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 $\mathbf{M}$ whose $(i,j)$th entry $\mathbf{M}(i,j)$ is $\mathcal{T}(i,j,:)$. 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 $\mathcal{T} = \text{fft}(\mathcal{T}, [3])$ 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 $(\mathcal{T}^{(l)})$ denote its lth frontal slice $\mathcal{T}(\cdot,\cdot;l)$. The t-product of $\mathcal{T}_1$ and $\mathcal{T}_2$ in original domain is equivalent to frontal slice-wise matrix product of $\mathcal{T}_1$ and $\mathcal{T}_2$ in spectral domain, i.e.,

$$\mathcal{T} = \mathcal{T}_1 \ast \mathcal{T}_2 \iff \tilde{\mathcal{T}}^{(l)} = \tilde{\mathcal{T}}_1^{(l)} \tilde{\mathcal{T}}_2^{(l)}, \forall l \in [d_3]. \quad (15)$$

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

$$\mathcal{T} = \mathcal{U} \ast \mathcal{S} \ast \mathcal{V}^\top, \quad (16)$$

where $\mathcal{U} \in \mathbb{R}^{d_1 \times d_1 \times d_3}$, $\mathcal{S} \in \mathbb{R}^{d_2 \times d_2 \times d_3}$ are orthogonal tensors, and $\mathcal{S} \in \mathbb{R}^{d_3 \times d_3 \times d_3}$ 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.](image)

**Definition 9** (Identity tensor (Kilmer et al. 2013)). The identity tensor $\mathcal{I} \in \mathbb{R}^{d \times d \times d}$ 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 $\mathcal{I}^{(l)} = \mathbf{I}_d \in \mathbb{R}^{d \times d}$, $\forall l \in [d_3]$.

**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 $\mathcal{Q} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is orthogonal if

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

In spectral domain, we have

$$(\mathcal{Q}^{(l)})^\top \mathcal{Q}^{(l)} = (\mathcal{Q}^{(l)} \mathcal{Q}^{(l)}) = \mathbf{I}_d \in \mathbb{R}^{d \times d}, \forall l \in [d_3], \quad (17)$$

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 $\mathcal{T}$ (or $\mathcal{T}'$) denote the block-diagonal matrix of the tensor $\mathcal{T}$ in the Fourier domain, i.e.,

$$\tilde{\mathcal{T}} := \begin{bmatrix} \tilde{\mathcal{T}}^{(1)} & \cdots & \tilde{\mathcal{T}}^{(d_3)} \end{bmatrix} \in \mathbb{C}^{d_1 \times d_2 \times d_3} \quad (18)$$

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

$$\mathcal{T} = \mathcal{T}_1 \ast \mathcal{T}_2 \iff \tilde{\mathcal{T}} = \tilde{\mathcal{T}}_1 \tilde{\mathcal{T}}_2.$$

We further have the following relationship for t-SVD

$$\mathcal{T} = \mathcal{U} \ast \mathcal{S} \ast \mathcal{V}^\top \iff \tilde{\mathcal{T}} = \tilde{\mathcal{U}} \tilde{\mathcal{S}} \tilde{\mathcal{V}}^\top,$$

which also indicates that the average rank and TNF satisfy

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

$$\|\mathcal{T}\|_* = \frac{1}{d_3} \cdot \|\tilde{\mathcal{T}}\|_* = \frac{1}{d_3} \cdot \|\tilde{\mathcal{S}}\|_*.$$

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

$$\text{rank}_{\text{avg}}(\mathcal{T}) := \# \{ i \mid \mathcal{S}(i,i,:) \neq \mathbf{0} \} = \sum_{l=1}^{d_3} \text{rank}(\mathcal{S}^{(l)}) \geq \frac{1}{d_3} \sum_{l=1}^{d_3} \text{rank}(\mathcal{S}^{(l)}) \geq \text{rank}_{\text{avg}}(\mathcal{S}). \quad (19)$$

The inner product between two tensors $\mathcal{T}_1$ and $\mathcal{T}_2$ is defined as $\langle \mathcal{T}_1, \mathcal{T}_2 \rangle := \text{vec}(\mathcal{T}_1)^\top \text{vec}(\mathcal{T}_2)$. 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 $T_1, T_2 \in \mathbb{C}^{d_1 \times d_2 \times d_3}$ has the relationship
\[ \langle T_1, T_2 \rangle = \frac{1}{d_3} \langle \tilde{T}_1, \tilde{T}_2 \rangle = \frac{1}{d_3} \langle T_1^\top, T_2 \rangle. \] (20)

The relationship between matrix nuclear norm and matrix F-norm holds for any $M$:
\[ \|M\|_* \leq \sqrt{\text{rank}(M)} \|M\|_F. \] (21)

Similar relationship between TN and F-norm also holds for any tensor $T \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ as follows:
\[ \|T\|_* = \frac{1}{d_3} \|T\|_* \leq \frac{1}{d_3} \sqrt{d_3 \text{rank}_0(T)} \|T\|_F \]
\[ = \frac{1}{d_3} \sqrt{\text{rank}_0(T)} \sqrt{d_3 \|T\|_F} \]
\[ = \sqrt{\text{rank}_0(T)} \|T\|_F. \] (22)

It is also known that, for any tensor $T$, the $l_1$-norm and the F-norm has the following relationship
\[ \|T\|_1 = \sqrt{\|T\|_0 \|T\|_F}. \] (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)**

