Supplementary Materials

In this section, we give proofs of three theorems in paper.
Proof of theorem 3.1

$$\begin{aligned}
&\sum_{i=1}^{n}\|\hat{\boldsymbol{\xi}}_i-\boldsymbol{\gamma}_i\times{}_1\hat{\boldsymbol{\beta}}_1\times{}_2\cdots\times{}_K\hat{\boldsymbol{\beta}}_K\|_F^2\\
=&\sum_{i=1}^{n}[\mathrm{vec}(\hat{\boldsymbol{\xi}}_i)-(\hat{\boldsymbol{\beta}}_K\otimes\cdots\otimes\hat{\boldsymbol{\beta}}_1)\mathrm{vec}(\boldsymbol{\gamma}_i)]^T[\mathrm{vec}(\hat{\boldsymbol{\xi}}_i)-(\hat{\boldsymbol{\beta}}_K\otimes\cdots\otimes\hat{\boldsymbol{\beta}}_1)\mathrm{vec}(\boldsymbol{\gamma}_i)] \qquad (16)\\
=&\sum_{i=1}^{n}[\mathrm{vec}(\hat{\boldsymbol{\xi}}_i)^T\mathrm{vec}(\hat{\boldsymbol{\xi}}_i)-\mathrm{vec}(\boldsymbol{\gamma}_i)^T(\hat{\boldsymbol{\beta}}_K\otimes\cdots\otimes\hat{\boldsymbol{\beta}}_1)\mathrm{vec}(\boldsymbol{\gamma}_i)\\
&-\mathrm{vec}(\hat{\boldsymbol{\xi}}_i)^T(\hat{\boldsymbol{\beta}}_K\otimes\cdots\otimes\hat{\boldsymbol{\beta}}_1)^T\mathrm{vec}(\boldsymbol{\gamma}_i)\\
&+\mathrm{vec}(\boldsymbol{\gamma}_i)^T(\hat{\boldsymbol{\beta}}_K\otimes\cdots\otimes\hat{\boldsymbol{\beta}}_1)^T(\hat{\boldsymbol{\beta}}_K\otimes\cdots\otimes\hat{\boldsymbol{\beta}}_1)\mathrm{vec}(\boldsymbol{\gamma}_i)]\\
=&\sum_{i=1}^{n}\mathrm{vec}(\hat{\boldsymbol{\xi}}_i)^T\mathrm{vec}(\hat{\boldsymbol{\xi}})-2\sum_{i=1}^{n}\mathrm{vec}(\hat{\boldsymbol{\xi}}_i)^T(\hat{\boldsymbol{\beta}}_K\otimes\cdots\otimes\hat{\boldsymbol{\beta}}_1)\mathrm{vec}(\boldsymbol{\gamma}_i)\\
&+\sum_{i=1}^{n}\mathrm{vec}(\boldsymbol{\gamma}_i)^T\mathrm{vec}(\boldsymbol{\gamma}_i), \qquad (17)
\end{aligned}$$

where equation (16) holds because of the fourth conclusion in *proposition* 2.3 and equation (17) holds due to $\hat{\boldsymbol{\beta}}_k^T\hat{\boldsymbol{\beta}}_k=I_{d_k}, k=1,\cdots,K$. Also, since the last term in equation (17) is a constant, the minimization in equation (11) is equivalent to minimizing

$$Q(\boldsymbol{\gamma})=\sum_{i=1}^{n}\mathrm{vec}(\boldsymbol{\gamma}_i)^T\mathrm{vec}(\boldsymbol{\gamma}_i)-2\sum_{i=1}^{n}\mathrm{vec}(\hat{\boldsymbol{\xi}}_i)^T(\hat{\boldsymbol{\beta}}_K\otimes\cdots\otimes\boldsymbol{\beta})\mathrm{vec}(\boldsymbol{\gamma}_i). \qquad (18)$$

