Tensor Recovery via Nonconvex Low-Rank Approximation

Lin Chen*, Xue Jiang*, Xingzhao Liu*, Zhixin Zhou†
*Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China
†Space Engineering University, Beijing, China

***Abstract*—The low-rank tensor recovery is a powerful approach to depict the intrinsic structure within high-dimensional data, and has been extensively leveraged in many real-world applications. Conventional techniques of low-rank recovery formulate it as a rank minimization problem, then approximate the rank function with the convex relaxation. In this paper, we propose a new tensor logarithmic norm as the nonconvex rank surrogate. Compared with the convex surrogate of nuclear norm, the proposed logarithmic norm is proved to be a tighter approximation to the tensor average rank, and thus is more sparsity-encouraging to extract the underlying low-rank information. Although minimizing the logarithmic norm leads to a nonconvex optimization problem, we rigorously derive its closed-form solution with the guarantee of local optimality. Experimental results demonstrate the strong convergence behavior of the proposed algorithm. In the real-world application of video recovery, our method outperforms several state-of-the-art methods and shows the remarkable recovery accuracy.**

***Index Terms*—low-rank, tensor recovery, nonconvex approximation, augmented Lagrange multiplier**

I. Introduction

Low-rank tensor recovery aim at extracting the underlying low-rank structure thus filling in the missing entries from the degenerate tensor observation [1]. It can be viewed as an extension from the problem of low-rank matrix recovery, which is usually formulated as the rank minimization problem [2]. Unlike the unique definition of matrix rank, the tensor rank can be defined in different forms corresponding to different tensor decomposition strategies. Among these decomposition strategies, the tensor singular value decomposition (t-SVD) [3] has attracted more and more attention. It employs the block circulant operator in the matricization of tensor, and hence favors the preservation of tensor structure [1].

Based upon t-SVD, the tensor multi-rank and tensor average rank is proposed to depict the rank information [4]. However, minimizing the tensor rank function leads to an NP-hard problem, whose solution is difficult to obtain within the polynomial time. Generally, the tensor rank can be replaced by its convex surrogate, which can be solved with powerful theoretical guarantees. In [4], the tensor robust principal component analysis (TRPCA) is proposed to use the tensor nuclear norm as the tightest convex surrogate of tensor average rank.

However, tensor nuclear norm treats different singular values with the same penalty degree, which will over-penalize large singular values thus suffering from a biased solution [5]. Since the low-rank property of a tensor is mainly affected by some large singular values, it would be better to enforce a less severe punishment to a larger singular value, which is consistent with the sparsity-driven singular value minimization [6].

To enhance the flexibility of singular value punishment, the reweighted tensor nuclear norm is proposed in [7]. It remodels the conventional nuclear norm with the weight allocation to improve the low-rank approximation performance. In addition, the tensor Schatten-p norm is proposed in [5] as the nonconvex relaxation of rank function, which adopts the nonconvex ℓ_p-norm with $0 < p < 1$ on the singular value vector rather than the ℓ_1-norm used in the nuclear norm. Inspired by the nonconvex surrogate for the tensor average rank, in this paper, we propose a new definition of tensor logarithmic norm and prove it to be a better approximation to the low-rank property of tensor than the convex surrogate. By employing the tensor logarithmic norm minimization along with the ℓ_1-ball noise rejection, we then develop an efficient and robust algorithm to recover the low-rank tensor from incomplete and noisy observations. The resultant optimization problem is reformulated into three subproblems and solved alternately by the augmented Lagrange multiplier (ALM) method [8]. In particular, the closed-form solution for the logarithmic norm-based subproblem is rigorously derived, and the experimental result empirically shows the good convergence behavior of our proposed algorithm. More importantly, experiments on both synthetic and real data demonstrate that using the proposed tensor logarithmic norm could reconstruct the low-rank structures in tensors with the higher precision than using tensor nuclear norm or tensor Schatten-p norm.

Notation: Fonts $\boldsymbol{\mathcal{A}}$, $\mathbf{A}$, $\mathbf{a}$, and a denote a tensor, a matrix, a vector, and a scalar, respectively. For any third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, we denote its ith frontal slice as $\mathbf{X}^{(i)} \in \mathbb{R}^{m_1 \times m_2}$, and its discrete Fourier transformation (DFT) along the third dimension as $\bar{\boldsymbol{\mathcal{X}}} \in \mathbb{C}^{m_1 \times m_2 \times m_3}$. Similarly, $\bar{\mathbf{X}}^{(i)} \in \mathbb{C}^{m_1 \times m_2}$ stands for the ith frontal slice of $\bar{\boldsymbol{\mathcal{X}}}$. The tensor multi-rank and tensor average rank are denoted by $\mathrm{rank}_m(\boldsymbol{\mathcal{X}}) = [\mathrm{rank}(\bar{\mathbf{X}}^{(1)}), \mathrm{rank}(\bar{\mathbf{X}}^{(2)}), \cdots, \mathrm{rank}(\bar{\mathbf{X}}^{(m_3)})]^T$ and $\mathrm{rank}_a(\boldsymbol{\mathcal{X}}) = \frac{1}{m_3}\|\mathrm{rank}_m(\boldsymbol{\mathcal{X}})\|_1$, respectively. For two tensors $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{m_2 \times m_4 \times m_3}$, their tensor-tensor product (t-product) is represented by $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}} \in \mathbb{R}^{m_1 \times m_4 \times m_3}$, which is equivalent to the matrix multiplication in the DFT domain as $\bar{\mathbf{C}}^{(i)} = \bar{\mathbf{A}}^{(i)}\bar{\mathbf{B}}^{(i)}$ $(i = 1, \cdots, m_3)$ [3].