*Given $A$ and $B$ of the same dimension, if $AB^\top = 0$ and $A^\top B = 0$, then $\|A + B\|_* = \|A\|_* + \|B\|_*$.\*

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

**Lemma 4 (Additivity of Tubal Nuclear Norm).** *Given $T_1, T_2 \in \mathbb{R}^{d_1 \times d_2 \times d_3}$. If $T_1 \ast T_2^\top = 0$ and $T_1^\top \ast T_2 = 0$, then
\[ \|T_1 + T_2\|_* = \|T_1\|_* + \|T_2\|_* \] (24)

*Proof*. Using the relationship between a 3D tensor and its block-diagonal matrix, we have
\[ T_1 \ast T_2^\top = 0 \Rightarrow T_1^\top T_2^\top = 0, \]
\[ T_1^\top \ast T_2 = 0 \Rightarrow T_1^\top T_2 = 0. \] (25)

Thus, we obtain
\[ \|T_1 + T_2\|_* = \frac{1}{d_3} \|T_1 + T_2\|_* \]
\[ = \frac{1}{d_3} \left( \|T_1\|_* + \|T_2\|_* \right) \]
\[ = \|T_1\|_* + \|T_2\|_* \]. \] (26)

Suppose $X \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ with tubal rank $r^*$ has reduced t-SVD as follows
\[ X = U_X \ast S_X \ast V_X^\top, \] (27)

where $U_X \in \mathbb{R}^{d_1 \times r^* \times d_3}$ and $V_X \in \mathbb{R}^{d_2 \times r^* \times d_3}$ are orthogonal and $S_X \in \mathbb{R}^{r^* \times r^* \times d_3}$ is f-diagonal. Define a tensor space $T$ as follows:
\[ T = \left\{ U_X \ast A + B \ast V_X^\top : A, B \in \mathbb{R}^{d_1 \times r^* \times d_3} \right\}. \]

We further define the projectors to $T$ and $T^\perp$ as $P_T: \mathbb{R}^{d_1 \times d_2 \times d_3} \rightarrow \mathbb{R}^{d_1 \times d_2 \times d_3}$ and $P_{T^\perp}: \mathbb{R}^{d_1 \times d_2 \times d_3} \rightarrow \mathbb{R}^{d_1 \times d_2 \times d_3}$ respectively
\[ P_T (T) = U_X \ast \bar{U}_X \ast T \ast + \bar{T} \ast + \bar{V}_X \ast \bar{V}_X^\top \]
\[ - U_X \ast \bar{U}_X \ast \bar{T} \ast \bar{V}_X \ast \bar{V}_X^\top, \] (28)
\[ P_{T^\perp} (T) = (I - U_X \ast \bar{U}_X) \ast T \ast (I - \bar{V}_X \ast \bar{V}_X^\top). \]

Thus, we have
\[ \text{rank}_0(P_T (T)) \leq \text{rank}_0(U_X \ast U_X^\top) + \text{rank}_0((I - U_X \ast U_X^\top) \ast T \ast (I - \bar{V}_X \ast \bar{V}_X^\top)) \leq 2\text{rank}_0(X). \] (29)

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

**Lemma 5.** *Given $T \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, we have
\[ \|X + P_{T^\perp} (T)\|_* = \|X\|_* + \|P_{T^\perp} (T)\|_* \]. (30)*

*Proof*. It is easy to check that $X \ast P_{T^\perp} (T) \ast = 0$ and $X^\top \ast P_{T^\perp} (T) \ast = 0$. By Lemma 4, we have
\[ \|L^\ast + P_{T^\perp} (T)\|_* \leq \|L^\ast\|_* + \|P_{T^\perp} (T)\|_* \]. \]

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 $t_\varepsilon(T) \leq \min \{ \varepsilon(T), \varepsilon_{\text{tub}}(T) \}$. 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 $T \in \mathbb{R}^{d_1 \times \cdots \times d_K}$, let $K = [K] \subseteq \mathbb{R}^{d_1 \times \cdots \times d_K}$ denote its mode-$(k, k + 1)$ 3d-unfolding. We first show that $t_\varepsilon(T) \leq \varepsilon(T)$. Indeed, it holds that
\[ (t_\varepsilon(T))_k = \text{rank}_0(K)^{(i)} \]
\[ \leq \text{rank}_0(K) \leq \varepsilon(T)_k \]
\[ (\varepsilon(T))_k = \text{rank}_0(K) \leq \text{rank}(K) \] (31)
\[ \text{where inequality (i) holds due to Eq. (19).} \]
\[ \text{Then, we show } t_\varepsilon(T) \leq \varepsilon_{\text{tub}}(T). \text{ On the one hand, according to (Lu et al. 2019), we have}
\[ (t_\varepsilon(T))_k = \text{rank}_0(K) \leq \text{rank}(K) \] (32)
\[ \text{where } K_1 \text{ is the mode-1 matricization of } K. \text{ On the other hand, since } K \text{ is the mode-}(k, k + 1) \text{ 3d-unfolding of } T, \text{ it holds naturally that}
\[ \text{rank}(K_1) = \text{rank}(T(k)) = (\varepsilon_{\text{tub}}(T))_k \] (33)
where $T(k)$ is the mode-$k$ matricization of $T$. Thus, we have $\tilde{r}_a(T) \leq \tilde{r}_{\text{Tucker}}(T)$. Putting things together, we obtain $\tilde{r}_a(T) \leq \min\{\tilde{r}_a(T), r_{\text{Tucker}}(T)\}$.

