On the optimality of kernels for high dimensional clustering - supplementary

February 28, 2020

1 Notation and Preliminaries

For any $d \in \mathbb{N}$, Let $g : \{(i,d)\}_{i\in[m],d\in[p]} \to [mp]$ be an injective mapping. For any (i,d), for ease of notation, we simply refer to $g(i,d)$ as id when it occurs as an index. For any $i \in [m]$, $d \in [p]$, x_{id} refers to the d^{th} component of x_i. For any two tensors x, y $x \otimes y$ refers to the outer product between x, y.

2 Definitions

2.1 $\alpha-$ sub-Exponential random variables

Definition 1 ($\alpha-$ **sub-Exponential random variables** (Götze, Sambale, and Sinulis 2019))**.** A centered random variable X is said to be $\alpha-$ sub-Exponential if there exist two constants c, C and some $\alpha > 0$ such that for all $t \geq 0$,

$$\Pr(|X| \geq t) \leq c \exp\left(-\frac{t^\alpha}{C}\right)$$

The corresponding $\alpha-$ sub-Exponential norm of X is given by:

$$\|X\|_{\psi_\alpha} = \inf\left\{t > 0 : \mathbb{E}\exp\left(\frac{|X|^\alpha}{t^\alpha}\right) \leq 2\right\}$$

$\alpha-$ sub-Exponential random variables with $\alpha = 2$ are referred to as sub-Gaussian random variables. Random variables with $\alpha = 1$ sub-Exponential decay are referred to simply as sub-Exponential random variables.

Definition 2 (**Tensor norms** (Götze, Sambale, and Sinulis 2019))**.** For any d^{th} order, symmetric tensor $A \in \mathbb{R}^{n^d}$, let $\mathcal{J} = \{J_1, J_2, ..., J_k\}$ be any partition of $[d]$. Then for any $x = x^1 \otimes x^2 \otimes \cdots \otimes x^k$, where $x^i \in \mathbb{R}^{n^{|J_i|}}$:

$$\|A\|_{\mathcal{J}} := \sup\left\{\sum_{i_1,\dots,i_d} a_{i_1\dots i_d} \prod_{j=1}^{k} x^j_{\mathbf{i}_{J_j}} : \|x^j\|_2 \leq 1\right\}. \tag{1}$$

2.1.1 Properties of sub-Gaussian random variables:

Proposition 1 (**Sums of sub-Gaussian random variables** (Vershynin 2018))**.** *Let $\{X_1, X_2, ..., X_m\}$ be m independent, centered, sub-Gaussian random variables. Then $\sum_{i\in[m]} X_i$ is a sub-Gaussian random variable and,*

$$\|\sum_{i\in[m]} X_i\|_{\psi_2}^2 \leq C \sum_{i\in[m]} \|X_i\|_{\psi_2}^2.$$

Proposition 2 (**Products of sub-Gaussian random variables** (Vershynin 2018))**.** *Let X_1 and X_2 be sub-Gaussian random variables. Then $X_1 \cdot X_2$ is a sub-Exponential random variable and,*

$$\|X_1 \cdot X_2\|_{\psi_1} \leq \|X_1\|_{\psi_2} \cdot \|X_2\|_{\psi_2}.$$

Proposition 3 (**Squares of sub-Gaussian random variables**(Vershynin 2018))**.** *Let X be sub-Gaussian random variables. Then X_1^2 is a sub-Exponential random variable and,*

$$\|X_1^2\|_{\psi_1} = \|X_1\|_{\psi_2}^2.$$

3 Useful concentration results

Proposition 4 (**Bernstein's inequality** (Vershynin 2018))**.** *Let $\{X_1, X_2, ..., X_m\}$ be a set of independent, centered, sub-Exponential random variables. Then for any $t > 0$, we have:*

$$\Pr\left(|\sum_{i\in[m]} X_i| \geq t\right) \leq 2\exp\left(-C\min\left(\frac{t^2}{\sum_{i\in[m]}\|X_i\|_{\psi_1}^2}, \frac{t}{\max_{i\in[m]}\|X_i\|_{\psi_1}}\right)\right)$$

for some fixed constant $C > 0$.

Proposition 5 (**Tail bounds for chi-squared distributions** (Birgé 2001))**.** *The following lower and upper tail bounds hold for non-central chi-squared distributions : For any $t > 0$,*

$$\Pr\left(\chi_d^2\left(\mu^2\right) < d + \mu^2 - 2\sqrt{(d + 2\mu^2)\,t}\right) < \exp(-t) \tag{2}$$

$$\Pr\left(\chi_d^2\left(\mu^2\right) > d + \mu^2 + 2\sqrt{(d + 2\mu^2)\,t} + 2t\right) < \exp(-t) \tag{3}$$

Proposition 6 (**Polynomials of $\alpha-$ sub-Exponentials**(Götze, Sambale, and Sinulis 2019))**.** *Let $X_1, \ldots, X_n$ be a set of independent random variables satisfying $\|X_i\|_{\psi_2} \leq b$ for some $b > 0$. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a polynomial of total degree $D \in \mathbb{N}$. Then, for any $t > 0$,*

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq t) \leq 2\exp\Big(-\frac{1}{C_D}\min_{1\leq s\leq D}\min_{\mathcal{J}\in P_s}\Big(\frac{t}{b^s\|\mathbb{E}f^{(s)}(X)\|_{\mathcal{J}}}\Big)^{\frac{2}{|\mathcal{J}|}}\Big).$$

where, for any for any $s \in D$, $f^{(s)}$ denotes the symmetric s^{th} order tensor of its sth order partial derivatives and P_s denotes the set of all possible partitions of $[s]$.

4 Other useful results

Proposition 7 (**Grothendieck's inequality** (Grothendieck 1956))**.** *For any matrix* $A \in \mathbb{R}^{m\times m}$*,*

$$\sup_{\substack{X\succeq 0\\ diag(X)\leq 1}} |\langle X, A\rangle| \leq K_G\|A\|_{\infty\to 1}.$$

where $K_G \approx 1.783$ *is the Grothendieck's constant.*

5 Proofs of lemmas

By an application of Bernstein's inequality for sub-exponential random variables, followed by an union bound over all $s, s' \in [k]$, it can be verified that, with high probability,

$$\min_{s\in[k]} \|\mu_s\|^2 = p + O(\sqrt{p\log p}); \min_{s\neq s'\in[k]} \langle \mu_s, \mu_{s'}\rangle = \frac{-p}{k-1} + O(\sqrt{p\log p}) \quad (4)$$

