

## A Derivation of Moments for the IBP (Section 3)

### A.1 Moments for the Linear Gaussian Latent Feature Model

Key to the derivation is the fact that  $z$  and  $\epsilon$  are independent random variables, hence their expectations can be taken independently.

**Order 1 tensor:** By using Equation (4), we have

$$S_1 := M_1 = \mathbf{E}_x[x] = \mathbf{E}_z[Az + \epsilon] = A\mathbf{E}_z[z] = T(\pi, A). \quad (25)$$

To infer the *number* of latent variables  $k$  and deal with the noise term, we need to determine the rank of the covariance matrix  $\mathbf{E}_x[(x - \mathbf{E}_x[x]) \otimes (x - \mathbf{E}_x[x])]$ . Because there is additive noise, the smallest  $(d - K)$  eigenvalues will not be exactly zero. Instead, they amount to the variance arising from  $\epsilon$  since

$$\text{cov}[Az + \epsilon] = A^\top \text{cov}[z]A + \text{cov}[\epsilon]. \quad (26)$$

Consequently the smallest eigenvalues of the covariance matrix of  $x$  allow us to read off the variance  $\sigma^2$ : for any normal vector  $v$  corresponding to the  $d - k$  smallest eigenvalues we have

$$\mathbf{E}_x \left[ \left( v^\top (x - \mathbf{E}[X]) \right)^2 \right] = v^\top A^\top \text{cov}[z]A v + v^\top \text{cov}[\epsilon]v = \sigma^2. \quad (27)$$

**Order 2 tensor:** For the second-order tensor, we plug in (6) and use independence of  $z$  and  $\epsilon$ . Linear terms in  $\epsilon$  vanish. Hence we get

$$\begin{aligned} M_2 &= \mathbf{E}_x[x \otimes x] = T(\mathbf{E}_z[z \otimes z], A, A) + \sigma^2 \mathbf{1} = T(\pi \otimes \pi + \text{diag}(\pi - \pi^2), A, A) + \sigma^2 \mathbf{1} \\ &= S_1 \otimes S_1 + T(\text{diag}(\pi - \pi^2), A, A) + \sigma^2 \mathbf{1}. \end{aligned} \quad (28)$$

This yields the statement in Equation (13).

**Order 3 tensor:** As before, denote by  $v$  an eigenvector corresponding to the  $(d - k)$  smallest eigenvalues, i.e.  $v^\top A = 0$ . We first define an auxiliary term

$$\begin{aligned} m_1 &:= \mathbf{E}_x \left[ x \left( v^\top (x - \mathbf{E}[x]) \right)^2 \right] = \mathbf{E}_x \left[ x \left( v^\top (A(z - \pi) + \epsilon) \right)^2 \right] \\ &= \mathbf{E}_x \left[ x \left( v^\top \epsilon \right)^2 \right] = \sigma^2 T(\pi, A). \end{aligned} \quad (29)$$

Since the Normal Distribution is symmetric, only even moments of  $\epsilon$  survive. Using (9), the third order moments yield

$$M_3 = \mathbf{E}_x[x \otimes x \otimes x] = \mathbf{E}_z[Az \otimes Az \otimes Az] + \mathbf{E}_z[\mathfrak{S}_3[Az \otimes \epsilon \otimes \epsilon]] \quad (30)$$

$$= T(\mathbf{E}_z[z \otimes z \otimes z], A, A, A) + \mathfrak{S}_3(m_1 \otimes \mathbf{1}) \quad (31)$$

$$= S_1 \otimes S_1 \otimes S_1 + \mathfrak{S}_3[S_1 \otimes S_2] + T(\text{diag}(\pi - 3\pi_i^2 + 2\pi_i^3), A, A, A) + \mathfrak{S}_3(m_1 \otimes \mathbf{1})$$

Thus, we get Equation (14).

**Order 4 tensor:** We obtain the fourth-order tensor by first calculating an auxiliary variable related to the additive noise term

$$m_4 := \mathbf{E}_x \left[ \left( v^\top (x - \mathbf{E}_x[x]) \right)^4 \right] / 3 = \mathbf{E}[(v^\top \epsilon)^4] / 3 = \sigma^4. \quad (32)$$

Here the last equality followed from the isotropy of Gaussians. With Equation (10), the forth order moments are

$$\begin{aligned} M_4 &= \mathbf{E}_x[x \otimes x \otimes x \otimes x] \\ &= \mathbf{E}_z[Az \otimes Az \otimes Az \otimes Az] + \mathbf{E}_z[\mathfrak{S}_6[Az \otimes Az \otimes \epsilon \otimes \epsilon]] + \mathbf{E}[\epsilon \otimes \epsilon \otimes \epsilon \otimes \epsilon] \end{aligned} \quad (33)$$

$$= T(\mathbf{E}_z[z \otimes z \otimes z \otimes z], A, A, A, A) + \sigma^2 \mathfrak{S}_6[S_2 \otimes \mathbf{1}] + \sigma^4 \mathfrak{S}_3[\mathbf{1} \otimes \mathbf{1}] \quad (34)$$

$$\begin{aligned} &= S_1 \otimes S_1 \otimes S_1 \otimes S_1 + \mathfrak{S}_6[S_2 \otimes S_1 \otimes S_1] + \mathfrak{S}_3[S_2 \otimes S_2] + \mathfrak{S}_4[S_3 \otimes S_1] \\ &\quad + T(\text{diag}(-6\pi^4 + 12\pi^3 - 7\pi^2 + \pi), A, A, A) + \sigma^2 \mathfrak{S}_6[S_2 \otimes \mathbf{1}] + m_4 \mathfrak{S}_3[\mathbf{1} \otimes \mathbf{1}]. \end{aligned} \quad (35)$$

### A.2 Moments for the Infinite Sparse Factor Analysis Model

Since both  $Y$  and  $\epsilon$  are symmetric and have zero mean, the odd order tensors vanish. That is  $M_1 = 0$  and  $M_3 = 0$ . It suffices for us to focus on the even terms.

**Order 2 tensor:** Using covariance matrix of (6) yields

$$M_2 = \mathbf{E}_x [x \otimes x] = T(\mathbf{E}_z[(z \odot y) \otimes (z \odot y)], G, G) + \sigma^2 \mathbf{1} \quad (36)$$

$$= T((\mathbf{E}_z[z \otimes z] \odot \mathbf{E}_y[y^2] \mathbf{1}), G, G) + \sigma^2 \mathbf{1} \quad (37)$$

$$= T((\pi \otimes \pi + \text{diag}(\pi - \pi^2)) \odot \mathbf{E}_y[y^2] \mathbf{1}, G, G) + \sigma^2 \mathbf{1} \quad (38)$$

$$= T(\mathbf{E}_y[y^2] \text{diag}(\pi), G, G) + \sigma^2 \mathbf{1} = T(\text{diag}(\pi), G, G) + \sigma^2 \mathbf{1}, \quad (39)$$

As before, the variance  $\sigma^2$  of  $\epsilon$  can be inferred by Equation (27). Here we get Equation (18).

**Order 4 tensor:** With Equation (10) and  $\mathbf{E}_y[y^4] = 3$ , we have

$$M_4 = \mathbf{E}_x [x \otimes x \otimes x \otimes x] \quad (40)$$

$$\begin{aligned} &= \mathbf{E}_z [G(z \odot y) \otimes G(z \odot y) \otimes G(z \odot y) \otimes G(z \odot y)] \\ &\quad + \mathbf{E}_z [\mathfrak{S}_6[G(z \odot y) \otimes G(z \odot y) \otimes \epsilon \otimes \epsilon]] + \mathbf{E}[\epsilon \otimes \epsilon \otimes \epsilon \otimes \epsilon] \\ &= T(\mathbf{E}_z[z \otimes z \otimes z \otimes z] \odot \mathbf{E}_y[y^4] \mathbf{1}, A, A, A, A) + \sigma^2 \mathfrak{S}_6[S_2 \otimes \mathbf{1}] + \sigma^4 \mathfrak{S}_3[\mathbf{1} \otimes \mathbf{1}] \end{aligned} \quad (41)$$

$$\begin{aligned} &= \mathfrak{S}_3[S_2 \otimes S_2] + T(\text{diag}(\mathbf{E}_y[y^4] \pi_i - 3\mathbf{E}_y[y^2]^2 \pi_i^2), A, A, A, A) \\ &\quad + \sigma^2 \mathfrak{S}_6[S_2 \otimes \mathbf{1}] + \sigma^4 \mathfrak{S}_3[\mathbf{1} \otimes \mathbf{1}] \end{aligned} \quad (42)$$

$$= \mathfrak{S}_3[S_2 \otimes S_2] + T(3(\pi_i - \pi_i^2), A, A, A, A) + \sigma^2 \mathfrak{S}_6[S_2 \otimes \mathbf{1}] + m_4 \mathfrak{S}_3[\mathbf{1} \otimes \mathbf{1}] \quad (43)$$

where  $m_4$  can be inferred by (17).

If the prior on  $Y$  is drawn from a Laplace distribution the model is called a iICA. The lower-order moments are similar to that of isFA, except for  $\mathbf{E}_y[y^2] = 2$  and  $\mathbf{E}_y[y^4] = 24$ . Replacing these terms in Equation (39) and (42) yields the claim.

## B Concentration measure of bounded moments

### B.1 Proof of the Moment Bound of Theorem 2

Denote by  $X$  the  $m$ -sample used in generating  $\hat{M}_l$ . Moreover, denote by

$$\Xi[X] := \sup_{u: \|u\| \leq 1} |T[M_l, u, \dots, u] - T[\hat{M}_l, u, \dots, u]| \quad (44)$$

the largest deviation between empirical and expected moments, when applied to the test vectors  $u$ . Bounding this quantity directly is desirable since it allows us to avoid having to derive *pointwise* bounds with regard to  $M_l$ . We prove that  $\Xi[X]$  is concentrated using McDiarmid's bound [25]. Substituting single observations in  $\Xi[X]$  yields

$$|\Xi[X] - \Xi[(X \setminus \{x_j\}) \cup \{x'\}]| \leq \frac{1}{m} [T[\phi_l(x_j) - \phi_l(x'), u, \dots, u]] \quad (45)$$

$$\leq \frac{1}{m} [\|\phi(x_j)\|^l + \|\phi(x')\|^l] \leq \frac{2}{m} R^l. \quad (46)$$

Plugging the bound of  $2R^l/m$  into McDiarmid's theorem shows that the random variable  $\Xi[X]$  is concentrated for  $\Pr\{\Xi[X] - \mathbf{E}_X[\Xi[X]] > \epsilon\} \leq \delta$  with probability  $\delta \leq \exp\left(-\frac{m\epsilon^2}{2R^{2l}}\right)$ . Solving the bound for  $\epsilon$  shows that with probability at least  $1 - \delta$  we have that  $\epsilon \leq \sqrt{-2 \log \delta / m R^l}$ .

