Supplementary Material of "Fully-Connected Tensor Network Decomposition and Its Application to Higher-Order Tensor Completion"

Yu-Bang Zheng, Ting-Zhu Huang, Xi-Le Zhao, Qibin Zhao, and Tai-Xiang Jiang

1 Proofs of Theorems

Since Theorem 1 can be easily obtained by Definitions 1, 2, and 3, we only prove Theorems 2, 3, 4, and 5.

Theorem 2 (Transpositional Invariance) *Supposing that an Nth-order tensor $\mathcal{X}$ has the following FCTN decomposition: $\mathcal{X} = \text{FCTN}(\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_N)$. Then, its vector $\mathbf{n}$-based generalized tensor transposition $\vec{\mathcal{X}}^{\mathbf{n}}$ can be expressed as $\vec{\mathcal{X}}^{\mathbf{n}} = \text{FCTN}(\vec{\mathcal{G}}^{\mathbf{n}}_{n_1}, \vec{\mathcal{G}}^{\mathbf{n}}_{n_2}, \cdots, \vec{\mathcal{G}}^{\mathbf{n}}_{n_N})$, where $\mathbf{n} = (n_1, n_2, \cdots, n_N)$ is a reordering of the vector $(1, 2, \cdots, N)$.*

Proof. Supposing that $k_1, k_2 \in \{1, 2, \cdots, N\}$ ($k_1 < k_2$) and $\hat{\mathbf{n}} = (1, 2, \cdots, k_1-1, k_2, k_1+1, \cdots, k_2-1, k_1, k_2+1, \cdots, N)$, we have

