An L1-L2 Variant of Tubal Nuclear Norm for Guaranteed Tensor Recovery

Andong Wang[1,2] , **Guoxu Zhou**[1] , **Zhong Jin**[3] , **Qibin Zhao**[2]

[1] School of Automation, Guangdong University of Technology

[2] Tensor Learning Team, RIKEN AIP

[3] School of Computer Science and Engineering, Nanjing University of Science and Technology

w.a.d@outlook.com, gx.zhou@gdut.edu.cn, zhongjin@njust.edu.cn, qibin.zhao@riken.jp

Abstract

As a convex approximation of the tensor multi-rank which models low-rankness in the spectral domain, the Tubal Nuclear Norm (TNN) has shown superiority over traditional tensor nuclear norms in many tensor recovery tasks. However, it over-penalizes larger singular values of the Fourier block-diagonal matrix and may result in biased estimation. To this point, we define a non-convex $l_1 - \alpha l_2$ metric to approximate tensor multi-rank and introduce it into a new tensor sensing model with guaranteed recovery performance. The proximal operator of the metric is then proposed and utilized in an alternating direction method of multiplier (ADMM)-based algorithm to solve the problem. Effectiveness of the proposed metric is evaluated on both synthetic and real data.

1 Introduction

Tensor recovery from a few noisy linear measurements is an active topic in machine learning and signal processing [Liu *et al.*, 2020]. It is ill-posed when information quantity of the observations is lower than Degree of Freedom (DoF) of the signal, so the underlying tensor is often assumed to be low-rank to model a low DoF. Traditional low-rank tensor models based on CP/Tucker decompositions have been well studied. Recently, the tensor Singular Value Decomposition (t-SVD) [Kilmer *et al.*, 2013] based models have shown superior performance over traditional models in many tensor recovery tasks like image inpainting [Lu *et al.*, 2018].

Tensor recovery based on t-SVD often assumes the underlying tensor $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ has low tubal rank r, leading to a DoF at most $r(d_1 + d_2 - r)d_3$ which is significantly smaller than the entry number $d_1 d_2 d_3$ [Lu *et al.*, 2018]. It is proved that by TNN minimization, $O(r(d_1 + d_2 - r)d_3)$ noiseless Gaussian measurements are sufficient for exact recovery, and $O(r \max\{d_1, d_2\} d_3 \log^2(d_1 d_3 + d_2 d_3))$ observations sufficient for exact tensor completion [Lu *et al.*, 2018]. In the noisy setting, one needs $O(r \max\{d_1, d_2\} d_3 \log(d_1 d_3 + d_2 d_3))$ observations for approximate tensor completion via an iterative singular tube thresholding algorithm (ISTT) [Wang *et al.*, 2018]. The constrained [Lu *et al.*, 2018] and regularized [Zhang *et al.*, 2020] TNN-based models have been proposed for stable tensor compressive sensing with order-optimal sample size.

The above works are based on TNN which treats singular values of the Fourier block-diagonal matrix equally to pursue the convexity of the objective function. However, it may sometimes yield sub-optimal performance due to the biased approximation to multi-rank in the sense that TNN is dominated by singular values with large magnitudes, unlike multi-rank in which all nonzero singular values have equal contributions.

To address this issue, we define a non-convex $l_1 - \alpha l_2$ metric as a tighter multi-rank approximator than the convex TNN and propose a new estimator for tensor recovery in §3. Statistically, an upper bound on the estimation error is established in §4. Algorithmically, by deriving an analytical solution to the proximal operator of the metric in §5, simple ADMM can be used to compute the estimator. Experiments on both synthetic and real data in §6 show effectiveness of the proposed metric.

2 Notations and Preliminaries on t-SVD

First, main notations are listed in Table 1. For any matrix $\mathbf{M} \in \mathbb{C}^{d_1 \times d_2}$, define its Frobenius norm and nuclear norm as $\|\mathbf{M}\|_{\mathrm{F}} := (\sum_{ij} |\mathbf{M}_{ij}|^2)^{\frac{1}{2}}$ and $\|\mathbf{M}\|_* := \|\boldsymbol{\sigma}(\mathbf{M})\|_1$ respectively, where $\boldsymbol{\sigma}(\mathbf{M}) \in \mathbb{R}^{\min\{d_1, d_2\}}$ denotes the vector of singular values of $\mathbf{M}$ in non-ascending order.

