Appendix

Proof for Theorem 1

Theorem 1: Suppose that $\hat{\boldsymbol{X}} = \boldsymbol{x}_i \overline{\boldsymbol{R}}_j \boldsymbol{x}_k^\top$ for $j = 1, 2, \cdots, |\mathcal{R}|$, where $\boldsymbol{x}_i, \boldsymbol{x}_k, \boldsymbol{R}_j$ are real matrices and $\boldsymbol{R}_j$ is diagonal. Then, the following equation holds

$$\min \frac{1}{\sqrt{|\mathcal{R}|}} \sum_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} ||\boldsymbol{x}_i||_F^2 + ||\boldsymbol{x}_k||_F^2 + \|\boldsymbol{x}_i \mathbf{R}_j - \boldsymbol{x}_k \mathbf{R}_j||_F^2 = \|\hat{\mathcal{X}}\|_2.$$

Proof *Notice that*

$$\begin{aligned}
&\min \sum_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} ||\boldsymbol{x}_i||_F^2 + ||\boldsymbol{x}_k||_F^2 + \|\boldsymbol{x}_i \mathbf{R}_j - \boldsymbol{x}_k \mathbf{R}_j||_F^2 \\
\overset{a}{\leq} &\min \sum_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} ||\boldsymbol{x}_i||_F^2 + ||\boldsymbol{x}_k||_F^2 + \|\boldsymbol{x}_i \mathbf{R}_j\|_F^2 + \|\boldsymbol{x}_k \mathbf{R}_j||_F^2 \\
&- \|\boldsymbol{x}_i \mathbf{R}_j \boldsymbol{x}_k \mathbf{R}_j\| - \|\boldsymbol{x}_k \mathbf{R}_j \boldsymbol{x}_i \mathbf{R}_j\| \\
\leq &\min \sum_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} ||\boldsymbol{x}_i||_F^2 + ||\boldsymbol{x}_k||_F^2 + \|\boldsymbol{x}_i \mathbf{R}_j\|_F^2 + \|\boldsymbol{x}_k \mathbf{R}_j||_F^2 \\
&- 2\|\boldsymbol{x}_i \mathbf{R}_j \boldsymbol{x}_k \mathbf{R}_j\|_F^2 \\
\leq &\min \sum_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} ||\boldsymbol{x}_i||_F^2 + ||\boldsymbol{x}_k||_F^2 + \|\boldsymbol{x}_i \mathbf{R}_j\|_F^2 + \|\boldsymbol{x}_k \mathbf{R}_j||_F^2.
\end{aligned} \tag{17}$$

Since $\boldsymbol{x}_i \boldsymbol{R}_j$ and $\boldsymbol{x}_k \boldsymbol{R}_j$ are all vectors, we can have $\|\boldsymbol{x}_i \boldsymbol{R}_j \boldsymbol{x}_k \boldsymbol{R}_j\| = \|\boldsymbol{x}_k \boldsymbol{R}_j \boldsymbol{x}_i \boldsymbol{R}_j\|$. Then the inequality (a) holds.

We first prove that the following equation holds

$$\min_{\hat{X}_j = \boldsymbol{x}_i R_j x_k^\top} \frac{1}{\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} \Big(\|\boldsymbol{x}_i\|_F^2 + \|\boldsymbol{x}_k\|_F^2 + \|\boldsymbol{x}_i \boldsymbol{R}_j\|_F^2 + \|\boldsymbol{x}_k \boldsymbol{R}_j\|_F^2 \Big) = \|\hat{\mathcal{X}}\|_2. \tag{18}$$

Denote $\boldsymbol{x}_i$ as $\boldsymbol{p}$ and $\boldsymbol{x}_k$ as $\boldsymbol{q}$. Then we have that

$$\begin{aligned}
&\sum_{j=1}^{|\mathcal{R}|} \Big(\|\boldsymbol{x}_i \mathbf{R}_j\|_F^2 + \|\boldsymbol{x}_k\|_F^2 \Big) \\
=&\sum_{j=1}^{|\mathcal{R}|} \left(\sum_{d=1}^{D} \|\mathbf{q}_{:d}\|_F^2 + \sum_{i=1}^{I} \sum_{d=1}^{D} \mathbf{p}_{id}^2 \mathbf{r}_{jd}^2 \right) \\
=&\sum_{j=1}^{|\mathcal{R}|} \sum_{d=1}^{D} \|\mathbf{q}_{:d}\|_2^2 + \sum_{d=1}^{D} \|\mathbf{p}_{:d}\|_2^2 \|\mathbf{r}_{:d}\|_2^2 \\
=&\sum_{d=1}^{D} \Big(\|\mathbf{p}_{:d}\|_2^2 \|\mathbf{r}_{:d}\|_2^2 + |\mathcal{R}| \|\mathbf{q}_{:d}\|_2^2 \Big) \\
\geq&\sum_{d=1}^{D} 2\sqrt{|\mathcal{R}|} \|\mathbf{p}_{:d}\|_2 \|\mathbf{r}_{:d}\|_2 \|\mathbf{q}_{:d}\|_2 \\
=&2\sqrt{|\mathcal{R}|} \sum_{d=1}^{D} \|\mathbf{p}_{:d}\|_2 \|\mathbf{r}_{:d}\|_2 \|\mathbf{q}_{:d}\|_2
\end{aligned}$$