This work is supported by the National Natural Science Foundation of China under Grant 61971279.

II. Preliminaries

To make the paper self-contained, we briefly introduce some basic knowledge of tensors as follows. For the tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, we can perform the t-SVD based on the t-product as $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^*$, which corresponds to the SVD of $\bar{\mathbf{X}}^{(i)} = \bar{\mathbf{U}}^{(i)} \bar{\mathbf{S}}^{(i)} \bar{\mathbf{V}}^{(i),*}$ in the DFT domain ($i = 1, \cdots, m_3$) [3]. Based upon t-product and t-SVD, the tensor nuclear norm is proposed in [4] to surrogate the tensor average rank for low-rank approximation in a convex manner as

$$\|\boldsymbol{\mathcal{X}}\|_* = \tfrac{1}{m_3} \sum_{i=1}^{m_3} \sum_{j=1}^{\min\{m_1,m_2\}} \sigma_j\big(\bar{\mathbf{X}}^{(i)}\big). \tag{1}$$

Because the tensor nuclear norm consistently treats all singular values, it usually leads to the excessive punishment on large singular values, which have significant impacts on the low-rank property of a tensor. To address this problem, the reweighted tensor nuclear norm is proposed to remodel the nuclear norm with a flexible weight allocation as [7]

$$\|\boldsymbol{\mathcal{X}}\|_{\mathrm{w}} = \tfrac{1}{m_3} \sum_{i=1}^{m_3} \sum_{j=1}^{\min\{m_1,m_2\}} w_j^i \sigma_j\big(\bar{\mathbf{X}}^{(i)}\big) \tag{2}$$

where w_j^i represents the weight associated with $\sigma_j\big(\bar{\mathbf{X}}^{(i)}\big)$, and its selection plays the crucial role for reweighted tensor nuclear norm. To promote the sparsity of singular values for the low-rank approximation, a widely recommended weight allocation strategy is to set $w_j^i = \big(\sigma_j\big(\bar{\mathbf{X}}^{(i)}\big) + \epsilon\big)^{-1}$, where $\epsilon > 0$ is a small scalar to avoid dividing by zero [7]. The minimization of (2) derives the low-rank tensor recovery problem, whose solution can be obtained by optimizing each singular value and the weight alternately and iteratively [2]. With respect to the kth iteration, it yields

$$\begin{cases} \sigma_j\big(\bar{\mathbf{X}}_{k+1}^{(i)}\big) = \arg\min w_j^{i,k} \sigma_j\big(\bar{\mathbf{X}}^{(i)}\big) \\ w_j^{i,k+1} = \big(\sigma_j\big(\bar{\mathbf{X}}_{k+1}^{(i)}\big) + \epsilon\big)^{-1} \end{cases} \tag{3}$$

which can be further integrated into an unified form as [2]

$$\sigma_j\big(\bar{\mathbf{X}}_{k+1}^{(i)}\big) = \underset{\sigma_j(\bar{\mathbf{X}}^{(i)})}{\arg\min} \ \frac{\sigma_j\big(\bar{\mathbf{X}}^{(i)}\big)}{\sigma_j\big(\bar{\mathbf{X}}_k^{(i)}\big) + \epsilon}. \tag{4}$$

III. Proposed Low-Rank Tensor Recovery

A. Proposed Model

Intuitively, (4) can be viewed as a linearized solution to the following optimization problem:

$$\min_{\sigma_j(\bar{\mathbf{X}}^{(i)})} f\big(\sigma_j\big(\bar{\mathbf{X}}^{(i)}\big)\big) = \log\big(\sigma_j\big(\bar{\mathbf{X}}^{(i)}\big)/\epsilon + 1\big). \tag{5}$$

In other words, we are capable of linearizing the objective function of (5) at the point $\sigma_j\big(\bar{\mathbf{X}}_k^{(i)}\big)$ during the iteration by

$$\begin{aligned} f\big(\sigma_j\big(\bar{\mathbf{X}}^{(i)}\big)\big) \approx{}& \log\big(\sigma_j\big(\bar{\mathbf{X}}_k^{(i)}\big)/\epsilon + 1\big) \\ &+ \big\langle \nabla f\big(\sigma_j\big(\bar{\mathbf{X}}_k^{(i)}\big)\big), \sigma_j\big(\bar{\mathbf{X}}^{(i)}\big) - \sigma_j\big(\bar{\mathbf{X}}_k^{(i)}\big)\big\rangle \end{aligned} \tag{6}$$

where $\nabla f\big(\sigma_j\big(\bar{\mathbf{X}}_k^{(i)}\big)\big) = \big(\sigma_j\big(\bar{\mathbf{X}}_k^{(i)}\big) + \epsilon\big)^{-1}$ is the derivative of $f(\cdot)$ at $\sigma_j\big(\bar{\mathbf{X}}_k^{(i)}\big)$. After ignoring the constant terms of (6), we substitute it to (5), which explicitly derives an equivalent form of (4). Motivated by the above logarithmic heuristics, we propose a new definition of tensor logarithmic norm as the surrogate of tensor average rank.

