Supplementary Materials
for Generalization Error Bound for Hyperbolic Ordinal Embedding

A. Proof of Lemma 4

Proof. (The first equation) Let $\boldsymbol{Z} := \begin{bmatrix} \boldsymbol{z}_1 & \boldsymbol{z}_2 \ldots \boldsymbol{z}_N \end{bmatrix} \in \mathbb{R}^{D,N}$, and $\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ be the singular value decomposition of $\boldsymbol{Z}$. Regarding the Gramian matrix, we can obtain the singular value decomposition of $\boldsymbol{G}$ as follows.

$$\begin{aligned} \boldsymbol{G} &= \boldsymbol{Z}^\top \boldsymbol{Z} \\ &= \boldsymbol{V}\boldsymbol{\Sigma}^\top \boldsymbol{U}^\top \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \\ &= \boldsymbol{V}\boldsymbol{\Sigma}^2 \boldsymbol{V}^\top. \end{aligned} \tag{37}$$

Hence, we have $\|\boldsymbol{G}\|_* = \mathrm{Tr}(\boldsymbol{\Sigma}^2)$. Therefore, we have that

$$\begin{aligned} \|\boldsymbol{G}\|_* &= \mathrm{Tr}(\boldsymbol{\Sigma}^2) \\ &= \mathrm{Tr}\Big(\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma}\boldsymbol{V}^\top \boldsymbol{V}\Big) \\ &= \mathrm{Tr}\Big(\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^\top \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\Big) \\ &= \mathrm{Tr}\Big(\boldsymbol{Z}^\top \boldsymbol{Z}\Big) \\ &= \mathrm{Tr}(\boldsymbol{G}) \\ &= \sum_{n=1}^{N} [\boldsymbol{G}]_{n,n} \\ &= \sum_{n=1}^{N} \boldsymbol{z}_n^\top \boldsymbol{z}_n \\ &= \sum_{n=1}^{N} \left(\Delta_{\mathbb{R}^D}(\boldsymbol{0}, \boldsymbol{z}_n)\right)^2, \end{aligned} \tag{38}$$

which completes the proof of the first equation.

(The second equation) For any $i, j \in [N]$, we have that

$$\begin{aligned} [\boldsymbol{G}]_{i,j} &= \boldsymbol{z}_i^\top \boldsymbol{z}_j \\ &\le \|\boldsymbol{z}_i\|_2 \|\boldsymbol{z}_j\|_2 \\ &\le \max\left\{\|\boldsymbol{z}_i\|_2^2, \|\boldsymbol{z}_j\|_2^2\right\} \\ &\le \max_{n\in[N]} \|\boldsymbol{z}_n\|_2^2 \\ &= \max_{n\in[N]} [\boldsymbol{G}]_{n,n} \\ &\le \max_{n\in[N]} \left(\Delta_{\mathbb{R}^D}(\boldsymbol{0}, \boldsymbol{z}_n)\right)^2, \end{aligned} \tag{39}$$

where the first inequality holds from the Cauchy Schwartz inequality. Hence, we have $\max_{i,j\in[N]} [\boldsymbol{G}]_{i,j} \le \max_{n\in[N]} [\boldsymbol{G}]_{n,n}$. Conversely, obviously, $\max_{i,j\in[N]} [\boldsymbol{G}]_{i,j} \ge \max_{n\in[N]} [\boldsymbol{G}]_{n,n}$ is valid. Therefore, we have that $\max_{i,j\in[N]} [\boldsymbol{G}]_{i,j} = \max_{n\in[N]} [\boldsymbol{G}]_{n,n}$. Since the left hand side equals $\|\boldsymbol{G}\|_{\max}$, and right hand side equals $\max_{n\in[N]} \left(\Delta_{\mathbb{R}^D}(\boldsymbol{0}, \boldsymbol{z}_n)\right)^2$, we have that $\|\boldsymbol{G}\|_{\max} = \max_{n\in[N]} \left(\Delta_{\mathbb{R}^D}(\boldsymbol{0}, \boldsymbol{z}_n)\right)^2$, which completes the proof of the second equation.

□

B. Proof of Lemma 6

Proof. Define

$$\begin{aligned} &\mu := \min\{\Delta^*(i,j) | i \neq j\}, \\ &\xi := \min\left\{\left|1 - \frac{\Delta^*(i,j)}{\Delta^*(i',j')}\right| \middle| \begin{array}{l} (i,j),(i',j') \in [N]\times[N], \\ i \neq j, i' \neq j', (i,j) \neq (i',j') \end{array}\right\}. \end{aligned} \tag{40}$$

We assume that $\mu, \xi > 0$ holds as in the discussion in Section 2.1. Let $\epsilon := \frac{1}{3}\xi$. Let $\nu, \eta_{\max}$ be the constants determined on the weighted graph defined by (Sarkar, 2011). Let $\tau := \max\left\{\eta_{\max}, \frac{\nu(1+\epsilon)}{\mu\epsilon}, \frac{1}{\mu\epsilon}\right\}$. Then by the $(1+\epsilon)$-distortion algorithm, we can obtain representations $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_N \in \mathbb{L}^2$ such that

$$(1-\epsilon)\tau\Delta^*(i,j) < \Delta_{\mathbb{L}^2}(\boldsymbol{z}_i, \boldsymbol{z}_j) \le (1+\epsilon)\tau\Delta^*(i,j), \tag{41}$$

for any $i, j \in [N]$. Here, the following is valid for $i, j, i', j' \in [N]$: if $\Delta^*(i,j) > \Delta^*(i',j')$, then

$$\begin{aligned} &\Delta_{\mathbb{L}^2}(\boldsymbol{z}_i, \boldsymbol{z}_j) - \Delta_{\mathbb{L}^2}(\boldsymbol{z}_{i'}, \boldsymbol{z}_{j'}) \\ &\ge \tau(1-\epsilon)\Delta^*(i,j) - (1+\epsilon)\Delta^*(i',j') \\ &\ge \tau\Delta^*(i,j)\left[(1-\epsilon) - (1+\epsilon)\frac{\Delta^*(i',j')}{\Delta^*(i,j)}\right] \\ &\ge \tau\Delta^*(i,j)\left[\left(1 - \frac{1}{3}\xi\right) - \left(1 + \frac{1}{3}\xi\right)(1-\xi)\right] \\ &> \tau\Delta^*(i,j)\frac{1}{3}\xi \\ &\ge \tau\mu\epsilon \\ &\ge 1. \end{aligned} \tag{42}$$

