1 Some Lemmas

Lemma 1.1. *Consider independent $W_1, ..., W_p \sim N(0, d^{-1}I_d)$ and $x_1, ..., x_n \sim N(0, I_d)$. Assume n^2/p and $n \log n/d$ are sufficiently small, and then there exists some constant $c > 0$, such that*

$$\inf_{\|\Delta\|=1} \frac{1}{p} \sum_{j=1}^{p} \left| \sum_{i=1}^{n} \psi(W_j^T x_i)\Delta_i \right| \geq c,$$

with high probability.

Proof. Let us adopt the notation that

$$f(W, X, \Delta) = \frac{1}{p} \sum_{j=1}^{p} \left| \sum_{i=1}^{n} \psi(W_j^T x_i)\Delta_i \right|.$$

Define $g(X, \Delta) = \mathbb{E}(f(W, X, \Delta)|X)$. We then have

$$\begin{aligned}
\inf_{\|\Delta\|=1} f(W, X, \Delta) &\geq \inf_{\|\Delta\|=1} \mathbb{E}f(W, X, \Delta) - \sup_{\|\Delta\|=1} |f(W, X, \Delta) - \mathbb{E}f(W, X, \Delta)| \\
&\geq \inf_{\|\Delta\|=1} \mathbb{E}f(W, X, \Delta) \qquad (1) \\
&\quad - \sup_{\|\Delta\|=1} |f(W, X, \Delta) - \mathbb{E}(f(W, X, \Delta)|X)| \qquad (2) \\
&\quad - \sup_{\|\Delta\|=1} |g(X, \Delta) - \mathbb{E}g(X, \Delta)|. \qquad (3)
\end{aligned}$$

We will analyze the three terms above separately.

Analysis of (1). Define $h(W_j) = \mathbb{E}(\psi(W_j^T x_i)|W_j)$ and $\bar{\psi}(W_j^T x_i) = \psi(W_j^T x_i) - h(W_j)$. We then have

$$\mathbb{E}f(W, X, \Delta) = \mathbb{E} \left| \sum_{i=1}^{n} \bar{\psi}(W^T x_i)\Delta_i + h(W) \sum_{i=1}^{n} \Delta_i \right|. \qquad (4)$$

A lower bound of (4) is

$$\mathbb{E}f(W, X, \Delta) \geq \left| \sum_{i=1}^{n} \Delta_i \right| |\mathbb{E}h(W)| - \mathbb{E} \left| \sum_{i=1}^{n} \bar{\psi}(W^T x_i)\Delta_i \right|,$$

where the second term can be bounded by

$$
\begin{aligned}
\mathbb{E}\left|\sum_{i=1}^n \bar\psi(W^T x_i)\Delta_i\right| &\le \sqrt{\mathbb{E}\left|\sum_{i=1}^n \bar\psi(W^T x_i)\Delta_i\right|^2} \\
&= \sqrt{\mathbb{E}\mathsf{Var}\left(\left|\sum_{i=1}^n \psi(W^T x_i)\Delta_i\right| \Big| W\right)} \\
&= \sqrt{\mathbb{E}\sum_{i=1}^n \Delta_i^2 \mathsf{Var}(\psi(W^T x_i)|W)} \\
&= \sqrt{\mathbb{E}\sum_{i=1}^n \Delta_i^2 \mathbb{E}(|\psi(W^T x_i)|^2|W)} \\
&= \sqrt{\mathbb{E}|\psi(W^T x)|^2} \le \sqrt{\mathbb{E}|W^T x|^2} = 1.
\end{aligned}
$$

Since

$$
\mathbb{E}h(W) = \frac{1}{\sqrt{2\pi}}\mathbb{E}\|W\| = \frac{1}{\sqrt{\pi}}\frac{\Gamma((d+1)/2)}{\sqrt{d}\,\Gamma(d/2)} \ge \frac{1}{\sqrt{2\pi}}\sqrt{\frac{d-1}{d}}.
$$

Therefore, as long as $d \ge 3$ and $\left|\sum_{i=1}^n \Delta_i\right| \ge 7$, we have $\mathbb{E}f(W, X, \Delta) \ge 1$, and we thus can conclude that

$$
\inf_{\|\Delta\|=1, |\sum_{i=1}^n \Delta_i| \ge 7} \mathbb{E}f(W, X, \Delta) \gtrsim 1. \tag{5}
$$

Now we consider the case $\left|\sum_{i=1}^n \Delta_i\right| < 7$. A lower bound for $f(W, X, \Delta)$ is

$$
f(W, X, \Delta) \ge \left|\sum_{i=1}^n \bar\psi(W^T x_i)\Delta_i\right| - 7h(W) = \left|\sum_{i=1}^n \bar\psi(W^T x_i)\Delta_i\right| - \frac{7}{\sqrt{2\pi}}\|W\|. \tag{6}
$$

Thus,

$$
\begin{aligned}
\mathbb{E}f(W, X, \Delta) &\ge \mathbb{E}\left(f(W, X, \Delta)\mathbb{I}\left\{\left|\sum_{i=1}^n \bar\psi(W^T x_i)\Delta_i\right| \ge 6, 1/2 \le \|W\|^2 \le 2\right\}\right) \\
&\ge \mathbb{P}\left(\left|\sum_{i=1}^n \bar\psi(W^T x_i)\Delta_i\right| \ge 6, 1/2 \le \|W\|^2 \le 2\right) \\
&= \mathbb{P}\left(\left|\sum_{i=1}^n \bar\psi(W^T x_i)\Delta_i\right| \ge 6 \Big| 1/2 \le \|W\|^2 \le 2\right)\mathbb{P}\left(1/2 \le \|W\|^2 \le 2\right) \\
&\ge \mathbb{P}\left(\left|\sum_{i=1}^n \bar\psi(W^T x_i)\Delta_i\right| \ge 6 \Big| 1/2 \le \|W\|^2 \le 2\right)\left(1 - 2\exp(-d/16)\right),
\end{aligned}
$$

where the last inequality is by Lemma 2.3. By direct calculations, we have

$$
\mathsf{Var}\left(\bar\psi(W^T x)|W\right) = \|W\|^2 \mathsf{Var}(\max(0, W^T x/\|W\|)|W) = \|W\|^2\frac{1-\pi^{-1}}{2}, \tag{7}
$$

and

$$\mathbb{E}\left(|\bar{\psi}(W^Tx)|^3|W\right) \leq 3\mathbb{E}\left(|\psi(W^Tx)|^3|W\right) + 3|h(W)|^3 \leq \frac{3}{2}\|W\|^3.$$