The next step is to bound the expectation of  $\Xi[X]$ . For this we exploit the ghost sample trick and the convexity of expectations. This leads to the following:

$$\begin{aligned} \mathbf{E}_X [\Xi[X]] &\leq \mathbf{E}_{X, X'} \left[ \sup_{u: \|u\| \leq 1} |T[M_l, u, \dots, u] - T[\hat{M}_l, u, \dots, u]| \right] \\ &= \mathbf{E}_\sigma \mathbf{E}_{X, X'} \left[ \sup_{u: \|u\| \leq 1} \left| \frac{1}{m} \sum_{j=1}^m \sigma_j (T[\phi_l(x_j), u, \dots, u] - T[\phi_l(x'_j), u, \dots, u]) \right| \right] \\ &\leq \frac{2}{m} \mathbf{E}_\sigma \mathbf{E}_X \left[ \sup_{u: \|u\| \leq 1} \left| \sum_{j=1}^m \sigma_j T[\phi_l(x_j), u, \dots, u] \right| \right] \end{aligned} \quad (47)$$

$$\leq \frac{2}{m} \mathbf{E}_\sigma \mathbf{E}_X \left[ \left\| \sum_{j=1}^m \sigma_j \phi_l(x_j) \right\| \right] \leq \frac{2}{m} \mathbf{E}_X \left[ \mathbf{E}_\sigma \left[ \left\| \sum_{j=1}^m \sigma_j \phi_l(x_j) \right\|^2 \right] \right]^{\frac{1}{2}} \leq \frac{2R^l}{\sqrt{m}} \quad (48)$$

Here the first inequality follows from convexity of the argument. The subsequent equality is a consequence of the fact that  $X$  and  $X'$  are drawn from the same distribution, hence a swapping permutation with the ghost-sample leaves terms unchanged; The following inequality is an application of the triangle inequality. Next we use the Cauchy-Schwartz inequality, convexity and last the fact that  $\|\phi(x)\| \leq R$ . Combining both bounds yields  $\epsilon \leq [2 + \sqrt{-2 \log \delta}] R^l / \sqrt{m}$ .

## B.2 Tools for bounding tensors with bounded moments

To prove the guarantees for tensors, we rely on the triangle inequality on tensorial reductions

$$\sup_u |T(A + B, u) - T(A' + B', u)| \leq \sup_u |T(A, u) - T(A', u)| + \sup_u |T(B, u) - T(B', u)| \quad (49)$$

and moreover, the fact that for products of bounded random variables the guarantees are additive, as stated in the lemma below:

**Lemma 4** *Denote by  $f_i$  random variables and by  $\hat{f}_i$  their estimates. Moreover, assume that each of them is bounded via  $|f_i| \leq R_i$  and  $|\hat{f}_i| \leq R_i$  and*

$$\Pr \left\{ |\mathbf{E}[f_i] - \hat{f}_i| > \epsilon_i \right\} \leq \delta_i. \quad (50)$$

*In this case the product is bounded via*

$$\Pr \left\{ \left| \prod_i \mathbf{E}[f_i] - \prod_i \hat{f}_i \right| > \epsilon \right\} \leq \sum_i \delta_i \quad \text{where } \epsilon = \left[ \prod_i R_i \right] \left[ \sum_i \frac{\epsilon_i}{R_i} \right] \quad (51)$$

**Proof** We prove the claim for two variables, say  $f_1$  and  $f_2$ . We have

$$\left| \mathbf{E}[f_1] \mathbf{E}[f_2] - \hat{f}_1 \hat{f}_2 \right| \leq \left| (\mathbf{E}[f_1] - \hat{f}_1) \mathbf{E}[f_2] \right| + \left| \hat{f}_1 (\mathbf{E}[f_2] - \hat{f}_2) \right| \leq \epsilon_1 R_2 + R_1 \epsilon_2$$

with probability at least  $1 - \delta_1 - \delta_2$ , when applying the union bound over  $\mathbf{E}[f_1] - \hat{f}_1$  and  $\mathbf{E}[f_2] - \hat{f}_2$  respectively. Rewriting terms yields the claim for  $n = 2$ . To see the claim for  $n > 2$  simply use the fact that we can decompose the bound into a chain of inequalities involving exactly one difference, say  $\mathbf{E}[f_i] - \hat{f}_i$  and  $n - 1$  instances of  $\mathbf{E}[f_j]$  or  $\hat{f}_j$  respectively. We omit details since they are straightforward to prove (and tedious). ■

## C Proof of Equation (21)

For simplicity in the proof, in Equation (13) (14) (15), we define the diagonal coefficients for  $S_i$  to be  $C_i \in \mathbb{R}^K$ , i.e.,  $C_2 = \pi - \pi^2$ ,  $C_3 = \pi - 3\pi^2 + 2\pi^3$  and  $C_4 = \pi - 7\pi^2 + 12\pi^3 - 6\pi^4$ , so that

$$S_2 = T(\text{diag}(C_2), A, A), \quad S_3 = T(\text{diag}(C_3), A, A, A), \quad S_4 = T(\text{diag}(C_4), A, A, A, A).$$

Following step 6 in Algorithm 1, we obtain whitening matrix  $W$  by doing svd on  $S_2$ . Suppose the svd of matrix  $T(\text{diag}(\sqrt{C_2}), A) = U \Sigma^{1/2} V^\top$ , we have  $S_2 = U \Sigma^{1/2} V^\top V \Sigma^{1/2} U^\top = U S U^\top$  and  $W = U \Sigma^{-1/2}$ . Using the fact that  $S_3 = T(\text{diag}(C_3 C_2^{-3/2}), \text{diag}(\sqrt{C_2}) A, \text{diag}(\sqrt{C_2}) A, \text{diag}(\sqrt{C_2}) A)$ , we have

$$\begin{aligned} W_3 &= T(S_3, W, W, W) \\ &= T\left(\text{diag}\left(C_3 C_2^{-3/2}\right), \Sigma^{-1/2} U^\top (U \Sigma^{1/2} V^\top), \Sigma^{-1/2} U^\top (U \Sigma^{1/2} V^\top), \Sigma^{-1/2} U^\top (U \Sigma^{1/2} V^\top)\right) \\ &= T\left(\text{diag}\left(C_3 C_2^{-3/2}\right), V^\top, V^\top, V^\top\right). \end{aligned} \quad (52)$$

The diagonalized tensor  $W_3$ , with some permutation  $\tau$  on  $[K]$  and  $s_i \in \{\pm 1\}$ , has eigenvalues and eigenvectors:

$$\lambda_i = s_i C_{3,i} C_{2,i}^{-3/2}, \quad v_i = s_i (V^\top), e_{\tau(i)}, \quad (53)$$

where  $C_{i,j}$  representing the  $j$ -th element in  $C_i$ . After obtaining  $v_i$ , we multiply  $v_i$  by  $(W^\dagger)^\top$  to rotate it back to  $A_i$  as describing in step 15 in Algorithm 1, where  $W^\dagger = (W^\top W)^{-1} W^\top = \Sigma^{1/2} U^\top$ , we get

$$(W^\dagger)^\top v_i = s_i U \Sigma^{1/2} V^\top e_{\tau(i)} = s_i T(\text{diag}(\sqrt{C_2}), A) e_{\tau(i)} = s_i \sqrt{C_{2,i}} A_{\tau(i)}, \quad (54)$$

which yields  $A_{\tau(i)} = \frac{(W^\dagger)^\top v_i}{s_i \sqrt{C_{2,i}}}$ . With the fact that  $s_i = C_{3,i} C_{2,i}^{-3/2} \lambda_i^{-1}$  from Equation (53), we have

$$A_{\tau(i)} = \frac{\lambda_i}{(C_{3,i} C_{2,i}^{-1})} (W^\dagger)^\top v_i = C_{2,i}^{-1/2} (W^\dagger)^\top v_i. \quad (55)$$

Plug in the definition of  $C_2$ , we get the scale factor for  $i \in [K_1]$ . For  $A_i$  which are recovered by conducting tensor decomposition on  $W_4$ , we first examine

$$W_4 = T(S_4, W, W, W, W) = T(\text{diag}(C_{4,i} C_{2,i}^{-2}), V^\top, V^\top, V^\top, V^\top), \quad (56)$$

and obtain

$$\lambda_i = C_{4,i} C_{2,i}^{-2}, \quad v_i = s_i (V^\top) e_{\tau(i)}. \quad (57)$$

By using the fact that  $s_i = s_i C_{4,i} C_{2,i}^{-2} \lambda_i^{-1}$  and Equation (54), we have

$$A_{\tau(i)} = \frac{(W^\dagger)^\top v_i}{s_i \sqrt{C_{2,i}}} = \frac{s_i}{(C_{4,i} C_{2,i}^{-3/2}) \lambda_i^{-1}} = s_i (W^\dagger)^\top v_i = s_i C_{2,i}^{-1/2} (W^\dagger)^\top v_i, \quad \forall i \in [K_1 + 1, \dots, K]. \quad (58)$$

Note that the value of  $\pi_i$  used to construct  $C_j$  can be recovered by Equation (53) and (57) after obtaining  $\lambda_i$ .

