table 4. Correct recovery for random problems of varying intensity using TRPCA [23].

$r = \text{rank}_t(\mathcal{F}) = 0.1n, m = \ \mathcal{S}\ _0 = 0.1n^3, n = 100$					
σ_w^2	r	m	$\ \tilde{\mathcal{S}}\ _0$	$\frac{\ \mathcal{F} - \tilde{\mathcal{F}}\ _F}{\ \mathcal{F}\ _F}$	$\frac{\ \mathcal{S} - \tilde{\mathcal{S}}\ _F}{\ \mathcal{S}\ _F}$
0.02	10	1e5	571,972	0.0013	0.1036
0.04	10	1e5	571,943	0.0024	0.2041
0.06	10	1e5	571,944	0.0036	0.3039
0.08	10	1e5	571,290	0.0048	0.4037
0.10	10	1e5	571,880	0.0059	0.5039
$r = \text{rank}_t(\mathcal{F}) = 0.1n, m = \ \mathcal{S}\ _0 = 0.1n^3, n = 200$					
σ_w^2	r	m	$\ \tilde{\mathcal{S}}\ _0$	$\frac{\ \mathcal{F} - \tilde{\mathcal{F}}\ _F}{\ \mathcal{F}\ _F}$	$\frac{\ \mathcal{S} - \tilde{\mathcal{S}}\ _F}{\ \mathcal{S}\ _F}$
0.02	20	8e5	4,573,505	3.0970e-04	0.1036
0.04	20	8e5	4,574,842	6.0571e-04	0.2043
0.06	20	8e5	4,573,573	9.0370e-04	0.3040
0.08	20	8e5	4,573,394	0.0012	0.4039
0.10	20	8e5	4,572,467	0.0015	0.5021

$n = 100, n_3 = 50$, (2) $n = 200, n_3 = 50$. We generate $\mathcal{A} = \mathcal{Q} * \mathcal{V}$, where the entries of $\mathcal{Q} \in \mathbb{R}^{n \times r \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{r \times n \times n_3}$ are independently sampled from a uniform distribution in interval $(0, 1/n)$. For \mathcal{E} , we still consider a Bernoulli model for its support and random signs as in Equation (27). The variance values σ_{w, i_3}^2 in each frontal slice $i_3 (1 \leq i_3 \leq n_3)$ are randomly selected from 0 to 0.1, and the mean variance values are both set to be 0.05.

We set r/n as all the choices in $[0.01 : 0.01 : 0.5]$, and ρ_s in $[0.01 : 0.01 : 0.5]$. For each (r, ρ_s) -pair, we simulate 10 test instances and declare a trial to be successful if the recovered $\tilde{\mathcal{A}}$ satisfies $\|\tilde{\mathcal{A}} - \mathcal{A}\|_F / \|\mathcal{A}\|_F \leq 10^{-3}$. **Figure 1** plots