Therefore, by Lemma 2.1, we have

$$\begin{aligned}
& \mathbb{P}\left(\left|\sum_{i=1}^n \bar{\psi}(W^Tx_i)\Delta_i\right| \geq 6 \middle| 1/2 \leq \|W\|^2 \leq 2\right) \\
\geq \ & \mathbb{P}\left(\frac{\left|\sum_{i=1}^n \bar{\psi}(W^Tx_i)\Delta_i\right|}{\|W\|\sqrt{\frac{1-\pi^{-1}}{2}}} \geq 21 \middle| 1/2 \leq \|W\|^2 \leq 2\right) \\
\geq \ & \mathbb{P}\left(N(0,1) > 21\right) - \sup_{1/2 \leq \|W\|^2 \leq 2} 2\sqrt{3\sum_{i=1}^n |\Delta_i|^3 \frac{\mathbb{E}\left(|\bar{\psi}(W^Tx_i)|^3|W\right)}{\|W\|^3\left(\frac{1-\pi^{-1}}{2}\right)^{3/2}}} \\
\geq \ & \mathbb{P}\left(N(0,1) > 21\right) - 10\sqrt{\sum_{i=1}^n |\Delta_i|^3} \\
\geq \ & \mathbb{P}\left(N(0,1) > 21\right) - 10\max_{1 \leq i \leq n} |\Delta_i|^{3/2}.
\end{aligned}$$

Hence, when $\max_{1 \leq i \leq n} |\Delta_i|^{3/2} \leq \delta_0^{3/2} := \mathbb{P}\left(N(0,1) > 21\right)/20$ and $\left|\sum_{i=1}^n \Delta_i\right| < 7$, we can lower bound $\mathbb{E}f(W,X,\Delta)$ by an absolute constant, and we conclude that

$$\inf_{\|\Delta\|=1, |\sum_{i=1}^n \Delta_i| \leq 7, \max_{1 \leq i \leq n} |\Delta_i| \leq \delta_0} \mathbb{E}f(W,X,\Delta) \gtrsim 1. \tag{8}$$

Finally, we consider the case when $\max_{1 \leq i \leq n} |\Delta_i| > \delta_0$ and $\left|\sum_{i=1}^n \Delta_i\right| < 7$. Without loss of generality, we can assume $\Delta_1 > \delta_0$. Note that the lower bound (6) still holds, and thus we have

$$f(W,X,\Delta) \geq \bar{\psi}(W^Tx_1)\Delta_1 - \left|\sum_{i=2}^n \bar{\psi}(W^Tx_i)\Delta_i\right| - \frac{7}{\sqrt{2\pi}}\|W\|.$$

We then lower bound $\mathbb{E}f(W,X,\Delta)$ by

$$\begin{aligned}
& \mathbb{E}\left(f(W,X,\Delta)\mathbb{I}\left\{\bar{\psi}(W^Tx_1)\Delta_1 \geq 8, \left|\sum_{i=2}^n \bar{\psi}(W^Tx_i)\Delta_i\right| \leq 8, 1/2 \leq \|W\|^2 \leq 2\right\}\right) \\
\geq \ & \mathbb{P}\left(\bar{\psi}(W^Tx_1)\Delta_1 \geq 8, \left|\sum_{i=2}^n \bar{\psi}(W^Tx_i)\Delta_i\right| \leq 2 \middle| 1/2 \leq \|W\|^2 \leq 2\right)\mathbb{P}\left(1/2 \leq \|W\|^2 \leq 2\right) \\
\geq \ & \mathbb{P}\left(\bar{\psi}(W^Tx_1)\Delta_1 \geq 8 \middle| 1/2 \leq \|W\|^2 \leq 2\right) \\
& \times\mathbb{P}\left(\left|\sum_{i=2}^n \bar{\psi}(W^Tx_i)\Delta_i\right| \leq 2 \middle| 1/2 \leq \|W\|^2 \leq 2\right)\left(1 - 2\exp(-d/16)\right).
\end{aligned}$$

For any W that satisfies $1/2 \leq \|W\|^2 \leq 2$, we have

$$\begin{aligned}
\mathbb{P}\left(\bar{\psi}(W^T x_1)\Delta_1 \geq 8 \middle| W\right) &\geq \mathbb{P}\left(\bar{\psi}(W^T x_1) \geq 8/\delta_0 \middle| W\right) \\
&\geq \mathbb{P}\left(\psi(W^T x_1) \geq 8/\delta_0 + 1/\sqrt{\pi} \middle| W\right) \\
&\geq \mathbb{P}\left(W^T x_1 \geq 8/\delta_0 + 1/\sqrt{\pi} \middle| W\right) \\
&\geq \mathbb{P}\left(N(0,1) \geq \sqrt{2}8/\delta_0 + \sqrt{2/\pi}\right),
\end{aligned}$$

which is a constant. We also have

$$\begin{aligned}
& \mathbb{P}\left(\left|\sum_{i=2}^n \bar{\psi}(W^T x_i)\Delta_i\right| \leq 2 \middle| 1/2 \leq \|W\|^2 \leq 2\right) \\
\geq\ & 1 - \frac{1}{4}\mathsf{Var}\left(\sum_{i=2}^n \bar{\psi}(W^T x_i)\Delta_i \middle| W\right) \\
\geq\ & \frac{1}{2},
\end{aligned}$$

where the last inequality is by (7). Therefore, we have

$$\mathbb{E}f(W,X,\Delta) \geq \frac{1}{2}\left(1 - 2\exp(-d/16)\right)\mathbb{P}\left(N(0,1) \geq \sqrt{2}8/\delta_0 + \sqrt{2/\pi}\right) \gtrsim 1,$$

and we can conclude that

$$\inf_{\|\Delta\|=1, |\sum_{i=1}^n \Delta_i| \leq 7, \max_{1\leq i\leq n}|\Delta_i| \geq \delta_0} \mathbb{E}f(W,X,\Delta) \gtrsim 1. \tag{9}$$

In the end, we combine the three cases (5), (8), and (9), and we obtain the conclusion that $\inf_{\|\Delta\|=1} \mathbb{E}f(W,X,\Delta) \gtrsim 1$.