Definition 1. *For any tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, its tensor logarithmic norm is defined as*

$$\|\boldsymbol{\mathcal{X}}\|_{\mathrm{L}} \triangleq \tfrac{1}{m_3} \sum_{i=1}^{m_3} \sum_{j=1}^{\min\{m_1,m_2\}} \log\big(\sigma_j\big(\bar{\mathbf{X}}^{(i)}\big)/\epsilon + 1\big). \tag{7}$$

Compared with tensor nuclear norm, the tensor logarithmic norm could approximate the tensor average rank better, as stated in the following lemma.

Lemma 1. *Within the unit ball of tensor spectral norm, i.e., $\{\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m_1 \times m_2 \times m_3} \,|\, \|\boldsymbol{\mathcal{X}}\| \le 1\}$, the tensor logarithmic norm with the constant $\epsilon = 1/(e-1)$ is a tighter envelope of tensor average rank than the tensor nuclear norm as $\|\boldsymbol{\mathcal{X}}\|_* \le \|\boldsymbol{\mathcal{X}}\|_{\mathrm{L}} \le \mathrm{rank}_a(\boldsymbol{\mathcal{X}})$.*

Proof. Noting that the function $f(x) = \log(x/\epsilon + 1) \le 1$ holds within the unit interval, we have:

$$\begin{aligned} \|\boldsymbol{\mathcal{X}}\|_{\mathrm{L}} &= \tfrac{1}{m_3} \sum_{i=1}^{m_3} \sum_{j=1}^{\min\{m_1,m_2\}} \log\big(\sigma_j\big(\bar{\mathbf{X}}^{(i)}\big)/\epsilon + 1\big) \\ &\le \tfrac{1}{m_3} \sum_{i=1}^{m_3} \mathrm{rank}\big(\bar{\mathbf{X}}^{(i)}\big) = \mathrm{rank}_a\big(\boldsymbol{\mathcal{X}}\big). \end{aligned} \tag{8}$$

In addition, the function $f(x)$ is concave within the unit interval, and hence the following inequality holds:

$$\begin{aligned} \|\boldsymbol{\mathcal{X}}\|_{\mathrm{L}} &= \tfrac{1}{m_3} \sum_{i=1}^{m_3} \sum_{j=1}^{\min\{m_1,m_2\}} \log\big(\sigma_j\big(\bar{\mathbf{X}}^{(i)}\big)/\epsilon + 1\big) \\ &\ge \tfrac{1}{m_3} \sum_{i=1}^{m_3} \sum_{j=1}^{\min\{m_1,m_2\}} \sigma_j\big(\bar{\mathbf{X}}^{(i)}\big) = \|\boldsymbol{\mathcal{X}}\|_*. \end{aligned} \tag{9}$$

Combining (8) and (9) completes the proof. ∎

Using the new defined tensor logarithmic norm to address the low-rank tensor recovery problem, we present the problem formulation as follows. Suppose that there are some missing entries in the tensor observation $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, while other observed entries are within the set Ω. Define $\mathcal{P}_\Omega$ as the linear projection to extract the observed entries in Ω while filling the missing entries not in Ω with zero values [1]. According to the above assumption, the underlying low-rank tensor can be recovered by

$$\min_{\boldsymbol{\mathcal{X}}} \|\boldsymbol{\mathcal{X}}\|_{\mathrm{L}} \quad \text{s.t. } \mathcal{P}_\Omega(\boldsymbol{\mathcal{X}}) = \mathcal{P}_\Omega(\boldsymbol{\mathcal{T}}). \tag{10}$$

In practical cases where input observations are corrupted by sparse noise or outliers, we robustify (10) by replacing the equality constraint with the ℓ_1-ball noise rejection scheme [9], then formulate the following tensor recovery model:

$$\min_{\boldsymbol{\mathcal{X}}} \|\boldsymbol{\mathcal{X}}\|_{\mathrm{L}} \quad \text{s.t. } \boldsymbol{\mathcal{X}} \in \mathcal{B}_1 \tag{11}$$

where $\mathcal{B}_1 := \big\{\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m_1 \times m_2 \times m_3} \big| \|\mathcal{P}_\Omega(\boldsymbol{\mathcal{T}}) - \mathcal{P}_\Omega(\boldsymbol{\mathcal{X}})\|_1 \le \delta\big\}$ denotes the ℓ_1-ball with the center of ball being $\mathcal{P}_\Omega(\boldsymbol{\mathcal{T}})$, and the parameter δ is used to control the fitting error.