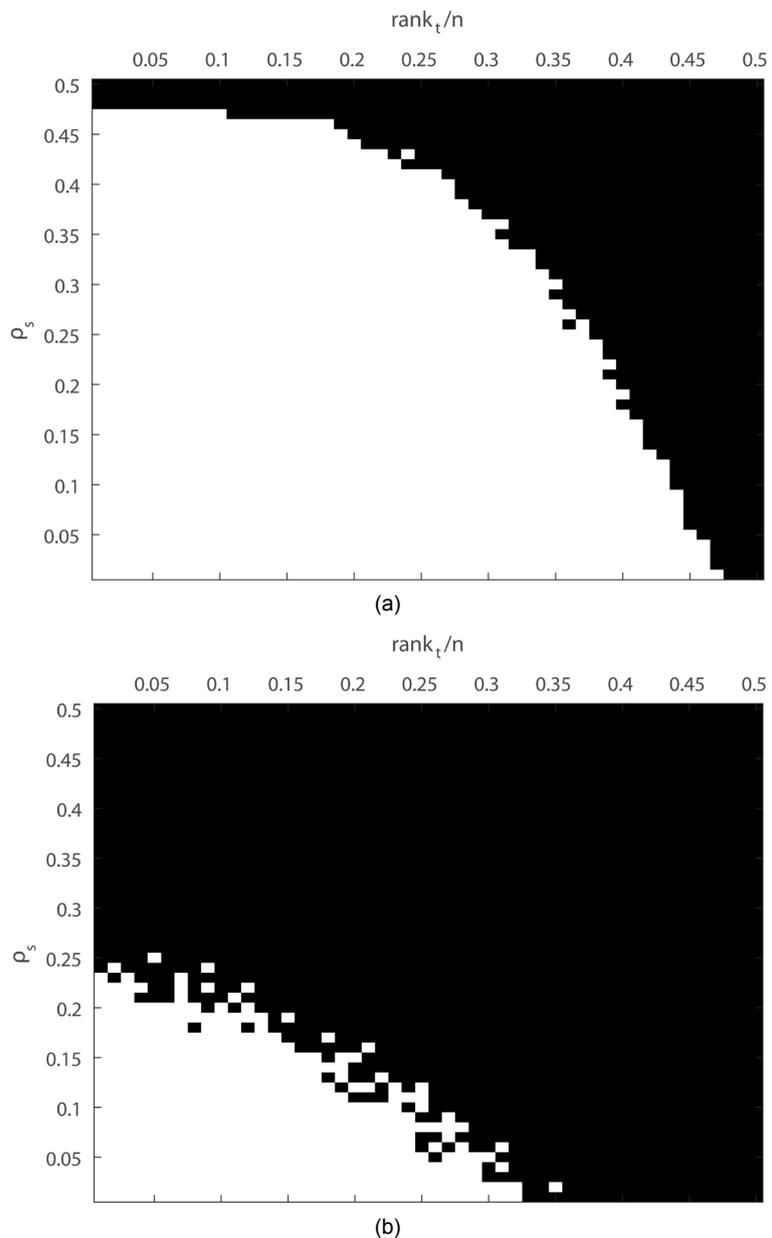


Figure 1. Correct recovery for varying rank and sparsity for RLRTA and TRPCA [23]. Fraction of correct recoveries, as a function of rank_t(\mathcal{A}) (x -axis) and sparsity of \mathcal{E} (y -axis).

the fraction of correct recovery for each pair (black = 0% and white = 100%). It can be seen that there is a large region in which the recovery is correct.

4.4. Phase Transition in Rank and Entry-Wise Noise Intensity

Now we exam the recovery phenomenon with varying rank of \mathcal{A} and varying intensity of noise \mathcal{F} . We still consider two sizes of $\mathcal{A} \in \mathbb{R}^{n \times n \times n_3}$: (1) $n = 100, n_3 = 50$, (2) $n = 200, n_3 = 50$. We generate $\mathcal{A} = \mathcal{Q} * \mathcal{V}$, where the entries of $\mathcal{Q} \in \mathbb{R}^{n \times r \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{r \times n \times n_3}$ independently sampled from a uniform distribution in interval $(0, 1/n)$. For \mathcal{E} , we still consider a Bernoulli model for its

