Supplementary Material of Guaranteed Multidimensional Time Series Prediction via Deterministic Tensor Completion Theory

Hao Shu, Jicheng Li, Yu Jin, Hailin Wang

I. Proof of Theorem III.1

To provide a complete proof of the theorem, we first present the definitions of tensor projection and the subgradient of the nuclear norm. Subsequently, we derive Lemmas I.1, I.2, I.3, I.4, and I.5, which lead us to Lemma I.6. Finally, we utilize Lemma I.6 to finish the proof.

Definition I.1. *[1] Let $\mathcal{M} \in \mathbb{R}^{m_1\times\cdots\times m_d}$ with $\operatorname{rank}_t(\mathcal{M}) = r$, and its skinny t-SVD is $\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\mathrm{T}}$, where $\mathcal{U} \in \mathbb{R}^{m_1\times r\times m_3\cdots\times m_d}$ and $\mathcal{V} \in \mathbb{R}^{m_2\times r\times m_3\cdots\times m_d}$ are the left and right singular tensor, respectively. Define $\mathbb{T}$ by the set $\mathbb{T} = \{\mathcal{U} * \mathcal{Y}^{\mathrm{T}} + \mathcal{Z} * \mathcal{V}^{\mathrm{T}} \mid \mathcal{Y} \in \mathbb{R}^{m_2\times r\times m_3\times\cdots\times m_d}, \mathcal{Z} \in \mathbb{R}^{n_1\times r\times m_3\times\cdots\times m_d}\}$ and by $\mathbb{T}^{\perp}$ its orthogonal complement. For any $\mathcal{X} \in \mathbb{R}^{m_1\times\cdots\times m_d}$, the projections onto $\mathbb{T}$ and its complementary set $\mathbb{T}^{\perp}$ are respectively denoted as*

$$\mathcal{P}_{\mathbb{T}}(\mathcal{X}) = \mathcal{U} * \mathcal{U}^{\mathrm{T}} * \mathcal{X} + \mathcal{X} * \mathcal{V} * \mathcal{V}^{\mathrm{T}} - \mathcal{U} * \mathcal{U} * \mathcal{X} * \mathcal{V} * \mathcal{V}^{\mathrm{T}},$$

and $\mathcal{P}_{\mathbb{T}^{\perp}}(\mathcal{X}) = \mathcal{X} - \mathcal{P}_{\mathbb{T}}(\mathcal{X})$. Similarly, define $\mathbb{U}, \mathbb{V}$ by the set

$$\mathbb{U} = \left\{\mathcal{U} * \mathcal{Y}^{\mathrm{T}} \mid \mathcal{Y} \in \mathbb{R}^{m_1\times r\times m_3\times\cdots\times m_d}\right\},$$
$$\mathbb{V} = \left\{\mathcal{Z} * \mathcal{V}^{\mathrm{T}} \mid \mathcal{Z} \in \mathbb{R}^{m_2\times r\times m_3\times\cdots\times m_d}\right\},$$

the projection on the set $\mathbb{U}, \mathbb{V}$ is as follows:

$$\mathcal{P}_{\mathbb{U}}(\mathcal{X}) = \mathcal{U} * \mathcal{U}^{\mathrm{T}} * \mathcal{X}, \mathcal{P}_{\mathbb{V}}(\mathcal{X}) = \mathcal{X} * \mathcal{V} * \mathcal{V}^{\mathrm{T}}.$$

Definition I.2 (Subgradient of tensor nuclear norm[1])**.** *Let $\mathcal{M} \in \mathbb{R}^{m_1\times\cdots\times m_d}$ with $\operatorname{rank}_t(\mathcal{M}) = r$, and it has skinny t-SVD $\mathcal{M} = \mathcal{U}*\mathcal{S}*\mathcal{V}^{\mathrm{T}}$. Then the subdifferential (the set of subgradients) of $\|\cdot\|_{\circledast}$ at $\mathcal{M}$ is $\partial\|\mathcal{M}\|_{\circledast} = \left\{\mathcal{U} * \mathcal{V}^{\mathrm{T}} + \mathcal{W} \mid \mathcal{U}^{\mathrm{T}} * \mathcal{W} = 0, \mathcal{W} * \mathcal{V} = 0, \|\mathcal{W}\| \le 1\right\}$.*

Lemma I.1 (Quantitative relationship between tensor norms)**.** *Let $\mathcal{X} \in \mathbb{R}^{m_1\times\cdots\times m_d}$ with $\operatorname{rank}_t(\mathcal{X}) = r$, the multi-rank of tensor $\mathcal{X}$ is denoted by $\mathbf{r}$, and the sum of the elements of $\mathbf{r}$ is represented by r_s, then*

$$\frac{1}{m}\|\mathcal{X}\| \le \frac{1}{\sqrt{m}}\|\mathcal{X}\|_F \le \|\mathcal{X}\|_{\circledast} \le \sqrt{r}\|\mathcal{X}\|_F \le \sqrt{\frac{r r_s}{m}}\|\mathcal{X}\|,$$

where $m = m_3 \times \cdots \times m_d$.

For a rectangular matrix $X \in \mathbb{R}^{m_1\times m_2}$ with rank r_0, the Frobenius norm, spectral norm and nuclear norm of the matrix X have the following relationship [2]:

$$\|X\| \le \|X\|_F \le \|X\|_* \le \sqrt{r_0}\|X\|_F \le r_0\|X\|.$$

According to the definition of tensor Frobenius norm, tensor spectral norm and tensor nuclear norm, we have

$$\|\mathcal{X}\| = \|\operatorname{bdiag}(\bar{\mathcal{X}})\| \le \|\operatorname{bdiag}(\bar{\mathcal{X}})\|_F = \sqrt{m}\|\mathcal{X}\|_F,$$

$$\begin{aligned}\|\mathcal{X}\|_F &= \frac{1}{\sqrt{m}}\|\operatorname{bdiag}(\bar{\mathcal{X}})\|_F \le \frac{1}{\sqrt{m}}\|\operatorname{bdiag}(\bar{\mathcal{X}})\|_* \\ &= \sqrt{m}\|\mathcal{X}\|_{\circledast},\end{aligned}$$

$$\begin{aligned}\|\mathcal{X}\|_{\circledast} &= \frac{1}{m}\|\operatorname{bdiag}(\bar{\mathcal{X}})\|_* \le \frac{1}{\sqrt{m}}\sqrt{rm}\|\operatorname{bdiag}(\bar{\mathcal{X}})\|_F \\ &= \sqrt{r}\|\mathcal{X}\|_F,\end{aligned}$$

$$\begin{aligned}\|\mathcal{X}\|_F &= \frac{1}{\sqrt{m}}\|\operatorname{bdiag}(\bar{\mathcal{X}})\|_F \le \frac{1}{\sqrt{m}}\sqrt{rm}\|\operatorname{bdiag}(\bar{\mathcal{X}})\| \\ &= \sqrt{r}\|\mathcal{X}\|,\end{aligned}$$

which implies that

$$\frac{1}{m}\|\mathcal{X}\| \leq \frac{1}{\sqrt{m}}\|\mathcal{X}\|_F \leq \|\mathcal{X}\|_{\circledast} \leq \sqrt{r}\|\mathcal{X}\|_F \\ \leq r\|\mathcal{X}\|.$$

