

Discussion of “Vintage factor analysis with Varimax performs statistical inference”

Rungang Han* and Anru R. Zhang†

(May 11, 2022)

We wholeheartedly congratulate Drs. Rohe and Zeng for their insightful paper on vintage factor analysis with Varimax rotation. Varimax rotation is a basic scheme to simplify the expression of a particular subspace and is included in build-in standard packages *stats* in R and *PROC FACTOR* statement in SAS. Drs. Rohe and Zeng’s paper nicely shows that the principal component analysis with Varimax rotation actually performs statistical inference for the explainable factors.

In the paper, Drs. Rohe and Zeng suggested that leptokurtosis is a key identifiability condition for Varimax rotation. In addition, the number of factors often increases as the data dimension and sample size grow. It is natural to ask whether Varimax works with a vanishing leptokurtosis and/or a growing number of factors. In this discussion, we consider when Varimax recovers the subspace rotation in such the high-dimensional regimes. As a first step, we assume the factor matrix $Z \in \mathbb{R}^{n \times k}$ includes a collection of i.i.d. centered random variables satisfying

$$\mathbb{E}Z_{ij} = \mathbb{E}Z_{ij}^3 = 0, \quad \mathbb{E}Z_{ij}^2 = 1, \quad \mathbb{E}Z_{ij}^4 = \kappa > 3. \quad (1)$$

We also assume Z_{ij} ’s are sub-Gaussian such that $\mathbb{E}\exp(\lambda Z_{ij}) \leq e^{c\lambda^2}, \forall \lambda \in \mathbb{R}$ for some constant $c > 0$. Let $\hat{Z} \in \mathbb{R}^{n \times r}$ be the observed factors generated as

$$\hat{Z} = ZR^*,$$

where R^* is an unknown k -dimensional orthogonal matrix that represents the rotation to be recovered. Due to vanishing mean of Z , we focus on the following centered Varimax:

$$\hat{R} = \arg \max_{R \in \mathcal{O}(k)} \sum_{i=1}^n \sum_{j=1}^k \left((\hat{Z}R^\top)_{ij} \right)^4,$$

*Department of Statistical Science, Duke University

†Departments of Biostatistics & Bioinformatics, Computer Science, Mathematics, and Statistical Science, Duke University

where $\mathcal{O}(k)$ is the set of k -by- k orthogonal matrices. Consider the following error metric:

$$\text{dist}(R^*, \hat{R}) = \min_{P \in \mathcal{P}(k)} \frac{\|\hat{R} - PR^*\|_F}{\|R^*\|_F} = k^{-1/2} \min_{P \in \mathcal{P}(k)} \|\hat{R}R^{*\top} - P\|_F,$$

where $\mathcal{P}(k) = \mathcal{O}(k) \cap \{0, \pm 1\}^{k^2}$ is the set of orthogonal matrices that allow for column reordering and sign changes as defined in Eqn. (12) in Rohe and Zeng. The following theorem characterizes the conditions of n, k, κ under which Varimax works or fails.

Theorem 1. *Let $\delta \in (0, 1/2]$ be any fixed value and $\kappa \leq C_0$ for some universal constant $C_0 > 3$.*

- *If $n \gtrsim \max \left\{ \frac{k \log n}{(\kappa-3)^2}, \frac{k^2 \log^2 n}{\kappa-3} \right\}$, then*

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left(\text{dist}(R^*, \hat{R}) \geq \delta \right) = 0;$$

- *if $k < n \lesssim k^2$, then*

$$\liminf_{n, k \rightarrow \infty} \mathbb{P} \left(\text{dist}(R^*, \hat{R}) \geq \delta \right) = 1.$$

If the kurtosis κ is a constant, Theorem 1 suggests that a necessary and nearly sufficient condition for consistent rotation recovery is $n \gtrsim k^2$ – this bound is tight up to $\log^2(n)$. Theorem 1 also shows when the leptokurtosis of factor is insignificant, i.e., $\kappa \rightarrow 3^+$, a sufficient sample size to ensure rotation recovery by Varimax is $\max\{k \log(n)/(\kappa-3)^2, k^2 \log^2(n)/(\kappa-3)\}$, while it is unclear if this bound is sharp. It is of future interest to investigate the tight condition that guarantees the consistency of Varimax.