Proof of Lemma 9. Let $X = \{x_{id}\}_{i\in[m],d\in[p]}$ and let $f(X) =$

$$\sum_{i\neq j\in[m]} y_i z_j(\langle x_i, x_j\rangle^2 - \frac{\rho^2}{p^2}\langle \mu_i, \mu_j\rangle^2 - p) + \sum_{i\in[m]} y_i z_i(\|x_i\|^2 - (p + \frac{\rho}{p}\|\mu_i\|^2)^2 - p).$$

Since f is a 4^{th} order polynomial in X, from Proposition 6, for any $t > 0$:

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq t) \leq 2\exp\Big(-\frac{1}{C_4}\min_{1\leq s\leq 4}\min_{\mathcal{J}\in P_s}\Big(\frac{t}{b^s\|\mathbb{E}f^{(s)}(X)\|_{\mathcal{J}}}\Big)^{\frac{2}{|\mathcal{J}|}}\Big).$$

We have $\mathbb{E}f(X) = C_f(m^2\rho + m)$ for some constant $C_f > 0$.
We need the following tensors and their corresponding tensor norms to establish the results of Lemma 9 via an application of Proposition 6.

- For each $s \in [4]$, s^{th} order tensors A_s of expectations of s^{th} order derivatives of f with respect to each $\{x_{i_1,d_1}, ..., x_{i_s,d_s}\}_{i_j\in[m],d_j\in[p]}$.
- Tensor norms for A_s with respect to each $\mathcal{J}$ in P_s - the set of all possible partitions of [s].

Computing A^1**:** The first order derivative of f with respect to x_{id} for any $i \in [m]$ and $d \in [p]$: $\frac{\partial f(X)}{\partial x_{id}} =$

$$4x_{id}^3 y_i z_i + \sum_{d'\neq d} 2x_{id}x_{id'}^2 y_i z_i + \sum_{j\neq i} 2x_{id}x_{jd}^2 y_i z_j + \sum_{d'\neq d}\sum_{j\neq i} x_{id'}x_{jd}x_{jd'}y_i z_j. \quad (5)$$

$$\mathbb{E}(\frac{\partial f(X)}{\partial x_{id}}) = O(\sqrt{p\log p}). \quad (6)$$

Therefore, $A^1 = O(\sqrt{p\log p})\mathbb{J}_{mp}$, where $\mathbb{J}_{mp} \in \mathbb{R}^{mp}$ denotes the vector of ones.

Tensor norms of A^1. $P_1 = \{1\}$ and

$$\|A^1\|_{\{1\}} = \sup\left\{ \sum_{1\in[m],d\in[p]} A^1_{id}x^1_{id} : \|x^1\|_2 \le 1 \right\} = \|A^1\|_2 = O(p\sqrt{m\log p}).$$

All the inequalities in this proof are obtained from multiple applications of Hölder's inequality with $p = 1$ and $q = \infty$, CauchySchwarz inequality and the inequality: for any $x \in \mathbb{R}^n$, $\|x\|_1 \le \sqrt{n}\,\|x\|_2$.
Computing A^2: The second order derivative of f with respect to $x_{id}, x_{k\beta}$ for any $i, k \in [m]$ and $d, \beta \in [p]$: $\frac{\partial f(X)}{\partial x_{id}\partial x_{k\beta}} =$

$$\begin{cases} 12x_{id}^2 y_i z_i + \sum_{d'\neq d} 2x_{id'}^2 y_i z_i + \sum_{j\neq i} 2x_{jd}^2 y_j z_j & \text{if } k = i; \beta = d, \\ 4x_{id}x_{i\beta}y_i z_i + \sum_{j\neq i} 2x_{jd}x_{j\beta}y_i z_j & \text{if } k = i; \beta \neq d, \\ 4x_{id}x_{kd}y_i z_k + \sum_{d'\neq d} x_{id'}x_{kd'}y_i z_k & \text{if } k \neq i; \beta = d, \\ x_{i\beta}x_{kd}y_i z_k & \text{otherwise .} \end{cases} \tag{7}$$

Then $A^2(i,j) =$

$$\begin{cases} O(p) & \text{if } k = i; \beta = d, \\ O(\log p) & \text{if } k = i; \beta \neq d, \\ O(1) & \text{if } k \neq i; \beta = d, \\ O(\log p/p) & \text{otherwise .} \end{cases} \tag{8}$$

Tensor norms of A^2
$P_2 = \{\{1,2\}, \{\{1\}\,\{2\}\}\}$.
From the definition, its clear that $A^2_{\{1,2\}} = \|A\|_2 = O(p^2)$.

$$A^2_{\{\{1\}\{2\}\}} = \sup\left\{ \sum_{i,j\in[m],d,d'\in[p]} A^2_{id,jd'}x^1_{id}x^2_{jd'} : \|x^1\|_2, \|x^2\|_2 \le 1 \right\}$$

$$\begin{aligned}
&\le \sup_{\forall l, \|x^l\|_2\le 1} \Big\{ \sum_{i\in[m]}\sum_{d\in p} O(p)|x^1_{id}x^2_{id}| + \sum_{i\in[m]}\sum_{d\neq d'\in p} O(\log p)|x^1_{id}x^2_{id'}| \\
&\qquad + \sum_{i\neq j\in[m]}\sum_{d\in p} O(1)|x^1_{id}x^2_{jd}| + \sum_{i\neq j\in[m]}\sum_{d\neq d'\in p} \rho O(\frac{\log p}{p})|x^1_{id}x^2_{jd'}| \Big\} \\
&\le \sup_{\forall l, \|x^l\|_2\le 1} \Big\{ O(p)\|x^1\|_2\|x^2\|_2 + O(\log p)\sum_{d\neq d'\in p}\sqrt{\sum_{i\in[m]}(x^1_{id})^2 \sum_{i\in[m]}(x^2_{id'})^2} \\
&\quad + \sum_{i\neq j\in[m]}\sum_{d\in p} O(1)\sqrt{\sum_{d\in[p]}(x^1_{id})^2\sum_{d\in[p]}(x^2_{jd})^2} + \rho O(\frac{\log p}{p})(\sqrt{mp}\|x^1\|_2)(\sqrt{mp}\|x^2\|_2) \Big\} \\
&\le \sup_{\forall l, \|x^l\|_2\le 1} \Big\{ O(p)\|x^1\|_2\|x^2\|_2 + O(\log p)\sqrt{p}\|x^1\|_2\sqrt{p}\|x^2\|_2 \\
&\quad + O(1)\sqrt{m}\|x^1\|_2\sqrt{m}\|x^2\|_2 + \rho O(\frac{\log p}{p})(\sqrt{mp}\|x^1\|_2)(\sqrt{mp}\|x^2\|_2) \Big\} \\
&\le O(p\log p).
\end{aligned}$$