Analysis of (2). We shorthand the conditional expectation operator $\mathbb{E}(\cdot|X)$ by $\mathbb{E}^X$. Let $\widetilde{W}$ be an independent copy of W, and we first bound the moment generating function via a standard symmetrization argument. For any $\lambda > 0$,

$$\begin{aligned}
& \mathbb{E}^X \exp\left(\lambda \sup_{\|\Delta\|=1}\left|f(W,X,\Delta) - \mathbb{E}^X f(W,X,\Delta)\right|\right) \\
\leq\ & \mathbb{E}^X \exp\left(\lambda \mathbb{E}^{X,W} \sup_{\|\Delta\|=1}\left|f(W,X,\Delta) - f(\widetilde{W},X,\Delta)\right|\right) \\
\leq\ & \mathbb{E}^X \exp\left(\lambda \sup_{\|\Delta\|=1}\left|f(W,X,\Delta) - f(\widetilde{W},X,\Delta)\right|\right) \\
=\ & \mathbb{E}^X \exp\left(\lambda \sup_{\|\Delta\|=1}\left|\frac{1}{p}\sum_{j=1}^p \epsilon_j \left(\left|\sum_{i=1}^n \psi(W_j^T x_i)\Delta_i\right| - \left|\sum_{i=1}^n \psi(\widetilde{W}_j^T x_i)\Delta_i\right|\right)\right|\right) \\
\leq\ & \mathbb{E}^X \exp\left(2\lambda \sup_{\|\Delta\|=1}\left|\frac{1}{p}\sum_{j=1}^p \epsilon_j \left|\sum_{i=1}^n \psi(W_j^T x_i)\Delta_i\right|\right|\right),
\end{aligned} \tag{10}$$

where $\epsilon_1, ..., \epsilon_p$ are independent Rademacher random variables. Let us adopt the notation that

$$F(\epsilon, W, X, \Delta) = \frac{1}{p}\sum_{j=1}^{p}\epsilon_j\left|\sum_{i=1}^{n}\psi(W_j^T x_i)\Delta_i\right|.$$

We use a discretization argument. For the Euclidean sphere $S^{n-1} = \{\Delta \in \mathbb{R}^n : \|\Delta\| = 1\}$, there exists a subset $\mathcal{N} \subset S^{n-1}$, such that for any $\Delta \in S^{n-1}$, there exists a $\Delta' \in \mathcal{N}$ that satisfies $\|\Delta - \Delta'\| \leq 1/2$, and we also have the bound $\log|\mathcal{N}| \leq 2n$. See, for example, Lemma 5.2 of Vershynin [2010]. For any $\Delta \in S^{n-1}$ and the corresponding $\Delta' \in \mathcal{N}$ that satisfies $\|\Delta - \Delta'\| \leq 1/2$, we have

$$\begin{aligned}|F(\epsilon, W, X, \Delta)| &\leq |F(\epsilon, W, X, \Delta')| + |F(\epsilon, W, X, \Delta - \Delta')| \\ &\leq |F(\epsilon, W, X, \Delta')| + \frac{1}{2}\sup_{\|\Delta\|=1}|F(\epsilon, W, X, \Delta)|,\end{aligned}$$

which, by taking supremum over both sides, implies

$$\sup_{\|\Delta\|=1}|F(\epsilon, W, X, \Delta)| \leq 2\max_{\Delta\in\mathcal{N}}|F(\epsilon, W, X, \Delta)|.$$

Define $\bar{F}(\epsilon, X, \Delta) = \mathbb{E}^{\epsilon, X}F(\epsilon, W, X, \Delta)$, and then

$$\max_{\Delta\in\mathcal{N}}|F(\epsilon, W, X, \Delta)| \leq \max_{\Delta\in\mathcal{N}}|F(\epsilon, W, X, \Delta) - \bar{F}(\epsilon, X, \Delta)| + \max_{\Delta\in\mathcal{N}}|\bar{F}(\epsilon, X, \Delta)|.$$

In view of (10), we obtain the bound

$$\begin{aligned}&\mathbb{E}^X\exp\left(\lambda\sup_{\|\Delta\|=1}\left|f(W, X, \Delta) - \mathbb{E}^X f(W, X, \Delta)\right|\right) \\ \leq\ &\mathbb{E}^X\exp\left(4\lambda\max_{\Delta\in\mathcal{N}}|F(\epsilon, W, X, \Delta) - \bar{F}(\epsilon, X, \Delta)| + 4\lambda\max_{\Delta\in\mathcal{N}}|\bar{F}(\epsilon, X, \Delta)|\right) \\ \leq\ &\frac{1}{2}\sum_{\Delta\in\mathcal{N}}\mathbb{E}^X\exp\left(4\lambda|F(\epsilon, W, X, \Delta) - \bar{F}(\epsilon, X, \Delta)|\right) \qquad (11) \\ &+\frac{1}{2}\sum_{\Delta\in\mathcal{N}}\mathbb{E}^X\exp\left(4\lambda|\bar{F}(\epsilon, X, \Delta)|\right). \qquad (12)\end{aligned}$$

We will bound the two terms above on the event $E = \left\{ \sum_{i=1}^{n} \|x_i\|^2 \le 3nd \right\}$. For any $W, \widetilde{W}$, we have

$$
\begin{aligned}
\left| F(\epsilon, W, X, \Delta) - F(\epsilon, \widetilde{W}, X, \Delta) \right| &\le \frac{1}{p} \sum_{j=1}^{p} \sum_{i=1}^{n} \left| (\psi(W_j^T x_i) - \psi(\widetilde{W}_j^T x_i)) \Delta_i \right| \\
&\le \frac{1}{p} \sum_{j=1}^{p} \sum_{i=1}^{n} |(W_j - \widetilde{W}_j)^T x_i| |\Delta_i| \\
&\le \frac{1}{p} \sum_{j=1}^{p} \sum_{i=1}^{n} \|W_j - \widetilde{W}_j\| \|x_i\| |\Delta_i| \\
&\le \frac{1}{\sqrt{p}} \sqrt{\sum_{j=1}^{p} \|W_j - \widetilde{W}_j\|^2} \sqrt{\sum_{i=1}^{n} \|x_i\|^2} \\
&\le \sqrt{\frac{3n}{p}} \sqrt{\sum_{j=1}^{p} \|\sqrt{d} W_j - \sqrt{d} \widetilde{W}_j\|^2},
\end{aligned}
$$

where the last inequality holds under the event E. By Lemma 2.2, we have for any X such that E holds,

