Supplementary material of AAAI 2020

Robust Tensor Decomposition via Orientation Invariant Tubal Nuclear Norms

Andong Wang, Chao Li, Zhong Jin, Qibin Zhao

In this file, proofs of theorems and lemmas in the main body are first given in **Supp-§-A**. Then, the proposed Algorithm 1 and Algorithm 2 are presented in **Supp-§-B**.

Supp-§-A. Proofs of Theorems and Lemmas

More Preliminaries of t-SVD

Before proving the theorems and lemmas, we will introduce more preliminaries omitted in the main submission due to space limitation.

Tensor Singular Value Decomposition At a high level, the framework of t-SVD treats a 3-way tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ as a matrix M whose $(i,j)^{\text{th}}$ entry $\mathsf{M}(i,j)$ is $\mathcal{T}(i,j,:)$ (i.e., the $(i,j)^{\text{th}}$ tube of $\mathcal{T}$).

By using circular convolution of tube vectors instead of product of scalars, the t-product (in Definition 1) is an extension of standard matrix multiplication. This t-product can be implemented efficiently in the Fourier domain according to the relationship between circular convolution and DFT (Kilmer et al. 2013). Specifically, let $\widetilde{\mathcal{T}} = \text{fft}(\mathcal{T}, [], 3)$ denote its Fourier version obtained conducting 1D-DFT on the tubes of $\mathcal{T}$. Given $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, let $\mathbf{T}^{(i)}$ (or $\mathcal{T}^{(i)}$) denote its i^{th} frontal slice $\mathcal{T}(:,:,i)$. The t-product of $\mathcal{T}_1$ and $\mathcal{T}_2$ in original domain is equivalent to frontal slice-wise matrix product of $\widetilde{\mathcal{T}}_1$ and $\widetilde{\mathcal{T}}_2$ in spectral domain, i.e,

$$\mathcal{T} = \mathcal{T}_1 * \mathcal{T}_2 \Leftrightarrow \widetilde{\mathbf{T}}^{(l)} = \widetilde{\mathbf{T}}_1^{(l)} \widetilde{\mathbf{T}}_2^{(l)}, \ \forall l \in [d_3]. \tag{15}$$

The t-SVD (in Definition 2) is a 3-way extension of standard SVD. It decomposes any tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ as

$$\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top, \tag{16}$$

where $\mathcal{U} \in \mathbb{R}^{d_1 \times d_1 \times d_3}, \mathcal{V} \in \mathbb{R}^{d_2 \times d_2 \times d_3}$ are orthogonal tensors, and $\mathcal{S} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is an f-diagonal tensor (see Fig. 9 (Wang and Jin 2017)). The t-SVD is indeed constructed in the spectral (Fourier) domain by using the relationship between circular convolution and DFT (Kilmer et al. 2013; Lu et al. 2019). Relevant concepts including tensor transpose, f-diagonal tensor and orthogonal tensor, are defined as follows.

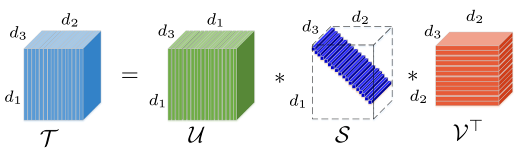

Figure 9: Illustration of t-SVD.

Definition 8 (Tensor transpose (Kilmer et al. 2013)). *Let $\mathcal{T}$ be a tensor of size $d_1 \times d_2 \times d_3$, then $\mathcal{T}^\top$ is the $d_2 \times d_1 \times d_3$ tensor obtained by transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through d_3. In spectral domain, we have $(\widetilde{\mathcal{T}^\top})^{(l)} = (\widetilde{\mathbf{T}}^{(l)})^{\mathsf{H}}, \ \forall l \in [d_3]$.*

Definition 9 (Identity tensor (Kilmer et al. 2013)). *The identity tensor $\mathcal{I} \in \mathbb{R}^{d \times d \times d_3}$ is a tensor whose first frontal slice is the $d \times d$ identity matrix and all other frontal slices are zero. In spectral domain, we have $\widetilde{\mathcal{I}}^{(l)} = \mathbf{I}_d \in \mathbb{R}^{d \times d}, \ \forall l \in [d_3]$.*

Definition 10 (f-diagonal tensor (Kilmer et al. 2013)). *A tensor is called f-diagonal if each frontal slice of the tensor is a diagonal matrix.*

Definition 11 (Orthogonal tensor (Kilmer et al. 2013)). *A tensor $\mathcal{Q} \in \mathbb{R}^{d \times d \times d_3}$ is orthogonal if*

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

In spectral domain, we have

$$(\widetilde{\mathbf{Q}}^{(l)})^{\mathsf{H}} \widetilde{\mathbf{Q}}^{(l)} = \widetilde{\mathbf{Q}}^{(l)} (\widetilde{\mathbf{Q}}^{(l)})^{\mathsf{H}} = \mathbf{I}_d \in \mathbb{R}^{d \times d}, \ \forall l \in [d_3], \tag{17}$$

which means all frontal slices of the Fourier version of an orthogonal tensor $\mathcal{Q}$ are unitary matrices.

The block diagonal matrix of 3-way tensors are further defined for the convenience of analysis.

Definition 12. (Kilmer et al. 2013). *Let $\overline{\mathbf{T}}$ (or $\overline{\mathcal{T}}$) denote the block-diagonal matrix of the tensor $\widetilde{\mathcal{T}}$ in the Fourier domain, i.e.,*

$$\overline{\mathbf{T}} := \begin{bmatrix} \widetilde{\mathbf{T}}^{(1)} & & \\ & \ddots & \\ & & \widetilde{\mathbf{T}}^{(d_3)} \end{bmatrix} \in \mathbb{C}^{d_1 d_3 \times d_2 d_3} \tag{18}$$

Then, it holds naturally according to Eq. (15)

$$\mathcal{T} = \mathcal{T}_1 * \mathcal{T}_2 \Leftrightarrow \overline{\mathbf{T}} = \overline{\mathbf{T}}_1 \overline{\mathbf{T}}_2.$$

We further have the following relationship for t-SVD

$$\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top \Leftrightarrow \overline{\mathbf{T}} = \overline{\mathbf{U}}\, \overline{\mathbf{S}}\, \overline{\mathbf{V}}^{\mathsf{H}},$$

which also indicates that the average rank and TNN satisfy

$$\text{rank}_{\text{avg}}(\mathcal{T}) = \frac{1}{d_3} \text{rank}(\overline{\mathbf{T}}) = \frac{1}{d_3} \text{rank}(\overline{\mathbf{S}}),$$

$$\|\mathcal{T}\|_\star = \frac{1}{d_3} \|\overline{\mathbf{T}}\|_* = \frac{1}{d_3} \|\overline{\mathbf{S}}\|_*.$$

Further, the property of DFT indicates that the tubal rank of $\mathcal{T}$ defined in Eq. (2) is lower bounded by the average rank:

$$\begin{aligned} \text{rank}_{\text{tb}}(\mathcal{T}) &:= \#\{i \,|\mathcal{S}(i,i,:) \neq \mathbf{0}\} \\ &= \#\{i \,|\widetilde{\mathcal{S}}(i,i,:) \neq \mathbf{0}\} \\ &= \max_{l \in [d_3]} \text{rank}(\mathbf{S}^{(l)}) \\ &\geq \frac{1}{d_3} \sum_{l=1}^{d_3} \text{rank}(\mathbf{S}^{(l)}) \\ &\geq \text{rank}_{\text{avg}}(\mathcal{S}). \end{aligned} \tag{19}$$

The inner product between two tensors $\mathcal{T}_1$ and $\mathcal{T}_2$ is defined as $\langle \mathcal{T}_1, \mathcal{T}_2 \rangle := \text{vec}(\mathcal{T}_1)^{\mathsf{H}} \text{vec}(\mathcal{T}_2)$. The the inner product of two 3-D tensors $\mathcal{T}_1, \mathcal{T}_2 \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ and the inner product of their corresponding block diagonal matrices $\overline{\mathbf{T}_1}, \overline{\mathbf{T}_2} \in \mathbb{C}^{d_1 d_3 \times d_2 d_3}$ has the relationship

$$\langle \mathcal{T}_1, \mathcal{T}_2 \rangle = \frac{1}{d_3} \left\langle \widetilde{\mathcal{T}_1}, \widetilde{\mathcal{T}_2} \right\rangle = \frac{1}{d_3} \left\langle \overline{\mathbf{T}_1}, \overline{\mathbf{T}_2} \right\rangle. \tag{20}$$

The relationship between matrix nuclear norm and matrix F-norm holds for any $\mathbf{M}$:

$$\|\mathbf{M}\|_* \leq \sqrt{\text{rank}(\mathbf{M})} \|\mathbf{M}\|_{\text{F}}. \tag{21}$$

Similar relationship between TNN and F-norm also holds for any tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ as follows:

$$\begin{aligned} \|\mathcal{T}\|_\star &= \frac{1}{d_3} \|\overline{\mathcal{T}}\|_* \leq \frac{1}{d_3} \sqrt{d_3 \text{rank}_{\text{tb}}(\mathcal{T})} \|\overline{\mathbf{T}}\|_{\text{F}} \\ &= \frac{1}{d_3} \sqrt{d_3 \text{rank}_{\text{tb}}(\mathcal{T})} (\sqrt{d_3} \|\mathcal{T}\|_{\text{F}}) \\ &= \sqrt{\text{rank}_{\text{tb}}(\mathcal{T})} \|\mathcal{T}\|_{\text{F}}. \end{aligned} \tag{22}$$

It is also known that, for any tensor $\mathcal{T}$, the l_1-norm and the F-norm has the following relationship

$$\|\mathcal{T}\|_1 = \sqrt{\|\mathcal{T}\|_{l_0}} \|\mathcal{T}\|_{\text{F}}. \tag{23}$$

Decomposability of Tubal Nuclear Norm In (Recht, Fazel, and Parrilo 2007), the matrix nuclear norm is proved to have the following property, called additivity.

Lemma 3 (Additivity of Matrix Nuclear Norm(Recht, Fazel, and Parrilo 2007))**.** *Given* $\mathbf{A}$ *and* $\mathbf{B}$ *of the same dimension, if* $\mathbf{A}\mathbf{B}^{\mathsf{H}} = \mathbf{0}$ *and* $\mathbf{A}^{\mathsf{H}}\mathbf{B} = \mathbf{0}$*, then* $\|\mathbf{A} + \mathbf{B}\|_* = \|\mathbf{A}\|_* + \|\mathbf{B}\|_*$.

Here, we will show that the tubal nuclear norm also has the property.