We can have the equality holds if and only if $\|\mathbf{p}_{:d}\|_2^2 \|\mathbf{r}_{:d}\|_2^2 = |\mathcal{R}| \|\mathbf{q}_{:d}\|_2^2$, i.e., $\|\mathbf{p}_{:d}\|_2 \|\mathbf{r}_{:d}\|_2 = \sqrt{|\mathcal{R}|} \|\mathbf{q}_{:d}\|_2$.

For all CP decomposition $\hat{\mathcal{X}} = \sum_{d=1}^{D} \mathbf{p}_{:d} \otimes \mathbf{r}_{:d} \otimes \mathbf{q}_{:d}$, we can always let $\boldsymbol{p}'_{:d} = \boldsymbol{p}_{:d}$, $\boldsymbol{r}'_{:d} = \sqrt{\frac{\|\mathbf{q}_d\|_2 \sqrt{|\mathcal{R}|}}{\|\mathbf{p}_{:d}\|_2 \|\mathbf{r}_{:d}\|_2}} \mathbf{r}_{:d}$ and $\boldsymbol{q}'_{:d} = \sqrt{\frac{\|\mathbf{p}_{:d}\|_2 \|\mathbf{r}_{:d}\|_2}{\|\mathbf{q}_{\cdot d}\|_2 \sqrt{|\mathcal{R}|}}} \mathbf{p}_{:d}$, then we have

$$\|\mathbf{p}'_{:d}\|_2 \|\mathbf{r}'_{:d}\|_2 = \sqrt{|\mathcal{R}|} \|\mathbf{q}'_{:d}\|_2,$$

and at the same time we make sure that $\hat{\mathcal{X}} = \sum_{d=1}^{D} \mathbf{p}'_{:d} \otimes \mathbf{r}'_{:d} \otimes \mathbf{q}'_{:d}$. Then, we can have

$$\begin{aligned}
\frac{1}{\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} \left\| \hat{\mathcal{X}}_j \right\|_2 &= \frac{1}{2\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} \min_{\hat{\mathcal{X}}_j = \boldsymbol{x}_i \mathbf{R}_j \boldsymbol{x}_k^\top} \Big(\|\boldsymbol{x}_i \mathbf{R}_j\|_F^2 + \|\boldsymbol{x}_k\|_F^2 \Big) \\
&\leq \frac{1}{2\sqrt{|\mathcal{R}|}} \min_{\hat{\mathcal{X}}_j = \boldsymbol{x}_i \mathbf{R}_j \boldsymbol{x}_k^\top} \sum_{j=1}^{|\mathcal{R}|} \Big(\|\boldsymbol{x}_i \mathbf{R}_j\|_F^2 + \|\boldsymbol{x}_k\|_F^2 \Big) \\
&= \min_{\hat{\mathcal{X}} = \sum_{d=1}^{D} \mathbf{p}_j \otimes \mathbf{r}_{id} \otimes \mathbf{q}_{:d}} \sum_{d=1}^{D} \|\mathbf{p}_{:d}\|_2 \|\mathbf{r}_{:d}\|_2 \|\mathbf{q}_{:d}\|_2 \\
&= \|\hat{\mathcal{X}}\|_2.
\end{aligned}$$

And in the same manner, we can have that

$$\min \frac{1}{\sqrt{|\mathcal{R}|}} \sum_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} ||\boldsymbol{x}_i||_F^2 + ||\boldsymbol{x}_k||_F^2 + \|\boldsymbol{x}_i \mathbf{R}_j - \boldsymbol{x}_k \mathbf{R}_j||_F^2 = \|\hat{\mathcal{X}}\|_2.$$

We can have the equality holds if and only if $\|\mathbf{q}_{:d}\|_2 \|\mathbf{r}_{:d}\|_2 = \sqrt{|\mathcal{R}|} \|\mathbf{p}_{:d}\|_2$. Then we can see that the conclusion holds if and only if $\|\mathbf{p}_{:d}\|_2 \|\mathbf{r}_{:d}\|_2 = \sqrt{|\mathcal{R}|} \|\mathbf{q}_{:d}\|_2$ and $||\boldsymbol{q}_{:d}||_2 ||\boldsymbol{r}_{:,d}||_2 = \sqrt{|\mathcal{R}|} \|\mathbf{p}_{:d}\|_2, \forall d \in \{1, 2, \ldots, D\}$. Then the proof of Theorem 1 completes.

