A Useful Technical Lemmas

Lemma 5 (Weyl's inequality)**.** *Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$ with $\sigma_1(\boldsymbol{A}) \geq \cdots \geq \sigma_r(\boldsymbol{A})$ and $\sigma_1(\boldsymbol{B}) \geq \cdots \geq \sigma_r(\boldsymbol{B})$, where $r = \min(m, n)$. Then,*

$$\max_{i\in[r]} |\sigma_i(\boldsymbol{A}) - \sigma_i(\boldsymbol{A})| \leq \|\boldsymbol{A} - \boldsymbol{B}\|_2 .$$

Lemma 6. *Let us define $\odot$ as the Hadamard product. Given two positive semi-definite (PSD) matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, it holds that*

$$\lambda_{\min}(\boldsymbol{A} \odot \boldsymbol{B}) \geq \left(\min_i \boldsymbol{B}_{ii}\right) \lambda_{\min}(\boldsymbol{A}).$$

Lemma 7 ([34])**.** *Let $h_r(x) = \frac{1}{\sqrt{r!}}(-1)^r e^{x^2/2} \frac{d^r}{dx^r} e^{-x^2/2}$ be normalized probabilist's hermite polynomials. Let $\phi(\cdot)$ denote ReLU , we define $\mu_r(\phi) = \int_{-\infty}^{\infty} \phi(x) h_r(x) \frac{e^{-x^2/2}}{\sqrt{\pi}} dx$. It holds that*

$$Q(x) = \sum_{r=0}^{\infty} \mu_r^2(\phi) x^r = \frac{\sqrt{1-x^2} + (\pi - \arccos x)\, x}{\pi}.$$

Moreover, it holds that $\sup\{r : \mu_r^2(\phi) > 0\} = \infty$.

B Proof for Section 3.1

We first present several useful inequalities. The proof mainly relies on basic norm inequalities and the Lipschitz property of ReLU.

Lemma 8. *For each $s \in [0, \tau]$, suppose that $\|\boldsymbol{W}(s)\|_2 \leq \bar{\rho}_w$, $\|\boldsymbol{U}(s)\|_2 \leq \bar{\rho}_u$, and $\|\boldsymbol{a}(s)\|_2 \leq \bar{\rho}_a$. It holds that*

$$\|\boldsymbol{Z}(s)\|_F \leq c_a \|\boldsymbol{X}\|_F , \tag{15}$$

and

$$\begin{cases} \|\nabla_{\boldsymbol{W}} \Phi(s)\|_F \leq c_w \|\boldsymbol{X}\|_F \|\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}\|_2 \\ \|\nabla_{\boldsymbol{U}} \Phi(s)\|_F \leq c_u \|\boldsymbol{X}\|_F \|\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}\|_2 \\ \|\nabla_{\boldsymbol{a}} \Phi(s)\|_2 \leq c_a \|\boldsymbol{X}\|_F \|\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}\|_2 . \end{cases} \tag{16}$$

Furthermore, for each $k, s \in [0, \tau]$, it holds that

$$\|\boldsymbol{Z}(k) - \boldsymbol{Z}(s)\|_F \leq \bar{\rho}_a^{-1} \left(c_w \|\boldsymbol{W}(k) - \boldsymbol{W}(s)\|_2 + c_u \|\boldsymbol{U}(k) - \boldsymbol{U}(s)\|_2\right) \|\boldsymbol{X}\|_F , \tag{17}$$

and

$$\begin{aligned} &\|\hat{\boldsymbol{y}}(k) - \hat{\boldsymbol{y}}(s)\|_2 \\ &\leq (c_w \|\boldsymbol{W}(k) - \boldsymbol{W}(s)\|_2 + c_u \|\boldsymbol{U}(k) - \boldsymbol{U}(s)\|_2 + c_a \|\boldsymbol{a}(k) - \boldsymbol{a}(s)\|_2) \|\boldsymbol{X}\|_F . \end{aligned} \tag{18}$$

Proof. (1) Proof of Eq. (15): Note that $\boldsymbol{Z}(s) = \phi(\boldsymbol{W}(s)\boldsymbol{Z}(s) + \boldsymbol{U}(s)\boldsymbol{X})$. Using the fact that $|\phi(x)| \leq |x|$, we have

$$\|\boldsymbol{Z}(s)\|_F \leq (\|\boldsymbol{W}(s)\|_2 \|\boldsymbol{Z}(s)\|_F + \|\boldsymbol{U}(s)\|_2 \|\boldsymbol{X}\|_F) \leq \bar{\rho}_w \|\boldsymbol{Z}(s)\|_F + \bar{\rho}_u \|\boldsymbol{X}\|_F .$$

Note that $\|\boldsymbol{W}(s)\|_2 \leq \bar{\rho}_w < 1$, for each $s \in [0, \tau]$, and thus it holds

$$\|\boldsymbol{Z}(s)\|_F \leq \frac{\bar{\rho}_u}{1 - \bar{\rho}_w} \|\boldsymbol{X}\|_F = c_a \|\boldsymbol{X}\|_F .$$

(2) Proof of Eq. (16): First, we have

$$\|\boldsymbol{J}(\tau)^{-1}\|_2 \leq \frac{1}{1 - \bar{\rho}_w},$$

and thus it holds that

$$\|\boldsymbol{R}(\tau)\|_2 \leq \|\|\boldsymbol{a}(\tau)\|_2\|\boldsymbol{J}(\tau)^{-1}\|_2\|\boldsymbol{D}(\tau)\|_2 \leq \frac{\bar{\rho}_a}{1-\bar{\rho}_w}.$$

Then, we have

$$\begin{aligned}
\|\nabla_{\boldsymbol{W}}\Phi(\tau)\|_F &= \|\operatorname{vec}(\nabla_{\boldsymbol{W}}\Phi(\tau))\|_2 \\
&= \left\|(\boldsymbol{Z}(\tau)\otimes \boldsymbol{I}_m)\,\boldsymbol{R}(\tau)^\top\,(\hat{\boldsymbol{y}}(\tau)-\boldsymbol{y})\right\|_2 \\
&\leq \|\boldsymbol{Z}(\tau)\|_2\,\|\boldsymbol{R}(\tau)\|_2\,\|\hat{\boldsymbol{y}}(\tau)-\boldsymbol{y}\|_2 \\
&\leq \frac{\bar{\rho}_u\bar{\rho}_a}{(1-\bar{\rho}_w)^2}\,\|\boldsymbol{X}\|_F\,\|\hat{\boldsymbol{y}}(\tau)-\boldsymbol{y}\|_2\,,
\end{aligned}$$

$$\begin{aligned}
\|\nabla_{\boldsymbol{U}}\Phi(\tau)\|_F &= \|\operatorname{vec}(\nabla_{\boldsymbol{U}}\Phi(\tau))\|_2 \\
&= \left\|(\boldsymbol{X}(\tau)\otimes \boldsymbol{I}_m)\,\boldsymbol{R}(\tau)^\top\,(\hat{\boldsymbol{y}}(\tau)-\boldsymbol{y})\right\|_2 \\
&\leq \frac{\bar{\rho}_a}{1-\bar{\rho}_w}\,\|\boldsymbol{X}\|_F\,\|\hat{\boldsymbol{y}}(\tau)-\boldsymbol{y}\|_2\,,
\end{aligned}$$

$$\|\nabla_{\boldsymbol{a}}\Phi(\tau)\|_2 = \|\boldsymbol{Z}(\hat{\boldsymbol{y}}(\tau)-\boldsymbol{y})\|_2 \leq \frac{\bar{\rho}_u}{1-\bar{\rho}_w}\,\|\boldsymbol{X}\|_F\,\|\hat{\boldsymbol{y}}(\tau)-\boldsymbol{y}\|_2\,.$$