Lemma 4 (Additivity of Tubal Nuclear Norm)**.** *Given* $\mathcal{T}_1, \mathcal{T}_2 \in \mathbb{R}^{d_1 \times d_2 \times d_3}$*, If* $\mathcal{T}_1 * \mathcal{T}_2{}^\top = \mathbf{0}$ *and* $\mathcal{T}_1{}^\top * \mathcal{T}_2 = \mathbf{0}$*, then*

$$\|\mathcal{T}_1 + \mathcal{T}_2\|_\star = \|\mathcal{T}_1\|_\star + \|\mathcal{T}_2\|_\star. \tag{24}$$

Proof. Using the relationship between a 3D tensor and its block-diagonal matrix we have

$$\begin{aligned} \mathcal{T}_1 * \mathcal{T}_2{}^\top = \mathbf{0} \Rightarrow \overline{\mathbf{T}_1}\, \overline{\mathbf{T}_2}^{\mathsf{H}} = \mathbf{0}, \\ \mathcal{T}_1{}^\top * \mathcal{T}_2 = \mathbf{0} \Rightarrow \overline{\mathbf{T}_1}^{\mathsf{H}} \overline{\mathbf{T}_2} = \mathbf{0}. \end{aligned} \tag{25}$$

Thus, we obtain

$$\begin{aligned} \|\mathcal{T}_1 + \mathcal{T}_2\|_\star &= \frac{1}{d_3} \|\overline{\mathbf{T}_1 + \mathbf{T}_2}\|_* \\ &= \frac{1}{d_3} \left(\|\overline{\mathbf{T}_1}\|_* + \|\overline{\mathbf{T}_2}\|_* \right) \\ &= \|\mathcal{T}_1\|_\star + \|\mathcal{T}_2\|_\star. \end{aligned} \tag{26}$$

□

Suppose $\mathcal{X} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ with tubal rank r^* has reduced t-SVD as follows

$$\mathcal{X} = \mathcal{U}_{\mathcal{X}} * \mathcal{S}_{\mathcal{X}} * \mathcal{V}_{\mathcal{X}}{}^\top, \tag{27}$$

where $\mathcal{U}_{\mathcal{X}} \in \mathbb{R}^{d_1 \times r^* \times d_3}$ and $\mathcal{V}_{\mathcal{X}} \in \mathbb{R}^{d_2 \times r^* \times d_3}$ are orthogonal and $\mathcal{S}_{\mathcal{X}} \in \mathbb{R}^{r^* \times r^* \times d_3}$ is f-diagonal. Define a tensor space T as follows:

$$T = \Big\{ \mathcal{U}_{\mathcal{X}} * \mathcal{A} + \mathcal{B} * \mathcal{V}_{\mathcal{X}}^\top \ : \ \text{where } \mathcal{A} \in \mathbb{R}^{r^* \times d_2 \times d_3}, \mathcal{B} \in \mathbb{R}^{d_1 \times r^* \times d_3} \Big\}.$$

We further define the projectors to T and $T^\perp$ as $\mathcal{P}_T : \mathbb{R}^{d_1 \times d_2 \times d_3} \to \mathbb{R}^{d_1 \times d_2 \times d_3}$ and $\mathcal{P}_{T^\perp} : \mathbb{R}^{d_1 \times d_2 \times d_3} \to \mathbb{R}^{d_1 \times d_2 \times d_3}$ respectively

$$\begin{aligned} \mathcal{P}_T(\mathcal{T}) &= \mathcal{U}_{\mathcal{X}} * \mathcal{U}_{\mathcal{X}}^\top * \mathcal{T} + \mathcal{T} * \mathcal{V}_{\mathcal{X}} * \mathcal{V}_{\mathcal{X}}^\top \\ &\quad - \mathcal{U}_{\mathcal{X}} * \mathcal{U}_{\mathcal{X}}^\top * \mathcal{T} * \mathcal{V}_{\mathcal{X}} * \mathcal{V}_{\mathcal{X}}^\top, \\ \mathcal{P}_{T^\perp}(\mathcal{T}) &= (\mathcal{I} - \mathcal{U}_{\mathcal{X}} * \mathcal{U}_{\mathcal{X}}^\top) * \mathcal{T} * (\mathcal{I} - \mathcal{V}_{\mathcal{X}} * \mathcal{V}_{\mathcal{X}}^\top). \end{aligned} \tag{28}$$

Thus, we have

$$\begin{aligned} &\text{rank}_{\text{tb}}(\mathcal{P}_T(\mathcal{T})) \\ &\leq \text{rank}_{\text{tb}}(\mathcal{U}_{\mathcal{X}} * \mathcal{U}_{\mathcal{X}}^\top * \mathcal{T}) + \text{rank}_{\text{tb}}((\mathcal{I} - \mathcal{U}_{\mathcal{X}} * \mathcal{U}_{\mathcal{X}}^\top) * \mathcal{T} * \mathcal{V}_{\mathcal{X}} * \mathcal{V}_{\mathcal{X}}^\top) \\ &\leq 2\text{rank}_{\text{tb}}(\mathcal{X}). \end{aligned} \tag{29}$$

Equipped with Lemma 4, we will present an inequality frequently used in our work as follows.

Lemma 5. *Given* $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$*, we have*

$$\|\mathcal{X} + \mathcal{P}_{T^\perp}(\mathcal{T})\|_\star = \|\mathcal{X}\|_\star + \|\mathcal{P}_{T^\perp}(\mathcal{T})\|_\star. \tag{30}$$

Proof. It is easy to check that $\mathcal{X} * \mathcal{P}_{T^\perp}(\mathcal{T})^\top = \mathbf{0}$ and $\mathcal{X}^\top * \mathcal{P}_{T^\perp}(\mathcal{T}) = \mathbf{0}$. By Lemma 4, we have $\|\mathcal{L}^* + \mathcal{P}_{T^\perp}(\mathcal{T})\|_\star = \|\mathcal{X}\|_\star + \|\mathcal{P}_{T^\perp}(\mathcal{T})\|_\star$. □

Note that Lemmas 4 and 5 indicate that the tubal nuclear norm belongs to the class of decomposable norms (Negahban et al. 2009).

Proofs of Lemma 1

Lemma 1 asserts that $\vec{\mathfrak{r}}_{\text{a}}(\mathcal{T}) \leq \min\{\vec{\mathfrak{r}}_{\text{t}}(\mathcal{T}), \vec{\mathfrak{r}}_{\text{Tucker}}(\mathcal{T})\}$. Thus the low OIAR assumption is weaker than the commonly used low Tucker rank assumption. In other words, many data like color images which satisfy low Tucker rank assumption also satisfy the low OIAR assumption. Here, we prove Lemma 1.

Proof of Lemma 1. Given any tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$, let $\mathcal{K} = \mathcal{T}_{[k]} \in \mathbb{R}^{d_k \times (D d_k^{-1} d_{k+1}^{-1}) \times d_{k+1}}$ denote its mode-$(k, k+1)$ 3d-unfolding ($k \in [K]$). We first show that $\vec{\mathfrak{r}}_{\text{a}}(\mathcal{T}) \leq \vec{\mathfrak{r}}_{\text{t}}(\mathcal{T})$. Indeed, it holds that

$$(\vec{\mathfrak{r}}_{\text{a}}(\mathcal{T}))_k = \text{rank}_{\text{avg}}(\mathcal{K}) \overset{(i)}{\leq} \text{rank}_{\text{tb}}(\mathcal{K}) = (\vec{\mathfrak{r}}_{\text{t}}(\mathcal{T}))_k, \tag{31}$$

where inequality *(i)* holds due to Eq. (19).

Then, we show $\vec{\mathfrak{r}}_{\text{a}}(\mathcal{T}) \leq \vec{\mathfrak{r}}_{\text{Tucker}}(\mathcal{T})$. On the one hand, according to (Lu et al. 2019), we have

$$(\vec{\mathfrak{r}}_{\text{a}}(\mathcal{T}))_k = \text{rank}_{\text{avg}}(\mathcal{K}) \leq \text{rank}(\mathbf{K}_{(1)}), \tag{32}$$

where $\mathbf{K}_{(1)}$ is the mode-1 matricization of $\mathcal{K}$. On the other hand, since $\mathcal{K}$ is the mode-$(k, k+1)$ 3d-unfolding of $\mathcal{T}$, it holds naturally that

$$\text{rank}(\mathbf{K}_{(1)}) = \text{rank}(\mathbf{T}_{(k)}) = (\vec{\mathfrak{r}}_{\text{Tucker}}(\mathcal{T}))_k, \tag{33}$$

where $\mathbf{T}_{(k)}$ is the mode-k matricization of $\mathcal{T}$. Thus, we have $\vec{\mathfrak{r}}_{\mathrm{a}}(\mathcal{T}) \leq \vec{\mathfrak{r}}_{\mathrm{Tucker}}(\mathcal{T})$. Putting things together, we obtain $\vec{\mathfrak{r}}_{\mathrm{a}}(\mathcal{T}) \leq \min\{\vec{\mathfrak{r}}_{\mathrm{t}}(\mathcal{T}), \vec{\mathfrak{r}}_{\mathrm{Tucker}}(\mathcal{T})\}$.

□

Proofs of Lemma 2

Before proving Lemma 2, we need the following lemma.

Lemma 6 (Dual norm of TNN (Lu et al. 2018)). *The tubal nuclear norm and the tensor spectral norm are dual to each other.*

Proof of Lemma 2. The lemma can be proved through formulating the following maximization problems:

$$\|\mathcal{T}\|_{\star\mathrm{o}}^* = \sup_{\mathcal{M}} \langle \mathcal{M}, \mathcal{T}\rangle, \text{ s.t. } \|\mathcal{M}\|_{\star\mathrm{o}} \leq 1, \tag{34}$$

and

$$\|\mathcal{T}\|_{\star\iota}^* = \sup_{\mathcal{M}} \langle \mathcal{M}, \mathcal{T}\rangle\,, \text{ s.t. } \|\mathcal{M}\|_{\star\iota} \leq 1. \tag{35}$$

They are constrained maximization problems. We prove the first part. Since Problem (34) satisfies Slatter's condition, the strong duality holds. Thus, we only need to show that its dual problem agrees with

$$\inf_{\sum_k \mathcal{T}^{(k)}=\mathcal{T}} \max_k \{w_k^{-1}\|\mathcal{T}^{(k)}_{[k]}\|\}. \tag{36}$$

Dual to Fenchel's duality theorem, we have the following equality:

$$\begin{aligned}&\sup_{\mathcal{M}} \left(\langle \mathcal{M}, \mathcal{T}\rangle + \delta(\|\mathcal{M}\|_{\star\mathrm{o}} \leq 1)\right)\\ &= \inf_{\{\mathcal{T}^{(k)}\}_k} \left(\delta(\sum_k \mathcal{T}^{(k)} = \mathcal{T}) + \max_k\{w_k^{-1}\|\mathcal{T}^{(k)}_{[k]}\|\}\right),\end{aligned}$$

where $\delta(C)$ is the indicator of condition C (0 if C is true and $+\infty$ otherwise). In this way, the first part is proved. The second part can be proved similarly.

□

Proofs of Theorem 1 and Theorem 3

For notational simplicity, we also define 3d-unfolding operator for $\mathcal{T} \in \mathbb{R}^{d_1\times\cdots\times d_K}$ as $\mathfrak{F}_k(\mathcal{T}) := \mathcal{T}_{[k]}$ and let $\mathfrak{F}_k^{-1}(\cdot)$ denote its inverse, i.e., $\mathfrak{F}_k^{-1}(\mathcal{T}_{[k]}) = \mathcal{T}$.

Proof of Theorem 1 In this subsection, we provide the proof of Theorem 1.

Proof. Recall model OITNN-O:

$$\begin{aligned}(\hat{\mathcal{L}}_{\mathrm{o}}, \hat{\mathcal{S}}_{\mathrm{o}}) \in &\operatorname*{argmin}_{\mathcal{L},\mathcal{S}}\ f(\mathcal{L},\mathcal{S}),\\ &\text{s.t. } \|\mathcal{L}\|_{l_\infty} \leq \alpha,\end{aligned}$$

where $f(\mathcal{L},\mathcal{S}) := \frac{1}{2}\|\mathcal{Y} - \mathcal{L} - \mathcal{S}\|_{\mathrm{F}} + \lambda_{\mathrm{o}}\|\mathcal{L}\|_{\star\mathrm{o}} + \mu_{\mathrm{o}}\|\mathcal{S}\|_{l_1}$.