Proof for Theorem 2

Theorem 2. Suppose that $\hat{\boldsymbol{X}} = \boldsymbol{x}_i \overline{\boldsymbol{R}}_j \boldsymbol{x}_k^\top$ for $j = 1, 2, \cdots, |\mathcal{R}|$, where $\boldsymbol{x}_i, \boldsymbol{x}_k, \boldsymbol{R}_j$ are real matrices and $\boldsymbol{R}_j$ is diagonal. Then, the following equation holds

$$\min \frac{1}{2\sqrt{|\mathcal{R}|}} \sum_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} ||\boldsymbol{x}_i||_F^2 + ||\boldsymbol{x}_k||_F^2 + \|\boldsymbol{x}_i \mathbf{R}_j + \boldsymbol{x}_k \mathbf{R}_j||_F^2 = \|\hat{\mathcal{X}}\|_2.$$

Proof *First we have*

$$\min \frac{1}{2\sqrt{|\mathcal{R}|}} \sum_{(\boldsymbol{x}_i,\boldsymbol{R}_j,\boldsymbol{x}_k)\in\mathcal{S}} ||\boldsymbol{x}_i||_F^2 + ||\boldsymbol{x}_k||_F^2 + \|\boldsymbol{x}_i\mathbf{R}_j + \boldsymbol{x}_k\mathbf{R}_j\|_F^2$$

$$\leq \min_{(\boldsymbol{x}_i,\boldsymbol{R}_j,\boldsymbol{x}_k)\in\mathcal{S}} \frac{1}{2\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} \|\boldsymbol{x}_i\|_F^2 + \|\boldsymbol{x}_k\|_F^2 + \|\boldsymbol{x}_i\boldsymbol{R}_j\|_F^2 + \|\boldsymbol{x}_k\boldsymbol{R}_j\|_F^2 + \|\boldsymbol{x}_i\boldsymbol{R}_j\boldsymbol{x}_k\boldsymbol{R}_j\| + \|\boldsymbol{x}_k\boldsymbol{R}_j\boldsymbol{x}_i\boldsymbol{R}_j\|$$

$$\stackrel{a}{=} \min_{(\boldsymbol{x}_i,\boldsymbol{R}_j,\boldsymbol{x}_k)\in\mathcal{S}} \frac{1}{2\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} \|\boldsymbol{x}_i\|_F^2 + \|\boldsymbol{x}_k\|_F^2 + \|\boldsymbol{x}_i\boldsymbol{R}_j\|_F^2 + \|\boldsymbol{x}_k\boldsymbol{R}_j\|_F^2 + 2\|\boldsymbol{x}_i\boldsymbol{R}_j\boldsymbol{x}_k\boldsymbol{R}_j\|$$

$$\leq \min_{(\boldsymbol{x}_i,\boldsymbol{R}_j,\boldsymbol{x}_k)\in\mathcal{S}} \frac{1}{2\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} \|\boldsymbol{x}_i\|_F^2 + \|\boldsymbol{x}_k\|_F^2 + \|\boldsymbol{x}_i\boldsymbol{R}_j\|_F^2 + \|\boldsymbol{x}_k\boldsymbol{R}_j\|_F^2 + \|\boldsymbol{x}_i\boldsymbol{R}_j\|_F^2 + \|\boldsymbol{x}_k\boldsymbol{R}_j\|_F^2$$

$$\leq \min_{(\boldsymbol{x}_i,\boldsymbol{R}_j,\boldsymbol{x}_k)\in\mathcal{S}} \frac{1}{2\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} \|\boldsymbol{x}_i\|_F^2 + \|\boldsymbol{x}_k\|_F^2 + 2\|\boldsymbol{x}_i\boldsymbol{R}_j\|_F^2 + 2\|\boldsymbol{x}_k\boldsymbol{R}_j\|_F^2$$

$$\leq \min_{(\boldsymbol{x}_i,\boldsymbol{R}_j,\boldsymbol{x}_k)\in\mathcal{S}} \frac{1}{\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} \|\boldsymbol{x}_i\|_F^2 + \|\boldsymbol{x}_k\|_F^2 + \|\boldsymbol{x}_i\boldsymbol{R}_j\|_F^2 + \|\boldsymbol{x}_k\boldsymbol{R}_j\|_F^2$$

Since $\boldsymbol{x}_i\boldsymbol{R}_j$ and $\boldsymbol{x}_k\boldsymbol{R}_j$ are all vectors, we can have $\|\boldsymbol{x}_i\boldsymbol{R}_j\boldsymbol{x}_k\boldsymbol{R}_j\| = \|\boldsymbol{x}_k\boldsymbol{R}_j\boldsymbol{x}_i\boldsymbol{R}_j\|$. Then the equality (a) holds.