B. Optimization Method

Equation (11) leads to a constrained minimization problem, which can be efficiently solved via alternating optimization approaches. Here we adopt the ALM method [8] to solve (11), and first introduce an auxiliary variable $\boldsymbol{\mathcal{F}}$ to reformulate it as

$$\min_{\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{F}}} \|\boldsymbol{\mathcal{X}}\|_{\mathrm{L}} \quad \text{s.t. } \boldsymbol{\mathcal{X}}+\boldsymbol{\mathcal{F}}=\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{T}}),\ \boldsymbol{\mathcal{F}}\in\mathcal{B}_2 \tag{12}$$

where $\mathcal{B}_2 := \big\{\boldsymbol{\mathcal{F}}\in\mathbb{R}^{m_1\times m_2\times m_3}\big|\|\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{F}})\|_1\le\delta\big\}$.

The augmented Lagrangian for (12) is

$$\begin{aligned}\mathcal{L}_{\mu}(\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{F}},\boldsymbol{\mathcal{Z}}) = \|\boldsymbol{\mathcal{X}}\|_{\mathrm{L}} + \langle\boldsymbol{\mathcal{Z}},\boldsymbol{\mathcal{X}}+\boldsymbol{\mathcal{F}}-\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{T}})\rangle \\ +\tfrac{\mu}{2}\|\boldsymbol{\mathcal{X}}+\boldsymbol{\mathcal{F}}-\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{T}})\|_F^2\end{aligned} \tag{13}$$

where $\boldsymbol{\mathcal{Z}}\in\mathbb{R}^{m_1\times m_2\times m_3}$ is the dual variable and μ is a small positive scalar. The ALM method alternately updates each optimization variable until convergence. In the kth iteration, the optimization procedure can be divided into three subproblems as follows.

1) Updating $\boldsymbol{\mathcal{X}}^{k+1}$ by Minimizing:

$$\boldsymbol{\mathcal{X}}^{k+1}=\arg\min_{\boldsymbol{\mathcal{X}}}\|\boldsymbol{\mathcal{X}}\|_{\mathrm{L}}+\frac{\mu_k}{2}\|\boldsymbol{\mathcal{X}}+\boldsymbol{\mathcal{F}}^k-\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{T}})+\mu_k^{-1}\boldsymbol{\mathcal{Z}}^k\|_F^2. \tag{14}$$

Although the nonconvexity of tensor logarithmic norm poses the hindrance to solve (14), we can obtain its closed-form solution with the local optimality via the following theorem.

Lemma 2. *For $\lambda>0$ and $\boldsymbol{\mathcal{Y}}\in\mathbb{R}^{m_1\times m_2\times m_3}$ with the t-SVD of $\boldsymbol{\mathcal{Y}}=\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{S}}_1*\boldsymbol{\mathcal{V}}^*$, the closed-form solution to the following problem*

$$\min_{\boldsymbol{\mathcal{X}}}\frac{1}{2}\|\boldsymbol{\mathcal{X}}-\boldsymbol{\mathcal{Y}}\|_F^2+\lambda\|\boldsymbol{\mathcal{X}}\|_{\mathrm{L}} \tag{15}$$

is given by $\boldsymbol{\mathcal{X}}=\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{S}}_2*\boldsymbol{\mathcal{V}}^*$. Especially, for the ith frontal slices of DFT of $\boldsymbol{\mathcal{S}}_1$ and $\boldsymbol{\mathcal{S}}_2$, i.e., $\bar{\mathbf{S}}_1^{(i)}=\mathrm{diag}\big(\sigma_1(\bar{\mathbf{Y}}^{(i)}),\cdots,\sigma_j(\bar{\mathbf{Y}}^{(i)}),\cdots,\sigma_d(\bar{\mathbf{Y}}^{(i)})\big)$ and $\bar{\mathbf{S}}_2^{(i)}=\mathrm{diag}\big(\sigma_1(\bar{\mathbf{X}}^{(i)}),\cdots,\sigma_j(\bar{\mathbf{X}}^{(i)}),\cdots,\sigma_d(\bar{\mathbf{X}}^{(i)})\big)$ with $d=\min\{m_1,m_2\}$, they must satisfy

$$\sigma_j\big(\bar{\mathbf{X}}^{(i)}\big)=\begin{cases}\frac{1}{2}\big(\sqrt{\Delta_{ij}}+\sigma_j\big(\bar{\mathbf{Y}}^{(i)}\big)-\epsilon\big), \text{ if}\\ \Delta_{ij}>0 \text{ and } \phi<\frac{1}{2}\sigma_j^2\big(\bar{\mathbf{Y}}^{(i)}\big);\\ 0, \text{ otherwise.}\end{cases} \tag{16}$$

where $\Delta_{ij}=\big(\sigma_j\big(\bar{\mathbf{Y}}^{(i)}\big)+\epsilon\big)^2-4\lambda$ and $\phi=\frac{1}{8}\big(\sqrt{\Delta_{ij}}-\sigma_j\big(\bar{\mathbf{Y}}^{(i)}\big)-\epsilon\big)^2+\lambda\log\big(\frac{\sigma_j(\bar{\mathbf{Y}}^{(i)})+\sqrt{\Delta_{ij}}+\epsilon}{2\epsilon}\big)$.