$$
\mathbb{P}\left(|F(\epsilon, W, X, \Delta) - \bar{F}(\epsilon, X, \Delta)| > t \middle| X \right) \le 2 \exp\left(-\frac{pt^2}{6n} \right),
$$

for any $t > 0$. The sub-Gaussian tail implies a bound for the moment generating function. By Lemma 5.5 of Vershynin [2010], we have

$$
\mathbb{E}^X \exp\left(4\lambda |F(\epsilon, W, X, \Delta) - \bar{F}(\epsilon, X, \Delta)| \right) \le \exp\left(C_1 \frac{n}{p} \lambda^2 \right),
$$

for some constant $C_1 > 0$. To bound the moment generating function of $\bar{F}(\epsilon, X, \Delta)$, we note that

$$
\begin{aligned}
|\bar{F}(\epsilon, X, \Delta)| &\le \left| \frac{1}{p} \sum_{j=1}^{p} \epsilon_j \right| \mathbb{E}^X \left| \sum_{i=1}^{n} \psi(W^T x_i) \Delta_i \right| \\
&\le \left| \frac{1}{p} \sum_{j=1}^{p} \epsilon_j \right| \sqrt{\sum_{i=1}^{n} \mathbb{E}^X |\psi(W^T x_i)|^2} \\
&\le \left| \frac{1}{p} \sum_{j=1}^{p} \epsilon_j \right| \sqrt{\sum_{i=1}^{n} \|x_i\|^2 / d} \le \sqrt{3n} \left| \frac{1}{p} \sum_{j=1}^{p} \epsilon_j \right|,
\end{aligned}
$$

where the last inequality holds under the event E. With an application of Hoeffding-type inequality (Lemma 5.9 of Vershynin [2010]), we have

$$
\mathbb{E}^X \exp\left(4\lambda |\bar{F}(\epsilon, X, \Delta)| \right) \le \mathbb{E} \exp\left(4\lambda \sqrt{3n} \left| \frac{1}{p} \sum_{j=1}^{p} \epsilon_j \right| \right) \le \exp\left(C_1 \frac{n}{p} \lambda^2 \right).
$$

Note that we can use the same constant C_1 by making its value sufficiently large. Plug the two moment generating function bounds into (11) and (12), and we obtain the bound

$$\mathbb{E}^X \exp\left(\lambda \sup_{\|\Delta\|=1} \left|f(W,X,\Delta) - \mathbb{E}^X f(W,X,\Delta)\right|\right) \leq \exp\left(C_1 \frac{n}{p}\lambda^2 + 2n\right),$$

for any X such that E holds. To bound (2), we apply Chernoff bound, and then

$$\mathbb{P}\left(\sup_{\|\Delta\|=1} |f(W,X,\Delta) - \mathbb{E}(f(W,X,\Delta)|X)| > t\right) \leq \exp\left(-\lambda t + C_1 \frac{n}{p}\lambda^2 + 2n\right).$$

Optimize over λ, set $t \asymp \sqrt{\frac{n^2}{p}}$, and we have

$$\sup_{\|\Delta\|=1} |f(W,X,\Delta) - \mathbb{E}(f(W,X,\Delta)|X)| \lesssim \sqrt{\frac{n^2}{p}},$$

with high probability.

Analysis of (3). We use a discretization argument. There exists a subset $\mathcal{N}_\zeta \subset S^{n-1}$, such that for any $\Delta \in S^{n-1}$, there exists a $\Delta' \in \mathcal{N}_\zeta$ that satisfies $\|\Delta - \Delta'\| \leq \zeta$, and we also have the bound $\log|\mathcal{N}| \leq n\log(1+2/\zeta)$ according to Lemma 5.2 of Vershynin [2010]. For any $\Delta \in S^{n-1}$ and the corresponding $\Delta' \in \mathcal{N}_\zeta$ that satisfies $\|\Delta - \Delta'\| \leq \zeta$, we have

$$\begin{aligned}
|g(X,\Delta) - \mathbb{E}g(X,\Delta)| &\leq |g(X,\Delta') - \mathbb{E}g(X,\Delta')| \\
&\quad + |g(X,\Delta-\Delta') - \mathbb{E}g(X,\Delta-\Delta')| \\
&\quad + 2\mathbb{E}g(X,\Delta-\Delta') \\
&\leq |g(X,\Delta') - \mathbb{E}g(X,\Delta')| \\
&\quad + \zeta \sup_{\|\Delta\|=1} |g(X,\Delta) - \mathbb{E}g(X,\Delta)| \\
&\quad + 2\zeta \sup_{\|\Delta\|=1} \mathbb{E}g(X,\Delta).
\end{aligned}$$

Take supremum over both sides, arrange the inequality, and we obtain the bound

$$\begin{aligned}
\sup_{\|\Delta\|=1} |g(X,\Delta) - \mathbb{E}g(X,\Delta)| &\leq (1-\zeta)^{-1} \max_{\Delta \in \mathcal{N}_\zeta} |g(X,\Delta) - \mathbb{E}g(X,\Delta)| && (13) \\
&\quad 2\zeta(1-\zeta)^{-1}\mathbb{E}g(X,\Delta). && (14)
\end{aligned}$$

To bound (13), we will use Lemma 2.2 together with a union bound argument. For any $X, \widetilde{X}$, we have

$$\begin{aligned} |g(X,\Delta) - g(\widetilde{X},\Delta)| &\le \mathbb{E}^X \left| \sum_{i=1}^n (\psi(W_j^T x_i) - \psi(W_j^T \widetilde{x}_j))\Delta_i \right| \\ &\le \mathbb{E}^X \sqrt{\sum_{i=1}^n \left(\psi(W_j^T x_i) - \psi(W_j^T \widetilde{x}_j)\right)^2} \\ &\le \sqrt{\sum_{i=1}^n \mathbb{E}^X \left(W_j^T (x_i - \widetilde{x}_i)\right)^2} \\ &= \frac{1}{\sqrt{d}} \sqrt{\sum_{i=1}^n \|x_i - \widetilde{x}_i\|^2}\,. \end{aligned}$$