In addition, the following formula proves that $\|\mathcal{X}\|_F \leq \sqrt{\frac{r_s}{m}}\|\mathcal{X}\|$:

$$\begin{aligned}\|\mathcal{X}\|_F &= \frac{1}{\sqrt{m}}\|\mathrm{bdiag}(\bar{\mathcal{X}})\|_F = \frac{1}{\sqrt{m}}\sqrt{\sum_{i=1}^{m}\|\bar{\mathcal{X}}^{(i)}\|_F^2} \\ &\leq \frac{1}{\sqrt{m}}\sqrt{\sum_{i=1}^{m}\mathbf{r}_{(i)}\|\bar{\mathcal{X}}^{(i)}\|^2} \leq \frac{1}{\sqrt{m}}\sqrt{\sum_{i=1}^{m}\mathbf{r}_{(i)}\max_i\|\bar{\mathcal{X}}^{(i)}\|^2} \\ &= \frac{\sqrt{\sum_{i=1}^{m}\mathbf{r}_{(i)}}}{\sqrt{m}}\max_i\|\bar{\mathcal{X}}^{(i)}\| = \sqrt{\frac{r_s}{m}}\|\mathcal{X}\|,\end{aligned}$$

where $\bar{\mathcal{X}}^{(i)} = \bar{\mathcal{X}}_{(:,:,i_3,\cdots,i_d)}$ for $i = i_3 + i_4 m_3 + \cdots + i_d m_3 \cdots m_{d-1}$. This completes the proof.

Lemma I.2. *Suppose that $\mathcal{M} \in \mathbb{R}^{m_1\times\cdots\times m_d}$ obeys the standard order-d tensor incoherence conditions and $\Omega \subseteq [m_1]\otimes[m_2]\otimes\cdots\otimes[m_d]$, $\rho(\Omega)$ is the minimum slice sampling rate of the sampling set Ω and $m = m_3\times\cdots\times m_d$, then*

$$\|\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^{\perp}}\mathcal{P}_{\mathbb{U}}\|_{op} \leq (1-\rho(\Omega))\mu r, \|\mathcal{P}_{\mathbb{V}}\mathcal{P}_{\Omega^{\perp}}\mathcal{P}_{\mathbb{V}}\|_{op} \leq (1-\rho(\Omega))\mu r.$$

Proof. For $i_1 \in [m_1], \cdots, i_d \in [m_d]$, define $\mathfrak{e}_{i_1\cdots i_d}$ as an $m_1\times\cdots\times m_d$ sized tensor with its $(i_1,\cdots,i_d)$-th entry equaling to 1 and the rest equaling to 0. Given any order- d tensor $\mathcal{X} \in \mathbb{R}^{m_1\times\cdots\times m_d}$, we have the following decomposition

$$\mathcal{X} = \sum_{i_1\cdots i_d}\langle\mathcal{X},\mathfrak{e}_{i_1\cdots i_d}\rangle\mathfrak{e}_{i_1\cdots i_d}.$$

For any $\mathcal{X} \in \mathbb{R}^{m_1\times\cdots\times m_d}$, the projection onto Ω is defined as

$$\mathcal{P}_{\Omega}(\mathcal{X}) = \sum_{i_1\cdots i_d}\delta_{i_1\cdots i_d}\langle\mathcal{X},\mathfrak{e}_{i_1\cdots i_d}\rangle\mathfrak{e}_{i_1\cdots i_d},$$

where $\delta_{i_1\cdots i_d} = 1_{(i_1,\cdots,i_d)\in\Omega}$ and $1_{(.)}$denotes the indicator function. We decompose $\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{U}}(\mathcal{X})$ into the following form $\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{U}}(\mathcal{X}) = \sum_{i_1\cdots i_d}\delta_{i_1\cdots i_d}\langle\mathcal{X},\mathcal{P}_{\mathbb{U}}(\mathfrak{e}_{i_1\cdots i_d})\rangle\mathcal{P}_{\mathbb{U}}(\mathfrak{e}_{i_1\cdots i_d})$ Since $\Omega^{\perp}$ is the complement of Ω, we have $\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^{\perp}}\mathcal{P}_{\mathbb{U}}(\mathcal{X}(:,i_2,:,\cdots,:)) = \sum_{i_1,i_3\cdots i_d}(1-\delta_{i_1\cdots i_d})\langle\mathcal{X}(:,i_2,;,\cdots,:),\mathcal{P}_{\mathbb{U}}(\mathfrak{e}_{i_1\cdots i_d})\rangle\mathcal{P}_{\mathbb{U}}(\mathfrak{e}_{i_1\cdots i_d})$, which gives that

$$\begin{aligned}&\|\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^{\perp}}\mathcal{P}_{\mathbb{U}}(\mathcal{X}(:,i_2,:,\cdots,:))\|_F \\ &\leq (1-\rho(\Omega))m_1 m_3\cdots m_d\|\mathcal{X}(:,i_2,;,\cdots,:)\|_F\|\mathcal{P}_{\mathbb{U}}(\mathfrak{e}_{i_1\cdots i_d})\|_F^2 \\ &\leq (1-\rho(\Omega))\mu r\|\mathcal{X}(:,i_2,;,\cdots,:)\|_F.\end{aligned}$$

the second inequality holds because

$$\begin{aligned}\|\mathcal{P}_{\mathbb{U}}(\mathfrak{e}_{i_1\cdots i_d})\|_F &= \|\mathcal{U}*\mathcal{U}^{\mathrm{T}}*\mathfrak{e}_{i_1\cdots i_d}\|_F = \|\mathcal{U}^{\mathrm{T}}*\mathfrak{e}_{i_1\cdots i_d}\|_F \\ &= \|\mathcal{U}^{\mathrm{T}}*\mathring{\mathfrak{e}}_1^{(i_1)}\|_F \leq \sqrt{\frac{\mu r}{m_1 m}}.\end{aligned}$$

Notice that

$$\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^{\perp}}\mathcal{P}_{\mathbb{U}}(\mathcal{X}(:,i_2,:,\cdots,:)) = (\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^{\perp}}\mathcal{P}_{\mathbb{U}}\mathcal{X})(:,i_2,:,\cdots,:),$$

we can obtain

$$\frac{\|(\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^{\perp}}\mathcal{P}_{\mathbb{U}}\mathcal{X})(:,i_2,:,\cdots,:)\|_F}{\|\mathcal{X}(:,i_2,:,\cdots,:)\|_F} \leq (1-\rho(\Omega))\mu r, \\ \forall i_2 \in [m_2], \mathcal{X} \in \mathbb{R}^{m_1\times\cdots\times m_d}.$$

Therefore, we get the bound of the operator norm of $\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^{\perp}}\mathcal{P}_{\mathbb{U}}$

$$\begin{aligned}\|\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^{\perp}}\mathcal{P}_{\mathbb{U}}\|_{op} &= \sup_{\mathcal{X}\in\mathbb{R}^{m_1\times\cdots\times m_d}}\sqrt{\frac{\|\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^{\perp}}\mathcal{P}_{\mathbb{U}}\mathcal{X}\|_F^2}{\|\mathcal{X}\|_F^2}} \\ &= \sup_{\mathcal{X}\in\mathbb{R}^{m_1\times\cdots\times m_d}}\sqrt{\frac{\sum_{i_2=1}^{m_2}\|(\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^{\perp}}\mathcal{P}_{\mathbb{U}}\mathcal{X})(:,i_2,:,\cdots,:)\|_F^2}{\sum_{i_2=1}^{m_2}\|\mathcal{X}(:,i_2,:,\cdots,:)\|_F^2}} \\ &\leq (1-\rho(\Omega))\mu r.\end{aligned}$$