(3) Proof of Eq. (17):

$$\begin{aligned}
&\|\boldsymbol{Z}(k)-\boldsymbol{Z}(s)\|_F \\
=&\|\phi(\boldsymbol{W}(k)\boldsymbol{Z}(k)+\boldsymbol{U}(k)\boldsymbol{X})-\phi(\boldsymbol{W}(s)\boldsymbol{Z}(s)+\boldsymbol{U}(s)\boldsymbol{X})\|_F \\
\leq&\|\boldsymbol{W}(k)\boldsymbol{Z}(k)+\boldsymbol{U}(k)\boldsymbol{X}-\boldsymbol{W}(s)\boldsymbol{Z}(s)-\boldsymbol{U}(s)\boldsymbol{X}\|_F \\
\leq&(\|\boldsymbol{W}(k)\boldsymbol{Z}(k)-\boldsymbol{W}(k)\boldsymbol{Z}(s)\|_F+\|\boldsymbol{W}(k)\boldsymbol{Z}(s)-\boldsymbol{W}(s)\boldsymbol{Z}(s)\|_F+\|\boldsymbol{U}(k)\boldsymbol{X}-\boldsymbol{U}(s)\boldsymbol{X}\|_F) \\
\leq&\|\boldsymbol{W}(k)\|_2\|\boldsymbol{Z}(k)-\boldsymbol{Z}(s)\|_F+(\|\boldsymbol{W}(k)-\boldsymbol{W}(s)\|_2\,\|\boldsymbol{Z}(s)\|_F+\|\boldsymbol{U}(k)-\boldsymbol{U}(s)\|_2\,\|\boldsymbol{X}\|_F) \\
\leq&\bar{\rho}_w\|\boldsymbol{Z}(k)-\boldsymbol{Z}(s)\|_F+\left(\frac{\bar{\rho}_u}{1-\bar{\rho}_w}\,\|\boldsymbol{W}(k)-\boldsymbol{W}(s)\|_2\,\|\boldsymbol{X}\|_F+\|\boldsymbol{U}(k)-\boldsymbol{U}(s)\|_2\,\|\boldsymbol{X}\|_F\right)
\end{aligned}$$

Consequently, we have

$$\|\boldsymbol{Z}(k)-\boldsymbol{Z}(s)\|_F \leq \bar{\rho}_a^{-1}\left(c_w\,\|\boldsymbol{W}(k)-\boldsymbol{W}(s)\|_2+c_u\,\|\boldsymbol{U}(k)-\boldsymbol{U}(s)\|_2\right)\|\boldsymbol{X}\|_F$$

(4) Proof of Eq. (18):

$$\begin{aligned}
&\|\hat{\boldsymbol{y}}(k)-\hat{\boldsymbol{y}}(s)\|_2 \\
=&\|\boldsymbol{a}(k)\boldsymbol{Z}(k)-\boldsymbol{a}(s)\boldsymbol{Z}(s)\|_F \\
\leq&\|\boldsymbol{a}(k)\boldsymbol{Z}(k)-\boldsymbol{a}(k)\boldsymbol{Z}(s)\|_F+\|\boldsymbol{a}(k)\boldsymbol{Z}(s)-\boldsymbol{a}(s)\boldsymbol{Z}(s)\|_F \\
\leq&\|\boldsymbol{a}(k)\|_2\|\boldsymbol{Z}(k)-\boldsymbol{Z}(s)\|_F+\|\boldsymbol{a}(k)-\boldsymbol{a}(s)\|_2\|\boldsymbol{Z}\|_F \\
\leq&(c_w\,\|\boldsymbol{W}(k)-\boldsymbol{W}(s)\|_2+c_u\,\|\boldsymbol{U}(k)-\boldsymbol{U}(s)\|_2+c_a\|\boldsymbol{a}(k)-\boldsymbol{a}(s)\|_2)\,\|\boldsymbol{X}\|_F\,,
\end{aligned}$$

where the last inequality follows from Eq. (17). $\square$

B.1 Proof of Theorem 1

Proof. We show by induction for every $\tau > 0$,

$$\begin{cases}
\|\boldsymbol{W}(s)\|_2 \leq \bar{\rho}_w,\ \|\boldsymbol{U}(s)\|_2 \leq \bar{\rho}_u,\ \|\boldsymbol{a}(s)\|_2 \leq \bar{\rho}_a,\ s\in[0,\tau] \\
\lambda_s \geq \frac{\lambda_0}{2},\quad s\in[0,\tau] \\
\Phi(s+1) \leq \left(1-\eta\frac{\lambda_0}{2}\right)^s\Phi(0),\quad s\in[0,\tau]
\end{cases} \tag{19}$$

For $\tau = 0$, it is clear that Eq. (19) holds. Assume that Eq. (19) holds up to τ iterations.

(1) With the triangle inequality,

$$\begin{aligned}\|\boldsymbol{W}(\tau+1) - \boldsymbol{W}(0)\|_F \leq& \sum_{s=0}^{\tau} \|\boldsymbol{W}(s+1) - \boldsymbol{W}(s)\|_F \\ =& \sum_{s=0}^{\tau} \eta \|\nabla_{\boldsymbol{W}} \Phi(s)\|_F \\ \leq& \eta c_w \|\boldsymbol{X}\|_F \sum_{s=0}^{\tau} \|\hat{\boldsymbol{y}}(s) - \boldsymbol{y}\|_2 \\ \leq& \eta c_w \|\boldsymbol{X}\|_F \sum_{s=0}^{\tau} \left(1 - \eta\frac{\lambda_0}{2}\right)^{s/2} \|\hat{\boldsymbol{y}}(0) - \boldsymbol{y}\|_2,\end{aligned}$$

where the second inequality follows from Eq. (16), and the last one follows from induction assumption. Let $u \triangleq \sqrt{1 - \eta\lambda_0/2}$. Then $\|\boldsymbol{W}(\tau+1) - \boldsymbol{W}(0)\|_F$ can be bounded with

$$\frac{2}{\lambda_0}(1-u^2)\frac{1-u^{\tau+1}}{1-u} c_w \|\boldsymbol{X}\|_F \|\hat{\boldsymbol{y}}(0) - \boldsymbol{y}\|_2 \leq \frac{4}{\lambda_0} c_w \|\boldsymbol{X}\|_F \|\hat{\boldsymbol{y}}(0) - \boldsymbol{y}\|_2 \leq \delta, \quad \text{by Eq. (7).}$$

With Weyl's inequality, it is easy to have $\|\boldsymbol{W}(\tau+1)\|_2 \leq \bar{\rho}_w < 1$.