Then in the same manner with Eq.(18), we can have that

$$\min \frac{1}{2\sqrt{|\mathcal{R}|}} \sum_{(\boldsymbol{x}_i,\boldsymbol{R}_j,\boldsymbol{x}_k)\in\mathcal{S}} ||\boldsymbol{x}_i||_F^2 + ||\boldsymbol{x}_k||_F^2 + \|\boldsymbol{x}_i\mathbf{R}_j + \boldsymbol{x}_k\mathbf{R}_j\|_F^2 = \|\hat{\mathcal{X}}\|_2.$$

We can have the equality holds if and only if $\|\mathbf{q}_{:d}\|_2\|\mathbf{r}_{:d}\|_2 = \sqrt{|\mathcal{R}|}\|\mathbf{p}_{:d}\|_2$. Then we can see that the conclusion holds if and only if $\|\mathbf{p}_{:d}\|_2\|\mathbf{r}_{:d}\|_2 = \sqrt{|\mathcal{R}|}\|\mathbf{q}_{:d}\|_2$ and $||\boldsymbol{q}_{:d}||_2||\boldsymbol{r}_{:,d}||_2 = \sqrt{|\mathcal{R}|}\|\mathbf{p}_{:d}\|_2, \forall d \in \{1,2,\ldots,D\}$. Then the proof of Theorem 2 completes.

Proof for Theorem 3

Here we denote $|\boldsymbol{x}|^3$ as $||\boldsymbol{x}||_3^3$ and denote $|\boldsymbol{x}|^{\frac{3}{2}}$ as $||\boldsymbol{x}||_C$. Then we have Theorem 3 and Theorem 4 with their proofs as follows:

Theorem 3 Suppose that $\hat{\boldsymbol{X}} = \boldsymbol{x}_i\overline{\boldsymbol{R}}_j\boldsymbol{x}_k^\top$ for $j = 1,2,\cdots,|\mathcal{R}|$, where $\boldsymbol{x}_i,\boldsymbol{x}_k,\boldsymbol{R}_j$ are real matrices and $\boldsymbol{R}_j$ is diagonal. Then, the following equation holds

$$\frac{1}{\sqrt{|\mathcal{R}|}}\min \sum_{(\boldsymbol{x}_i,\boldsymbol{R}_j,\boldsymbol{x}_k)\in\mathcal{S}} ||\boldsymbol{x}_i||_3^3 + ||\boldsymbol{x}_k||_3^3 + \|\boldsymbol{x}_i\mathbf{R}_j - \boldsymbol{x}_k\mathbf{R}_j\|_3^3 = \|\hat{\mathcal{X}}\|_3 \tag{19}$$

Proof *Notice that*

$$\min \sum_{(\boldsymbol{x}_i,\boldsymbol{R}_j,\boldsymbol{x}_k)\in\mathcal{S}} ||\boldsymbol{x}_i||_3^3 + ||\boldsymbol{x}_k||_3^3 + \|\boldsymbol{x}_i\mathbf{R}_j - \boldsymbol{x}_k\mathbf{R}_j\|_3^3$$

$$\leq \min \sum_{(\boldsymbol{x}_i,\boldsymbol{R}_j,\boldsymbol{x}_k)\in\mathcal{S}} ||\boldsymbol{x}_i||_3^3 + ||\boldsymbol{x}_k||_3^3 + \|\boldsymbol{x}_i\mathbf{R}_j\|_3^3 + \|\boldsymbol{x}_k\boldsymbol{R}_j\|_3^3.$$

We first prove that the following equation holds

$$\min_{\hat{X}_j=\boldsymbol{x}_iR_jx_k^\top} \frac{1}{\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} \left(\|\boldsymbol{x}_i\|_3^3 + \|\boldsymbol{x}_k\|_3^3 + \|\boldsymbol{x}_i\boldsymbol{R}_j\|_3^3 + \|\boldsymbol{x}_k\boldsymbol{R}_j\|_3^3\right) = \|\hat{\mathcal{X}}\|_3. \tag{20}$$

Denote $\boldsymbol{x}_i$ as $\mathbf{p}$ and $\boldsymbol{x}_k$ as $\mathbf{q}$. Then we have that

$$\begin{aligned}
&\sum_{j=1}^{|\mathcal{R}|}\left(\|\boldsymbol{x}_i\mathbf{R}_j\|_3^3 + \|\boldsymbol{x}_k\|_3^3\right)\\
=&\sum_{j=1}^{|\mathcal{R}|}\left(\sum_{d=1}^{D}\|\mathbf{q}_{:d}\|_3^3 + \sum_{i=1}^{I}\sum_{d=1}^{D}\mathbf{p}_{id}^3\mathbf{r}_{jd}^3\right)\\
=&\sum_{j=1}^{|\mathcal{R}|}\sum_{d=1}^{D}\|\mathbf{q}_{:d}\|_3^3 + \sum_{d=1}^{D}\|\mathbf{p}_{:d}\|_3^3\|\mathbf{r}_{:d}\|_3^3\\
=&\sum_{d=1}^{D}\left(\|\mathbf{p}_{:d}\|_3^3\|\mathbf{r}_{:d}\|_3^3 + |\mathcal{R}|\|\mathbf{q}_{:d}\|_3^3\right)\\
\geq&\sum_{d=1}^{D}2\sqrt{|\mathcal{R}|}\|\mathbf{p}_{:d}\|_C\|\mathbf{r}_{:d}\|_C\|\mathbf{q}_{:d}\|_C\\
=&2\sqrt{|\mathcal{R}|}\sum_{d=1}^{D}\|\mathbf{p}_{:d}\|_C\|\mathbf{r}_{:d}\|_C\|\mathbf{q}_{:d}\|_C
\end{aligned}.$$

We can have the equality holds if and only if $\|\mathbf{p}_{:d}\|_3^3\|\mathbf{r}_{:d}\|_3^3 = |\mathcal{R}|\|\mathbf{q}_{:d}\|_3^3$, i.e., $\|\mathbf{p}_{:d}\|_C\|\mathbf{r}_{:d}\|_C = \sqrt{|\mathcal{R}|}\|\mathbf{q}_{:d}\|_C$.

