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

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

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*

$$\partial\|\mathcal{M}\|_{\circledast} = \left\{\mathcal{U} * \mathcal{V}^{\mathrm{T}} + \mathcal{W} \mid \mathcal{U}^{\mathrm{T}} * \mathcal{W} = 0, \mathcal{W} * \mathcal{V} = 0, \|\mathcal{W}\| \leq 1\right\}$$

.

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*

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

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

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

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

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

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

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

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

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

which implies that

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

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

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

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*

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

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

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

the second inequality holds because

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

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

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

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

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

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 $\mathcal{P}_{\mathbb{V}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{V}}(\mathcal{X}(i_1,:,\cdots,:)) = (\mathcal{P}_{\mathbb{V}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{V}}\mathcal{X})(i_1,:,\cdots,:)$. $\square$

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,

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

$\square$

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

$$D = \mathrm{diag}\left(\delta_{1\ldots1}, \delta_{2\ldots1}, \ldots, \delta_{i_1\ldots i_d}, \ldots, \delta_{m_1\ldots m_d}\right) \in \mathbb{R}^{m_1\cdots m_d\times m_1\cdots m_d}.$$

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 $[\mathrm{vec}\left(\mathcal{P}_{\mathbb{T}}\left(\mathfrak{e}_{1\cdots1}\right)\right), \mathrm{vec}\left(\mathcal{P}_{\mathbb{T}}\left(\mathfrak{e}_{2\cdots1}\right)\right), \ldots, \mathrm{vec}\left(\mathcal{P}_{\mathbb{T}}\left(\mathfrak{e}_{m_1\cdots m_d}\right)\right)] = PP^T$. 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 $\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\| < 1, \mathcal{I} + \sum_{i=1}^{\infty} \left(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\right)^i$ is well defined. Notice that, for any $\mathcal{X} \in \mathcal{P}_{\mathbb{T}}$, the following holds:

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

Similarly, it can be also proven that $\left(\mathcal{I} + \sum_{i=1}^{\infty} \left(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\right)^i\right) \mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathcal{X}) = \mathcal{X}$. Hence, $\mathcal{I} + \sum_{i=1}^{\infty} \left(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\right)^i$ 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 $\operatorname{vec}\left(\mathcal{P}_{\mathbb{T}}(\mathcal{X})\right) = PP^T \operatorname{vec}(\mathcal{X}), \forall \mathcal{X}, P^T P = \boldsymbol{I}$. Since $\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}$ is invertible, it follows that $P^T D P$ 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

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

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

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

where $\|\widetilde{\mathcal{M}} - \mathcal{S}^j(\widetilde{\mathcal{M}})\|_F \leq j\|\widetilde{\mathcal{M}} - \mathcal{S}(\widetilde{\mathcal{M}})\|_F \leq j\sqrt{t}\eta(\mathcal{M})$. 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 $1 \leq b_i \leq \lceil \frac{k}{r} \rceil$ 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

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

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

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

$\square$

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 $\text{rank}_t(\mathcal{Y}) \leq r$,

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

$\square$

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 $|(\Omega_\mathcal{T})_{i_k}| \geq (k-h)n_1\cdots n_p$ 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

$$\mathbf{0} \in \partial \|\mathcal{Y}^{\ell+1}\|_{\circledast} + \mu_\ell(\mathcal{Y}^{\ell+1} - \mathcal{T}_k(\mathcal{X}^{\ell+1}) + \mathcal{N}^{\ell}/\mu_\ell) \tag{4}$$

and

$$\mathcal{X}^{\ell+1} = \mathcal{P}_{\Omega}(\mathcal{M}) + \mathcal{P}_{\Omega^\perp}(\mathcal{T}_k^{-1}(\mathcal{Y}^{\ell+1} + \mathcal{N}^{\ell}/\mu^{\ell})). \tag{5}$$