Differentiate $Q(\boldsymbol{\gamma})$ with respect to $\mathrm{vec}(\boldsymbol{\gamma}_i)$

$$\begin{aligned}
\frac{\partial Q(\boldsymbol{\gamma})}{\partial\mathrm{vec}(\boldsymbol{\gamma}_i)^T}&=2\mathrm{vec}(\boldsymbol{\gamma}_i)^T-2\mathrm{vec}(\hat{\boldsymbol{\xi}}_i)^T(\hat{\boldsymbol{\beta}}_K\otimes\cdots\otimes\hat{\boldsymbol{\beta}}_1)\\
&=0.
\end{aligned}$$

So we can obtain $\mathrm{vec}(\boldsymbol{\gamma}_i)=(\hat{\boldsymbol{\beta}}_K\otimes\cdots\otimes\hat{\boldsymbol{\beta}}_1)^T\mathrm{vec}(\hat{\boldsymbol{\xi}}_i)$. Then by the fourth conclusion in *proposition* 2.3, we can get $\hat{\boldsymbol{\gamma}}_i=\hat{\boldsymbol{\xi}}_i\times{}_1\hat{\boldsymbol{\beta}}_1^T\times{}_2\cdots\times{}_K\hat{\boldsymbol{\beta}}_K^T$. ■.

Proof of theorem 3.2. From Theorem 3.1, we know that

$$\hat{\boldsymbol{\gamma}}_i=\hat{\boldsymbol{\xi}}_i\times{}_1\boldsymbol{\beta}_1^T\times{}_2\cdots\times{}_K\boldsymbol{\beta}_K^T$$

for every i, plugging $\boldsymbol{\gamma}_i\triangleq\hat{\boldsymbol{\gamma}}_i=\hat{\boldsymbol{\xi}}_i\times{}_1\boldsymbol{\beta}_1^T\times{}_2\cdots\times{}_K\boldsymbol{\beta}_K^T$ into $\sum_{i=1}^{n}\|\hat{\boldsymbol{\xi}}_i-\boldsymbol{\gamma}_i\times{}_1\boldsymbol{\beta}_1\times{}_2\cdots\times{}_K\boldsymbol{\beta}_K\|_F^2$,