## 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:

$$\left\| T \right\|_{*o}^* = \sup_{M} \left\langle M, T \right\rangle, \text{ s.t. } \|M\|_{*o} \leq 1,$$

and

$$\left\| T \right\|_{*o}^* = \sup_{M} \left\langle M, T \right\rangle, \text{ s.t. } \|M\|_{*e} \leq 1.$$

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

$$\sum_k \inf_{T^{(k)}} \max \left\{ w_k^{-1} \left\| T^{(k)} \right\| \right\}.$$

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

$$\sup_{M} \left\langle \left\{ M, T \right\} + \delta(\|M\|_{*o} \leq 1) \right\rangle = \inf_{\left\{ T^{(k)} \right\}} \left( \delta \sum_k \left\| T^{(k)} \right\| + \max \left\{ w_k^{-1} \left\| T^{(k)} \right\| \right\} \right),$$

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 $T \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ as $T_k(T) := T[k]$ and let $T_k^{-1}$ denote its inverse, i.e., $T_k^{-1}(T[k]) = T$.

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

**Proof.** Recall model OITNN-O:

$$\hat{(\mathcal{L}_o, \hat{S}_o)} \in \arg \min_{\mathcal{L}, S} f(\mathcal{L}, S),$$

$$\text{s.t. } \|\mathcal{L}\|_{1w} \leq \alpha,$$

where $f(\mathcal{L}, S) := \frac{1}{2}\left\| \mathcal{Y} - \mathcal{L} - \mathcal{S}\right\|_F + \lambda_d \|\mathcal{L}\|_{*o} + \mu_d \|\mathcal{S}\|_{1l}.$

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

$$f(\hat{\mathcal{L}}_o, \hat{\mathcal{S}}_o) \leq f(\mathcal{L}^*, \mathcal{S}^*).$$

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

$$\frac{1}{2} \left( \|\Delta_o^L\|_F^2 + \|\Delta_o^S\|_F^2 \right) \leq \lambda_d \|\mathcal{L}^*\|_{1w} - \|\mathcal{L}^* + \Delta_o^L\|_{1w} + \mu_d \|\mathcal{L}^*\|_{1l} - \|\mathcal{L}^* + \Delta_o^L\|_{1l}$$

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 $\mathcal{P}_k(\cdot) = \mathcal{P}_{\mathcal{E}_k}(\cdot)$ (see the definition of $\mathcal{P}_T$ in Eq. (28)). For any tensor $\Delta \in \mathbb{R}^{d_i \times \cdots \times d_K}$, define $\Delta_k = \mathcal{P}_k(\Delta)$, and $\Delta_k = \Delta - \Delta_k$.

Using Lemma 4 directly yields

$$\|\mathcal{L}^* - \Delta\|_F^2 = \|\mathcal{L}^*\|_F^2 + \|\Delta\|_F^2.$$

It leads to

$$\|\Delta\|_F^2 \leq \|\Delta_k\|_F^2 + \|\Delta_k\|_F^2 \leq \|\Delta_k\|_F^2 + \|\Delta_k\|_F^2,$$

Thus, we have

$$I = \lambda_d \sum_k \omega_k \left( \|\mathcal{L}^*\|_{1l} - \|\mathcal{L}^* - \Delta_k\|_{1l} \right) \leq \lambda_d \sum_k \omega_k \left( \|\mathcal{L}^*\|_{1l} - \|\mathcal{L}^* - \Delta_k\|_{1l} \right)$$

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

$$\|\mathcal{T}\|_{1l} = \|\mathcal{T}\|_{l} + \|\mathcal{T}\|_{1l}.$$
Combining Eq. (38) and Eqs (39)-(41) yields
\[
\frac{1}{2} (\|\Delta_L\|_F^2 + \|\Delta_S\|_F^2) \\
\leq \lambda_0 \sum_k w_k \|\Delta'_L\|_* + \lambda_0 \sum_k w_k \|\Delta'_S\|_* + \mu_0 \|\Delta_{S, l}\|_{l_1} \\
- \mu_0 \|\Delta_{S, l}\|_{l_1} + \|\mathcal{E}\|_{l_1} + (\|\mathcal{E}\|_{l_1} + 2\alpha)\|\Delta_{S, l}\|_{l_1},
\]
with probability at least $1 - \exp(-cD)$.

Choosing
\[
\lambda_0 > 2\|\mathcal{E}\|_{l_1},
\]
and
\[
\mu_0 \geq 2\|\mathcal{E}\|_{l_1} + 2\alpha,
\]
we have
\[
\lambda_0 \sum_k w_k \|\Delta'_L\|_* + \mu_0 \|\Delta_{S, l}\|_{l_1} \\
\leq 3(\lambda_0 \sum_k w_k \|\Delta'_L\|_* + \mu_0 \|\Delta_{S, l}\|_{l_1}).
\]