Let $\Delta_{\mathrm{o}}^L = \mathcal{L}^* - \hat{\mathcal{L}}_{\mathrm{o}}$ and $\Delta_{\mathrm{o}}^S = \mathcal{S}^* - \hat{\mathcal{S}}_{\mathrm{o}}$. Using the optimality of $(\hat{\mathcal{L}}_{\mathrm{o}}, \hat{\mathcal{S}}_{\mathrm{o}})$, we have:

$$f(\hat{\mathcal{L}}_{\mathrm{o}}, \hat{\mathcal{S}}_{\mathrm{o}}) \leq f(\mathcal{L}^*, \mathcal{S}^*). \tag{37}$$

By the observation model of RTD, we have $\mathcal{Y} - \mathcal{L}^* - \mathcal{S}^* = \mathcal{E}$. According to Eq. (37), we obtain

$$\begin{aligned}&\frac{1}{2}\left(\|\Delta_{\mathrm{o}}^L\|_{\mathrm{F}}^2 + \|\Delta_{\mathrm{o}}^S\|_{\mathrm{F}}^2\right)\\ &\leq \underbrace{\lambda_{\mathrm{o}}\left(\|\mathcal{L}^*\|_{\star\mathrm{o}} - \|\mathcal{L}^* - \Delta_{\mathrm{o}}^L\|_{\star\mathrm{o}}\right)}_{\mathrm{I}} + \underbrace{\mu_{\mathrm{o}}\left(\|\mathcal{S}^*\|_{l_1} - \|\mathcal{S}^* - \Delta_{\mathrm{o}}^S\|_{l_1}\right)}_{\mathrm{II}}\\ &\quad + \underbrace{\left\langle \Delta_{\mathrm{o}}^L, \Delta_{\mathrm{o}}^S\right\rangle}_{\mathrm{III}} + \underbrace{\left\langle \Delta_{\mathrm{o}}^L, \mathcal{E}\right\rangle}_{\mathrm{IV}} + \underbrace{\left\langle \Delta_{\mathrm{o}}^S, \mathcal{E}\right\rangle}_{\mathrm{V}},\end{aligned} \tag{38}$$

where the right hand side involves 5 items I to V. We will upper bound **Items I** to **V** as follows.

Bound Item I: Let $\mathcal{P}_k(\cdot) = \mathcal{P}_{\mathfrak{F}_k(\mathcal{L}^*)}(\cdot)$ (see the definition of $\mathcal{P}_T$ in Eq. (28)). For any tensor $\mathbf{\Delta} \in \mathbb{R}^{d_1\times\cdots\times d_K}$, define

$$\Delta_k' = \mathcal{P}_k(\mathfrak{F}_k(\mathbf{\Delta})), \text{ and } \Delta_k'' = \mathbf{\Delta}_{[k]} - \Delta_k'.$$

Using Lemma 4 directly yields

$$\|\mathcal{L}^*{}_{[k]} - \Delta_k''\|_\star = \|\mathcal{L}^*{}_{[k]}\|_\star + \|\Delta_k''\|_\star.$$

It leads to

$$\begin{aligned}\|\mathcal{L}^*{}_{[k]} - (\Delta_{\mathrm{o}}^L)_{[k]}\|_\star &= \|(\mathcal{L}^*{}_{[k]} - (\Delta_{\mathrm{o}}^L)_k'') - (\Delta_{\mathrm{o}}^L)_k'\|_\star\\ &\geq \|(\mathcal{L}^*{}_{[k]} - (\Delta_{\mathrm{o}}^L)_k'')\|_\star - \|(\Delta_{\mathrm{o}}^L)_k'\|_\star\\ &= \|\mathcal{L}^*{}_{[k]}\|_\star + \|(\Delta_{\mathrm{o}}^L)_k''\|_\star - \|(\Delta_{\mathrm{o}}^L)_k'\|_\star.\end{aligned}$$

Thus, we have

$$\begin{aligned}\mathrm{I} &= \lambda_{\mathrm{o}} \sum_k w_k \left(\|\mathcal{L}^*{}_{[k]}\|_\star - \|(\mathcal{L}^* - \Delta_{\mathrm{o}}^L)_{[k]}\|_\star\right)\\ &\leq \lambda_{\mathrm{o}} \sum_k w_k \Big(\|\mathcal{L}^*{}_{[k]}\|_\star - \left(\|\mathcal{L}^*{}_{[k]}\|_\star + \|(\Delta_{\mathrm{o}}^L)_k''\|_\star - \|(\Delta_{\mathrm{o}}^L)_k'\|_\star\right)\Big)\\ &= \lambda_{\mathrm{o}} \sum_k w_k \|(\Delta_{\mathrm{o}}^L)_k'\|_\star - \lambda_{\mathrm{o}} \sum_k w_k \|(\Delta_{\mathrm{o}}^L)_k''\|_\star.\end{aligned} \tag{39}$$

Bound Item II: Let S be the true sparse tensor $\mathcal{S}^*$, i.e., $S = \mathrm{supp}(\mathcal{S}^*) = \left\{(i_1, i_2, \cdots, i_K) | \mathcal{S}^*_{i_1 i_2 \cdots i_K} \neq 0\right\}$. According to the decomposability of l_1-norm (Negahban et al. 2009), any tensor $\mathcal{T} \in \mathbb{R}^{d_1\times\cdots\times d_K}$ satisfies

$$\|\mathcal{T}\|_{l_1} = \|\mathcal{T}_S\|_{l_1} + \|\mathcal{T}_{S^\perp}\|_{l_1}.$$

Then, we have

$$\begin{aligned}\|\mathcal{S}^* - \Delta_{\mathrm{o}}^S\|_{l_1} &= \|(\mathcal{S}^* - (\Delta_{\mathrm{o}}^S)_{S^\perp}) - (\Delta_{\mathrm{o}}^S)_S\|_{l_1}\\ &\geq \|\mathcal{S}^* - (\Delta_{\mathrm{o}}^S)_{S^\perp}\|_{l_1} - \|(\Delta_{\mathrm{o}}^S)_S\|_{l_1}\\ &\geq \|\mathcal{S}^*\|_{l_1} + \|(\Delta_{\mathrm{o}}^S)_{S^\perp}\|_{l_1} - \|(\Delta_{\mathrm{o}}^S)_S\|_{l_1},\end{aligned}$$

leading to the bound

$$\mathrm{II} \leq \mu_{\mathrm{o}}\|(\Delta_{\mathrm{o}}^S)_S\|_{l_1} - \mu_{\mathrm{o}}\|(\Delta_{\mathrm{o}}^S)_{S^\perp}\|_{l_1}. \tag{40}$$

Bound Items III, IV and V. Due to the feasibility of $\hat{\mathcal{L}}$, we have $\|\hat{\mathcal{L}}\|_{l_\infty} \leq \alpha$. Then, by the triangular inequality, we have

$$\|\Delta_{\mathrm{o}}^L\|_{l_\infty} = \|\mathcal{L}^* - \hat{\mathcal{L}}_{\mathrm{o}}\|_{l_\infty} \leq \|\mathcal{L}^*\|_{l_\infty} + \|\hat{\mathcal{L}}_{\mathrm{o}}\|_{l_\infty} \leq 2\alpha.$$

Using the definition of dual norm, we have

$$\begin{aligned}\mathrm{III} &\leq \|\Delta_{\mathrm{o}}^L\|_{l_\infty}\|\Delta_{\mathrm{o}}^S\|_{l_1} \leq 2\alpha\|\Delta_{\mathrm{o}}^S\|_{l_1},\\ \mathrm{IV} &\leq \|\Delta_{\mathrm{o}}^L\|_{\star\mathrm{o}}\|\mathcal{E}\|_{\star\mathrm{o}}^*,\\ \mathrm{V} &\leq \|\Delta_{\mathrm{o}}^S\|_{l_1}\|\mathcal{E}\|_{l_\infty}.\end{aligned} \tag{41}$$

Combining Eq. (38) and Eqs (39)-(41) yields

$$\begin{aligned}
&\frac{1}{2}\big(\|\Delta_{\rm o}^L\|_{\rm F}^2+\|\Delta_{\rm o}^S\|_{\rm F}^2\big)\\
&\le \lambda_{\rm o}\sum_k w_k\|(\Delta_{\rm o}^L)'_k\|_\star - \lambda_{\rm o}\sum_k w_k\|(\Delta_{\rm o}^L)''_k\|_\star + \mu_{\rm o}\|(\Delta_{\rm o}^S)_S\|_{l_1}\\
&\quad - \mu_{\rm o}\|(\Delta_{\rm o}^S)_{S^\perp}\|_{l_1} + \|\mathcal{E}\|_{\star{\rm o}}^*\|\Delta_{\rm o}^L\|_{\star{\rm o}} + (\|\mathcal{E}\|_{l_\infty}+2\alpha)\|\Delta_{\rm o}^S\|_{l_1}\\
&\le (\lambda_{\rm o}+\|\mathcal{E}\|_{\star{\rm o}}^*)\sum_k w_k\|(\Delta_{\rm o}^L)'_k\|_\star - (\lambda_{\rm o}-\|\mathcal{E}\|_{\star{\rm o}}^*)\sum_k w_k\|(\Delta_{\rm o}^L)''_k\|_\star\\
&+ \big(\mu_{\rm o}+(\|\mathcal{E}\|_{l_\infty}+2\alpha)\big)\|(\Delta_{\rm o}^S)_S\|_{l_1} - \big(\mu_{\rm o}-(\|\mathcal{E}\|_{l_\infty}+2\alpha)\big)\|(\Delta_{\rm o}^S)_{S^\perp}\|_{l_1}.
\end{aligned}$$

Choosing

$$\lambda_{\rm o} > 2\|\mathcal{E}\|_{\star{\rm o}}^*, \tag{42}$$

and

$$\mu_{\rm o} \ge 2(\|\mathcal{E}\|_{l_\infty}+2\alpha), \tag{43}$$

we have

$$\begin{aligned}
&\lambda_{\rm o}\sum_k w_k\|(\Delta_{\rm o}^L)''_k\|_\star + \mu_{\rm o}\|(\Delta_{\rm o}^S)_{S^\perp}\|_{l_1}\\
&\le 3\big(\lambda_{\rm o}\sum_k w_k\|(\Delta_{\rm o}^L)'_k\|_\star + \mu_{\rm o}\|(\Delta_{\rm o}^S)\|_{l_1}\big).
\end{aligned} \tag{44}$$

Note that according to Eq. (38) and the triangular inequality, we obtain

$$\begin{aligned}
&\frac{1}{2}\big(\|\Delta_{\rm o}^L\|_{\rm F}^2+\|\Delta_{\rm o}^S\|_{\rm F}^2\big)\\
&\le (\lambda_{\rm o}+\|\mathcal{E}\|_{\star{\rm o}}^*)\big(\sum_k w_k\|(\Delta_{\rm o}^L)'_k\|_\star + \sum_k w_k\|(\Delta_{\rm o}^L)''_k\|_\star\big)\\
&+ \big(\mu_{\rm o}+(\|\mathcal{E}\|_{l_\infty}+2\alpha)\big)\big(\|(\Delta_{\rm o}^S)_S\|_{l_1}+\|(\Delta_{\rm o}^S)_{S^\perp}\|_{l_1}\big).
\end{aligned} \tag{45}$$

That leads to

$$\|\Delta_{\rm o}^L\|_{\rm F}^2+\|\Delta_{\rm o}^S\|_{\rm F}^2 \le 16\lambda_{\rm o}\sum_k w_k\|(\Delta_{\rm o}^L)'_k\|_\star + 16\mu_{\rm o}\|(\Delta_{\rm o}^S)_S\|_{l_1}. \tag{46}$$