$$
\begin{aligned}
&\vec{\mathcal{X}}^{\hat{\mathbf{n}}}(i_1, i_2, \cdots, i_{k_1-1}, i_{k_2}, i_{k_1+1} \cdots, i_{k_2-1}, i_{k_1}, i_{k_2+1}, \cdots, i_N) \\
&= \mathcal{X}(i_1, i_2, \cdots, i_{k_1-1}, i_{k_1}, i_{k_1+1} \cdots, i_{k_2-1}, i_{k_2}, i_{k_2+1}, \cdots, i_N) \\
&= \sum_{r_{1,2}=1}^{R_{1,2}} \sum_{r_{1,3}=1}^{R_{1,3}} \cdots \sum_{r_{1,N}=1}^{R_{1,N}} \sum_{r_{2,3}=1}^{R_{2,3}} \cdots \sum_{r_{2,N}=1}^{R_{2,N}} \cdots \sum_{r_{k_1,k_1+1}=1}^{R_{k_1,k_1+1}} \cdots \sum_{r_{k_1,k_2}=1}^{R_{k_1,k_2}} \cdots \sum_{r_{k_1,N}=1}^{R_{k_1,N}} \cdots \sum_{r_{k_2,k_2+1}=1}^{R_{k_2,k_2+1}} \cdots \sum_{r_{k_2,N}=1}^{R_{k_2,N}} \cdots \sum_{r_{N-1,N}=1}^{R_{N-1,N}} \\
&\quad \big\{\mathcal{G}_1(i_1, r_{1,2}, r_{1,3}, \cdots, r_{1,N}) \\
&\quad \mathcal{G}_2(r_{1,2}, i_2, r_{2,3}, \cdots, r_{2,N}) \cdots \\
&\quad \mathcal{G}_{k_1}(r_{1,k_1}, r_{2,k_1}, \cdots, r_{k_1-1,k_1}, i_{k_1}, r_{k_1,k_1+1}, \cdots, r_{k_1,k_2-1}, r_{k_1,k_2}, r_{k_1,k_2+1}, \cdots, r_{k_1,N}) \cdots \\
&\quad \mathcal{G}_{k_2}(r_{1,k_2}, r_{2,k_2}, \cdots, r_{k_1-1,k_2}, r_{k_1,k_2}, r_{k_1+1,k_2}, \cdots, r_{k_2-1,k_2}, i_{k_2}, r_{k_2,k_2+1}, \cdots, r_{k_2,N}) \cdots \\
&\quad \mathcal{G}_N(r_{1,N}, r_{2,N}, \cdots, r_{N-1,N}, i_N)\big\}. \\
&= \sum_{r_{1,2}=1}^{R_{1,2}} \sum_{r_{1,3}=1}^{R_{1,3}} \cdots \sum_{r_{1,N}=1}^{R_{1,N}} \sum_{r_{2,3}=1}^{R_{2,3}} \cdots \sum_{r_{2,N}=1}^{R_{2,N}} \cdots \sum_{r_{k_1,k_1+1}=1}^{R_{k_1,k_1+1}} \cdots \sum_{r_{k_1,k_2}=1}^{R_{k_1,k_2}} \cdots \sum_{r_{k_1,N}=1}^{R_{k_1,N}} \cdots \sum_{r_{k_2,k_2+1}=1}^{R_{k_2,k_2+1}} \cdots \sum_{r_{k_2,N}=1}^{R_{k_2,N}} \cdots \sum_{r_{N-1,N}=1}^{R_{N-1,N}} \\
&\quad \big\{\vec{\mathcal{G}}^{\hat{\mathbf{n}}}_1(i_1, r_{1,2}, r_{1,3}, \cdots, r_{1,k_1-1}, r_{1,k_2}, r_{1,k_1+1}, \cdots, r_{1,k_2-1}, r_{1,k_1}, r_{1,k_2+1}, \cdots, r_{1,N}) \\
&\quad \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_2(r_{1,2}, i_2, r_{2,3}, \cdots, r_{2,k_1-1}, r_{2,k_2}, r_{2,k_1+1}, \cdots, r_{2,k_2-1}, r_{2,k_1}, r_{2,k_2+1}, \cdots, r_{2,N}) \cdots \\
&\quad \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_{k_1-1}(r_{1,k_1-1}, \cdots, r_{k_1-2,k_1-1}, i_{k_1-1}, r_{k_1-1,k_2}, r_{k_1-1,k_1+1}, \cdots, r_{k_1-1,k_2-1}, r_{k_1-1,k_1}, r_{k_1-1,k_2+1}, \cdots, r_{k_1-1,N}) \\
&\quad \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_{k_2}(r_{1,k_2}, r_{2,k_2}, \cdots, r_{k_1-1,k_2}, i_{k_2}, r_{k_1+1,k_2}, \cdots, r_{k_2-1,k_2}, r_{k_1,k_2}, r_{k_2,k_2+1}, \cdots, r_{k_2,N}) \\
&\quad \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_{k_1+1}(r_{1,k_1+1}, \cdots, r_{k_1-1,k_1+1}, r_{k_1+1,k_2}, i_{k_1+1}, r_{k_1+1,k_1+2}, \cdots, r_{k_1+1,k_2-1}, r_{k_1,k_1+1}, r_{k_1+1,k_2+1}, \cdots, r_{k_1+1,N}) \cdots \\
&\quad \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_{k_2-1}(r_{1,k_2-1}, \cdots, r_{k_1-1,k_2-1}, r_{k_2-1,k_2}, r_{k_1+1,k_2-1}, \cdots, r_{k_2-2,k_2-1}, i_{k_2-1}, r_{k_1,k_2-1}, r_{k_2-1,k_2+1}, \cdots, r_{k_2-1,N}) \\
&\quad \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_{k_1}(r_{1,k_1}, r_{2,k_1}, \cdots, r_{k_1-1,k_1}, r_{k_1,k_2}, r_{k_1,k_1+1}, \cdots, r_{k_1,k_2-1}, i_{k_1}, r_{k_1,k_2+1}, \cdots, r_{k_1,N}) \\
&\quad \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_{k_2+1}(r_{1,k_2+1}, \cdots, r_{k_1-1,k_2+1}, r_{k_2,k_2+1}, r_{k_1+1,k_2+1}, \cdots, r_{k_2-1,k_2+1}, r_{k_1,k_2+1}, i_{k_2+1}, r_{k_2+1,k_2+2}, \cdots, r_{k_2+1,N}) \cdots \\
&\quad \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_N(r_{1,N}, r_{2,N}, \cdots, r_{k_1-1,N}, r_{k_2,N}, r_{k_1+1,N}, \cdots, r_{k_2-1,N}, r_{k_1,N}, r_{k_2+1,N}, \cdots, r_{N-1,N}, i_N)\big\}.
\end{aligned}
$$