## D Reconstruction accuracy for Algorithm 1

In this section, we provides bounds for moments of linear gaussian latent feature model. The concentration behavior is more complicated than that of the bounded moments in Theorem 2 due to the additive Gaussian noise. Here we restate the model as

$$x = Az + \epsilon \quad (59)$$

where  $x \in \mathbb{R}^d$  is the observation,  $z \in \{0, 1\}^K$  is a binary vector indicating the possession of certain latent vector and  $\epsilon$  is gaussian noise drawn from  $N(0, \sigma^2 \mathbf{1})$ .

### D.1 Concentration measure of unbounded moments

In order to utilize the bounds for gaussian random vectors, we need to subtract the term  $Az$  from  $x$  by operating  $[\hat{M} - M]$ . The bounds for observation generated by different  $z$  are examined separately. Let  $B = \{x_1, x_2, \dots, x_n\}$  and, for a specific  $z_i \in \{0, 1\}^K$ , write  $B_{z_i} := \{x \in B : z = z_i\}$  and  $\hat{w}_{z_i} = |B_{z_i}| / |B|$  for  $i \in \{0, 1 \dots 2^K - 1\}$  and  $z_i = \text{binary}(i)$ . Define the conditional moments to be

$$\begin{aligned} M_{1,z_i} &:= \mathbf{E}[x|z = z_i], \quad M_{2,z_i} := \mathbf{E}[x \otimes x|z = z_i], \quad M_{3,z_i} := \mathbf{E}[x \otimes x \otimes x|z = z_i], \\ M_{4,z_i} &:= \mathbf{E}[x \otimes x \otimes x \otimes x|z = z_i], \end{aligned}$$

while the empirical moments are

$$\begin{aligned} \hat{M}_{1,z_i} &:= |B_{z_i}|^{-1} \sum_{x \in B_{z_i}} x, \quad \hat{M}_{2,z_i} := |B_{z_i}|^{-1} \sum_{x \in B_{z_i}} x \otimes x, \quad \hat{M}_{3,z_i} := |B_{z_i}|^{-1} \sum_{x \in B_{z_i}} x \otimes x \otimes x, \\ \hat{M}_{4,z_i} &:= |B_{z_i}|^{-1} \sum_{x \in B_{z_i}} x \otimes x \otimes x \otimes x. \end{aligned}$$

**Lemma 5** (Concentration of conditional empirical moments) *With probability greater than  $1 - \delta$ , pick any  $\delta \in (0, 1)$  and any random matrix  $V \in \mathbb{R}^{d \times r}$  of rank  $r$ , the following guarantee holds*

1. *For the first-order moments, we have*

$$\left\| T(\hat{M}_{1,z_i} - M_{1,z_i}, V) \right\|_2 \leq \sigma \|V\|_2 \sqrt{\frac{r + 2\sqrt{r \ln(2^K/\delta)} + 2 \ln(2^K/\delta)}{\hat{w}_{z_i} n}}, \quad \forall i \in \{0, 1 \dots 2^K - 1\}$$

2. *For the second-order moments, we have*

$$\begin{aligned} &\left\| T(\hat{M}_{2,z_i} - M_{2,z_i}, V, V) \right\|_2 \\ &\leq \sigma^2 \|V\|_2^2 \left( \sqrt{\frac{128(r \ln 9 + \ln(2^{K+2}/\delta))}{\hat{w}_{z_i} n}} + \frac{4(r \ln 9 + \ln(2^{K+2}/\delta))}{\hat{w}_{z_i} n} \right) \\ &\quad + 2\sigma \left\| V^\top M_{1,z_i} \right\|_2 \|V\|_2 \sqrt{\frac{r + 2\sqrt{r \ln(2^{K+1}/\delta)} + 2 \ln(2^{K+1}/\delta)}{\hat{w}_{z_i} n}}, \quad \forall i \in \{0, 1 \dots 2^K - 1\} \end{aligned}$$

3. For the third-order moments, we have

$$\begin{aligned}
& \left\| T \left( \hat{M}_{3,z_i} - M_{3,z_i}, V, V, V \right) \right\|_2 \leq \sigma^3 \|V\|_2^3 \left( \sqrt{\frac{108e^3 [r \ln 13 + \ln(3 \cdot 2^K/\delta)]^3}{\hat{w}_{z_i} n}} \right) \\
& + 3\sigma^2 \left\| V^\top M_{1,z_i} \right\|_2 \|V\|_2^2 \left( \sqrt{\frac{128(r \ln 9 + \ln(3 \cdot 2^{K+1}/\delta))}{\hat{w}_{z_i} n}} + \frac{4(r \ln 9 + \ln(3 \cdot 2^{K+1}/\delta))}{\hat{w}_{z_i} n} \right) \\
& + 3\sigma \left\| V^\top M_{1,z_i} \right\|_2^2 \|V\|_2 \sqrt{\frac{r + 2\sqrt{r \ln(3 \cdot 2^K/\delta)} + 2 \ln(3 \cdot 2^K/\delta)}{\hat{w}_{z_i} n}}, \quad \forall i \in \{0, 1 \dots 2^K - 1\}
\end{aligned}$$

4. For the fourth-order moments, we have

$$\begin{aligned}
& \left\| T \left( \hat{M}_{4,z_i} - M_{4,z_i}, V, V, V, V \right) \right\|_2 \\
& \leq \sigma^4 \|V\|_2^4 \left( \sqrt{\frac{8192(r \ln 17 + \ln(4 \cdot 2^{K+1}/\delta))^2}{n^2}} + \frac{32(r \ln 17 + \ln(4 \cdot 2^{K+1}/\delta))^3}{n^3} \right) \\
& + 4\sigma^3 \left\| V^\top M_{1,z_i} \right\|_2 \|V\|_2^3 \left( \sqrt{\frac{108e^3 [r \ln 13 + \ln(4 \cdot 2^K/\delta)]^3}{\hat{w}_{z_i} n}} \right) \\
& + 6\sigma^2 \left\| V^\top M_{1,z_i} \right\|_2^2 \|V\|_2^2 \left( \sqrt{\frac{128(r \ln 9 + \ln(4 \cdot 2^{K+1}/\delta))}{\hat{w}_{z_i} n}} + \frac{4(r \ln 9 + \ln(4 \cdot 2^{K+1}/\delta))}{\hat{w}_{z_i} n} \right) \\
& + 4\sigma \left\| V^\top M_{1,z_i} \right\|_2^3 \|V\|_2 \sqrt{\frac{r + 2\sqrt{r \ln(4 \cdot 2^K/\delta)} + 2 \ln(4 \cdot 2^K/\delta)}{\hat{w}_{z_i} n}}, \quad \forall i \in \{0, 1 \dots 2^K - 1\},
\end{aligned}$$

**Proof** Here we only show the derivation of the fourth-order conditional moments. The other inequalities can be found in [23]. Under the stated model, the fourth-order conditional moment can be expended as

$$M_{4,z_i} = M_{1,z_i} \otimes M_{1,z_i} \otimes M_{1,z_i} \otimes M_{1,z_i} + \sigma^2 \mathfrak{S}_6 [M_{1,z_i} \otimes M_{1,z_i} \otimes \mathbf{1}] + \mathbf{E} [\epsilon \otimes \epsilon \otimes \epsilon \otimes \epsilon],$$

which yields

$$\begin{aligned}
& \hat{M}_{4,z_i} - M_{4,z_i} \\
& = \frac{1}{\hat{w}_{z_i} n} \left( \sum_{x \in B_{z_i}} (x_j - M_{1,z_i}) \otimes (x_j - M_{1,z_i}) \otimes (x_j - M_{1,z_i}) \otimes (x_j - M_{1,z_i}) - \sigma^4 \mathfrak{S}_3 [\mathbf{1} \otimes \mathbf{1}] \right. \\
& + \sum_{x \in B_{z_i}} (\mathfrak{S}_4 [M_{1,z_i} \otimes (x_j - M_{1,z_i}) \otimes (x_j - M_{1,z_i}) \otimes (x_j - M_{1,z_i})]) \\
& + \sum_{x \in B_{z_i}} (\mathfrak{S}_6 [M_{1,z_i} \otimes M_{1,z_i} \otimes ((x_j - M_{1,z_i}) \otimes (x_j - M_{1,z_i}) - \sigma^2 \mathbf{1})]) \\
& \left. + \sum_{x \in B_{z_i}} \mathfrak{S}_4 [M_{4,z_i} \otimes M_{4,z_i} \otimes M_{4,z_i} \otimes (x_j - M_{1,z_i})] \right) \quad (60)
\end{aligned}$$

Suppose  $V = V_1 \Sigma V_2^\top$  is the SVD of  $V$ , where  $V_1 \in \mathbb{R}^{d \times r}$  consists of orthonormal columns. With  $y_{j,z_i} = V_1^\top (x_j - M_{1,z_i})$ , applying triangle inequalities to Equation (60) yields

$$\begin{aligned}
& \left\| T \left( \hat{M}_{4,z_i} - M_{4,z_i}, V, V, V, V \right) \right\|_2 \\
& \leq \|V\|_2^4 \left\| \frac{1}{\hat{w}_{z_i} n} \sum_{x \in B_{z_i}} (y_{j,z_i} \otimes y_{j,z_i} \otimes y_{j,z_i} \otimes y_{j,z_i} - \sigma^4 \mathfrak{S}_3 [\mathbf{1} \otimes \mathbf{1}]) \right\|_2 \\
& + 4 \left\| V^\top M_{1,z_i} \right\|_2 \left\| \frac{1}{\hat{w}_{z_i} n} \sum_{x \in B_{z_i}} y_{j,z_i} \otimes y_{j,z_i} \otimes y_{j,z_i} \right\|_2 \\
& + 6 \left\| V^\top M_{1,z_i} \right\|_2^2 \left\| \frac{1}{\hat{w}_{z_i} n} \sum_{x \in B_{z_i}} (y_{j,z_i} \otimes y_{j,z_i} - \sigma^2 \mathbf{1}) \right\|_2 + 4 \left\| V^\top M_{1,z_i} \right\|_2^3 \left\| \frac{1}{\hat{w}_{z_i} n} \sum_{x \in B_{z_i}} y_{j,z_i} \right\|_2,
\end{aligned}$$

By using Lemma 13, we bound the first term by

$$\Pr \left[ \left\| \frac{1}{\hat{w}_{z_i} n} \sum_{x \in B_{z_i}} (y_{j,z_i} \otimes y_{j,z_i} \otimes y_{j,z_i} \otimes y_{j,z_i} - \mathbf{E}[\epsilon \otimes \epsilon \otimes \epsilon \otimes \epsilon]) \right\|_2 \right. \\ \left. > \sigma^4 \sqrt{\frac{8192 (r \ln 17 + \ln(2K/\delta))^2}{n^2} + \frac{32 (r \ln 17 + \ln(2K/\delta))^3}{n^3}} \right] \leq \delta. \quad (61)$$

■

The other norm can be bounded by using the bounds for low-order conditional moments. We finish the proof by adding the bounds for every term. By using inequalities for conditional moments, We get the bounds for completed moments stated in the following Lemma.