By the definition of $(\Delta_{\rm o}^L)'_k$, we have

$$\text{rank}_{\rm tb}((\Delta_{\rm o}^L)'_k) \le 2\text{rank}_{\rm tb}(\mathcal{L}^*{}_{[k]}) = 2\underline{r}_k^{\rm o}, \tag{47}$$

and

$$\|(\Delta_{\rm o}^L)'_k\|_{\rm F} \le \|(\Delta_{\rm o}^L)_{[k]}\|_{\rm F} = \|\Delta_{\rm o}^L\|_{\rm F}. \tag{48}$$

We also have $\|(\Delta_{\rm o}^S)_S\|_{l_0} \le |S| = s$. Then we reach the inequality:

$$\begin{aligned}
&\|\Delta_{\rm o}^L\|_{\rm F}^2+\|\Delta_{\rm o}^S\|_{\rm F}^2\\
&\le 16\lambda_{\rm o}\sum_k w_k\sqrt{2\underline{r}_k^{\rm o}}\|\Delta_{\rm o}^L\|_{\rm F} + 16\mu_{\rm o}\sqrt{s}\|\Delta_{\rm o}^S\|_{\rm F}.
\end{aligned} \tag{49}$$

The usage of $ab \le a^2/4 + b^2$ leading to the conclusion of Theorem 1. □

Proof of Theorem 3 The key of proving Theorem 3 is to bound the quantity $\|\mathcal{E}\|_{\star{\rm o}}^*$, when $\mathcal{E}$ denotes the tensor whose entries are *i.i.d.* Gaussian $\mathcal{N}(0,\sigma^2)$. To bound this quantity, we need the following two lemmas:

Lemma 7. *For any K-way ($K \ge 3$) tensor $\mathcal{T} \in \mathbb{R}^{d_1\times\cdots\times d_K}$, the following inequality holds:*

$$\|\mathcal{T}\|_{\star\iota}^* \le \frac{1}{K^2}\sum_k w_k^{-1}\|\mathcal{T}_{[k]}\|. \tag{50}$$

Proof. Recall the formulation of $\|\mathcal{T}\|_{\star\iota}^*$ as follows

$$\|\mathcal{T}\|_{\star{\rm o}}^* := \inf_{\sum_k \mathcal{T}^{(k)}=\mathcal{T}} \max_k\{w_k^{-1}\|\mathcal{T}_{[k]}^{(k)}\|\}. \tag{51}$$

Letting

$$\mathcal{T}^{(k)} = \frac{w_k\|\mathcal{T}_{[k]}^{(k)}\|^{-1}}{\sum_k w_k\|\mathcal{T}_{[k]}^{(k)}\|^{-1}}\mathcal{T}, \tag{52}$$

then for any $k\in[K]$,

$$\begin{aligned}
w_k^{-1}\|\mathcal{T}_{[k]}^{(k)}\| &= w_k^{-1}\frac{w_k\|\mathcal{T}_{[k]}^{(k)}\|^{-1}}{\sum_k w_k\|\mathcal{T}_{[k]}^{(k)}\|^{-1}}\|\mathcal{T}_{[k]}^{(k)}\|\\
&= \frac{1}{\sum_k w_k\|\mathcal{T}_{[k]}^{(k)}\|^{-1}}\\
&\le \frac{1}{K^2}\sum_k w_k^{-1}\|\mathcal{T}_{[k]}^{(k)}\|,
\end{aligned} \tag{53}$$

where the last inequality holds because the “harmonic mean" is no larger than the “arithmetic mean". In this way, the lemma is proved. □

Lemma 8. *Let $\mathcal{T} \in \mathbb{R}^{d_1\times d_2\times d_3}$ be random tensors with i.i.d. Gaussian entries $\mathcal{N}(0,1)$. Then the following inequality holds*

$$\|\mathcal{T}\| \le \sqrt{d_1 d_3} + \sqrt{d_2 d_3} + t, \tag{54}$$

with probability at least $1-\exp\big(-ct^2/d_3\big)$.

Proof. By letting $\mathcal{U}$ and $\mathcal{V}$ in Lemma 9 of (Lu et al. 2018) be the identity tensors, this lemma can be proved directly. □

We also have the following lemma to bound the l_∞-norm of $\mathcal{E}$:

Lemma 9. *Let $\mathcal{T} \in \mathbb{R}^{d_1\times\cdots\times d_K}$ be random tensors with i.i.d. Gaussian entries $\mathcal{N}(0,1)$. Then for the following inequality hold*

$$\|\mathcal{G}\|_{l_\infty} \le \sqrt{2\log(2D)} + t, \tag{55}$$

with probability at least $1-\exp\big(-ct^2\big)$.

Proof of Theorem 3. Using Lemma III.2, we have for any $k\in[K]$:

$$\|\mathcal{E}_{[k]}\| \le 2\sigma\tilde{d}_k, \tag{56}$$

with probability at least $1-\exp(-c_k\tilde{d}_k^2/d_{k+1})$. Taking union bound, we have with probability at least $1-\sum_k\exp(-c_k\tilde{d}_k^2/d_{k+1})$,

$$\|\mathcal{E}\|_{\star{\rm o}}^* \le \frac{2\sigma}{K^2}\sum_k w_k^{-1}\tilde{d}_k. \tag{57}$$

Accodring to Lemma 9, we also have

$$\|\mathcal{E}\|_{l_\infty} \le 4\sigma\sqrt{\log D}, \tag{58}$$

with probability at least $1-\exp(-c'D)$.

Combing Eqs. (57)-(58) and Theorem 1, Theorem 3 can be proved. □

Proofs of Theorem 2 and Theorem 4

For notational simplicity, we recall the definition 3d-unfolding operator for $\mathcal{T} \in \mathbb{R}^{d_1\times\cdots\times d_K}$ as $\mathfrak{F}_k(\mathcal{T}) := \mathcal{T}_{[k]}$ and its $\mathfrak{F}_k^{-1}(\cdot)$ such that $\mathfrak{F}_k^{-1}(\mathcal{T}_{[k]}) = \mathcal{T}$.

Proof of Theorem2 In this subsection, we provide the proof of Theorem 2.

Proof. Recall model OITNN-L:

$$(\{\hat{\mathcal{L}}^{(k)}\}_k, \hat{\mathcal{S}}_\iota) \in \underset{\{\mathcal{L}^{(k)}\}_k, \mathcal{S}}{\text{argmin}} \; g(\{\mathcal{L}^{(k)}\}_k, \mathcal{S}), \tag{59}$$

where

$$\begin{aligned} &g(\{\mathcal{L}^{(k)}\}_k, \mathcal{S}) \\ &= \frac{1}{2}\|\mathcal{Y} - \textstyle\sum_k \mathcal{L}^{(k)} - \mathcal{S}\|_{\text{F}} + \lambda_\iota \textstyle\sum_k v_k \|\mathcal{L}^{(k)}_{[k]}\|_\star + \mu_\iota \|\mathcal{S}\|_{l_1}. \end{aligned} \tag{60}$$

Let $\Delta^L_{\iota,k} = \mathcal{L}^{(k)*} - \hat{\mathcal{L}}^{(k)}\}_k$ and $\Delta^S_\iota = \mathcal{S}^* - \hat{\mathcal{S}}_\iota$. Using the optimality of $(\{\hat{\mathcal{L}}^{(k)}\}_k, \hat{\mathcal{S}}_\iota)$, we have:

$$g(\{\hat{\mathcal{L}}^{(k)}\}_k, \hat{\mathcal{S}}_\iota) \le g(\{\mathcal{L}^{(k)*}\}_k, \mathcal{S}^*). \tag{61}$$

Note that $\sum_k \mathcal{L}^{(k)*} = \mathcal{L}^*$. Through the observation model of RTD, we have $\mathcal{Y} - \mathcal{L}^* - \mathcal{S}^* = \mathcal{E}$. After some algebra, we obtain

$$\begin{aligned} &\frac{1}{2}\Big(\|\sum_k \Delta^L_{\iota,k}\|_{\text{F}}^2 + \|\Delta^S_\iota\|_{\text{F}}^2\Big) \\ &\le \underbrace{\lambda_\iota \sum_k v_k \Big(\|\mathcal{L}^{(k)*}_{[k]}\|_\star - \|\mathcal{L}^{(k)*}_{[k]} - \Delta\mathcal{L}^{(k)}_{[k]}\|_\star\Big)}_{\text{I}} + \underbrace{\mu_\iota\Big(\|\mathcal{S}^*\|_{l_1} - \|\mathcal{S}^* - \Delta^S_\iota\|_{l_1}\Big)}_{\text{II}} \\ &+ \underbrace{\Big\langle \sum_k \Delta^L_{\iota,k}, \Delta^S_\iota \Big\rangle}_{\text{III}} + \underbrace{\Big\langle \sum_k \Delta^L_{\iota,k}, \mathcal{E} \Big\rangle}_{\text{IV}} + \underbrace{\langle \Delta^S_\iota, \mathcal{E}\rangle}_{\text{V}}. \end{aligned} \tag{62}$$

and

$$\begin{aligned} &\frac{1}{2}\Big(\sum_k \|\Delta^L_{\iota,k}\|_{\text{F}}^2 + \|\Delta^S_\iota\|_{\text{F}}^2\Big) \\ &\le \text{I} + \text{II} + \text{III} + \text{IV} + \text{V} + \underbrace{\sum_k \sum_{l\neq k} \Big\langle \Delta\mathcal{L}^{(k)}, \Delta\mathcal{L}^{(l)} \Big\rangle}_{\text{VI}}. \end{aligned} \tag{63}$$

We will bound **Items I** to **VI** as follows.

Bound Item I. Let $\mathcal{P}^k(\cdot) = \mathcal{P}_{\mathfrak{F}_k(\mathcal{L}^{(k)*})}(\cdot)$ (see the definition of $\mathcal{P}_T$ in Eq. (28)). For $\Delta\mathcal{L}^{(k)} \in \mathbb{R}^{d_1\times\cdots\times d_K}$, define

$$\Delta'\mathcal{L}^{(k)}_{[k]} = \mathcal{P}^k(\mathcal{L}^{(k)}_{[k]}), \text{ and } \Delta''\mathcal{L}^{(k)}_{[k]} = \Delta\mathcal{L}^{(k)}_{[k]} - \Delta'\mathcal{L}^{(k)}_{[k]}.$$

Using Lemma 4 directly yields

$$\|\mathcal{L}^{(k)*}_{[k]} - \Delta''\mathcal{L}^{(k)}_{[k]}\|_\star = \|\mathcal{L}^{(k)*}_{[k]}\|_\star + \|\Delta''\mathcal{L}^{(k)}_{[k]}\|_\star,$$

leading to

$$\begin{aligned} \|\mathcal{L}^{(k)*}_{[k]} - \Delta\mathcal{L}^{(k)}_{[k]}\|_\star &= \|(\mathcal{L}^{(k)*}_{[k]} - \Delta''\mathcal{L}^{(k)}_{[k]}) - \Delta'\mathcal{L}^{(k)}_{[k]}\|_\star \\ &\ge \|(\mathcal{L}^{(k)*}_{[k]} - \Delta''\mathcal{L}^{(k)}_{[k]})\|_\star - \|\Delta'\mathcal{L}^{(k)}_{[k]}\|_\star \\ &= \|\mathcal{L}^{(k)*}_{[k]}\|_\star + \|\Delta''\mathcal{L}^{(k)}_{[k]}\|_\star - \|\Delta'\mathcal{L}^{(k)}_{[k]}\|_\star. \end{aligned}$$

Thus, we have

$$\begin{aligned} \text{I} &= \lambda_\iota \sum_k v_k \Big(\|\mathcal{L}^{(k)*}_{[k]}\|_\star - \|\mathcal{L}^{(k)*}_{[k]} - \Delta\mathcal{L}^{(k)}_{[k]}\|_\star\Big) \\ &\le \lambda_\iota \sum_k v_k \|\Delta'\mathcal{L}^{(k)}_{[k]}\|_\star - \lambda_\iota \sum_k v_k \|\Delta''\mathcal{L}^{(k)}_{[k]}\|_\star. \end{aligned} \tag{64}$$