Similarly, we can get $\|\mathcal{P}_{\mathbb{V}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{V}}\|_{op} \leq (1-\rho(\Omega))\mu r$ due to $\mathcal{P}_{\mathbb{V}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{V}}(\mathcal{X}(i_1,:,\cdots,:)) = (\mathcal{P}_{\mathbb{V}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{V}}\mathcal{X})(i_1,:,\cdots,:)$. □

Lemma I.3. *Let $\mathcal{M} \in \mathbb{R}^{m_1\times\cdots\times m_d}$ with skinny t-SVD $\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\mathrm{T}}$, $\Omega^\perp \subseteq \{1,\ldots,m_1\}\times\cdots\times\{1,\ldots,m_d\}$, $\mathcal{P}_{\mathbb{T}}, \mathcal{P}_{\mathbb{U}}, \mathcal{P}_{\mathbb{V}}$ is given by Definition I.1, then we have*

$$\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\|_{op} \leq \|\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{U}}\|_{op} + \|\mathcal{P}_{\mathbb{V}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{V}}\|_{op}.$$

Proof. First, we can prove that $\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\|_{op} = \|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\|_{op}^2$ using the self-conjugate property of the projection operator as follows,

$$\begin{aligned}\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\|_{op} &= \|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\|_{op}\\ &= \|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp})^*\|_{op}\\ &= \|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\|_{op}^2.\end{aligned}$$

Then we get the target conclusion based on $\mathcal{P}_{\mathbb{U}^\perp}\mathcal{X} = \mathcal{X} - \mathcal{P}_{\mathbb{U}}\mathcal{X}$ and the definition of the operator norm,

$$\begin{aligned}&\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\|_{op} = \|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\|^2 = \sup_{\|\mathcal{X}\|_F=1}\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}(\mathcal{X})\|_F^2\\ &= \sup_{\|\mathcal{X}\|_F=1}\|\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^\perp}(\mathcal{X}) + \mathcal{P}_{\mathbb{U}^\perp}\mathcal{P}_{\mathbb{V}}\mathcal{P}_{\Omega^\perp}(\mathcal{X})\|_F^2\\ &= \sup_{\|\mathcal{X}\|_F=1}(\|\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^\perp}(\mathcal{X})\|_F^2 + \|\mathcal{P}_{\mathbb{U}^\perp}\mathcal{P}_{\mathbb{V}}\mathcal{P}_{\Omega^\perp}(\mathcal{X})\|_F^2)\\ &\leq \sup_{\|\mathcal{X}\|_F=1}\|\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^\perp}(\mathcal{X})\|_F^2 + \sup_{\|\mathcal{X}\|_F=1}\|\mathcal{P}_{\mathbb{V}}\mathcal{P}_{\Omega^\perp}(\mathcal{X})\|_F^2\\ &= \|\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^\perp}\|_{op}^2 + \|\mathcal{P}_{\mathbb{V}}\mathcal{P}_{\Omega^\perp}\|_{op}^2\\ &= \|\mathcal{P}_{\mathbb{U}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{U}}\|_{op} + \|\mathcal{P}_{\mathbb{V}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{V}}\|_{op}.\end{aligned}$$

□

Lemma I.4. *Let $\mathcal{M} \in \mathbb{R}^{m_1\times\cdots\times m_d}$ with its skinny t-SVD $\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\mathrm{T}}$, $\mathcal{P}_{\mathbb{T}}$ is given by $\mathcal{P}_{\mathbb{T}}(\cdot) = \mathcal{U} * \mathcal{U}^{\mathrm{T}} * (\cdot) + (\cdot) * \mathcal{V} * \mathcal{V}^{\mathrm{T}} - \mathcal{U} * \mathcal{U} * (\cdot) * \mathcal{V} * \mathcal{V}^{\mathrm{T}}$, $\Omega \subseteq [m_1]\otimes[m_2]\otimes\cdots\otimes[m_d]$, then $\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}$ is invertible and $\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\|_{op} < 1$ are equivalent.*

Proof. On the one hand, we prove that $\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\|_{op} < 1$ can be derived from $\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}$ is invertible. we denote the vectorization of the tensor $\mathcal{X} \in \mathbb{R}^{m_1\times\cdots\times m_d}$ as $vec(\mathcal{X}) \in \mathbb{R}^{m_1\cdots m_d\times 1}$ following references [3, 4]. Suppose that the basis matrix associated with $\mathcal{P}_{\mathbb{T}}$ is given by $P \in \mathbb{R}^{m_1\cdots m_d\times a}$, $P^TP = I$; namely,

$$\mathrm{vec}(\mathcal{P}_{\mathbb{T}}(\mathcal{X})) = PP^T\,\mathrm{vec}(\mathcal{X}), \forall \mathcal{X} \in \mathbb{R}^{m_1\times\cdots\times m_d}.$$

Denote $\delta_{i_1\cdots i_d} = 1_{(i_1,\cdots,i_d)\in\Omega}$ where $1_{(.)}$denotes the indicator function, and define a diagonal matrix D as $D = \mathrm{diag}\,(\delta_{1\ldots1}, \delta_{2\ldots1}, \ldots, \delta_{i_1\ldots i_d}, .$ $\mathbb{R}^{m_1\cdots m_d\times m_1\cdots m_d}$. Notice that

$$\begin{aligned}\mathcal{P}_{\mathbb{T}}(\mathcal{X}) &= \mathcal{P}_{\mathbb{T}}\left(\sum_{i_1\cdots i_d}\langle\mathcal{X}, \mathfrak{e}_{i_1\cdots i_d}\rangle\,\mathfrak{e}_{i_1\cdots i_d}\right)\\ &= \sum_{i_1\cdots i_d}\langle\mathcal{X}, \mathfrak{e}_{i_1\cdots i_d}\rangle\,\mathcal{P}_{\mathbb{T}}\left(\mathfrak{e}_{i_1\cdots i_d}\right),\end{aligned}$$

where $\mathfrak{e}_{i_1\cdots i_d}$ an $m_1\times\cdots\times m_d$ sized tensor with its $(i_1,\cdots,i_d)$-th entry equaling to 1 and the rest equaling to 0, $\langle\cdot\rangle$ denotes the inner product between two tensors. With this notation, it is easy to see that $[\mathrm{vec}\left(\mathcal{P}_{\mathbb{T}}\left(\mathfrak{e}_{1\ldots1}\right)\right), \mathrm{vec}\left(\mathcal{P}_{\mathbb{T}}\left(\mathfrak{e}_{2\ldots1}\right)\right), \ldots, \mathrm{vec}\left(\mathcal{P}_{\mathbb{T}}\left(\mathfrak{e}_{m_1\cdots m_d}\right)\right)] = PP^T$. Similarly, we have

$$\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{X}) = \sum_{i_1\cdots i_d}\delta_{i_1\cdots i_d}\langle\mathcal{P}_{\mathbb{T}}(\mathcal{X}), \mathfrak{e}_{i_1\cdots i_d}\rangle\,\mathcal{P}_{\mathbb{T}}\left(\mathfrak{e}_{i_1\cdots i_d}\right),$$