That is $\vec{\mathcal{X}}^{\hat{\mathbf{n}}} = \text{FCTN}\big(\vec{\mathcal{G}}^{\hat{\mathbf{n}}}_1, \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_2, \cdots, \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_{k_1-1}, \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_{k_2}, \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_{k_1+1}, \cdots, \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_{k_2-1}, \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_{k_1}, \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_{k_2+1}, \cdots, \vec{\mathcal{G}}^{\hat{\mathbf{n}}}_N\big)$. According to the idea of recursion, we can obtain $\vec{\mathcal{X}}^{\mathbf{n}} = \text{FCTN}\big(\vec{\mathcal{G}}^{\mathbf{n}}_{n_1}, \vec{\mathcal{G}}^{\mathbf{n}}_{n_2}, \cdots, \vec{\mathcal{G}}^{\mathbf{n}}_{n_N}\big)$ since the vector $\mathbf{n}$ can be obtained by continually exchanging the elements of $(1, 2, \cdots, N)$. □

Theorem 3 *Supposing that an Nth-order tensor $\mathcal{X}$ can be represented by Equation (1), the following inequality holds:*

$$\texttt{Rank}(\mathbf{X}_{[n_{1:d};n_{d+1:N}]}) \leq \prod_{i=1}^{d}\prod_{j=d+1}^{N} R_{n_i,n_j},$$

where $R_{n_i,n_j} = R_{n_j,n_i}$ if $n_i > n_j$ and $(n_1, n_2, \cdots, n_N)$ is a reordering of the vector $(1, 2, \cdots, N)$.

Proof. Supposing that $\mathcal{G}_k \in \mathbb{R}^{R_{1,k}\times R_{2,k}\times\cdots\times R_{k-1,k}\times I_k\times R_{k,k+1}\times\cdots\times R_{k,N}}$ $(k = 1, 2, \cdots N)$ are the FCTN factors of the tensor $\mathcal{X}$, i.e., $\mathcal{X} = \text{FCTN}(\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_N)$. We define the tensor $\mathcal{A}$ as the composition of the FCTN factors $\mathcal{G}_i$ $(i = 1, 2, \cdots, d)$ and the tensor $\mathcal{B}$ as the composition of the FCTN factors $\mathcal{G}_j$ $(j = d+1, d+2, \cdots, N)$. According to the definition of tensor contraction, the tensor $\mathcal{A}$ is of size $I_1 \times R_{1,d+1} \times R_{1,d+2} \times \cdots \times R_{1,N} \times I_2 \times R_{2,d+1} \times R_{2,d+2} \times \cdots \times R_{2,N} \times \cdots \times I_d \times R_{d,d+1} \times R_{d,d+2} \times \cdots \times R_{d,N}$ and the tensor $\mathcal{B}$ is of size $R_{1,d+1} \times R_{2,d+1} \times \cdots \times R_{d,d+1} \times I_{d+1} \times R_{1,d+2} \times R_{2,d+2} \times \cdots \times R_{d,d+2} \times I_{d+2} \times \cdots \times R_{1,N} \times R_{2,N} \times \cdots \times R_{d,N} \times I_N$, satisfying

$$\mathcal{X} = \mathcal{A} \times_{n_{1:l}}^{m_{1:l}} \mathcal{B},$$

with

$$\begin{aligned} m_i &= i + (q-1), && \text{if } (q-1)d+1 \leq i \leq qd, \\ n_i &= q + 1 + \big(i - [(q-1)d+1]\big)(N-d+1), && \text{if } (q-1)d+1 \leq i \leq qd, \end{aligned}$$