Computing A^3: The third order derivative of f with respect to $x_{id}, x_{k\beta}, x_{\alpha l}$ for any $i, k \in [m]$ and $d, \beta \in [p]$: $\frac{\partial f(X)}{\partial x_{id} \partial x_{k\beta} \partial x_{\alpha l}} =$

$$\begin{cases} 24 x_{id} y_i z_i & \text{if } \alpha = k = i; l = \beta = d, \\ 4 x_{il} y_i z_i & \text{if } \alpha = k = i; l \neq \beta = d, \\ 4 x_{\alpha d} y_i z_\alpha & \text{if } \alpha \neq k = i; l = \beta = d, \\ x_{\alpha\beta} y_i z_\alpha & \text{if } \alpha \neq k = i; l = d \neq \beta, \\ 0 & \text{otherwise} . \end{cases} \tag{9}$$

Then $A^3_{id,k\beta,\alpha l} =$

$$\begin{cases} O(\frac{\log p}{p}) y_i z_i & \text{if } \alpha = k = i; l = \beta = d, \\ O(\frac{\log p}{p}) y_i z_i & \text{if } \alpha = k = i; l \neq \beta = d, \\ O(\frac{\log p}{p}) y_i z_\alpha & \text{if } \alpha \neq k = i; l = \beta = d, \\ O(\frac{\log p}{p}) y_i z_\alpha & \text{if } \alpha \neq k = i; l = d \neq \beta, \\ 0 & \text{otherwise} . \end{cases} \tag{10}$$

Tensor norms of A^3: The set of all possible partitions of [3] up to symmetries is $P_3 =$

$$\{\{\{1,2,3\}\}, \{\{1\}, \{2\}, \{3\}\}, \{\{1\}, \{2,3\}\}\}$$

.
Computing $\|A^3\|_{\{\{1,2,3\}\}}$: It follows from the definition that $\|A^3\|_{\{\{1,2,3\}\}} = \|A^3\|_2 = O(m\sqrt{p \log p})$.
Computing $\|A^3\|_{\{\{1\},\{2\},\{3\}\}}$:

$$\begin{aligned}
&= \sup \left\{ \sum_{i_1,i_2,i_3} a_{i_1,i_2,i_3} x^1_{i_1} x^2_{i_2} x^2_{i_3} : \forall l, \|x^l\|_2 \leq 1 \right\} \\
&\leq \sup_{\forall l, \|x^l\|_2 \leq 1} \left\{ \sum_{i,j \in [m]} \sum_{d,d' \in [p]} |x^1_{id}||x^2_{id'}||x^3_{jd}| \right\} \cdot O\left(\sqrt{\frac{\log p}{p}} \right) \\
&\leq \sup_{\forall l, \|x^l\|_2 \leq 1} \left\{ \sum_{i,j \in [m]} \sqrt{\sum_{d \in [p]} (x^1_{id})^2} \sqrt{\sum_{d \in [p]} (x^2_{jd})^2} \sqrt{p} \sqrt{\sum_{d \in [p]} (x^3_{id'})^2} \right\} \cdot O\left(\sqrt{\frac{\log p}{p}} \right) \\
&\leq \sup_{\forall l, \|x^l\|_2 \leq 1} \left\{ \sqrt{mp} \|x^1\|_2 \|x^2\|_2 \|x^3\|_2 \right\} \cdot O\left(\sqrt{\frac{\log p}{p}} \right) \\
&\leq O(\sqrt{p \log p}).
\end{aligned}$$

Computing $\|A^3\|_{\{\{1,2\},\{3\}\}}$:

$$
\begin{aligned}
&= \sup\left\{ \sum_{i_1,i_2,i_3} a_{i_1,i_2,i_3} x^1_{i_1} x^2_{i_2,i_3} : \forall l, \|x^l\|_2 \le 1 \right\} \\
&\le \sup_{\forall l, \|x^l\|_2 \le 1} \left\{ \sum_{i,j\in[m]} \sum_{d,d'\in[p]} |x^1_{id}||x^2_{id',jd}| \right\} \cdot \|A^3\|_\infty \\
&\le \sup_{\forall l, \|x^l\|_2 \le 1} \left\{ \sum_{i,j\in[m]} \sqrt{\sum_{d\in[p]} (x^1_{id})^2} \sqrt{p} \sqrt{\sum_{d'\in[p]} (x^2_{id',jd})^2} \right\} \cdot O\left(\sqrt{\frac{\log p}{p}}\right) \\
&\le \sup_{\forall l, \|x^l\|_2 \le 1} \left\{ \sqrt{mp} \|x^1\|_2 \|x^2\|_2 \right\} \cdot O\left(\sqrt{\frac{\log p}{p}}\right) \\
&\le O(\sqrt{p \log p}).
\end{aligned}
$$

Computing A^4: The fourth order derivative of f with respect to $x_{id}, x_{k\beta}, x_{\alpha l}, x_{q\gamma}$ for any $i, k, \alpha, q \in [m]$ and $d, \beta, l, \gamma \in [p]$: $\frac{\partial f(X)}{\partial x_{id} \partial x_{k\beta} \partial x_{\alpha l} \partial x_{q\gamma}} =$

$$
\begin{cases}
24 y_i z_i & \text{if } q = \alpha = k = i; \gamma = l = \beta = d, \\
4 y_i z_i & \text{if } q = \alpha = k = i; \gamma = l \neq \beta = d, \\
4 y_i z_\alpha & \text{if } q = \alpha \neq k = i; \gamma = l = \beta = d, \\
y_i z_\alpha & \text{if } q = \alpha \neq k = i; l = d \neq \beta = \gamma, \\
0 & \text{otherwise .}
\end{cases} \tag{11}
$$