and thereby $\mathrm{vec}\left(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{X})\right) = PP^TD\,\mathrm{vec}(\mathcal{P}_{\mathbb{T}}(\mathcal{X})) = PP^TDPP^T\,\mathrm{vec}(\mathcal{X})$. For $\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}$ to be invertible, the matrix P^TDP must be positive definite. Because, whenever P^TDP is singular, there exists $z \in \mathbb{R}^{m_1\cdots m_d\times 1}$ that satisfies $z \neq 0$ and $P^TDPz = 0$, and thus there exists $\mathcal{X} \in \mathcal{P}_{\mathbb{T}}$ and $\mathcal{X} \neq 0$ such that $\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{X}) = 0$; this contradicts the assumption that $\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}$ is invertible. Denote the minimal singular value of P^TDP as $0 < \sigma_{\min} \leq 1$. Since P^TDP is positive definite, we have

$$\begin{aligned}&\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}(\mathcal{X})\|_F = \|\mathrm{vec}\left(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}(\mathcal{X})\right)\|_2\\ &= \left\|\left(I - P^TDP\right)P^T\,\mathrm{vec}(\mathcal{X})\right\|_2 \leq (1-\sigma_{\min})\left\|P^T\,\mathrm{vec}(\mathcal{X})\right\|_2\\ &= (1-\sigma_{\min})\,\|\mathcal{P}_{\mathbb{T}}(\mathcal{X})\|_F,\end{aligned}$$

which gives that $\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\|_{op} \leq 1 - \sigma_{\min} < 1$.

On the other hand, we prove that $\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}$ is invertible can be derived from $\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\|_{op} < 1$. Provided that $\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\| < 1, \mathcal{I} + \sum_{i=1}^{\infty} (\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}})^i$ is well defined. Notice that, for any $\mathcal{X} \in \mathcal{P}_{\mathbb{T}}$, the following holds:

$$\begin{aligned}
&\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}} \left(\mathcal{I} + \sum_{i=1}^{\infty} (\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}})^i\right)(\mathcal{X}) \\
&= \mathcal{P}_{\mathbb{T}} (\mathcal{I} - \mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}) \left(\mathcal{I} + \sum_{i=1}^{\infty} (\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}})^i\right)(\mathcal{X}) \\
&= \mathcal{P}_{\mathbb{T}}(\mathcal{X}) = \mathcal{X}.
\end{aligned}$$

Similarly, it can be also proven that $\left(\mathcal{I} + \sum_{i=1}^{\infty} (\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}})^i\right) \mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{X}) = \mathcal{X}$. Hence, $\mathcal{I} + \sum_{i=1}^{\infty} (\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}})^i$ is indeed the inverse operator of $\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}$. □

Lemma I.5. *Let $\mathcal{M} \in \mathbb{R}^{m_1 \times \cdots \times m_d}$ with $\operatorname{rank}_t(\mathcal{M}) = r$, and it has the skinny t-SVD $\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\mathrm{T}}$, $\mathcal{P}_{\mathbb{T}}$ is given by $\mathcal{P}_{\mathbb{T}}(\cdot) = \mathcal{U} * \mathcal{U}^{\mathrm{T}} * (\cdot) + (\cdot) * \mathcal{V} * \mathcal{V}^{\mathrm{T}} - \mathcal{U} * \mathcal{U} * (\cdot) * \mathcal{V} * \mathcal{V}^{\mathrm{T}}$, if the operator $\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}$ is invertible, then we have*

$$\|\mathcal{P}_{\mathbb{T}^\perp}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}})^{-1}\|_{op} = \sqrt{\frac{1}{1 - \|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\|_{op}} - 1}.$$

Proof. We shall use again the two notations, vec(.) and D, defined in the proof of Lemma I.4. Let $P \in \mathbb{R}^{m_1 \cdots m_d \times a}$ be a column-wisely orthonormal matrix such that $\operatorname{vec}(\mathcal{P}_{\mathbb{T}}(\mathcal{X})) = PP^T \operatorname{vec}(\mathcal{X}), \forall \mathcal{X}, P^T P = \boldsymbol{I}$. Since $\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}$ is invertible, it follows that $P^T DP$ is positive definite. Denote by $\sigma_{\min}(\cdot)$ the smallest singular value of a matrix. Then we have the following:

$$\begin{aligned}
&\|\mathcal{P}_{\mathbb{T}^\perp}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}})^{-1}\|_{op}^2 \\
&= \sup_{\|\mathcal{X}\|_F = 1} \|\mathcal{P}_{\mathbb{T}^\perp}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}})^{-1}(\mathcal{X})\|_F^2 \\
&= \sup_{\|\operatorname{vec}(\mathcal{X})\|_F = 1} \|(I - PP^T)DPP^T(PP^TDPP^T)^{-1}\operatorname{vec}(\mathcal{X})\|_F^2 \\
&= \|(I - PP^T)DPP^T(PP^TDPP^T)^{-1}\|^2 \\
&= \|(I - PP^T)DP(P^TDP)^{-1}P^T\|^2 \\
&= \|P(P^TDP)^{-1} - I)P^T\| = \|(P^TDP)^{-1} - I)\| \\
&= \frac{1}{\sigma_{\min}(P^TDP)} - 1 = \frac{1}{1 - \|P^T(I - D)P\|} - 1 \\
&= \frac{1}{1 - \|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\|_{op}} - 1.
\end{aligned}$$

□

Lemma I.6. *Suppose that $\mathcal{M} \in \mathbb{R}^{m_1 \times \cdots \times m_d}$ obeys the standard order-d tensor incoherence conditions and $\Omega \subseteq [m_1] \otimes [m_2] \otimes \cdots \otimes [m_d]$, $\mathcal{P}_{\mathbb{T}}$ is given by $\mathcal{P}_{\mathbb{T}}(\cdot) = \mathcal{U} * \mathcal{U}^{\mathrm{T}} * (\cdot) + (\cdot) * \mathcal{V} * \mathcal{V}^{\mathrm{T}} - \mathcal{U} * \mathcal{U} * (\cdot) * \mathcal{V} * \mathcal{V}^{\mathrm{T}}$, if $\rho(\Omega) > 1 - \frac{1}{2\mu r(r_s+1)}$, then the following conditions hold: 1. $\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\| < 1$. 2. There exists a dual certificate $\Lambda \in \mathbb{R}^{m_1 \times \cdots \times m_d}$ such that $\mathcal{P}_{\Omega}(\Lambda) = \Lambda$ and (a) $\|\mathcal{P}_{\mathbb{T}}(\Lambda)\| < 1$. (b)$\mathcal{P}_{\mathbb{T}}(\Lambda) = \mathcal{U} * \mathcal{V}^{\mathrm{T}}$.*

Proof. Since $\rho(\Omega) > 1 - \frac{1}{2\mu r(r_s+1)}$, it follows from Lemma I.2, Lemma I.3 and Lemma I.4 that $\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\|_{op} < \frac{1}{r_s+1} < 1$ and the operator $\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}$ is invertible. Then we define

$$\Lambda = \mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}})^{-1}(\mathcal{U} * \mathcal{V}^{\mathrm{T}}),$$