□

C. Proof of Lemma 7

Proof. Let c be the center of the 6-star subgraph and $n_1, n_2, \ldots, n_6$ be its neighborhood. Define $\Delta_m^* := \Delta^*(c, n_m)$. Let $B_\Delta(\boldsymbol{z})$ be an open ball of radius Δ centered at $\boldsymbol{z}$ in $\mathbb{R}^2$. Assume that $\boldsymbol{z}_c, \boldsymbol{z}_{n_1}, \ldots, \boldsymbol{z}_{n_6} \in \mathbb{R}^2$ satisfies (2) and define $\Delta_m := \Delta_{\mathbb{R}^2}(\boldsymbol{z}_c, \boldsymbol{z}_{n_m})$. For m, m' such that $\Delta_m^* < \Delta_{m'}^*$, $\boldsymbol{z}_{n_m} \in B_{\Delta_{m'}}(\boldsymbol{z}_c)$ and $\boldsymbol{z}_{n_m} \notin B_{\Delta_{m'}}(\boldsymbol{z}_{n_{m'}})$

are necessary. Hence, $\angle \boldsymbol{z}_{n_m} \boldsymbol{z}_c \boldsymbol{z}_{n_{m'}} > 60°$. Thus, segments $\boldsymbol{z}_c\boldsymbol{z}_{n_1}, \boldsymbol{z}_c\boldsymbol{z}_{n_2}, \dots, \boldsymbol{z}_c\boldsymbol{z}_{n_6}$ partition $360°$ into 6 angles larger than $60°$, which is contradiction. □

D. Proof of Proposition 8

Proof. If $a, b, c > 0$ and assume $x > 0$, then we have that

$$ax^2 + bx < c \Leftrightarrow 0 < x < -b + \sqrt{b^2 + 4ac}. \tag{43}$$

We are going to find S that satisfies

$$\begin{aligned} &O\left(L_\phi (\exp R)^2 \left(\sqrt{\frac{N \ln N}{S}} + \frac{N \ln N}{S} + \sqrt{\frac{\ln \frac{2}{\delta}}{S}}\right)\right) \\ &< 2\alpha \frac{V_{\min}^{\mathbb{R}^2}}{|\mathcal{T}|}, \end{aligned} \tag{44}$$

where $L_\phi = 1$. This corresponds to $x = \frac{1}{\sqrt{S}}$, $a = O(N \ln N)$, $b = O\left(\sqrt{N \ln N} + \sqrt{\ln \frac{2}{\delta}}\right)$, $c = O\left(\frac{2\alpha V_{\min}^{\mathbb{R}^2}}{(\exp R)^2 |\mathcal{T}|}\right)$. We can formulate the condition as follows:

$$\begin{aligned} &0 < \frac{1}{\sqrt{S}} < -b + \sqrt{b^2 + 4ac} \\ &\Leftrightarrow S > \frac{1}{\left(-b + \sqrt{b^2 + 4ac}\right)^2}. \end{aligned} \tag{45}$$

Here, we have that

$$\begin{aligned} &\frac{1}{\left(-b + \sqrt{b^2 + 4ac}\right)^2} \\ &= \frac{\left(b + \sqrt{b^2 + 4ac}\right)^2}{(4ac)^2} \\ &\leq \frac{\left(b + b\left(1 + \frac{2ac}{b^2}\right)\right)^2}{(4ac)^2} \\ &= \left(\frac{b\left(2 + \frac{2ac}{b^2}\right)}{4ac}\right)^2 \\ &= \left(\frac{b}{2ac} + \frac{1}{2b}\right)^2. \\ &\leq \left(\frac{b}{2ac} + \frac{1}{2b}\right)^2. \\ &= O\left(\frac{(\exp R)^2 |\mathcal{T}|}{\alpha V_{\min}^{\mathbb{R}^2} N \ln N} \left(\sqrt{N \ln N} + \sqrt{\ln \frac{2}{\delta}}\right)\right. \\ &\qquad \left. + \frac{1}{\sqrt{N \ln N} + \sqrt{\ln \frac{2}{\delta}}}\right)^2, \end{aligned} \tag{46}$$

which completes the proof. Here, the first inequality holds since $\sqrt{1 + y} \leq 1 + \frac{1}{2} y$. □

E. Proof of Lemma 9

Proof. The statement (iv) follows from (ii) and (iii). Therefore, we prove (i)-(iii) in the following.

(Sufficiency) (i) Assume that $(\boldsymbol{z}_n)_{n=1}^N \in \left(\mathbb{L}^D\right)^N$ is valid. For $d = 0, 1, \dots, D$ and $n = 1, 2, \dots, N$, we denote the d-th element of $\boldsymbol{z}_n \in \mathbb{L}^D$ by $z_{d,n}$, and for $n = 1, 2, \dots, N$, we define $\boldsymbol{z}_n^- \in \mathbb{R}^1$ and $\boldsymbol{z}_n^+$ by

$$\boldsymbol{z}_n^- := \begin{bmatrix} z_{0,n} \end{bmatrix}, \boldsymbol{z}_n^+ := \begin{bmatrix} z_{1,n} & z_{2,n} & \cdots & z_{D,n} \end{bmatrix}^\top. \tag{47}$$

Also, we define $\boldsymbol{Z}^- \in \mathbb{R}^{1,N}$ and $\boldsymbol{Z}^+ \in \mathbb{R}^{D-1,N}$ by

$$\begin{aligned} \boldsymbol{Z}^- &:= \begin{bmatrix} \boldsymbol{z}_1^- & \boldsymbol{z}_2^- & \cdots & \boldsymbol{z}_N^- \end{bmatrix}, \\ \boldsymbol{Z}^+ &:= \begin{bmatrix} \boldsymbol{z}_1^+ & \boldsymbol{z}_2^+ & \cdots & \boldsymbol{z}_N^+ \end{bmatrix}, \end{aligned} \tag{48}$$

respectively. Define $\boldsymbol{L}^-, \boldsymbol{L}^+ \in \mathbb{R}^{N,N}$ by $\boldsymbol{L}^- := \left(\boldsymbol{Z}^-\right)^\top \boldsymbol{Z}^-$ and $\boldsymbol{L}^+ := \left(\boldsymbol{Z}^+\right)^\top \boldsymbol{Z}^+$, respectively. For all $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{x}^\top \boldsymbol{L}^- \boldsymbol{x} = \left(\boldsymbol{L}^- \boldsymbol{x}\right)^\top \boldsymbol{L}^- \boldsymbol{x} \geq 0$. Therefore, we have $\boldsymbol{L}^- \succeq \mathbf{O}$. Likewise, $\boldsymbol{L}^+ \succeq \mathbf{O}$ is valid, and thus we obtain **(a)**. Because $\boldsymbol{Z}^- \in \mathbb{R}^{1,N}$ and $\boldsymbol{Z}^+ \in \mathbb{R}^{D-1,N}$, we have $\operatorname{rank} \boldsymbol{L}^- = 1$ and $\operatorname{rank} \boldsymbol{L}^+ \leq D$, respectively. As $\boldsymbol{z}_n \in \mathbb{L}^D$, $\boldsymbol{z}_{0,n} \geq 1$ is valid, $\operatorname{rank} \boldsymbol{L}^- \neq 0$, and therefore $\operatorname{rank} \boldsymbol{L}^- = 1$. Thus, we have **(b)**. If $i, j \in [N]$, then the following inequality holds:

$$\begin{aligned} &\left[\boldsymbol{L}^+ - \boldsymbol{L}^-\right]_{i,j} \\ &= \left(\boldsymbol{z}_i^+\right)^\top \boldsymbol{z}_j^+ - \left(\boldsymbol{z}_i^-\right)^\top \boldsymbol{z}_j^- \\ &= \left(\boldsymbol{z}_i^+\right)^\top \boldsymbol{z}_j^+ - \sqrt{1 + \left(\boldsymbol{z}_i^+\right)^\top \left(\boldsymbol{z}_i^+\right)} \sqrt{1 + \left(\boldsymbol{z}_j^+\right)^\top \left(\boldsymbol{z}_j^+\right)} \\ &= \left(\boldsymbol{z}_i^+\right)^\top \boldsymbol{z}_j^+ - \left\| \begin{bmatrix} 1 & \left(\boldsymbol{z}_i^+\right)^\top \end{bmatrix}^\top \right\|_2 \left\| \begin{bmatrix} 1 & \left(\boldsymbol{z}_j^+\right)^\top \end{bmatrix}^\top \right\|_2 \\ &\leq \left(\boldsymbol{z}_i^+\right)^\top \boldsymbol{z}_j^+ - \begin{bmatrix} 1 & \left(\boldsymbol{z}_i^+\right)^\top \end{bmatrix} \begin{bmatrix} 1 & \left(\boldsymbol{z}_i^+\right)^\top \end{bmatrix}^\top \\ &= 1, \end{aligned} \tag{49}$$

where the inequality comes from the Cauchy Schwarz inequality, and the equality holds if $i = j$. These imply **(c)** and **(d)**, which completes the proof of the sufficiency in (i).

(ii) Assume $(\boldsymbol{z}_n)_{n=1}^N \in \mathcal{B}_R$, that is, for all $n \in [N]$, $\Delta_{\mathbb{L}^D}(\boldsymbol{z}_0, \boldsymbol{z}_n) \leq R$ is valid. Since $\mathcal{B}_R \subset \mathbb{L}^D$, conditions **(a)-(d)** holds true from the above discussion. Since $\Delta_{\mathbb{L}^D}(\boldsymbol{z}_0, \boldsymbol{z}_n) = \operatorname{arcosh}\left(-\langle \boldsymbol{z}_0, \boldsymbol{z}_n \rangle_{\mathrm{M}}\right) = \operatorname{arcosh}\left(z_{0,n}\right) = \operatorname{arcosh}\left(\sqrt{1 + \left(\boldsymbol{z}_n^+\right)^\top \boldsymbol{z}_n^+}\right)$, the followings are valid:

$$\begin{aligned} \left(\boldsymbol{z}_n^-\right)^\top \boldsymbol{z}_n^- &= \left(z_{0,n}\right)^2 = \cosh^2 \Delta_{\mathbb{L}^D}(\boldsymbol{z}_0, \boldsymbol{z}_n), \\ \left(\boldsymbol{z}_n^+\right)^\top \boldsymbol{z}_n^+ &= 1 + \cosh^2 \Delta_{\mathbb{L}^D}(\boldsymbol{z}_0, \boldsymbol{z}_n) = \sinh^2 \Delta_{\mathbb{L}^D}(\boldsymbol{z}_0, \boldsymbol{z}_n). \end{aligned} \tag{50}$$

Therefore, we have

$$\begin{aligned}\left[\boldsymbol{L}^{-}\right]_{n,n} &= \left(\boldsymbol{z}_n^{-}\right)^{\top}\boldsymbol{z}_n^{-} \leq \cosh^2 R \\ \left[\boldsymbol{L}^{+}\right]_{n,n}, &= \left(\boldsymbol{z}_n^{+}\right)^{\top}\boldsymbol{z}_n^{+} \leq \sinh^2 R.\end{aligned} \tag{51}$$

For all $i, j \in [N]$,

$$\begin{aligned}\left[\boldsymbol{L}^{-}\right]_{i,j} &= \left(\boldsymbol{z}_i^{-}\right)^{\top}\boldsymbol{z}_j^{-} \\ &\leq \left\|\boldsymbol{z}_i^{-}\right\|_2\left\|\boldsymbol{z}_j^{-}\right\|_2 \\ &\leq \max_{n=i,j}\left\{\left(\boldsymbol{z}_n^{-}\right)^{\top}\boldsymbol{z}_n^{-}\right\} \\ &\leq \max_{n\in[N]}\left\{\left(\boldsymbol{z}_n^{-}\right)^{\top}\boldsymbol{z}_n^{-}\right\} \\ &\leq \cosh^2 R\end{aligned} \tag{52}$$

is valid, where the first inequality is from the Cauchy Schwarz inequality. Thus, we have $\left\|\boldsymbol{L}^{-}\right\|_{\max} \leq \cosh^2 R$, and likewise, we also have $\left\|\boldsymbol{L}^{+}\right\|_{\max} \leq \sinh^2 R$, which imply condition **(e)**. We have proved the sufficiency in (ii).

(iii) If $(\boldsymbol{z}_n)_{n=1}^{N} \in \mathcal{B}_R^{\rho}$, since $\mathcal{B}_R^{\rho} \subset \mathcal{B}_R$, conditions **(a)-(e)** follows from the above discussion. Let $\boldsymbol{Z}^{-} = \boldsymbol{U}^{-}\boldsymbol{\Sigma}^{-}\left(\boldsymbol{V}^{-}\right)^{\top}$ be the singular value decomposition of $\boldsymbol{Z}^{-}$, where $\boldsymbol{U} \in \mathbb{R}^{1,1}, \boldsymbol{V} \in \mathbb{R}^{N,N}$ are orthogonal and $\boldsymbol{\Sigma}^{-} \in \mathbb{R}^{1,N}$ is diagonal. The singular decomposition of $\boldsymbol{L}^{-}$ is given by

$$\begin{aligned}\boldsymbol{L}^{-} &= \boldsymbol{V}^{-}\left(\boldsymbol{\Sigma}^{-}\right)^{\top}\left(\boldsymbol{U}^{-}\right)^{\top}\boldsymbol{U}^{-}\boldsymbol{\Sigma}^{-}\left(\boldsymbol{V}^{-}\right)^{\top} \\ &= \boldsymbol{V}^{-}\left(\boldsymbol{\Sigma}^{-}\right)^{\top}\boldsymbol{\Sigma}^{-}\left(\boldsymbol{V}^{-}\right)^{\top},\end{aligned} \tag{53}$$

where the diagonal elements of $\left(\boldsymbol{\Sigma}^{-}\right)^{\top}\boldsymbol{\Sigma}^{-}$ indicate the singular values of $\boldsymbol{L}^{-}$. Hence, $\left\|\boldsymbol{L}^{-}\right\|_{*} = \operatorname{Tr}\left(\left(\boldsymbol{\Sigma}^{-}\right)^{\top}\boldsymbol{\Sigma}^{-}\right)$. As $\boldsymbol{V}$ is orthogonal, we have

$$\begin{aligned}\operatorname{Tr}\left(\left(\boldsymbol{\Sigma}^{-}\right)^{\top}\boldsymbol{\Sigma}^{-}\right) &= \operatorname{Tr}\left(\left(\boldsymbol{\Sigma}^{-}\right)^{\top}\boldsymbol{\Sigma}^{-}\left(\boldsymbol{V}^{-}\right)^{\top}\boldsymbol{V}^{-}\right) \\ &= \operatorname{Tr}\left(\boldsymbol{V}^{-}\left(\boldsymbol{\Sigma}^{-}\right)^{\top}\boldsymbol{\Sigma}^{-}\left(\boldsymbol{V}^{-}\right)^{\top}\right) \\ &= \operatorname{Tr}\left(\boldsymbol{L}^{-}\right).\end{aligned} \tag{54}$$