**Lemma 6** (Lemma 6 in [23]; Concentration of empirical moments) For a fixed matrix  $V \in \mathbb{R}^{d \times r}$ ,

$$\begin{aligned} & \left\| T(\hat{M}_i - M_i, V, \dots, V) \right\|_2 \\ & \leq (1 + 2^{K/2} \epsilon_w) \max_{z_j} \left\| T(\hat{M}_{i,z_j} - M_{i,z_j}, V, \dots, V) \right\|_2 + 2^{K/2} \max_{z_j} \left\| T(M_{i,z_j}, V, \dots, V) \right\|_2 \epsilon_w \\ & \quad \forall i \in [4], \forall j \in \{0, 1 \dots 2^K - 1\} \end{aligned}$$

where  $\epsilon_w = \left( \sum_{z_j} (\hat{w}_{z_j} - w_{z_j})^2 \right)^{\frac{1}{2}} \leq \frac{1 + \sqrt{\ln(1/\delta)}}{\sqrt{n}}.$

## D.2 Estimation of $\sigma, S_2, S_3, S_4$

Note that we have  $\sigma^2 = \lambda_{\min}[M_2 - M_1 \otimes M_1] = \varsigma_K[M_2 - M_1 \otimes M_1]$ , where  $\varsigma_t[M]$  denoting the  $t$ -th singular value of matrix  $M$  which is defined in Theorem 3. Here we define  $\hat{S}_{2,K}$  to be the best rank  $k$  approximation of  $\hat{M}_2 - \hat{M}_1 \otimes \hat{M}_1 - \hat{\sigma}^2 \mathbf{1}$ , which is the truncated matrix  $S_2$  in Algorithm 1.  $\hat{S}_i$  denotes the empirical tensors derived from summation of  $\hat{M}_i$  and  $\hat{\sigma}$ .  $S_i$  denotes the theoretical values.

**Lemma 7** (Accuracy of  $\sigma^2, \sigma^4$  and  $M_{2,K}$ )

$$|\hat{\sigma}^2 - \sigma^2| \leq \left\| \hat{M}_2 - M_2 \right\|_2 + \left\| \hat{M}_1 - M_1 \right\|_2^2 + 2 \left\| \hat{M}_1 - M_1 \right\|_2 \|M_1\|_2 \quad (62)$$

$$|\hat{\sigma}^4 - \sigma^4| \leq |\hat{\sigma}^2 - \sigma^2|^2 + 2\sigma^2 |\hat{\sigma}^2 - \sigma^2| \quad (63)$$

$$\left\| \hat{S}_{2,k} - S_2 \right\|_2 \leq 4 \left( \left\| \hat{M}_2 - M_2 \right\|_2 + \left\| \hat{M}_1 - M_1 \right\|_2^2 + 2 \|M_1\|_2 \left\| \hat{M}_1 - M_1 \right\|_2 \right) \quad (64)$$

### Proof

For the first order tensor, the inequality holds trivially due to the guarantees for  $\left\| \hat{M}_1 - M_1 \right\|_2$ . Next we bound the difference in variance estimates. Using the fact that differences in the  $k$ -th eigenvalues are bounded by the matrix norm of the difference we have that

$$|\hat{\sigma}^2 - \sigma^2| = \left| \varsigma_k[\hat{M}_2 - \hat{M}_1 \otimes \hat{M}_1] - \varsigma_k[M_2 - M_1 \otimes M_1] \right| \quad (65)$$

$$\leq \left\| [\hat{M}_2 - \hat{M}_1 \otimes \hat{M}_1] - [M_2 - M_1 \otimes M_1] \right\|_2 \quad (66)$$

$$\leq \left\| \hat{M}_2 - M_2 \right\|_2 + \left\| \hat{M}_1 - M_1 \right\|_2^2 + 2 \left\| \hat{M}_1 - M_1 \right\|_2 \|M_1\|_2. \quad (67)$$

The second inequality follows the Weyl's inequality and the last inequality is obtained by the triangle inequality. For estimation of  $\sigma^4$ ,

$$|\hat{\sigma}^4 - \sigma^4| \leq |(\hat{\sigma}^2 - \sigma^2)^2 + 2\sigma^2(\hat{\sigma}^2 - \sigma^2)| \leq |\hat{\sigma}^2 - \sigma^2|^2 + 2\sigma^2 |\hat{\sigma}^2 - \sigma^2|. \quad (68)$$

For the last claimed inequality, with Weyl's inequality,

$$\left\| \hat{S}_{2,k} - (\hat{M}_2 - \hat{M}_1 \otimes \hat{M}_1 - \hat{\sigma}^2 \mathbf{1}) \right\|_2 \leq \varsigma_{k+1}[\hat{M}_2 - \hat{M}_1 \otimes \hat{M}_1 - \hat{\sigma}^2 \mathbf{1}] \quad (69)$$

$$= \left\| \varsigma_{k+1}[\hat{M}_2 - \hat{M}_1 \otimes \hat{M}_1 - \hat{\sigma}^2 \mathbf{1}] - \varsigma_{k+1}[M_2 - M_1 \otimes M_1 - \sigma^2 \mathbf{1}] \right\|_2 \quad (70)$$

$$\leq \left\| \hat{M}_2 - \hat{M}_1 \otimes \hat{M}_1 - \hat{\sigma}^2 \mathbf{1} - (M_2 - M_1 \otimes M_1 - \sigma^2 \mathbf{1}) \right\|_2 \quad (71)$$

, which yields

$$\begin{aligned} \left\| \hat{S}_{2,k} - S_2 \right\|_2 &\leq \left\| \hat{S}_{2,k} - (\hat{M}_2 - \hat{M}_1 \otimes \hat{M}_1 - \hat{\sigma}^2 \mathbf{1}) \right\|_2 \\ &\quad + \left\| \hat{M}_2 - \hat{M}_1 \otimes \hat{M}_1 - \hat{\sigma}^2 \mathbf{1} - (M_2 - M_1 \otimes M_1 - \sigma^2 \mathbf{1}) \right\|_2 \end{aligned} \quad (72)$$

$$\leq 2 \left( \left\| \hat{M}_2 - M_2 \right\|_2 + \left\| \hat{M}_1 - M_1 \right\|_2^2 + 2 \|M_1\|_2 \left\| \hat{M}_1 - M_1 \right\|_2 + |\hat{\sigma}^2 - \sigma^2| \right) \quad (73)$$

$$\leq 4 \left( \left\| \hat{M}_2 - M_2 \right\|_2 + \left\| \hat{M}_1 - M_1 \right\|_2^2 + 2 \|M_1\|_2 \left\| \hat{M}_1 - M_1 \right\|_2 \right). \quad (74)$$

■

The inequalities for  $\sigma$  can be used for bounding the tensors  $S_2$ ,  $S_3$  and  $S_4$ , which will be shown next, and the inequality for  $S_{2,k}$  will be used in bounding whitened tensor in Section D.3.

**Lemma 8** (Accuracy of  $S_2$ ,  $S_3$  and  $S_4$ ) For a fixed matrix  $V \in \mathbb{R}^{d \times K}$

$$\begin{aligned} \left\| T(\hat{S}_2 - S_2, V, V) \right\|_2 &\leq \left\| T(\hat{M}_2 - M_2, V, V) \right\|_2 + \left\| T(\hat{M}_1 - M_1, V) \right\|_2^2 \\ &\quad + 2 \|T(M_1, V)\|_2 \left\| T(\hat{M}_1 - M_1, V) \right\|_2 + \|V\|_2^2 |\hat{\sigma}^2 - \sigma^2| \end{aligned} \quad (75)$$

$$\begin{aligned} &\left\| T(\hat{S}_3 - S_3, V, V, V) \right\|_2 \\ &\leq \left\| T(\hat{M}_3 - M_3, V, V, V) \right\|_2 + \left( \left\| T(\hat{M}_1 - M_1, V) \right\|_2 + \|T(M_1, V)\|_2 \right)^3 - \|T(M_1, V)\|_2^3 \\ &\quad + 3 \left( \left\| T(\hat{M}_1 - M_1, V) \right\|_2 \left\| T(\hat{S}_2 - S_2, V, V) \right\|_2 + \|T(M_1, V)\|_2 \left\| T(\hat{S}_2 - S_2, V, V) \right\|_2 \right. \\ &\quad \left. + \left\| T(\hat{M}_1 - M_1, V) \right\|_2 \|T(S_2, V, V)\|_2 \right) + 3 \|V\|_2^2 \left( |\hat{\sigma}^2 - \sigma^2| \left\| T(\hat{M}_1 - M_1, V) \right\|_2 \right. \\ &\quad \left. + \sigma^2 \left\| T(\hat{M}_1 - M_1, V) \right\|_2 + |\hat{\sigma}^2 - \sigma^2| \|T(M_1, V)\|_2 \right) \end{aligned} \quad (76)$$