Tensor norms of A^4: The list of all possible partitions of [4] is the following(up to symmetries). $P_{[4]} =$
$\{\{\{1,2,3,4\}\}, \{\{1\},\{2\},\{3\},\{4\}\}, \{\{1,2\},\{3,4\}\}, \{\{1\},\{2,3,4\}\}, \{\{1\},\{2\},\{3,4\}\}\}$.
Computation of $\|A^4\|_{\{\{1,2,3,4\}\}}$:
Its clear from the definition that $\|A^4\|_{\{\{1,2,3,4\}\}} = \|A\|_2 \le 24mp$.
Computation of $\|A^4\|_{\{\{1\},\{2\},\{3\},\{4\}\}}$:

$$
\begin{aligned}
&= \sup\left\{ \sum_{i_1,i_2,i_3,i_4} a_{i_1,i_2,i_3,i_4} x^1_{i_1} x^2_{i_2} x^2_{i_3} x^4_{i_4} : \forall l, \|x^l\|_2 \le 1 \right\} \\
&= \sup\left\{ \sum_{i,j\in[m]} \sum_{d,d'\in[p]} A^4_{id,id',jd,jd'} x^1_{id} x^2_{id'} x^3_{jd} x^4_{jd'} : \forall l, \|x^l\|_2 \le 1 \right\} \\
&\le \sup\Bigg\{ \sum_{i\in[m]} \left(\sqrt{\sum_{d\in[p]} (x^1_{id})^2 \sum_{d'\in[p]} (x^2_{id'})^2} \right) \sum_{j\in[m]} \left(\sqrt{\sum_{d\in[p]} (x^3_{jd})^2 \sum_{d'\in[p]} (x^4_{jd'})^2} \right) \\
&\qquad\qquad : \forall l, \|x^l\|_2 \le 1 \Bigg\} \|A^4\|_\infty. \\
&\le \|A^4\|_\infty = 24.
\end{aligned}
$$

Note that the first inequality follows from a simultaneous application of the Holder's inequality with $p = 1$ and $q = \infty$ and the Cauchy schwarz inequality. The last step follows from an application of the Cauchy-Schwarz inequality.

Computation of $A^4_{\{\{1,2\},\{3,4\}\}}$ **:**

$$
\begin{aligned}
&= \sup\left\{ \sum_{i_1,i_2,i_3,i_4} a_{i_1,i_2,i_3,i_4} x^1_{i_1,i_2} x^2_{i_3,i_4} : \forall l, \|x^l\|_2 \le 1 \right\} \\
&= \sup_{\forall l, \|x^l\|_2 \le 1} \left\{ \sum_{i,j\in[m]} \sum_{d,d'\in[p]} A^4_{id,id',jd,jd'} \ \left(x^1_{id,id'} x^2_{jd,jd'} + x^1_{id,jd} x^2_{id',jd'} + x^1_{id,jd'} x^2_{id',jd} \right.\right. \\
&\qquad\qquad \left.\left. + x^1_{id,id'} x^2_{jd',jd} + x^1_{id,jd} x^2_{jd',id'} + x^1_{id,jd'} x^2_{jd,id'} \right) \right\} \\
&\le \sup_{\forall l, \|x^l\|_2 \le 1} \left\{ \sum_{i,j\in[m]} \sqrt{\sum_{d,d'\in[p]} (x^1_{id,id'})^2 \sum_{d,d'\in[p]} (x^2_{jd,jd'})^2} + \right. \\
&\qquad \sum_{d,d'\in[p]} \sqrt{\sum_{i,j\in[m]} (x^1_{id,jd})^2 \sum_{i,j\in[m]} (x^2_{id',jd'})^2} + \\
&\qquad \left. \sqrt{\sum_{i,j\in[m]} \sum_{d,d'\in[p]} (x^1_{id,jd'})^2 \sum_{i,j\in[m]} \sum_{d,d'\in[p]} (x^2_{id',jd})^2} \right\} 2\|A^4\|_\infty. \\
&\le 2\|A^4\|_\infty (m + p + 1) = 48(m + p + 1).
\end{aligned}
$$

Computation of $A^4_{\{\{1\},\{2,3,4\}\}}$**:**

$$
\begin{aligned}
&= \sup\left\{ \sum_{i_1,i_2,i_3,i_4} a_{i_1,i_2,i_3,i_4} x^1_{i_1} x^2_{i_2,i_3,i_4} : \forall l, \|x^l\|_2 \le 1 \right\} \\
&\le \sup_{\forall l, \|x^l\|_2 \le 1} \left\{ \sum_{i,j\in[m]} \sum_{d,d'\in[p]} A^4_{id,id',jd,jd'} \ \left(x^1_{id} x^2_{id',jd,jd'} + x^1_{id} x^2_{id',jd',jd} + x^1_{id} x^2_{id',jd,jd'} \right.\right. \\
&\qquad\qquad \left.\left. + x^1_{id} x^2_{jd',jd,id'} + x^1_{id} x^2_{jd,jd',id'} + x^1_{id} x^2_{jd',jd,id'} \right) \right\} \cdot \|A^4\|_\infty, \\
&\le \sup_{\forall l, \|x^l\|_2 \le 1} \left\{ \sum_{i\in[m]} \sum_{d\in[p]} |x^1_{id}| |\sum_{d'\in[p]} x^2_{id'jdjd'}| \right\} \cdot 6\|A^4\|_\infty, \\
&\le \sup_{\forall l, \|x^l\|_2 \le 1} \left\{ \sqrt{p} \sum_{i\in[m]} \sqrt{\sum_{d\in[p]} (x^1_{id})^2} \left(\sum_{j\in[m]} \sqrt{\sum_{d,d'\in[p]} (x^2_{id'jdjd'})^2} \right) \right\} \cdot 6\|A^4\|_\infty, \\
&\le \sup_{\forall l, \|x^l\|_2 \le 1} \left\{ \sqrt{mp} \sum_{i\in[m]} \sqrt{\sum_{d\in[p]} (x^1_{id})^2} \left(\sqrt{\sum_{j\in[m]} \sum_{d,d'\in[p]} (x^2_{id'jdjd'})^2} \right) \right\} \cdot 6\|A^4\|_\infty, \\
&\le \sup_{\forall l, \|x^l\|_2 \le 1} \left\{ \sqrt{mp} \sqrt{\sum_{i\in[m]} \sum_{d\in[p]} (x^1_{id})^2} \left(\sqrt{\sum_{i,j\in[m]} \sum_{d,d'\in[p]} (x^2_{id'jdjd'})^2} \right) \right\} \cdot 6\|A^4\|_\infty \\
&\le 6\|A^4\|_\infty (\sqrt{mp}) = 144\sqrt{mp}.
\end{aligned}
$$

