Johnson-Lindenstrauss Lemma

Richard Xu

December 9, 2021

1 Johnson-Lindenstrauss Lemma

Lemma 1 *let* $\boldsymbol{y} = \frac{1}{\sqrt{k}}\mathbf{W}\mathbf{x}$*, where* $W_{i,j} \sim \mathcal{N}(0,1)$*:*

$$\Pr\big((1-\epsilon)\|\mathbf{x}\|^2 \leq \|\boldsymbol{y}\|^2 \leq (1+\epsilon)\|\mathbf{x}\|^2\big) \geq 1 - 2\exp^{\frac{(2\epsilon^3 - 3\epsilon^2)}{12}k} \tag{1}$$

2 Proof of lemma 1

Let $W_{k\times d}$ be the random matrix where $W_{ij} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$ and $\mathbf{x} \in \mathbb{R}^d$ and:

$$\begin{aligned} \boldsymbol{y} &= \frac{1}{\sqrt{k}}\mathbf{W}\mathbf{x} \\ \implies y_i &= \frac{1}{\sqrt{k}}\sum_{j=1}^{d} W_{i,j}x_j \end{aligned} \tag{2}$$

2.1 some helper properties

2.1.1 $\mathbb{E}[\|\boldsymbol{y}\|^2] = \|\mathbf{x}\|^2$

we know $\mathbb{E}[y_i] = \frac{1}{\sqrt{k}}\sum_{j=1}^{d}\mathbb{E}[W_{i,j}]x_j = 0$

$$\begin{aligned} \mathbb{E}[y_i^2] &= \frac{1}{k}\mathbb{E}\Big[\Big(\sum_{j=1}^{d} W_{i,j}x_j\Big)^2\Big] \\ &= \frac{1}{k}\sum_{j=1}^{d}\mathbb{E}[W_{i,j}^2]x_j^2 \qquad \text{all terms for } \mathbb{E}[W_{i,j}W_{i,j'}] = 0 \\ &= \frac{1}{k}\sum_{j=1}^{d} x_j^2 = \frac{1}{k}\|\mathbf{x}\|^2 \end{aligned} \tag{3}$$

This implies that $y_i \sim \mathcal{N}\big(0, \frac{1}{k}\|\mathbf{x}\|^2\big)$

$$\begin{aligned} \mathbb{E}[\|\boldsymbol{y}\|^2] &= \mathbb{E}\big[\sum_i y_i^2\big] = \sum_i \mathbb{E}[y_i^2] \\ &= k\frac{1}{k}\|\mathbf{x}\|^2 \\ &= \|\mathbf{x}\|^2 \end{aligned} \tag{4}$$

2.1.2 introduce "normalized" $\hat{\boldsymbol{y}}$ such that $\hat{\boldsymbol{y}} \sim \mathcal{N}(0, \mathbf{I}_k)$:

$$
\begin{aligned}
&\text{let } \hat{\boldsymbol{y}} = \frac{\sqrt{k}}{\|\mathbf{x}\|}\boldsymbol{y} \quad \text{be the "normalized" version of } \boldsymbol{y} \\
&= \frac{\sqrt{k}}{\|\mathbf{x}\|}\frac{1}{\sqrt{k}}\mathbf{W}\mathbf{x} = \frac{1}{\|\mathbf{x}\|}\mathbf{W}\mathbf{x} \\
\implies \hat{y}_i &= \frac{\sqrt{k}}{\|\mathbf{x}\|}y_i = \frac{1}{\|\mathbf{x}\|^2}W_{i,:}\mathbf{x} \\
&\sim \mathcal{N}(0,1) \quad \text{using } y_i \sim \mathcal{N}\big(0, \frac{1}{k}\|\mathbf{x}\|^2\big)
\end{aligned} \tag{5}
$$

one can interchange $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ (no need to use properties above):

$$
\begin{aligned}
\hat{\boldsymbol{y}} &= \frac{\sqrt{k}}{\|\mathbf{x}\|}\boldsymbol{y} \\
\|\hat{\boldsymbol{y}}\|^2 &= \frac{k\boldsymbol{y}^\top\boldsymbol{y}}{\|\mathbf{x}\|^2} \\
&= \frac{k\|\boldsymbol{y}\|^2}{\|\mathbf{x}\|^2} \\
\implies \|\boldsymbol{y}\|^2 &= \frac{\|\hat{\boldsymbol{y}}\|^2\|\mathbf{x}\|^2}{k}
\end{aligned} \tag{6}
$$