For all CP decomposition $\hat{\mathcal{X}} = \sum_{d=1}^{D}\mathbf{p}_{:d}\otimes\mathbf{r}_{:d}\otimes\mathbf{q}_{:d}$, we can always let $\boldsymbol{p}'_{:d} = \boldsymbol{p}_{:d}$, $\mathbf{r}'_{:d} = \sqrt{\frac{\|\mathbf{q}_d\|_C\sqrt{|\mathcal{R}|}}{\|\mathbf{p}_{:d}\|_C\|\mathbf{r}_{:d}\|_C}}\mathbf{r}_{:d}$ and $\boldsymbol{q}'_{:d} = \sqrt{\frac{\|\mathbf{p}_{:d}\|_C\|\mathbf{r}_{:d}\|_C}{\|\mathbf{q}_{\cdot d}\|_C\sqrt{|\mathcal{R}|}}}\mathbf{p}_{:d}$, then we have

$$\|\mathbf{p}'_{:d}\|_C\|\mathbf{r}'_{:d}\|_C = \sqrt{|\mathcal{R}|}\|\mathbf{q}'_{:d}\|_C,$$

and at the same time we make sure that that $\hat{\mathcal{X}} = \sum_{d=1}^{D}\mathbf{p}'_{:d}\otimes\mathbf{r}'_{:d}\otimes\mathbf{q}'_{:d}$. Therefore, we can have

$$\begin{aligned}
&\frac{1}{\sqrt{|\mathcal{R}|}}\sum_{j=1}^{|\mathcal{R}|}\left\|\hat{\mathcal{X}}_j\right\|_C\\
=&\frac{1}{2\sqrt{|\mathcal{R}|}}\sum_{j=1}^{|\mathcal{R}|}\min_{\hat{\mathcal{X}}_j=\boldsymbol{x}_i\mathbf{R}_j\boldsymbol{x}_k^\top}\left(\|\boldsymbol{x}_i\mathbf{R}_j\|_3^3 + \|\boldsymbol{x}_k\|_3^3\right)\\
\leq&\frac{1}{2\sqrt{|\mathcal{R}|}}\min_{\hat{\mathcal{X}}_j=\boldsymbol{x}_i\mathbf{R}_j\boldsymbol{x}_k^\top}\sum_{j=1}^{|\mathcal{R}|}\left(\|\boldsymbol{x}_i\mathbf{R}_j\|_3^3 + \|\boldsymbol{x}_k\|_3^3\right)\\
=&\min_{\hat{\mathcal{X}}=\sum_{d=1}^{D}\mathbf{p}_j\otimes\mathbf{r}_{id}\otimes\mathbf{q}_{:d}}\sum_{d=1}^{D}\|\mathbf{p}_{:d}\|_C\|\mathbf{r}_{:d}\|_C\|\mathbf{q}_{:d}\|_C\\
=&\|\hat{\mathcal{X}}\|_3.
\end{aligned}$$

And in the same manner, we can have that

$$\min \frac{1}{\sqrt{|\mathcal{R}|}} \sum_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} ||\boldsymbol{x}_i||_3^3 + ||\boldsymbol{x}_k||_3^3 + \|\boldsymbol{x}_i \mathbf{R}_j - \boldsymbol{x}_k \mathbf{R}_j\|_3^3 = \|\hat{\mathcal{X}}\|_3.$$

We can have that the equality holds if and only if $\|\mathbf{q}_{:d}\|_C \|\mathbf{r}_{:d}\|_C = \sqrt{|\mathcal{R}|}\, \|\mathbf{p}_{:d}\|_C$. *Then we can see that the conclusion holds if and only if* $\|\mathbf{p}_{:d}\|_C \|\mathbf{r}_{:d}\|_C = \sqrt{|\mathcal{R}|}\, \|\mathbf{q}_{:d}\|_C$ *and* $||\boldsymbol{q}_{:d}||_C ||\boldsymbol{r}_{:,d}||_C = \sqrt{|\mathcal{R}|}\, \|\mathbf{p}_{:d}\|_C, \forall d \in \{1, 2, \dots, D\}$. *Then the proof of Theorem 3 completes.*