For Peer Review Only

then we have

$$
\begin{aligned}
&\sum_{i=1}^{n} \|\hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\xi}}_i \times {}_1\boldsymbol{\beta}_1^T \times {}_2 \cdots \times {}_K\boldsymbol{\beta}_K^T \times {}_1\boldsymbol{\beta}_1 \times {}_2 \cdots \times {}_K\boldsymbol{\beta}_K\|_F^2 \\
&= \sum_{i=1}^{n} \|\hat{\boldsymbol{\xi}}_i - \hat{\boldsymbol{\xi}}_i \times {}_1(\boldsymbol{\beta}_1\boldsymbol{\beta}_1^T) \times {}_2 \cdots \times {}_K(\boldsymbol{\beta}_K\boldsymbol{\beta}_K^T)\|_F^2 \\
&= \sum_{i=1}^{n} \left[\mathrm{vec}(\hat{\boldsymbol{\xi}}_i) - \mathrm{vec}\big(\hat{\boldsymbol{\xi}}_i \times {}_1(\boldsymbol{\beta}_1\boldsymbol{\beta}_1^T) \times {}_2 \cdots \times {}_K(\boldsymbol{\beta}_K\boldsymbol{\beta}_K^T)\big)\right]^T \\
&\quad \times \left[\mathrm{vec}(\hat{\boldsymbol{\xi}}_i) - \mathrm{vec}\big(\hat{\boldsymbol{\xi}}_i \times {}_1(\boldsymbol{\beta}_1\boldsymbol{\beta}_1^T) \times {}_2 \cdots \times {}_K(\boldsymbol{\beta}_K\boldsymbol{\beta}_K^T)\big)\right] \\
&= \sum_{i=1}^{n} \mathrm{vec}(\hat{\boldsymbol{\xi}})^T \mathrm{vec}(\hat{\boldsymbol{\xi}}_i) - \sum_{i=1}^{n} \mathrm{vec}(\hat{\boldsymbol{\xi}}_i)^T \big((\boldsymbol{\beta}_K\boldsymbol{\beta}_K^T) \otimes \cdots \otimes (\boldsymbol{\beta}_1\boldsymbol{\beta}_1^T)\big)\mathrm{vec}(\hat{\boldsymbol{\xi}}_i) \\
&\quad - \sum_{i=1}^{n} \mathrm{vec}(\hat{\boldsymbol{\xi}}_i)^T \big((\boldsymbol{\beta}_K\boldsymbol{\beta}_K^T) \otimes \cdots \otimes (\boldsymbol{\beta}_1\boldsymbol{\beta}_1^T)\big)\mathrm{vec}(\hat{\boldsymbol{\xi}}_i) \\
&\quad - \sum_{i=1}^{n} \mathrm{vec}(\hat{\boldsymbol{\xi}}_i)^T \big((\boldsymbol{\beta}_K\boldsymbol{\beta}_K^T) \otimes \cdots \otimes (\boldsymbol{\beta}_1\boldsymbol{\beta}_1^T)\big) \times \big((\boldsymbol{\beta}_K\boldsymbol{\beta}_K^T) \otimes \cdots \otimes (\boldsymbol{\beta}_1\boldsymbol{\beta}_1^T)\big)\mathrm{vec}(\hat{\boldsymbol{\xi}}_i) \\
&= \sum_{i=1}^{n} \mathrm{vec}(\hat{\boldsymbol{\xi}}_i)^T \mathrm{vec}(\boldsymbol{\xi}_i) - \sum_{i=1}^{n} \mathrm{vec}(\boldsymbol{\xi}_i)^T \big((\boldsymbol{\beta}_K\boldsymbol{\beta}_K^T) \otimes \cdots \otimes (\boldsymbol{\beta}_1\boldsymbol{\beta}_1^T)\big)\mathrm{vec}(\hat{\boldsymbol{\xi}}_i) \\
&= \sum_{i=1}^{n} \|\hat{\boldsymbol{\xi}}_i\|_F^2 - \sum_{i=1}^{n} \|\hat{\boldsymbol{\xi}}_i \times {}_1\boldsymbol{\beta}_1^T \times {}_2 \cdots \times {}_K\boldsymbol{\beta}_K^T\|_F^2.
\end{aligned}
$$

Hence the minimization of equation (11) is equivalent to the following optimizing problem

$$
\{\hat{\boldsymbol{\beta}}_k\}_{k=1}^{K} = \underset{\boldsymbol{\beta}_k^T\boldsymbol{\beta}_k = I_{d_k}}{\arg\max} \sum_{i=1}^{n} \|\hat{\boldsymbol{\xi}}_i \times {}_1\boldsymbol{\beta}_1^T \times {}_2 \cdots \times {}_K\boldsymbol{\beta}_K^T\|_F^2,
$$

which completes the proof of the theorem. ■.

Proof of theorem 3.3. By Theorem 3.2, $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_K$ maximize

$$
\sum_{i=1}^{n} \|\hat{\boldsymbol{\xi}}_i \times {}_1\boldsymbol{\beta}_1^T \times {}_2 \cdots \times {}_K\boldsymbol{\beta}_K^T\|_F^2. \tag{19}
$$

Let $\boldsymbol{Y}_i = \hat{\boldsymbol{\xi}}_i \times {}_1\boldsymbol{\beta}_1^T \times {}_2 \cdots \times {}_K\boldsymbol{\beta}_K^T$, then $\boldsymbol{Y}_{i(k)} = \boldsymbol{\beta}_k^T \hat{\boldsymbol{\xi}}_{i(k)}(\boldsymbol{\beta}_K \otimes \cdots \otimes \boldsymbol{\beta}_{k+1} \otimes \boldsymbol{\beta}_{k-1} \cdots \otimes \boldsymbol{\beta}_1)$,