where $q = 1, 2, \cdots, N-d$ and $l = d(N-d)$. According to Theorem 1, we have

$$\mathbf{X}_{[1:d;d+1:N]} = \mathbf{A}_{[n'_{1:d};n_{1:l}]}\mathbf{B}_{[m_{1:l};m'_{1:N-d}]},$$

with

$$\begin{aligned} m'_i &= i + id, && i = 1, 2, \cdots, N-d, \\ n'_i &= i + (i-1)(N-d), && i = 1, 2, \cdots, d, \end{aligned}$$

where $\mathbf{A}_{[n'_{1:d};n_{1:l}]}$ is size of $\prod_{i=1}^{d} I_i \times \prod_{i=1}^{d}\prod_{j=d+1}^{N} R_{i,j}$ and $\mathbf{B}_{[m_{1:l};m'_{1:N-d}]}$ is size of $\prod_{i=1}^{d}\prod_{j=d+1}^{N} R_{i,j} \times \prod_{j=d+1}^{N} I_j$. Therefore, we obtain

$$\begin{aligned} \texttt{Rank}(\mathbf{X}_{[1:d;d+1:N]}) &\leq \min\left\{\texttt{Rank}(\mathbf{A}_{[n'_{1:d};n_{1:l}]}), \texttt{Rank}(\mathbf{B}_{[m_{1:l};m'_{1:N-d}]})\right\}, \\ &\leq \min\left\{\prod_{i=1}^{d} I_i, \prod_{i=1}^{d}\prod_{j=d+1}^{N} R_{i,j}, \prod_{j=d+1}^{N} I_j\right\}, \\ &\leq \prod_{i=1}^{d}\prod_{j=d+1}^{N} R_{i,j}. \end{aligned}$$

And since $\vec{\mathcal{X}}^{\mathbf{n}} = \text{FCTN}\big(\vec{\mathcal{G}}_{n_1}^{\mathbf{n}}, \vec{\mathcal{G}}_{n_2}^{\mathbf{n}}, \cdots, \vec{\mathcal{G}}_{n_N}^{\mathbf{n}}\big)$ (Theorem 2), we can easily obtain

$$\texttt{Rank}(\mathbf{X}_{[n_{1:d};n_{d+1:N}]}) \leq \prod_{i=1}^{d}\prod_{j=d+1}^{N} R_{n_i,n_j}.$$

□

Theorem 4 *Supposing that $\mathcal{G}_k$ for $k = 1, 2, \cdots, N$ are the FCTN factors of an Nth-order tensor, $\mathcal{X} = FCTN\big(\{\mathcal{G}_k\}_{k=1}^N\big)$, and $\mathcal{M}_t = FCTN\big(\{\mathcal{G}_k\}_{k=1}^N, /\mathcal{G}_t\big)$, we can obtain that*

$$\mathbf{X}_{(t)} = (\mathbf{G}_t)_{(t)}(\mathbf{M}_t)_{[m_{1:N-1};n_{1:N-1}]},$$

where

$$m_i = \begin{cases} 2i, & \text{if } i < t, \\ 2i-1, & \text{if } i \geq t, \end{cases} \quad \text{and} \quad n_i = \begin{cases} 2i-1, & \text{if } i < t, \\ 2i, & \text{if } i \geq t. \end{cases}$$

Proof. According to the definition of tensor contraction, the tensor $\mathcal{M}_t = \text{FCTN}\big(\{\mathcal{G}_k\}_{k=1}^N, /\mathcal{G}_t\big)$ is of size $I_1 \times R_{1,t} \times I_2 \times R_{2,t} \times \cdots \times I_{t-1} \times R_{t-1,t} \times R_{t,t+1} \times I_{t+1} \times \cdots \times R_{t,N} \times I_N$, satisfying

$$\vec{\mathcal{X}}^{\mathbf{p}} = \mathcal{G}_t \times_{q_{1:N-1}}^{m_{1:N-1}} \mathcal{M}_t,$$

with $\mathbf{q} = (1, 2, \cdots, t-1, t+1, t+2, \cdots, N)$ and $\mathbf{p} = (t, 1, 2, \cdots, t-1, t+1, t+2, \cdots, N)$. According to Theorem 1, we have

$$\mathbf{X}_{(t)} = (\mathbf{G}_t)_{(t)}(\mathbf{M}_t)_{[m_{1:N-1};n_{1:N-1}]}.$$