Hence, we get

$$\begin{aligned}\left\|\boldsymbol{L}^{-}\right\|_{*} &= \operatorname{Tr}\left(\boldsymbol{L}^{-}\right) \\ &= \sum_{n=1}^{N}\left(\boldsymbol{z}_n^{-}\right)^{\top}\boldsymbol{z}_n^{-} \\ &= \sum_{n=1}^{N}\cosh^2\Delta_{\mathbb{L}^D}(\boldsymbol{z}_0, \boldsymbol{z}_n).\end{aligned} \tag{55}$$

Likewise, we have

$$\begin{aligned}\left\|\boldsymbol{L}^{+}\right\|_{*} &= \operatorname{Tr}\left(\boldsymbol{L}^{+}\right) \\ &= \sum_{n=1}^{N}\left(\boldsymbol{z}_n^{+}\right)^{\top}\boldsymbol{z}_n^{+} \\ &= \sum_{n=1}^{N}\sinh^2\Delta_{\mathbb{L}^D}(\boldsymbol{z}_0, \boldsymbol{z}_n).\end{aligned} \tag{56}$$

By the definition of $\mathcal{B}_R^{\rho}$, we have $\left\|\boldsymbol{L}^{+}\right\|_{*} \leq N\sinh^2\rho$ and $\left\|\boldsymbol{L}^{-}\right\|_{*} \leq N\cosh^2\rho$, which imply condition **(f)**. This completes the proof of the sufficiency in (iii).

(Necessity) (i) Assume that conditions **(a)-(d)** are satisfied. Noting that $\boldsymbol{L}^{-} \succeq \mathbf{O}$, let $\boldsymbol{L}^{-} = \tilde{\boldsymbol{V}}^{-}\boldsymbol{T}^{-}\left(\tilde{\boldsymbol{V}}^{-}\right)^{\top}$ be a singular value decomposition of $\boldsymbol{L}^{-}$, where $\tilde{\boldsymbol{V}}^{-} \in \mathbb{R}^{N.N}$ is orthogonal and $\boldsymbol{T}^{-}$ is diagonal, that is, $\left[\boldsymbol{T}^{-}\right]_{i,j} = 0$ if $i \neq j$. Since $\operatorname{rank}\boldsymbol{L}^{-} = 1$ and $\boldsymbol{L}^{-} \succeq \mathbf{O}$, we can assume that $\left[\boldsymbol{T}^{-}\right]_{1,1} > 0$ and $\left[\boldsymbol{T}^{-}\right]_{n,n} = 0$ for all $N = 2, 3, \ldots, N$. Therefore, $[\boldsymbol{L}^{-}]_{i,j} = \left[\tilde{\boldsymbol{V}}^{-}\right]_{i,1}\left[\boldsymbol{T}^{-}\right]_{1,1}\left[\tilde{\boldsymbol{V}}^{-}\right]_{i,1}$, for $i, j \in [N]$. In particular, $[\boldsymbol{L}^{-}]_{1,1} = \left(\left[\tilde{\boldsymbol{V}}^{-}\right]_{1,1}\right)^2\left[\boldsymbol{T}^{-}\right]_{1,1}$. As $\boldsymbol{L}^{+}$ is positive semi-definite, its diagonal entries are all non-negative. In particular, $\left[\boldsymbol{L}^{+}\right]_{1,1} \geq 0$. Since $[\boldsymbol{L}^{+}]_{1,1} - [\boldsymbol{L}^{-}]_{1,1} = -1$ from **(c)**, we have $\left[\boldsymbol{L}^{-}\right]_{1,1} \geq 1$. Hence, we have $\left[\tilde{\boldsymbol{V}}^{-}\right]_{1,1} \neq 0$. Define $\boldsymbol{V}^{-}$ by

$$\boldsymbol{V}^{-} \coloneqq \begin{cases}\boldsymbol{V}^{-} & \text{if } \left[\tilde{\boldsymbol{V}}^{-}\right]_{1,1} > 0, \\ -\boldsymbol{V}^{-} & \text{if } \left[\tilde{\boldsymbol{V}}^{-}\right]_{1,1} < 0.\end{cases} \tag{57}$$

Then $\left[\boldsymbol{V}^{-}\right]_{1,1} > 0$ and $\boldsymbol{L}^{-} = \boldsymbol{V}^{-}\boldsymbol{T}^{-}\left(\boldsymbol{V}^{-}\right)^{\top}$ is valid. Let $\boldsymbol{L}^{+} = \boldsymbol{V}^{+}\boldsymbol{T}^{+}\left(\boldsymbol{V}^{+}\right)^{\top}$ be a singular value decomposition of $\boldsymbol{L}^{+}$, where $\tilde{\boldsymbol{V}}^{+} \in \mathbb{R}^{N.N}$ is orthogonal and $\boldsymbol{T}^{+}$ is diagonal. Since $\operatorname{rank}\boldsymbol{L}^{+} \leq D$ and $\boldsymbol{L}^{+} \succeq \mathbf{O}$, we can assume that $\left[\boldsymbol{T}^{-}\right]_{n,n} = 0$ for all $n = D+1, D+2, \ldots, N$. Define $\boldsymbol{z}_n^{-} \in \mathbb{R}^1$ and $\boldsymbol{z}_n^{+} \in \mathbb{R}^D$ by

$$\begin{aligned}\boldsymbol{z}_n^{-} &\coloneqq \left[z_{0,n}\right], \\ \boldsymbol{z}_n^{+} &\coloneqq \begin{bmatrix} z_{1,n} & z_{2,n} & \cdots & z_{D,n}\end{bmatrix}^{\top},\end{aligned} \tag{58}$$

respectively, where

$$z_{d,n} = \begin{cases}\sqrt{\left[\boldsymbol{T}^{-}\right]_{1,1}}\left[\boldsymbol{V}^{-}\right]_{n,1} & \text{if } d = 0, \\ \sqrt{\left[\boldsymbol{T}^{+}\right]_{d,d}}\left[\boldsymbol{V}^{+}\right]_{n,d} & \text{if } d = 1, 2, \ldots, D.\end{cases} \tag{59}$$

respectively, and define $\boldsymbol{Z}^{-} \in \mathbb{R}^{1,N}$ and $\boldsymbol{Z}^{+} \in \mathbb{R}^{D,N}$ by

$$\begin{aligned}\boldsymbol{Z}^{-} &= \begin{bmatrix}\boldsymbol{z}_1^{-} & \boldsymbol{z}_2^{-} & \cdots & \boldsymbol{z}_N^{-}\end{bmatrix}, \\ \boldsymbol{Z}^{+} &= \begin{bmatrix}\boldsymbol{z}_1^{+} & \boldsymbol{z}_2^{+} & \cdots & \boldsymbol{z}_N^{+}\end{bmatrix},\end{aligned} \tag{60}$$

respectively. Now, for $n \in [N]$, we define $\boldsymbol{z}_n \in \mathbb{R}^{1+D}$ by

$$\boldsymbol{z}_n = \begin{bmatrix} \boldsymbol{z}_n^- \\ \boldsymbol{z}_n^+ \end{bmatrix}. \tag{61}$$