Using the similar technique, one can show that

$$\begin{aligned}&\|\boldsymbol{U}(\tau+1) - \boldsymbol{U}(0)\|_F \leq \sum_{s=0}^{\tau} \|\boldsymbol{U}(s+1) - \boldsymbol{U}(s)\|_F \\ &= \sum_{s=0}^{\tau} \eta \|\nabla_{\boldsymbol{U}} \Phi(s)\|_F \leq c_u \|\boldsymbol{X}\|_F \sum_{s=0}^{\tau} \|\hat{\boldsymbol{y}}(s) - \boldsymbol{y}\|_2 \\ &\leq \eta c_w \|\boldsymbol{X}\|_F \sum_{s=0}^{\tau} \left(1 - \eta\frac{\lambda_0}{2}\right)^{s/2} \|\hat{\boldsymbol{y}}(0) - \boldsymbol{y}\|_2 \\ &\leq \frac{4}{\lambda_0} c_u \|\boldsymbol{X}\|_F \|\hat{\boldsymbol{y}}(0) - \boldsymbol{y}\|_2 \leq \delta, \quad \text{by Eq. (8),}\end{aligned}$$

$$\begin{aligned}&\|\boldsymbol{a}(\tau+1) - \boldsymbol{a}(0)\|_F \leq \sum_{s=0}^{\tau} \|\boldsymbol{a}(s+1) - \boldsymbol{a}(s)\|_F \\ &= \sum_{s=0}^{\tau} \eta \|\nabla_{\boldsymbol{a}} \Phi(s)\|_F \leq c_a \|\boldsymbol{X}\|_F \sum_{s=0}^{\tau} \|\hat{\boldsymbol{y}}(s) - \boldsymbol{y}\|_2 \\ &\leq \eta c_w \|\boldsymbol{X}\|_F \sum_{s=0}^{\tau} \left(1 - \eta\frac{\lambda_0}{2}\right)^{s/2} \|\hat{\boldsymbol{y}}(0) - \boldsymbol{y}\|_2 \\ &\leq \frac{4}{\lambda_0} c_a \|\boldsymbol{X}\|_F \|\hat{\boldsymbol{y}}(0) - \boldsymbol{y}\|_2 \leq \delta, \quad \text{by Eq. (9).}\end{aligned}$$

By Weyl's inequality, it holds that $\|\boldsymbol{U}(\tau+1)\|_2 \leq \bar{\rho}_u$, and $\|\boldsymbol{a}(\tau+1)\|_2 \leq \bar{\rho}_a$.

(2) Next, using Eq. (17), we have

$$\begin{aligned}&\|\boldsymbol{Z}(\tau+1) - \boldsymbol{Z}(0)\|_F \\ \leq& \bar{\rho}_a^{-1} \left(c_w \|\boldsymbol{W}(\tau+1) - \boldsymbol{W}(0)\|_2 + c_u \|\boldsymbol{U}(\tau+1) - \boldsymbol{U}(0)\|_2\right) \|\boldsymbol{X}\|_F \\ \leq& \frac{4}{\lambda_0} \bar{\rho}_a^{-1} \left(c_w^2 + c_u^2\right) \|\boldsymbol{X}\|_F^2 \|\hat{\boldsymbol{y}}(0) - \boldsymbol{y}\|_2 \\ \leq& \frac{2-\sqrt{2}}{2}\sqrt{\lambda_0}, \quad \text{by Eq. (8)}\end{aligned}$$

By Wely's inequality, it implies that $\sigma_{\min}\left(\boldsymbol{Z}(\tau+1)\right) \geq \sqrt{\frac{\lambda_0}{2}}$. Thus, it holds $\lambda_{\tau+1} \geq \frac{\lambda_0}{2}$.

(3) Furthermore, we define $\boldsymbol{g} \triangleq \boldsymbol{a}(\tau+1)^\top \boldsymbol{Z}(\tau)$ and note that

$$\begin{aligned}
&\Phi(\tau+1) - \Phi(\tau) \\
=&\frac{1}{2}\left\|\hat{\boldsymbol{y}}(\tau+1) - \hat{\boldsymbol{y}}(\tau)\right\|_2^2 + \left(\hat{\boldsymbol{y}}(\tau+1) - \boldsymbol{g}\right)^\top \left(\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}\right) + \left(\boldsymbol{g} - \hat{\boldsymbol{y}}(\tau))\right)^\top \left(\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}\right).
\end{aligned}$$

We bound each term of the RHS of this equation individually. Firstly, using Eq. (18), we have

$$\begin{aligned}
&\left\|\hat{\boldsymbol{y}}(\tau+1) - \hat{\boldsymbol{y}}(\tau)\right\|_2 \\
\leq&\left(c_w\left\|\boldsymbol{W}(\tau+1) - \boldsymbol{W}(\tau)\right\|_2 + c_u\left\|\boldsymbol{U}(\tau+1) - \boldsymbol{U}(\tau)\right\|_2 + c_a\|\boldsymbol{a}(\tau+1) - \boldsymbol{a}(\tau)\|_2\right)\left\|\boldsymbol{X}\right\|_F \\
\leq&\eta \cdot C_1\left\|\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}\right\|_2,
\end{aligned}$$

where $C_1 \triangleq \left(c_w^2 + c_u^2 + c_a^2\right)\left\|\boldsymbol{X}\right\|_F^2$.

Secondly, by Eq. (17), we have

$$\begin{aligned}
&\left(\hat{\boldsymbol{y}}(\tau+1) - \boldsymbol{g}\right)^\top \left(\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}\right) \\
\leq&\left\|\boldsymbol{a}(\tau+1)\right\|_2 \left\|\boldsymbol{Z}(\tau+1) - \boldsymbol{Z}(\tau)\right\|_F \left\|\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}\right\|_2 \\
\leq&\left(c_w\left\|\boldsymbol{W}(\tau+1) - \boldsymbol{W}(\tau)\right\|_2 + c_u\left\|\boldsymbol{U}(\tau+1) - \boldsymbol{U}(\tau)\right\|_2\right)\left\|\boldsymbol{X}\right\|_F \left\|\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}\right\|_2 \\
\leq&\eta\left(c_w^2 + c_u^2\right)\left\|\boldsymbol{X}\right\|_F^2 \left\|\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}\right\|_2^2 \\
\leq&\eta \cdot C_2\left\|\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}\right\|_2^2,
\end{aligned}$$

where $C_2 \triangleq \left(c_w^2 + c_u^2\right)\left\|\boldsymbol{X}\right\|_F^2$.