It can be verified that $\mathcal{P}_{\Omega}(\Lambda) = \Lambda$ and $\mathcal{P}_{\mathbb{T}}(\Lambda) = \mathcal{U} * \mathcal{V}^{\mathrm{T}}$. Moreover, according to Lemma I.1 and Lemma I.5, we have

$$\begin{aligned}
\|\mathcal{P}_{\mathbb{T}^\perp}\Lambda\| &\leq \sqrt{m}\|\mathcal{P}_{\mathbb{T}^\perp}\Lambda\|_F \\
&= \sqrt{m}\|\mathcal{P}_{\mathbb{T}^\perp}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}})^{-1}(\mathcal{U} *_{\mathfrak{L}} \mathcal{V}^{\mathrm{T}})\|_F \\
&\leq \sqrt{m}\|\mathcal{P}_{\mathbb{T}^\perp}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}})^{-1}\|_{op}\|\mathcal{U} * \mathcal{V}^{\mathrm{T}}\|_F \\
&\leq \sqrt{m}\sqrt{\frac{1}{1 - \|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\|_{op}} - 1}\frac{\sqrt{r_s}}{\sqrt{m}}\|\mathcal{U} * \mathcal{V}\| \\
&< \sqrt{r_s}\sqrt{\frac{1}{1 - \frac{1}{r_s+1}} - 1} = 1.
\end{aligned}$$

□

Based on the foreshadowing of Lemma I.6 , we can obtain the proof of Theorem III.3:
Suppose that $\mathcal{K} \in \mathbb{R}^{m_1 \times \cdots \times m_d}$ meets the following conditions:

$$\|\mathcal{P}_{\mathbb{T}^\perp}\mathcal{K}\| = 1, \langle \mathcal{P}_{\mathbb{T}^\perp}\mathcal{K}, \mathcal{P}_{\mathbb{T}^\perp}(\mathcal{X} - \mathcal{M})\rangle = \|\mathcal{P}_{\mathbb{T}^\perp}(\mathcal{X} - \mathcal{M})\|_{\circledast}. \tag{1}$$

Such $\mathcal{K}$ always exists by duality between the tensor nuclear norm and tensor spectral norm. Note that $\mathcal{U} * \mathcal{V}^{\mathrm{T}} + \mathcal{P}_{\mathbb{T}^\perp}\mathcal{K}$ is a subgradient of $\partial\|\mathcal{M}\|_{\circledast}$, according to the definition of subgradient, we have

$$\|\mathcal{X}\|_{\circledast} - \|\mathcal{M}\|_{\circledast} \geq \left\langle \mathcal{U} * \mathcal{V}^{\mathrm{T}} + \mathcal{P}_{\mathbb{T}^\perp}\mathcal{K}, \mathcal{X} - \mathcal{M}\right\rangle. \tag{2}$$

It follows from $\rho(\Omega) > 1 - \frac{1}{2\mu r(r_s+1)}$ and Lemma I.6 that $\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\| < 1$ and there exists a dual certificate $\Lambda \in \mathbb{R}^{m_1 \times \cdots \times m_d}$ such that

$$\mathcal{P}_{\Omega}(\Lambda) = \Lambda, \|\mathcal{P}_{\mathbb{T}}(\Lambda)\| < 1, \ and\ \mathcal{P}_{\Omega}(\Lambda) = \mathcal{U} * \mathcal{V}^{\mathrm{T}}. \tag{3}$$

Now utilizing (1),(2)and (3), we have

$$\begin{aligned}
&\|\mathcal{X}\|_{\circledast} - \|\mathcal{M}\|_{\circledast} \geq \left\langle \mathcal{U} * \mathcal{V}^{\mathrm{T}} + \mathcal{P}_{\mathbb{T}^\perp}\mathcal{K}, \mathcal{X} - \mathcal{M}\right\rangle \\
&= \left\langle \mathcal{U} * \mathcal{V}^{\mathrm{T}} + \mathcal{P}_{\mathbb{T}^\perp}\mathcal{K} - \Lambda, \mathcal{X} - \mathcal{M}\right\rangle \\
&= \|\mathcal{P}_{\mathbb{T}^\perp}(\mathcal{X} - \mathcal{M})\|_{\circledast} + \left\langle \mathcal{U} * \mathcal{V}^{\mathrm{T}} - \Lambda, \mathcal{X} - \mathcal{M}\right\rangle \\
&= \|\mathcal{P}_{\mathbb{T}^\perp}(\mathcal{X} - \mathcal{M})\|_{\circledast} - \langle \mathcal{P}_{\mathbb{T}^\perp}\Lambda, \mathcal{P}_{\mathbb{T}^\perp}(\mathcal{X} - \mathcal{M})\rangle \\
&\geq \|\mathcal{P}_{\mathbb{T}^\perp}(\mathcal{X} - \mathcal{M})\|_{\circledast} - ||\mathcal{P}_{\mathbb{T}^\perp}\Lambda||||\mathcal{P}_{\mathbb{T}^\perp}(\mathcal{X} - \mathcal{M})||_{\circledast} \\
&= (1 - ||\mathcal{P}_{\mathbb{T}^\perp}\Lambda||)\|\mathcal{P}_{\mathbb{T}^\perp}(\mathcal{X} - \mathcal{M})\|_{\circledast}.
\end{aligned}$$

When $\|\mathcal{P}_{\mathbb{T}^\perp}(\mathcal{X} - \mathcal{M})\|_{\circledast} \neq 0$, $\|\mathcal{X}\|_{\circledast} - \|\mathcal{M}\|_{\circledast} > 0$. When $\|\mathcal{P}_{\mathbb{T}^\perp}(\mathcal{X} - \mathcal{M})\|_{\circledast} = 0$, $\mathcal{P}_{\mathbb{T}^\perp}(\mathcal{X} - \mathcal{M}) = 0$, then $\mathcal{P}_{\mathbb{T}}(\mathcal{X} - \mathcal{M}) = \mathcal{X} - \mathcal{M}$, thus $\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}(\mathcal{X} - \mathcal{M}) = \mathcal{P}_{\Omega^\perp}(\mathcal{X} - \mathcal{M}) = \mathcal{X} - \mathcal{M}$. Therefore, $(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}})(\mathcal{X} - \mathcal{M}) = (\mathcal{X} - \mathcal{M})$, which means $||\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}||_{op} \geq 1$. This contradicts condition $||\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}||_{op} < 1$.