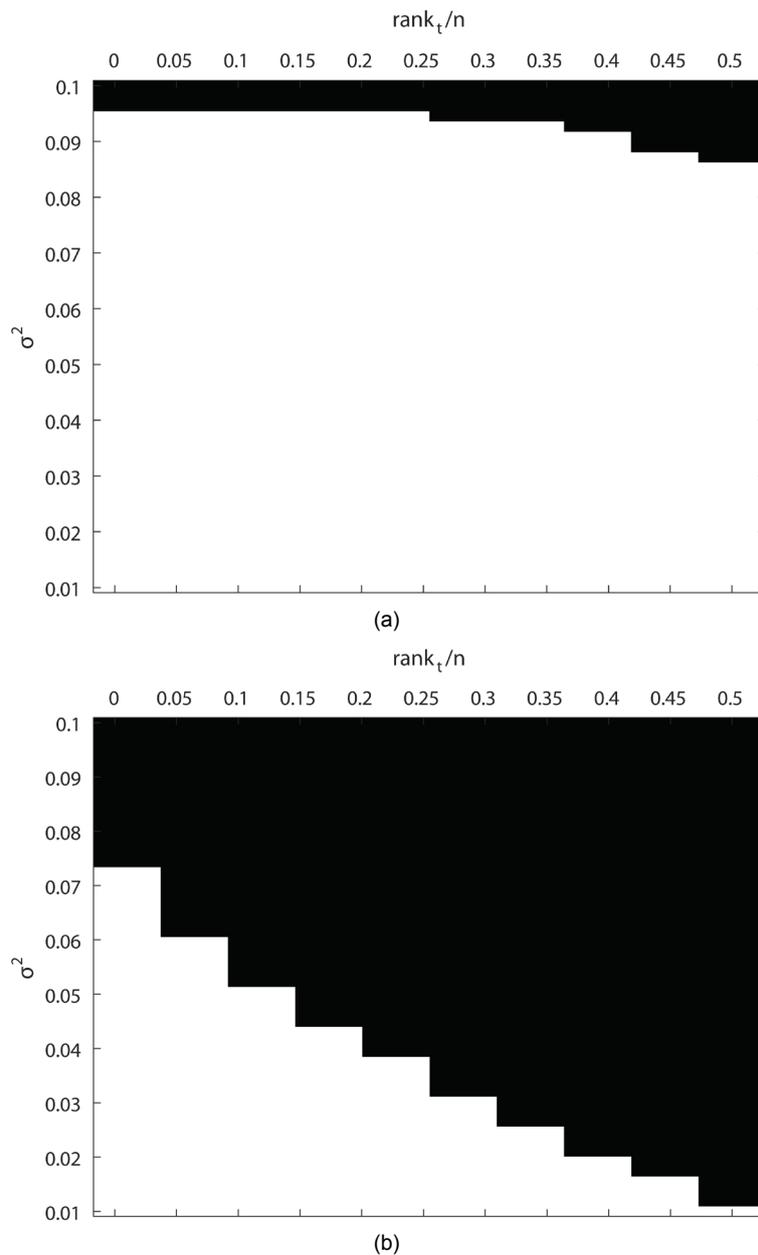


Figure 2. Correct recovery for varying rank and noise variance for RLRTA and TRPCA [23]. Fraction of correct recoveries, as a function of $\text{rank}_t(\mathcal{A})$ (x -axis) and variance of \mathcal{F} (y -axis).

support and random signs as in Equation (27) and sparsity parameter ρ_s is fixed at 0.1. The generation of \mathcal{F} is similar to Equation (29).

We set r/n as all the choices in $[0.01:0.05:0.5]$. The noise variance values σ_w^2 are in $[0.01:0.01:0.1]$. For each (r, σ_w^2) -pair, we simulate 10 test instances and declare a trial to be successful if the recovered $\tilde{\mathcal{A}}$ satisfies $\|\tilde{\mathcal{A}} - \mathcal{A}\|_F / \|\mathcal{A}\|_F \leq 10^{-3}$. **Figure 2** plots the fraction of correct recovery for each pair (black = 0% and white = 100%). It can be seen that there is a large region in which the recovery is correct.

5. Conclusion

This work verifies the ability of convex optimization for the recovery of low-rank tensors corrupted by both impulse and Gaussian noise. The problem is tackled by integrating the tensor nuclear norm, l_1 -norm and least square term in a unified convex relaxation framework. Parameters are selected to comprise the low-rank component, the sparse component and the Gaussian-noise term. Besides, the convergence of the proposed algorithm is discussed. Numerical experiments have been conducted to demonstrate the efficacy of our proposed denoising approach.

Acknowledgements

The authors would like to thank Canyi Lu for providing the code for TRPCA algorithm.

References

- [1] Jolliffe, I. (2002) Principal Component Analysis. Wiley Online Library.
- [2] Candes, E.J., Li, X.D., Ma, Y. and Wright, J. (2011) Robust Principal Component Analysis? *Journal of the ACM*, **58**, Article Number: 11. <https://doi.org/10.1145/1970392.1970395>
- [3] Chandrasekaran, V., Sanghavi, S., Parrilo, P.A. and Willsky, A.S. (2011) Rank-Sparsity Incoherence for Matrix Decomposition. *SIAM Journal on Optimization*, **21**, 572-596. <https://doi.org/10.1137/090761793>
- [4] Wright, J., Ganesh, A., Rao, S.Y.P. and Ma, Y. (2009) Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization Neural Information Processing Systems (NIPS).
- [5] Tao, M. and Yuan, X. (2011) Recovering Low-Rank and Sparse Components of Matrices from Incomplete and Noisy Observations. *SIAM Journal on Optimization*, **21**, 57-81. <https://doi.org/10.1137/100781894>
- [6] Candes, E.J. and Recht, B. (2009) Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, **9**, 717-772. <https://doi.org/10.1007/s10208-009-9045-5>
- [7] Candes, E.J. and Tao, T. (2010) The Power of Convex Relaxation: Near-Optimal Matrix Completion. *IEEE Transaction on Information Theory*, **56**, 2053-2080. <https://doi.org/10.1109/TIT.2010.2044061>
- [8] Wang, H. and Ahuja, N. (2004) Compact Representation of Multidimensional Data Using Tensor Rank-One Decomposition.
- [9] Bloy, L. and Verma, R. (2008) On Computing the Underlying Fiber Directions from

the Diffusion Orientation Distribution Function. *11th International Conference on Medical Image Computing and Computer-Assisted Intervention*, New York, 6-10 September 2008, 1-8.