Bound Item II. Similar to the proof of Theorem 1, we have

$$\text{II} \le \mu_\iota \|(\Delta^S_\iota)_S\|_{l_1} - \mu_\iota \|(\Delta^S_\iota)_{S^\perp}\|_{l_1}. \tag{65}$$

Bound Items III and V. Using the definition of dual norm, we have

$$\text{III+V} \le (\|\mathcal{E}\|_{l_\infty} + 2\alpha)\|\Delta^S_\iota\|_{l_1}. \tag{66}$$

Bound Item IV.

$$\begin{aligned} \Big\langle \sum_k \Delta\mathcal{L}^{(k)}, \mathcal{E} \Big\rangle &= \sum_k \Big\langle \Delta\mathcal{L}^{(k)}, \mathcal{E} \Big\rangle \\ &\le \sum_k \|\Delta\mathcal{L}^{(k)}_{[k]}\|_\star \|\mathcal{E}_{[k]}\| \\ &= \sum_k (v_k \|\Delta\mathcal{L}^{(k)}_{[k]}\|_\star)(\|\mathcal{E}_{[k]}\|/v_k) \\ &\le (\sum_k v_k \|\Delta\mathcal{L}^{(k)}_{[k]}\|_\star) \max_k (\|\mathcal{E}_{[k]}\|/v_k) \\ &= \|\mathcal{E}\|^*_{\star\iota} \sum_k v_k \|\Delta\mathcal{L}^{(k)}_{[k]}\|_\star. \end{aligned} \tag{67}$$

Bound Item VI.

$$\begin{aligned} &\sum_k \sum_{l\neq k} \Big\langle \Delta\mathcal{L}^{(k)}, \Delta\mathcal{L}^{(l)} \Big\rangle \\ &\le \sum_k \sum_{l\neq k} \|\Delta\mathcal{L}^{(k)}_{[k]}\|_\star \|\Delta\mathcal{L}^{(l)}_{[k]}\| \\ &\le \sum_k (K-1)\beta \tilde{d}_k \|\Delta\mathcal{L}^{(k)}_{[k]}\|_\star \\ &= (K-1)\beta \sum_k (\tilde{d}_k/v_k)(v_k \|\Delta\mathcal{L}^{(k)}_{[k]}\|_\star) \\ &= (K-1)\beta \max_k (\tilde{d}_k/v_k) \sum_k v_k \|\Delta\mathcal{L}^{(k)}_{[k]}\|_\star. \end{aligned} \tag{68}$$

Combining Eq. (63) and the above bounds yields

$$\begin{aligned} &\frac{1}{2}\Big(\sum_k \|\Delta\mathcal{L}^{(k)}\|_{\text{F}}^2 + \|\Delta^S_\iota\|_{\text{F}}^2\Big) \\ &\le \lambda_\iota \sum_k v_k \|\Delta'\mathcal{L}^{(k)}_{[k]}\|_\star - \lambda_\iota \sum_k v_k \|\Delta''\mathcal{L}^{(k)}_{[k]}\|_\star \\ &\quad + \mu_\iota \|(\Delta^S_\iota)_S\|_{l_1} - \mu_\iota \|(\Delta^S_\iota)_{S^\perp}\|_{l_1} + (\|\mathcal{E}\|_{l_\infty} + 2\alpha)\|\Delta^S_\iota\|_{l_1} \\ &\quad + (\|\mathcal{E}\|^*_{\star\iota} + (K-1)\beta \max_k (\tilde{d}_k/v_k)) \sum_k v_k \|\Delta\mathcal{L}^{(k)}_{[k]}\|_\star \\ &\le \big(\lambda_\iota + (\|\mathcal{E}\|^*_{\star\iota} + (K-1)\beta \max_k (\tilde{d}_k/v_k))\big) \sum_k v_k \|\Delta'\mathcal{L}^{(k)}_{[k]}\|_\star \\ &\quad - \big(\lambda_\iota - (\|\mathcal{E}\|^*_{\star\iota} + (K-1)\beta \max_k (\tilde{d}_k/v_k))\big) \sum_k v_k \|\Delta''\mathcal{L}^{(k)}_{[k]}\|_\star \\ &\qquad + \big(\mu_\iota + (\|\mathcal{E}\|_{l_\infty} + 2\alpha)\big)\|(\Delta^S_\iota)_S\|_{l_1} \\ &\qquad - \big(\mu_\iota - (\|\mathcal{E}\|_{l_\infty} + 2\alpha)\big)\|(\Delta^S_\iota)_{S^\perp}\|_{l_1}. \end{aligned}$$

Choosing

$$\lambda_\iota > 2(\|\mathcal{E}\|_{\star\iota}^{*} + (K-1)\beta \max_k(\tilde{d}_k/v_k)) \tag{69}$$

and

$$\mu_\iota \geq \|\mathcal{E}\|_{l_\infty} + 2\alpha, \tag{70}$$

we have

$$\begin{aligned}&\lambda_\iota \sum_k v_k \|\Delta'' \mathcal{L}_{[k]}^{(k)}\|_\star + \mu_\iota \|(\Delta_\iota^S)_{S^\perp}\|_{l_1} \\ &\leq 3\big(\lambda_\iota \sum_k v_k \|\Delta' \mathcal{L}_{[k]}^{(k)}\|_\star + \mu_\iota \|(\Delta_\iota^S)\|_{l_1}\big).\end{aligned} \tag{71}$$

Note that according to Eq. (63) and the triangular inequality, we have

$$\begin{aligned}&\frac{1}{2}\big(\sum_k \|\Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2 + \|\Delta_\iota^S\|_{\mathrm{F}}^2\big) \\ &\leq \big(\lambda_\iota + (\|\mathcal{E}\|_{\star\iota}^{*} + (K-1)\beta \max_k(\tilde{d}_k/v_k))\big) \\ &\quad \big(\sum_k v_k \|\Delta' \mathcal{L}_{[k]}^{(k)}\|_\star + \sum_k v_k \|\Delta'' \mathcal{L}_{[k]}^{(k)}\|_\star\big) \\ &\quad + \big(\mu_\iota + (\|\mathcal{E}\|_{l_\infty} + 2\alpha)\big)\big(\|(\Delta_\iota^S)_S\|_{l_1} + \|(\Delta_\iota^S)_{S^\perp}\|_{l_1}\big).\end{aligned} \tag{72}$$

That leads to

$$\sum_k \|\Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2 + \|\Delta_\iota^S\|_{\mathrm{F}}^2 \leq 16\lambda_\iota \sum_k v_k \|\Delta' \mathcal{L}_{[k]}^{(k)}\|_\star + 16\mu_\iota \|(\Delta_\iota^S)_S\|_{l_1}. \tag{73}$$

By the definition of $\Delta' \mathcal{L}_{[k]}^{(k)}$, we have

$$\mathrm{rank}_{\mathrm{tb}}(\Delta' \mathcal{L}_{[k]}^{(k)}) \leq 2\mathrm{rank}_{\mathrm{tb}}(\mathcal{L}_{[k]}^{(k)*}) = 2\overline{r}_k^\iota, \tag{74}$$

and

$$\|\Delta' \mathcal{L}_{[k]}^{(k)}\|_{\mathrm{F}} \leq \|\Delta \mathcal{L}_{[k]}^{(k)}\|_{\mathrm{F}} = \|\Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}. \tag{75}$$

We also have $\|(\Delta_\iota^S)_S\|_{l_0} \leq |S| = s$. Then, we reach the inequality:

$$\begin{aligned}&\sum_k \|\Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2 + \|\Delta_\iota^S\|_{\mathrm{F}}^2 \\ &\leq 16\lambda_\iota \sum_k v_k \sqrt{2\overline{r}_k^\iota} \|\Delta \mathcal{L}^{(k)}\|_{\mathrm{F}} + 16\mu_\iota \sqrt{s} \|\Delta_\iota^S\|_{\mathrm{F}} \\ &\leq 16\lambda_\iota \sqrt{\sum_k (v_k \sqrt{2\overline{r}_k^\iota})^2} \sqrt{\sum_k \|\Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2} + 16\mu_\iota \sqrt{s} \|\Delta_\iota^S\|_{\mathrm{F}}\end{aligned} \tag{76}$$

The usage of $ab \leq a^2/4 + b^2$ leading to the first part of Theorem 2, i.e.,

$$\sum_k \|\Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2 + \|\Delta_\iota^S\|_{\mathrm{F}}^2 \leq c_3 \lambda_\iota^2 \sum_k v_k^2 \overline{r}_k^\iota + c_4 \mu_\iota^2 s. \tag{77}$$

To prove the second part of Theorem 2. First, we discuss in two cases:

Case 1: If $\|\sum_k \Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2 \leq \sum_k \|\Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2$, according to Eq. (77) we have

$$\|\sum_k \Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2 + \|\Delta_\iota^S\|_{\mathrm{F}}^2 \leq c_3 \lambda_\iota^2 \sum_k v_k^2 \overline{r}_k^\iota + c_4 \mu_\iota^2 s. \tag{78}$$

Case 2: If $\|\sum_k \Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2 > \sum_k \|\Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2$, according to Eq. (62), we have

$$\begin{aligned}&\frac{1}{2}\big(\|\sum_k \Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2 + \|\Delta_\iota^S\|_{\mathrm{F}}^2\big) \\ &\leq \big(\lambda_\iota + \|\mathcal{E}\|_{\star\iota}^{*}\big)\big(\sum_k v_k \|\Delta' \mathcal{L}_{[k]}^{(k)}\|_\star + \sum_k v_k \|\Delta'' \mathcal{L}_{[k]}^{(k)}\|_\star\big) \\ &\quad + \big(\mu_\iota + (\|\mathcal{E}\|_{l_\infty} + 2\alpha)\big)\big(\|(\Delta_\iota^S)_S\|_{l_1} + \|(\Delta_\iota^S)_{S^\perp}\|_{l_1}\big),\end{aligned} \tag{79}$$

which leads to

$$\begin{aligned}&\|\sum_k \Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2 + \|\Delta_\iota^S\|_{\mathrm{F}}^2 \\ &\leq 16\lambda_\iota \sum_k v_k \sqrt{2\overline{r}_k^\iota} \|\Delta \mathcal{L}^{(k)}\|_{\mathrm{F}} + 16\mu_\iota \sqrt{s} \|\Delta_\iota^S\|_{\mathrm{F}} \\ &\leq 16\lambda_\iota \sqrt{\sum_k (v_k \sqrt{2\overline{r}_k^\iota})^2} \sqrt{\sum_k \|\Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2} + 16\mu_\iota \sqrt{s} \|\Delta_\iota^S\|_{\mathrm{F}} \\ &\leq 16\lambda_\iota \sqrt{\sum_k (v_k \sqrt{2\overline{r}_k^\iota})^2} \sqrt{\|\sum_k \Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2} + 16\mu_\iota \sqrt{s} \|\Delta_\iota^S\|_{\mathrm{F}} \\ &\leq c_3 \lambda_\iota^2 \sum_k v_k^2 \overline{r}_k^\iota + c_4 \mu_\iota^2 s.\end{aligned} \tag{80}$$

Select $k^* \in \mathrm{argmax}_k\, v_k^2 \mathrm{rank}_{\mathrm{tb}}(\mathcal{L}_{[k]}^*)$. Letting $\mathcal{L}^{(k^*)} = \mathcal{L}^*$ and $\mathcal{L}^{(l)*} = \mathcal{O}\ , \forall l \neq k^*$, then $(\{\mathcal{L}^{(k)*}\}_k, \mathcal{S}^*)$ is feasible. In this case, $\overline{r}_{k^*}^\iota = \underline{r}_{k^*}^{\mathrm{o}}$ and $\overline{r}_l^\iota = 0, \forall l \neq k^*$. Then, we obtain

$$\begin{aligned}&\|\sum_k \Delta \mathcal{L}^{(k)}\|_{\mathrm{F}}^2 + \|\Delta_\iota^S\|_{\mathrm{F}}^2 \\ &\leq c_3 \lambda_\iota^2 \sum_k v_k^2 \overline{r}_k^\iota + c_4 \mu_\iota^2 s \\ &\leq c_3 \lambda_\iota \min_k v_k^2 \underline{r}_k^{\mathrm{o}} + c_4 \mu_\iota^2 s.\end{aligned} \tag{81}$$

Then, the proof is completed. □

Proof of Theorem 4 Since the proof of Theorem 4 differs from Theorem 3 only in bounding the maximum of the tensor spectral norms instead of their sum, we simply omit it.