Computation of $A^4_{\{\{1\},\{2\},\{3,4\}\}}$:

$$= \sup\left\{ \sum_{i_1,i_2,i_3,i_4} a_{i_1,i_2,i_3,i_4} \cdot x^1_{i_1}, x^2_{i_2}, x^3_{i_3,i_4} : \forall l, \|x^l\|_2 \leq 1 \right\}$$

$$= \sup_{\forall l, \|x^l\|_2\leq 1} \left\{ \sum_{i,j\in[m]} \sum_{d,d'\in[p]} A^4_{id,id',jd,jd'} \;\; (x^1_{id}, x^2_{id'} x^3_{jd,jd'} + x^1_{id}, x^2_{jd} x^3_{id',jd'} + x^1_{id}, x^2_{jd'} x^3_{id',jd} \right.$$

$$\left. + x^1_{id}, x^2_{id'} x^3_{jd',jd} + x^1_{id}, x^2_{jd} x^3_{jd',id'} + x^1_{id}, x^2_{jd'} x^3_{jd,id'}) \right\}$$

$$\leq \sup_{\forall l, \|x^l\|_2\leq 1} \left\{ \sum_{i,j\in[m]} \sqrt{\sum_{d,d'\in[p]} (x^1_{id} x^2_{id'})^2 \sum_{d,d'\in[p]} (x^3_{jd,jd'})^2} + \right.$$

$$\sum_{d,d'\in[p]} \sqrt{\sum_{i,j\in[m]} (x^1_{id} x^2_{jd})^2 \sum_{i,j\in[m]} (x^3_{id',jd'})^2}$$

$$\left. + \sqrt{\sum_{i,j\in[m]} \sum_{d,d'\in[p]} (x^1_{id} x^2_{jd'})^2 \sum_{i,j\in[m]} \sum_{d,d'\in[p]} (x^3_{id',jd})^2} \right\} \cdot 2\|A^4\|_\infty.$$

$$\leq 2\|A^4\|_\infty (m+p+1) = 48(m+p+1).$$

□

Gathering all the norms, we have that for any fixed $y, z \in \{\pm 1\}^m$:

$$\mathbb{P}(f(X) \geq C_f(m^2\rho + m) + t) \leq 2\exp\left(-\frac{1}{C} \min\left(\left(\frac{t}{24mp}\right)^2, \left(\frac{t}{24}\right)^{\frac{1}{2}}, \left(\frac{t}{4(m+p+1)}\right), \right.\right.$$

$$\left.\left. \left(\frac{t}{\sqrt{mp}}\right), \left(\frac{t}{4(m+p+1)}\right)^{\frac{2}{3}}, \left(\frac{t}{p\sqrt{p\log p}}\right)^2, \left(\frac{t}{p^2}\right), \left(\frac{t}{m\sqrt{p\log p}}\right)^2, \left(\frac{t}{\sqrt{p\log p}}\right)^{\frac{2}{3}}, \left(\frac{t}{\sqrt{mp}}\right) \right) \right). \tag{12}$$

Applying a union bound over all possible $y, z \in \{\pm 1\}^m$ and the setting the R.H.S of Equation 12 to $\exp(-(1+\epsilon)m\log 2)$, for some arbitrarily small constant $\epsilon > 0$ we have that w.h.p,

$$\sup_{\{z,y\in\pm 1\}^m} \kappa \sum_{i,j=1}^m y_i z_j R^{(2)}_{i,j} \leq$$

$$\frac{C_2\kappa}{p^2}(\rho m(m-1) + m + (mp\sqrt{m} \;\vee\; m^2\sqrt{m} \;\vee\; p^2\sqrt{m})). \tag{13}$$

for some constant $C_2 > 0$.

Proof of Lemma 8. For any fixed $y, z \in \{\pm 1\}^m$,

$$\sum_{i,j\in[m]} y_i z_j \left(\langle x_i, x_j\rangle - \mathbb{E}\langle x_i, x_j\rangle\right) =$$
$$\sum_{d\in[p]} \left[\left(\sum_{i\in[m]} y_i x_{id}\right)\left(\sum_{j\in[m]} z_j x_{jd}\right) - \mathbb{E}\left(\sum_{i\in[m]} y_i x_{id}\right)\left(\sum_{j\in[m]} z_j x_{jd}\right)\right]. \quad (14)$$

Since each x_{id} is a normally distributed random variable, $\sum_{i\in[m]} y_i x_{id}$ is a sub-Gaussian random variable with

$$\|\sum_{i\in[m]} y_i x_{id}\|_{\psi_2} \le \sqrt{m}\|x_{id}\|_{\psi_2} \le \sqrt{m}(1 + O(\log p/p)).$$

Therefore, for each $d \in [p]$, $\left(\sum_{i\in[m]} y_i x_{id}\right)\left(\sum_{j\in[m]} z_j x_{jd}\right)$ is a sub-exponential random variable with sub-exponential norm:

$$\|(\sum_{i\in[m]} y_i x_{id})(\sum_{j\in[m]} z_j x_{jd})\|_{\psi_1} \le \|\sum_{i\in[m]} y_i x_{id}\|_{\psi_2}\|\sum_{j\in[m]} z_j x_{jd}\|_{\psi_2} \le m(1+O(\log p/p))^2.$$

Applying Bernstein's inequality for sums of independent sub-exponential random variables, we have that $\forall t > 0$,

$$\Pr\left(\sum_{d\in[p]}\left[(\sum_{i\in[m]} y_i x_{id})(\sum_{j\in[m]} z_j x_{jd}) - \mathbb{E}(\sum_{i\in[m]} y_i x_{id})(\sum_{j\in[m]} z_j x_{jd})\right] > t\right) \le$$
$$\exp\left(-c\min\left(\frac{t^2}{pm^2(1+O(\log p/p))^4}, \frac{t}{m(1+O(\log p/p))^2}\right)\right). \quad (15)$$

Applying a union bound over all possible partitions $y, z \in \{\pm 1\}^m$, we can see that w.h.p

$$\sup_{y,z\in\{\pm 1\}^m} \sum_{i,j\in[m]} y_i z_j \left(\langle x_i, x_j\rangle - \mathbb{E}\langle x_i, x_j\rangle\right) \le C_1\left(\frac{m^2}{\sqrt{\alpha}} \vee m^2\right) \quad (16)$$

for some constant $C_1 > 0$.

$\square$

Proof of Lemma 1. Since for each $i \in [m]$ and each $d \in [p]$, x_{id} is a normally distributed random variable, for each $i, j \in [m]$, $x_{id}x_{jd}$ is a sub-exponential random variable with:

$$\|x_{id}x_{jd}\|_{\psi_1} \le \|x_{id}\|_{\psi_2}\|x_{jd}\|_{\psi_2} \le (1 + O(\log p/p))^2.$$

From an application of Bernstein's inequality for sub-exponential random variables, we have that:

$$\Pr\left(|\langle x_i, x_j\rangle - \mathbb{E}\langle x_i, x_j\rangle|\right) > t \le 2\exp\left(-c\left(\frac{t^2}{p(1+O(\log p/p))^4} \wedge \frac{t}{(1+O(\log p/p))^2}\right)\right).$$

Taking a union bound over all $i, j \in [m]$, we have that:

$$\max_{i,j\in[m]} \langle x_i, x_j\rangle \le \mathbb{E}\langle x_i, x_j\rangle + O(\sqrt{p\log p}).$$

and

$$\min_{i,j\in[m]} \langle x_i, x_j\rangle \ge \mathbb{E}\langle x_i, x_j\rangle - O(\sqrt{p\log p}).$$