Therefore, by Lemma 2.2,

$$\mathbb{P}\left(|g(X,\Delta) - g(\widetilde{X},\Delta)| > t\right) \le 2\exp\left(-\frac{dt^2}{2}\right),$$

for any $t > 0$. A union bound argument leads to

$$\mathbb{P}\left(\max_{\Delta\in\mathcal{N}_\zeta} |g(X,\Delta) - \mathbb{E}g(X,\Delta)| > t\right) \le 2\exp\left(-\frac{dt^2}{2} + n\log\left(1+\frac{2}{\zeta}\right)\right),$$

which implies that

$$\max_{\Delta\in\mathcal{N}_\zeta} |g(X,\Delta) - \mathbb{E}g(X,\Delta)| \lesssim \sqrt{\frac{n\log(1+2/\zeta)}{d}},$$

with high probability. For (14), we have

$$\mathbb{E}g(X,\Delta) \le \mathbb{E}\sqrt{\sum_{i=1}^n |\psi(W^T x_i)|^2} \le \sqrt{\sum_{i=1}^n \mathbb{E}|\psi(W^T x_i)|^2} \le \sqrt{n}.$$

Combining the bounds for (13) and (14), we have

$$\sup_{\|\Delta\|=1} |g(X,\Delta) - \mathbb{E}g(X,\Delta)| \lesssim \sqrt{\frac{n\log(1+2/\zeta)}{d}} + \zeta\sqrt{n},$$

with high probability as long as $\zeta \le 1/2$. We choose $\zeta = \frac{c}{\sqrt{n}}$ with a sufficiently small constant $c > 0$, and thus the bound is sufficiently small as long as $\frac{n\log n}{d}$ is sufficiently small.

Finally, combine results for (1), (2) and (3), and we obtain the desired conclusion as long as n^2/p and $n\log n/d$ are sufficiently small. $\square$

Lemma 1.2. *Consider independent* $W_1, ..., W_p \sim N(0, d^{-1}I_d)$ *and* $x_1, ..., x_n \sim N(0, I_d)$. *We define the matrices* $G, \bar{G} \in \mathbb{R}^{n\times n}$ *by*

$$G_{il} = \frac{1}{p}\sum_{j=1}^{p}\psi(W_j^T x_i)\psi(W_j^T x_l),$$

and

$$\bar{G}_{il} = \begin{cases} \frac{1}{2}, & i = l, \\ \frac{1}{2\pi} + \frac{1}{4}\frac{\bar{x}_i^T\bar{x}_l}{d} + \frac{1}{2\pi}\left(\frac{\|x_i\|}{\sqrt{d}} - 1 + \frac{\|x_l\|}{\sqrt{d}} - 1\right), & i \neq l. \end{cases}$$

Assume $d/\log n$ *is sufficiently large, and then*

$$\|G - \bar{G}\|_{\mathrm{op}}^2 \lesssim \frac{n^2}{p} + \frac{\log n}{d} + \frac{n^2}{d^2},$$

with high probability.

Proof. Define $\widetilde{G} \in \mathbb{R}^{n\times n}$ with entries $\widetilde{G}_{il} = \mathbb{E}\left(\psi(W^T x_i)\psi(W^T x_l)|X\right)$, and we first bound the difference between G and $\widetilde{G}$. Note that

$$\mathbb{E}(G_{il} - \widetilde{G}_{il})^2 = \mathbb{E}\mathsf{Var}(G_{il}|X) \leq \frac{1}{p}\mathbb{E}|\psi(W^T x_i)\psi(W^T x_l)|^2 = \frac{3}{2p}\mathbb{E}\|W\|^4 \leq 5p^{-1}.$$

We then have

$$\mathbb{E}\|G - \widetilde{G}\|_{\mathrm{op}}^2 \leq \mathbb{E}\|G - \widetilde{G}\|_{\mathrm{F}}^2 \leq \frac{5n^2}{p}.$$

By Markov's inequality,

$$\|G - \widetilde{G}\|_{\mathrm{op}}^2 \lesssim \frac{n^2}{p}, \tag{15}$$

with high probability.

Next, we study the diagonal entries of $\widetilde{G}$. For any $i \in [n]$, $\widetilde{G}_{ii} = \mathbb{E}(|\psi(W^T x_i)|^2|X) = \frac{\|x_i\|^2}{2d}$. By Lemma 2.3 and a union bound argument, we have

$$\mathbb{P}\left(\max_{1\leq i\leq n}|\widetilde{G}_{ii} - \bar{G}_{ii}| > \sqrt{\frac{t}{d}} + \frac{t}{d}\right) \leq 2ne^{-t},$$

for any $t > 0$. Choosing $t \asymp \log n$, we obtain the bound

$$\max_{1\leq i\leq n}|\widetilde{G}_{ii} - \bar{G}_{ii}| \lesssim \sqrt{\frac{\log n}{d}} + \frac{\log n}{d}, \tag{16}$$

with high probability.

Now we analyze the off-diagonal entries. We use the notation $\bar{x}_i = \frac{\sqrt{d}}{\|x_i\|}x_i$. For any $i \neq l$, we have

$$\begin{aligned}
\widetilde{G}_{il} \quad &= \quad \mathbb{E}\left(\psi(W^T\bar{x}_i)\psi(W^T\bar{x}_l)|X\right) && (17)\\
&\quad +\mathbb{E}\left((\psi(W^T x_i) - \psi(W^T\bar{x}_i))\psi(W^T\bar{x}_l)|X\right) && (18)\\
&\quad +\mathbb{E}\left(\psi(W^T\bar{x}_i)(\psi(W^T x_l) - \psi(W^T\bar{x}_l))|X\right) && (19)\\
&\quad +\mathbb{E}\left((\psi(W^T x_i) - \psi(W^T\bar{x}_i))(\psi(W^T x_l) - \psi(W^T\bar{x}_l))|X\right). && (20)
\end{aligned}$$