Table 1: Some notations

Notation	Descriptions	Notation	Descriptions
x	scalar	$\mathbf{v}$	vector
$\mathbf{M}$	matrix	$\mathbf{T}$	tensor
$\widetilde{\mathcal{T}}$	$\mathrm{fft}(\mathcal{T}, [], 3)$	$\|\mathcal{T}\|_{\circledast}$	tubal nuclear norm
$\mathcal{T}_{ijk}$	$(i,j,k)_{th}$ entry of $\mathcal{T}$	$\|\mathcal{T}\|_{\mathrm{F}}$	$\sqrt{\sum_{ijk} \mathcal{T}_{ijk}^2}$
$\mathcal{T}(i,j,:)$	$(i,j)_{th}$ tube of $\mathcal{T}$	$\|\mathcal{T}\|_{\infty}$	$\max_{ijk} \lvert\mathcal{T}_{ijk}\rvert$
$\mathcal{T}(:,:,k)$ or $\mathcal{T}^{(k)}$	k_{th} frontal slice of $\mathcal{T}$	$\langle \mathcal{A}, \mathcal{B} \rangle$	$\sum_{ijk} \mathcal{A}_{ijk} \mathcal{B}_{ijk}$

Then, some notions of t-SVD will be defined.

Definition 1 ([Kilmer *et al.*, 2013])**.** *Any $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ has tensor singular value decomposition (t-SVD)*

$$\mathcal{T} := \mathcal{U} * \mathcal{S} * \mathcal{V}^{\top}, \tag{1}$$

where $*$ denotes the tensor t-product, $\mathcal{U} \in \mathbb{R}^{d_1 \times d_1 \times d_3}$ and $\mathcal{V} \in \mathbb{R}^{d_2 \times d_2 \times d_3}$ are orthogonal, $\mathcal{S} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is f-diagonal, $(\cdot)^{\top}$ denotes the tensor transpose.

Let $\widetilde{\mathcal{T}} := \mathrm{fft}(\mathcal{T}, [], 3) \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ be the tensor obtained after conducing DFT along the third mode of $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$.

Definition 2 ([Kilmer *et al.*, 2013]). *The tubal rank of $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is defined as the number of non-zero tubes of $\mathcal{S}$ in its t-SVD in Eq. (1), i.e., $r_{tubal}(\mathcal{T}) := \sum_i \mathbf{1}(\mathcal{S}(i,i,:) \neq \mathbf{0})$.*

The multi-rank of $\mathcal{T}$ is defined as the vector $\mathbf{r}_{\mathrm{m}}(\mathcal{T}) := (r_1, \cdots, r_{d_3}) \in \mathbb{R}^{d_3}$ whose i_{th} element $r_i = \mathrm{rank}(\widetilde{\mathcal{T}}^{(i)})$.

Definition 3 ([Kilmer *et al.*, 2013]). *Define the Fourier block-diagonal matrix of $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ as follows:*

$$\overline{\mathbf{T}} := \mathrm{blkdiagDFT}(\mathcal{T}) = \begin{bmatrix} \widetilde{\mathcal{T}}^{(1)} & & \\ & \ddots & \\ & & \widetilde{\mathcal{T}}^{(d_3)} \end{bmatrix} \in \mathbb{C}^{d_1 d_3 \times d_2 d_3}.$$

Definition 4 ([Lu *et al.*, 2016]). *The tubal nuclear norm (TNN) of $\mathcal{T}$ is defined as the scaled nuclear norm of its Fourier block-diagonal matrix $\overline{\mathbf{T}}$,i.e.,*

$$\|\mathcal{T}\|_{\circledast} := \frac{1}{d_3}\|\overline{\mathbf{T}}\|_* = \frac{1}{d_3}\sum_{i=1}^{d_3} \|\widetilde{\mathcal{T}}^{(i)}\|_*.$$

According to Definition 4, TNN is a rescaled l_1-norm of singular values of the Fourier block-diagonal matrix, and thus encourages a low multi-rank structure which models low-rankness in the spectral domain.