$\square$

Theorem 5 (Convergence) *The sequence $\{\mathcal{G}^{(s)}, \mathcal{X}^{(s)}\}_{s\in\mathbb{N}}$ obtained by the Algorithm 1 globally converges to a critical point of the problem (2).*

Proof. To prove the Theorem 5, we only need to justify that the following four conditions hold [1]:
1) $\mathcal{G}^{(s)}$ and $\mathcal{X}^{(s)}$ $(s \in \mathbb{N})$ are bounded;
2) $f(\mathcal{G}, \mathcal{X})$ is a proper lower semi-continuous function;
3) $f(\mathcal{G}, \mathcal{X})$ satisfies the K-Ł property at $\{\mathcal{G}^{(s)}, \mathcal{X}^{(s)}\}_{s\in\mathbb{N}}$;
4) $\{\mathcal{G}^{(s)}, \mathcal{X}^{(s)}\}_{s\in\mathbb{N}}$ satisfies Lemmas 1 and 2.

Lemma 1 (Sufficient Decrease) *Letting $\{\mathcal{G}^{(s)}, \mathcal{X}^{(s)}\}_{s\in\mathbb{N}}$ be the sequence obtained by the Algorithm 1, then it satisfies*

$$f(\mathcal{G}^{(s+1)}_{1:k}, \mathcal{G}^{(s)}_{k+1:N}, \mathcal{X}^{(s)}) + \frac{\rho}{2}\|\mathcal{G}^{(s+1)}_k - \mathcal{G}^{(s)}_k\|_F^2 \leq f(\mathcal{G}^{(s+1)}_{1:k-1}, \mathcal{G}^{(s)}_{k:N}, \mathcal{X}^{(s)}),\ k = 1, 2, \cdots, N;$$
$$f(\mathcal{G}^{(s+1)}, \mathcal{X}^{(s+1)}) + \frac{\rho}{2}\|\mathcal{X}^{(s+1)} - \mathcal{X}^{(s)}\|_F^2 \leq f(\mathcal{G}^{(s+1)}, \mathcal{X}^{(s)}).$$

Lemma 2 (Relative Error) *Letting $\{\mathcal{G}^{(s)}, \mathcal{X}^{(s)}\}_{s\in\mathbb{N}}$ be the sequence obtained by the Algorithm 1, then there exists $\mathcal{A}^{(s+1)}_k \in \mathbf{0}$ and $\mathcal{A}^{(s+1)} \in \partial_{\mathcal{X}}\iota_{\mathbb{S}}(\mathcal{X}^{(s+1)})$ satisfied*

$$\|\mathcal{A}^{(s+1)}_k + \nabla_{\mathcal{G}_k} h(\mathcal{G}^{(s+1)}_{1:k}, \mathcal{G}^{(s)}_{k+1:N}, \mathcal{X}^{(s)})\|_F \leq \rho\|\mathcal{G}^{(s+1)}_k - \mathcal{G}^{(s)}_k\|_F,\ k = 1, 2, \cdots, N;$$
$$\|\mathcal{A}^{(s+1)} + \nabla_{\mathcal{X}} h(\mathcal{G}^{(s+1)}, \mathcal{X}^{(s+1)})\|_F \leq \rho\|\mathcal{X}^{(s+1)} - \mathcal{X}^{(s)}\|_F,$$

where $h(\mathcal{G}, \mathcal{X}) = \frac{1}{2}\|\mathcal{X} - \mathit{FCTN}\big(\{\mathcal{G}_k\}_{k=1}^N\big)\|_F^2$.