2.2 prove $\Pr\big(\|\boldsymbol{y}\|^2 \geq (1+\epsilon)\|\mathbf{x}\|^2\big) \leq \exp^{\frac{(2\epsilon^3 - 3\epsilon^2)}{12}}k$

$$
\begin{aligned}
&\Pr\big(\|\boldsymbol{y}\|^2 \geq (1+\epsilon)\|\mathbf{x}\|^2\big) \qquad \text{not in terms of } k \\
&= \Pr\big(\frac{\|\hat{\boldsymbol{y}}\|^2\|\mathbf{x}\|^2}{k} \geq (1+\epsilon)\|\mathbf{x}\|^2\big) \\
&= \Pr\big(\|\hat{\boldsymbol{y}}\|^2 \geq k(1+\epsilon)\big) \qquad \text{in terms of } k
\end{aligned} \tag{7}
$$

using Chernoff bound for $\|\hat{\boldsymbol{y}}\|^2$:

$$
\begin{aligned}
\Pr\big(\|\hat{\boldsymbol{y}}\|^2 \geq k(1+\epsilon)\big) &\leq \frac{\mathbb{E}\big[\exp\big(\lambda\|\hat{\boldsymbol{y}}\|^2\big)\big]}{\exp(\lambda k(1+\epsilon))} \\
&= \frac{\mathbb{E}\big[\exp\big(\lambda\sum_{i=1}^k \hat{y}_i^2\big)\big]}{\exp(\lambda k(1+\epsilon))} \qquad \sum_i^k \hat{y}_i^2 \sim \chi^2(k) \\
&= \frac{(1-2\lambda)^{-\frac{k}{2}}}{\exp(\lambda k(1+\epsilon))} \\
&= \left(\frac{1}{\sqrt{1-2\lambda}\,\exp(\lambda(1+\epsilon))}\right)^k \\
&= \big((1+\epsilon)\exp^{-\epsilon}\big)^{\frac{k}{2}}
\end{aligned} \tag{8}
$$

this is because, by solving for λ to minimize the bound which obtains $\lambda = \frac{\epsilon}{2(1+\epsilon)}$ and then simplify, one obtains:

$$
\begin{aligned}
&\frac{1}{\sqrt{1-2\frac{\epsilon}{2(1+\epsilon)}}\ \exp\left(\frac{\epsilon}{2(1+\epsilon)}(1+\epsilon)\right)} \\
=&\frac{1}{\sqrt{1-\frac{\epsilon}{1+\epsilon}}\ \exp(\frac{\epsilon}{2})} \\
=&\frac{1}{\sqrt{\frac{1}{1+\epsilon}}}\ \exp^{-\frac{\epsilon}{2}} \\
=&(1+\epsilon)^{\frac{1}{2}}\ \left(\exp^{-\epsilon}\right)^{\frac{1}{2}} \\
=&\left((1+\epsilon)\exp^{-\epsilon}\right)^{\frac{1}{2}}
\end{aligned}
\tag{9}
$$