By combining equation 4 with the update rule

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

we obtain

$$\mathcal{N}^{\ell+1} \in \partial \|\mathcal{Y}^{\ell+1}\|_{\circledast}.$$

Similarly, by combining equation 5 with the constraint

$$\mathcal{P}_{\Omega}(\mathcal{X}^{\ell+1}) = \mathcal{P}_{\Omega}(\mathcal{M})$$

in Algorithm 1, we obtain

$$\mathcal{P}_{\Omega^\perp}\left(\mathcal{T}_k^*(\mathcal{N}^{\ell+1})\right) = 0,$$

where $\mathcal{T}_k^*(\cdot) = k\mathcal{T}_k^{-1}(\cdot)$ denotes 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 $\|\mathcal{N}^{\ell+1}\|_{\circledast}$:

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

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

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

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. Proof of Lemma VI.1

Proof. Consider the transformed tensor

$$\bar{\mathcal{X}} := \mathcal{F}_{d-2}(\mathcal{X}) = \mathcal{X} \times_3 \mathrm{F}_{m_3} \times_4 \cdots \times_d \mathrm{F}_{m_d},$$

where F_{m_j} is the DFT matrix along mode j. For fixed $(p_3, \ldots, p_d)$ we have

$$\bar{\mathcal{X}}(\cdot, \cdot, p_3, \ldots, p_d) = \sum_{i_3=1}^{m_3} \cdots \sum_{i_d=1}^{m_d} \left(\prod_{j=3}^{d} \mathrm{F}_{m_j}(i_j, p_j) \right) \mathcal{X}(\cdot, \cdot, i_3, \ldots, i_d).$$

Thus each transformed frontal slice is a linear combination of circulant slices, and hence remains circulant. Every circulant matrix admits a diagonalization

$$\mathcal{X}(\cdot, \cdot, i_3, \ldots, i_d) = \mathrm{F}_{m_1} \Lambda_{i_3,\ldots,i_d} (\mathrm{F}_{m_2})^H,$$

where $\Lambda_{i_3,\ldots,i_d}$ is diagonal with at most r nonzero entries at the first r diagonal elements. Substituting yields

$$\bar{\mathcal{X}}(\cdot, \cdot, p_3, \ldots, p_d) = \mathrm{F}_{m_1} \left[\sum_{i_3=1}^{m_3} \cdots \sum_{i_d=1}^{m_d} \left(\prod_{j=3}^{d} \mathrm{F}_{m_j}(i_j, p_j) \right) \Lambda_{i_3,\ldots,i_d} \right] (\mathrm{F}_{m_2})^H.$$

Define

$$\bar{\Lambda}_{p_3,\ldots,p_d} := \sum_{i_3=1}^{m_3} \cdots \sum_{i_d=1}^{m_d} \left(\prod_{j=3}^{d} \mathrm{F}_{m_j}(i_j, p_j) \right) \Lambda_{i_3,\ldots,i_d}.$$

Since each $\Lambda_{i_3,\ldots,i_d}$ has at most r nonzeros at the first r diagonal elements, $\bar{\Lambda}_{p_3,\ldots,p_d}$ also has at most r nonzeros at the first r diagonal elements, implying

$$\mathrm{rank}(\bar{\mathcal{X}}(\cdot, \cdot, p_3, \ldots, p_d)) \leq r.$$

By definition, the t-SVD rank of $\mathcal{X}$ is the maximum rank among all frontal slices of $\bar{\mathcal{X}}$. Therefore,

$$\mathrm{rank}_t(\mathcal{X}) = \max_{p_3,\ldots,p_d} \mathrm{rank}(\bar{\mathcal{X}}(\cdot, \cdot, p_3, \ldots, p_d)) \leq r.$$