Proof of Theorem 1. For any random matrix Z and fixed $a \in \mathbb{R}^k$, define the following stochastic functions

$$v(Z; a) = \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^k a_j Z_{ij} \right)^4. \quad (2)$$

Note that the Varimax object function can be then written as $v(\hat{Z}; R^\top) = \sum_{j=1}^k v(\hat{Z}; R_j)$ where R_j is the j th row of R . Our proof relies on the following lemmas.

Lemma 1. *Suppose $A \in \mathcal{O}(k)$ and $\min_{P \in \mathcal{P}(k)} \|A - P\|_F > t$ for some $t > 0$. Then,*

$$\sum_{i=1}^k \left(\sum_{j=1}^k A_{ij}^4 - 1 \right) \leq -\frac{1}{16} t^2.$$

Lemma 2. Suppose Z is a n -by- k random matrix with i.i.d. mean-zero, unit-variance and sub-Gaussian entries. Then, with probability at least $1 - n^{-ck} - n^{-4}$,

$$\sup_{a: \|a\|_2 \leq 1} v(Z; a) - \mathbb{E}v(Z; a) \lesssim \sqrt{\frac{k \log n}{n}} + \frac{k^2 \log^2 n}{n}.$$

Lemma 3. Suppose X_1, \dots, X_p are independent random variables with mean zero, variance one, and sub-Gaussian tails. Then there exists constants $C, c > 0$ such that

$$\mathbb{P} \left(\left| \sum_{i=1}^p (X_i^4 - \mathbb{E}X_i^4) \right| \geq C\sqrt{pt} + Ct^2 \right) \leq 2\exp(-ct).$$

We start by proving the sufficient condition. Define

$$\delta := \frac{C_0}{\sqrt{\kappa - 3}} \max \left\{ \left(\frac{k^2 \log^2 n}{n} \right)^{1/2}, \left(\frac{k \log n}{n} \right)^{1/4} \right\}, \quad (3)$$

for some positive constant C_0 to be specified later. It suffices to show that with high probability, for any $R \in \mathcal{O}(k)$ such that $\min_{P \in \mathcal{P}(k)} \|R^* R^\top - P\|_F > \delta k^{1/2}$, R cannot be the solution of Varimax.

By Lemma 2, we know with probability at least $1 - o(1/n)$,

$$\sup_{a: \|a\|_2 \leq 1} v(Z; a) - \mathbb{E}v(Z; a) \leq C \left(\sqrt{\frac{k \log n}{n}} + \frac{k^2 \log^2 n}{n} \right) \leq c(\kappa - 3), \quad (4)$$

where c is some small positive constant. In the meantime, Define $A = RR^{*\top} = (a_1, a_2, \dots, a_k)$, where a_j is the j -th orthogonal column of A . Note that the Varimax loss function can be written as $v(\hat{Z}; R) = \sum_{j=1}^k v(Z; a_j)$. By calculating its expectation, we have:

$$\begin{aligned} \mathbb{E}v(\hat{Z}; R) &= \mathbb{E} \sum_{j=1}^k v(Z; a_j) = 3k + (\kappa - 3) \sum_{i=1}^k \sum_{j=1}^k a_{ij}^4 \\ &= \kappa k + (\kappa - 3) \sum_{i=1}^k \left(\sum_{j=1}^k a_{ij}^4 - 1 \right) \stackrel{\text{Lemma 1}}{\leq} \kappa k - \frac{(\kappa - 3)\delta^2}{16} k. \end{aligned} \quad (5)$$

Meanwhile, note that

$$\mathbb{E}v(\hat{Z}, R^*) = \kappa k. \quad (6)$$

Then we have

$$\begin{aligned}
v(\hat{Z}; R^*) - v(\hat{Z}; R) &= \left(\mathbb{E}v(\hat{Z}; R^*) - \mathbb{E}v(\hat{Z}; R) \right) + \left(v(\hat{Z}; R^*) - \mathbb{E}(\hat{Z}; R^*) \right) - \left(v(\hat{Z}; R) - \mathbb{E}(\hat{Z}; R) \right) \\
&\stackrel{(5),(6)}{\geq} \frac{(\kappa - 3)\delta^2}{16} k - 2k \sup_{a: \|a\| \leq 1} |v(Z; a) - \mathbb{E}v(Z; a)| \\
&\stackrel{(3),(4)}{>} 0.
\end{aligned}$$

Therefore, R is not a solution of varimax. In other words, given $\hat{R} = \arg \max_{R \in \mathcal{O}(k)} v(\hat{Z}, R)$, one must have $\min_{P \in \mathcal{P}(k)} \|R^* R^\top - P\|_F < \delta k^{1/2}$, which completes the proof of sufficient condition.