For first term on the right hand side of (17), we observe that $\mathbb{E}\left(\psi(W^T\bar{x}_i)\psi(W^T\bar{x}_l)|X\right)$ is a function of $\frac{\bar{x}_i^T\bar{x}_l}{d}$, and thus we can write

$$\mathbb{E}\left(\psi(W^T\bar{x}_i)\psi(W^T\bar{x}_l)|X\right) = f\left(\frac{\bar{x}_i^T\bar{x}_l}{d}\right),$$

where

$$f(\rho) = \begin{cases} \mathbb{E}\psi(\sqrt{1-\rho}U + \sqrt{\rho}Z)\psi(\sqrt{1-\rho}V + \sqrt{\rho}Z), & \rho \geq 0, \\ \mathbb{E}\psi(\sqrt{1+\rho}U - \sqrt{-\rho}Z)\psi(\sqrt{1+\rho}V + \sqrt{-\rho}Z), & \rho < 0, \end{cases}$$

with $U, V, Z \overset{iid}{\sim} N(0,1)$. By some direct calculations, we have $f(0) = \frac{1}{2\pi}$, $f'(0) = \frac{1}{4}$, and $\sup_{|\rho|\leq 0.2} \frac{|f'(\rho) - f'(0)|}{|\rho|} \lesssim 1$. Therefore, as long as $|\bar{x}_i^T\bar{x}_l|/d \leq 1/5$,

$$\left| f\left(\frac{\bar{x}_i^T\bar{x}_l}{d}\right) - \frac{1}{2\pi} - \frac{1}{4}\frac{\bar{x}_i^T\bar{x}_l}{d}\right| \leq C_1 \left|\frac{\bar{x}_i^T\bar{x}_l}{d}\right|^2,$$

for some constant $C_1 > 0$. By Lemma 2.4, we know that $\max_{i\neq l} |\bar{x}_i^T\bar{x}_l|/d \lesssim \sqrt{\frac{\log n}{d}} \leq 1/5$ with high probability, which then implies

$$\sum_{i\neq l}\left(\mathbb{E}\left(\psi(W^T\bar{x}_i)\psi(W^T\bar{x}_l)|X\right) - \bar{G}_{il}\right)^2 \leq C_1 \sum_{i\neq l}\left|\frac{\bar{x}_i^T\bar{x}_l}{d}\right|^2.$$

The term on the right hand side can be bounded by

$$\sum_{i\neq l}\left|\frac{\bar{x}_i^T\bar{x}_l}{d}\right|^4 \leq \frac{d}{\min_{1\leq i\leq n}\|x_i\|^2}\sum_{i\neq l}\left|\frac{x_i^T x_l}{d}\right|^4.$$

By Lemma 2.3, $\frac{d}{\min_{1\leq i\leq n}\|x_i\|^2} \lesssim 1$ with high probability. By integrating out the probability tail bound of $|x_i^T x_l|$ given in Lemma 2.4, we have $\sum_{i\neq l}\mathbb{E}\left|\frac{x_i^T x_l}{d}\right|^4 \lesssim \frac{n^2}{d^2}$, and by Markov's inequality, we have $\sum_{i\neq l}\left|\frac{x_i^T x_l}{d}\right|^4 \lesssim \frac{n^2}{d^2}$ with high probability.

We also need to analyze the contributions of (18) and (19). Observe the fact that $\mathbb{I}\{W^T x_i \geq 0\} = \mathbb{I}\{W^T\bar{x}_i \geq 0\}$, which implies

$$\begin{aligned} \psi(W^T x_i) - \psi(W^T\bar{x}_i) &= W^T(x_i - \bar{x}_i)\mathbb{I}\{W^T\bar{x}_i \geq 0\}\psi(W^T\bar{x}_l) \\ &= \left(\frac{\|x_i\|}{\sqrt{d}} - 1\right)\psi(W^T\bar{x}_i)\psi(W^T\bar{x}_l). \end{aligned} \tag{21}$$

Then, the sum of (18) and (19) can be written as

$$\left(\frac{\|x_i\|}{\sqrt{d}} - 1 + \frac{\|x_l\|}{\sqrt{d}} - 1\right) f\left(\frac{\bar{x}_i^T\bar{x}_l}{d}\right).$$

Note that

$$\begin{aligned}&\sum_{i\neq l}\left(\frac{\|x_i\|}{\sqrt{d}}-1+\frac{\|x_l\|}{\sqrt{d}}-1\right)^2\left[f\left(\frac{\bar{x}_i^T\bar{x}_l}{d}\right)-\frac{1}{2\pi}\right]^2\\ \lesssim\ &\sum_{i\neq l}\left(\frac{\|x_i\|}{\sqrt{d}}-1+\frac{\|x_l\|}{\sqrt{d}}-1\right)^4+\sum_{i\neq l}\left|\frac{\bar{x}_i^T\bar{x}_l}{d}\right|^4.\end{aligned}$$

We have already shown that $\sum_{i\neq l}\left|\frac{\bar{x}_i^T\bar{x}_l}{d}\right|^4\lesssim\frac{n^2}{d^2}$ with high probability. By integrating out the probability tail bound of Lemma 2.3, we have $\mathbb{E}\left(\frac{\|x_i\|}{\sqrt{d}}-1\right)^4\lesssim d^{-2}$, which then implies $\sum_{i\neq l}\mathbb{E}\left(\frac{\|x_i\|}{\sqrt{d}}-1+\frac{\|x_l\|}{\sqrt{d}}-1\right)^4\lesssim\frac{n^2}{d^2}$ and the corresponding high-probability bound by Markov's inequality.

Finally, we show that the contribution of (20) is negligible. By (21), we can write (20) as

$$\left(\frac{\|x_i\|}{\sqrt{d}}-1\right)\left(\frac{\|x_l\|}{\sqrt{d}}-1\right)\mathbb{E}\left(\psi(W^T\bar{x}_i)^2\psi(W^T\bar{x}_l)^2\Big|X\right),$$

whose absolute value can be bounded by $\frac{3}{2}\left|\frac{\|x_i\|}{\sqrt{d}}-1\right|\left|\frac{\|x_l\|}{\sqrt{d}}-1\right|$. Since

$$\sum_{i\neq l}\mathbb{E}\left(\frac{\|x_i\|}{\sqrt{d}}-1\right)^2\mathbb{E}\left(\frac{\|x_l\|}{\sqrt{d}}-1\right)^2\lesssim\frac{n^2}{d^2},$$

we can conclude that (20) is bounded by $O\left(\frac{n^2}{d^2}\right)$ with high probability by Markov's inequality.

Combining the analyses of (17), (18), (19) and (20), we conclude that $\sum_{i\neq l}(\widetilde{G}_{il}-\bar{G}_{il})^2\lesssim\frac{n^2}{d^2}$ with high probability. Together with (15) and (16), we obtain the desired bound for $\|G-\bar{G}\|_{\mathrm{op}}$. □

Lemma 1.3. *Consider independent $W_1,...,W_p\sim N(0,d^{-1}I_d)$, $x_1,...,x_n\sim N(0,I_d)$, and $\beta_1,...,\beta_p\sim N(0,1)$. We define the matrices $H,\bar{H}\in\mathbb{R}^{n\times n}$ by*

$$\begin{aligned}H_{il}&=\frac{x_i^Tx_l}{d}\frac{1}{p}\sum_{j=1}^p\beta_j^2\mathbb{I}\{W_j^Tx_i\geq 0,W_j^Tx_l\geq 0\},\\ \bar{H}_{il}&=\frac{1}{4}\frac{x_i^Tx_l}{\|x_i\|\|x_l\|}+\frac{1}{4}\mathbb{I}\{i=l\}.\end{aligned}$$

Assume $d/\log n$ is sufficiently large, and then

$$\|H-\bar{H}\|_{\mathrm{op}}^2\lesssim\frac{n^2}{pd}+\frac{n}{p}+\frac{\log n}{d}+\frac{n^2}{d^2},$$

with high probability.