Note that according to Eq. (38) and the triangular inequality, we obtain
\[
\frac{1}{2} (\|\Delta_L\|_F^2 + \|\Delta_S\|_F^2) \\
\leq (\lambda_0 + \|\mathcal{E}\|_{l_1}) (\sum_k w_k \|\Delta'_L\|_* + \sum_k w_k \|\Delta'_S\|_* + \mu_0 \|\Delta_{S, l}\|_{l_1} + \mu_0 \|\Delta_{S, l}\|_{l_1}) \\
+ (\mu_0 + \|\mathcal{E}\|_{l_1} + 2\alpha) (\|\Delta_{S, l}\|_{l_1} + \|\Delta_{S, l}\|_{l_1}).
\]
That leads to
\[
\|\Delta_L\|_F^2 + \|\Delta_S\|_F^2 \leq 16\lambda_0 \sum_k w_k \|\Delta'_L\|_* + 16\mu_0 \|\Delta_{S, l}\|_{l_1}. \tag{46}
\]

By the definition of $(\Delta'_L)'_k$, we have
\[
\text{rank}_k((\Delta'_L)'_k) \leq 2\text{rank}_k(\mathcal{E}' [k]) = 22^k, \tag{47}
\]
and
\[
\|((\Delta'_L)'_k)\|_F \leq \|((\mathcal{E}' [k])\|_F = \|\Delta_{L, l}\|_F. \tag{48}
\]
We also have $\|((\Delta'_S)\|_{l_1} \leq |S| = s$. Then we reach the inequality:
\[
\|\Delta_L\|_F^2 + \|\Delta_S\|_F^2 \leq 16\lambda_0 \sum_k w_k \sqrt{22^k\|\Delta_L\|_F^2} + 16\mu_0 \sqrt{s}\|\Delta_{S, l}\|_{F}. \tag{49}
\]

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

**Proof of Theorem 3.** The key of proving Theorem 3 is to bound the quantity $\|\mathcal{E}\|_{l_1}^*$, 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 ($K \geq 3$) tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$, the following inequality holds:
\[
\|\mathcal{T}\|_{l_1}^* \leq \frac{1}{K^2} \sum_k w_k^{-1}\|\mathcal{T}_k\|_1. \tag{50}
\]

**Proof.** Recall the formulation of $\|\mathcal{T}\|_{l_1}^*$ as follows:
\[
\|\mathcal{T}\|_{l_1}^* := \inf_{\sum_k \mathcal{T}^{(k)} = \mathcal{T}} \max_k \|w_k^{-1}\|_{\mathcal{T}^{(k)}_1}\|_{l_1}^*.
\]

Letting
\[
\mathcal{T}^{(k)} = \frac{w_k\|\mathcal{T}_k^{(k)}\|_1^{-1}}{\sum_k w_k\|\mathcal{T}_k^{(k)}\|_1^{-1}} \mathcal{T},
\]
then for any $k \in [K]$,
\[
w_k^{-1}\|\mathcal{T}_k^{(k)}\|_1 = \frac{w_k^{-1}\|\mathcal{T}_k^{(k)}\|_1^{-1} - \sum_k w_k\|\mathcal{T}_k^{(k)}\|_1^{-1}}{\sum_k w_k\|\mathcal{T}_k^{(k)}\|_1^{-1}} \mathcal{T}_k^{(k)}
\]
\[
\leq \frac{1}{K^2} \sum_k w_k^{-1}\|\mathcal{T}_k^{(k)}\|_1,
\]
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 $\mathcal{T} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ be random tensors with i.i.d. Gaussian entries $\mathcal{N}(0, 1)$. Then the following inequality holds
\[
\|\mathcal{T}\| \leq \sqrt{d_1d_2 + \cdots + d_K^2} + t , \tag{54}
\]
with probability at least $1 - \exp(-c\sigma^2)$.

**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 $l_{\infty}$-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 holds
\[
\|\mathcal{E}\|_{l_{\infty}} \leq \sqrt{2\log(2D)} + t , \tag{55}
\]
with probability at least $1 - \exp(-c\sigma^2)$.

**Proof of Theorem 3.** Using Lemma III.2, we have for any $k \in [K]$,
\[
\|\mathcal{E}_k\| \leq 2\sigma d_k,
\]
with probability at least $1 - \exp(-c\sigma^2)$. Taking union bound, we have with probability at least $1 - \sum_k \exp(-c\sigma^2 d_k^2/d_{k+1})$,
\[
\|\mathcal{E}\|_{l_{\infty}} \leq \frac{2\sigma}{K} \sum_k \sqrt{d_k}. \tag{56}
\]

According to Lemma 9, we also have
\[
\|\mathcal{E}\|_{l_{\infty}} \leq 4\sigma \sqrt{\log D}, \tag{58}
\]
with probability at least $1 - \exp(-c'\sigma D)$.

Combining 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 of the unfolding operator for $T \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ as $\hat{g}_k(T) := T[k]$ and its inverse such that $\hat{g}_k^{-1}(T[k]) = T$.