Supp-§-B. Optimization Algorithms

Due to space limitation, the description of Algorithm 1 and Algorithm 2 is omitted. In this section, we present the proposed Algorithms 1 and 2 for Model I and Model II, respectively. In Algorithms 1 and 2, each sub-problem has a closed-form solution.

For notational simplicity, recall the definition 3d-unfolding operator for $\mathcal{T} \in \mathbb{R}^{d_1\times\cdots\times d_K}$ as $\mathfrak{F}_k(\mathcal{T}) := \mathcal{T}_{[k]}$ and its $\mathfrak{F}_k^{-1}(\cdot)$ such that $\mathfrak{F}_k^{-1}(\mathcal{T}_{[k]}) = \mathcal{T}$.

Algorithm 1 ADMM for Model I

Input: Observation $\mathcal{Y}$, parameters $\lambda_o, \mu_o, \{w_k\}_k, \rho > 0, \varepsilon > 0$.

1: Initialize $\mathcal{L}^0 = \mathcal{K}^0 = \mathcal{W}^0 = \mathcal{S}^0 = \mathcal{T}^0 = \mathcal{Z}^0 = \mathbf{0}, \mathcal{K}_k^0 = \mathcal{Y}_k^0 = \mathbf{0}, \forall k$.

2: **while** not converged **do**

3: Update $(\mathcal{L}^{t+1}, \mathcal{S}^{t+1})$ simultaneously by:

$$\min_{\mathcal{L},\mathcal{S}} \; l(\mathcal{L},\mathcal{S}) + \sum\nolimits_k \frac{\rho}{2}\|\mathcal{L} - \mathfrak{F}_k^{-1}(\mathcal{K}_k^t + \frac{\mathcal{Y}^t}{\rho})\|_{\mathrm{F}}^2 + \frac{\rho}{2}\|\mathcal{S} - (\mathcal{T}^t + \frac{\mathcal{Z}^t}{\rho})\|_{\mathrm{F}}^2 + \frac{\rho}{2}\|\mathcal{L} - (\mathcal{K}^t + \frac{\mathcal{W}^t}{\rho})\|_{\mathrm{F}}^2$$

4: Update $\{\mathcal{K}_k^{t+1}\}_k$, $\mathcal{T}^{t+1}$ and $\mathcal{K}^{t+1}$ simultaneously by:

$$\min_{\mathcal{K}_k} \; \lambda_o w_k \|\mathcal{K}_k\|_\star + \frac{\rho}{2}\|\mathcal{K}_k - \mathfrak{F}_k\big(\mathcal{L}^{t+1}\big) + \frac{\mathcal{Y}_k^t}{\rho}\|_{\mathrm{F}}^2$$

$$\min_{\mathcal{T}} \; \mu_o\|\mathcal{T}\|_{l_1} + \frac{\rho}{2}\|\mathcal{T} - (\mathcal{S}^{t+1} - \frac{\mathcal{Z}^t}{\rho})\|_{\mathrm{F}}^2$$

$$\min_{\mathcal{K}} \; \delta_\alpha^{l_\infty}(\mathcal{K}) + \frac{\rho}{2}\|\mathcal{K} - (\mathcal{L}^{t+1} - \frac{\mathcal{W}^t}{\rho})\|_{\mathrm{F}}^2$$

5: Dual update: $\mathcal{Z}^{k+1} = \mathcal{Z}^t + \rho(\mathcal{T}^{t+1} - \mathcal{S}^{t+1})$, $\mathcal{W}^{k+1} = \mathcal{W}^t + \rho(\mathcal{K}^{t+1} - \mathcal{L}^{t+1})$ and $\mathcal{Y}_k^{t+1} = \mathcal{Y}_k^t + \rho(\mathcal{K}_k^{t+1} - \mathfrak{F}_k(\mathcal{L}^{t+1}))$, $\forall k \in [K]$;

6: Check the convergence conditions: $\|\mathcal{X}^{t+1} - \mathcal{X}^t\|_{l_\infty} \le \varepsilon, \; \forall \mathcal{X} \in \{\mathcal{L},\mathcal{S},\mathcal{T},\mathcal{K},\{\mathcal{K}_k\}\}$; $\|\mathcal{T}^{t+1} - \mathcal{S}^{t+1}\|_{l_\infty} \le \varepsilon$ $\|\mathcal{K}^{t+1} - \mathcal{L}^{t+1}\|_{l_\infty} \le \varepsilon$; $\|\mathcal{K}_k^{t+1} - \mathfrak{F}_k\big(\mathcal{L}^{t+1}\big)\|_{l_\infty} \le \varepsilon, \forall k \in [K]$;

7: $t = t + 1$.

8: **end while**

Several operators

Before giving solutions to the sub-problems in Algorithm 1 and Algorithm 2, we briefly give the proximal operators of TNN $\|\cdot\|_\star$ as follows:

Lemma 10. (Wang and Jin 2017). *Let tensor $\mathcal{T} \in \mathbb{R}^{d_1\times d_2\times d_3}$ with t-SVD $\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$, where $\mathcal{U} \in \mathbb{R}^{d_1\times r\times d_3}$ and $\mathcal{V} \in \mathbb{R}^{d_2\times r\times d_3}$ are orthogonal tensors and $\mathcal{S} \in \mathbb{R}^{r\times r\times d_3}$ is the f-diagonal tensor of singular tubes. Then the proximal operator of function $\tau\|\cdot\|_\star$ at point $\mathcal{T}_0$, denoted by $Prox_\tau^{\|\cdot\|_\star}(\mathcal{T}_0)$, can be computed as follows*

$$\begin{aligned} Prox_\tau^{\|\cdot\|_\star}(\mathcal{T}_0) &= \operatorname*{argmin}_{\mathcal{T}} \frac{1}{2}\|\mathcal{T}_0 - \mathcal{T}\|_F^2 + \tau\|\mathcal{T}\|_\star \\ &= \mathcal{U} * \mathrm{ifft3}(\max(\mathrm{fft3}(\mathcal{S}) - \tau, 0)) * \mathcal{V}^\top. \end{aligned} \tag{82}$$

Algorithm 2 ADMM for Model II

Input: Observation $\mathcal{Y}$, parameters $\lambda_\iota, \mu_\iota, \{v_k\}_k, \rho > 0, \varepsilon > 0$.

1: Initialize $\mathcal{S}^0 = \mathcal{T}^0 = \mathcal{Z}^0 = \mathcal{K}^0 = \mathcal{W}^0 = \mathbf{0}, (\mathcal{L}^{(k)})^0 = \mathcal{K}_k^0 = \mathcal{Y}_k^0 = \mathbf{0}, \forall k$.

2: **while** not converged **do**

3: Update $\{(\mathcal{L}^{(k)})^{t+1}\}_k$ and $\mathcal{S}^{t+1}$ simultaneously by:

$$\min_{\{\mathcal{L}^{(k)}\}_k,\mathcal{S}} \; l(\sum\nolimits_k \mathcal{L}^{(k)}, \mathcal{S}) + \sum\nolimits_k \frac{\rho}{2}\|\mathcal{L}^{(k)} - \mathfrak{F}_k^{-1}(\mathcal{K}_k^t + \frac{\mathcal{Y}^t}{\rho})\|_{\mathrm{F}}^2 + \frac{\rho}{2}\|\mathcal{S} - (\mathcal{T}^t + \frac{\mathcal{Z}^t}{\rho})\|_{\mathrm{F}}^2 + \frac{\rho}{2}\|\sum\nolimits_k \mathcal{L}^{(k)} - (\mathcal{K}^t + \frac{\mathcal{W}^t}{\rho})\|_{\mathrm{F}}^2;$$

4: Update $\{\mathcal{K}_k^{t+1}\}_k$, $\mathcal{T}^{t+1}$ and $\mathcal{K}^{t+1}$ simultaneously by:

$$\min_{\mathcal{K}_k} \; \lambda_\iota v_k \|\mathcal{K}_k\|_\star + \frac{\rho}{2}\|\mathcal{K}_k - \mathfrak{F}_k\big((\mathcal{L}^{(k)})^{t+1}\big) + \frac{\mathcal{Y}_k^t}{\rho}\|_{\mathrm{F}}^2$$

$$\min_{\mathcal{T}} \; \mu_\iota\|\mathcal{T}\|_{l_1} + \frac{\rho}{2}\|\mathcal{T} - (\mathcal{S}^{t+1} - \frac{\mathcal{Z}^t}{\rho})\|_{\mathrm{F}}^2$$

$$\min_{\mathcal{K}} \; \delta_\alpha^{l_\infty}(\mathcal{K}) + \frac{\rho}{2}\|\mathcal{K} - \sum_k (\mathcal{L}^{(k)})^{t+1} + \frac{\mathcal{W}^t}{\rho}\|_{\mathrm{F}}^2$$

5: Dual update: $\mathcal{Z}^{t+1} = \mathcal{Z}^t + \rho(\mathcal{T}^{t+1} - \mathcal{S}^{t+1})$, $\mathcal{W}^{t+1} = \mathcal{W}^t + \rho(\mathcal{K}^{t+1} - \sum_k (\mathcal{L}^{(k)})^{t+1})$ and $\mathcal{Y}_k^{t+1} = \mathcal{Y}_k^t + \rho\big(\mathcal{K}_k^{t+1} - \mathfrak{F}_k\big((\mathcal{L}^{(k)})^{t+1}\big)\big)$, $\forall k \in [K]$;

6: Check the convergence conditions: $\|\mathcal{X}^{t+1} - \mathcal{X}^t\|_{l_\infty} \le \varepsilon, \; \forall \mathcal{X} \in \big\{\{\mathcal{L}^{(k)}\}_k, \mathcal{S}, \mathcal{T}, \mathcal{K}, \{\mathcal{K}_k\}\big\}$; $\|\mathcal{T}^{t+1} - \mathcal{S}^{t+1}\|_{l_\infty} \le \varepsilon$; $\|\mathcal{K}^{t+1} - \sum_k (\mathcal{L}^{(k)})^{t+1}\|_{l_\infty} \le \varepsilon$; $\|\mathcal{K}_k^{t+1} - \mathfrak{F}_k\big((\mathcal{L}^{(k)})^{t+1}\big)\|_{l_\infty} \le \varepsilon, \forall k \in [K]$;

7: $t = t + 1$.