Next, we prove the necessary condition. Let $Z_1 \in \mathbb{R}^{\lceil k/2 \rceil \times k}$ be the first $\lceil k/2 \rceil$ rows of the factor matrix Z . Consider the QR-decomposition of Z_1^\top : $Z_1^\top = U_1 R_1$, where U_1 is a k -by- $\lceil k/2 \rceil$ orthogonal matrix and R_1 is an $\lceil k/2 \rceil$ -by- $\lceil k/2 \rceil$ upper triangular matrix. Let $U_{1\perp} \in \mathbb{O}_{k, k-\lceil k/2 \rceil}$ be the perpendicular subspace of U_1 (such that $U_1^\top U_{1\perp} = 0$), and denote $A = [U_1 \ U_{1\perp}]^\top R^*$. Then,

$$(\hat{Z} A^\top)_{[1:\lceil k/2 \rceil, :]} = Z_1 [U_1 \ U_{1\perp}] = [R_1^\top \ O]$$

and it follows that

$$\begin{aligned}
v(\hat{Z}, A) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k \left((\hat{Z} A^\top)_{ij} \right)^4 \\
&\geq \frac{1}{n} \left(\underbrace{\sum_{j=1}^{\lceil k/2 \rceil} (R_1)_{jj}^4}_{D_1} + \underbrace{\sum_{i=k+1}^n \sum_{j=1}^k \left((\hat{Z} A^\top)_{ij} \right)^4}_{D_2} \right). \tag{7}
\end{aligned}$$

Now we establish the probabilistic bounds for D_1 and D_2 .

- D_1 . By random matrix theory [2], we know that with probability at least $1 - e^{-ck}$ that

$$\sigma_{\lceil k/2 \rceil}(Z_1) \geq \sqrt{k} - \sqrt{\lceil k/2 \rceil} - c\sqrt{k} \geq \sqrt{0.1k}. \tag{8}$$

Note that R_1 shares the same non-zeros singular values with X_1 . Thus it follows that

$$\prod_{j=1}^{\lceil k/2 \rceil} (R_1)_{jj}^2 = |\det(R_1)|^2 = \prod_{j=1}^{\lceil k/2 \rceil} \sigma_j^2(Z_1) \stackrel{(8)}{\geq} (0.1k)^{\lceil k/2 \rceil}.$$

Then by Inequality of arithmetic and geometric means,

$$D_1 = \sum_{j=1}^{\lceil k/2 \rceil} (R_1)_{jj}^4 \geq \lceil k/2 \rceil \left(\prod_{j=1}^{\lceil k/2 \rceil} (R_1)_{jj}^4 \right)^{1/\lceil k/2 \rceil} \geq 0.005k^3. \quad (9)$$

- D_2 . Denote $B := R^* A^\top$ and we can rewrite

$$\begin{aligned} D_2 &= \sum_{i=k+1}^n \sum_{j=1}^k \left((\hat{Z} A^\top)_{ij} \right)^4 = \sum_{i=k+1}^n \sum_{j=1}^k ((ZB)_{ij})^4 \\ &= \sum_{j=1}^k \sum_{i=k+1}^n \left(\sum_{l=1}^k B_{lj} Z_{il} \right)^4. \end{aligned}$$

Then, conditioning on A (or B), we have

$$\begin{aligned} \mathbb{E}[D_2|A] &= \mathbb{E} \left[\sum_{j=1}^k \sum_{i=k+1}^n \left(\sum_{l=1}^k B_{lj} Z_{il} \right)^4 \middle| B \right] \\ &= \sum_{j=1}^k (n-k) \mathbb{E} \left[\left(\sum_{l=1}^k B_{lj} Z_{1l} \right)^4 \middle| B \right] \\ &= (n-k) \left(3k + (\kappa - 3) \sum_{l=1}^k \sum_{j=1}^k B_{lj}^4 \right) \\ &\geq 3(n-k)k. \end{aligned}$$