$$\begin{aligned} &\left\| T(\hat{S}_4 - S_4, V, V, V, V) \right\|_2 \\ &\leq \left\| T(\hat{M}_4 - M_4, V, V, V, V) \right\|_2 + \left( \left\| T(\hat{M}_1 - M_1, V) \right\|_2 + \|T(M_1, V)\|_2 \right)^4 - \|T(M_1, V)\|_2^4 \\ &\quad + 6 \left\| T(\hat{S}_2 - S_2, V, V) \right\|_2 \|T(M_1, V)\|_2^2 + 6 \left( \left\| T(\hat{S}_2 - S_2, V, V) \right\|_2 + \|T(S_2, V, V)\|_2 \right) \\ &\quad \left( 2 \|T(M_1, V)\|_2 \left\| T(\hat{M}_1 - M_1, V) \right\|_2 + \left\| T(\hat{M}_1 - M_1, V) \right\|_2^2 \right) + 3 \left( \left\| T(\hat{S}_2 - S_2, V, V) \right\|_2^2 \right. \\ &\quad \left. + 2 \left\| T(\hat{S}_2 - S_2, V, V) \right\|_2 \|T(S_2, V, V)\|_2 \right) + 6 \|V\|_2^2 \left( \sigma^2 \left\| T(\hat{S}_2 - S_2, V, V) \right\|_2 + \right. \\ &\quad \left. + |\hat{\sigma}^2 - \sigma^2| \left( \left\| T(\hat{S}_2 - S_2, V, V) \right\|_2 + \|T(S_2, V, V)\|_2 \right) \right) + 3 |\hat{\sigma}^4 - \sigma^4| \|V\|_2^4 \\ &\quad + 4 \left( \left\| T(\hat{S}_3 - S_3, V, V, V) \right\|_2 \left\| T(\hat{M}_1 - M_1, V) \right\|_2 + \|T(M_1, V)\|_2 \left\| T(\hat{S}_3 - S_3, V, V, V) \right\|_2 \right. \\ &\quad \left. + \|T(S_3, V, V, V)\|_2 \left\| T(\hat{M}_1 - M_1, V) \right\|_2 \right) \end{aligned} \quad (77)$$

**Proof** To bound the second order tensor, we use the inequality for bounding  $\hat{\sigma}$  in Lemma 7 and get

$$\begin{aligned} &\left\| T(\hat{S}_2, V, V) - T(S_2, V, V) \right\|_2 \\ &\leq \left\| T(\hat{M}_2 - M_2, V, V) \right\|_2 + \left\| T((\hat{M}_1 - M_1) \otimes (\hat{M}_1 - M_1), V, V) \right\|_2 \\ &\quad + 2 \left\| T(M_1 \otimes (\hat{M}_1 - M_1), V, V) \right\|_2 + \|V\|_2^2 |\hat{\sigma}^2 - \sigma^2| \end{aligned} \quad (78)$$

$$\begin{aligned} &\leq \left\| T(\hat{M}_2 - M_2, V, V) \right\|_2 + \left\| T(\hat{M}_1 - M_1, V, V) \right\|_2^2 + 2 \|T(M_1, V)\|_2 \left\| T(\hat{M}_1 - M_1, V, V) \right\|_2 \\ &\quad + \|V\|_2^2 |\hat{\sigma}^2 - \sigma^2|. \end{aligned} \quad (79)$$

Similarly, for  $\hat{S}_3$ , we have that

$$\begin{aligned} & \left\| T(\hat{S}_3, V, V, V) - T(S_3, V, V, V) \right\|_2 \\ & \leq \left\| T(\hat{M}_3 - M_3, V, V, V) \right\|_2 + \left\| T(\hat{M}_1 \otimes \hat{M}_1 \otimes \hat{M}_1 - M_1 \otimes M_1 \otimes M_1, V, V, V) \right\|_2 \\ & \quad + 3 \left\| T(\hat{S}_1 \otimes \hat{S}_2 - S_1 \otimes S_2, V, V, V) \right\|_2 + 3 \left\| T((\hat{\sigma}^2 \hat{M}_1 - \sigma^2 M_1) \otimes \mathbf{1}, V, V, V) \right\|_2. \end{aligned} \quad (80)$$

Note that the second term can be written as

$$\begin{aligned} & \hat{M}_1 \otimes \hat{M}_1 \otimes \hat{M}_1 - M_1 \otimes M_1 \otimes M_1 \\ & = (\hat{M}_1 - M_1) \otimes (\hat{M}_1 - M_1) \otimes (\hat{M}_1 - M_1) \\ & \quad + \mathfrak{S}_3 [M_1 \otimes (\hat{M}_1 - M_1) \otimes (\hat{M}_1 - M_1)] + \mathfrak{S}_3 [M_1 \otimes M_1 \otimes (\hat{M}_1 - M_1)]. \end{aligned} \quad (81)$$

Using the same expansion trick, the third term becomes

$$\hat{S}_1 \otimes \hat{S}_2 - S_1 \otimes S_2 = (\hat{S}_1 - S_1) \otimes (\hat{S}_2 - S_2) + S_1 \otimes (\hat{S}_2 - S_2) + (\hat{S}_1 - S_1) \otimes S_2. \quad (82)$$

Using triangle inequality, the bound for Equation (81) is

$$\begin{aligned} & \left\| T(\hat{M}_1 \otimes \hat{M}_1 \otimes \hat{M}_1 - M_1 \otimes M_1 \otimes M_1, V, V, V) \right\|_2 \\ & \leq \left\| T(\hat{M}_1 - M_1, V) \right\|_2^3 + 3 \left\| T(M_1, V) \right\|_2 \left\| T(\hat{M}_1 - M_1, V) \right\|_2^2 \\ & \quad + 3 \left\| T(M_1, V) \right\|_2^2 \left\| T(\hat{M}_1 - M_1, V) \right\|_2, \end{aligned} \quad (83)$$

and the bound for Equation (82) is

$$\begin{aligned} & \left\| T(\hat{S}_1 \otimes \hat{S}_2 - S_1 \otimes S_2, V, V, V) \right\|_2 \\ & \leq \left\| T(\hat{S}_1 - S_1, V) \right\|_2 \left\| T(\hat{S}_2 - S_2, V, V) \right\|_2 + \left\| T(S_1, V) \right\|_2 \left\| T(\hat{S}_2 - S_2, V, V) \right\|_2 \\ & \quad + \left\| T(\hat{S}_1 - S_1, V) \right\|_2 \left\| T(S_2, V, V) \right\|_2 \\ & \left\| T((\hat{\sigma}^2 \hat{M}_1 - \sigma^2 M_1) \otimes \mathbf{1}, V, V, V) \right\|_2 \\ & \leq \|V\|_2^2 (|\hat{\sigma}^2 - \sigma^2| \left\| T(\hat{M}_1 - M_1, V) \right\|_2 + \sigma^2 \left\| T(\hat{M}_1 - M_1, V) \right\|_2 + |\hat{\sigma}^2 - \sigma^2| \left\| T(M_1, V) \right\|_2). \end{aligned} \quad (84)$$

By combining all the inequalities, we get the bound for  $S_3$ . The bound for  $S_4$  can be derived by similar procedure.  $\blacksquare$

To complete the bounds, we need to examine the bounds for the whitening matrix and also the whitened tensors.

### D.3 Properties with whitening matrix

Note that in Algorithm 1 we have  $W_3 := T(S_3, W, W, W)$ ,  $W_4 := T(S_4, W, W, W)$ . To bound  $\|W_3\|$  and  $\|W_4\|$ , we use the fact stated in Section C that these tensor are diagonalized so that finding the norm is actually equivalent to finding the largest eigenvalue of  $T(S_3, W, W, W)$  and  $T(S_4, W, W, W)$ , respectively. Note that in Algorithm 1, the first  $K_1$  eigenvectors and their corresponding eigenvalues are solved by conducting tensor decomposition on  $W_3$ , while the others are extracted from  $W_4$ . With Equation (53) and (57),

$$\lambda_i = \begin{cases} \frac{-2\pi_i + 1}{\sqrt{\pi_i - \pi_i^2}} & \text{if } i \leq K_1 \\ \frac{6\pi_i^2 - 6\pi_i + 1}{\pi_i - \pi_i^2} & \text{otherwise.} \end{cases} \quad (86)$$

As we have mentioned previously, eigenvalues of  $S_3$  degenerate to zero at the value of  $\pi_i = 0.5$  while eigenvalues of  $S_4$  degenerate to zero at the value of  $\pi_i \approx 0.2, 0.8$ . So here we define thresholds,  $\pi_{Th_{up}}$  and  $\pi_{Th_{down}}$ , such that

$$\frac{-2\pi_{Th_{down}} + 1}{\sqrt{\pi_{Th_{down}} - \pi_{Th_{down}}^2}} = 1, \quad \frac{-2\pi_{Th_{up}} + 1}{\sqrt{\pi_{Th_{up}} - \pi_{Th_{up}}^2}} = -1. \quad (87)$$



In other words, we solve the latent factors by the third-order moments if  $\pi_i < \pi_{Th_{down}}$  or  $\pi_i > \pi_{Th_{up}}$ , otherwise we turn to the fourth-order moments. Since  $\lambda_i$  is a symmetric function of  $\pi_i$  on the  $\pi_i = 0.5$  axis for  $i \in [K]$ , we set  $\pi_{Th} = \pi_{Th_{down}}$  to simplify the proof. Here we have

$$1 = \left| \frac{-2\pi_{Th} + 1}{\sqrt{\pi_{Th} - \pi_{Th}^2}} \right| \leq |\lambda_i| \leq \frac{-2\pi_{min} + 1}{\sqrt{\pi_{min} - \pi_{min}^2}} \quad \text{if } i \leq K_1 \quad (88)$$

$$-2 \leq \lambda_i \leq \frac{6\pi_{Th}^2 - 6\pi_{Th} + 1}{\pi_{Th} - \pi_{Th}^2} \approx -1 \quad \text{otherwise,} \quad (89)$$

where  $\pi_{min} = \arg\max_{i \in [K_1]} |\pi_i - 0.5|$ . Since  $W_3$  and  $W_4$  are diagonalized tensor, we have that

$$\|W_3\|_2 \leq \frac{-2\pi_{min} + 1}{\sqrt{\pi_{min} - \pi_{min}^2}}, \quad \|W_4\|_2 \leq 2. \quad (90)$$

Next, in order to bound  $[\hat{W}_i - W_i]$ , we need to consider the bounds using empirical whitening matrix. Let  $\hat{W}$  denotes the empirical whitening matrix in our algorithm. Here we define  $W := \hat{W}(\hat{W}S_2\hat{W})^{-\frac{1}{2}}$  and  $\epsilon_{S_2} := \|\hat{S}_{2,k} - S_2\|_2 / \varsigma_k[S_2]$  in order to use the bounds for whitening matrix stated in lemma 10 in [23].