First, as shown in the Algorithm 1, the initial $\mathcal{G}^{(0)}_k$ $(k = 1, 2, \cdots, N)$ and $\mathcal{X}^{(0)}$ are apparently bounded. Therefore, we only need to justify that $\mathcal{G}^{(s+1)}_k$ and $\mathcal{X}^{(s+1)}$ are bounded when $\mathcal{G}^{(s)}_k$ and $\mathcal{X}^{(s)}$ are bounded. Supposing that $\|\mathcal{G}^{(s)}_k\|_F \leq c$ and $\|\mathcal{X}^{(s)}\|_F \leq d$, and according to (5), we have

$$\begin{aligned}\|\mathcal{G}^{(s+1)}_1\|_F &\leq (\|\mathcal{X}^{(s)}\|_F\|\mathcal{M}^{(s)}_1\|_F + \rho\|\mathcal{G}^{(s)}_1\|_F)\|(\mathbf{Q}^{(s)}_1 + \rho\mathbf{I})^{-1}\|_F \\ &\leq (dc^{N-1} + \rho c)\sqrt{\sum\nolimits_{i=1}^{j}\big(1/(\varepsilon_i + \rho)\big)^2} \\ &\leq (dc^{N-1} + \rho c)\sqrt{j}/\rho,\end{aligned}$$

where $\mathbf{Q}^{(s)}_1 = (\mathbf{M}^{(s)}_1)_{[m_{1:N-1};n_{1:N-1}]}(\mathbf{M}^{(s)}_1)_{[n_{1:N-1};m_{1:N-1}]}$ and $\varepsilon_i \geq 0$ $(i = 1, 2, \cdots, j = \prod_{t=2}^N R_{1,t})$ are the eigenvalues of $\mathbf{Q}^{(s)}_1$. Thus, $\mathcal{G}^{(s+1)}_1$ is bounded. Similarly, we can obtain that $\mathcal{G}^{(s+1)}_2$, $\mathcal{G}^{(s+1)}_3$, $\cdots$, and $\mathcal{G}^{(s+1)}_N$ are bounded. Supposing that $\|\mathcal{G}^{(s+1)}_k\|_F \leq e$ and according to (6), we have

$$\|\mathcal{X}^{(s+1)}\|_F \leq (e^N + \rho d)/(1 + \rho) + \|\mathcal{F}\|_F.$$

Therefore, $\mathcal{X}^{(s+1)}$ is bounded and the condition 1) holds.

Second, $f(\mathcal{G}, \mathcal{X})$ is the sum of a Frobenius-norm-based function $h(\mathcal{G}, \mathcal{X})$ and an indicator function $\iota_{\mathbb{S}}(\mathcal{X})$. It is not hard to see that $h(\mathcal{G}, \mathcal{X})$ is a C^1 function whose gradient is Lipschitz continuous and $\iota_{\mathbb{S}}(\mathcal{X})$ is a proper lower semi-continuous function. Therefore, the condition 2) holds.

Third, since the semi-algebraic real-valued function satisfies the K-Ł property [2], we only need to illustrate that $f(\mathcal{G}, \mathcal{X})$ is a semi-algebraic function. It is easily obtained since the sum of two semi-algebraic functions is still a semi-algebraic function, and $h(\mathcal{G}, \mathcal{X})$ and $\iota_{\mathbb{S}}(\mathcal{X})$ are the semi-algebraic functions. Therefore, the condition 3) holds.

Fourth, we first prove the Lemma 1. Since $\mathcal{G}^{(s+1)}_k$ is the optimal solution of the $\mathcal{G}_k$-subproblem, we have

$$\begin{aligned}f(\mathcal{G}^{(s+1)}_{1:k}, \mathcal{G}^{(s)}_{k+1:N}, \mathcal{X}^{(s)}) + \frac{\rho}{2}\|\mathcal{G}^{(s+1)}_k - \mathcal{G}^{(s)}_k\|_F^2 &\leq f(\mathcal{G}^{(s+1)}_{1:k-1}, \mathcal{G}^{(s)}_{k:N}, \mathcal{X}^{(s)}) + \frac{\rho}{2}\|\mathcal{G}^{(s)}_k - \mathcal{G}^{(s)}_k\|_F^2 \\ &= f(\mathcal{G}^{(s+1)}_{1:k-1}, \mathcal{G}^{(s)}_{k:N}, \mathcal{X}^{(s)}).\end{aligned}$$