This completes the proof. □

VII. Additional discussions on experiments

A. Multi-sample time series analysis

The previous experiments are all based on predictions in small sample scenarios, and do not consider the case of multiple samples. We sample the extended version of the NYC taxi dataset, which is recorded from May 1st to May 18th, 2018, with a temporal resolution of 1 hour. We structure this data into a tensor with dimensions 216 × 30 × 30, and then set the prediction domain to 2, 4, 6, 8, and 10 for testing. To better analyze our model's capability for multi-sample forecasting, we benchmark it against four classical predictive models: *Double Exponential Smoothing* (DES) [6], *Holt–Winters Exponential Smoothing* (HWES) [7], *Autoregressive Integrated Moving Average* (ARIMA) [8], and *Vector Autoregression* (VAR) [9]. The first two are exponential smoothing methods, while the latter two are autoregressive models. The experimental results are shown in Table I. Among all methods, BTTF achieves the highest prediction accuracy, as it effectively learns parameters from sufficient samples while fully exploiting the spatiotemporal tensor structure. TCTNN ranks second, benefiting from its ability to capture temporal dynamics together with structural spatiotemporal information. In contrast, the four classical baselines fall short due to their limited ability to capture spatiotemporal dependencies. For example, ARIMA can model temporal autoregression and moving-average effects at each location, but its lack of spatiotemporal dependency modeling leads to inferior accuracy.

TABLE I: Performance comparison (in MAE/RMSE) of different methods on the NYC taxi dataset for multi-sample time series analysis. The first two are exponential smoothing methods, while the third and fourth are autoregressive models.

FH	DES	HWES	ARIMA	VAR	BTTF	CNNM	TCTNN
h=2	4.51/7.31	5.24/9.11	4.39/7.08	3.24/5.14	3.21/4.81	3.70/6.07	3.17/4.43
h=4	5.75/10.27	6.64/12.53	5.44/9.53	3.42/5.39	3.34/4.95	4.50/7.35	3.37/4.77
h=6	6.08/11.09	7.20/14.80	5.77/10.31	3.76/6.23	3.45/5.22	5.51/9.27	3.61/5.27
h=8	6.47/12.31	6.00/12.18	6.10/11.32	3.94/6.76	3.33/5.07	5.62/9.96	3.63/5.35
h=10	6.14/11.39	5.40/10.16	5.66/10.30	3.64/6.26	3.44/5.27	5.50/10.12	3.63/5.50

B. Application to multivariate time series

TABLE II: 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	**1.49/1.91**
h=4	41.61/42.75	1.76/2.24	1.92/2.52	**1.61/2.12**
h=6	41.57/42.70	1.68/2.14	1.90/2.50	**1.62/2.12**
h=8	41.49/42.61	1.66/2.15	1.93/2.55	**1.62/2.15**
h=10	41.49/42.62	1.84/2.35	1.93/2.56	**1.69/2.24**

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. We conduct prediction experiments on the *Low-rank matrix completion* (LRMC) model [10], *temporal regularized matrix factorization* (TRMF) model [11], CNNM model and TCTNN model on the Guangzhou dataset, and the experimental results are summarized in Table II. It is obvious that the TCTNN model has the best prediction accuracy, followed by the TRMF model. This demonstrates that the TCTNN model exhibits superior capabilities in forecasting multivariate time series compared to other matrix-based models.

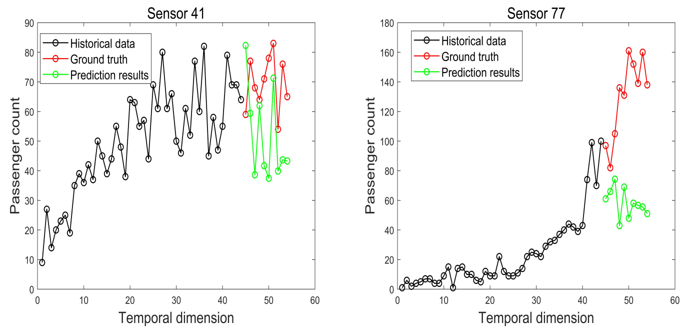
C. Limitations in handling non-periodic and non-smooth data

In real-world data exhibiting abrupt changes, non-stationarity, and highly nonlinear patterns, the temporal convolution low-rankness assumption is often violated due to insufficient smoothness or periodicity. As a result, TCTNN fails to achieve accurate predictions. To verify this, we conduct experiments using the Hangzhou metro passenger flow dataset, which contains data from 80 metro stations in Hangzhou between 6:00 and 15:00, with a temporal resolution of 10 minutes. The prediction results are shown in Table III, where it is evident that TCTNN performs very poorly, even underperforming classical methods such as *Double Exponential Smoothing* (DES) [6] and *Autoregressive Integrated Moving Average* (ARIMA) [8]. To visually illustrate the predictions, we plot the time-varying outputs for sensors 41 and 77 with a prediction horizon of h = 10, as shown in Figure 1. From the curves in Figure 1, it is clear that the TCTNN predictions deviate significantly from the true values. This indicates that TCTNN offers no advantage when forecasting abrupt, non-stationary time series, further confirming that the temporal convolution low-rankness is weak when smoothness and periodicity are not satisfied.