Since $\forall i \neq j$, $\mathbb{E}\langle x_i, x_j\rangle = O(1)$, we have that w.h.p:

$$\max_{i\neq j\in[m]} \frac{\langle x_i, x_j\rangle}{p} \le O(\log p/\sqrt{p}), \ \min_{i\neq j\in[m]} \frac{\langle x_i, x_j\rangle}{p} \ge -O(\sqrt{\log p/p})$$

and $\forall i \in [m]$, $\mathbb{E}\|x_i\|^2 = p + O(1)$. So,

$$\max_{i\in[m]} \frac{\|x_i\|}{p} \le 1 + O(\log p/\sqrt{p}), \ \min_{i\in[m]} \frac{\|x_i\|}{p} \ge 1 - O(\sqrt{\log p/p})$$

$\square$

Proof of Lemma 2. For any partition σ such that $\|\beta(\sigma,\sigma_*)\|_F^2 \le 1+(k-1)\epsilon$,

$$\frac{k}{m}\sum_{s=1}^{k}\sum_{\substack{\sigma(i)=s\\ \sigma(j)=s}} \langle x_i, x_j\rangle = \frac{k}{m}\sum_{s=1}^{k}\left\|\sum_{\sigma(i)=s} x_i\right\|^2.$$

$\sum\limits_{\sigma(i)=s} x_i$ is the sum of independent normally distributed random variable and is also normally distributed. Therefore, $\|\sum\limits_{\sigma(i)=s} x_i\|^2$ follows a non central chi-square distribution with non-centrality:

$$\frac{\alpha\rho}{k}\sum_{s'\in[k]}\sum_{s,t\in[k]} \beta_{s,s'}\beta_{t,s'}\langle\mu_s,\mu_t\rangle = \frac{p\alpha\rho}{k-1}(\|\beta\|_F^2 - 1) + O(\sqrt{p\log p})$$

and pk degrees of freedom. Applying upper tail bounds from proposition 5, followed a union bound over all such partitions and setting $t = (1+\epsilon)m\log k$, we obtain the following inequality which holds with high probability:

$$\max_{\substack{\sigma:\|\beta(\sigma,\sigma_*)\|_F^2\\ \le 1+(k-1)\epsilon}} \frac{k}{m}\sum_{s=1}^{k}\sum_{\substack{\sigma(i)=s\\ \sigma(j)=s}} Q_{i,j}^{1\sigma} \le k + \alpha\rho\epsilon + 2(1+\epsilon)\alpha\log k + 2\sqrt{(1+\epsilon)\,(k+2\alpha\rho\epsilon)\,\alpha\log k} + O(\sqrt{\log p/p}). \quad (17)$$

Similarly, the random variable $\sum\limits_{i=1}^{m}\sum\limits_{d=1}^{p} x_{id}^2$ is distributed according to a non-central chi-squared distribution with non-centrality(μ^2) $p\alpha\rho$ and mp degrees of freedom(d). Note that it is independent of the partition. Using the lower tail bounds from proposition 5 and setting $t = \log(p)$, w.p.a.l $(1 - \frac{1}{p})$.

$$\max_{\substack{\sigma:\|\beta(\sigma,\sigma_*)\|_F^2\\ \le 1+(k-1)\epsilon}} -\gamma_{\max}Q_i^5 \le -\frac{k\gamma_{\max}\tau}{mp}\left(mp + p\alpha\rho - 2\sqrt{(mp+2p\alpha\rho)\log p}\right). \quad (18)$$

$\square$

Proof of Lemma 4. Using the inequality, $\sum_{i=1}^{n} a_i \cdot b_i \leq \sup_{i\in[n]} |b_i| \cdot \sum_{i=1}^{n} |a_i|$, we have:

$$\frac{k}{m}\sum_{i\in[m]}(\frac{\|x_i\|^2}{p}-\tau)^2 \leq k\max_{i\in[m]}(\frac{\|x_i\|^2}{p}-\tau)^2 \leq kO(\frac{\log p}{p}).$$

Therefore,

$$\max_{\substack{\sigma:\|\beta(\sigma,\sigma_*)\|_F^2\\ \leq 1+(k-1)\epsilon}} \sum_{i\in[m]} \frac{k\gamma_{\max}(e^{\tau}-1)}{2m} Q_i^{4\sigma} \leq C_0 k\gamma_{\max}(e^{\tau}-1)(\log p)^2/2p. \tag{19}$$

$\square$

Proof of Lemma 5. For the true partition σ^*, $\|\sum_{\sigma^*(i)=s} x_i\|^2$ follows a non central chi-square distribution with non-centrality:

$$p\alpha\rho + O(\sqrt{p\log p})$$

and pk degrees of freedom. Applying lower tail bounds from proposition 5 and setting $t = \log p$, we obtain the following inequality which holds with high probability:

$$\frac{k}{m}\sum_{s=1}^{k}\sum_{\substack{\sigma^*(i)=s\\ \sigma^*(j)=s}} Q_{i,j}^{1\sigma} \geq k + \alpha\rho - O(\sqrt{\log p/p}). \tag{20}$$

Similarly, as noted earlier, the random variable $\sum_{i=1}^{m}\sum_{d=1}^{p} x_{id}^2$ is distributed according to a non-central chi-squared distribution with non-centrality(μ^2) $p\alpha\rho$ and mp degrees of freedom(d). Using the upper tail bounds from proposition 5 and setting $t = \log(p)$, w.p.a.l $(1-\frac{1}{p})$:

$$-\gamma_{\min}Q^5 > -\frac{k\gamma_{\min}\tau}{mp}\left(mp + p\alpha\rho + 2\log p - 2\sqrt{(mp+2p\alpha\rho)\log p}\right). \tag{21}$$

$\square$

Proof of Lemma 7.