**Lemma 9** (Lemma 10 in [23]) Assume  $\epsilon_{S_2} \leq 1/3$ . We have

$$\begin{aligned} 1. & W^\top S_2 W = I, \quad 2. \quad \|\hat{W}\|_2 \leq \frac{1}{\sqrt{(1 - \epsilon_{S_2})\varsigma[S_2]}}, \\ 3. & \left\| \left( \hat{W} S_2 \hat{W} \right)^{1/2} - I \right\|_2 \leq 1.5\epsilon_{S_2}, \quad \left\| \left( \hat{W} S_2 \hat{W} \right)^{-1/2} - I \right\|_2 \leq 1.5\epsilon_{S_2} \\ & \left\| (\hat{W})^\top \text{Adiag}(\pi - \pi^2)^{1/2} \right\|_2 \leq \sqrt{1 + 1.5\epsilon_{S_2}}, \\ & \left\| (\hat{W} - W)^\top \text{Adiag}(\pi - \pi^2)^{1/2} \right\|_2 \leq \sqrt{1 + 1.5\epsilon_{S_2}}. \end{aligned}$$

Using Lemma 9, we can complete the bounds for empirical whitened tensors.

**Lemma 10** Assume  $\epsilon_{S_2} \leq 1/3$ . Then

$$\begin{aligned} \|\hat{W}_3 - W_3\|_2 &\leq \|T(S_3 - \hat{S}_3, \hat{W}, \hat{W}, \hat{W})\|_2 + 3 \frac{-2\pi_{min} + 1}{\sqrt{\pi_{min} - \pi_{min}^2}} \\ \|\hat{W}_4 - W_4\|_2 &\leq \|T(S_4 - \hat{S}_4, \hat{W}, \hat{W}, \hat{W})\|_2 + 10 \end{aligned}$$

**Proof** Here we only show the second inequality, the first one can be derived with similar procedure.

$$\begin{aligned} \|\hat{W}_4 - W_4\|_2 &= \|T(S_4, W, W, W, W) - T(\hat{S}_4, \hat{W}, \hat{W}, \hat{W}, \hat{W})\|_2 \\ &\leq \|T(S_4, W, W, W, W) - T(S_4, \hat{W}, \hat{W}, \hat{W}, \hat{W})\|_2 + \|T(S_4 - \hat{S}_4, \hat{W}, \hat{W}, \hat{W}, \hat{W})\|_2 \end{aligned} \quad (91)$$

For the first term, using Lemma 9 and Equation (90), we have:

$$\begin{aligned} & \|T(S_4, W, W, W, W) - T(S_4, \hat{W}, \hat{W}, \hat{W}, \hat{W})\|_2 \\ & \leq \|T(S_4, \hat{W} - W, \hat{W}, \hat{W}, \hat{W})\|_2 + \|T(S_4, W, \hat{W} - W, \hat{W}, \hat{W})\|_2 \\ & \quad + \|T(S_4, W, W, \hat{W} - W, \hat{W})\|_2 + \|T(S_4, W, W, W, \hat{W} - W)\|_2 \end{aligned} \quad (92)$$

$$\begin{aligned} & \leq \|T(S_4, W, W, W, W)\|_2 \left\| (\hat{W}^\top S_2 \hat{W})^{1/2} - \mathbf{1} \right\|_2 \left( \left\| (\hat{W}^\top S_2 \hat{W})^{1/2} \right\|_2^3 + \dots + \left\| (\hat{W}^\top S_2 \hat{W})^{1/2} \right\|_2^0 \right) \\ & \leq \|T(S_4, W, W, W, W)\|_2 \cdot (1.5\epsilon_{S_2}) \left( (1 + 1.5\epsilon_{S_2})^3 + \dots + (1 + 1.5\epsilon_{S_2}) + 1 \right) \leq 5 \cdot 2 = 10 \end{aligned} \quad (93)$$

■

#### D.4 Reconstruction analysis

Before putting everything together, we utilize the eigendecomposition analysis in Appendix C.7 of [23]. First, we consider the case where  $A_i$  is recovered by applying tensor decomposition on  $W_3$ , i.e., for  $i \leq K_1$ . Note that, in Algorithm 1,  $Z_i = \sqrt{\pi_i - \pi_i^2}$ . Following the approach in [23], define

$$\gamma_{S_3} := \frac{1}{2 \max_{i \in [K_1]} \sqrt{(\pi_i - \pi_i^2)} \sqrt{eK} \binom{K+1}{2}}, \quad \epsilon_{S_3} := \frac{\|W_3 - \hat{W}_3\|}{\gamma_{S_3}},$$

$$\gamma_{S_4} := \frac{1}{2 \max_{i > K_1} \sqrt{(\pi_i - \pi_i^2)} \sqrt{eK} \binom{K+1}{2}}, \quad \epsilon_{S_4} := \frac{\|W_4 - \hat{W}_4\|}{\gamma_{S_4}}.$$

**Lemma 11** (Reconstruction Accuracy in [23]) Assume  $\epsilon_{S_2} \leq 1/3$ ,  $\epsilon_{S_3} \leq 1/4$  and  $\epsilon_{S_4} \leq 1/4$ , and  $\epsilon_1 \leq 1/3$ . With  $\kappa[S_2] := \varsigma_1[S_2]/\varsigma_K[S_2]$ , there exists a permutation  $\pi$  on  $[K]$  such that

$$\|A_{\pi(i)} - \hat{A}_i\| \leq 3 \|A_{\pi(i)}\|_2 \epsilon_{1,i} + 2 \|S_2\|_2^{1/2} \epsilon_{0,i}, \quad \forall i \in [K]$$

where

$$\epsilon_{0,i} = \begin{cases} (5.5\epsilon_{S_2} + 7\epsilon_{S_3}) / \sqrt{\pi_{\min} - \pi_{\min}^2} & \text{if } i \in [K_1] \\ 13.75\epsilon_{S_2} + 17.5\epsilon_{S_4} & \text{otherwise} \end{cases},$$

$$\epsilon_{1,i} = \begin{cases} \left( \left( (6.875\kappa[S_2]^{1/2} + 2) \epsilon_{S_2} + (8.75\kappa[S_2]^{1/2} + \gamma_{S_3} \sqrt{\pi_{\min} - \pi_{\min}^2}) \epsilon_{S_3} \right) / (\gamma_{S_3} \sqrt{\pi_{\min} - \pi_{\min}^2}) \right) & \text{if } i \in [K_1], \\ 2.5 \left( (6.875\kappa[S_2]^{1/2} + 2) \epsilon_{S_2} + (8.75\kappa[S_2]^{1/2} + 0.4\gamma_{S_4}) \epsilon_{S_4} \right) / \gamma_{S_4} & \text{otherwise} \end{cases}$$

#### D.5 Proof of Theorem 3

The proof can be accomplished by finding out the sufficient sample size that satisfies the assumption in Lemma 11. First, assume that  $n \geq Ck \log(1/\delta)$  to make sure  $\varepsilon_w \leq 1$ . In this proof, we use  $c, c_1, c_2, \dots$  to denote some positive constant. By Lemma 5 and 6, with probability greater than  $1 - \delta$ , bounds for first and second-order moments are

$$\|\hat{M}_1 - M_1\|_2 \leq c_1 \sigma \sqrt{\frac{d + \log(2^k/\delta)}{\tilde{\pi}n}} + c_1 \sum_{i=1}^K \|A_i\|_2 \sqrt{\frac{2^K \log(1/\delta)}{n}} \quad (94)$$

$$\|\hat{M}_2 - M_2\|_2 \leq c_1 \left( \sigma^2 \sqrt{\frac{d + \log(2^k/\delta)}{\tilde{\pi}n}} + \sigma^2 \frac{d + \log(2^k/\delta)}{\tilde{\pi}n} + \sigma \sqrt{\frac{d + \log(2^k/\delta)}{\tilde{\pi}n}} \right) \quad (95)$$

$$+ c_1 \left( \sum_{i=1}^K \|A_i\|_2^2 + \sigma^2 \right) \sqrt{\frac{2^K \log(1/\delta)}{n}} \quad (96)$$

$$\leq c_1 \left( 2 \left( \sum_{i=1}^K \|A_i\|_2^2 + \sigma^2 \right) \sqrt{\frac{d + \log(2^k/\delta)}{\tilde{\pi}n}} + \sigma^2 \frac{d + \log(2^k/\delta)}{\tilde{\pi}n} \right) \quad (97)$$

Using Lemma 7, we derive the bound for  $\sigma$  and approximate rank-k second-order tensor as

$$\begin{aligned} & \max \left\{ |\hat{\sigma}^2 - \sigma^2|, \|\hat{S}_{2,K} - S_2\|_2 \right\} \\ & \leq 4c_1 \left( 2 \left( \sum_{i=1}^K \|A_i\|_2^2 + \sigma^2 \right) \sqrt{\frac{d + \log(2^k/\delta)}{\tilde{\pi}n}} + \sigma^2 \frac{d + \log(2^k/\delta)}{\tilde{\pi}n} \right) \\ & \quad + 8c_1^2 \left( \sigma \sqrt{\frac{d + \log(2^k/\delta)}{\tilde{\pi}n}} + \sum_{i=1}^K \|A_i\|_2 \sqrt{\frac{2^K \log(1/\delta)}{n}} \right)^2 \\ & \quad + 8c_1 \|M_1\|_2 \left( \sigma \sqrt{\frac{d + \log(2^k/\delta)}{\tilde{\pi}n}} + \sum_{i=1}^K \|A_i\|_2 \sqrt{\frac{2^K \log(1/\delta)}{n}} \right) \end{aligned} \quad (98)$$

$$\leq c_2 \left( \sum_{i=1}^K \|A_i\|_2^2 + \sigma^2 \right) \left( \sqrt{\frac{d + \log(2^k/\delta)}{\tilde{\pi}n}} + \frac{d + \log(2^k/\delta)}{\tilde{\pi}n} \right). \quad (99)$$