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

**Proof.** Recall model OITNN-L:
\[
\left(\{L^{(k)}\}_k, S_\ast\right) \in \text{argmin} \ g(\{L^{(k)}\}_k, S),
\]
where
\[
g(\{L^{(k)}\}_k, S) = \frac{1}{2} \left\| Y - \sum_k L^{(k)}_S - S \right\|_F^2 + \lambda \sum_k \left\| L^{(k)}_S \right\|_1 + \mu_1 \left\| S \right\|_1.
\]

Let $\Delta^{l,k}_e = L^{(k)}_{S+} - L^{(k)}_S$ and $S^* = S - S_\ast$. Using the optimality of $\left(\{L^{(k)}\}_k, S_\ast\right)$, we have:
\[
g(\{L^{(k)}\}_k, S_\ast) \leq g(\{L^{(k)}\}_k, S^*).\]

Note that $\sum_k L^{(k)}_{S^*} = L^*$. Through the observation model of RTD, we have $Y - L^* - S^* = E$. After some algebra, we obtain
\[
\frac{1}{2} \left( \sum_k \left\| L^{(k)}_{S^*} \right\|_F^2 + \left\| \Delta^{l,k}_S \right\|_F^2 \right) \leq \lambda \sum_k \left\| L^{(k)}_{S^*} \right\|_1 - \left\| \Delta^{l,k}_S \right\|_1 + \mu_1 \left\| S \right\|_1 - \left\| S^* - \Delta^{l,k}_S \right\|_1.
\]

\[
\left( \sum_k \Delta^{l,k}_S, \Delta^S \right) = \left( \sum_k \left\| \Delta^{l,k}_S \right\|_1 + \left\| \Delta^S \right\|_1 \right).
\]

and
\[
\frac{1}{2} \left( \sum_k \left\| \Delta^{l,k}_S \right\|_F^2 + \left\| \Delta^S \right\|_F^2 \right) \leq I + II + III + IV + V + \sum_k \left\| \Delta^{l,k}_S \right\|_1 + \left\| \Delta^S \right\|_1.
\]

We will bound Items I to VI as follows.

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

Using Lemma 1 directly yields
\[
\left\| L^{(k)}_{S^*} - \Delta^{l,k}_S \right\|_1 = \left\| L^{(k)}_{S^*} - \Delta^{l,k}_S \right\|_1 + \left\| \Delta^{l,k}_S \right\|_1.
\]

leading to
\[
\left\| L^{(k)}_{S^*} - \Delta^{l,k}_S \right\|_1 = \left\| L^{(k)}_{S^*} - \Delta^{l,k}_S \right\|_1 + \left\| \Delta^{l,k}_S \right\|_1.
\]

Thus, we have
\[
I = \lambda \sum_k v_k \left( \left\| L^{(k)}_{S^*} \right\|_1 - \left\| L^{(k)}_S \right\|_1 \right) + \mu \left\| S \right\|_1.
\]

**Bound Item II.** Similar to the proof of Theorem 1, we have
\[
II \leq \mu \left\| \Delta^S \right\|_1 + \mu_1 \left\| \Delta^S \right\|_1.
\]

**Bound Items III and V.** Using the definition of dual norm, we have
\[
III + V \leq \left\| E \right\|_{1,\infty} + 2\alpha \left\| \Delta^S \right\|_1.
\]

**Bound Item VI.**
\[
\sum_k \sum_{l \neq k} \left\langle \Delta^{l,k}_S, \Delta^{l,k}_S \right\rangle \leq \sum_k \sum_{l \neq k} \left\| \Delta^{l,k}_S \right\|_1 \left\| \Delta^{l,k}_S \right\|_1.
\]

Combining Eq. (63) and the above bounds yields
\[
\frac{1}{2} \left( \sum_k \left\| \Delta^{l,k}_S \right\|_F^2 + \left\| \Delta^S \right\|_F^2 \right) \leq \lambda \sum_k v_k \left\| \Delta^{l,k}_S \right\|_1 + \mu \left\| S \right\|_1.
\]


Choosing
\[ \lambda_i > 2\|E\|_{\infty}^\star (\sum \lambda_{i_k} + (K-1)\beta \max_i (\tilde{d}_i/v_i)) \] (69)
and
\[ \mu_i \geq \|E\|_{l\infty} + 2\alpha, \] (70)
we have
\[ \lambda_i \sum_k v_k \|\Delta\lambda^{(k)}\|_i^\star + \mu_i \|\Delta S\|_S^i \|S\|_i \leq 3(\lambda_i \sum_k v_k \|\Delta\lambda^{(k)}\|_i^\star + \mu_i \|\Delta S\|_S^i \|S\|_i). \] (71)

Note that according to Eq. (63) and the triangular inequality, we have
\[ \frac{1}{2} \left( \sum_k \|\Delta\lambda^{(k)}\|^2_F + \|\Delta S\|^2_F \right) \leq (\lambda_i + \|E\|_{\infty}^\star + (K-1)\beta \max_i (\tilde{d}_i/v_i))) \left( \sum_k v_k \|\Delta\lambda^{(k)}\|_i^\star + \sum_k v_k \|\Delta S\|_S^i \|S\|_i \right) + \left( \mu_i + \|E\|_{l\infty} + 2\alpha \right) \left( \|\Delta S\|_S^i \|S\|_i + \|\Delta S\|_S^i \|S\|_i + \|\Delta S\|_S^i \|S\|_i \right). \] (72)

That leads to
\[ \sum_k \|\Delta\lambda^{(k)}\|^2_F + \|\Delta S\|^2_F \leq 16\lambda_i \sum_k v_k \|\Delta\lambda^{(k)}\|_i^\star + 16\mu_i \|\Delta S\|_S^i \|S\|_i. \] (73)

By the definition of \( \Delta\lambda^{(k)} \), we have
\[ \text{rank}_k(\Delta\lambda^{(k)}_i) \leq 2\text{rank}_k(\lambda^{(k)}_i) = 2\tau_k, \] (74)
and
\[ \|\Delta\lambda^{(k)}\|_F \leq \|\Delta\lambda^{(k)}\|_F = \|\Delta\lambda^{(k)}\|_F. \] (75)