$$\tilde{K}(i,j) = f(0) + \begin{cases} \frac{f'(0)\rho\langle\mu_i,\mu_j\rangle}{p^2} + \frac{\kappa\rho^2\langle\mu_i,\mu_j\rangle^2}{p^4} + \frac{\kappa}{p} & \text{if } i \neq j \\ \frac{f'(0)(p^2+\rho\|\mu_i\|^2)}{p^2} + \frac{\kappa(p^2+\rho\|\mu_i\|^2)^2}{p^4} + \frac{\kappa}{p} & \text{otherwise.} \end{cases}$$

$$\begin{aligned} \langle\tilde{K}, X^*-\hat{X}\rangle &= \sum_{S}\sum_{i\neq j\in S}\tilde{K}_{i,j}(1-\hat{X}_{i,j}) + \sum_{S,S'}\sum_{\substack{i\in S\\ j\in S'}}\tilde{K}_{i,j}(-\hat{X}_{i,j}) \\ &\geq \min_{S}\min_{i\neq j\in S}\tilde{K}_{i,j}\sum_{S}\sum_{i\neq j\in S}(1-\hat{X}_{i,j}) - \max_{S,S'}\max_{\substack{i\in S\\ j\in S'}}\tilde{K}_{i,j}\sum_{\substack{i\in S\\ j\in S'}}\sum_{\substack{i\in S\\ j\in S'}}(\hat{X}_{i,j}) \\ &= (\min_{S}\min_{i\neq j\in S}\tilde{K}_{i,j} - \max_{S,S'}\max_{\substack{i\in S\\ j\in S'}}\tilde{K}_{i,j})\sum_{S}\sum_{i\neq j\in S}(1-\hat{X}_{i,j}). \end{aligned}$$

The last inequality is obtained using the property that the sum of entries of each row of $\hat{X}$ is equal to $\frac{m}{k}$. Similarly,

$$\begin{aligned}\|X^* - \hat{X}\|_1 &= \sum_S \sum_{i \neq j \in S} (1 - \hat{X}_{i,j}) + \sum_{S,S'} \sum_{\substack{i \in S \\ j \in S'}} (\hat{X}_{i,j}) \\ &\leq 2 \sum_S \sum_{i \neq j \in S} (1 - \hat{X}_{i,j}).\end{aligned}$$

Therefore,

$$\|X^* - \hat{X}\|_1 \leq \frac{2}{(\min_S \min_{i \neq j \in S} \tilde{K}_{i,j} - \max_{S,S'} \max_{\substack{i \in S \\ j \in S'}} \tilde{K}_{i,j})} \langle \tilde{K}, X^* - \hat{X} \rangle. \tag{22}$$

Substituting the values of $\tilde{K}_{i,j}$, we have that

$$\begin{aligned}\min_S \min_{i \neq j \in S} \tilde{K}_{i,j} &= f(0) + f'(0) \min_S \frac{\rho \|\mu_s\|^2}{p^2} + e^\tau \gamma_{\max} \min_S \frac{\rho^2 \|\mu_s\|^4}{p^4} \\ &= f(0) + f'(0) \frac{\rho(1 + O(\sqrt{\frac{\log p}{p}}))}{p} + e^\tau \gamma_{\max} \frac{\rho^2 (1 + O(\sqrt{\frac{\log p}{p}}))^2}{p^2}.\end{aligned}$$

$$\begin{aligned}\max_{S \neq S'} \max_{i \in S, j \in S'} \tilde{K}_{i,j} &= f(0) + f'(0) \max_{S \neq S'} \frac{\rho \langle \mu_s, \mu_s' \rangle}{p^2} + e^\tau \gamma_{\max} \max_{S \neq S'} \frac{\rho^2 \langle \mu_s, \mu_s' \rangle^2}{p^4} \\ &= f(0) + f'(0) \frac{\rho(\frac{-1}{k-1} + O(\sqrt{\frac{\log p}{p}}))}{p} + e^\tau \gamma_{\max} \frac{\rho^2 (\frac{-1}{k-1} + O(\sqrt{\frac{\log p}{p}}))^2}{p^2}.\end{aligned}$$

where, the second equalities for both quantities arise from substituting the values of $\min_s \|\mu_s\|^2$ and $\min_{s,s'} \langle \mu_s, \mu_{s'} \rangle$. Therefore,

$$\min_S \min_{i \neq j \in S} \tilde{K}_{i,j} - \max_{S \neq S'} \max_{i \in S, j \in S'} \tilde{K}_{i,j} = \frac{\rho}{p} \left(\frac{k}{k-1} + O(\sqrt{\frac{\log p}{p}}) + O(1/p) \right) \text{ and}$$

$$\|X^* - \hat{X}\|_1 \leq \frac{2 \langle \tilde{K}, X^* - \hat{X} \rangle}{\frac{\rho}{p} \left(\frac{k}{k-1} + O(\sqrt{\frac{\log p}{p}}) + O(1/p) \right)}.$$

□

Proofs of Lemma 3,6. $f_{\sigma_*}(X) = \sum_{s \in [k]} \sum_{i,j \in \sigma_*^{-1}(s)} \langle x_i, x_j \rangle^2$.

$$\mathbb{E} \sum_{s \in [k]} \sum_{i,j \in \sigma_*^{-1}(s)} \langle x_i, x_j \rangle^2 = \frac{mp^2}{k} (k + \frac{\alpha}{k} + O(\frac{1}{p})).$$

We need the following tensors and their corresponding tensor norms to establish the results of lemma 3 and 6 via an application of Proposition 6.

- For each $s \in [4]$, s^{th} order tensors A_s of expectations of s^{th} order derivatives of f with respect to each $\{x_{i_1,d_1}, ..., x_{i_s,d_s}\}_{i_j \in [m], d_j \in [p]}$.

- Tensor norms for A_s with respect to each $\mathcal{J}$ in P_s.

Computing A^1**:** The first order derivative of f with respect to x_{id} for any $i \in [m]$ and $d \in [p]$: $\frac{\partial f_{\sigma_*}(X)}{\partial x_{id}} =$

$$4x_{id}^3 + \sum_{d' \neq d} 2x_{id}x_{id'}^2 + \sum_{s \in [k]} \sum_{j \neq i \in \sigma_*^{-1}(s)} 2x_{id}x_{jd}^2 + \sum_{d' \neq d} \sum_{j \neq i \in \sigma_*^{-1}(s)} x_{id'}x_{jd}x_{jd'}. \quad (23)$$

$$\mathbb{E}(\frac{\partial f_{\sigma_*}(X)}{\partial x_{id}}) = O(\sqrt{p \log p}). \quad (24)$$

Therefore,

$$A^1 = O(\sqrt{p \log p})\mathbb{J}_{mp}$$

, where $\mathbb{J}_{mp} \in \mathbb{R}^{mp}$ denotes the vector of ones.