Proof. Define $\widetilde{H} \in \mathbb{R}^{n\times n}$ with entries $\widetilde{H}_{il} = \frac{x_i^T x_l}{d}\mathbb{E}\left(\beta^2\mathbb{I}\{W^T x_i \geq 0, W^T x_l \geq 0\}\big|X\right)$, and we first bound the difference between H and $\widetilde{H}$. Note that

$$\mathbb{E}(H_{il} - \widetilde{H}_{il})^2 = \mathbb{E}\mathsf{Var}(H_{il}|X) \leq \frac{1}{p}\mathbb{E}\left(\frac{|x_i^T x_l|^2}{d^2}\beta^4\right) \leq \begin{cases} \frac{3}{pd}, & i \neq l, \\ 9p^{-1}, & i = l. \end{cases}$$

We then have

$$\mathbb{E}\|H - \widetilde{H}\|_{\rm op}^2 \leq \mathbb{E}\|H - \widetilde{H}\|_{\rm F}^2 \leq \frac{3n^2}{pd} + \frac{9n}{p}.$$

By Markov's inequality,

$$\|H - \widetilde{H}\|_{\rm op}^2 \lesssim \frac{n^2}{pd} + \frac{n}{p}, \tag{22}$$

with high probability.

Next, we study the diagonal entries of $\widetilde{H}$. For any $i \in [n]$, $\widetilde{H}_{ii} = \frac{\|x_i\|^2}{d}\mathbb{E}(\beta^2\mathbb{I}\{W^T x_i \geq 0\}|X) = \frac{\|x_i\|^2}{2d}$. The same analysis that leads to the bound (16) also implies that

$$\max_{1\leq i\leq n} |\widetilde{H}_{ii} - \bar{H}_{ii}| \lesssim \sqrt{\frac{\log n}{d}} + \frac{\log n}{d}, \tag{23}$$

with high probability.

Now we analyze the off-diagonal entries. Recall the notation $\bar{x}_i = \frac{\sqrt{d}}{\|x_i\|}x_i$. For any $i \neq l$, we have

$$\begin{aligned} \widetilde{H}_{il} &= \frac{\|x_i\|\|x_l\|}{d}\frac{\bar{x}_i^T\bar{x}_l}{d}\mathbb{P}\left(W^T\bar{x}_i \geq 0, W^T\bar{x}_l \geq 0|X\right) \\ &= \frac{\bar{x}_i^T\bar{x}_l}{d}\mathbb{P}\left(W^T\bar{x}_i \geq 0, W^T\bar{x}_l \geq 0|X\right) \\ &\quad + \left(\frac{\|x_i\|\|x_l\|}{d} - 1\right)\frac{\bar{x}_i^T\bar{x}_l}{d}\mathbb{P}\left(W^T\bar{x}_i \geq 0, W^T\bar{x}_l \geq 0|X\right). \end{aligned} \tag{24}$$

Since $\mathbb{P}\left(W^T\bar{x}_i \geq 0, W^T\bar{x}_l \geq 0|X\right)$ is a function of $\frac{\bar{x}_i^T\bar{x}_l}{d}$, we can write

$$\frac{\bar{x}_i^T\bar{x}_l}{d}\mathbb{P}\left(W^T\bar{x}_i \geq 0, W^T\bar{x}_l \geq 0|X\right) = f\left(\frac{\bar{x}_i^T\bar{x}_l}{d}\right), \tag{25}$$

where for $\rho > 0$,

$$\begin{aligned} f(\rho) &= \rho\mathbb{P}\left(\sqrt{1-\rho}U + \sqrt{\rho}Z \geq 0, \sqrt{1-\rho}V + \sqrt{\rho}Z\right) \\ &= \rho\mathbb{E}\mathbb{P}\left(\sqrt{1-\rho}U + \sqrt{\rho}Z \geq 0, \sqrt{1-\rho}V + \sqrt{\rho}Z|Z\right) \\ &= \rho\mathbb{E}\Phi\left(\sqrt{\frac{\rho}{1-\rho}}Z\right)^2, \end{aligned}$$

with $U, V, Z \overset{iid}{\sim} N(0,1)$ and $\Phi(\cdot)$ being the cumulative distribution function of $N(0,1)$. Similarly, for $\rho < 0$,

$$f(\rho) = \rho\mathbb{E}\left[\Phi\left(\sqrt{\frac{-\rho}{1+\rho}}Z\right)\left(1 - \Phi\left(\sqrt{\frac{-\rho}{1+\rho}}Z\right)\right)\right].$$

By some direct calculations, we have $f(0) = 0$, $f'(0) = \frac{1}{4}$, and

$$\sup_{|\rho|\leq 1/5} |f''(\rho)| \lesssim \sup_{|t|\leq 1/2} |\mathbb{E}\phi(tZ)\Phi(tZ)Z/t| + \sup_{|t|\leq 1/2} |\mathbb{E}\phi(tZ)Z/t|,$$

where $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$. For any $|t| \leq 1/2$,

$$|\mathbb{E}\phi(tZ)Z/t| = \left|\mathbb{E}\frac{\phi(tZ) - \phi(0)}{tZ}Z^2\right| = \left|\mathbb{E}\xi\phi(\xi)Z^2\right| \leq \frac{|t|}{\sqrt{2\pi}}\mathbb{E}|Z|^3 \lesssim 1,$$

where ξ is a scalar between 0 and tZ so that $|\xi| \leq |tZ|$. By a similar argument, we also have $\sup_{|t|\leq 1/2} |\mathbb{E}\phi(tZ)\Phi(tZ)Z/t| \lesssim 1$ so that $\sup_{|\rho|\leq 1/5} |f''(\rho)| \lesssim 1$. Therefore, as long as $|\bar{x}_i^T\bar{x}_l|/d \leq 1/5$,