Derive the fact:

$$
\begin{aligned}
\log(1+x) &= \int_x \frac{1}{1+x}\mathrm{d}x = \int_x (1+x)^{-1}\mathrm{d}x \\
&= \int_x (1+0)^{-1} - (1+0)^{-2}x + \frac{2(1+0)^{-3}}{2!}x^2 - \frac{2\times 3(1+0)^{-4}}{3!}x^3 + \dots \quad \because \text{ Taylor expand about } x=0 \\
&= \int_x 1 - x + x^2 - x^3 + x^4 \\
&\le \int_x 1 - x + x^2 \quad \because\ 0<x<1 \quad \text{and} \quad (-x^3+x^4)<0,\ (-x^5+x^6)<0\dots \\
&= x - \frac{x^2}{2} + \frac{x^3}{3}
\end{aligned}
\tag{10}
$$

using the fact: $\log(1+x) < x - \frac{x^2}{2} + \frac{x^3}{3}$:

$$
\begin{aligned}
&(1+\epsilon) < \exp^{\epsilon-\frac{\epsilon^2}{2}+\frac{\epsilon^3}{3}} \\
\implies &\left((1+\epsilon)\exp^{-\epsilon}\right)^{\frac{k}{2}} \le \left(\exp^{\epsilon-\frac{\epsilon^2}{2}+\frac{\epsilon^3}{3}}\ \exp^{-\epsilon}\right)^{\frac{k}{2}} \\
&= \left(\exp^{-\frac{\epsilon^2}{2}+\frac{\epsilon^3}{3}}\right)^{\frac{k}{2}} \qquad \text{alternatively } \le \left(\exp^{-\frac{\epsilon^2}{2}+\frac{\epsilon^3}{2}}\right)^{\frac{k}{2}} = \exp^{-(\epsilon^2-\epsilon^3)\frac{k}{4}} \\
&= \left(\exp^{\frac{2\epsilon^3-3\epsilon^2}{6}}\right)^{\frac{k}{2}} \\
&= \exp^{\frac{(2\epsilon^3-3\epsilon^2)}{12}k}
\end{aligned}
\tag{11}
$$

so we proved that:

$$
\begin{aligned}
\Pr\left(\|\boldsymbol{y}\|^2 \ge (1+\epsilon)\|\mathbf{x}\|^2\right) &= \Pr\left(\|\hat{\boldsymbol{y}}\|^2 \ge k(1+\epsilon)\right) \quad \text{from Eq.(7)} \\
&\le \left((1+\epsilon)\exp^{-\epsilon}\right)^{\frac{k}{2}} \quad \text{from Eq.(8)} \\
&\le \exp^{\frac{(2\epsilon^3-3\epsilon^2)}{12}k} \quad \text{from Eq.(11)}
\end{aligned}
\tag{12}
$$

2.3 prove $\Pr\big(\|\boldsymbol{y}\|^2 \leq (1-\epsilon)\|\mathbf{x}\|^2\big) \leq (\exp^{-\frac{\epsilon^2}{2}+\frac{\epsilon^3}{3}})^{\frac{k}{2}}$

$$\begin{aligned}
&\Pr\big(\|\boldsymbol{y}\|^2 \leq (1-\epsilon)\|\mathbf{x}\|^2\big)\\
&=\Pr\Big(\frac{\|\hat{\boldsymbol{y}}\|^2\|\mathbf{x}\|^2}{k} \leq (1-\epsilon)\|\mathbf{x}\|^2\Big)\\
&=\Pr\big(\|\hat{\boldsymbol{y}}\|^2 \leq k(1-\epsilon)\big)\\
&=\Pr\big(\exp^{-\lambda\|\hat{\boldsymbol{y}}\|^2} \geq \exp^{-\lambda k(1-\epsilon)}\big) \quad \lambda > 0
\end{aligned} \tag{13}$$

using Chernoff bound:

$$\begin{aligned}
\Pr\big(\|\hat{\boldsymbol{y}}\|^2 \leq k(1-\epsilon)\big) &\leq \frac{\mathbb{E}\big[\exp^{-\lambda\|\hat{\boldsymbol{y}}\|^2}\big]}{\exp^{-\lambda k(1+\epsilon)}} \quad \lambda > 0\\
&= \frac{\mathbb{E}\big[\exp\big(-\lambda\sum_{i=1}^k \hat{y}_i^2\big)\big]}{\exp(-\lambda k(1+\epsilon))} \quad \because \sum_i^k \hat{y}_i^2 \sim \chi^2(k)\\
&= \frac{(1+2\lambda)^{-\frac{k}{2}}}{\exp^{\lambda k(1-\epsilon)}}\\
&= \left(\frac{1}{\sqrt{1+2\lambda}\ \exp^{-\lambda(1-\epsilon)}}\right)^k\\
&= \big((1-\epsilon)\exp^{-\epsilon}\big)^{\frac{k}{2}}
\end{aligned} \tag{14}$$

this is because, by solving for λ to minimize the bound which obtains $\lambda = \frac{\epsilon}{2(1-\epsilon)}$ and then simplify, one obtains:

$$\begin{aligned}
&\frac{1}{\sqrt{1+2\frac{\epsilon}{2(1-\epsilon)}}\ \exp\Big(\frac{-\epsilon}{2(1-\epsilon)}(1-\epsilon)\Big)}\\
=&\frac{1}{\sqrt{1-\frac{\epsilon}{1-\epsilon}}\ \exp(\frac{-\epsilon}{2})}\\
=&\frac{1}{\sqrt{\frac{1}{1-\epsilon}}}\ \exp^{\frac{\epsilon}{2}}\\
=&(1-\epsilon)^{\frac{1}{2}}\ (\exp^{\epsilon})^{\frac{1}{2}}\\
=&\big((1-\epsilon)\exp^{\epsilon}\big)^{\frac{1}{2}}
\end{aligned} \tag{15}$$

Derive the fact:

$$\begin{aligned}
&\log(1-x) < -x - \frac{x^2}{2} < -x - \frac{x^2}{2} - \frac{x^3}{3} \quad \text{for } x > 0\\
\implies &(1-\epsilon) < \exp^{-\epsilon-\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3}}
\end{aligned} \tag{16}$$

the above inequality can be proven as:

$$
\begin{aligned}
\log(1-x) &= \int_x \frac{-1}{1-x}\mathrm{d}x = -\int_x (1-x)^{-1}\mathrm{d}x \\
&= -\int_x (1-0)^{-1} + (1-0)^{-2}x + \frac{2(1-0)^{-3}}{2!}x^2 + \frac{2\times 3(1+0)^{-4}}{3!}x^3 + \dots \quad \because \text{ Taylor expand about } x=0 \\
&= \int_x -1 - x - x^2 - x^3 - x^4 - \dots \\
&\leq \int_x -1 - x - x^2 \quad \because\ 0 < x < 1 \\
&= -x - \frac{x^2}{2} - \frac{x^3}{3}
\end{aligned}
\tag{17}
$$

$$
\begin{aligned}
\Pr\big(\|\boldsymbol{y}\|^2 \leq (1-\epsilon)\|\mathbf{x}\|^2\big) &\leq ((1-\epsilon)\exp^{\epsilon})^{\frac{k}{2}} \\
&\leq \big(\exp^{-\epsilon-\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3}}\exp^{\epsilon}\big)^{\frac{k}{2}} \\
&= \big(\exp^{-\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3}}\big)^{\frac{k}{2}} \\
&\leq \big(\exp^{-\frac{\epsilon^2}{2}+\frac{\epsilon^3}{3}}\big)^{\frac{k}{2}} \quad \because \epsilon > 0 \\
&\leq \exp^{-(\epsilon^2-\epsilon^3)\frac{k}{4}} \quad \because \text{ alternative version, see Eq.(11)}
\end{aligned}
\tag{18}
$$