The Lorentz Gramian of $(\boldsymbol{z}_n)_{n=1}^N$ is given by

$$\begin{aligned}(\boldsymbol{Z}^+)^\top \boldsymbol{Z}^+ - (\boldsymbol{Z}^-)^\top \boldsymbol{Z}^- &= \boldsymbol{V}^+ \boldsymbol{T}^+ (\boldsymbol{V}^+)^\top - \boldsymbol{V}^- \boldsymbol{T}^- (\boldsymbol{V}^-)^\top \\ &= \boldsymbol{L}^+ - \boldsymbol{L}^-.\end{aligned} \tag{62}$$

In the following, we prove $\boldsymbol{z}_n \in \mathbb{L}^D$ for all $n \in [N]$. Since $\langle \boldsymbol{z}_n, \boldsymbol{z}_n \rangle_{\mathrm{M}} = [\boldsymbol{L}]_{n,n} = [\boldsymbol{L}^+ - \boldsymbol{L}^-]_{n,n} = -1$, it is sufficient to prove that $z_{0.n} > 0$. From $[\boldsymbol{V}^-]_{1,1} > 0$ and $z_{0,n} = \sqrt{[\boldsymbol{T}^-]_{1,1}} [\boldsymbol{V}^-]_{n,1}$, we have $z_{0,1} > 0$. For general $n \in [N]$, the following is valid:

$$\begin{aligned}|z_{0,1}||z_{0,n}| - z_{0,1} z_{0,n} &> \|\boldsymbol{z}_1^+\|_2 \|\boldsymbol{z}_n^+\|_2 - z_{0,1} z_{0,n} \\ &\geq (\boldsymbol{z}_1^+)^\top \boldsymbol{z}_n^+ - z_{0,1} z_{0,n} \\ &= -\langle \boldsymbol{z}_1, \boldsymbol{z}_n \rangle_{\mathrm{M}} \\ &= [\boldsymbol{L}]_{1,n} \\ &= [\boldsymbol{L}^+ - \boldsymbol{L}^-]_{1,n} \\ &\geq 1 \\ &> 0.\end{aligned} \tag{63}$$

Therefore, $z_{0,1}$ and $z_{0,n}$ must have the same sign. Hence, $z_{0,n} > 0$, which completes the proof of the necessity in (i).

(ii) Assume that conditions **(a)-(e)** are satisfied. Define $(\boldsymbol{z}_n)_{n=1}^N$ as in (58). Then, since **(a)-(d)** are satisfied, $(\boldsymbol{z}_n)_{n=1}^N \in (\mathbb{L}^D)^N$ and its Lorentz Gramian is $\boldsymbol{L} = \boldsymbol{L}^+ - \boldsymbol{L}^-$. Thus, it suffices to show that condition **(e)** implies that $\Delta_{\mathbb{L}^D}(\boldsymbol{z}_0, \boldsymbol{z}_N) \leq R$ is valid for all $n \in N$. Since $[\boldsymbol{L}^-]_{n,n} = (z_{0,n})^2 = \cosh^2 \Delta_{\mathbb{L}^D}(\boldsymbol{z}_0, \boldsymbol{z}_N)$, condition **(e)** implies $\cosh^2 \Delta_{\mathbb{L}^D}(\boldsymbol{z}_0, \boldsymbol{z}_N) \leq \cosh^2 R$, which is equivalent to $\Delta_{\mathbb{L}^D}(\boldsymbol{z}_0, \boldsymbol{z}_N) \leq R$. Hence, we have $(\boldsymbol{z}_n)_{n=1}^N \in \mathcal{B}_R$.

(iii) Assume that conditions **(a)-(f)** are satisfied. Define $(\boldsymbol{z}_n)_{n=1}^N$ as in (58). Then, since **(a)-(e)** are satisfied, $(\boldsymbol{z}_n)_{n=1}^N \in \mathcal{B}_R$ and its Lorentz Gramian is $\boldsymbol{L} = \boldsymbol{L}^+ - \boldsymbol{L}^-$. Also, condition **(f)** implies $\frac{1}{N} \sum_{n=1}^N \cosh^2 \Delta_{\mathbb{L}^D}(\boldsymbol{z}_0, \boldsymbol{z}_n) \leq \cosh^2 \rho$, since the left hand side is given by $\frac{1}{N} \|\boldsymbol{L}^-\|_*$. □

Remark 3. The statement of Lemma 9 (i) is equivalent to that of Proposition 1 in (Tabaghi & Dokmanic, 2020). However, the proof there does not consider the case where the representations given by the decomposition of the Lorentz Gramian are all in $-\mathbb{L}^D$ instead of $\mathbb{L}^D$. The above construction of representations solves this problem by forcing $z_{0,n}$ to be positive.

F. Rademacher Complexity and Proof of Theorem 1

First, we define Rademacher complexity (Koltchinskii, 2001; Koltchinskii & Panchenko, 2000; Bartlett et al., 2002), the key tool to obtain HOE's generalization error bound. Let $\mathcal{T}$ be our input space and $\mathcal{H} \subset \{h | h : \mathcal{T} \to \mathbb{R}\}$ be our hypothesis space. Let $S \in \mathbb{Z}_{\geq 0}$ be the number of data points, and suppose that data points $(t_1, y_1), (t_2, y_2), \ldots, (t_S, y_S) \in \mathcal{T} \times \{-1, +1\}$ are independently distributed according to some unknown fixed distribution μ. The Rademacher complexity of h is defined as follows:

Definition 3. Let $\sigma_1, \sigma_2, \ldots, \sigma_S$ be random values such that $\sigma_1, \sigma_2, \ldots, \sigma_S, (t_1, y_1), (t_2, y_2), \ldots, (t_S, y_S)$ are mutually independent and each of $\sigma_1, \sigma_2, \ldots, \sigma_S$ takes values $\{-1, +1\}$ with equal probability. The Rademacher complexity $\mathfrak{R}_S(\mathcal{H})$ is defined by

$$\mathfrak{R}_S(\mathcal{H}) := \mathbb{E}_{(t_s)_{s=1}^S} \mathbb{E}_{(\sigma_s)_{s=1}^S} \left[\frac{1}{S} \sup_{h \in \mathcal{H}} \sum_{s=1}^S \sigma_s h(t_s) \right]. \tag{64}$$

We use the following theorem provided by Bartlett & Mendelson (2002) and arranged by Kakade et al. (2008).