To ensure that

$$\max \left\{ \frac{|\hat{\sigma}^2 - \sigma^2|}{\varsigma_K(S_2)}, \epsilon_{S_2} \right\} \leq c_3 \frac{\gamma_{S_3}^2 \tilde{\pi}}{\kappa[S_2]^{1/2}} \leq 1/3, \quad (100)$$

we set the sample size as

$$n \geq c \frac{d + \log(2^k/\delta)}{\tilde{\pi}} \left( \left[ \frac{\kappa[S_2]^{1/2} \left( \sum_{i=1}^K \|A_i\|_2^2 + \sigma^2 \right)}{\gamma_{S_3}^2 \tilde{\pi} \varsigma_K[S_2] \epsilon} \right]^2 + \left[ \frac{\kappa[S_2]^{1/2} \left( \sum_{i=1}^K \|A_i\|_2^2 + \sigma^2 \right)}{\gamma_{S_3}^2 \tilde{\pi} \varsigma_K[S_2] \epsilon} \right] \right).$$

To examine the moments after multiplying whitening matrix  $W$ , by Lemma 9,

$$\|\hat{W}\|_2 \leq \sqrt{1.5/\varsigma_K[S_2]} \quad (101)$$

$$\max_{z_i \in [2^K]} \|T(M_{1,z_i}, \hat{W})\| \leq \|\hat{W}^\top \text{Adiag}(\pi - \pi^2)^{1/2}\|_2 / \sqrt{\pi_{\min} - \pi_{\min}^2} \quad (102)$$

$$\leq \sqrt{1.5/(\pi_{\min} - \pi_{\min}^2)} \quad (103)$$

$$\max_{z_i \in [2^K]} \|T(M_{2,z_i}, \hat{W}, \hat{W})\| \leq 1.5/(\pi_{\min} - \pi_{\min}^2) + \sigma^2(1.5/\varsigma_K[S_2]) \quad (104)$$

$$\begin{aligned} \max_{z_i \in [2^K]} \|T(M_{3,z_i}, \hat{W}, \hat{W}, \hat{W})\| &\leq (1.5/(\pi_{\min} - \pi_{\min}^2))^{3/2} \\ &\quad + 3\sigma^2 \sqrt{1.5/(\pi_{\min} - \pi_{\min}^2)} (1.5/\varsigma_K[S_2]) \end{aligned} \quad (105)$$

$$\begin{aligned} \max_{z_i \in [2^K]} \|T(M_{4,z_i}, \hat{W}, \hat{W}, \hat{W}, \hat{W})\| &\leq (1.5/(\pi_{\min} - \pi_{\min}^2))^2 + 6\sigma^2 \frac{2.25}{(\pi_{\min} - \pi_{\min}^2) \varsigma_K[S_2]} \\ &\quad + 3\sigma^4 (1.5/\varsigma_K[S_2])^2 \end{aligned} \quad (106)$$

Using Lemma 6,

$$\|T(\hat{M}_1 - M_1, \hat{W})\| \leq c_4 \frac{\sigma}{\varsigma_K[S_2]^{1/2}} \sqrt{\frac{K + \log(2^K/\delta)}{\tilde{\pi}n}} + c_4 \frac{1}{\sqrt{\pi_{\min} - \pi_{\min}^2}} \sqrt{\frac{2^K \log(1/\delta)}{n}} \quad (107)$$

$$\begin{aligned} \|T(\hat{M}_2 - M_2, \hat{W}, \hat{W})\| &\leq c_4 \frac{\sigma^2}{\varsigma_K[S_2]} \left( \sqrt{\frac{K + \log(2^K/\delta)}{\tilde{\pi}n}} + \frac{K + \log(2^K/\delta)}{\tilde{\pi}n} \right) \\ &\quad + c_4 \frac{1}{\sqrt{\pi_{\min} - \pi_{\min}^2} \varsigma_K[S_2]^{1/2}} \sqrt{\frac{2^K \log(1/\delta)}{n}} \\ &\quad + c_4 \left( \frac{1}{\varsigma_K[S_2]} + \frac{1}{\sqrt{\pi_{\min} - \pi_{\min}^2}} \right) \sqrt{\frac{K \log(1/\delta)}{n}} \end{aligned} \quad (108)$$

$$\begin{aligned} \|T(\hat{M}_3 - M_3, \hat{W}, \hat{W}, \hat{W})\| &\leq c_4 \frac{\sigma^3}{\varsigma_K[S_2]^{3/2}} \sqrt{\frac{(K + \log(2^K/\delta))^3}{\tilde{\pi}n}} \\ &\quad + c_4 \frac{\sigma^2}{\varsigma_K[S_2] \sqrt{\pi_{\min} - \pi_{\min}^2}} \left( \sqrt{\frac{K + \log(2^K/\delta)}{\tilde{\pi}n}} + \frac{K + \log(2^K/\delta)}{\tilde{\pi}n} \right) \\ &\quad + c_4 \frac{\sigma}{\varsigma_K[S_2]^{1/2} (\pi_{\min} - \pi_{\min}^2)} \sqrt{\frac{K + \log(2^K/\delta)}{\tilde{\pi}n}} \\ &\quad + c_4 \left( \frac{\sigma^2}{\varsigma_K[S_2] \sqrt{\pi_{\min} - \pi_{\min}^2}} + \frac{1}{(\pi_{\min} - \pi_{\min}^2)^{3/2}} \right) \sqrt{\frac{K \log(1/\delta)}{n}} \end{aligned} \quad (109)$$

$$\begin{aligned}
\left\| T\left(\hat{M}_4 - M_4, \hat{W}, \hat{W}, \hat{W}\right) \right\| &\leq c_4 \frac{\sigma^4}{\varsigma_K [S_2]^2} \left( \left( \frac{K + \log(2^K/\delta)}{\tilde{\pi}n} \right) + \left( \frac{K + \log(2^K/\delta)}{\tilde{\pi}n} \right)^{3/2} \right) \\
&\quad + c_4 \frac{\sigma^3}{\varsigma_K [S_2]^{3/2}} \sqrt{\frac{(K + \log(2^K/\delta))^3}{\tilde{\pi}n}} \\
&\quad + c_4 \frac{\sigma^2}{\varsigma_K [S_2]} \left( \sqrt{\frac{K + \log(2^K/\delta)}{\tilde{\pi}n}} + \frac{K + \log(2^K/\delta)}{\tilde{\pi}n} \right) \\
&\quad + c_4 \frac{\sigma}{\varsigma_K [S_2]^{1/2}} \sqrt{\frac{K + \log(2^K/\delta)}{\tilde{\pi}n}} \\
&\quad + c_4 \left( \frac{\sigma^4}{\varsigma_K [S_2]} + \frac{\sigma^2}{\varsigma_K [S_2]} \right) \sqrt{\frac{K \log(1/\delta)}{n}}. \tag{110}
\end{aligned}$$

With Lemma 8 and 10,

$$\begin{aligned}
\left\| T\left(\hat{S}_2 - S_2, \hat{W}, \hat{W}\right) \right\|_2 &\leq \left\| T\left(\hat{M}_2 - M_2, \hat{W}, \hat{W}\right) \right\|_2 + \left\| T\left(\hat{M}_1 - M_1, \hat{W}\right) \right\|_2^2 \\
&\quad + 2 \frac{\pi_{max}}{\sqrt{\pi_{max} - \pi_{max}^2}} \left\| T\left(\hat{M}_1 - M_1, \hat{W}\right) \right\|_2 + \frac{1.5}{\varsigma_K [S_2]} |\hat{\sigma}^2 - \sigma^2| \tag{111}
\end{aligned}$$

$$\begin{aligned}
\left\| T\left(\hat{S}_3 - S_3, \hat{W}, \hat{W}, \hat{W}\right) \right\| &\leq \left\| T\left(\hat{M}_3 - M_3, \hat{W}, \hat{W}, \hat{W}\right) \right\|_2 \\
&\quad + \left( \left\| T\left(\hat{M}_1 - M_1, \hat{W}\right) \right\|_2 + \frac{\pi_{max}}{\sqrt{\pi_{max} - \pi_{max}^2}} \right)^3 - \left( \frac{\pi_{max}}{\sqrt{\pi_{max} - \pi_{max}^2}} \right)^3 \\
&\quad + 3 \left( \left\| T\left(\hat{M}_1 - M_1, \hat{W}\right) \right\|_2 \left\| T\left(\hat{S}_2 - S_2, \hat{W}, \hat{W}\right) \right\|_2 + \frac{\pi_{max}}{\sqrt{\pi_{max} - \pi_{max}^2}} \left\| T\left(\hat{S}_2 - S_2, \hat{W}, \hat{W}\right) \right\|_2 \right. \\
&\quad \left. + \left\| T\left(\hat{M}_1 - M_1, \hat{W}\right) \right\|_2 \frac{-2\pi_{min} + 1}{\sqrt{\pi_{min} - \pi_{min}^2}} \right) + \frac{4.5}{\varsigma_K [S_2]} \left( |\hat{\sigma}^2 - \sigma^2| \left\| T\left(\hat{M}_1 - M_1, \hat{W}\right) \right\|_2 \right. \\
&\quad \left. + \sigma^2 \left\| T\left(\hat{M}_1 - M_1, \hat{W}\right) \right\|_2 + |\hat{\sigma}^2 - \sigma^2| \frac{\pi_{max}}{\sqrt{\pi_{max} - \pi_{max}^2}} \right). \tag{112}
\end{aligned}$$

Plug this in Lemma 10, we get the overall bounds for  $\left\| W_3 - \hat{W}_3 \right\|$ . To get  $\epsilon_{S_3} \leq c_5 \frac{\gamma_{S_3} \sqrt{\tilde{\pi}}}{\kappa[S_2]^{1/2}} \epsilon$ , we set

$$n \geq \text{poly} \left( d, K, \frac{1}{\epsilon}, \log(1/\delta), \frac{1}{\tilde{\pi}}, \frac{\varsigma_1 [S_2]}{\varsigma_K [S_2]}, \frac{\sum_{i=1}^K \|A_i\|_2^2}{\varsigma_K [S_2]}, \frac{\sigma^2}{\varsigma_K [S_2]}, \frac{1}{\sqrt{\pi_{min} - \pi_{min}^2}}, \frac{\pi_{max}}{\sqrt{\pi_{max} - \pi_{max}^2}} \right) \tag{113}$$

Similarly, for  $A_i$  reconstructed by  $\hat{W}_4$ ,  $n$  should be set to

$$n \geq \text{poly} \left( d, K, \frac{1}{\epsilon}, \log(1/\delta), \frac{1}{\tilde{\pi}}, \frac{\varsigma_1 [S_2]}{\varsigma_K [S_2]}, \frac{\sum_{i=1}^K \|A_i\|_2^2}{\varsigma_K [S_2]}, \frac{\sigma^2}{\varsigma_K [S_2]} \right), \tag{114}$$

in order to  $\epsilon_{S_4} \leq c_6 \frac{\gamma_{S_4} \sqrt{\tilde{\pi}}}{\kappa[S_2]^{1/2}} \epsilon$ . The overall bounds can be obtained by Equation 113, 114 and Lemma 11.