2.4 Putting together:

$$
\begin{aligned}
&\Pr\big(\|\boldsymbol{y}\|^2 \geq (1+\epsilon)\|\mathbf{x}\|^2 \cup \|\boldsymbol{y}\|^2 \leq (1-\epsilon)\|\mathbf{x}\|^2\big) \leq 2\exp^{\frac{(2\epsilon^3-3\epsilon^2)}{12}k} \\
\implies &\Pr\big(\|\boldsymbol{y}\|^2 \leq (1+\epsilon)\|\mathbf{x}\|^2 \cap \|\boldsymbol{y}\|^2 \geq (1-\epsilon)\|\mathbf{x}\|^2\big) \geq 1 - 2\exp^{\frac{(2\epsilon^3-3\epsilon^2)}{12}k} \\
&\implies \Pr\big((1-\epsilon)\|\mathbf{x}\|^2 \leq \|\boldsymbol{y}\|^2 \leq (1+\epsilon)\|\mathbf{x}\|^2\big) \geq 1 - 2\exp^{\frac{(2\epsilon^3-3\epsilon^2)}{12}k} \\
&\implies \Pr\big((1-\epsilon)\|\mathbf{x}\|^2 \leq \|\boldsymbol{y}\|^2 \leq (1+\epsilon)\|\mathbf{x}\|^2\big) \geq 1 - 2\exp^{-(\epsilon^2-\epsilon^3)\frac{k}{4}} \quad \text{alternative version}
\end{aligned}
\tag{19}
$$

another alternative version:

$$
\begin{aligned}
\implies \Pr\big((1-\epsilon)\|\mathbf{x}\|^2 \leq \|\boldsymbol{y}\|^2 \leq (1+\epsilon)\|\mathbf{x}\|^2\big) &\geq 1 - 2\exp^{-(\epsilon^2-\epsilon^3)\frac{k}{4}} \quad \text{alternative version} \\
&\geq 1 - 2\exp^{-\epsilon^2\frac{k}{4}} \quad \text{looser bound}
\end{aligned}
\tag{20}
$$

2.4.1 alternative expression of J-L Lemma

let $\delta = \exp^{\frac{(2\epsilon^3-3\epsilon^2)}{12}k}$ or any other variants:

$$
\begin{aligned}
\Pr\Big((1-\epsilon)\|\mathbf{x}\|^2 \leq \big\|\frac{1}{\sqrt{k}}W\mathbf{x}\big\|^2 \leq (1+\epsilon)\|\mathbf{x}\|^2\Big) &\geq 1 - 2\delta \\
\implies \Pr\Big((1-\epsilon) \leq \frac{1}{k}\frac{\big\|W\mathbf{x}\big\|^2}{\|\mathbf{x}\|^2} \leq (1+\epsilon)\Big) &\geq 1 - 2\delta \\
\implies \Pr\Big(\Big|\frac{1}{k}\frac{\big\|W\mathbf{x}\big\|^2}{\|\mathbf{x}\|^2} - 1\Big| \leq \epsilon\Big) &\geq 1 - 2\delta \\
\implies \Pr\Big(\Big|\frac{1}{k}\big\|W\mathbf{x}\big\|^2 - \|\mathbf{x}\|^2\Big| \leq \epsilon\|\mathbf{x}\|^2\Big) &\geq 1 - 2\delta
\end{aligned}
\tag{21}
$$

under special case where $\|\mathbf{x}\| = 1$

$$\implies \Pr\left(\left|\frac{1}{k}\|W\mathbf{x}\|^2 - 1\right| \leq \epsilon\right) \geq 1 - 2\delta \tag{22}$$

in the strong bound paper: letting $\epsilon = \frac{\gamma}{2}$ and $\|\mathbf{x}\| = 1$:

$$\begin{aligned} &\Pr\left(\left|\frac{1}{k}\|W\mathbf{x}\|^2 - 1\right| \leq \epsilon\right) \geq 1 - 2\delta \\ \implies &\Pr\left(\left|\frac{1}{k}\|W\mathbf{x}\|^2 - 1\right| \leq \frac{\gamma}{2}\right) \geq 1 - 2\exp^{-(\frac{\gamma}{2})^2 \frac{k}{4}} \\ &\geq 1 - 2\exp^{-\frac{k\gamma^2}{16}} \end{aligned} \tag{23}$$