Theorem 10 ((Bartlett & Mendelson, 2002; Kakade et al., 2008))**.** *Let $\phi : \mathcal{T} \times \{-1. +1\} \to \mathbb{R}$ be a loss function. Define the empirical risk function $\hat{\mathcal{R}}_S(h)$ and expected risk function $\mathcal{R}(h)$ by*

$$\begin{aligned}\hat{\mathcal{R}}_S(h) &:= \frac{1}{S} \sum_{s=1}^S \phi((h(t_s), y_s)), \\ \mathcal{R}(h) &:= \mathbb{E}_{t,y} \phi(h(t), y)).\end{aligned} \tag{65}$$

Assume that $\phi(\cdot, -1)$ and $\phi(\cdot, +1)$ are L_ϕ-Lipschitz and bounded. Define

$$c_\phi := \sup_{\substack{t \in \mathcal{T}, \\ y \in \{-1, +1\}, \\ h \in \mathcal{H}}} \phi(h(t), y) - \inf_{\substack{t \in \mathcal{T}, \\ y \in \{-1, +1\}, \\ h \in \mathcal{H}}} \phi(h(t), y). \tag{66}$$

Then for any $\delta \in \mathbb{R}_{>0}$ and with probability at least $1 - \delta$ simultaneously for all $h \in \mathcal{H}$ we have that

$$\mathcal{R}(h) - \hat{\mathcal{R}}_S(h) \leq 2 \mathfrak{R}_S(\mathcal{H}) + c_\phi \sqrt{\frac{\ln(1/\delta)}{2S}}. \tag{67}$$

From this theorem, we can easily derive an upper bound for the excess risk of the empirical risk minimizer as follows.

Corollary 11. *Define the empirical risk minimizer $\hat{h} \in \mathcal{H}$ and expected loss minimizer by $h^* \in \mathcal{H}$ by*

$$\hat{h} := \operatorname*{argmin}_{h \in \mathcal{H}} \hat{\mathcal{R}}_S(h), \quad h^* := \operatorname*{argmin}_{h \in \mathcal{H}} \mathcal{R}(h), \tag{68}$$

and we call $\mathcal{R}\left(\hat{h}\right) - \mathcal{R}(h^*)$ the excess risk of $\hat{h}$. Then for any $\delta \in \mathbb{R}_{>0}$ and with probability at least $1-\delta$ we have that

$$\mathcal{R}\left(\hat{h}\right) - \mathcal{R}(h^*) \leq 2\mathfrak{R}_S(\mathcal{H}) + 2c_\phi \sqrt{\frac{\ln(2/\delta)}{2S}}. \tag{69}$$

Proof. We have that

$$\begin{aligned}
&\mathcal{R}\left(\hat{h}\right) - \mathcal{R}(h^*) \\
&= \left(\mathcal{R}\left(\hat{h}\right) - \hat{\mathcal{R}}_{\mathcal{S}}\left(\hat{h}\right)\right) + \left(\hat{\mathcal{R}}_{\mathcal{S}}\left(\hat{h}\right) - \hat{\mathcal{R}}_{\mathcal{S}}(h^*)\right) \\
&\quad + \left(\hat{\mathcal{R}}_{\mathcal{S}}(h^*) - \mathcal{R}(h^*)\right) \\
&\leq \left(\mathcal{R}\left(\hat{h}\right) - \hat{\mathcal{R}}_{\mathcal{S}}\left(\hat{h}\right)\right) + \left(\hat{\mathcal{R}}_{\mathcal{S}}(h^*) - \mathcal{R}(h^*)\right),
\end{aligned} \tag{70}$$

where the last inequality holds from the definition of $\hat{h}$. We complete the proof by evaluating the first and second term by Theorem 10 and Hoeffding's inequality, respectively. □

Therefore, it suffices to obtain the upper bound of HOE model's Rademacher complexity, which is given below. Let $\mathcal{B} \subset \mathbb{L}^D$. We define a hypothesis function class $h(\cdot;\mathcal{B})$ by

$$h(\cdot;\mathcal{B}) := \left\{h\left(\cdot;(\boldsymbol{z}_n)_{n=1}^N\right) \middle| (\boldsymbol{z}_n)_{n=1}^N \in \mathcal{B}\right\}. \tag{71}$$

We evaluate the Rademacher complexity of the hypothesis function class $h(\cdot;\mathcal{B}^\rho)$ defined by

$$\begin{aligned}
&\mathfrak{R}_S(h(\cdot;\mathcal{B}^\rho)) \\
&:= \mathbb{E}_{(i,j,k)}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{(\boldsymbol{z}_n)_{n=1}^N \in \mathcal{B}^\rho} \frac{1}{S}\sum_{s=1}^S \sigma_s h\left(i_s, j_s, k_s; (\boldsymbol{z}_n)_{n=1}^N\right)\right].
\end{aligned} \tag{72}$$

The following evaluates the complexity.

Lemma 12. *The Rademacher complexity of the hypothesis function class $h(\cdot;\mathcal{B}^\rho)$ satisfies the following inequality:*

$$\begin{aligned}
&\mathfrak{R}_S(h(\cdot;\mathcal{B}^\rho)) \\
&\leq \left(\cosh^2\rho + \sinh^2\rho\right)\left(\sqrt{\frac{2N\ln N}{S}} + \frac{N\ln N}{6S}\right).
\end{aligned} \tag{73}$$

Proof. Define $\mathcal{G}_R^-, \mathcal{G}_R^+, \overline{\mathcal{G}}_R^-, \overline{\mathcal{G}}_R^+ \subset \mathbb{S}^{N,N}$ by

$$\begin{aligned}
\mathcal{G}^\rho &:= \left\{\boldsymbol{L}^+ - \boldsymbol{L}^- \middle| \boldsymbol{L}^-, \boldsymbol{L}^+ \in \mathbb{S}^{N,N} \text{ satisfy } \textbf{(a)-(d), (f)}.\right\}, \\
\overline{\mathcal{G}}^\rho &:= \left\{\boldsymbol{L}^+ - \boldsymbol{L}^- \middle| \boldsymbol{L}^- \in \mathcal{Q}^{N\cosh^2\rho}, \boldsymbol{L}^+ \in \mathcal{Q}^{N\sinh^2\rho}.\right\},
\end{aligned} \tag{74}$$

where **(a)-(f)** are the conditions defined in Lemma 9, and for $\lambda \in \mathbb{R}_{\geq 0}$, $\mathcal{Q}^\lambda$ is defined by

$$\mathcal{Q}^\lambda := \left\{\boldsymbol{L} \in \mathbb{S}^{N,N} \middle| \boldsymbol{L} \succeq \mathbf{O}, \|\boldsymbol{L}\|_* \leq \lambda.\right\}. \tag{75}$$