Proof for Theorem 4

Theorem 4 Suppose that $\hat{\boldsymbol{X}} = \boldsymbol{x}_i \overline{\boldsymbol{R}}_j \boldsymbol{x}_k^\top$ for $j = 1, 2, \cdots, |\mathcal{R}|$, where $\boldsymbol{x}_i, \boldsymbol{x}_k, \boldsymbol{R}_j$ are real matrices and $\boldsymbol{R}_j$ is diagonal. Then, the following equation holds

$$\frac{1}{4\sqrt{|\mathcal{R}|}} \min \sum_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} ||\boldsymbol{x}_i||_3^3 + ||\boldsymbol{x}_k||_3^3 + \|\boldsymbol{x}_i \mathbf{R}_j + \boldsymbol{x}_k \mathbf{R}_j\|_3^3 = \|\hat{\mathcal{X}}\|_3 \tag{21}$$

First we prove $\|\boldsymbol{x}_i \mathbf{R}_j\|^2 |\boldsymbol{x}_k \mathbf{R}_j| + |\boldsymbol{x}_i \mathbf{R}_j| ||\boldsymbol{x}_k \mathbf{R}_j||^2 \leq \|\boldsymbol{x}_i \mathbf{R}_j\|_3^3 + \|\boldsymbol{x}_k \mathbf{R}_j\|_3^3$ holds. Notice that $|\boldsymbol{x}_i \mathbf{R}_j| + |\boldsymbol{x}_k \mathbf{R}_j| \geq 0$ and $(|\boldsymbol{x}_i \mathbf{R}_j| - |\boldsymbol{x}_k \mathbf{R}_j|)^2 \geq 0$. Then we have

$$\begin{aligned}
(|\boldsymbol{x}_i \mathbf{R}_j| + |\boldsymbol{x}_k \mathbf{R}_j|)(|\boldsymbol{x}_i \mathbf{R}_j| - |\boldsymbol{x}_k \mathbf{R}_j|)^2 &\geq 0 \\
(|\boldsymbol{x}_i \mathbf{R}_j| + |\boldsymbol{x}_k \mathbf{R}_j|)(|\boldsymbol{x}_i \mathbf{R}_j| - |\boldsymbol{x}_k \mathbf{R}_j|)(|\boldsymbol{x}_i \mathbf{R}_j| - |\boldsymbol{x}_k \mathbf{R}_j|) &\geq 0 \\
(|\boldsymbol{x}_i \mathbf{R}_j|^2 - |\boldsymbol{x}_k \mathbf{R}_j|^2)(|\boldsymbol{x}_i \mathbf{R}_j| - |\boldsymbol{x}_k \mathbf{R}_j|) &\geq 0
\end{aligned} \tag{22}$$

Then we can derive the following formula:

$$\begin{aligned}
|\boldsymbol{x}_i \mathbf{R}_j|^2 (|\boldsymbol{x}_i \mathbf{R}_j| - |\boldsymbol{x}_k \mathbf{R}_j|) &\geq |\boldsymbol{x}_k \mathbf{R}_j|^2 (|\boldsymbol{x}_i \mathbf{R}_j| - |\boldsymbol{x}_k \mathbf{R}_j|) \\
|\boldsymbol{x}_i \mathbf{R}_j|^3 - |\boldsymbol{x}_i \mathbf{R}_j|^2 |\boldsymbol{x}_k \mathbf{R}_j| &\geq |\boldsymbol{x}_k \mathbf{R}_j|^2 |\boldsymbol{x}_i \mathbf{R}_j| - |\boldsymbol{x}_i \mathbf{R}_j|^3 \\
\|\boldsymbol{x}_i \mathbf{R}_j\|_3^3 + \|\boldsymbol{x}_k \mathbf{R}_j\|_3^3 &\geq \|\boldsymbol{x}_i \mathbf{R}_j\|^2 |\boldsymbol{x}_k \mathbf{R}_j^\top| + |\boldsymbol{x}_i \mathbf{R}_j| ||\boldsymbol{x}_k \mathbf{R}_j||^2
\end{aligned} \tag{23}$$

Then we have:

$$\begin{aligned}
&\min \frac{1}{4\sqrt{|\mathcal{R}|}} \sum_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} ||\boldsymbol{x}_i||_3^3 + ||\boldsymbol{x}_k||_3^3 + \|\boldsymbol{x}_i \mathbf{R}_j + \boldsymbol{x}_k \mathbf{R}_j\|_3^3 \\
\overset{a}{\leq} &\min_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} \frac{1}{4\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} \|\boldsymbol{x}_i\|_3^3 + \|\boldsymbol{x}_k\|_3^3 + \|\boldsymbol{x}_i \boldsymbol{R}_j\|_3^3 \\
&+ \|\boldsymbol{x}_k \boldsymbol{R}_j\|_3^3 + 3\|\boldsymbol{x}_i \mathbf{R}_j\|^2 |\boldsymbol{x}_k \mathbf{R}_j| + 3|\boldsymbol{x}_i \mathbf{R}_j| ||\boldsymbol{x}_k \mathbf{R}_j||^2 \\
\overset{b}{\leq} &\min_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} \frac{1}{4\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} \|\boldsymbol{x}_i\|_3^3 + \|\boldsymbol{x}_k\|_3^3 + \|\boldsymbol{x}_i \boldsymbol{R}_j\|_3^3 \\
&+ \|\boldsymbol{x}_k \boldsymbol{R}_j\|_3^3 + 3\, \|\boldsymbol{x}_i \boldsymbol{R}_j\|_3^3 + 3\|\boldsymbol{x}_k \boldsymbol{R}_j\|_3^3 \\
\leq &\min_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} \frac{1}{4\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} \|\boldsymbol{x}_i\|_3^3 + \|\boldsymbol{x}_k\|_3^3 + 4\, \|\boldsymbol{x}_i \boldsymbol{R}_j\|_3^3 \\
&+ 4\|\boldsymbol{x}_k \boldsymbol{R}_j\|_3^3 \\
\leq &\min_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} \frac{1}{\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} \|\boldsymbol{x}_i\|_3^3 + \|\boldsymbol{x}_k\|_3^3 + \|\boldsymbol{x}_i \boldsymbol{R}_j\|_3^3 + \|\boldsymbol{x}_k \boldsymbol{R}_j\|_3^3
\end{aligned}$$

Datasets	WN18RR	FB15K237	YAGO3-10
dimension	2000	2000	2000
batch size	100	100	500
learning rate	0.1	0.05	0.1

Table 2: Hyperparameters found by grid search for CP mdoel.

Datasets	WN18RR	FB15K237	YAGO3-10
dimension	2000	2000	2000
batch size	200	200	1000
learning rate	0.05	0.1	0.05

Table 3: Hyperparameters found by grid search for ComplEx mdoel.

Since $\boldsymbol{x}_i \boldsymbol{R}_j$ and $\boldsymbol{x}_k \boldsymbol{R}_j$ are all vectors, we can have $\|\boldsymbol{x}_i \boldsymbol{R}_j \boldsymbol{x}_k \boldsymbol{R}_j\| = \|\boldsymbol{x}_k \boldsymbol{R}_j \boldsymbol{x}_i \boldsymbol{R}_j\|$. Then the inequality (a) holds. The inequality (b) holds due to the Eq.(22).

Then in the same manner with Eq.(20), we can have that

$$\min \frac{1}{4\sqrt{|\mathcal{R}|}} \sum_{(\boldsymbol{x}_i, \boldsymbol{R}_j, \boldsymbol{x}_k) \in \mathcal{S}} ||\boldsymbol{x}_i||_3^3 + ||\boldsymbol{x}_k||_3^3 + \|\boldsymbol{x}_i \mathbf{R}_j + \boldsymbol{x}_k \mathbf{R}_j\|_3^3 = \|\hat{\mathcal{X}}\|_3.$$

We can have that the equality holds if and only if $\|\mathbf{q}_{:d}\|_C \|\mathbf{r}_{:d}\|_C = \sqrt{|\mathcal{R}|}\, \|\mathbf{p}_{:d}\|_C$. Then we can see that the conclusion holds if and only if $\|\mathbf{p}_{:d}\|_C \|\mathbf{r}_{:d}\|_C = \sqrt{|\mathcal{R}|}\, \|\mathbf{q}_{:d}\|_C$ and $||\boldsymbol{q}_{:d}||_C ||\boldsymbol{r}_{:,d}||_C = \sqrt{|\mathcal{R}|}\, \|\mathbf{p}_{:d}\|_C, \forall d \in \{1, 2, \dots, D\}$. Then the proof of Theorem 4 completes.

Experimental Details and Appendix

We implement our model using PyTorch and test it on a single GPU. Here Table 7 shows statistics of the datasets used in this paper. The hypermeters for CP, ComplEx, RESCAL RotatE models are shown in Table 2, Table 3, Table 4 and Table 5 respectively. We have counted the running time of each epoch for different models with ER in WN18RR as follows: CP with ER takes 58s, ComplEx with ER takes 84s and RESCAL with ER takes 73s.

Study on semantic-similarity hyperparameter ϵ. In the experiments above, we provide the semantic-similarity parameter ϵ_j for each relation $\boldsymbol{R}_j$ in ER. To characterize the similarity between entities adequately and study the impact of ϵ, here we also conduct another version of ER where we provide ϵ_{ik} for a_{ik} (in Eq.(3)), which we denote as $ER*$. From Table 6, we can see ER and $ER*$ have similar performance. It shows providing ϵ_j for each relation $\boldsymbol{R}_j$ in ER is proper.