Lastly, using the fact $\left(\boldsymbol{a}(\tau+1) - \boldsymbol{a}(\tau)\right)^\top = -\eta\nabla_{\boldsymbol{a}}\Phi(\tau)$, we have

$$\begin{aligned}
&\left(\boldsymbol{g} - \hat{\boldsymbol{y}}(\tau))\right)^\top \left(\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y})\right) \\
=&-\eta\left(\nabla_{\boldsymbol{a}}\Phi(\tau)\boldsymbol{Z}(\tau)\right)^\top \left(\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}\right) \\
=&-\eta(\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y})^\top \boldsymbol{Z}(\tau)^\top \boldsymbol{Z}(\tau)(\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}) \\
\leq&-\eta\frac{\lambda_0}{2}\|\hat{\boldsymbol{y}}(\tau) - \boldsymbol{y}\|_2^2,
\end{aligned}$$

where we use the induce assumption $\lambda_\tau > \frac{\lambda_0}{2}$.

Putting all bounds together, we have

$$\begin{aligned}
\Phi(\tau+1) &= \left(1 - \eta(\lambda_0 - \eta C_1^2 - 2C_2)\right)\Phi(\tau) \\
&\leq \left(1 - \eta(\lambda_0 - 4C_2)\right)\Phi(\tau), \quad \text{by the condition on } \eta \\
&\leq \left(1 - \eta\frac{\lambda_0}{2}\right)\Phi(\tau), \quad \text{by Eq. (9)}.
\end{aligned}$$

□

C Proof for Section 4.1

C.1 Proof of Lemma 3

Proof of Lemma 3. By Eq. (6), it is easy to show that for all $i, j \in [n]$ and $l \geq 1$,

$$\boldsymbol{K}_{ii}^{(l)} = \boldsymbol{K}_{jj}^{(l)}, \quad \boldsymbol{K}_{ij}^{(l)} = \boldsymbol{K}_{ji}^{(l)}.$$

Recall that we define $\cos\theta_{ij}^{(l)} = \frac{\sigma_w^2 \boldsymbol{K}_{ij}^{(l-1)} + d^{-1}\boldsymbol{x}_i^\top \boldsymbol{x}_j}{\sigma_w^2 \boldsymbol{K}_{ij}^{(l-1)}+1}$ and it holds that

$$\begin{aligned}
\boldsymbol{\Lambda}_{ij}^{(l)} &= \begin{bmatrix} \sigma_w^2 \boldsymbol{K}_{ii}^{(l-1)} + 1 & \sigma_w^2 \boldsymbol{K}_{ij}^{(l-1)} + \frac{1}{d}\boldsymbol{x}_i^\top \boldsymbol{x}_j \\ \sigma_w^2 \boldsymbol{K}_{ji}^{(l-1)} + \frac{1}{d}\boldsymbol{x}_j^\top \boldsymbol{x}_i & \sigma_w^2 \boldsymbol{K}_{ii}^{(l-1)} + 1 \end{bmatrix} \\
&= \left(\sigma_w^2 \boldsymbol{K}_{ii}^{(l-1)} + 1\right) \begin{bmatrix} 1 & \cos\theta_{ij}^{(l)} \\ \cos\theta_{ij}^{(l)} & 1 \end{bmatrix} \\
&= \rho^{(l)} \begin{bmatrix} 1 & \cos\theta_{ij}^{(l)} \\ \cos\theta_{ij}^{(l)} & 1 \end{bmatrix}.
\end{aligned}$$

For $i = j$, $\boldsymbol{\Lambda}_{ij}^{(l)} = \left(\sigma_w^2 \boldsymbol{K}_{ii}^{(l-1)} + 1\right) \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. By the homogeneity of ReLU, we have

$$\begin{aligned}
\boldsymbol{K}_{ii}^{(l)} &= 2\mathbb{E}_{(\mathrm{u},\mathrm{v})^\top \sim \mathcal{N}(0, \boldsymbol{\Lambda}_{ii}^{(l)})} \left[\phi(\mathrm{u})\phi(\mathrm{v})\right] \\
&= 2\left(\sigma_w^2 \boldsymbol{K}_{ii}^{(l-1)} + 1\right) \mathbb{E}_{(\mathrm{u}',\mathrm{v}')^\top \sim \mathcal{N}\left(0, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right)} \left[\phi(\mathrm{u}')\phi(\mathrm{v}')\right] \\
&= \left(\sigma_w^2 \boldsymbol{K}_{ii}^{(l-1)} + 1\right) \cdot Q(1) \\
&= \sigma_w^2 \boldsymbol{K}_{ii}^{(l-1)} + 1,
\end{aligned}$$

Note that $\boldsymbol{K}_{ii}^{(0)} = 0$, and it is easy to show that for all $i \in [n]$ and $l \geq 1$, it holds

$$\rho^{(l)} = \boldsymbol{K}_{ii}^{(l)} = \frac{1 - \sigma_w^{2l}}{1 - \sigma_w^2}.$$

For all $(i, j) \in [n] \times [n]$, we have

$$\begin{aligned}
\boldsymbol{K}_{ij}^{(l)} &= 2\mathbb{E}_{(\mathrm{u},\mathrm{v})^\top \sim \mathcal{N}(0, \boldsymbol{\Lambda}_{ii}^{(l)})} \left[\phi(\mathrm{u})\phi(\mathrm{v})\right] \\
&= 2\left(\sigma_w^2 \boldsymbol{K}_{ii}^{(l-1)} + 1\right) \mathbb{E}_{(\mathrm{u}',\mathrm{v}')^\top \sim \mathcal{N}\left(0, \begin{bmatrix} 1 & \cos\theta_{ij}^{(l)} \\ \cos\theta_{ij}^{(l)} & 1 \end{bmatrix}\right)} \left[\phi(\mathrm{u}')\phi(\mathrm{v}')\right] \\
&= \left(\sigma_w^2 \boldsymbol{K}_{ii}^{(l-1)} + 1\right) \cdot Q\left(\cos\theta_{ij}^{(l)}\right) \\
&= \rho^{(l)} Q\left(\cos\theta_{ij}^{(l)}\right).
\end{aligned}$$

Consequently, we prove Eq. (12).

By substituting Eq. (12) into the definition of $\cos\theta_{ij}^{(l)}$, one can show that

$$\cos\theta_{ij}^{(l)} = \frac{\sigma_w^2 \boldsymbol{K}_{ij}^{(l-1)} + \frac{1}{d}\boldsymbol{x}_i^\top \boldsymbol{x}_j}{\sigma_w^2 \boldsymbol{K}_{ij}^{(l-1)} + 1} = \frac{\left(\boldsymbol{K}_{ij}^{(l-1)} - 1\right) Q\left(\cos\theta_{ij}^{(l-1)}\right) + \frac{1}{d}\boldsymbol{x}_i^\top \boldsymbol{x}_j}{\boldsymbol{K}_{ij}^{(l-1)}}.$$

Therefore, we have

$$\cos\theta_{ij}^{(l)} = \frac{\left(\rho^{(l)} - 1\right) Q\left(\cos\theta_{ij}^{(l-1)}\right) + \frac{1}{d}\boldsymbol{x}_i^\top \boldsymbol{x}_j}{\rho^{(l)}}.$$

Letting $l \to \infty$, Eq. (14) is proved. □

C.2 Proof of Theorem 3

Proof of Theorem 3. (i) About $\left\|\boldsymbol{K} - \boldsymbol{K}^{(l)}\right\|_F$.