Note that $\mathcal{G}_R^\rho \subset \overline{\mathcal{G}}^\rho$. Let $\sigma_1, \sigma_2, \ldots, \sigma_S$ be i.i.d Rademacher random variables and $\boldsymbol{\sigma} = \begin{bmatrix}\sigma_1 & \sigma_2 & \ldots & \sigma_S\end{bmatrix}^\top$. The Rademacher complexity is calculated as follows:

$$\begin{aligned}
&\mathfrak{R}_S(h(\cdot;\mathcal{B}^\rho)) \\
&:= \mathbb{E}_{(i,j,k)}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{(\boldsymbol{z}_n)_{n=1}^N \in \mathcal{B}^\rho} \frac{1}{S}\sum_{s=1}^S \sigma_s h\left(i_s, j_s, k_s; (\boldsymbol{z}_n)_{n=1}^N\right)\right] \\
&= \mathbb{E}_{(i,j,k)}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{L} \in \mathcal{G}^\rho} \frac{1}{S}\sum_{s=1}^S \sigma_s \left\langle \boldsymbol{L}, \boldsymbol{T}_{i_s,j_s,k_s}\right\rangle_{\mathrm{F}}\right] \\
&\leq \mathbb{E}_{(i,j,k)}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{L} \in \overline{\mathcal{G}}^\rho} \frac{1}{S}\sum_{s=1}^S \sigma_s \left\langle \boldsymbol{L}, \boldsymbol{T}_{i_s,j_s,k_s}\right\rangle_{\mathrm{F}}\right],
\end{aligned} \tag{76}$$

where the last inequality results from $\mathcal{G}_R^\rho \subset \overline{\mathcal{G}}^\rho$. We can decompose the integrand of the above expectation operator as follows:

$$\begin{aligned}
&\sup_{\boldsymbol{L} \in \overline{\mathcal{G}}^\rho} \frac{1}{S}\sum_{s=1}^S \sigma_s \left\langle \boldsymbol{L}, \boldsymbol{T}_{i_s,j_s,k_s}\right\rangle_{\mathrm{F}} \\
&= \sup_{\substack{\boldsymbol{L}^- \in \mathcal{Q}^{N\sinh^2\rho}, \\ \boldsymbol{L}^+ \in \mathcal{Q}^{N\cosh^2\rho}}} \frac{1}{S}\sum_{s=1}^S \sigma_s \left\langle \boldsymbol{L}^+ - \boldsymbol{L}^-, \boldsymbol{T}_{i_s,j_s,k_s}\right\rangle_{\mathrm{F}} \\
&\leq \sup_{\boldsymbol{L}^+ \in \mathcal{Q}^{N\sinh^2\rho}} \frac{1}{S}\sum_{s=1}^S \sigma_s \left\langle \boldsymbol{L}^+, \boldsymbol{T}_{i_s,j_s,k_s}\right\rangle_{\mathrm{F}} \\
&\quad + \sup_{\boldsymbol{L}^- \in \mathcal{Q}^{N\cosh^2\rho}} \frac{1}{S}\sum_{s=1}^S \sigma_s \left\langle \boldsymbol{L}^-, \boldsymbol{T}_{i_s,j_s,k_s}\right\rangle_{\mathrm{F}}.
\end{aligned} \tag{77}$$

Hence, we have

$$\begin{aligned}
&\mathfrak{R}_S(h(\cdot;\mathcal{B}^\rho)) \\
&\leq \mathfrak{R}_S^{\mathrm{G}}\left(\left\langle \mathcal{Q}^{N\sinh^2\rho}, \cdot\right\rangle_{\mathrm{F}}\right) + \mathfrak{R}_S^{\mathrm{G}}\left(\left\langle \mathcal{Q}^{N\cosh^2\rho}, \cdot\right\rangle_{\mathrm{F}}\right),
\end{aligned} \tag{78}$$

where $\mathfrak{R}_S^{\mathrm{G}}\left(\left\langle \mathcal{Q}^\lambda, \cdot\right\rangle_{\mathrm{F}}\right)$ is defined as

$$\mathbb{E}_{(i,j,k)}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\boldsymbol{L} \in \mathcal{Q}^\lambda} \frac{1}{S}\sum_{s=1}^S \sigma_s \left\langle \boldsymbol{L}, \boldsymbol{T}_{i_s,j_s,k_s}\right\rangle_{\mathrm{F}}\right]. \tag{79}$$

We can bound $\mathfrak{R}_S^{\mathrm{G}}\left(\left\langle \mathcal{Q}^\lambda, \cdot\right\rangle_{\mathrm{F}}\right)$ by the following lemma.

Lemma 13.

$$\mathfrak{R}_S^{\mathrm{G}}\left(\left\langle \mathcal{Q}^\lambda, \cdot\right\rangle_{\mathrm{F}}\right) \leq \frac{\lambda}{N}\left(\sqrt{\frac{2(N+1)\ln N}{S}} + \cdot \frac{N\ln N}{\sqrt{12}S}\right). \tag{80}$$

We prove Lemma 13 later. By Lemma 13, we obtain

$$\begin{aligned}&\mathfrak{R}_S(h(\cdot;\mathcal{B}^\rho))\\&\leq \left(\cosh^2\rho+\sinh^2\rho\right)\left(\sqrt{\frac{2(N+1)\ln N}{S}}+\cdot\frac{N\ln N}{\sqrt{12}S}\right),\end{aligned} \tag{81}$$

which completes the proof. □

Proof of Lemma 13. For $\lambda\in\mathbb{R}_{\geq 0}$, define $\mathcal{Q}_1^\lambda$ by

$$\left\{\lambda\boldsymbol{u}\boldsymbol{u}^\top\middle|\boldsymbol{u}\in\mathbb{R}^N,\|\boldsymbol{u}\|_2=1.\right\}. \tag{82}$$

Since $\mathcal{Q}_1^\lambda$ is the convex hull of $\mathcal{Q}^\lambda$ and the Rademacher complexity of a function class equals that of its convex hull (Bartlett & Mendelson, 2002, Theorem 12-2), we have that

$$\begin{aligned}&\mathfrak{R}_S^{\mathrm{G}}\left(\left\langle\mathcal{Q}^\lambda,\cdot\right\rangle_{\mathrm{F}}\right)\\&=\mathfrak{R}_S^{\mathrm{G}}\left(\left\langle\mathcal{Q}_1^\lambda,\cdot\right\rangle_{\mathrm{F}}\right)\\&=\mathbb{E}_{(i,j,k)}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\|\boldsymbol{u}\|_2\leq 1}\frac{1}{S}\sum_{s=1}^S\sigma_s\left\langle\lambda\boldsymbol{u}\boldsymbol{u}^\top,\boldsymbol{T}_{i_s,j_s,k_s}\right\rangle_{\mathrm{F}}\right]\\&=\mathbb{E}_{(i,j,k)}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\|\boldsymbol{u}\|_2\leq 1}\frac{1}{S}\sum_{s=1}^S\sigma_s\lambda\operatorname{Tr}\left(\boldsymbol{u}\boldsymbol{u}^\top\boldsymbol{T}_{i_s,j_s,k_s}\right)\right]\\&=\frac{\lambda}{S}\mathbb{E}_{(i,j,k)}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\|\boldsymbol{u}\|_2\leq 1}\sum_{s=1}^S\sigma_s\operatorname{Tr}\left(\boldsymbol{u}^\top\boldsymbol{T}_{i_s,j_s,k_s}\boldsymbol{u}\right)\right]\\&=\frac{\lambda}{S}\mathbb{E}_{(i,j,k)}\mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\|\boldsymbol{u}\|_2\leq 1}\operatorname{Tr}\left(\boldsymbol{u}^\top\left(\sum_{s=1}^S\sigma_s\boldsymbol{T}_{i_s,j_s,k_s}\right)\boldsymbol{u}\right)\right]\\&=\frac{\lambda}{S}\mathbb{E}_{(i,j,k)}\mathbb{E}_{\boldsymbol{\sigma}}\left[\left\|\sum_{s=1}^S\sigma_s\boldsymbol{T}_{i_s,j_s,k_s}\right\|_{\mathrm{op},2}\right],\end{aligned} \tag{83}$$

where $\|\cdot\|_{\mathrm{op},2}$ denotes the operator norm with respect to the 2-norm defined by

$$\|\boldsymbol{A}\|_{\mathrm{op},2}\coloneqq\max_{\|\boldsymbol{u}\|_2\leq 1}\|\boldsymbol{A}\boldsymbol{u}\|_2 \tag{84}$$

To evaluate this, we can apply the following the matrix Bernstein inequality.