Similarly, since $\mathcal{X}^{(s+1)}$ is the optimal solution of the $\mathcal{X}$-subproblem, we have

$$f(\mathcal{G}^{(s+1)}, \mathcal{X}^{(s+1)}) + \frac{\rho}{2}\|\mathcal{X}^{(s+1)} - \mathcal{X}^{(s)}\|_F^2 \le f(\mathcal{G}^{(s+1)}, \mathcal{X}^{(s)}).$$

Then we prove the Lemma 2. For each subproblem, we have

$$\begin{aligned}
&0 \in \nabla_{\mathcal{G}_k} h(\mathcal{G}_{1:k-1}^{(s+1)}, \mathcal{G}_k, \mathcal{G}_{k+1:N}^{(s)}, \mathcal{X}^{(s)}) + \rho(\mathcal{G}_k - \mathcal{G}_k^{(s)}),\\
&0 \in \nabla_{\mathcal{X}} h(\mathcal{G}^{(s+1)}, \mathcal{X}) + \rho(\mathcal{X} - \mathcal{X}^{(s)}) + \partial_{\mathcal{X}} \iota_{\mathbb{S}}(\mathcal{X}).
\end{aligned}$$

Letting

$$\begin{aligned}
&\mathcal{A}_k^{(s+1)} = -\nabla_{\mathcal{G}_k} h(\mathcal{G}_{1:k}^{(s+1)}, \mathcal{G}_{k+1:N}^{(s)}, \mathcal{X}^{(s)}) - \rho(\mathcal{G}_k^{(s+1)} - \mathcal{G}_k^{(s)}) \in \mathbf{0},\\
&\mathcal{A}^{(s+1)} = -\nabla_{\mathcal{X}} h(\mathcal{G}^{(s+1)}, \mathcal{X}^{(s+1)}) - \rho(\mathcal{X}^{(s+1)} - \mathcal{X}^{(s)}) \in \partial_{\mathcal{X}} \iota_{\mathbb{S}}(\mathcal{X}^{(s+1)}),
\end{aligned}$$

which are evidently satisfied the conditions in the Lemma 2. Therefore, the condition 4) holds. □

2 Numerical Experiments for the Storage Cost

We test the storage cost of the Tucker decomposition and the proposed FCTN decomposition on a real hyperspectral video[1] (HSV) of size $60 \times 60 \times 20 \times 20$ (spatial height×spatial width×band×frame) [3]. All methods are solved by PAM to get rid of the influence of the algorithm. When the error bound is 10^{-2}, we find that Tucker decomposition needs 384652 (26.71% of total elements) parameters (Tucker rank is (38, 39, 16, 16)) to express the testing data, while FCTN decomposition only needs 20000 (1.39% of total elements) parameters (FCTN rank is (5,5,5,5,5,5)). Here, the Tucker rank is set as different values to obtain the minimum parameters, and the FCTN rank is set as the same value for reducing hyper-parameters in our method. This testing result provides empirical evidence for the analysis (Section 3.2 in the main body) regarding the superiorities of the FCTN decomposition over the Tucker decomposition for the storage cost.

References

[1] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods," *Mathematical Programming*, vol. 137, no. 1, pp. 91–129, 2013.

[2] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota, "Clarke subgradients of stratifiable functions," *SIAM Journal on Optimization*, vol. 18, no. 2, pp. 556–572, 2007.

[3] A. Mian and R. Hartley, "Hyperspectral video restoration using optical flow and sparse coding," *Optics Express*, vol. 20, no. 10, pp. 10 658–10 673, 2012.

[1] The data is available at: *http://openremotesensing.net/kb/data/*.