## E Tail Inequalities

Here we derive the tail inequality for the fourth-order subgaussian random tensor.

**Lemma 12** *Let  $x_1, x_2, \dots, x_n$  be i.i.d. random variables such that*

$$\mathbf{E}_i [\exp(\eta x_i)] \leq \exp(\gamma \eta^2/2) \quad \forall \eta \in \mathbb{R} \tag{115}$$

Then for any  $t > 0$  and  $\frac{\gamma t}{n} < \frac{1}{4}$ ,

$$\Pr \left[ \frac{1}{n} \sum_{i=1}^n (x_i^4 - \mathbf{E}_i [x_i^4]) > \gamma \sqrt{\frac{64t^2}{n^2} \left( 8\gamma - \frac{16\gamma^2 t}{n} \right) \frac{1}{(1 - 4\gamma \frac{t}{n})^2}} \right] \leq e^{-t}, \quad (116)$$

$$\Pr \left[ \frac{1}{n} \sum_{i=1}^n (x_i^4 - \mathbf{E}_i [x_i^4]) < -\gamma \sqrt{\frac{8t^2}{n^2} \left( 2\gamma + \frac{\gamma^2 t}{n} \right) \frac{1}{(1 + \gamma \frac{t}{n})^2}} \right] \leq e^{-t}, \quad (117)$$

**Proof** We use Chernoff's bounding method to derive the inequality. For  $\eta < \frac{1}{2\epsilon\gamma}$ , set  $\eta = \frac{1-\sigma}{2\gamma\epsilon}$  for some  $\sigma > 0$ , we have

$$\mathbf{E}_i [\exp(\eta x_i^4)] = 1 + \eta \mathbf{E}_i [x_i^4] + \eta \int_0^\infty (\exp(\eta \epsilon^2) - 1) \mathbf{E}_i [\mathbb{1}_{\{x_i^4 > \epsilon^2\}}] d\epsilon^2 \quad (118)$$

$$\leq 1 + \eta \mathbf{E}_i [x_i^4] + 2\eta \int_0^\infty (\exp(\eta \epsilon^2) - 1) \exp\left(\frac{-\epsilon}{2\gamma}\right) 2\epsilon d\epsilon \quad (119)$$

$$\leq 1 + \eta \mathbf{E}_i [x_i^4] + 4\eta \left( \int_0^\infty \epsilon \exp\left(\frac{-\sigma\epsilon}{2\gamma}\right) d\epsilon - \int_0^\infty \epsilon \exp\left(\frac{-\epsilon}{2\gamma}\right) d\epsilon \right) \quad (120)$$

$$\leq 1 + \eta \mathbf{E}_i [x_i^4] + 4\eta \left( 4\gamma^2 \left( \frac{1}{\sigma^2} - 1 \right) \right) \quad (121)$$

$$\leq \exp \left( \eta \mathbf{E}_i [x_i^4] + 4\eta \left( 4\gamma^2 \left( \frac{1}{\sigma^2} - 1 \right) \right) \right) \quad (122)$$

The second line uses the fact that  $\Pr [x_i^4 > \epsilon^2] \leq \frac{\mathbf{E}[\exp(\alpha |x_i|)]}{\exp(\alpha \epsilon^{1/2})} \leq 2 \frac{\exp(\gamma \alpha^2/2)}{\exp(\alpha \epsilon^{1/2})} = 2 \exp\left(-\frac{\epsilon}{2\gamma}\right)$  with  $\alpha = \frac{\epsilon^{1/2}}{\gamma}$ . Since the above inequality holds for  $i = 1, 2, \dots, n$ ,

$$\mathbf{E} \left[ \exp \left( \eta \sum_{i=1}^n (x_i^4 - \mathbf{E}_i [x_i^4]) \right) \right] = \prod_{i=1}^n \mathbf{E}_i [\exp(\eta (x_i^4 - \mathbf{E}_i [x_i^4]))] \quad (123)$$

$$\leq \exp \left( 16n\eta\gamma^2 \left( \frac{1}{\sigma^2} - 1 \right) \right) \quad (124)$$

With Chernoff's inequality, for  $0 \leq \eta < \frac{1}{2\epsilon\gamma}$  and  $\epsilon \geq 0$ ,

$$\Pr \left[ \frac{1}{n} \sum_{i=1}^n (x_i^4 - \mathbf{E}_i [x_i^4]) > \epsilon \right] \leq \exp \left( -\eta n \epsilon + 16n\eta\gamma^2 \left( \frac{1}{\sigma^2} - 1 \right) \right). \quad (125)$$

Setting  $\eta = \frac{1-\sigma}{2\gamma\epsilon}$  and  $\sigma = 1 - \frac{4\gamma t}{n}$ , for  $\frac{\gamma t}{n} < \frac{1}{4}$ , we get the first inequality. For  $\eta < 0$  and  $\epsilon \geq 0$ ,

$$\Pr \left[ \frac{1}{n} \sum_{i=1}^n (x_i^4 - \mathbf{E}_i [x_i^4]) < -\epsilon \right] \leq \exp \left( \eta n \epsilon + 16n\eta\gamma^2 \left( \frac{1}{\sigma^2} - 1 \right) \right). \quad (126)$$

Setting  $\sigma = 1 + \gamma t$  gives the claimed inequality. ■

**Lemma 13** (Fourth-order normal random vectors). Let  $y_1, y_2, \dots, y_n \in \mathbb{R}^d$  be i.i.d.  $N(0, I)$  random vectors. For  $\epsilon_0 \in (0, 1/4)$  and  $\delta \in (0, 1)$ ,

$$\Pr \left[ \left\| \frac{1}{n} \sum_{i=1}^n y_i \otimes y_i \otimes y_i \otimes y_i - \mathbf{E}[\epsilon \otimes \epsilon \otimes \epsilon \otimes \epsilon] \right\|_2 > \frac{1}{1 - 4\epsilon_0} \epsilon_{\epsilon_0, t, n} \right] \leq 2\delta \quad (127)$$

where

$$\epsilon_{\epsilon_0, t, n} = \sqrt{\frac{2048 \ln((1 + 2/\epsilon_0)^d / \delta)^2}{n^2} + \frac{8 \ln((1 + 2/\epsilon_0)^d / \delta)^3}{n^3}} \quad (128)$$

**Proof** We follow the approach of [22]. Let  $Y := \frac{1}{n} \sum_{i=1}^n y_i \otimes y_i \otimes y_i \otimes y_i - \mathbf{E}[\epsilon \otimes \epsilon \otimes \epsilon \otimes \epsilon]$ . By [28], there exists  $Q \subseteq \mathcal{S}^{d-1} := \{\alpha \in \mathbb{R}^d : \|\alpha\|_2 = 1\}$  with cardinality at most  $(1 + 2\epsilon)^d$  such that  $\forall \alpha \in \mathcal{S}^{d-1} \exists q \in Q$   $\|\alpha - q\|_2 \leq \epsilon_0$ . Since, for any  $q \in Q$ ,  $y_i^\top q$  is distributed as  $N(0, 1)$ , with union bounds and Lemma 12, for

,  $\Pr [\exists q \in Q |T(Y, q, q, q, q)| > \epsilon_{\epsilon_0, t, n}] \leq 2\delta$ . So we assume with probability greater than  $1 - 2\delta$ ,  $\forall q \in Q$ ,  $|T(Y, q, q, q, q)| \leq \epsilon_{\epsilon_0, t, n}$ . Let  $\alpha_0 = \operatorname{argmax}_{\alpha \in \mathcal{S}^{d-1}} |T(Y, \alpha, \alpha, \alpha, \alpha)|$ , we have

$$\|Y\|_2 = |T(Y, \alpha_0, \alpha_0, \alpha_0, \alpha_0)| \quad (129)$$

$$\begin{aligned} &\leq \min_{q \in Q} |T(Y, q, q, q, q)| + |T(Y, \alpha_0 - q, q, q, q)| + |T(Y, \alpha_0, \alpha_0 - q, q, q)| \\ &\quad + |T(Y, \alpha_0, \alpha_0, \alpha_0 - q, q)| + |T(Y, \alpha_0, \alpha_0, \alpha_0, \alpha_0 - q)| \end{aligned} \quad (130)$$

$$\leq \min_{q \in Q} |T(Y, q, q, q, q)| + 4 \|\alpha_0 - q\| \|Y\|_2 \quad (131)$$

$$\leq \epsilon_{\epsilon_0, t, n} + 4\epsilon_0 \|Y\|_2, \quad (132)$$

which yields

$$\|Y\|_2 \leq \frac{1}{1 - 4\epsilon_0} \cdot \epsilon_{\epsilon_0, t, n} \quad (133)$$

■