II. Proof of Proposition III.1

Under the assumption of Bernoulli sampling $\Omega \sim \mathrm{Ber}(p)$,

$$\mathbb{P}(\frac{x_1 + \cdots + x_n}{n} \leq p - a) \leq \exp\left(-2a^2 n\right)$$

holds for each horizontal/lateral mask sub-tensor of $\mathcal{M} \in \mathbb{R}^{m_1 \times \cdots \times m_d}$ according to the Hoefding inequality, where $n = m_1 m_3 \cdots m_d$ for each lateral mask sub-tensor and $n = m_2 m_3 \cdots m_d$ for each horizontal mask sub-tensor. Assume that the sampling of each sub-tensor is independent, then

$$\mathbb{P}(\rho(\Omega) \leq p - a) \leq \exp\left(-4a^2 m_0\right),$$

where $m_0 = m_1 \times \times m_d$. Setting $a = p - 1 + 1/2\mu r(r_s + 1)$ implies

$$\mathbb{P}\left(\rho(\Omega) \leq 1 - \frac{1}{2ur\left(r_s + 1\right)}\right) \leq \exp\left(-4a^2 m\right),$$

which in turn means that

$$\mathbb{P}\left(\rho(\Omega) > 1 - \frac{1}{2ur\left(r_s + 1\right)}\right) \geqslant 1 - \exp\left(-4a^2 m_0\right)$$

III. Proof of Lemma IV.1 and Lemma IV.2

Let $\widetilde{\mathcal{M}} \in \mathbb{R}^{t \times 1 \times n_1 \times \cdots \times n_p}$ be a shape variant of $\mathcal{M} \in \mathbb{R}^{t \times n_1 \times \cdots \times n_p}$(i.e.,$\widetilde{\mathcal{M}} = reshape(\mathcal{M}, t, 1, n_1, ..., n_p)$). In other words, $[\mathcal{T}_k(\mathcal{M})]_{(:,j+1,:,\cdots,:)}$ is given by $\mathcal{S}^j(\widetilde{\mathcal{M}})$, i.e.,

$$\mathcal{T}_k(\mathcal{M}) = \left[\widetilde{\mathcal{M}}, \mathcal{S}^1(\widetilde{\mathcal{M}}), \mathcal{S}^2(\widetilde{\mathcal{M}}), \cdots, \mathcal{S}^{k-1}(\widetilde{\mathcal{M}})\right].$$

where $\mathcal{S}$ is an operator that circularly shifts the elements of a tensor by one position along the first dimension; namely,

$$\mathcal{S}(\widetilde{\mathcal{M}}) = \begin{bmatrix} \mathcal{M}^t \\ \mathcal{M}^1 \\ \vdots \\ \mathcal{M}^{t-1} \end{bmatrix}, \widetilde{\mathcal{M}} = \begin{bmatrix} \mathcal{M}^1 \\ \mathcal{M}^2 \\ \vdots \\ \mathcal{M}^t \end{bmatrix}.$$

Proof of Lemma IV.1:

Proof. According to the definition of the temporal convolution tensor, we have

$$\begin{aligned}\mathcal{T}_k(\mathcal{M}) &= \begin{bmatrix} \mathcal{M}^1 & \mathcal{M}^t & \cdots & \mathcal{M}^{t-k+2} \\ \mathcal{M}^2 & \mathcal{M}^1 & \cdots & \mathcal{M}^{t-k+3} \\ \vdots & \vdots & \vdots & \vdots \\ \mathcal{M}^t & \mathcal{M}^{t-1} & \cdots & \mathcal{M}^{t-k+1} \end{bmatrix} \\ &= \left[\widetilde{\mathcal{M}}, \mathcal{S}^1(\widetilde{\mathcal{M}}), \mathcal{S}^2(\widetilde{\mathcal{M}}), \cdots, \mathcal{S}^{k-1}(\widetilde{\mathcal{M}})\right].\end{aligned}$$

where $\|\widetilde{\mathcal{M}} - \mathcal{S}^j(\widetilde{\mathcal{M}})\|_F \le j\|\widetilde{\mathcal{M}} - \mathcal{S}(\widetilde{\mathcal{M}})\|_F \le j\sqrt{t}\eta(\mathcal{M})$. Decompose $\mathcal{T}_k(\mathcal{M})$ into the concatenation of r subtensors, namely $\mathcal{T}_k(\mathcal{M}) = [\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_r]$, such that $\text{size}(\mathcal{A}_i, 2) = b_i$ with $1 \le b_i \le \lceil \frac{k}{r} \rceil$ and $\sum_{i=1}^{r} b_i = k$. For $\mathcal{A}_i \in \mathbb{R}^{t\times b_i \times n_1 \times \cdots \times n_p}$, construct a rank-1 tensor $\hat{\mathcal{A}}_i \in \mathbb{R}^{t\times b_i \times n_1 \times \cdots \times n_p}$ by repeating $\mathcal{A}_i(:, 1, :, \cdots, :)$ for b_i times. Then we have

$$\begin{aligned}\left\|\mathcal{A}_i - \hat{\mathcal{A}}_i\right\|_F &= \sqrt{\sum_{j=0}^{b_i-1} \|\widetilde{\mathcal{M}} - \mathcal{S}^j(\widetilde{\mathcal{M}})\|_F^2} \le \sqrt{t}\eta(\mathcal{M})\sqrt{\sum_{j=0}^{b_i-1} j^2} \\ &= \sqrt{t}\eta(\mathcal{M})\sqrt{\frac{b_i(b_i+1)(2b_i-1)}{6}} \le \sqrt{t}\eta(\mathcal{M}) b_i \sqrt{\frac{b_i+1}{3}}\end{aligned}$$

We consider $\mathcal{Y} = \left[\hat{\mathcal{A}}_1, \hat{\mathcal{A}}_2, \cdots, \hat{\mathcal{A}}_r\right]$. Obviously, $\text{rank}_t(\mathcal{Y}) \le r$. Thus,

$$\begin{aligned}\varepsilon_r(\mathcal{T}_k(\mathcal{M})) &\le \|\mathcal{T}_k(\mathcal{M}) - \mathcal{Y}\|_F = \sqrt{\sum_{i=1}^{r} \|\mathcal{A}_i - \hat{\mathcal{A}}_i\|_F^2} \\ &= \sqrt{\sum_{i=1}^{r} b_i^2 \frac{b_i+1}{3}} \sqrt{t}\eta(\mathcal{M}) \le \sqrt{\frac{t(k+r)}{3}} \left\lceil \frac{k}{r} \right\rceil \eta(\mathcal{M})\end{aligned}$$

□

Proof of Lemma IV.2:

Proof. Similarly, set $a = \lceil \frac{k}{\tau} \rceil$, then decompose $\mathcal{T}_k(\mathcal{M})$ into the concatenation of a subtensors, namely $\mathcal{T}_k(\mathcal{M}) = [\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_a]$, such that $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_{a-1} \in \mathbb{R}^{t\times\tau\times n_1\times\cdots\times n_p}$ and $\mathcal{A}_a \in \mathbb{R}^{t\times(k-(a-1)\tau)\times n_1\times\cdots\times n_p}$. We consider $\mathcal{Y} = [\mathcal{A}_1, \mathcal{A}_1, ..., \mathcal{A}_1^{'}]$, where $\mathcal{A}_1^{'} = [\mathcal{A}_1](:, 1 : k - (a-1)\tau, :, ..., :)$. Since $\text{rank}_t(\mathcal{Y}) \le r$,

$$\begin{aligned}\varepsilon_\tau\left(\mathcal{T}_k(\mathcal{M})\right) &\le \|\mathcal{T}_k(\mathcal{M}) - \mathcal{Y}\|_F \le \sqrt{\sum_{i=1}^{a} \|\mathcal{A}_i - \mathcal{A}_1\|_F^2} \\ &\le (a-1)\tau t \beta_\tau(\mathcal{M}) = \tau t (\left\lceil \frac{k}{\tau} \right\rceil - 1)\beta_\tau(\mathcal{M}).\end{aligned}$$

□

IV. Proof of Theorem IV.1

By applying deterministic tensor completion recovery theory (Theorem III.1), we can then directly prove that $\mathcal{T}_k(\mathcal{M})$ is the unique solution to the convex model

$$\min_{\mathcal{Y}\in\mathbb{R}^{t\times k\times n_1\times\cdots\times n_p}} \|\mathcal{Y}\|_{\circledast}, \quad \text{s.t. } \mathcal{P}_{\Omega_\mathcal{T}}(\mathcal{Y}) = \mathcal{P}_{\Omega_\mathcal{T}}(\mathcal{T}_k(\mathcal{M}))$$

when

$$\rho(\Omega_\mathcal{T}) > 1 - 1/(2\mu_\mathcal{T} r_\mathcal{T}((r_s)_\mathcal{T} + 1)).$$

We observe that $|(\Omega_\mathcal{T})_{i_k}| \ge (k-h)n_1\cdots n_p$ and $|(\Omega_\mathcal{T})_{i_t}| = (t-h)n_1\cdots n_p$, then $\mathcal{M}$ is the unique solution to the proposed TCTNN model when h satisfies $h < k/(2\mu_\mathcal{T} r_\mathcal{T}((r_s)_\mathcal{T} + 1))$.