Proof. We first convert the quadratic term of (15) into the DFT domain as

$$\begin{aligned}&\|\boldsymbol{\mathcal{X}}-\boldsymbol{\mathcal{Y}}\|_F^2=\tfrac{1}{m_3}\sum_{i=1}^{m_3}\|\bar{\mathbf{X}}^{(i)}-\bar{\mathbf{Y}}^{(i)}\|_F^2\\ &=\tfrac{1}{m_3}\sum_{i=1}^{m_3}\mathrm{tr}\big(\bar{\mathbf{X}}^{(i),*}\bar{\mathbf{X}}^{(i)}\big)+\mathrm{tr}\big(\bar{\mathbf{Y}}^{(i),*}\bar{\mathbf{Y}}^{(i)}\big)-2\mathrm{tr}\big(\bar{\mathbf{X}}^{(i),*}\bar{\mathbf{Y}}^{(i)}\big)\\ &=\tfrac{1}{m_3}\sum_{i=1}^{m_3}\Big(\sum_{j=1}^{d}\sigma_j^2\big(\bar{\mathbf{X}}^{(i)}\big)+\sigma_j^2\big(\bar{\mathbf{Y}}^{(i)}\big)\Big)-2\mathrm{tr}\big(\bar{\mathbf{X}}^{(i),*}\bar{\mathbf{Y}}^{(i)}\big).\end{aligned} \tag{17}$$

Before the next step of proof, we first introduce the following lemma.

Lemma 3. *[10] For matrices $\mathbf{A},\mathbf{B}\in\mathbb{C}^{m_1\times m_2}$, $\mathrm{tr}(\mathbf{A}^*\mathbf{B})\le\sum_{i=1}^{d}\sigma_i(\mathbf{A})\sigma_i(\mathbf{B})$ holds. The equality occurs if and only if there exist unitaries $\mathbf{U}$ and $\mathbf{V}$ that consistently singular value decompose $\mathbf{A}=\mathbf{U}\mathbf{S}_1\mathbf{V}^*$ and $\mathbf{B}=\mathbf{U}\mathbf{S}_2\mathbf{V}^*$ with $\mathbf{S}_1=\mathrm{diag}(\sigma_1(\mathbf{A}),\cdots,\sigma_d(\mathbf{A}))$ and $\mathbf{S}_1=\mathrm{diag}(\sigma_1(\mathbf{B}),\cdots,\sigma_d(\mathbf{B}))$, respectively.*

Inspired by Lemma 3, we perform the SVD of $\bar{\mathbf{X}}^{(i)}$ by using the same unitaries as those in the SVD of $\bar{\mathbf{Y}}^{(i)}$, which reaches the equality $\mathrm{tr}\big(\bar{\mathbf{X}}^{(i),*}\bar{\mathbf{Y}}^{(i)}\big)=\sum_{i=1}^{d}\sigma_i\big(\bar{\mathbf{X}}^{(i)}\big)\sigma_i\big(\bar{\mathbf{Y}}^{(i)}\big)$. Then, (17) can be rewritten into

$$\|\boldsymbol{\mathcal{X}}-\boldsymbol{\mathcal{Y}}\|_F^2=\frac{1}{m_3}\sum_{i=1}^{m_3}\sum_{j=1}^{d}\big(\sigma_j^2\big(\bar{\mathbf{X}}^{(i)}\big)-\sigma_j^2\big(\bar{\mathbf{Y}}^{(i)}\big)\big)^2. \tag{18}$$

Based upon Definition 1 and (18), we can equivalently convert (15) to solve each singular value minimization problem in parallel as

$$\min_{\sigma_j(\bar{\mathbf{X}}^{(i)})}\frac{1}{2}\big(\sigma_j\big(\bar{\mathbf{X}}^{(i)}\big)-\sigma_j\big(\bar{\mathbf{Y}}^{(i)}\big)\big)^2+\lambda\log\big(\sigma_j\big(\bar{\mathbf{X}}^{(i)}\big)/\epsilon+1\big). \tag{19}$$

According to the first-order optimality condition, the local optimality of (19) is reached either by the boundary point 0 or the stationary point, which is derived from the following quadratic equation

$$\sigma_j^2\big(\bar{\mathbf{X}}^{(i)}\big)-\big(\sigma_j\big(\bar{\mathbf{Y}}^{(i)}\big)-\epsilon\big)\sigma_j\big(\bar{\mathbf{X}}^{(i)}\big)+\lambda-\sigma_j\big(\bar{\mathbf{Y}}^{(i)}\big)\epsilon=0. \tag{20}$$

If $\Delta_{ij}\le 0$, the stationary point does not exist and hence the boundary point gives the optimal solution. If $\Delta_{ij}>0$, stationary points are two roots of (20). Considering the monotonicity of the objective function of (19) around those two stationary points, we find that only the larger stationary point, i.e., $\frac{1}{2}\big(\sqrt{\Delta_{ij}}+\sigma_j\big(\bar{\mathbf{Y}}^{(i)}\big)-\epsilon\big)$, is likely to be the optimal solution. Note that this stationary point leads to an objective value of ϕ, we compare it with the boundary point, which leads to an objective value of $\sigma_j^2\big(\bar{\mathbf{A}}^{(i)}\big)/2$. If $\phi<\sigma_j^2\big(\bar{\mathbf{A}}^{(i)}\big)/2$, the optimal solution is the stationary point, and vice versa. ■