3 corollary of J-L lemma

3.1 norm preserving for n data points after dimension reduction

Corollary 1.1 *let* $k \geq \left(\frac{24}{3\epsilon^2 - 2\epsilon^3} \log n\right)$ *then:* $\exists f : \mathbb{R}^d \to \mathbb{R}^k$, *s.t.:*

$$(1-\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|^2 \leq \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|^2 \leq (1+\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|^2 \qquad \forall \mathbf{x}_i, \mathbf{x}_j, \quad \forall 0 < \epsilon < 1 \tag{24}$$

https://cs.stanford.edu/people/mmahoney/cs369m/Lectures/lecture1.pdf
A few points about J-L lemma:

1. Lemma (1.2) in its raw form made no suggestion about what form $f(\cdot)$ must take. However, since we only try to prove existence, therefore, we only need to show when $f(\mathbf{x}) \equiv \mathbf{W}\mathbf{x}$
2. Since we try to prove $\forall \mathbf{x}_i, \mathbf{x}_j$ (or more generically, $\forall e_i$), then it is useful to use union bound the complement case for proof, where by De Morgan:

$$\Pr(e_1 \cap \cdots \cap e_n) \geq \delta \Leftrightarrow \Pr(\neg e_1 \cup \cdots \cup \neg e_n) \leq 1 - \delta \tag{25}$$

3. if we can show $\delta > 0$ it implies existence $\exists$ of condtion $(e_1 \cap \cdots \cap e_n)$ critieria. In addition, usually $\delta \equiv \delta(n, \epsilon)$, so we can show it's also true for all n, ϵ
4. It is easier to work with $\cup_i \neg e_i$, as we can just use union bound

3.1.1 choosing k in terms of n

looking at the (one side) J-L Lemma in Eq.(12). j is the only turnable parameter, so we need to choose an appropriate k. The clever choice is to let $k = \frac{24}{3\epsilon^2 - 2\epsilon^3} \log n$, so we have:

$$\begin{aligned} \Pr\left(\|\boldsymbol{y}\|^2 \geq (1+\epsilon)\|\mathbf{x}\|^2\right) &\leq \exp^{\frac{(2\epsilon^3 - 3\epsilon^2)}{12} k} \\ &= \exp^{\frac{(2\epsilon^3 - 3\epsilon^2)}{12}\left(\frac{24}{3\epsilon^2 - 2\epsilon^3} \log n\right)} \qquad \text{let } k = \frac{24}{3\epsilon^2 - 2\epsilon^3} \log n \\ &= \exp^{-2\log n} \\ &= n^{-2} \end{aligned} \tag{26}$$

Point to note here:

1. first point to note here is that n does not appear anywhere in J-L lemma proof upon til now. It is irrelvant for the purpose of just a **single** random projection. The purpose of introducing n is really pave the way for the future use of $\binom{n}{2}$ pairs.

2. choice of constant $\left(k = \frac{24}{3\epsilon^2 - 2\epsilon^3} \log n\right)$ term leaves n^{-2} after derivation. It removes both the ϵ and k

3.1.2 back to the dimension reduction case

let $\boldsymbol{y} = f(\mathbf{x}_i) - f(\mathbf{x}_j) \quad \mathbf{x} = \mathbf{x}_i - \mathbf{x}_j$:

$$\begin{aligned}
&\Pr\left(\|\boldsymbol{y}\|^2 \notin \left\{(1-\epsilon)\|\mathbf{x}\|^2 \cup (1+\epsilon)\|\mathbf{x}\|^2\right\}\right) \leq \frac{2}{n^2} \\
\implies &\Pr\left(\|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|^2 \notin \left[(1-\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|^2 \cup (1+\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|^2\right]\right) \leq \frac{2}{n^2} \\
\implies &\Pr\left(\exists_{i,j}\left\{\|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|^2 \notin \left[(1-\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|^2 \cup (1+\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|^2\right]\right\}\right) \leq \binom{n}{k}\frac{2}{n^2} \qquad \text{union bound} \\
&= \frac{n(n-1)}{2}\frac{2}{n^2} = \frac{n-1}{n} \\
&= 1 - \frac{1}{n} = 1 - \delta \\
\implies &\Pr\left(\forall_{i,j}\left\{(1-\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|2 \leq \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|2 \leq (1+\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|2\right\}\right) \geq \frac{1}{n}
\end{aligned} \tag{27}$$