Theorem 14 ((Tropp, 2015) Theorem 6.6.1)**.** *Let $\boldsymbol{A}_1,\boldsymbol{A}_2,\dots,\boldsymbol{A}_S\in\mathbb{S}^{N,N}$ be independent random matrices that satisfies*

$$\mathbb{E}\boldsymbol{A}_s=\mathbf{O},\quad\|\boldsymbol{A}_s\|_{\mathrm{op},2}\leq\sigma. \tag{85}$$

Then

$$\mathbb{E}\left\|\sum_{s=1}^S\boldsymbol{A}_s\right\|_{\mathrm{op},2}\leq\sqrt{2v\left(\sum_{s=1}^S\boldsymbol{A}_s\right)\ln N}+\frac{1}{3}\sigma\ln N, \tag{86}$$

where v is the matrix variance statistics defined by

$$v(\boldsymbol{A})\coloneqq\left\|\mathbb{E}\boldsymbol{A}^2\right\|_{\mathrm{op},2}. \tag{87}$$

Note that $v\left(\sum_{s=1}^S\boldsymbol{A}_s\right)=\left\|\sum_{s=1}^S\mathbb{E}\boldsymbol{A}_s^2\right\|_{\mathrm{op},2}$ is valid since $\boldsymbol{A}_1,\boldsymbol{A}_2,\dots,\boldsymbol{A}_S\in\mathbb{S}^{N,N}$ are independent. We apply Theorem 14 to the right hand side of (83) by substituting $\boldsymbol{A}_s$ by $\sigma_s\boldsymbol{T}_{i_s,j_s,k_s}$. Here, $\mathbb{E}\sigma_s\boldsymbol{T}_{i_s,j_s,k_s}=\mathbf{O}$ is valid because $\mathbb{E}\sigma_s=0$. The singular values of $\sigma_s\boldsymbol{T}_{i_s,j_s,k_s}$ is equal to those of $\sigma_s\tilde{\boldsymbol{T}}$, where

$$\tilde{\boldsymbol{T}}\coloneqq\begin{bmatrix}0&-\frac{1}{2}&+\frac{1}{2}\\-\frac{1}{2}&0&0\\+\frac{1}{2}&0&0\end{bmatrix}. \tag{88}$$

Since

$$\left(\sigma_s\tilde{\boldsymbol{T}}\right)^2=\left(\sigma_s\tilde{\boldsymbol{T}}\right)^\top\left(\sigma_s\tilde{\boldsymbol{T}}\right)=\begin{bmatrix}+\frac{1}{2}&0&0\\0&+\frac{1}{4}&-\frac{1}{4}\\0&-\frac{1}{4}&+\frac{1}{4}\end{bmatrix}, \tag{89}$$

and its eigenvalues are $0,+\frac{1}{2},+\frac{1}{2}$, the singular values of $\tilde{\boldsymbol{T}}$ are $0,+\frac{1}{\sqrt{2}},+\frac{1}{\sqrt{2}}$. Hence, we have that $\|\sigma_s\boldsymbol{T}_{i_s,j_s,k_s}\|_{\mathrm{op},2}\leq\frac{1}{\sqrt{2}}$. Lastly, we evaluate $\left\|\sum_{s=1}^S\mathbb{E}(\sigma_s\boldsymbol{T}_{i_s,j_s,k_s})^2\right\|_{\mathrm{op},2}=S\left\|\mathbb{E}\boldsymbol{T}_{i_s,j_s,k_s}^2\right\|_{\mathrm{op},2}$. In general, since $\|\cdot\|_{\mathrm{op},2}$ is a convex function, we have that $\left\|\mathbb{E}\boldsymbol{T}_{i_s,j_s,k_s}^2\right\|_{\mathrm{op},2}\leq\mathbb{E}\left\|\boldsymbol{T}_{i_s,j_s,k_s}^2\right\|_{\mathrm{op},2}=\frac{1}{2}$.

The diagonal elements and off-diagonal elements of $\mathbb{E}\boldsymbol{T}_{i_s,j_s,k_s}^2$ are all $\frac{1}{N}$ and all $-\frac{1}{2}\frac{1}{N(N-1)}$, respectively, because the sum of the diagonal elements and that of the off-diagonal elements in $\boldsymbol{T}_{i_s,j_s,k_s}$ are always 1 and $-\frac{1}{2}$, respectively, and from the symmetricity among the N diagonal elements and $N(N-1)$ off-diagonal elements. Hence, we have $\mathbb{E}\boldsymbol{T}_{i_s,j_s,k_s}^2=\left(\frac{1}{N}+\frac{1}{2}\frac{1}{N(N-1)}\right)\mathbf{I}-\frac{1}{2}\frac{1}{N(N-1)}\mathbf{1}\mathbf{1}^\top$. whose eigenvalues (singular values) are $\frac{1}{N}+\frac{1}{2}\frac{1}{N(N-1)}$ (multiplicity $N-1$) and $\frac{1}{N}$ (multiplicity 1). Thus, we have

$$\begin{aligned}v\left(\sum_{s=1}^S\sigma_s\boldsymbol{T}_{i_s,j_s,k_s}\right)&=S\left\|\mathbb{E}\boldsymbol{T}_{i_s,j_s,k_s}^2\right\|_{\mathrm{op},2}\\&=S\left(\frac{1}{N}+\frac{1}{2}\frac{1}{N(N-1)}\right)\\&=\frac{S}{2}\left(\frac{1}{N-1}+\frac{1}{N}\right)\\&=\frac{S}{2}\frac{1}{N^2}\left(\frac{N^2}{N-1}+N\right)\\&\leq\frac{S}{N^2}(N+1)\end{aligned} \tag{90}$$

. Therefore we have

$$\begin{aligned}&\mathbb{E}_{(i,j,k)}\mathbb{E}_{\boldsymbol{\sigma}}\left[\left\|\sum_{s=1}^S\sigma_s\boldsymbol{T}_{i_s,j_s,k_s}\right\|_{\mathrm{op},2}\right]\\&\leq\frac{1}{N}\sqrt{2S(N+1)\ln N}+\frac{1}{3}\cdot\frac{1}{\sqrt{2}}\ln N,\end{aligned} \tag{91}$$

which completes the proof. □

Proof of Theorem 1. We complete the proof by applying Corollary 11 to Lemma 12. □