8: **end while**

The proximal operator of l_1-norm $\|\cdot\|_{l_1}$ is given as

$$\begin{aligned} \mathrm{Prox}_\tau^{\|\cdot\|_{l_1}}(\mathcal{T}_0) &= \operatorname*{argmin}_{\mathcal{T}} \frac{1}{2}\|\mathcal{T}_0 - \mathcal{T}\|_{\mathrm{F}}^2 + \tau\|\mathcal{T}\|_{l_1} \\ &= \mathrm{sgn}(\mathcal{T}_0) \circledast \max\big(|\mathcal{T}_0| - \tau, 0\big), \end{aligned} \tag{83}$$

and the proximal operator of indicator function of l_∞-norm ball $\delta_\alpha^{l_\infty}(\cdot)$ is a projector:

$$\begin{aligned} \mathrm{Proj}_\alpha^{\|\cdot\|_{l_\infty}}(\mathcal{T}_0) &= \operatorname*{argmin}_{\mathcal{T}} \frac{1}{2}\|\mathcal{T}_0 - \mathcal{T}\|_{\mathrm{F}}^2 + \delta_\alpha^{l_\infty}(\mathcal{T}_0) \\ &= \mathrm{sgn}(\mathcal{T}_0) \circledast \min\big(|\mathcal{T}_0|, \alpha\big). \end{aligned} \tag{84}$$

Solutions to Sub-problems in Algorithm 1

In this subsection, we derive solutions to sub-problems in Algorithm 1.

First, adding auxiliary variables to Problem (12), we get

$$\min_{\substack{\mathcal{L},\mathcal{S},\mathcal{T},\\ \mathcal{K},\{\mathcal{K}_k\}_k}} \frac{1}{2}\|\mathcal{Y} - \mathcal{L} - \mathcal{S}\|_{\mathrm{F}}^2 + \lambda_o \sum\nolimits_k w_k\|\mathcal{K}_k\|_\star + \mu_o\|\mathcal{T}\|_{l_1} + \delta_\alpha^{l_\infty}(\mathcal{K})$$

$$\text{s.t. } \mathcal{K}_k = \mathfrak{F}_k\big(\mathcal{L}\big), \forall k; \mathcal{T} = \mathcal{S}; \mathcal{K} = \mathcal{L}.$$

Then, the augmented Lagrangian is given as follows

$$
\begin{aligned}
&L^{\mathrm{I}}_{\rho}(\mathcal{L},\mathcal{S},\mathcal{T},\mathcal{K},\{\mathcal{K}_k\}_k,\{\mathcal{Y}_k\}_k,\mathcal{Z},\mathcal{W})\\
=&\frac{1}{2}\|\mathcal{Y}-\mathcal{L}-\mathcal{S}\|_{\mathrm{F}}^2+\lambda_{\mathrm{o}}\sum\nolimits_k w_k\|\mathcal{K}_k\|_\star+\mu_{\mathrm{o}}\|\mathcal{T}\|_{l_1}+\delta_\alpha^{l_\infty}(\mathcal{K})\\
&+\sum_k\left(\langle\mathcal{Y}_k,\mathcal{K}_k-\mathfrak{F}_k(\mathcal{L})\rangle+\frac{\rho}{2}\|\mathcal{K}_k-\mathfrak{F}_k(\mathcal{L})\|_{\mathrm{F}}\right)\\
&+\langle\mathcal{Z},\mathcal{T}-\mathcal{S}\rangle+\frac{\rho}{2}\|\mathcal{T}-\mathcal{S}\|_{\mathrm{F}}^2+\langle\mathcal{W},\mathcal{K}-\mathcal{L}\rangle+\frac{\rho}{2}\|\mathcal{K}-\mathcal{L}\|_{\mathrm{F}}^2.
\end{aligned}
$$

Further, we update blocks $(\mathcal{L},\mathcal{S})$ and $(\{\mathcal{K}_k\},\mathcal{T},\mathcal{K})$ alternatively by fixing the other variables.

Update $(\mathcal{L},\mathcal{S})$**.** Fixing $(\{\mathcal{K}_k\},\mathcal{T},\mathcal{K})$, we update $(\mathcal{L},\mathcal{S})$ by minimizing the augmented Lagrangian $L^{\mathrm{I}}_r ho$ with respect to $(\mathcal{L},\mathcal{S})$, which can be simplified as follows

$$
\begin{aligned}
\min_{\mathcal{L},\mathcal{S}}\; & l(\mathcal{L},\mathcal{S})+\sum\nolimits_k\frac{\rho}{2}\|\mathcal{L}-\mathfrak{F}_k^{-1}(\mathcal{K}_k^t+\frac{\mathcal{Y}^t}{\rho})\|_{\mathrm{F}}^2\\
&+\frac{\rho}{2}\|\mathcal{S}-(\mathcal{T}^t+\frac{\mathcal{Z}^t}{\rho})\|_{\mathrm{F}}^2+\frac{\rho}{2}\|\mathcal{L}-(\mathcal{K}^t+\frac{\mathcal{W}^t}{\rho})\|_{\mathrm{F}}^2
\end{aligned}
\tag{85}
$$

Taking the derivatives with respect to $\mathcal{L}$ and $\mathcal{S}$ and setting the derivatives to zero, we obtain

$$(K\rho+\rho+1)\mathcal{L}+\mathcal{S}=\rho\tilde{\mathcal{K}}+\rho\sum_k\tilde{\mathcal{K}}_k+\mathcal{Y};\tag{86}$$

and

$$\mathcal{L}+(1+\rho)\mathcal{S}=\mathcal{Y}+\mu\tilde{\mathcal{T}}.\tag{87}$$

where

$$\tilde{\mathcal{K}}=\mathcal{K}^t+\frac{\mathcal{W}^t}{\rho},\tilde{\mathcal{K}}_k=\mathcal{K}_k^t+\frac{\mathcal{Y}^t}{\rho}\text{ and }\tilde{\mathcal{T}}=\mathcal{T}^t+\frac{\mathcal{Z}^t}{\rho}.$$

By solving matrix equation group, we get the closed-form solution of $\mathcal{L}^{t+1}$ and $\mathcal{S}^{t+1}$

$$
\begin{aligned}
\mathcal{L}^{t+1}=&\frac{(1+\rho)\tilde{\mathcal{K}}+(1+\rho)\sum_k\tilde{\mathcal{K}}_k+\mathcal{Y}-\mathcal{T}}{(K+1)(\rho+1)+1},\\
\mathcal{S}^{t+1}=&\frac{(K+1)\mathcal{Y}+(K\rho+\rho+1)\tilde{\mathcal{T}}-\tilde{\mathcal{K}}-\sum_k\tilde{\mathcal{T}}_k}{(K+1)(\rho+1)+1}.
\end{aligned}
$$

Update $(\{\mathcal{K}_k\},\mathcal{T},\mathcal{K})$**.** Fixing $(\mathcal{L},\mathcal{S})$, we update $\{\mathcal{K}_k\}_k$, $\mathcal{T}$, and $\mathcal{K}$ by minimizing the augmented Lagrangian $L^{\mathrm{I}}_r ho$ with respect to $(\{\mathcal{K}_k\},\mathcal{T},\mathcal{K})$. The problem can be solved separately as follows.

$$
\begin{aligned}
\mathcal{K}_k^{t+1}&=\operatorname*{argmin}_{\mathcal{K}_k}\;\lambda_{\mathrm{o}}w_k\|\mathcal{K}_k\|_\star+\frac{\rho}{2}\|\mathcal{K}_k-\mathfrak{F}_k(\mathcal{L}^{t+1})+\frac{\mathcal{Y}_k^t}{\rho}\|_{\mathrm{F}}^2\\
&=\mathrm{Prox}^{\|\cdot\|_\star}_{\lambda_{\mathrm{o}}w_k/\rho}(\mathfrak{F}_k(\mathcal{L}^{t+1})-\frac{\mathcal{Y}_k^t}{\rho})\\
\mathcal{T}^{t+1}&=\operatorname*{argmin}_{\mathcal{T}}\;\mu_{\mathrm{o}}\|\mathcal{T}\|_{l_1}+\frac{\rho}{2}\|\mathcal{T}-(\mathcal{S}^{t+1}-\frac{\mathcal{Z}^t}{\rho})\|_{\mathrm{F}}^2\\
&=\mathrm{Prox}^{\|\cdot\|_{l_1}}_{\mu_{\mathrm{o}}/\rho}(\mathcal{S}^{t+1}-\frac{\mathcal{Z}^t}{\rho}).\\
\mathcal{K}^{t+1}&\operatorname*{argmin}_{\mathcal{K}}\;\delta_\alpha^{l_\infty}(\mathcal{K})+\frac{\rho}{2}\|\mathcal{K}-(\mathcal{L}^{t+1}-\frac{\mathcal{W}^t}{\rho})\|_{\mathrm{F}}^2\\
&=\mathrm{Proj}^{\|\cdot\|_{l_\infty}}_{\alpha}(\mathcal{L}^{t+1}-\frac{\mathcal{W}^t}{\rho}))
\end{aligned}
$$

Solutions to Sub-problems in Algorithm 2

We solve the sub-problems in Algorithm 2 as follows. First, adding auxiliary variables Problem (13) yields

$$
\begin{aligned}
\min_{\substack{\{\mathcal{L}^{(k)}\}_k,\mathcal{S},\\\{\mathcal{K}_k\}_k,\mathcal{T},\mathcal{K}}}\; & l(\sum\nolimits_k\mathcal{L}^{(k)},\mathcal{S})+\lambda_{\mathrm{o}}\sum\nolimits_k v_k\|\mathcal{K}_k\|_\star+\mu_\iota\|\mathcal{T}\|_{l_1}+\delta_\alpha^{l_\infty}(\mathcal{K})\\
\text{s.t. }& \mathcal{K}_k=\mathfrak{F}_k(\mathcal{L}^{(k)}),\forall k;\mathcal{T}=\mathcal{S};\mathcal{K}=\sum\nolimits_k\mathcal{L}^{(k)}.
\end{aligned}
$$

Then, the augmented Lagrangian is given as follows

$$
\begin{aligned}
&L^{\mathrm{II}}_{\rho}(\{\mathcal{L}^{(k)}\}_k,\mathcal{S},\mathcal{T},\mathcal{K},\{\mathcal{K}_k\}_k,\{\mathcal{Y}_k\}_k,\mathcal{Z},\mathcal{W})\\
&=\frac{1}{2}\|\mathcal{Y}-\sum_k\mathcal{L}^{(k)}-\mathcal{S}\|_{\mathrm{F}}^2+\lambda_{\mathrm{o}}\sum\nolimits_k w_k\|\mathcal{K}_k\|_\star+\mu_\iota\|\mathcal{T}\|_{l_1}\\
&+\sum_k\left(\left\langle\mathcal{Y}_k,\mathcal{K}_k-\mathfrak{F}_k(\mathcal{L}^{(k)})\right\rangle+\frac{\rho}{2}\|\mathcal{K}_k-\mathfrak{F}_k(\mathcal{L}^{(k)})\|_{\mathrm{F}}\right)\\
&+\delta_\alpha^{l_\infty}(\mathcal{K})+\langle\mathcal{Z},\mathcal{T}-\mathcal{S}\rangle+\frac{\rho}{2}\|\mathcal{T}-\mathcal{S}\|_{\mathrm{F}}^2\\
&+\left\langle\mathcal{W},\mathcal{K}-\mathcal{L}^{(k)}\right\rangle+\frac{\rho}{2}\|\mathcal{K}-\sum_k\mathcal{L}^{(k)}\|_{\mathrm{F}}^2.
\end{aligned}
$$

Further, we update blocks $(\{(\mathcal{L}^{(k)})\},\mathcal{S})$ and $(\{\mathcal{K}_k\},\mathcal{T},\mathcal{K})$ alternatively by fixing the other variables.