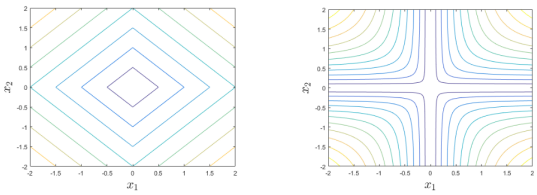
3 The Proposed Model For Tensor Recovery

3.1 A New Tensor Low-rank Regularizer

The $l_1 - \alpha l_2$ metric was proposed to impose sparser solutions in sparse signal recovery than the commonly used l_1-norm.

Definition 5 ([Lou and Yan, 2018]). *With $\alpha > 0$, define the $l_1 - \alpha l_2$ metric of any $\mathbf{x} \in \mathbb{R}^d$ as $\|\mathbf{x}\|_{\alpha,1-2} := \|\mathbf{x}\|_1 - \alpha\|\mathbf{x}\|_2$.*

A 2-D case is plotted in Fig. 1 [Yao *et al.*, 2016], showing $l_1 - \alpha l_2$ metric approaches the axes closer for smaller values.



(a) l_1-norm (b) $l_1 - \alpha l_2$ metric ($\alpha = 1$)

Figure 1: Level curves of the l_1-norm and $l_1 - \alpha l_2$ metric ($\alpha = 1$).

Using $l_1 - \alpha l_2$ metric to impose sparsity in singular values, we define a matrix metric to encourage low-rankness.

Definition 6. *The $* - \alpha$F metric of a matrix $\mathbf{M}$ is defined as the $l_1 - \alpha l_2$-metric of the vector of its singular values, i.e.,*

$$\|\mathbf{M}\|_{\alpha,*-\mathrm{F}} := \|\boldsymbol{\sigma}(\mathbf{M})\|_{\alpha,1-2} = \|\mathbf{M}\|_* - \alpha\|\mathbf{M}\|_{\mathrm{F}}. \quad (2)$$

The matrix $* - \alpha$F metric is further extended to tensors as a non-convex surrogate of TNN to impose spectral low-rankess.

Definition 7. *The $\circledast - \alpha$F metric of a tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is defined as the scaled $* - \alpha$F metric of its Fourier block-diagonal matrix $\overline{\mathbf{T}}$, i.e.*

$$\|\mathcal{T}\|_{\alpha,\circledast-\mathrm{F}} := \frac{1}{d_3}\|\overline{\mathbf{T}}\|_{\alpha,*-\mathrm{F}} = \|\mathcal{T}\|_{\circledast} - \frac{\alpha}{\sqrt{d_3}}\|\mathcal{T}\|_{\mathrm{F}}. \quad (3)$$

3.2 The Proposed Estimator

Suppose one has N noisy observations of $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$:

$$y_i = \langle \mathcal{L}^*, \mathcal{X}_i \rangle + \xi_i, \quad i = 1, \cdots, N, \quad (4)$$

where $\mathcal{X}_i$'s are known design tensors and ξ_i's are noises.

Let $\mathbf{y} = (y_1, \cdots, y_N)^\top$ and $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_N)^\top$. Define the observation operator $\mathfrak{X}(\cdot) := (\langle \cdot, \mathcal{X}_1 \rangle, \cdots, \langle \cdot, \mathcal{X}_N \rangle)^\top \in \mathbb{R}^N$ with adjoint operator $\mathfrak{X}^*(\mathbf{z}) := \sum_{i=1}^N z_i \mathcal{X}_i, \forall \mathbf{z} \in \mathbb{R}^N$. Then, the model (4) can be rewritten as $\mathbf{y} = \mathfrak{X}(\mathcal{L}^*) + \boldsymbol{\xi}$.

When $\mathfrak{X}$ is the *random Gaussian design*, i.e., $\mathcal{X}_i$'s are random tensors with *i.i.d.* standard Gaussian entries, Eq. (4) gives the tensor compressive sensing model in [Lu *et al.*, 2018].

To recover $\mathcal{L}^*$, we use the proposed $\circledast - \alpha$F metric to impose spectral low-rankness and define the following estimator:

$$\hat{\mathcal{L}} \in \underset{\mathcal{L}}{\mathrm{argmin}} \|\mathcal{L}\|_{\alpha,\circledast-\mathrm{F}} \quad \text{s.t. } \|\mathbf{y} - \mathfrak{X}(\mathcal{L})\| \leq \tau, \quad (5)$$

where constant τ is chosen as a noise level satisfying $\|\boldsymbol{\xi}\| \leq \tau$.