2) Updating $\boldsymbol{\mathcal{F}}^{k+1}$ by Minimizing:

$$\boldsymbol{\mathcal{F}}^{k+1}=\underset{\|\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{F}})\|_1\le\delta}{\arg\min}\|\boldsymbol{\mathcal{F}}-(\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{T}})-\boldsymbol{\mathcal{X}}^{k+1}-\mu_k^{-1}\boldsymbol{\mathcal{Z}}^k)\|_F^2 \tag{21}$$

We define $\boldsymbol{\mathcal{A}}=\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{T}})-\boldsymbol{\mathcal{X}}^{k+1}-\mu_k^{-1}\boldsymbol{\mathcal{Z}}^k$ for simplicity, and then distinguish two distinct possibilities for solving (21). If $\|\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{A}})\|_1\le\delta$, the optimal solution is exactly $\boldsymbol{\mathcal{F}}^{k+1}=\boldsymbol{\mathcal{A}}$. If $\|\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{A}})\|_1>\delta$, the optimality can be obtained by the Lagrange multiplier method as $\boldsymbol{\mathcal{F}}^{k+1}=\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{F}})+\mathcal{P}_{\Omega_\perp}(\boldsymbol{\mathcal{A}})$. To be clearer, $\Omega_\perp$ denotes the complementary set of Ω, and the entry-wise $\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{F}})$ is given by [9]

$$[\boldsymbol{\mathcal{F}}]_{ijh}=\mathrm{sgn}([\boldsymbol{\mathcal{A}}]_{ijh})\max\{0,|[\boldsymbol{\mathcal{A}}]_{ijh}|-\theta\},\ \text{for } (i,j,h)\in\Omega \tag{22}$$

where $[\boldsymbol{\mathcal{F}}]_{ijh}$ denotes the (i,j,h)th entry of $\boldsymbol{\mathcal{F}}$, and $\theta\in(0,\|\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{A}})\|_\infty)$ is the Lagrange multiplier. Using the bisection or Newton's method to the boundary condition $\sum_{(i,j,h)\in\Omega}\max\{0,|[\boldsymbol{\mathcal{A}}]_{ijh}|-\theta\}=\delta$, we can solve for θ efficiently with a computational complexity of $\mathcal{O}(1)$.

3) Updating $\boldsymbol{\mathcal{Z}}^{k+1}$ by:

$$\boldsymbol{\mathcal{Z}}^{k+1}=\boldsymbol{\mathcal{Z}}^k+\mu_k(\boldsymbol{\mathcal{X}}^{k+1}+\boldsymbol{\mathcal{F}}^{k+1}-\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{T}})). \tag{23}$$

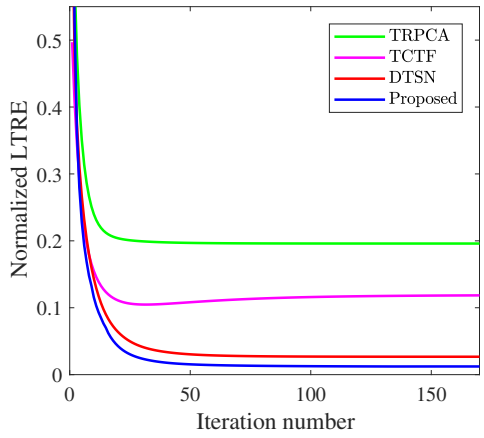


Fig. 1. Normalized LTRE versus iteration number for different methods.

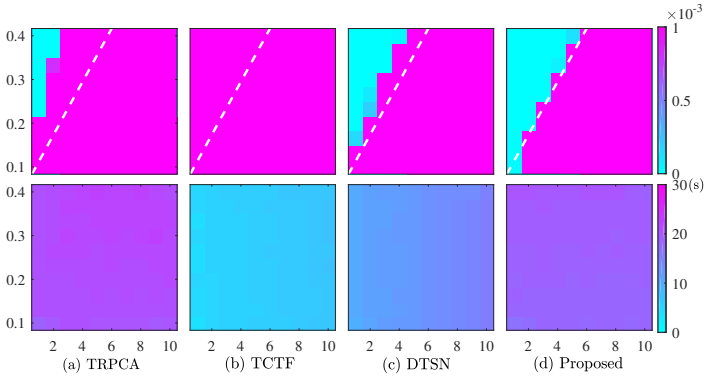


Fig. 2. Phase transition results achieved by different methods. X-axis and Y-axis represent the scene setting of rank r from 1 to 10, and the scene setting of sampling ratio from 10% to 40%, respectively. The color magnitudes in the first and second rows show the LTRE values and the execution times, respectively. Notice that the white dotted lines are provided as a guide for easier comparison.