On the other hand, for each fixed $j \in [k]$, $\left\{ \sum_{l=1}^k B_{lj} Z_{il} \right\}_{i=k+1}^n$ are i.i.d. mean-0, variance-1 sub-Gaussian random variables (since B is independent of the last $n-k$ rows of Z). By Lemma 3, we have

$$\mathbb{P} \left(\sum_{i=k+1}^n \left(\sum_{l=1}^k B_{lj} Z_{il} \right)^4 - \mathbb{E} \left[\sum_{i=k+1}^n \left(\sum_{l=1}^k B_{lj} Z_{il} \right)^4 \middle| A \right] < -C(\sqrt{nt} + t^2) \middle| A \right) \leq 2e^{-ct}.$$

Taking $t = c_0 k$ for some small constant c_0 and applying union bound for each $j \in [k]$, we know that

$$\mathbb{P} \left(D_2 - \mathbb{E}[D_2|A] < -c_0 \left(\sqrt{nk^3} + k^3 \right) \middle| A \right) \leq 2ke^{-ck}$$

and it follows that

$$\mathbb{P}\left(D_2 < 3(n-k)k - c_0\left(\sqrt{nk^3} + k^3\right) \middle| \mathbf{A}\right) \leq 2ke^{-ck}.$$

Integrate over the density of A and we obtain

$$\mathbb{P}\left(D_2 < 3(n-k)k - c_0\left(\sqrt{nk^3} + k^3\right)\right) \leq 2ke^{-ck} \leq e^{-c'k}. \quad (10)$$

In conclusion, we proved that with probability at least $1 - e^{-ck}$,

$$D_1 \geq 0.005k^3, \quad D_2 \geq 3(n-k)k - c_0\left(\sqrt{nk^3} + k^3\right). \quad (11)$$

Since $n \leq ck^2$, by taking c_0 to be sufficiently small, we have

$$v(\hat{Z}, A) \geq \frac{1}{n} (0.004k^3 + 3nk). \quad (12)$$

Next we show that for any R such that $\text{dist}(R, R^*) \leq \delta$, $v(\hat{Z}, R)$ is upper bounded by the right-hand side of (12) with high probability. Note that

$$\begin{aligned} v(\hat{Z}; R) &= \mathbb{E}v(\hat{Z}; R^*) + (v(\hat{Z}, R^*) - \mathbb{E}v(\hat{Z}; R^*)) + (v(\hat{Z}, R) - v(\hat{Z}, R^*)) \\ &= \kappa k + \left(v(\hat{Z}; R^*) - \mathbb{E}v(\hat{Z}; R^*)\right) + \sup_{R: \text{dist}(R, R^*) \leq \delta} \left|v(\hat{Z}, R) - v(\hat{Z}, R^*)\right|. \end{aligned} \quad (13)$$

Following the same argument as (10), we know that with probability at least $1 - e^{-ck}$ that

$$v(\hat{Z}; R^*) - \mathbb{E}v(\hat{Z}; R^*) \leq \frac{c_0}{n} \left(\sqrt{nk^3} + k^3\right).$$

In the meantime,

$$\begin{aligned} \sup_{R: \text{dist}(R, R^*) \leq \delta} \left|v(\hat{Z}, R) - v(\hat{Z}, R^*)\right| &= \sup_{\substack{\tilde{R} \in \mathcal{O}(k): \\ \|\tilde{R} - I\|_F \leq \delta\sqrt{k}}} \left|v(Z; \tilde{R}) - v(Z; I)\right| \\ &\leq \sup_{\substack{\tilde{R} \in \mathcal{O}(k): \\ \|\tilde{R} - I\|_F \leq \delta\sqrt{k}}} \sum_{i=1}^k \left|v(Z; \tilde{R}_{[i,:]} - v(Z; e_i)\right|, \end{aligned} \quad (14)$$

where e_i is the i -th canonical basis in \mathbb{R}^k . Let $\delta_i := \|\tilde{R}_{[i,:]} - e_i\|_2$. Then, by the proof of (20) in Lemma 2, for each $i \in [k]$, we have