By the triangle inequality, we have

$$\begin{aligned}
&\left|\boldsymbol{K}_{ij}^{(l+1)} - \boldsymbol{K}_{ij}^{(l)}\right| \\
=&\left|\rho^{(l+1)} Q\left(\cos\theta_{ij}^{(l+1)}\right) - \rho^{(l)} Q\left(\cos\theta_{ij}^{(l)}\right)\right| \\
\leq&\left|\rho^{(l+1)} Q\left(\cos\theta_{ij}^{(l+1)}\right) - \rho^{(l+1)} Q\left(\cos\theta_{ij}^{(l)}\right)\right| + \left|\rho^{(l+1)} Q\left(\cos\theta_{ij}^{(l)}\right) - \rho^{(l)} Q\left(\cos\theta_{ij}^{(l)}\right)\right|.
\end{aligned}$$

We bound each term individually.

Firstly, using the fact that $|Q'(x)| \leq 1$, we have

$$\begin{aligned}
&\left|\rho^{(l+1)} Q\left(\cos\theta_{ij}^{(l+1)}\right) - \rho^{(l+1)} Q\left(\cos\theta_{ij}^{(l)}\right)\right| \\
\leq&\left|\rho^{(l+1)}\cos\theta_{ij}^{(l+1)} - \rho^{(l+1)}\cos\theta_{ij}^{(l)}\right| \\
\leq&\left|\rho^{(l+1)}\cos\theta_{ij}^{(l+1)} - \rho^{(l)}\cos\theta_{ij}^{(l)}\right| + \left|\rho^{(l)}\cos\theta_{ij}^{(l)} - \rho^{(l+1)}\cos\theta_{ij}^{(l)}\right| \\
\leq&\left|\rho^{(l+1)}\cos\theta_{ij}^{(l+1)} - \rho^{(l)}\cos\theta_{ij}^{(l)}\right| + \left|\rho^{(l)} - \rho^{(l+1)}\right| \\
=&\left|\sigma_w^2 \boldsymbol{K}_{ij}^{(l)} + \frac{1}{d}\boldsymbol{x}_i^\top \boldsymbol{x}_j - \left(\sigma_w^2 \boldsymbol{K}_{ij}^{(l-1)} + \frac{1}{d}\boldsymbol{x}_i^\top \boldsymbol{x}_j\right)\right| + \left|\frac{1-\sigma_w^{2l}}{1-\sigma_w^2} - \frac{1-\sigma_w^{2(l+1)}}{1-\sigma_w^2}\right| \\
=&\sigma_w^2\left|\boldsymbol{K}_{ij}^{(l)} - \boldsymbol{K}_{ij}^{(l-1)}\right| + \sigma_w^{2l},
\end{aligned}$$

where the first equality follows from the fact that $\rho^{(l+1)} = \sigma_w^2 \boldsymbol{K}_{ii}^{(l)} + 1$, and $\cos\theta_{ij}^{(l+1)} = \frac{\sigma_w^2 \boldsymbol{K}_{ij}^{(l)} + d^{-1}\boldsymbol{x}_i^\top \boldsymbol{x}_j}{\sigma_w^2 \boldsymbol{K}_{ii}^{(l)} + 1}$.

Secondly, using the fact that $|Q(x)| \leq 1$, we have

$$\left|\rho^{(l+1)} Q\left(\cos\theta_{ij}^{(l)}\right) - \rho^{(l)} Q\left(\cos\theta_{ij}^{(l)}\right)\right| \leq \left|\rho^{(l+1)} - \rho^{(l)}\right| = \sigma_w^{2l}.$$

Consequently, for $l \geq 1$, it holds that

$$\left|\boldsymbol{K}_{ij}^{(l+1)} - \boldsymbol{K}_{ij}^{(l)}\right| \leq \sigma_w^2\left|\boldsymbol{K}_{ij}^{(l)} - \boldsymbol{K}_{ij}^{(l-1)}\right| + 2\sigma_w^{2l}.$$

This implies that, for $l \geq 1$, we have

$$\left|\boldsymbol{K}_{ij}^{l} - \boldsymbol{K}_{ij}^{(l-1)}\right| \leq (2l-1)\sigma_w^{2(l-1)}.$$

Therefore, it holds that

$$\left|\boldsymbol{K}_{ij} - \boldsymbol{K}_{ij}^{(l)}\right| = \mathcal{O}\big(l\sigma_w^{2l}\big),$$

which implies that

$$\left\|\boldsymbol{K} - \boldsymbol{K}^{(l)}\right\|_F = \mathcal{O}\Big(n\sigma_w^l l^{\frac{1}{2}}\Big).$$

(ii) About the positive definiteness of $\boldsymbol{K}$.

The proof of this part is similar with those of [16, 17] which are based on Hermite polynomials. We refer the reader to [34] for a detailed introduction about Hermite polynomials.

Following from Lemma 7, for $(i, j) \in [n] \times [n]$, it holds that

$$\boldsymbol{K}_{ij} = \frac{1}{1-\sigma_w^2} Q(\cos\theta_{ij}) = \frac{1}{1-\sigma_w^2}\sum_{r=0}^{\infty} \mu_r^2(\phi)(\cos\theta_{ij})^r.$$

Let $\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_n]$ where $\boldsymbol{h}_1, \cdots, \boldsymbol{h}_n$ be unit vectors such that $\cos\theta_{ij} = \boldsymbol{h}_i^\top \boldsymbol{h}_j$ for all $(i,j) \in [n] \times [n]$. It is easy to check that $[(\boldsymbol{H}^\top \boldsymbol{H})^{(\odot r)}]_{ij} = (\boldsymbol{h}_i^\top \boldsymbol{h}_j)^r$ holds for all $(i,j) \in [n] \times [n]$. Then, $\boldsymbol{K}$ can be rewritten as

$$\boldsymbol{K} = \frac{1}{1-\sigma_w^2} \sum_{r=0}^{\infty} \mu_r^2(\phi) (\boldsymbol{H}^\top \boldsymbol{H})^{(\odot r)}. \tag{20}$$

Following from Lemma 6, we show that $\boldsymbol{K}$ is a sum of a series of PSD matrices. Thus, it suffices to show that $\boldsymbol{K}$ is strictly positive definite if there exists a r such that $\mu_r^2(\phi) \neq 0$ and $(\boldsymbol{H}^\top \boldsymbol{H})^{(\odot r)}$ is strictly positive definite.