4 Statistical Analysis

The Restricted Isometry Property (RIP) is a powerful analysis tool for sparse and low-rank recovery. As for the t-SVD framework, [Zhang *et al.*, 2019] defined a tensor variant of RIP based on the tubal rank. Here, a different tensor RIP based on the multi-rank is given to analyze the proposed estimator.

Definition 8 (Tensor multi-rank based RIP (tm-RIP)). *The observation operator $\mathfrak{X} : \mathbb{R}^{d_1 \times d_2 \times d_3} \to \mathbb{R}^N$ is said to satisfy the tensor multi-rank based RIP with order $\mathbf{r}$ and constant $\delta_{\mathbf{r}}^{\mathfrak{X}}$, if $\delta_{\mathbf{r}}^{\mathfrak{X}}$ is the smallest $\delta \in (0,1)$ such that*

$$(1-\delta)\|\mathcal{T}\|_{\mathrm{F}}^2 \leq \|\mathfrak{X}(\mathcal{T})\|_2^2 \leq (1+\delta)\|\mathcal{T}\|_{\mathrm{F}}^2, \quad (6)$$

holds for all $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ with multi-rank $\mathbf{r}$.

The following theorem shows a random sub-Gaussian design with sufficient samples satisfies the tm-RIP.

Theorem 1 (tm-RIP for sub-Gaussian design). *Fix $\delta, \varepsilon \in (0,1)$ and suppose the sample size N in Model (4) satisfy*

$$N \geq C\delta^{-2}\max\{(d_1+d_2+1)\|\mathbf{r}\|_1, \log(\varepsilon^{-1})\}, \quad (7)$$

then with probability at least $1-\varepsilon$, any sub-Guassian design $\mathfrak{X}$ satisfies $\delta_{\mathbf{r}}^{\mathfrak{X}} \leq \delta$ for all $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ with multi-rank $\mathbf{r}$.

Remark 1. *Note that sub-Gaussian distributions is a larger class of random distributions, including zero-mean Gaussian distributions, symmetric Bernoulli distributions and all zero-mean bounded distributions. Thus, similar to Theorem 1 in [Zhang* et al.*, 2019], the proposed Theorem 1 characterizes the behavior of numerous random designs in term of the tm-RIP.*

Then, upper bound on the estimation error is established.

Theorem 2 (Stable recovery under tm-RIP). *Suppose the true tensor $\mathcal{L}^*$ in Model (4) has multi-rank $\mathbf{r} = (r_1, \cdots, r_{d_3})$. If there exists a positive integer s such that*

$$\Phi_{\mathbf{r},s} := 1 - \delta_{2\mathbf{r}+\mathbf{s}}^{\mathfrak{X}} - \frac{\sqrt{4\|\mathbf{r}\|_1 + 2\alpha^2 d_3}}{\sqrt{s}-\alpha}(\delta_{2\mathbf{r}+\mathbf{s}}^{\mathfrak{X}} + \delta_{2\mathbf{s}}^{\mathfrak{X}}) > 0, \quad (8)$$

where $\mathbf{s} = (s, \cdots, s) \in \mathbb{R}^{d_3}$, then any $\hat{\mathcal{L}}$ in Eq. (5) obeys

$$\|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\mathrm{F}} \leq \frac{2\sqrt{1+\delta_{2\mathbf{r}+\mathbf{s}}^{\mathfrak{X}}}}{\Phi_{\mathbf{r},s}}\tau. \quad (9)$$

Remark 2. *Theorem 2 indicates that $\hat{\mathcal{L}} = \mathcal{L}^*$ as $\tau = 0$, that is, Model (5) can guarantee exact recovery in the noiseless setting when the observation operator $\mathfrak{X}$ satisfies Eq. (8).*

Remark 3. *When $d_3 = 1$ and $\alpha = 1$, the results in Eqs. (8) and (9) are consistent with those of low-rank matrix recovery in Theorem 3.8 of [Ma* et al.*, 2017] (when $t = 0$).*

5 Optimization Algorithm

To solve Problem (5), we first propose the proximal operator of the $\circledast - \alpha$F metric and then design an ADMM-based algorithm.