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

TABLE III: Performance comparison (in MAE/RMSE) of different methods on the Hangzhou dataset.

FH	DES	ARIMA	CNNM	TCTNN
h=2	26.16/53.85	24.36/47.12	105.13/189.21	37.70/69.46
h=4	35.92/93.35	33.83/77.72	92.37/171.51	37.11/64.96
h=6	36.27/98.44	34.34/70.52	103.20/191.03	45.95/88.10
h=8	38.75/87.40	39.02/78.06	110.55/212.05	58.06/116.79
h=10	36.70/70.41	36.33/68.14	112.33/221.76	62.32/131.16



(a) Sensor 41 **(b) Sensor 77**

Fig. 1: Prediction results of the TCTNN model for sensors 41 and 77 in the Hangzhou dataset. The black curve represents the historical data, the red curve shows the true values to be predicted, and the green curve indicates the predictions made by the TCTNN model.

D. Comparison with deep learning methods

To assess the differences in predictive performance between TCTNN and deep learning methods, we have conducted comparative experiments on the Pacific surface temperature dataset. The baseline models included *Long Short-Term Memory* (LSTM) [12] and *Convolutional Long Short-Term Memory* (ConvLSTM) [13]. LSTM is a variant of the *Recurrent Neural Network* (RNN) designed to address the vanishing and exploding gradient issues encountered when training on long sequences. By introducing gating mechanisms, LSTM effectively captures long-term dependencies and is widely used for time series forecasting tasks. ConvLSTM combines the advantages of *Convolution Neural Networks* (CNN) and LSTM, enabling it to process both spatial and temporal information. It is particularly well-suited for time series data with spatial structure, such as weather forecasting. The experimental results in Table IV show that ConvLSTM slightly outperforms TCTNN in prediction accuracy, and both outperform LSTM and CNNM. This suggests that leveraging tensor structure information is essential for multi-dimensional time series modeling, which explains why ConvLSTM and TCTNN outperform LSTM and CNNM. Overall, in comparison with deep learning methods, the predictive performance of the TCTNN model remains impressive. Moreover, TCTNN enjoys a theoretical advantage, namely deterministic recovery guarantees, which deep learning methods such as LSTM and ConvLSTM do not provide.

TABLE IV: Performance comparison (in MAE/RMSE) of LSTM, ConvLSTM, CNNM, and TCTNN for different forecast horizons on the Pacific surface temperature dataset.

FH	LSTM	ConvLSTM	CNNM	TCTNN
h=2	1.78/2.25	0.70/0.88	1.67/2.10	0.63/0.83
h=4	2.01/2.59	0.82/1.05	1.70/2.21	0.84/1.13
h=6	2.18/2.83	0.85/1.16	1.95/2.46	1.07/1.39
h=8	2.20/2.85	0.99/1.38	2.23/2.79	1.28/1.65
h=10	2.22/2.85	1.21/1.59	2.29/2.90	1.34/1.70

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] C. C. Holt, “Forecasting seasonals and trends by exponentially weighted moving averages,” Tech. Rep., 1957.

[7] P. R. Winters, “Forecasting sales by exponentially weighted moving averages,” *Manag. Sci.*, vol. 6, no. 3, pp. 324–342, 1960.

[8] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th ed. Wiley, 2008.

[9] C. A. Sims, “Macroeconomics and reality,” *Econometrica*, pp. 1–48, 1980.

[10] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.
[11] 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.
[12] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[13] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, “Convolutional lstm network: A machine learning approach for precipitation nowcasting,” *Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.