For any unit vector $\boldsymbol{v} = [v_1, \cdots, v_n]^\top \in R^n$, it holds that

$$\begin{aligned}
\boldsymbol{v}^\top (\boldsymbol{H}^\top \boldsymbol{H})^{(\odot r)} \boldsymbol{v} &= \sum_{i,j} v_i v_j (\boldsymbol{h}_i^\top \boldsymbol{h}_j)^r \\
&= \sum_{i,j} v_i v_j (\cos\theta_{ij})^r \\
&= \sum_i v_i^2 + \sum_{i \neq j} v_i v_j (\cos\theta_{ij})^r \\
&= 1 + \sum_{i \neq j} v_i v_j (\cos\theta_{ij})^r
\end{aligned}$$

Let us define $\beta = \max_{i \neq j} |\cos\theta_{ij}|$. By Eq. (14), it holds that

$$\begin{aligned}
|\cos\theta_{ij}| &= \left| \sigma_w^2 Q(\cos\theta_{ij}) + (1-\sigma_w^2) \frac{1}{d} \boldsymbol{x}_i^\top \boldsymbol{x}_j \right| \\
&= \left| \sigma_w^2 Q(\cos\theta_{ij}) \right| + \left| (1-\sigma_w^2) \frac{1}{d} \boldsymbol{x}_i^\top \boldsymbol{x}_j \right| \\
&< \sigma_w^2 + 1 - \sigma_w^2 = 1.
\end{aligned}$$

for all $(i,j) \in [n] \times [n]$. The last inequality follows from that facts that $|Q(x)| \leq 1$ for $|x| \leq 1$ and $\max_{i \neq j} \left(\left| \frac{1}{d} \boldsymbol{x}_i^\top \boldsymbol{x}_j \right| \right) < 1$ (by Assumption 2). Therefore, it holds that

$$\beta < 1.$$

Taking $r > -\frac{\log(n)}{\log(\beta)}$, we have

$$\left| \sum_{i \neq j} v_i v_j (\cos\theta_{ij})^r \right| \leq \sum_{i \neq j} |v_i||v_j| \beta^r \leq \left(\sum_i |v_i| \right)^2 \beta^r \leq n\beta^r < 1.$$

By Weyl's inequality, it holds that $\boldsymbol{v}^\top (\boldsymbol{H}^\top \boldsymbol{H})^{(\odot r)} \boldsymbol{v} > 0$, *i.e.* $(\boldsymbol{H}^\top \boldsymbol{H})^{(\odot r)}$ is positive definite. Following from Lemma 7, it holds that $\mu_r^2(\phi) > 0$. Therefore, the positive definiteness of $\boldsymbol{K}$ is proved. $\square$

D Proof for Section 4.2

D.1 Proof of Theorem 4

Proof. Using standard bounds on the operator norm of Gaussian matrices, it holds *w.p.* $\geq 1 - \exp(m)$,

$$\left\| \boldsymbol{z}_i^{(l+1)} - \boldsymbol{z}_i^{(l)} \right\|_2 \leq 2\sqrt{2}\sigma_w \left\| \boldsymbol{z}_i^{(l)} - \boldsymbol{z}_i^{(l-1)} \right\|_2,$$

Therefore, it holds that

$$\left\| \boldsymbol{z}_i^{(l)} - \boldsymbol{z}_i^{(l-1)} \right\|_2 = \mathcal{O}\left(\left\| \boldsymbol{z}_i^{(1)} \right\|_2 \right),$$

and

$$\left\|z_i - z_i^{(l)}\right\|_2 = \mathcal{O}\left(\left(2\sqrt{2}\sigma_w\right)^l \left\|z_i^{(1)}\right\|_2\right).$$

For $z_i^{(1)}$, we have

$$\mathbb{E}\left[\frac{1}{m}\left(z_i^{(1)}\right)^\top z_i^{(1)}\right] = \mathbb{E}\left[\frac{1}{m}\phi(\boldsymbol{U}\boldsymbol{x}_i)^\top\phi(\boldsymbol{U}\boldsymbol{x}_i)\right] = 1.$$

Using Bernstein inequality, it holds *w.p.* $\geq 1 - \exp\{-\Omega\left(mt^2\right)\}$

$$\left|\frac{1}{m}\left(z_i^{(1)}\right)^\top z_i^{(1)} - 1\right| \leq t.$$

Consequently, we have

$$\begin{aligned}
\left|\boldsymbol{G}_{ij} - \boldsymbol{G}_{ij}^{(l)}\right| &= \left|z_i^\top z_j - \left(z_i^{(l)}\right)^\top\left(z_j^{(l)}\right)\right| \\
&\leq \left|z_i^\top z_j - z_i^\top z_j^{(l)}\right| + \left|z_i^\top z_j^{(l)} - \left(z_i^{(l)}\right)^\top\left(z_j^{(l)}\right)\right| \\
&\leq \|z_i\|_2\left\|z_j - z_j^{(l)}\right\|_2 + \left\|z_i^{(l)}\right\|_2\left\|z_i - z_i^{(l)}\right\|_2 \\
&\leq C\left(2\sqrt{2}\sigma_w\right)^L m\left(1+\sqrt{t}\right).
\end{aligned}$$

where C is an absolute positive constant. Lastly, letting t be an absolute positive constant, we prove Theorem 4 by applying the simple union bound. □

E Proof for Section 4.3

In this section, we define $\hat{\boldsymbol{G}}_{ij}^{(l)} = \mathbb{E}_{\mathbf{w}\sim\mathcal{N}(0,\boldsymbol{I})}[\phi(\mathbf{w}^\top\boldsymbol{h})\phi(\mathbf{w}^\top\boldsymbol{h}')]$. Combining Lemma 7 and the homogeneity of ReLU, we write $\hat{\boldsymbol{G}}_{ij}^{(l)}$ as

$$\begin{aligned}
\hat{\boldsymbol{A}}_{ij}^{(l)} &= \boldsymbol{h}^\top\boldsymbol{h}' \\
\cos\hat{\theta}_{ij}^{(l)} &= \frac{\hat{\boldsymbol{A}}_{ij}^{(l)}}{\sqrt{\hat{\boldsymbol{A}}_{ii}^{(l)}\hat{\boldsymbol{A}}_{jj}^{(l)}}} \\
\hat{\boldsymbol{G}}_{ij}^{(l)} &= \sqrt{\hat{\boldsymbol{A}}_{ii}^{(l)}\hat{\boldsymbol{A}}_{jj}^{(l)}}\,Q(\cos\hat{\theta}_{ij}^{(l)})
\end{aligned}$$

By the triangle inequality, we have

$$\left|\frac{1}{m}\boldsymbol{G}_{ij}^{(l)} - \boldsymbol{K}_{ij}^{(l)}\right| \leq \left|\frac{1}{m}\boldsymbol{G}_{ij}^{(l)} - \hat{\boldsymbol{G}}_{ij}^{(l)}\right| + \left|\hat{\boldsymbol{G}}_{ij}^{(l)} - \boldsymbol{K}_{ij}^{(l)}\right|. \tag{21}$$

E.1 Proof of Theorem 5

Lemma 9. *For $i = j$, with probability at least $1 - l\exp\{-\Omega(m\varepsilon^2) + \mathcal{O}\left(\frac{1}{\varepsilon}\right)\}$, it holds that*

$$\left|\frac{1}{m}\boldsymbol{G}_{ii}^{(l)} - \boldsymbol{K}_{ii}^{(l)}\right| \leq \varepsilon, \tag{22}$$

or equivalently, $\left|\frac{1}{m}\boldsymbol{G}_{ii}^{(l)} - \rho^{(l)}\right| \leq \varepsilon$.