5.1 Proximal Operator of The Proposed Metric

The proximal operator of $l_1 - \alpha l_2$ metric is first introduced.

Lemma 1 ([Lou and Yan, 2018])**.** *The proximal operator of $l_1 - \alpha l_2$ metric, i.e., $\mathfrak{P}_\lambda^{\alpha,1-2}(\mathbf{x}_0) := \operatorname{argmin}_{\mathbf{x}} \lambda\|\mathbf{x}\|_{\alpha,1-2} + \frac{1}{2}\|\mathbf{x} - \mathbf{x}_0\|_{\mathrm{F}}^2$, can be given in closed form*

$$\mathfrak{P}_\lambda^{\alpha,1-2}(\mathbf{x}_0) = \mathbf{z} + \lambda\alpha\frac{\mathbf{z}}{\|\mathbf{z}\|}, \tag{10}$$

where $\mathbf{z} = \mathfrak{P}_\lambda^{\|\cdot\|_1}(\mathbf{x}_0) := \operatorname{sign}(\mathbf{x}_0) \odot \max\{|\mathbf{x}_0| - \lambda, \mathbf{0}\}$, and let $\mathfrak{P}_\lambda^{\alpha,1-2}(\mathbf{x}_0) = \mathbf{0}$ if $\mathbf{z} = \mathbf{0}$.

We then extend Lemma 1 to the matrix $* - \alpha$F metric.

Lemma 2. *The proximal operator of $* - \alpha$F metric at any point $\mathbf{M}$ with SVD $\mathbf{M} = \mathbf{U}\operatorname{diag}(\boldsymbol{\sigma}(\mathbf{M}))\mathbf{V}^{\mathsf{H}}$ can be given as*

$$\mathfrak{P}_\lambda^{\alpha,*-\mathrm{F}}(\mathbf{M}) = \mathbf{U}\operatorname{diag}(\mathbf{s})\mathbf{V}^{\mathsf{H}}, \tag{11}$$

where $\mathbf{s} = \mathfrak{P}_\lambda^{\alpha,1-2}(\boldsymbol{\sigma}(\mathbf{M}))$.

We are now ready to establish the closed-form expression for the proximal operator of the proposed $\circledast - \alpha$F metric.

Lemma 3. *The proximal operator of $\|\cdot\|_{\alpha,\circledast-\mathrm{F}}$ is given by*

$$\mathfrak{P}_\lambda^{\alpha,\circledast-\mathrm{F}}(\mathcal{T}_0) := \frac{1}{d_3}\text{blkdiagDFT}^{-1}(\mathfrak{P}_\lambda^{\alpha,*-\mathrm{F}}(\overline{\mathbf{T}}_0)). \tag{12}$$

5.2 An ADMM-based Algorithm

The ADMM framework [Boyd *et al.*, 2011] is applied to solve the proposed model. Let $\mathbb{B}_\tau := \{\boldsymbol{\epsilon} | \|\boldsymbol{\epsilon}\|_2 \le \tau\}$. Adding auxiliary variables yields an equivalent formulation to Problem (5):

$$\begin{aligned} &\min_{\mathcal{L},\mathcal{K},\boldsymbol{\epsilon}} \|\mathcal{L}\|_{\alpha,\circledast-\mathrm{F}} \\ &\text{s.t. } \mathcal{K} = \mathcal{L},\ \mathfrak{X}(\mathcal{K}) + \boldsymbol{\epsilon} = \mathbf{y}, \boldsymbol{\epsilon} \in \mathbb{B}_\tau. \end{aligned} \tag{13}$$

To solve Problem (13), an ADMM-based algorithm is described in Algorithm 1. Convergence analysis of Algorithm 1 can be derived based on [Lou and Yan, 2018] for the proximal operator of $l_1 - \alpha l_2$ metric and [Boyd *et al.*, 2011] for the general ADMM framework.