$$\left|v(Z; \tilde{R}_{[i,:]} - v(Z; e_i)\right| \leq 6\delta_i \max_{j,l} Z_{jl}^4 \leq C\delta_i \log^2 n \quad (15)$$

hold with probability at least $1 - n^{-4}$. Combining (14) and (15), we obtain

$$\begin{aligned} \sup_{R: \text{dist}(R, R^*) \leq \delta} \left| v(\hat{Z}, R) - v(\hat{Z}, R^*) \right| &\leq C \log^2 n \sum_{i=1}^k \delta_i \leq C \log^2 n \left(k \sum_{i=1}^k \delta_i^2 \right)^{1/2} \\ &= C \sqrt{k} \log^2 n \left\| \tilde{R} - I \right\|_{\text{F}} \leq C \delta k \log^2 n. \end{aligned} \quad (16)$$

Thus, when $n \lesssim \frac{k^2}{\log^2 n}$, we have

$$\sup_{R: \text{dist}(R, R^*) \leq \delta} v(\hat{Z}, R) \stackrel{(13), (16)}{\leq} \kappa k + \frac{c_0 \left(\sqrt{nk^3} + k^3 \right)}{n} + \delta k \log^2 n < \frac{0.004k^3 + 3nk}{n} \stackrel{(12)}{\leq} v(\hat{Z}, A).$$

Here, the second inequality is obtained by choosing c_0 to be sufficiently small and $\delta = 1/2$. This suggests that the solution of Varimax function \hat{R} must satisfy:

$$\text{dist}(\hat{R}, R^*) \geq \delta = 1/2.$$

□

Proof of Lemma 1. Let $P^* = \arg \min_{P \in \mathcal{O}(k)} \|A - P\|_{\text{F}}^2$. Without loss of generality, we can assume $P^* = I$ and $A_{11} \geq A_{22} \geq \dots \geq A_{kk} \geq 0$. This suggests that

$$t^2 < \sum_{i=1}^k \left((1 - A_{ii})^2 + \sum_{j \neq i} A_{ij}^2 \right) = 2 \sum_{i=1}^k (1 - A_{ii}) \leq 2 \sum_{i=1}^k (1 - A_{ii}^2). \quad (17)$$

It suffices to show that for each $i \in [k]$,

$$\sum_{j=1}^k A_{ij}^4 - 1 \leq -\frac{1}{8}(1 - A_{ii}^2).$$

Now consider the following two situations for each $i \in [k]$:

- $A_{ii}^2 \geq 1/16$.

$$\begin{aligned} \sum_{j=1}^k A_{ij}^4 - 1 &\leq A_{ii}^4 + \left(\sum_{j \neq i} A_{ij}^2 \right)^2 - 1 \\ &= A_{ii}^4 + (1 - A_{ii}^2)^2 - 1 \\ &= 2A_{ii}^2 (A_{ii}^2 - 1) \leq -\frac{1}{8}(1 - A_{ii}^2). \end{aligned}$$

- $A_{ii}^2 < 1/16$. We claim that $\max_j A_{ij}^2 < 3/4$. Suppose there exists $l \in [k]/i$ such that $A_{il}^2 \geq 3/4$. Since $I = \arg \min_{P \in \mathcal{P}(k)} \|A - P\|_F^2$, we must have

$$\|A - I\|_F^2 \leq \|A - D^{(i,l)}\|_F^2, \quad (18)$$

where $D^{(i,l)}$ is the permutation matrix such that $D_{il}^{(i,l)} = \text{sgn}(A_{il})$, $D_{li}^{(i,l)} = \text{sgn}(A_{li})$ and $D_{jj}^{(i,l)} = 1$ for $j \in [k]/\{i, l\}$. This yields

$$(A_{ll} - 1)^2 + (A_{ii} - 1)^2 + A_{il}^2 + A_{li}^2 \leq (|A_{li}| - 1)^2 + (|A_{il}| - 1)^2 + A_{ii}^2 + A_{ll}^2,$$

which is equivalent to

$$|A_{li}| + |A_{il}| \leq A_{ll} + A_{ii}.$$

Then it follows that

$$A_{ll} \geq |A_{il}| - A_{ii} \geq \frac{2\sqrt{3} - 1}{4}.$$

Consequently, we get

$$1 \geq A_{ll}^2 + A_{ii}^2 \geq \frac{13 - 4\sqrt{3}}{16} + \frac{3}{4} > 1.$$

This contradiction shows that we must have $\max_j A_{ij}^2 < 3/4$. Therefore,

$$\sum_{j=1}^k A_{ij}^4 - 1 \leq \max_j A_{ij}^2 \sum_{j=1}^k A_{ij}^2 - 1 = -\frac{1}{4} \leq -\frac{1}{4}(1 - A_{ii}^2).$$