C. Complexity Analysis

We analyse the computational complexity of the proposed algorithm, whose per-iteration requires the updates of $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{Z}}$ respectively. Updating $\boldsymbol{\mathcal{X}}$ needs to conduct the t-SVD, (inverse) DFT and matrix product [1], whose total complexities are at most $\mathcal{O}(m_1m_2m_3(\log m_3+\min\{m_1,m_2\})$. The update overhead of $\boldsymbol{\mathcal{F}}$ mainly depends on the entry-wise root finding for $\mathcal{P}_\Omega(\boldsymbol{\mathcal{F}})$ that consumes at most $\mathcal{O}(|\Omega|)$. To update $\boldsymbol{\mathcal{Z}}$ only requires an algebraic calculation with the low overhead. To sum up, the computational complexity of the proposed algorithm is $\mathcal{O}(m_1m_2m_3(\log m_3+\min\{m_1,m_2\})$ and thus is similar to that of TRPCA [4].

IV. Experimental Results

A. Synthetic Data Experiment

We first utilize the synthetic data to evaluate the low-rank tensor recovery performance of the proposed method, which is also compared with three state-of-the-art methods of TRPCA [4], tensor completion by tensor factorization (TCTF) [1], and double tensor Schatten-p norm minimization (DTSN) with $p_1=p_2=1$ [5]. As suggested in [1], we adopt the Matlab functions $\texttt{randn}(m_1,r,m_3)$ to generate $\boldsymbol{\mathcal{U}}\in\mathbb{R}^{m_1\times r\times m_3}$, and $\texttt{randn}(r,m_2,m_3)$ to generate $\boldsymbol{\mathcal{V}}\in\mathbb{R}^{r\times m_2\times m_3}$ with $m_1=m_2=m_3=50$ and $r=5$. Then the ground-truth low-rank tensor is synthesized by $\boldsymbol{\mathcal{X}}=\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{V}}$. To measure the robustness of different methods against sparse noise/outliers, we also add the noise tensor $\boldsymbol{\mathcal{F}}$ into the low-rank tensor to form the complete observation of $\boldsymbol{\mathcal{T}}=\boldsymbol{\mathcal{X}}+\boldsymbol{\mathcal{F}}$. For generating $\boldsymbol{\mathcal{F}}$, we randomly select 20% of its entries as outliers, which satisfy the standard normal distribution $\mathcal{N}(0,\sigma^2)$ with σ^2 being the variance corresponding to $\text{SNR}=\|\boldsymbol{\mathcal{X}}\|_F^2/(m_1m_2m_3\sigma^2)$. After randomly sampling 35% of the entries in $\boldsymbol{\mathcal{T}}$ as the known set Ω, the tensor recovery is applied to infer other unknown entries. We evaluate the recovery performance by the normalized low-rank tensor reconstruction error (LTRE) defined as

$$\text{LTRE}=\|\mathcal{P}_{\Omega_\perp}(\boldsymbol{\mathcal{X}})-\mathcal{P}_{\Omega_\perp}(\hat{\boldsymbol{\mathcal{X}}})\|_F^2/\|\mathcal{P}_{\Omega_\perp}(\boldsymbol{\mathcal{X}})\|_F^2 \qquad (24)$$

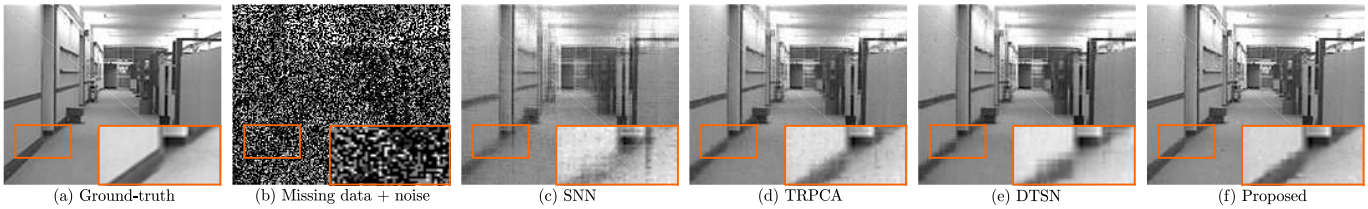
where $\hat{\boldsymbol{\mathcal{X}}}$ is the estimated result of the ground-truth $\boldsymbol{\mathcal{X}}$.

We measure the LTRE values of different methods versus iteration number at SNR=3dB. The parameters of all methods are finely tuned to obtain the best performance. As plotted in Fig. 1, the proposed method gains the stable and fast convergence after about 50 iterations. More importantly, our method yields the smaller reconstruction error than other competing methods, which verifies its efficiency to the low-rank recovery and its robustness to outliers. For further comparison, we carry out different scene settings for the rank r from 1 to 10, and for the set Ω with a sampling ratio from 10% to 40%. Fig. 2 shows the phase transition results of different methods with respect to the LTRE value and the execution time. All methods are conducted in MATLAB and we utilize a PC with Intel i7 CPU of 1.8 GHz and RAM of 8 GB. Note that TCTF and DTSN result in the faster execution speed since they employ the tensor factorization strategy, while TRPCA and the proposed method are performed on the undercomposed large-scale tensor. TCTF and TRPCA show the poor recovery performance since they are susceptible to noise. Compared with TRPCA, which uses the tensor nuclear norm, DTSN enforces the tensor Schatten-p norm to improve the reconstruction precision. However, its reconstruction precision is still inferior to our proposed method, which demonstrates the superior capacity of tensor logarithmic norm to approximate the low-rank structure over tensor nuclear norm and tensor Schatten-p norm.