The computational complexity is analyzed as follows. For simplicity, let $D = d_1 d_2 d_3$. Updating $\mathcal{L}$ involves the proximal operator of $\|\cdot\|_{\alpha,\circledast-\mathrm{F}}$ in Eq. (12) which costs $O\big(D(\min\{d_1, d_2\} + \log d_3)\big)$; By precomputing $(\mathbb{I} + \mathfrak{X}^*\mathfrak{X}\mathfrak{X}^*\mathfrak{X})^{-1}$ and $\mathfrak{X}^*\mathfrak{X}$ which costs $O(D^3 + ND^2)$, the cost of updating $\mathcal{K}$ is $O(D^2)$; Updating $\boldsymbol{\epsilon}$ involves computing the projection into l_2-ball $\text{Proj}_\tau^{\|\cdot\|_2}(\cdot)$ [Boyd *et al.*, 2011] and $\mathfrak{X}(\mathcal{K}^{t+1})$ which mainly costs $O(ND)$; Updating $\mathcal{Z}$ and $\mathbf{z}$ costs $O(D^2)$. Supposing the iteration number is T, the overall computational complexity will be $O\big(D^3 + TD^2 + TD(\min\{d_1, d_2\} + \log d_3)\big)$, which is very expensive for large tensors. In some special cases (like tensor completion) where $\langle \mathcal{X}_i, \mathcal{L}\rangle$ operates on an element of $\mathcal{L}$, $(\mathbb{I} + \mathfrak{X}^*\mathfrak{X}\mathfrak{X}^*\mathfrak{X})^{-1}$ and $\mathfrak{X}^*\mathfrak{X}$ can be computed in $O(D)$. Hence, the total complexity of Algorithm 1 will drop to $O\big(TD(\min\{d_1, d_2\} + \log d_3)\big)$.

Algorithm 1: ADMM for Problem (13)

Input: $\{\mathcal{X}_i\}_i, \mathbf{y}, \rho, \varepsilon, T_{\max}$.
1: $\mathcal{L}^0 = \mathcal{K}^0 = \boldsymbol{\epsilon}^0 = \mathcal{Z}^0 = \mathbf{z}^0 = \mathbf{0}$;
2: **for** $t = 0$ to $T_{\max} - 1$ **do**
3: Update $\mathcal{L}^{t+1} = \mathfrak{P}_{1/\rho}^{\alpha,\circledast-\mathrm{F}}(\mathcal{K}^{t+1} + \mathcal{Z}^{t+1}/\rho)$, and $\boldsymbol{\epsilon}^{t+1} = \text{Proj}_\tau^{\|\cdot\|_2}(\mathbf{y} - \mathfrak{X}(\mathcal{K}^{t+1}) - \mathbf{z}^{t+1}/\rho)$.
4: Update $\mathcal{K}^{t+1} = (\mathbb{I} + \mathfrak{X}^*\mathfrak{X}\mathfrak{X}^*\mathfrak{X})^{-1}\mathcal{K}_0$, where $\mathcal{K}_0 = \mathfrak{X}^*\mathfrak{X}\big(\mathfrak{X}^*(\mathbf{y} - \boldsymbol{\epsilon}^{t+1} - \mathbf{z}^{t+1}/\rho)\big) + \mathcal{L}^{t+1} - \mathcal{Z}^{t+1}/\rho$.
5: Check stopping criteria: $\|\mathcal{K}^{t+1} - \mathcal{L}^{t+1}\|_\infty \le \varepsilon$, $\|\boldsymbol{\epsilon}^{t+1} + \mathfrak{X}(\mathcal{K}^{t+1}) - \mathbf{y}\|_\infty \le \varepsilon$, and $\|\mathcal{T}^{t+1} - \mathcal{T}^t\|_\infty \le \varepsilon,\ \forall \mathcal{T} \in \{\mathcal{L}, \mathcal{K}, \boldsymbol{\epsilon}\}$.
6: Update $\mathcal{Z}^{t+1} = \mathcal{Z}^t + \rho(\mathcal{K}^{t+1} - \mathcal{L}^{t+1})$ and $\mathbf{z}^{t+1} = \mathbf{z}^t + \rho\big(\boldsymbol{\epsilon}^{t+1} + \mathfrak{X}(\mathcal{K}^{t+1}) - \mathbf{y}\big)$.
7: **end for**
Output: $\hat{\mathcal{L}} = \mathcal{L}^{t+1}$.