We also have \( \|\Delta S\|_S \|S\|_i \leq |S| = s \). Then, we reach the inequality:
\[ \sum_k \|\Delta\lambda^{(k)}\|^2_F + \|\Delta S\|^2_F \leq 16\lambda_i \sum_k v_k \sqrt{2\tau_k} \|\Delta\lambda^{(k)}\|_F + 16\mu_i \sqrt{s} \|\Delta S\|_F \leq 16\lambda_i \sum_k \left( \frac{v_k \sqrt{2\tau_k}}{\sqrt{\tau_k}} \right)^2 \|\Delta\lambda^{(k)}\|_F + 16\mu_i \sqrt{s} \|\Delta S\|_F \] (76)

The usage of \( ab \leq a^2/4 + b^2 \) leading to the first part of Theorem 2, i.e.,
\[ \sum_k \|\Delta\lambda^{(k)}\|^2_F + \|\Delta S\|^2_F \leq c_3 \lambda_i^2 \sum_k v_k^2 \tau_k + c_4 \mu_i^2 s. \] (77)

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

**Case 1:** If \( \sum_k \|\Delta\lambda^{(k)}\|^2_F \leq \sum_i \|\Delta\lambda^{(k)}\|_i^\star \), according to Eq. (77) we have
\[ \sum_k \|\Delta\lambda^{(k)}\|^2_F + \|\Delta S\|^2_F \leq c_3 \lambda_i^2 \sum_k v_k^2 \tau_k + c_4 \mu_i^2 s. \] (78)

**Case 2:** If \( \sum_k \|\Delta\lambda^{(k)}\|^2_F > \sum_i \|\Delta\lambda^{(k)}\|_i^\star \), according to Eq. (62), we have
\[ \frac{1}{2} \left( \sum_k \|\Delta\lambda^{(k)}\|^2_F + \|\Delta S\|^2_F \right) \leq \left( \lambda_i + \|E\|_{\infty}^\star \right) \left( \sum_k v_k \|\Delta\lambda^{(k)}\|_i^\star \right) + \left( \mu_i + \|E\|_{l\infty} + 2\alpha \right) \left( \|\Delta S\|_S^i \|S\|_i + \|\Delta S\|_S^i \|S\|_i \right), \] (79)
which leads to
\[ \sum_k \|\Delta\lambda^{(k)}\|^2_F + \|\Delta S\|^2_F \leq 16\lambda_i \sum_k v_k \sqrt{2\tau_k} \|\Delta\lambda^{(k)}\|_F + 16\mu_i \sqrt{s} \|\Delta S\|_F \]
\[ \leq 16\lambda_i \sqrt{\sum_k \left( \frac{v_k \sqrt{2\tau_k}}{\sqrt{\tau_k}} \right)^2 \|\Delta\lambda^{(k)}\|_F + 16\mu_i \sqrt{s} \|\Delta S\|_F \]
\[ \leq 16\lambda_i \sqrt{\sum_k \left( \frac{v_k \sqrt{2\tau_k}}{\sqrt{\tau_k}} \right)^2 \|\Delta\lambda^{(k)}\|_F + 16\mu_i \sqrt{s} \|\Delta S\|_F \]
\[ \leq c_3 \lambda_i^2 \sum_k v_k^2 \tau_k + c_4 \mu_i^2 s. \] (80)

Select \( k^* \in \arg \max_k v_k^2 \text{rank}_k(\lambda^{(k)}_i) \). Letting \( \lambda^{(k)} = \lambda^* \) and \( \lambda^{(l)} = \lambda^* \), \( \forall l \neq k^* \), then \( \{\lambda^{(k)}_i, S^*\} \) is feasible. In this case, \( \tau_{k^*} = \lambda^*_k \) and \( \tau_l = 0, \forall l \neq k^* \). Then, we obtain
\[ \sum_k \|\Delta\lambda^{(k)}\|^2_F + \|\Delta S\|^2_F \leq c_3 \lambda_i^2 \sum_k v_k^2 \tau_k + c_4 \mu_i^2 s. \] (81)

Then, the proof is completed. \( \square \)

**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 \( T \in \mathbb{R}^{d_1 \times \cdots \times d_K} \) as \( \mathcal{F}_k(T) := T_{[k]} \) and its \( \mathcal{F}_k^{-1}(\cdot) \) such that \( \mathcal{F}_k^{-1}(T_{[k]}) = T \).

**Algorithm 1 ADMM for Model I**

**Input:** Observation \( \mathcal{Y} \), parameters \( \lambda_k, \mu_k, \{ w_k \} \), \( \rho > 0 \), \( \varepsilon > 0 \).

1. Initialize \( F^0 = K^0 = W^0 = S^0 = Z^0 = 0 \).

2: while not converged do

3. Update \((L_{t+1}, S_{t+1})\) simultaneously by:

4. Update \( \{ K_{k+1} \}_k, T_{t+1} \) and \( \mathcal{K}^{t+1} \) simultaneously by:

5. Dual update:

6. Check the convergence conditions:

7. \( t = t + 1 \).

8: end while

**Algorithm 2 ADMM for Model II**

**Input:** Observation \( \mathcal{Y} \), parameters \( \lambda_k, \mu_k, \{ w_k \} \), \( \rho > 0 \), \( \varepsilon > 0 \).