B. Real Data Experiment

As mentioned in [11], a third-order tensor of grayscale video usually has the notable low-rank property, whose complete observation can be restored from the incomplete one. We apply the proposed method to restore video sequences from their sampled observations, and also compare our method with TRPCA, TCTF, DTSN, and sum of nuclear norm (SNN) [11]. The YUV Video Sequences database[1] is used here and we test the first ten frames of two sequences called "Hall" and "Akiyo"

[1] http://trace.eas.asu.edu/yuv/



(a) Ground-truth (b) Missing data + noise (c) SNN (d) TRPCA (e) DTSN (f) Proposed

Fig. 3. Restoration results achieved by different methods for the first frame of Hall sequence with 50% missing entries in salt-and-pepper noise at SNR=5dB.

TABLE I
RSNR RESULTS OF VIDEO RESTORATION ACHIEVED BY DIFFERENT METHODS. THE BEST RESULT IN EACH SCENE SETTING IS HIGHLIGHTED IN THE BOLD RESPECTIVELY.

Method	Hall				Akiyo			
	25% Sampling		50% Sampling		25% Sampling		50% Sampling	
	5dB	10dB	5dB	10dB	5dB	10dB	5dB	10dB
TCTF	16.42	19.32	18.48	22.83	16.67	20.59	18.42	19.76
SNN	19.58	24.03	24.24	29.33	21.64	26.74	26.47	30.46
TRPCA	24.63	29.68	28.62	35.74	27.35	32.36	30.94	38.15
DTSN	24.84	29.88	30.14	33.67	28.14	32.53	31.72	37.03
Proposed	**27.01**	**32.88**	**32.62**	**39.32**	**29.26**	**35.55**	**35.87**	**42.83**

respectively. The restoration performance is evaluated by the commonly-used peak signal-to-noise ratio (PSNR).

Table I shows the PSNR results from different scene settings, where input videos are randomly sampled at a ratio of 25% or 50% to generate Ω. To measure the noise robustness, the salt-and-pepper noise at SNR of 5dB or 10dB are further added into sampled videos. As listed in Table I, the proposed method gains the highest PSNR values over all experimental settings. It is observed that our method improves the PSNR values of 3.48dB and 3.42dB over TRPCA and DTSN on average. In addition to quantitative comparison, the visual comparison results for the first frame of Hall sequence are also displayed in Fig. 3. We find that the restoration results of SNN and TRPCA are sensitive to noise. DTSN shows the satisfactory noise robustness, but suffers from the blurry details. Our method achieves the remarkable preservation for image details as well as the enhanced noise suppression, which demonstrates the efficiency and accuracy of the proposed tensor logarithmic norm in video restoration.

CONCLUSIONS

In this paper, the tensor logarithmic norm was proposed as the nonconvex rank surrogate to address the problem of low-rank tensor approximation. We proved that compared with the convex surrogate of tensor nuclear norm, the proposed tensor logarithmic norm was a tighter envelope for the tensor average rank, and thus could enhance the sparsity of singular values in the tensor estimation. Synthetic experiment and real data experiment showed that our method was capable of reconstructing the low-rank structure of tensors with the better accuracy and outlier robustness against several state-of-the-art methods.

REFERENCES

[1] P. Zhou, C. Lu, Z. Lin, and C. Zhang, "Tensor factorization for low-rank tensor completion," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1152–1163, 2018.

[2] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 183–208, 2017.

[3] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 641–658, 2011.

[4] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5249–5257.

[5] H. Kong, X. Xie and Z. Lin, "t-Schatten-p norm for low-rank tensor recovery," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1405–1419, 2018.

[6] L. Chen, X. Jiang, X. Liu, T. Kirubarajan, and Z. Zhou "Outlier-Robust moving object and background decomposition via structured ℓ_p-regularized low-rank representation," to appear in *IEEE Trans. Emerging Topics Comput. Intell.*, pp. 1–19, 2019.

[7] Y. Huang, G. Liao, L. Zhang, Y. Xiang, J. Li, and A. Nehorai, "Efficient narrowband RFI mitigation algorithms for SAR systems with reweighted tensor structures," *IEEE Trans. Geosci. Remote Sens.*, pp. 9396–9409, 2019.

[8] M. J. D. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*, R. Fletcher, Ed. New York, NY, USA: Academic, 1969, pp. 283–298.

[9] X. Jiang, Z. Zhong, X. Liu, and H. C. So, "Robust matrix completion via alternating projection," *IEEE Sig. Process. Lett.*, vol. 24, no. 5, pp. 579–583, 2017.

[10] L. Mirsky, "A trace inequality of john von neumann," *Monatsh. Math.*, vol. 79, no. 4, pp. 303–306, 1975.

[11] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, 2013.