6 Experiments

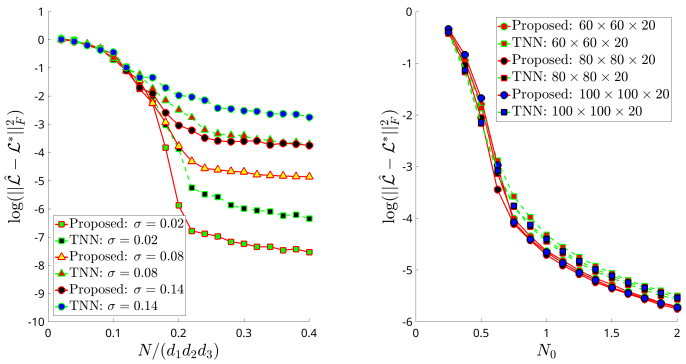
We evaluate the effectiveness of the proposed metric by first conducting tensor compressive sensing and tensor completion on synthetic data and then carrying out noisy inpainting on color images. We simply set the constant $\alpha = 1$ in Model (5) on synthetic data and $\alpha = 0.9$ on color images.

Tensor Compressive Sensing. We conduct tensor compressive sensing with random Gaussian design where the design tensors $\mathcal{X}_i$ are formed with *i.i.d.* standard Gaussian entries. The true tensors $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ with tubal rank r^* are generated by $\mathcal{L}^* = \mathcal{P} * \mathcal{Q}$, where $\mathcal{P} \in \mathbb{R}^{d_1 \times r^* \times d_3}$ and $\mathcal{Q} \in \mathbb{R}^{r^* \times d_2 \times d_3}$ are *i.i.d.* sampled from $\mathcal{N}(0, 1)$. We consider the tensors with square frontal slices (i.e. $d_1 = d_2 = d$) for simplicity. For tensors $\mathcal{L}^*$ of size $20 \times 20 \times 5$ with tubal rank $r^* = 1$, we consider additive Gaussian noises $\mathcal{N}(0, \sigma^2)$ with noise level $\sigma \in \{0.02, 0.08, 0.14\}$ along with the observation ratio $N/(d_1 d_2 d_3) \in [0.02 : 0.02 : 0.4]$.

We compare the proposed metric with TNN [Lu *et al.*, 2018]. The averaged estimation error in 50 runs are shown in log scale in Fig. 2-(a), showing that the proposed metric is more accurate than TNN.

Tensor Completion. We also consider tensor completion from noisy partial observations where the design tensors $\mathcal{X}_i$ are standard tensor bases. For an underlying tensor $\mathcal{L}^*$ with unit F-norm, we add *i.i.d.* $\mathcal{N}(0, \sigma^2)$ noises with level $\sigma = 0.1/\sqrt{d_1 d_2 d_3}$. By choosing $d_1 = d_2 = d \in \{60, 80, 100\}$, $d_3 = 20$ and $r^* = \lceil \log^{1/2} d \rceil$, we consider 3 different problem sizes. The estimation error is plotted versus the *rescaled sample size* defined as $N_0 := N/r^* d_1 d_3 \log(d_1 d_3 + d_2 d_3)$ like [Wang *et al.*, 2018].

The proposed metric is compared with TNN with results shown in Fig. 2-(b). We can see from Fig. 2-(b) that the

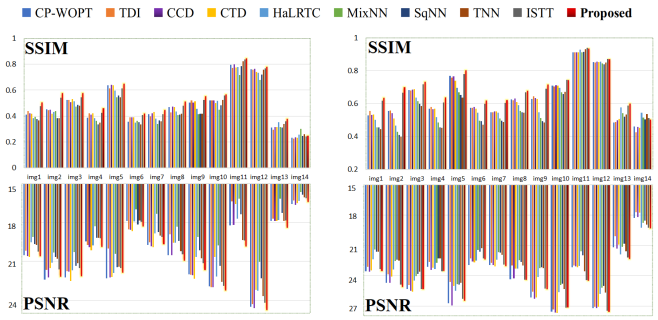


(a) Tensor compressive sensing (b) Tensor completion

Figure 2: Performance evaluation of the proposed metric and TNN in tensor compressive sensing and tensor completion.

proposed metric achieves higher accuracy than TNN.