The proof is finished by combining the two situations. \square

Proof of Lemma 2. We first construct an ε -net $\mathcal{N} = \{x^{(1)}, \dots, x^{(N)}\} \subset \mathbb{S}^{k-1}$ such that for any $a \in \mathbb{S}^{k-1}$, there exists some $i_a \in [N]$ with $\|a - x^{(i_a)}\| \leq \varepsilon$. By [2], we can choose such an ε -net with $N \leq (3/\varepsilon)^k$. Then, by triangle inequality,

$$\begin{aligned} & \sup_{a \in \mathbb{S}^{k-1}} |v(Z; a) - \mathbb{E}v(Z; a)| \\ & \leq \sup_{a \in \mathbb{S}^{k-1}} \left(\left| \mathbb{E}v(Z; x^{(i_a)}) - \mathbb{E}v(Z; a) \right| + \left| v(Z; a) - v(Z; x^{(i_a)}) \right| + \left| v(Z; x^{(i_a)}) - \mathbb{E}v(Z; x^{(i_a)}) \right| \right) \\ & \leq \underbrace{\sup_{a \in \mathbb{S}^{k-1}} \left| \mathbb{E}v(Z; x^{(i_a)}) - \mathbb{E}v(Z; a) \right|}_{D_1} + \underbrace{\sup_{a \in \mathbb{S}^{k-1}} \left| v(Z; a) - v(Z; x^{(i_a)}) \right|}_{D_2} + \underbrace{\max_{i \in N} \left| v(Z; x^{(i)}) - \mathbb{E}v(Z; x^{(i)}) \right|}_{D_3}. \end{aligned}$$

We specify $\varepsilon = 1/n$ and $N \leq (3n)^k$. Next we obtain deterministic bounds for D_1 and D_2 and a probabilistic bound for D_3 .

- $\sup_{a \in \mathbb{S}^{k-1}} |\mathbb{E}v(Z; x^{(i_a)}) - \mathbb{E}v(Z; a)|$. Note that for any fixed $a \in \mathbb{S}^{k-1}$, one can calculate that

$$\mathbb{E}v(Z; a) = \mathbb{E} \left(\sum_{j=1}^k a_j X_{1j} \right)^4 = (\kappa - 3) \sum_{j=1}^k a_j^4 + 3.$$

Therefore,

$$\begin{aligned} \left| \mathbb{E}v(Z; x^{(i_a)}) - \mathbb{E}v(Z; a) \right| &= (\kappa - 3) \left| \sum_{j=1}^k \left(a_j^4 - (x_j^{(i_a)})^4 \right) \right| \\ &\leq (\kappa - 3) \left(\max_j a_j^2 + (x_j^{(i_a)})^2 \right) \cdot \sum_{j=1}^k \left| a_j^2 - (x_j^{(i_a)})^2 \right| \quad (19) \\ &\leq 2(\kappa - 3) \sum_{j=1}^k \left| a_j^2 - (x_j^{(i_a)})^2 \right|. \end{aligned}$$

Denote $\delta = a - b$ for any $b \in \mathbb{S}^{k-1}$. Note that

$$\sum_{j=1}^k |a_j - b_j| = \sum_{j=1}^k |b_j^2 + \delta_j^2 + 2b_j\delta_j - b_j^2| \leq \|\delta\|_2^2 + 2 \sum_{j=1}^k |b_j\delta_j| \leq \|\delta\|_2^2 + 2\|\delta\|_2,$$

where we use Cauchy-Schwarz inequality and the fact $\|b\| = 1$ to obtain the last inequality. Now we replace b with $x^{(i_a)}$, then D_1 can be further bounded as

$$\begin{aligned} D_1 &\leq 2(\kappa - 3) \sum_{j=1}^k \left| a_j^2 - (x_j^{(i_a)})^2 \right| \leq 2(\kappa - 3) \max_j |a_j + x_j^{(i_a)}| \cdot \sum_{j=1}^j |a_j - x_j^{(i_a)}| \quad (20) \\ &\leq 12\varepsilon(\kappa - 3) \leq \frac{C}{n}. \end{aligned}$$