V. PROOF OF THEOREM V.1

Now we shall give the detailed proof of Theorem V.1 presented in the main text. To this end, we need the following two lemmas.

Lemma V.1. *The sequence of dual variables $\mathcal{N}^j$ in Algorithm 1 are bounded.*

Proof. According to the optimality principle, we have

$$\mathbf{0} \in \partial_{\mathcal{Y}} L(\mathcal{X}^{\ell+1}, \mathcal{Y}^{\ell+1}, \mathcal{N}^{\ell}) \ and \ \mathbf{0} \in \partial_{\mathcal{X}} L(\mathcal{X}^{\ell+1}, \mathcal{Y}^{\ell+1}, \mathcal{N}^{\ell}),$$

which leads to

$$\begin{aligned} &\mathbf{0} \in \partial \|\mathcal{Y}^{\ell+1}\|_{\circledast} + \mu_\ell(\mathcal{Y} - \mathcal{T}_k(\mathcal{X}) + \mathcal{N}^{\ell}/\mu_\ell), \\ &\mathcal{X}^{\ell+1} = \mathcal{P}_\Omega(\mathcal{M}) + \mathcal{P}_{\Omega^\perp}(\mathcal{T}_k^{-1}(\mathcal{Y}^{\ell+1} + \mathcal{N}_\ell/\mu^\ell)). \end{aligned}$$

Combining this with the update criterion $\mathcal{N}^{\ell+1} = \mathcal{N}^{\ell} + \mu_\ell(\mathcal{Y}^{\ell+1} - \mathcal{T}_k(\mathcal{X}^{\ell+1}))$ and the constraint equation $\mathcal{P}_\Omega(\mathcal{X}^{\ell+1}) = \mathcal{P}_\Omega(\mathcal{M})$ in Algorithm 1, we have

$$\begin{aligned} &\mathcal{N}^{\ell+1} \in \partial \left(\|\mathcal{Y}^{\ell+1}\|_{\circledast} \right), \\ &\mathcal{P}_{\Omega^\perp}(\mathcal{T}_k^*(\mathcal{N}^{\ell+1})) = 0. \end{aligned}$$

where $\mathcal{T}_k^*(\cdot) = k\mathcal{T}_k^{-1}(\cdot)$ is the Hermitian adjoint of $\mathcal{T}_k$. Note the fact that the dual norm of tensor nulclear norm $\|\cdot\|_{\circledast}$ is tensor spectral norm $\|\cdot\|$, and $\mathcal{N}^{\ell+1} \in \partial \left(\|\mathcal{Y}^{\ell+1}\|_{\circledast} \right)$. Thus, Following Theorem 4 in [5], we get that $\|\mathcal{N}^{\ell+1}\|$ is bounded. Considering the relationship between $\|\mathcal{N}^{\ell+1}\|$ and $\|\mathcal{N}^{\ell+1}\|_{\circledast}$:

$$\begin{aligned} \|\mathcal{N}^{\ell+1}\|_F &= \frac{1}{\sqrt{n}} \|\mathrm{bdiag}(\bar{\mathcal{N}}^{\ell+1}\|_F) \\ &\le \sqrt{r} \|\mathrm{bdiag}(\bar{\mathcal{N}}^{\ell+1})\| = \sqrt{r}\|\mathcal{N}^{\ell+1}\|, \end{aligned}$$

where $n = n_1 \times \cdots \times n_p$ we can conclude that $\|\mathcal{N}^{\ell+1}\|_F$ is bounded. □

Lemma V.2. *The accumulation point $\left(\mathcal{Y}^{\ell}, \mathcal{X}^{\ell}, \mathcal{N}^{\ell}\right)$ generated by Algorithm 1 is a feasible solution of the TCTNN model.*

Proof. Based on the general ADMM principle, we have

$$\|\mathcal{N}^{\ell+1} - \mathcal{N}^{\ell}\|_F = \mu_\ell \|\mathcal{Y}^{\ell+1} - \mathcal{T}_k(\mathcal{X}^{\ell+1})\|_F$$

Since $\{\mu_\ell\}$ is an increasing sequence and $\lim_{\ell\to+\infty} \mu_\ell = +\infty$, and according to Lemma V.1 , we have

$$\lim_{\ell\to+\infty} \|\mathcal{Y}^{\ell+1} - \mathcal{T}_k(\mathcal{X}^{\ell+1})\|_F = 0,$$

which means $\lim_{\ell\to+\infty} \mathcal{Y}^{\ell+1} = \mathcal{T}_k(\mathcal{X}^{\ell+1})$, and the constraint always holds in the iteration, so the proof $\mathcal{P}_\Omega(\mathcal{X}^{\ell+1}) = \mathcal{P}_\Omega(\mathcal{M})$ is complete. □

With the above lemmas, we next give the proof of Theorem V.1.

Proof. Suppose $(\mathcal{X}^*, \mathcal{Y}^*)$ is an optimal solution of the TCTNN model, and $\mathcal{N}^*$ is the optimal solution of its dual model, it thus get that $(\mathcal{X}^*, \mathcal{Y}^*, \mathcal{N}^*)$ forms the saddle point of the Lagrangian function (21). Then it is obvious that $p^* = \|\mathcal{Y}^*\|_{\circledast}$ gets the minimum value and the following equation holds

$$\mathcal{Y}^* = \mathcal{T}_k(\mathcal{X}^*), \ \mathcal{P}_\Omega(\mathcal{X}^*) = \mathcal{P}_\Omega(\mathcal{M}).$$