Proof. Following Lemma 4, we reconstruct $\boldsymbol{G}_{ii}^{(l+1)}$ as

$$\boldsymbol{G}_{ii}^{(l+1)} = \phi\left(\boldsymbol{M}\boldsymbol{h}\right)^{\top}\phi\left(\boldsymbol{M}\boldsymbol{h}\right),$$

where $\|\boldsymbol{h}\|_2^2 = \boldsymbol{h}^{\top}\boldsymbol{h} = \frac{\sigma_w^2}{m}\boldsymbol{G}_{ii}^{(l)} + 1$.

(1) For *fixed* $\boldsymbol{h}$, by the standard Bernstein inequality, it holds *w.p.* $\geq 1 - \exp\{-\Omega\left(m\varepsilon^2\right)\}$,

$$\left|\frac{1}{m}\boldsymbol{G}_{ii}^{(l+1)} - \hat{\boldsymbol{G}}_{ii}^{(l+1)}\right| \leq \varepsilon.$$

(2) For *all* $\boldsymbol{h}$, note that the ε-net size is at most $\exp\{\mathcal{O}\big(l\log\frac{1}{\varepsilon}\big)\}$. Therefore, it holds *w.p.* $\geq 1 - \exp\{-\Omega(m\varepsilon^2) + \mathcal{O}\big(l\log\frac{1}{\varepsilon}\big)\}$,

$$\left|\frac{1}{m}\boldsymbol{G}_{ii}^{(l+1)} - \hat{\boldsymbol{G}}_{ii}^{(l+1)}\right| \leq \varepsilon.$$

(3) Substitute the choice of $\boldsymbol{h}$ such that $\boldsymbol{h}^{\top}\boldsymbol{h} = \frac{\sigma_w^2}{m}\boldsymbol{G}_{ii}^{(l)} + 1$. We have

$$\left|\hat{\boldsymbol{G}}_{ii}^{(l+1)} - \boldsymbol{K}_{ii}^{(l+1)}\right| = \sigma_w^2\left|\frac{1}{m}\boldsymbol{G}_{ii}^{(l)} - \boldsymbol{K}_{ii}^{(l)}\right|.$$

And we have, *w.p.* $\geq 1 - \exp\{-\Omega(m\varepsilon^2) + \mathcal{O}\big(l\log\frac{1}{\varepsilon}\big)\}$,

$$\left|\frac{1}{m}\boldsymbol{G}_{ii}^{(l+1)} - \boldsymbol{K}_{ii}^{(l+1)}\right| \leq \left|\frac{1}{m}\boldsymbol{G}_{ii}^{(l+1)} - \hat{\boldsymbol{G}}_{ii}^{(l+1)}\right| + \left|\hat{\boldsymbol{G}}_{ii}^{(l+1)} - \boldsymbol{K}_{ii}^{(l+1)}\right| \leq \sigma_w^2\left|\frac{1}{m}\boldsymbol{G}_{ii}^{(l)} - \boldsymbol{K}_{ii}^{(l)}\right| + \varepsilon$$

which implies that with probability at least $1 - l\exp\{-\Omega(m\varepsilon^2) + \mathcal{O}\big(l\log\frac{1}{\varepsilon}\big)\}$, we have

$$\left|\boldsymbol{G}_{ii}^{(l)} - \boldsymbol{K}_{ii}^{(l)}\right| \leq \frac{1-\sigma_w^{2l}}{1-\sigma_w^2}\varepsilon. \tag{23}$$

□

Lemma 10. *For $i \neq j$, with probability at least $1 - l^2\exp\{-\Omega(m\varepsilon^2) + \mathcal{O}\big(l\log\frac{1}{\varepsilon}\big)\}$, it holds that*

$$\left|\frac{1}{m}\boldsymbol{G}_{ij}^{(l)} - \boldsymbol{K}_{ij}^{(l)}\right| \leq \varepsilon.$$

Proof. Following Lemma 4, we reconstruct $\boldsymbol{G}_{ij}^{(l+1)}$ as

$$\boldsymbol{G}_{ij}^{(l+1)} = \phi\left(\boldsymbol{M}\boldsymbol{h}\right)^{\top}\phi\left(\boldsymbol{M}\boldsymbol{h}'\right),$$

where $\boldsymbol{h}^{\top}\boldsymbol{h}' = \frac{\sigma_w^2}{m}\boldsymbol{G}_{ij}^{(l)} + \frac{1}{d}\boldsymbol{x}_i^{\top}\boldsymbol{x}_j$.

(1) For *fixed* $\boldsymbol{h}$ and $\boldsymbol{h}'$, by the standard Bernstein inequality, we have *w.p.* $\geq 1 - \exp\{-\Omega(m\varepsilon^2)\}$

$$\left|\frac{1}{m}\boldsymbol{G}_{ij}^{(l+1)} - \hat{\boldsymbol{G}}_{ij}^{(l+1)}\right| \leq \varepsilon.$$

(2) For *all* $\boldsymbol{h}, \boldsymbol{h}'$, note that the ε-net size is at most $\exp\{\mathcal{O}\big(l\log\frac{1}{\varepsilon}\big)\}$. Therefore, *w.p.* $\geq 1 - \exp\{-\Omega(m\varepsilon^2) + \mathcal{O}\big(l\log\frac{1}{\varepsilon}\big)\}$, it holds that

$$\left|\frac{1}{m}\boldsymbol{G}_{ij}^{(l+1)} - \hat{\boldsymbol{G}}_{ij}^{(l+1)}\right| \leq \varepsilon.$$

(3) Substituting the choice of $\boldsymbol{h}$ and $\boldsymbol{h}'$ such that $\boldsymbol{h}^\top \boldsymbol{h}' = \frac{\sigma_w^2}{m}\boldsymbol{G}_{ij}^{(l)} + \frac{1}{d}\boldsymbol{x}_i^\top \boldsymbol{x}_j$. We have