$$\left|f\left(\frac{\bar{x}_i^T\bar{x}_l}{d}\right) - \frac{1}{4}\frac{\bar{x}_i^T\bar{x}_l}{d}\right| \leq C_1\left|\frac{\bar{x}_i^T\bar{x}_l}{d}\right|^2,$$

for some constant $C_1 > 0$. By Lemma 2.4, we know that $\max_{i\neq l} |\bar{x}_i^T\bar{x}_l|/d \lesssim \sqrt{\frac{\log n}{d}} \leq 1/5$ with high probability. In view of the identities (24) and (25), we then have the high probability bound,

$$\begin{aligned}
\sum_{i\neq l}\left(\widetilde{H}_{il} - \frac{1}{4}\frac{\bar{x}_i^T\bar{x}_l}{d}\right)^2 &\leq 2\sum_{i\neq l}\left(\frac{\|x_i\|\|x_l\|}{d} - 1\right)^2\left|\frac{\bar{x}_i^T\bar{x}_l}{d}\right|^2 + 2C_1\sum_{i\neq l}\left|\frac{\bar{x}_i^T\bar{x}_l}{d}\right|^4 \\
&\leq \sum_{i\neq l}\left(\frac{\|x_i\|\|x_l\|}{d} - 1\right)^4 + (2C_1 + 1)\sum_{i\neq l}\left|\frac{\bar{x}_i^T\bar{x}_l}{d}\right|^4. \qquad (26)
\end{aligned}$$

For the first term on the right hand side of (26), we use Lemma 2.4 and obtain a probability tail bound for $|\|x_i\|\|x_l\| - d|$. By integrating out this tail bound, we have

$$\sum_{i\neq l}\mathbb{E}\left(\frac{\|x_i\|\|x_l\|}{d} - 1\right)^4 \lesssim \frac{n^2}{d^2},$$

which, by Markov's inequality, implies $\sum_{i\neq l}\left(\frac{\|x_i\|\|x_l\|}{d} - 1\right)^4 \lesssim \frac{n^2}{d^2}$ with high probability. Using the same argument in the proof of Lemma 1.2, we have $\sum_{i\neq l}\left|\frac{x_i^Tx_l}{d}\right|^4 \lesssim \frac{n^2}{d^2}$ with high probability. Finally, combining (22), (23), and the bound for (26), we obtain the desired bound for $\|H - \bar{H}\|_{\mathrm{op}}$. $\square$

2 Technical Lemmas

Lemma 2.1. *If* $Z \sim N(0,1)$ *and* $W = \sum_{i=1}^{n} X_i$ *where* X_i *are independent mean* 0 *and* $\mathsf{Var}(W) = 1$*, then*

$$\sup_z |\mathbb{P}(W \leq z) - \mathbb{P}(Z \leq z)| \leq 2\sqrt{3\sum_{i=1}^{n} \mathbb{E}|X_i|^3}.$$

Talagrand's inequality, see reference of Massart's paper About the Constants in Talagrand's Concentration Inequalities.

Lemma 2.2. *Let* $f : \mathbb{R}^k \to \mathbb{R}$ *be a Lipschitz function with constant* $L > 0$*. That is,* $|f(x) - f(y)| \leq L\|x - y\|$ *for all* $x, y \in \mathbb{R}^k$*. Then, for any* $t > 0$*,*

$$\mathbb{P}\left(|f(Z) - \mathbb{E}f(Z)| > t\right) \leq 2\exp\left(-\frac{t^2}{2L^2}\right),$$

where $Z \sim N(0, I_k)$*.*

Lemma 2.3. *For any* $t > 0$*, we have*

$$\begin{aligned} \mathbb{P}\left(\chi_k^2 \geq k + 2\sqrt{tk} + 2t\right) &\leq e^{-t}, \\ \mathbb{P}\left(\chi_k^2 \leq k - 2\sqrt{tk}\right) &\leq e^{-t}. \end{aligned}$$

Lemma 2.4. *Consider independent* $Y_1, Y_2 \sim N(0, I_k)$*. For any* $t > 0$*, we have*

$$\begin{aligned} \mathbb{P}\left(|\|Y_1\|\|Y_2\| - k| \geq 2\sqrt{tk} + 2t\right) &\leq 4e^{-t}, \\ \mathbb{P}\left(|Y_1^T Y_2| \geq \sqrt{2kt} + 2t\right) &\leq 2e^{-t}. \end{aligned}$$

Proof. By Lemma 2.3, we have

$$\begin{aligned} & \mathbb{P}\left(\|Y_1\|\|Y_2\| - k \geq 2\sqrt{tk} + 2t\right) \\ \leq\ & \mathbb{P}\left(\|Y_1\|^2 \geq k + 2\sqrt{tk} + 2t\right) + \mathbb{P}\left(\|Y_2\|^2 \geq k + 2\sqrt{tk} + 2t\right) \\ \leq\ & 2e^{-t}, \end{aligned}$$

and

$$\begin{aligned} & \mathbb{P}\left(\|Y_1\|\|Y_2\| - k \leq -2\sqrt{tk} - 2t\right) \\ \leq\ & \mathbb{P}\left(\|Y_1\|^2 \leq k - 2\sqrt{tk}\right) + \mathbb{P}\left(\|Y_2\|^2 \leq k - 2\sqrt{tk}\right) \\ \leq\ & 2e^{-t}. \end{aligned}$$

Summing up the two bounds above, we obtain the first conclusion. For the second conclusion, note that

$$\mathbb{P}\left(Y_1^T Y_2 \geq x\right) \leq e^{-\lambda x} \mathbb{E} e^{\lambda Y_1^T Y_2} = \exp\left(-\lambda x - \frac{k}{2}\log(1-\lambda^2)\right) \leq \exp\left(-\lambda x + \frac{k}{2}\lambda^2\right),$$

for any $x > 0$ and $\lambda \in (0,1)$. Optimize over $\lambda \in (0,1)$, and we obtain $\mathbb{P}\left(Y_1^T Y_2 > x\right) \leq e^{-\frac{1}{2}\left(\frac{x^2}{k} \wedge x\right)}$. Take $x = \sqrt{2kt} + 2t$, and then we obtain the bound

$$\mathbb{P}\left(Y_1^T Y_2 \geq \sqrt{2kt} + 2t\right) \leq e^{-t},$$

which immediately implies the second conclusion. □

References

R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.