Due to the definition of subgradient and $\mathcal{N}^{\ell+1} \in \partial \left(\|\mathcal{Y}^{\ell+1}\|_{\circledast} \right)$, we have

$$\begin{aligned} \|\mathcal{Y}^{\ell}\|_{\circledast} \le &\|\mathcal{Y}^*\|_{\circledast} + \langle \mathcal{N}^{\ell}, \mathcal{Y}^{\ell} - \mathcal{Y}^* \rangle \\ \overset{\ell\to+\infty}{=\!=\!=} &\|\mathcal{Y}^*\|_{\circledast} + \langle \mathcal{N}^{\ell}, \mathcal{T}_k(\mathcal{X}^{\ell} - \mathcal{X}^*) \rangle \\ = &\|\mathcal{Y}^*\|_{\circledast} + \langle \mathcal{T}_k^*\mathcal{N}^{\ell}, \mathcal{X}^{\ell} - \mathcal{X}^* \rangle \\ = &\|\mathcal{Y}^*\|_{\circledast} + \langle \mathcal{P}_{\Omega^\perp}(\mathcal{T}_k^*\mathcal{N}^{\ell}), \mathcal{X}^{\ell} - \mathcal{X}^* \rangle \\ &+ \langle \mathcal{T}_k^*\mathcal{N}^{\ell}, \mathcal{P}_\Omega(\mathcal{X}^{\ell} - \mathcal{X}^*) \rangle \\ = &\|\mathcal{Y}^*\|_{\circledast}, \end{aligned}$$

the last equality holds because $\mathcal{P}_{\Omega^\perp}(\mathcal{T}_k^*(\mathcal{N}^{\ell+1})) = 0$ and $\mathcal{P}_\Omega(\mathcal{X}^{\ell} - \mathcal{X}^*) = \mathcal{P}_\Omega(\mathcal{M} - \mathcal{M}) = 0$. Thus,

$$\lim_{\ell\to+\infty} \|\mathcal{Y}^{\ell}\|_{\circledast} = \|\mathcal{Y}^*\|_{\circledast}.$$

This completes the proof. □

VI. Other discussions of the experiment

A. Multi-sample time series analysis

The previous experiments were all based on predictions in small sample scenarios, and did not consider the case of multiple samples. We sample the extended version of the Abilene dataset, which was recorded from 12:00 AM to 5:00 PM on March 1, 2004, with a temporal resolution of 5 minutes. We structured this data into a tensor with dimensions 204 × 12 × 12, and then set the prediction domain to 10, 20, 30, and 40 for testing. The experimental results are plotted in Figure 1. The test results show that the prediction accuracy of the TCTNN model we proposed is higher than that of the CNNM model in the multi-sample scenario, but not as good as the BTTF model, which may be because the autoregressive characteristics used by BTTF can better model long time series.

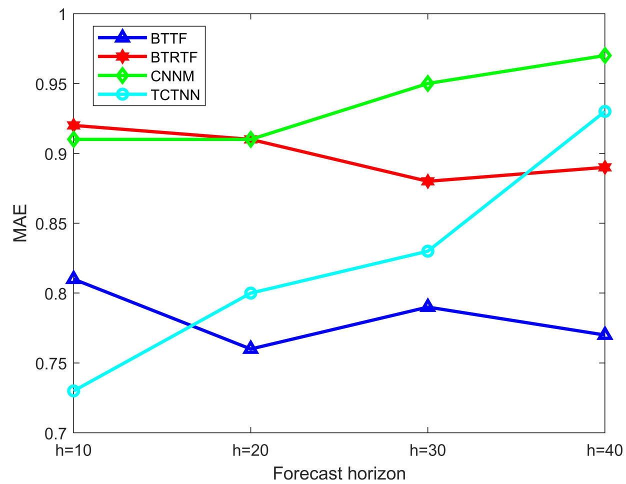

Fig. 1: The predicted MAE results of multi-sample time serie using the TCTNN model and other tensor-based models

B. Application to multivariate time series

TABLE I: Performance comparison (in MAE/RMSE)) of TCTNN and other baseline models for multivariate time series prediction across various forecast horizon scenarios. The forecast horizon (FH) column indicates the respective forecast horizons.

FH	LRMC	TRMF	CNNM	TCTNN
h=2	41.74/42.88	1.66/2.12	1.81/2.36	**1.40/1.90**
h=4	41.61/42.75	1.76/2.24	1.92/2.52	**1.51/2.04**
h=6	41.57/42.70	1.68/2.14	1.90/2.50	**1.48/2.00**
h=8	41.49/42.61	1.66/2.15	1.93/2.55	**1.50/2.04**
h=10	41.49/42.62	1.84/2.35	1.93/2.56	**1.58/2.16**

Although the focus of this work is on multidimensional time series prediction, our TCTNN model and its underlying theory are applicable to multivariate time series as well. By setting $p = 1$ in the TCTNN model, we can effectively achieve multivariate time series forecasting. We select a publicly available Guangzhou traffic dataset [1] for testing. This dataset records the speed of 214 road sections during the first half of August 2, 2016 with a 10-minute resolution in Guangzhou, China. Consequently, the size of the urban traffic data matrix for Guangzhou is 214×72. The RMSE results of the TCTNN and CNNM models applied to the Guangzhou traffic dataset are illustrated in TableI. The Table shows that compared with the CNNM model, the prediction results of the TCTNN model have a significant lead in different prediction domains. We conduct prediction experiments on the *Low-rank matrix completion* (LRMC) model [6], *temporal regularized matrix factorization* (TRMF) model [7], CNNM model and TCTNN model on the Guangzhou dataset, and the experimental results are summarized in Table 2. It is obvious that the TCTNN model has the best prediction accuracy, followed by the TRMF model. The LRMC prediction result is all 0, which is congruent with the experimental results of the SNN/TNN model for multi-dimensional time series prediction. To visualize the prediction results, we present a plot of the time-varying outputs from sensors 1 , utilizing a forecast horizon of h=10. This visualization is depicted in Figure 2 From the curve trend of Figure 2, we can see that the prediction results of the TCTNN model are closer to the actual values. This demonstrates that the TCTNN model exhibits superior capabilities in forecasting multivariate time series compared to other matrix-based models.

[1] https://doi.org/10.5281/zenodo.1205229

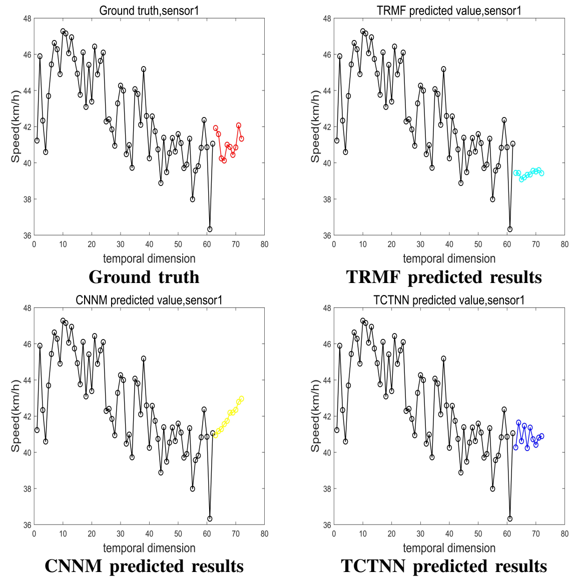

Fig. 2: The prediction results of the CNNM model and other matrix-based models in sensor 1; the red curve represents the true values to be predicted, the cyan curve represents the prediction results of the TRMF model, the yellow curve represents the prediction results of the CNNM model, and the blue curve represents the prediction results of the TCTNN model.

References

[1] W. Qin, H. Wang, F. Zhang, J. Wang, X. Luo, and T. Huang, "Low-rank high-order tensor completion with applications in visual data," *IEEE Trans. Image Process.*, vol. 31, pp. 2433–2448, 2022.

[2] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[3] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[4] T. G. Kolda, "Multilinear operators for higher-order decompositions." Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA, Tech. Rep., 2006.

[5] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[6] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[7] H.-F. Yu, N. Rao, and I. S. Dhillon, "Temporal regularized matrix factorization for high-dimensional time series prediction," in *Proc. Advances Neural Inf. Process. Syst.*, vol. 29, 2016.