Update $(\{(\mathcal{L}^{(k)})\},\mathcal{S})$**.** Fixing $(\{\mathcal{K}_k\},\mathcal{T},\mathcal{K})$, we update $(\{(\mathcal{L}^{(k)})\},\mathcal{S})$ by minimizing the augmented Lagrangian $L^{\mathrm{II}}_r ho$ with respect to $(\mathcal{L},\mathcal{S})$, which can be simplified to the following problem

$$
\begin{aligned}
\min_{\{\mathcal{L}^{(k)}\}_k,\mathcal{S}}\; & l(\sum\nolimits_k\mathcal{L}^{(k)},\mathcal{S})+\sum\nolimits_k\frac{\rho}{2}\|\mathcal{L}^{(k)}-\mathfrak{F}_k^{-1}(\mathcal{K}_k^t+\frac{\mathcal{Y}^t}{\rho})\|_{\mathrm{F}}^2\\
&+\frac{\rho}{2}\|\mathcal{S}-(\mathcal{T}^t+\frac{\mathcal{Z}^t}{\rho})\|_{\mathrm{F}}^2+\frac{\rho}{2}\|\sum_k\mathcal{L}^{(k)}-(\mathcal{K}^t+\frac{\mathcal{W}^t}{\rho})\|_{\mathrm{F}}^2.
\end{aligned}
$$

Taking the derivatives with respect to $\mathcal{L}^{(k)}$ and $\mathcal{S}$ and setting the derivatives to zero, we obtain

$$\sum_k\mathcal{L}^{(k)}+\mathcal{S}-\mathcal{Y}+\rho\mathcal{L}^{(k)}-\rho\tilde{\mathcal{K}}_k+\rho\sum_k\mathcal{L}^{(k)}-\rho\tilde{\mathcal{K}}=0\tag{88}$$

and

$$\sum_k\mathcal{L}^{(k)}+\mathcal{S}-\mathcal{Y}+\rho\mathcal{S}-\mu\tilde{\mathcal{T}}=0.\tag{89}$$

where

$$\tilde{\mathcal{K}}=\mathcal{K}^t+\frac{\mathcal{W}^t}{\rho},\tilde{\mathcal{K}}_k=\mathcal{K}_k^t+\frac{\mathcal{Y}^t}{\rho}\text{ and }\tilde{\mathcal{T}}=\mathcal{T}^t+\frac{\mathcal{Z}^t}{\rho}.$$

By solving matrix equation group, we get the closed-form solution of $\mathcal{L}^{t+1}$:

$$(\mathcal{L}^{(k)})^{t+1}=\rho^{-1}\left(\rho\tilde{\mathcal{K}}+\sum_k\tilde{\mathcal{K}}+\mathcal{Y}-(1+\rho)\mathcal{M}-\mathcal{S}^{t+1}\right)$$

with

$$\mathcal{S}^{t+1}=\frac{(1+K)\mathcal{Y}+(K+\rho+K\rho)\tilde{\mathcal{T}}-K\tilde{\mathcal{K}}-\sum_k\tilde{\mathcal{K}}_k}{(1+K)(1+\rho)+K},$$

where

$$\mathcal{M} = \frac{K(1+\rho)\tilde{\mathcal{K}} + (1+\rho)\sum_k \tilde{\mathcal{K}}_k + K\mathcal{Y} - K\tilde{\mathcal{T}}}{(1+K)(1+\rho)+K}.$$

Update $(\{\mathcal{K}_k\}, \mathcal{T}, \mathcal{K})$. Fixing $(\{\mathcal{L}^{(k)}\}, \mathcal{S})$, we update $\{\mathcal{K}_k\}_k$, $\mathcal{T}$, and $\mathcal{K}$ by minimizing the augmented Lagrangian $L^{\text{II}}_r ho$ with respect to $(\{\mathcal{K}_k\}, \mathcal{T}, \mathcal{K})$. The problem can be solved separately as follows.

$$\begin{aligned}
\mathcal{K}_k^{t+1} =& \min_{\mathcal{K}_k} \ \lambda_\iota v_k \|\mathcal{K}_k\|_\star + \frac{\rho}{2}\|\mathcal{K}_k - \mathfrak{F}_k\big((\mathcal{L}^{(k)})^{t+1}\big) + \frac{\mathcal{Y}_k^t}{\rho}\|_{\text{F}}^2 \\
=& \text{Prox}_{\lambda_\iota v_k/\rho}^{\|\cdot\|_\star}\big(\mathfrak{F}_k\big((\mathcal{L}^{(k)})^{t+1}\big) - \frac{\mathcal{Y}_k^t}{\rho}\big), \\
\mathcal{T}^{t+1} =& \operatorname*{argmin}_{\mathcal{T}} \ \mu_\iota \|\mathcal{T}\|_{l_1} + \frac{\rho}{2}\|\mathcal{T} - (\mathcal{S}^{t+1} - \frac{\mathcal{Z}^t}{\rho})\|_{\text{F}}^2 \\
=& \text{Prox}_{\mu_\iota/\rho}^{\|\cdot\|_{l_1}}(\mathcal{S}^{t+1} - \frac{\mathcal{Z}^t}{\rho}), \\
\mathcal{K}^{t+1} =& \operatorname*{argmin}_{\mathcal{K}} \ \delta_\alpha^{l_\infty}(\mathcal{K}) + \frac{\rho}{2}\|\mathcal{K} - \sum_k (\mathcal{L}^{(k)})^{t+1} + \frac{\mathcal{W}^t}{\rho}\|_{\text{F}}^2 \\
=& \text{Proj}_\alpha^{\|\cdot\|_{l_\infty}}(\sum_k (\mathcal{L}^{(k)})^{t+1} - \frac{\mathcal{W}^t}{\rho}).
\end{aligned}$$

References

[Agarwal, Negahban, and Wainwright 2012] Agarwal, A.; Negahban, S.; and Wainwright, M. J. 2012. Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions. *The Annals of Statistics* 1171–1197.

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

[Candès and Tao 2010] Candès, E. J., and Tao, T. 2010. The power of convex relaxation: near-optimal matrix completion. *IEEE TIT* 56(5):2053–2080.

[Candès et al. 2011] Candès, E. J.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis? *Journal of the ACM* 58(3):11.

[Fazel 2002] Fazel, M. 2002. *Matrix rank minimization with applications*. Ph.D. Dissertation, PhD thesis, Stanford University.

[Gu, Gui, and Han 2014] Gu, Q.; Gui, H.; and Han, J. 2014. Robust tensor decomposition with gross corruption. In *NIPS 2014*, 1422–1430.

[Harshman 1970] Harshman, R. A. 1970. Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis.

[Hillar and Lim 2009] Hillar, C. J., and Lim, L. 2009. Most tensor problems are np-hard. *Journal of the ACM* 60(6):45.

[Hu et al. 2017] Hu, W.; Tao, D.; Zhang, W.; Xie, Y.; and Yang, Y. 2017. The twist tensor nuclear norm for video completion. *IEEE TNNLS* 28(12):2961–2973.

[Huang et al. 2015] Huang, B.; Mu, C.; Goldfarb, D.; and Wright, J. 2015. Provable models for robust low-rank tensor completion. *Pacific Journal of Optimization* 11(2):339–364.

[Kilmer et al. 2013] Kilmer, M. E.; Braman, K.; Hao, N.; and Hoover, R. C. 2013. 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.

[Klopp, Lounici, and Tsybakov 2016] Klopp, O.; Lounici, K.; and Tsybakov, A. B. 2016. Robust matrix completion. *Probability Theory and Related Fields* 1–42.

[Kolda and Bader 2009] Kolda, T. G., and Bader, B. W. 2009. Tensor decompositions and applications. *SIAM Review* 51(3):455–500.

[Liu et al. 2013] Liu, J.; Musialski, P.; Wonka, P.; and Ye, J. 2013. Tensor completion for estimating missing values in visual data. *IEEE TPAMI* 35(1):208–220.

[Liu et al. 2016a] Liu, X. Y.; Aeron, S.; Aggarwal, V.; and Wang, X. 2016a. Low-tubal-rank tensor completion using alternating minimization. *arXiv preprint arXiv:1610.01690*.

[Liu et al. 2016b] Liu, X.; Aeron, S.; Aggarwal, V.; Wang, X.; and Wu, M. 2016b. Adaptive sampling of rf fingerprints for fine-grained indoor localization. *IEEE Transactions on Mobile Computing* 15(10):2411–2423.

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

[Lu et al. 2018] Lu, C.; Feng, J.; Lin, Z.; and Yan, S. 2018. Exact low tubal rank tensor recovery from gaussian measurements. In *IJCAI 2018*.

[Lu et al. 2019] Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; and Yan, S. 2019. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE TPAMI* 1–1.

[Mu et al. 2014] Mu, C.; Huang, B.; Wright, J.; and Goldfarb, D. 2014. Square deal: Lower bounds and improved relaxations for tensor recovery. In *ICML 2014*, 73–81.

[Negahban et al. 2009] Negahban, S.; Yu, B.; Wainwright, M. J.; and Ravikumar, P. K. 2009. A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. In *NIPS*, 1348–1356.

[Oseledets 2011] Oseledets, I. V. 2011. Tensor-train decomposition. *SIAM Journal on Scientific Computing* 33(5):2295–2317.

[Recht, Fazel, and Parrilo 2007] Recht, B.; Fazel, M.; and Parrilo, P. A. 2007. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review* 52(3):471–501.

[Tomioka and Suzuki 2013] Tomioka, R., and Suzuki, T. 2013. Convex tensor decomposition via structured schatten norm regularization. In *NIPS 2013*, 1331–1339.

[Tomioka et al. 2011] Tomioka, R.; Suzuki, T.; Hayashi, K.; and Kashima, H. 2011. Statistical performance of convex tensor decomposition. In *NIPS 2011*, 972–980.

[Tucker 1966] Tucker, L. R. 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika* 31(3):279–311.

[Wang and Jin 2017] Wang, A., and Jin, Z. 2017. Near-optimal noisy low-tubal-rank tensor completion via singular tube thresholding. In *IEEE ICDMW*, 553–560.

[Wei et al. 2018] Wei, D.; Wang, A.; Feng, X.; Wang, B.; and Wang, B. 2018. Tensor completion based on triple tubal nuclear norm. *Algorithms* 11(7).

[Xie et al. 2017] Xie, Y.; Tao, D.; Zhang, W.; Liu, Y.; Zhang, L.; and Qu, Y. 2017. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *IJCV* (4):1–23.

[Yokota et al. 2018] Yokota, T.; Erem, B.; Guler, S.; Warfield, S. K.; and Hontani, H. 2018. Missing slice recovery for tensors using a low-rank model in embedded space. *arXiv preprint arXiv:1804.01736*.

[Yuan and Zhang 2016] Yuan, M., and Zhang, C. H. 2016. On tensor completion via nuclear norm minimization. *Foundations of Computational Mathematics* 16(4):1–38.

[Zhang and Aeron 2017] Zhang, Z., and Aeron, S. 2017. Exact tensor completion using t-svd. *IEEE TSP* 65(6):1511–1526.

[Zhang et al. 2014] Zhang, Z.; Ely, G.; Aeron, S.; Hao, N.; and Kilmer, M. 2014. Novel methods for multilinear data completion and de-noising based on tensor-svd. In *CVPR 2014*, 3842–3849.

[Zhao et al. 2015a] Zhao, Q.; Meng, D.; Kong, X.; Xie, Q.; Cao, W.; Wang, Y.; and Xu, Z. 2015a. A novel sparsity measure for tensor recovery. In *ICCV 2015*, 271–279.

[Zhao et al. 2015b] Zhao, Q.; Zhou, G.; Zhang, L.; Cichocki, A.; and Amari, S.-I. 2015b. Bayesian robust tensor factorization for incomplete multiway data. *IEEE TNNLS* 27(4):736–748.

[Zhou and Feng 2017] Zhou, P., and Feng, J. 2017. Outlier-robust tensor pca. In *CVPR 2017*, 3938–3946.