$$
\begin{aligned}
&\left|\hat{\boldsymbol{G}}_{ij}^{(l+1)} - \boldsymbol{K}_{ij}^{(l+1)}\right| \\
=&\left|\sqrt{\hat{\boldsymbol{A}}_{ii}^{(l+1)}\hat{\boldsymbol{A}}_{jj}^{(l+1)}}Q(\cos\hat{\theta}_{ij}^{(l+1)}) - \rho^{(l+1)}Q(\cos\theta_{ij}^{(l+1)})\right| \\
\leq&\left|\left(\sqrt{\hat{\boldsymbol{A}}_{ii}^{(l+1)}\hat{\boldsymbol{A}}_{jj}^{(l+1)}} - \rho^{(l+1)}\right)Q(\cos\hat{\theta}_{ij}^{(l+1)})\right| + \left|\rho^{(l+1)}\left(Q(\cos\hat{\theta}_{ij}^{(l+1)}) - Q(\cos\theta_{ij}^{(l+1)})\right)\right| \\
\leq&\left|\sqrt{\hat{\boldsymbol{A}}_{ii}^{(l+1)}\hat{\boldsymbol{A}}_{jj}^{(l+1)}} - \rho^{(l+1)}\right| + \rho^{(l+1)}\left|\cos\hat{\theta}_{ij}^{(l+1)} - \cos\theta_{ij}^{(l+1)}\right|, \quad |Q(\cdot)| < 1 \text{ and } Q(\cdot) \text{ is 1-Lipschitz} \\
\leq&\left|\sqrt{\hat{\boldsymbol{A}}_{ii}^{(l+1)}\hat{\boldsymbol{A}}_{jj}^{(l+1)}} - \rho^{(l+1)}\right| + \left|\left(\sqrt{\hat{\boldsymbol{A}}_{ii}^{(l+1)}\hat{\boldsymbol{A}}_{jj}^{(l+1)}} + \rho^{(l+1)} - \sqrt{\hat{\boldsymbol{A}}_{ii}^{(l+1)}\hat{\boldsymbol{A}}_{jj}^{(l+1)}}\right)\cos\hat{\theta}_{ij}^{(l+1)} - \rho^{(l+1)}\cos\theta_{ij}^{(l+1)}\right| \\
\leq&2\left|\sqrt{\hat{\boldsymbol{A}}_{ii}^{(l+1)}\hat{\boldsymbol{A}}_{jj}^{(l+1)}} - \rho^{(l+1)}\right| + \left|\sqrt{\hat{\boldsymbol{A}}_{ii}^{(l+1)}\hat{\boldsymbol{A}}_{jj}^{(l+1)}}\cos\hat{\theta}_{ij}^{(l+1)} - \rho^{(l+1)}\cos\theta_{ij}^{(l+1)}\right|.
\end{aligned}
$$

From the definition of $\hat{\boldsymbol{G}}^{(l)}$, it holds that $\hat{\boldsymbol{A}}_{ii}^{(l+1)} = \frac{\sigma_w^2}{m}\boldsymbol{G}_{ii}^{(l)} + 1$. Applying Lemma 9, it holds *w.p.* $\geq 1 - l\exp\{-\Omega\left(m\varepsilon^2\right) + \mathcal{O}\left(l\log\frac{1}{\varepsilon}\right)\}$,

$$
\left|\sqrt{\hat{\boldsymbol{A}}_{ii}^{(l+1)}\hat{\boldsymbol{A}}_{jj}^{(l+1)}} - \rho^{(l+1)}\right| = \left|\sqrt{\left(\frac{\sigma_w^2}{m}\boldsymbol{G}_{ii}^{(l)} + 1\right)\left(\frac{\sigma_w^2}{m}\boldsymbol{G}_{jj}^{(l)} + 1\right)} - \left(\sigma_w^2\boldsymbol{K}_{ii}^{(l)} + 1\right)\right| \leq \varepsilon.
$$

Moreover, note that $\sqrt{\hat{\boldsymbol{A}}_{ii}^{(l+1)}\hat{\boldsymbol{A}}_{jj}^{(l+1)}}\cos\hat{\theta}_{ij}^{(l+1)} = \hat{\boldsymbol{A}}_{ij}^{(l+1)} = \frac{\sigma_w^2}{m}\boldsymbol{G}_{ij}^{(l)} + \frac{1}{d}\boldsymbol{x}_i^\top\boldsymbol{x}_j$ and $\rho^{(l+1)}\cos\theta_{ij}^{(l+1)} = \sigma_w^2\boldsymbol{K}_{ij}^{(l)} + \frac{1}{d}\boldsymbol{x}_i^\top\boldsymbol{x}_j$. Thus, it holds that

$$
\left|\sqrt{\hat{\boldsymbol{A}}_{ii}^{(l+1)}\hat{\boldsymbol{A}}_{jj}^{(l+1)}}\cos\hat{\theta}_{ij}^{(l+1)} - \rho^{(l+1)}\cos\theta_{ij}^{(l+1)}\right| = \sigma_w^2\left|\frac{1}{m}\boldsymbol{G}_{ij}^{(l)} - \boldsymbol{K}\right|.
$$

Thus, *w.p.* $\geq 1 - l\exp\{-\Omega\left(m\varepsilon^2\right) + \mathcal{O}\left(l\log\frac{1}{\varepsilon}\right)\}$, it holds that

$$
\left|\hat{\boldsymbol{G}}_{ij}^{(l+1)} - \boldsymbol{K}_{ij}^{(l+1)}\right| \leq \sigma_w^2\left|\frac{1}{m}\boldsymbol{G}_{ij}^{(l)} - \boldsymbol{K}_{ij}^{(l)}\right| + \varepsilon.
$$

Consequently, *w.p.* $\geq 1 - l\exp\{-\Omega(m\varepsilon^2) + \mathcal{O}\left(l\log\frac{1}{\varepsilon}\right)\}$, we have

$$
\left|\frac{1}{m}\boldsymbol{G}_{ij}^{(l+1)} - \boldsymbol{K}_{ij}^{(l+1)}\right| \leq \left|\frac{1}{m}\boldsymbol{G}_{ij}^{(l+1)} - \hat{\boldsymbol{G}}_{ij}^{(l+1)}\right| + \left|\hat{\boldsymbol{G}}_{ij}^{(l+1)} - \boldsymbol{K}_{ij}^{(l+)}\right| \leq \varepsilon + \sigma_w^2\left|\frac{1}{m}\boldsymbol{G}_{ij}^{(l)} - \boldsymbol{K}_{ij}^{(l)}\right|.
$$

By applying the induction argument, one can show that for $l \geq 1$, it holds *w.p.* $\geq 1 - l^2\exp\{-\Omega(m\varepsilon^2) + \mathcal{O}\left(l\log\frac{1}{\varepsilon}\right)\}$,

$$
\left|\frac{1}{m}\boldsymbol{G}_{ij}^{(l)} - \boldsymbol{K}_{ij}^{(l)}\right| \leq \varepsilon.
$$

□

Now we are ready to prove Theorem 5.

Proof of Theorem 5. Combing Lemmas 9 and 10 with the standard union bound, we have *w.p.* $\geq 1 - n^2 l^2\exp\{-\Omega(m\varepsilon^2) + \mathcal{O}\left(l\log\frac{1}{\varepsilon}\right)\}$

$$
\left\|\frac{1}{m}\boldsymbol{G}^{(l)} - \boldsymbol{K}^{(l)}\right\|_F \leq n\varepsilon.
$$

Take $\varepsilon = \left(2\sqrt{2}\sigma_w\right)^l$ and notice that $\sigma_w^2 < \frac{1}{8}$. It holds *w.p.* $\geq 1 - n^2 l^2 \exp\left\{-\Omega(8^l \sigma_w^{2l} m) + \mathcal{O}\left(l^2\right)\right\} \geq 1 - n^2 \exp\left\{-\Omega(8^l \sigma_w^{2l} m) + \mathcal{O}\left(l^2\right)\right\}$,

$$\left\| \frac{1}{m} \boldsymbol{G}^{(l)} - \boldsymbol{K}^{(l)} \right\|_F = \mathcal{O}\left(n \left(2\sqrt{2}\sigma_w\right)^l \right).$$

□