Noisy Color Image Inpainting. Noisy color image inpainting aims to recover a color image from its noisy incomplete observations. We test the 14 images of size $256 \times 256 \times 3$ used in [Wang *et al.*, 2018]. Given a color image $\mathcal{M} \in \mathbb{R}^{d_1 \times d_2 \times 3}$, it is first polluted by additive *i.i.d.* zero-mean Gaussian noise with standard deviation $\sigma = c_0 \|\mathcal{M}\|_F / \sqrt{3 d_1 d_2}$ and then sampled uniformly with ratio p. Specifically, we consider two settings $(c_0, p) \in \{(0.1, 0.1), (0.1, 0.3)\}$ for a given image. The proposed model is compared with nine models including CP-WOPT, TDI, CCD, CTD, HaLRTC, MixNN, SqNN, TNN and ISTT in Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) values (see [Wang *et al.*, 2018] for more descriptions of the competitors). The experimental results are reported in Fig. 3. As can be seen from Fig. 3, the proposed model outperforms its competitors in most cases.



(a) $(c_0, p) = (0.1, 0.1)$ (b) $(c_0, p) = (0.1, 0.3)$

Figure 3: Performance evaluation of ten tensor completion models for noisy color image inpainting in PSNR and SSIM values.

7 Conclusion

The $l_1 - \alpha l_2$ metric is first generalized to a tensor low-rank regularizer to exploit spectral low-rankness. Then, we define a constrained tensor estimator based on the new metric. Statistically, upper bound on the estimation error is established through the multi-rank based tensor RIP. Algorithmically, an ADMM-based algorithm is developed based on the proximal operator of proposed metric. Effectiveness of the new metric is demonstrated through experiments.

8 Acknowledgments

This work is partially supported by National Natural Science Foundation of China [Grant Nos. 61872188, 61703209], by the China Postdoctoral Science Foundation [Grant No. 2020M672536], and by the Natural Science Foundation of Guangdong Province [Grant No. 2020A1515010671].

References

[Boyd *et al.*, 2011] S. Boyd, N. Parikh, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[Kilmer *et al.*, 2013] M. E Kilmer, K. Braman, N. Hao, and R. C. Hoover. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM Journal on Matrix Analysis and Applications*, 34(1):148–172, 2013.

[Liu *et al.*, 2020] X. Liu, S. Aeron, V. Aggarwal, and X. Wang. Low-tubal-rank tensor completion using alternating minimization. *IEEE TIT*, 66(3):1714–1737, 2020.

[Lou and Yan, 2018] Yifei Lou and Ming Yan. Fast l1–l2 minimization via a proximal operator. *Journal of Scientific Computing*, 74(2):767–785, 2018.

[Lu *et al.*, 2016] 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 *CVPR*, pages 5249–5257, 2016.

[Lu *et al.*, 2018] C. Lu, J. Feng, Z. Lin, et al. Exact low tubal rank tensor recovery from gaussian measurements. In *IJCAI*, pages 1948–1954, 2018.

[Ma *et al.*, 2017] Tian-Hui Ma, Yifei Lou, and Ting-Zhu Huang. Truncated l_{1-2} models for sparse recovery and rank minimization. *SIAM Journal on Imaging Sciences*, 10(3):1346–1380, 2017.

[Wang *et al.*, 2018] A. Wang, D. Wei, B. Wang, and Z. Jin. Noisy low-tubal-rank tensor completion through iterative singular tube thresholding. *IEEE Access*, 6:35112–35128, 2018.

[Yao *et al.*, 2016] Quanming Yao, James T Kwok, and Xiawei Guo. Fast learning with nonconvex l1-2 regularization. *arXiv preprint arXiv:1610.09461*, 2016.

[Zhang *et al.*, 2019] Feng Zhang, Wendong Wang, Jingyao Hou, Jianjun Wang, and Jianwen Huang. Tensor restricted isometry property analysis for a large class of random measurement ensembles. *arXiv preprint arXiv:1906.01198*, 2019.

[Zhang *et al.*, 2020] Feng Zhang, Wendong Wang, Jianwen Huang, Jianjun Wang, and Yao Wang. Rip-based performance guarantee for low-tubal-rank tensor recovery. *Journal of Computational and Applied Mathematics*, 374:112767, 2020.