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

2: while not converged do

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

4. Update \( \{ K_{k+1} \}_k, T_{t+1} \) and \( \mathcal{K}^{t+1} \) simultaneously by:

5. Dual update: \( Z_{t+1} = Z_t + \rho (T_{t+1} - S_{t+1}) \), \( W_{t+1} = W_t + \rho (\mathcal{K}_{t+1} - \mathcal{L}_{t+1}) \), and \( Y_{t+1} = Y_t + \rho (\mathcal{K}_{t+1} - \mathcal{L}_{t+1}) \), \( \forall k \in [K] \).

6. Check the convergence conditions:

7. \( t = t + 1 \).

8: end while

The proximal operator of \( l_1\)-norm \( \| \cdot \|_1 \) is given as

\[
\text{Prox}_{\| \cdot \|_1}(\mathcal{T}_0) = \arg \min_{\mathcal{T}} \frac{1}{2} \| \mathcal{T}_0 - \mathcal{T} \|_2^2 + \| \mathcal{T} \|_1
\]

(83)

\[= \text{sgn}(-\mathcal{T}_0) \oplus \max (|\mathcal{T}_0| - \tau, 0),\]

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

\[
\text{Pro}_{\delta_{\alpha}^\infty}(\mathcal{T}_0) = \arg \min_{\mathcal{T}} \frac{1}{2} \| \mathcal{T}_0 - \mathcal{T} \|_2^2 + \delta_{\alpha}^\infty(\mathcal{T}_0)
\]

(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

\[
\min_{\mathcal{L}, \mathcal{S}, \mathcal{K}, \mathcal{T}} \frac{1}{2} \| \mathcal{Y} - \mathcal{L} - \mathcal{S} \|_2^2 + \lambda_o \sum_k w_k \| \mathcal{K}_k \|_* + \mu_o \| \mathcal{T} \|_1 + \delta_{\alpha}^\infty(\mathcal{K})
\]

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

**Several operators**

Before giving solutions to the sub-problems in Algorithm 1 and Algorithm 2, we briefly give the proximal operators of TNN \( \| \cdot \|_1 \) as follows:

**Lemma 10.** (Wang and Jin 2017). Let tensor \( \mathcal{T} \in \mathbb{R}^{d_1 \times \cdots \times d_K} \) with t-SVD \( \mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T \), where \( \mathcal{U} \in \mathbb{R}^{d_1 \times d_r} \) and \( \mathcal{V} \in \mathbb{R}^{d_K \times d_r} \) are orthogonal tensors and \( \mathcal{S} \in \mathbb{R}^{r \times r \times d_r} \) is the f-diagonal tensor of singular tubes. Then the proximal operator of function \( \| \cdot \|_1 \), at point \( \mathcal{T}_0 \), denoted by \( \text{Pro}_{\| \cdot \|_1}(\mathcal{T}_0) \), can be computed as follows:

\[
\text{Pro}_{\| \cdot \|_1}(\mathcal{T}_0) = \arg \min_{\mathcal{T}} \frac{1}{2} \| \mathcal{T}_0 - \mathcal{T} \|_2^2 + \| \mathcal{T} \|_1
\]

\[= \mathcal{U} * \text{ifft3}((\text{max}(\text{fft3}(\mathcal{S}) - \tau, 0)) + \mathcal{V}^T).
\]

(82)
Then, the augmented Lagrangian is given as follows
\[ L^1_{\rho}(L, S, T, K, \{K_k\}, \{Y_k\}, \mathcal{L}, \mathcal{W}) \]
\[ = \frac{1}{2} \|Y - L - S\|^2 + \lambda_0 \sum_k w_k \|K_k\|_* + \mu_\|T\|_{l_1} + \delta^\infty(K) \]
\[ + \sum_k \left( \langle Y_k, K_k - \tilde{s}_k(L) \rangle + \frac{\rho}{2} \|K_k - \tilde{s}_k(L)\|_F^2 \right) \]
\[ + \langle Z, T - S \rangle + \frac{\rho}{2} \|T - S\|^2_F + \langle \mathcal{W}, K - L \rangle + \frac{\rho}{2} \|K - L\|_F^2. \]

Further, we update blocks \((L, S, T, K)\) alternatively by fixing the other variables.

**Update** \((L, S, T, K)\).

Fixing \((\{K_k\}, \{Y_k\})\), we update \((L, S, T)\) by minimizing the augmented Lagrangian \(L^1_{\rho} \) with respect to \((L, S, T)\), which can be simplified as follows
\[
\min_{L, S} l(L, S) + \sum_k \frac{\rho}{2} \|L - \tilde{s}_k^1(L^t_k) + \frac{\rho}{2} T - S\|^2_F + \frac{\rho}{2} \|T - (K - L) + \frac{\rho}{2} T_{\infty} \|^2_F.
\]

Taking the derivatives with respect to \(L\) and \(S\) and setting the derivatives to zero, we obtain
\[
(K + 1)\mathcal{L} + S = \rho\mathcal{K} + \sum_k \mathcal{K}_k + \mathcal{Y} = \mathcal{T}.
\]