Models		WN18RR			FB15K-237			YAGO3-10	
	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10
CP-FRO	.460	-	.480	.340	-	.510	.540	-	.680
CP-N3	.470	.430	.544	.354	.261	.544	.577	.505	.705
CP-DURA	.478	.441	.552	.367	.272	.555	.579	.506	.709
CP-ER	**.482**	**.444**	**.557**	**.371**	**.275**	**.561**	**.584**	**.508**	**.712**
ComplEx-FRO	.470	-	.540	.350	-	.530	.573	-	.710
ComplEx-N3	.489	.443	.580	.366	.271	.558	.577	.502	.711
ComplEx-DURA	.491	.449	.571	.371	.276	.560	.584	.511	.713
ComplEx-ER	**.494**	**.453**	**.575**	**.374**	**.282**	**.563**	**.588**	**.515**	**.718**
RESCAL-FRO	.397	.363	.452	.323	.235	.501	.474	.392	.628
RESCAL-DURA	.498	.455	.577	.368	.276	.550	.579	.505	.712
RESCAL-ER	**.499**	**.458**	**.582**	**.373**	**.281**	**.554**	**.583**	**.509**	**.715**

Table 9: Comparison between DURA, the squared Frobenius norm (FRO), and the nuclear 3-norm (N3) regularizers (N3 does not apply to RESCAL). The best performance on each model are marked in bold.

Models		WN18RR			FB15K-237			YAGO3-10	
	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10
TransE-FRO	.259	.105	.532	.327	.231	.519	.478	.377	.665
TransE-N3	.265	.107	.533	.328	.232	.518	.483	**.385**	.664
TransE-DURA	.260	.105	.531	.328	.233	.518	.475	.371	.666
TransE-ER	**.268**	**.110**	**.536**	**.329**	**.235**	**.525**	**.489**	.384	**.669**
RotatE-FRO	.481	.434	.572	.337	.242	.528	.570	.481	.680
RotatE-N3	.483	.440	.580	346	.251	.538	.574	.498	.701
RotatE-DURA	.487	.443	.580	.342	.246	.533	.567	.491	.702
RotatE-ER	**.490**	**.445**	**.581**	**.352**	**.255**	**.547**	**.581**	**.505**	**.704**

Table 10: Comparison between DURA, the squared Frobenius norm (FRO), and the nuclear 3-norm (N3) regularizers. The best performance on each model are marked in bold.

Datasets	WN18RR	FB15K237	YAGO3-10
dimension	512	512	512
batch size	400	400	1000
learning rate	0.1	0.1	0.05

Table 4: Hyperparameters found by grid search for RESCAL mdoel.

Datasets	WN18RR	FB15K237	YAGO3-10
dimension	400	400	400
batch size	100	100	500
learning rate	0.1	0.05	0.05

Table 5: Hyperparameters found by grid search for RotatE mdoel.

Model	MRR	Hits@1	Hits@10
ComplEx-ER	.374	.282	.563
ComplEx-ER*	.375	.282	.565

Table 6: Evaluation results of ϵ on FB15K237.

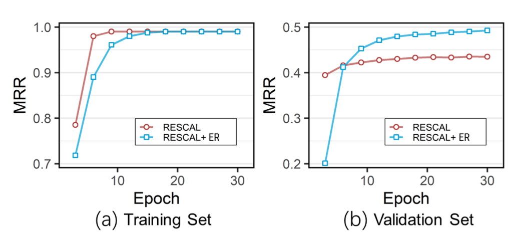

Figure 3: Study of training and validation curves.

Models	WN18RR MRR	Hits@1	Hits@10	FB15K-237 MRR	Hits@1	Hits@10	YAGO3-10 MRR	Hits@1	Hits@10
ComplEx-RHE	.469	.430	.538	.348	.262	.542	.570	.501	.708
ComplEx-LLE	.477	.442	.551	.363	.271	.552	.576	.504	.701
ComplEx-TFR	.473	.441	.545	.358	.264	.541	.573	.502	.702
ComplEx-EIA	.463	.345	.542	.356	.266	.529	.573	.501	.703
ComplEx-Pretrain	.479	.440	.553	.353	.268	.533	.578	.502	.704
ComplEx-ER	.494	.453	.575	.374	.282	.563	.588	.515	.718

Table 11: Evaluation results of different models on WN18RR, FB15k-237 and YAGO3-10 datasets.

Models	WN18RR MRR	Hits@1	Hits@10	FB15K-237 MRR	Hits@1	Hits@10	YAGO3-10 MRR	Hits@1	Hits@10
CP-ER	.479	.441	.556	.371	.273	.560	.582	.506	.709
ComplEx-ER	.492	.452	.574	.371	.275	.560	.586	.514	.712

Table 12: Evaluation results of ER based on 3-norm on WN18RR, FB15k-237 and YAGO3-10 datasets.

Dataset	#Entity	#Relation	#Training	#Valid	#Test
WN18RR	40,943	11	86,835	3,034	3,134
FB15K237	14,541	237	272,115	17,535	20,466
YAGO3-10	123,182	37	1,079,040	5,000	5,000

Table 7: Statistics of the datasets used in this paper.

Model	MRR	Hits@1	Hits@10
ComplEx_0	.355	.263	.542
ComplEx_1	.374	.282	.563
ComplEx_2	.378	.284	.569

Table 8: Evaluation results on FB15K237.