For Peer Review Only

so we rewrite the above formula (19) as

$$
\begin{aligned}
&\sum_{i=1}^{n} \|\hat{\boldsymbol{\xi}}_i \times {}_1\boldsymbol{\beta}_1^T \times {}_2 \cdots \times {}_K\boldsymbol{\beta}_K^T\|_F^2 = \sum_{i=1}^{n} \|\mathbf{Y}_i\|_F^2 \\
&= \sum_{i=1}^{n} \mathrm{Trace}(\mathbf{Y}_{i(k)}\mathbf{Y}_{i(k)}^T) \\
&= \sum_{i=1}^{n} \mathrm{Trace}\big(\boldsymbol{\beta}_k^T \hat{\boldsymbol{\xi}}_{i(k)}(\boldsymbol{\beta}_K \otimes \cdots \otimes \boldsymbol{\beta}_{k+1} \otimes \boldsymbol{\beta}_{k-1} \otimes \cdots \otimes \boldsymbol{\beta}_1) \\
&\quad \times (\boldsymbol{\beta}_K \otimes \cdots \otimes \boldsymbol{\beta}_{k+1} \otimes \boldsymbol{\beta}_{k-1} \otimes \cdots \otimes \boldsymbol{\beta}_1)^T \hat{\boldsymbol{\xi}}_{i(k)}^T \boldsymbol{\beta}_k\big) \\
&= \mathrm{Trace}\Big(\boldsymbol{\beta}_k^T \sum_{i=1}^{n} \hat{\boldsymbol{\xi}}_{i(k)}\big((\boldsymbol{\beta}_K\boldsymbol{\beta}_K^T) \otimes \cdots \otimes (\boldsymbol{\beta}_{k+1}\boldsymbol{\beta}_{k+1}^T) \otimes (\boldsymbol{\beta}_{k-1}\boldsymbol{\beta}_{k-1}^T) \otimes \cdots \otimes (\boldsymbol{\beta}_1\boldsymbol{\beta}_1^T)\big)\hat{\boldsymbol{\xi}}_{i(k)}^T)\boldsymbol{\beta}_k\Big) \\
&\triangleq \mathrm{Trace}(\boldsymbol{\beta}_k^T M_k \boldsymbol{\beta}_k),
\end{aligned}
$$

where $M_k \triangleq \sum_{i=1}^{n} \hat{\boldsymbol{\xi}}_{i(k)}\big((\boldsymbol{\beta}_K\boldsymbol{\beta}_K^T) \otimes \cdots \otimes (\boldsymbol{\beta}_{k+1}\boldsymbol{\beta}_{k+1}^T) \otimes (\boldsymbol{\beta}_{k-1}\boldsymbol{\beta}_{k-1}^T) \otimes \cdots \otimes (\boldsymbol{\beta}_1\boldsymbol{\beta}_1^T)\big)\hat{\boldsymbol{\xi}}_{i(k)}^T\big)$. Hence, given $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_{k-1}, \boldsymbol{\beta}_{k+1}, \cdots, \boldsymbol{\beta}_K$, the maximum of $\sum_{i=1}^{n} \|\hat{\boldsymbol{\xi}}_i \times_1 \boldsymbol{\beta}_1^T \times_2 \cdots \times_K \boldsymbol{\beta}_K^T\|_F^2 = \mathrm{Trace}(\boldsymbol{\beta}_k^T M_k \boldsymbol{\beta}_k)$ is obtained. Hence, $\boldsymbol{\beta} \in \mathbb{R}^{p_k \times d_k}$ consists of the d_k eigenvectors of the matrix M_k corresponding to the largest d_k eigenvalues. ■

For Peer Review Only