**Update** \((\{K_k\}, \mathcal{L}, \mathcal{W})\).

Fixing \((L, S, T)\), we update \((\{K_k\})\), \(\mathcal{L}\), and \(\mathcal{W}\) by minimizing the augmented Lagrangian \(L^1_{\rho} \) with respect to \((\{K_k\}, \mathcal{L}, \mathcal{W})\), which can be simplified to the following problem
\[
\min_{\mathcal{L}, \mathcal{W}} l(\mathcal{L}, \mathcal{W}) + \sum_k \frac{\rho}{2} \|\mathcal{L}^t - \tilde{s}_k^1(K^t_k) + \frac{\rho}{2} \mathcal{T} - \tilde{s}_k - \sum_k \mathcal{T}_k\|^2_F.
\]

**Solutions to Sub-problems in Algorithm 2**

We solve the sub-problems in Algorithm 2 as follows. First, adding auxiliary variables Equation (13) yields
\[
\begin{align*}
\min_{\{K_k\}, \mathcal{L}, \mathcal{W}} & \quad l(\sum_k \mathcal{L}^t_k, S) + \lambda_0 \sum_k w_k \|K_k\|_* + \mu_\|T\|_{l_1} + \delta^\infty(K) \\
\text{s.t.} & \quad K_k = \tilde{s}_k(L^t_k), \forall k; T = S; K = \sum_k \mathcal{L}^t_k.
\end{align*}
\]

Then, the augmented Lagrangian is given as follows
\[ L^1_{\rho}(\mathcal{L}^t_k, S, T, K, \{K_k\}, \{Y_k\}, \mathcal{L}, \mathcal{W}) \]
\[ = \frac{1}{2} \|Y - \mathcal{L}^t - S\|^2 + \lambda_0 \sum_k w_k \|K_k\|_* + \mu_\|T\|_{l_1} + \delta^\infty(K) \]
\[ + \sum_k \left( \langle Y_k, K_k - \tilde{s}_k(L^t_k) \rangle + \frac{\rho}{2} \|K_k - \tilde{s}_k(L^t_k)\|_F^2 \right) \]
\[ + \langle Z, T - S \rangle + \frac{\rho}{2} \|T - S\|^2_F + \langle \mathcal{W}, K - \mathcal{L}^t \rangle + \frac{\rho}{2} \|K - \mathcal{L}^t\|_F^2. \]

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

**Update** \((\{\mathcal{L}^t_k\}), S)\).

Fixing \((\{K_k\}, \mathcal{T}, K)\), we update \((\{\mathcal{L}^t_k\})\) by minimizing the augmented Lagrangian \(L^1_{\rho} \) with respect to \((\mathcal{L}, S)\), which can be simplified to the following problem
\[
\begin{align*}
\min_{\mathcal{L}, S} & \quad l(\sum_k \mathcal{L}^t_k, S) + \sum_k \frac{\rho}{2} \|\mathcal{L}^t_k - \tilde{s}_k^1(K^t_k) + \frac{\rho}{2} \mathcal{T} - \tilde{s}_k - \sum_k \mathcal{T}_k\|^2_F \\
& \quad + \frac{\rho}{2} \|S - (\mathcal{T} + \frac{\rho}{2} \mathcal{T}_{\infty})\|^2_F + \frac{\rho}{2} \sum_k \|\mathcal{L}^t_k - (\mathcal{L}^t + \frac{\rho}{2} \mathcal{T}_{\infty})\|^2_F.
\end{align*}
\]

Taking the derivatives with respect to \(\mathcal{L}^t\) and \(S\) and setting the derivatives to zero, we obtain
\[
\sum_k \mathcal{L} = S + \mathcal{Y} + \rho \mathcal{L}^t = \rho \mathcal{K} + \sum_k \mathcal{L}^t = 0.
\]

and
\[
\sum_k \mathcal{L}^t + S = \mathcal{Y} + \rho S - \mu \mathcal{T} = 0.
\]
where 

\[
M = \frac{K(1 + \rho)K}{(1 + K)(1 + \rho)} + \frac{1}{(1 + K)(1 + \rho)} \sum_k K_k + KY - K\hat{T}.
\]

**Update** \((\{K_k\}, T, K)\). Fixing \((\{L(k)\}, S)\), we update \(\{K_k\}, T, K\) by minimizing the augmented Lagrangian \(L_{ho}^t\) with respect to \((\{K_k\}, T, K)\). The problem can be solved separately as follows.

\[
\begin{align*}
K_k^{t+1} &= \min_{K_k} \lambda_v \|\text{Cov}(K_k)\|_* + \frac{\rho}{2} \|K_k - \tilde{K}_k((\text{Cov}(k))^{t+1}) + \frac{\|\gamma_l\|_F^2}{\rho}, \\
&= \text{Prox}_{\lambda_v \|\cdot\|_*/\rho}(\tilde{K}_k((\text{Cov}(k))^{t+1}) - \frac{\|\gamma_l\|_F^2}{\rho}), \\
T^{t+1} &= \min_{T} \|S(T - (S^{t+1} - \frac{Z^t}{\rho})^2 \|_F^2, \\
&= \text{Prox}_{\|\cdot\|_*/\rho}(S^{t+1} - \frac{Z^t}{\rho}), \\
K^{t+1} &= \min_{K} \delta_{\alpha_\infty}(K) + \frac{\rho}{2} \|K - \frac{1}{k} \sum_k (\text{Cov}(k))^{t+1} + \frac{\|W_l^t\|_F^2}{\rho}, \\
&= \text{Prox}_{\|\cdot\|_*/\rho}(\frac{1}{\rho} (\text{Cov}(k))^{t+1} - \frac{\|W_l^t\|_F^2}{\rho}).
\end{align*}
\]

**References**