- $\sup_{a \in \mathbb{S}^{k-1}} |v(Z; a) - v(Z; x^{(i_a)})|$. For any $a, b \in \mathbb{S}^{k-1}$,

$$\begin{aligned}
|v(Z; a) - v(Z; b)| &= \frac{1}{n} \left| \sum_{i=1}^n \left(\left(\sum_{j=1}^k a_j Z_{ij} \right)^4 - \left(\sum_{j=1}^k b_j Z_{ij} \right)^4 \right) \right| \\
&\leq \frac{1}{n} \max_i \left| \left(\sum_{j=1}^k a_j Z_{ij} \right)^2 + \left(\sum_{j=1}^k b_j Z_{ij} \right)^2 \right| \cdot \sum_{i=1}^n \left| \left(\sum_{j=1}^k a_j Z_{ij} \right)^2 - \left(\sum_{j=1}^k b_j Z_{ij} \right)^2 \right| \\
&\leq \frac{2}{n} \left(\max_i \sum_{j=1}^k Z_{ij}^2 \right) \cdot \max_i \left| \sum_{j=1}^k (a_j + b_j) Z_{ij} \right| \cdot \sum_{i=1}^n \sum_{j=1}^k |(a_j - b_j) Z_{ij}| \\
&\leq 4k^{3/2} (\max_{i,j} |Z_{ij}|)^4 \cdot \sum_{j=1}^k |a_j - b_j| \leq 4k^2 (\max_{i,k} |Z_{ik}|)^4 \cdot \sqrt{\sum_{j=1}^k (a_j - b_j)^2} \\
&\leq 4k^2 (\max_{i,j} |X_{ij}|)^4 \varepsilon \leq \frac{4k^2 (\max_{i,j} |X_{ij}|)^4}{n}.
\end{aligned}$$

- $\max_{i \in N} |v(Z; x^{(i)}) - \mathbb{E}v(Z; x^{(i)})|$. By Lemma 3, for fixed $x^{(l)}$, since $Y_i^{(l)} := \sum_{j=1}^k x_j^{(l)} X_{ij}$ are i.i.d. mean-zero, variance one, sub-Gaussian random variables, we have

$$\mathbb{P} \left(v(Z; x^{(j)}) - \mathbb{E}v(Z; x^{(j)}) \geq C \left(\sqrt{\frac{t}{n}} + \frac{t^2}{n} \right) \right) \leq e^{-ct}.$$

Taking $t = Ck \log n$ and applying union bounds on $x^{(l)}$ for each $l \in [N]$, it follows that

$$\mathbb{P} \left(D_3 \leq C \left(\sqrt{\frac{k \log n}{n}} + \frac{k^2 \log^2 n}{n} \right) \right) \geq 1 - e^{-ck \log n}. \quad (21)$$

In the meantime, by the tail bound of sub-Gaussian Extreme values, we have

$$\mathbb{P} \left(\max_{i,k} |X_{i,k}| \leq \sqrt{10 \log(nk)} \right) \leq 1 - \sqrt{\frac{2}{\pi \log n}} n^{-4}. \quad (22)$$

Combining (20), (19), (21) and (22), we know that with probability at least $1 - n^{-ck} - n^{-4}$ that

$$\sup_{a \in \mathbb{S}^{k-1}} |v(Z; a) - \mathbb{E}v(Z; a)| \lesssim \sqrt{\frac{k \log n}{n}} + \frac{k^2 \log^2 n}{n}.$$

Now the proof is finished. \square

Proof of Lemma 3. Note that $\{X_i^4\}_{i=1}^n$ are sub-Weibull(1/2) random variables and the result simply comes from the concentration developed by [1, Theorem 3.1]. \square

References

- [1] Arun Kumar Kuchibhotla and Abhishek Chakraborty. Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *arXiv preprint arXiv:1804.02605*, 2018. [10](#)
- [2] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010. [4](#), [8](#)