- [10] Ghosh, A., Tsigaridas, E., Descoteaux, M., Comon, P., Mourrain, B. and Deriche, R. (2008) A Polynomial Based Approach to Extract the Maxima of an Antipodally Symmetric Spherical Function and Its Application to Extract Fiber Directions from the Orientation Distribution Function in Diffusion MRI.
- [11] Qi, L., Yu, G. and Wu, E.X. (2010) Higher Order Positive Semi-Definite Diffusion Tensor Imaging. *SIAM Journal on Imaging Sciences*, **3**, 416-433.
<https://doi.org/10.1137/090755138>
- [12] Hilling, J.J. and Sudbery, A. (2010) The Geometric Measure of Multipartite Entanglement and the Singular Values of a Hypermatrix. *Journal of Mathematical Physics*, **51**, Article ID: 072102. <https://doi.org/10.1063/1.3451264>
- [13] Hu, S. and Qi, L. (2012) Algebraic Connectivity of an Even Uniform Hypergraph. *Journal of Combinatorial Optimization*, **24**, 564-579.
<https://doi.org/10.1007/s10878-011-9407-1>
- [14] Li, W. and Ng, M. (2011) Existence and Uniqueness of Stationary Probability Vector of a Transition Probability Tensor Technical Report. Department of Mathematics, the Hong Kong Baptist University, Hong Kong.
- [15] Liu, J., Musialski, P., Wonka, P. and Ye, J. (2009) Tensor Completion for Estimating Missing Values in Visual Data In ICCV.
- [16] Lubich, C. (2008) From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis EMS Zürich.
- [17] Beck, M.H., Jäckle, A., Worth, G.A. and Meyer, H.D. (1999) The Multiconfiguration Time-Dependent Hartree (MCTDH) Method: A Highly Efficient Algorithm for Propagating Wavepackets. *Physics Reports*, **324**, 1-105.
[https://doi.org/10.1016/S0370-1573\(99\)00047-2](https://doi.org/10.1016/S0370-1573(99)00047-2)
- [18] Wang, H. and Thoss, M. (2009) Numerically Exact Quantum Dynamics for Indistinguishable Particles: The Multilayer Multiconfiguration Time-Dependent Hartree Theory in Second Quantization Representation. *Journal of Chemical Physics*, **131**, Article ID: 024114.
- [19] Paredes, B.R., Aung, H., Berthouze, N.B. and Pontil, M. (2013) Multilinear Multi-task Learning. *Journal of Machine Learning Research*, **28**, 1444-1452.
<https://doi.org/10.1063/1.3173823>
- [20] Kolda, T.G. and Bader, B.W. (2009) Tensor Decompositions and Applications. *SIAM Review*, **51**, 455-500. <https://doi.org/10.1137/07070111X>
- [21] Zhang, Z., Ely, G., Aeron, S., Hao, N. and Kilmer, M.E. (2014) Novel Methods for Multilinear Data Completion and De-Noising Based on Tensor-SVD. *Computer Science*, **44**, 3842-3849. <https://doi.org/10.1109/cvpr.2014.485>
- [22] Semerci, O., Hao, N., Kilmer, M.E. and Miller, E.L. (2013) Tensor-Based Formulation and Nuclear Norm Regularization for Multienergy Computed Tomography. *IEEE Transactions on Image Processing*, **23**, 1678-1693.
<https://doi.org/10.1109/TIP.2014.2305840>
- [23] Lu, C., Feng, J., Chen, Y., Liu, W., Lin, Z. and Yan, S. (2016) Tensor Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Tensors via Convex Optimization.
- [24] Cai, X., Han, D. and Yuan, X. (2014) The Direct Extension of ADMM for Three-Block Separable Convex Minimization Models Is Convergent When One Function Is Strongly Convex.

- [25] Li, M., Sun, D. and Toh, K. (2014) A Convergent 3-Block Semi-Proximal ADMM for Convex Minimization Problems with One Strongly Convex Block. *Asia Pacific Journal of Operational Research*, **32**, 1550024. <https://doi.org/10.1142/S0217595915500244>
- [26] Candes, E.J. and Plan, Y. (2010) Matrix Completion with Noise. *Proceedings of the IEEE*, **98**, 925-936. <https://doi.org/10.1109/JPROC.2009.2035722>
- [27] Zhang, Z. and Aeron, S. (2015) Exact Tensor Completion Using t-SVD.



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact am@scirp.org