3.2 inner product

Corollary 1.2 *Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ and that $\|\mathbf{u}\| \leq 1$ and $\|\mathbf{v}\| \leq 1$. Let $\boldsymbol{y}(\mathbf{x}) = \frac{1}{\sqrt{k}}\mathbf{W}\mathbf{x}$ where $W_{i,j} \sim \mathcal{N}(0,1)$ (or $W_{i,j} \sim U(-1,+1)$), then:*

$$\Pr\left(|\mathbf{u}^\top\mathbf{v} - \boldsymbol{y}(\mathbf{u})^\top\boldsymbol{y}(\mathbf{v})| \geq \epsilon\right) \leq 4\exp^{-(\epsilon^2-\epsilon^3)\frac{k}{4}} \tag{28}$$

3.2.1 proof

$$\begin{aligned}
&\Pr\left((1-\epsilon)\|\mathbf{x}\|^2 \leq \|\boldsymbol{y}\|^2 \leq (1+\epsilon)\|\mathbf{x}\|^2\right) \geq 1 - 2\exp^{-(\epsilon^2-\epsilon^3)\frac{k}{4}} \qquad \text{alternative version} \\
\implies &\Pr\left((1-\epsilon)\|\mathbf{u}-\mathbf{v}\|^2 \leq \|\mathbf{u}'-\mathbf{v}'\|^2 \leq (1+\epsilon)\|\mathbf{u}-\mathbf{v}\|^2\right) \geq 1 - 2\exp^{-(\epsilon^2-\epsilon^3)\frac{k}{4}} \qquad \text{let } \mathbf{x} \equiv \mathbf{u}-\mathbf{v} \text{ and } \boldsymbol{y} \equiv \mathbf{u}'-\mathbf{v}' \\
\implies &\Pr\left((1-\epsilon)\|\mathbf{u}+\mathbf{v}\|^2 \leq \|\mathbf{u}'+\mathbf{v}'\|^2 \leq (1+\epsilon)\|\mathbf{u}+\mathbf{v}\|^2\right) \geq 1 - 2\exp^{-(\epsilon^2-\epsilon^3)\frac{k}{4}} \qquad \text{let } \mathbf{x} \equiv \mathbf{u}+\mathbf{v} \text{ and } \boldsymbol{y} \equiv \mathbf{u}'+\mathbf{v}'
\end{aligned} \tag{29}$$

therefore, for both of the above last two lines to satisfy simultaneously, we have (union of the complement) with probability of at least $1 - 4\exp^{-(\epsilon^2-\epsilon^3)\frac{k}{4}}$:

$$\begin{aligned}
&(1-\epsilon)\|\mathbf{u}-\mathbf{v}\|^2 \leq \|\mathbf{u}'-\mathbf{v}'\|^2 \leq (1+\epsilon)\|\mathbf{u}-\mathbf{v}\|^2) \quad \text{and} \\
&(1-\epsilon)\|\mathbf{u}+\mathbf{v}\|^2 \leq \|\mathbf{u}'+\mathbf{v}'\|^2 \leq (1+\epsilon)\|\mathbf{u}+\mathbf{v}\|^2)
\end{aligned} \tag{30}$$

when both of these satisfy, then the above implies:

3.2.2 Take left portion

$$
\begin{aligned}
&(1-\epsilon)\|\mathbf{u}-\mathbf{v}\|^2 \leq \|\mathbf{u}'-\mathbf{v}'\|^2 \leq \underline{(1+\epsilon)\|\mathbf{u}-\mathbf{v}\|^2} \quad \text{and} \\
&\underbrace{(1-\epsilon)\|\mathbf{u}+\mathbf{v}\|^2 \leq \|\mathbf{u}'+\mathbf{v}'\|^2} \leq (1+\epsilon)\|\mathbf{u}+\mathbf{v}\|^2)
\end{aligned} \tag{31}
$$

$$
\begin{aligned}
4\mathbf{u}'^\top\mathbf{v}' &= \|\mathbf{u}'+\mathbf{v}'\|^2 - \|\mathbf{u}'-\mathbf{v}'\|^2 \\
&\geq \underbrace{(1-\epsilon)\|\mathbf{u}+\mathbf{v}\|^2} - \|\mathbf{u}'-\mathbf{v}'\|^2 \\
&\geq (1-\epsilon)\|\mathbf{u}+\mathbf{v}\|^2 - \underline{(1+\epsilon)\|\mathbf{u}-\mathbf{v}\|^2} \\
&= (1-\epsilon)(\|\mathbf{u}\|^2 + 2\mathbf{u}^\top\mathbf{v} + \|\mathbf{v}\|^2) - (1+\epsilon)(\|\mathbf{u}\|^2 - 2\mathbf{u}^\top\mathbf{v} + \|\mathbf{v}\|^2) \\
&= (\|\mathbf{u}\|^2 + 2\mathbf{u}^\top\mathbf{v} + \|\mathbf{v}\|^2 - \|\mathbf{u}\|^2 + 2\mathbf{u}^\top\mathbf{v} - \|\mathbf{v}\|^2) - \epsilon(\|\mathbf{u}\|^2 + 2\mathbf{u}^\top\mathbf{v} + \|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 - 2\mathbf{u}^\top\mathbf{v} + \|\mathbf{v}\|^2) \\
&= 4\mathbf{u}^\top\mathbf{v} - \epsilon(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2) \\
&= 4\mathbf{u}^\top\mathbf{v} - 4\epsilon \quad \because \|\mathbf{u}\| = 1 \text{ and } \|\mathbf{u}\| = 1 \\
\implies \mathbf{u}'^\top\mathbf{v}' &\geq \mathbf{u}^\top\mathbf{v} - \epsilon
\end{aligned} \tag{32}
$$

3.2.3 Combine with the right portion

we also use the two existing conditions in Eq.(30): (therefore no additional union bound is needed):

3.3 Own idea

to design an "norm persevering" neural network. In terms of expressibility: at each input and output, we can bound $\|\mathbf{x}^{(l)}\|$ and $\|\mathbf{x}^{(l+1)}\|$ with a desirable δ by tuning a neuron number k, with either a $W_{i,j} \sim \mathcal{N}(0,1)$. This then becomes a neural network architecture hyper-parameter selection scheme.

key challenges is to find the norm for $\|\sigma(\mathbf{x}^{(l+1)})\|$

Firstly we need to find motivation for the "norm preserving" feed-forward (there was a paper about norm preserving for ResNet).

3.4 Own idea

In terms of Radmarcher complexity, which requires $\sigma_i \sim \text{Rad}(\cdot)$, then we need to design a "matrix worth" of Radmarcher variables $\{\sigma_{i,j}\}$, one may require to compute the following quantity, where $\Pi =$

$$\Pi_{i,j} = \sigma_{i,j} f_{i,j} \tag{33}$$

then one needs to bound the following quantity again:

$$\|\Pi\mathbf{x}\|^2 - \|\mathbf{x}\|^2 \tag{34}$$

it seems to be very limiting, so how about let's using identical array of function $\mathbf{f}$ for all Rademarcher variables, and let $\Pi_{ij} = \sigma_{i,j}$:

$$\|\Pi\mathbf{f}\|^2 - \|\mathbf{f}\|^2 \tag{35}$$

the only problem is that how can we justify applying different rows of $\Pi_{i,:}$ to the same $\mathbf{f}$