Computing A^2**:** The second order derivative of f with respect to $x_{id}, x_{k\beta}$ for any $i, k \in [m]$ and $d, \beta \in [p]$: $\frac{\partial f_{\sigma_*}(X)}{\partial x_{id} \partial x_{k\beta}} =$

$$\begin{cases} 12x_{id}^2 + \sum_{d' \neq d} 2x_{id'}^2 + \sum_{j \neq i \in \sigma_*^{-1}(s)} 2x_{jd}^2 & \text{if } k = i; \beta = d, \\ 4x_{id}x_{i\beta} + \sum_{j \neq i \in \sigma_*^{-1}(s)} 2x_{jd}x_{j\beta} & \text{if } k = i; \beta \neq d, \\ 4x_{id}x_{kd} + \sum_{d' \neq d} x_{id'}x_{kd'} & \text{if } k \neq i \in \sigma_*^{-1}(s); s \in [k]; \beta = d, \\ x_{i\beta}x_{kd} & \text{if } k \neq i \in \sigma_*^{-1}(s); s \in [k]; \beta \neq d, \\ 0 & \text{otherwise }. \end{cases} \quad (25)$$

Then $A^2(i, j) =$

$$\begin{cases} O(p) & \text{if } k = i; \beta = d, \\ O(\log p) & \text{if } k = i; \beta \neq d, \\ O(1) & \text{if } k \neq i \in \sigma_*^{-1}(s); s \in [k]; \beta = d, \\ O(\log p/p) & \text{if } k \neq i \in \sigma_*^{-1}(s); s \in [k]; \beta \neq d, \\ 0 & \text{otherwise }. \end{cases} \quad (26)$$

Computing A^3**:** The third order derivative of $f_{\sigma_*}(X)$ with respect to $x_{id}, x_{k\beta}, x_{\alpha l}$ for any $i, k \in [m]$ and $d, \beta \in [p]$: $\frac{\partial f_{\sigma_*}(X)}{\partial x_{id} \partial x_{k\beta} \partial x_{\alpha l}} =$

$$\begin{cases} 24x_{id} & \text{if } \alpha = k = i; l = \beta = d, \\ 4x_{il} & \text{if } \alpha = k = i; l \neq \beta = d, \\ 4x_{\alpha d} & \text{if } \alpha \neq k = i; l = \beta = d, \alpha \neq i \in \sigma_*^{-1}(s); s \in [k]; \\ x_{\alpha\beta} & \text{if } \alpha \neq k = i; l = d \neq \beta, \alpha \neq i \in \sigma_*^{-1}(s); s \in [k]; \\ 0 & \text{otherwise }. \end{cases} \quad (27)$$

Then $A^3_{id,k\beta,\alpha l} =$

$$\begin{cases} O(\frac{\log p}{p}) & \text{if } \alpha = k = i; l = \beta = d, \\ O(\frac{\log p}{p}) & \text{if } \alpha = k = i; l \neq \beta = d, \\ O(\frac{\log p}{p}) & \text{if } \alpha \neq k = i; l = \beta = d, \\ O(\frac{\log p}{p}) & \text{if } \alpha \neq k = i; l = d \neq \beta, \\ 0 & \text{otherwise }. \end{cases} \quad (28)$$

Computing A^4: The fourth order derivative of $f_{\sigma_*}(X)$ with respect to $x_{id}, x_{k\beta}, x_{\alpha l}, x_{q\gamma}$ for any $i, k, \alpha, q \in [m]$ and $d, \beta, l, \gamma \in [p]$: $\frac{\partial f_{\sigma_*}(X)}{\partial x_{id}\partial x_{k\beta}\partial x_{\alpha l}\partial x_{q\gamma}} =$

$$\begin{cases} 24 & \text{if } q = \alpha = k = i; \gamma = l = \beta = d, \\ 4 & \text{if } q = \alpha = k = i; \gamma = l \neq \beta = d, \\ 4 & \text{if } q = \alpha \neq k = i; \gamma = l = \beta = d, \alpha \neq i \in \sigma_*^{-1}(s); s \in [k]; \\ 1 & \text{if } q = \alpha \neq k = i; l = d \neq \beta = \gamma, \alpha \neq i \in \sigma_*^{-1}(s); s \in [k]; \\ 0 & \text{otherwise} . \end{cases} \tag{29}$$

Computing all the tensor norms, (see the proof of Lemma 9 for how the norms are computed) we have:

$\|A^1\|_{\{1\}} = O(p\sqrt{p\log p})$; $\|A^2\|_{\{1,2\}} = O(p^2)$; $\|A^2\|_{\{\{1\},\{2\}\}} = O(p\log p)$;
$\|A^3\|_{\{1,2,3\}} = O(m\sqrt{p\log p})$; $\|A^3\|_{\{1,2\},\{3\}} = O(\sqrt{mp})$; $\|A^3\|_{\{\{1\},\{2\},\{3\}\}} = O(\sqrt{p\log p})$;
$\|A^4\|_{\{1,2,3,4\}} = O(mp)$; $\|A^4\|_{\{1,2\},\{3,4\}} = O(p)$; $\|A^4\|_{\{\{1\},\{2\},\{3\},\{4\}\}} = O(1)$;
$\|A^4\|_{\{1\},\{2,3,4\}} = O(\sqrt{mp})$; $\|A^4\|_{\{\{1\},\{2\},\{3,4\}\}} = O(p)$;

Applying the lower tail bounds for $f_{\sigma_*}(X)$ from Proposition 6, and setting the R.H.S of the inequality to $\frac{1}{p}$, we derive the following upper bound that holds with probability at least $1 - 1/p$:

$$f_{\sigma_*}(X) \quad > \quad \frac{mp^2}{k}(k \ + \ \frac{\alpha}{k} \ + \ O(\frac{1}{p})) \ - \ (O(p^2\sqrt{\log p}) \ \vee \ O(mp\sqrt{\log p}))$$

Therefore,

$$\gamma_{\min} Q_{2\sigma_*} > \gamma_{\min}\left(1 + \frac{1}{k} + O(\frac{1}{p})\right) - C_2\gamma_{\min}\left(\sqrt{\frac{\log p}{p^2}} \vee \alpha\sqrt{\frac{\log p}{p^2}}\right).$$

Fix some $\epsilon > 0$ be an arbitrarily small constant. For any $\sigma : \|\beta(\sigma, \sigma_*)\|_F^2 < 1 + (k-1)\epsilon$, let $f_\sigma(X) = \sum_{s\in[k]} \sum_{i,j\in\sigma^{-1}(s)} \langle x_i, x_j\rangle^2$.

We can show that $\mathbb{E} f_\sigma(X) \leq \frac{mp^2}{k}(k + \frac{\alpha}{k} + O(\frac{1}{p}))$. Computing the tensors and their respective norms similarly as above, applying proposition 6, followed by an union bound over all such partitions and we can show that w.h.p,

$$\max_{\substack{\sigma:\|\beta(\sigma,\sigma_*)\|_F^2 \\ \leq 1+(k-1)\epsilon}} \gamma_{\max} Q_{2\sigma} \leq \gamma_{\max}\left(1 + \frac{1}{k} + O(\frac{1}{p})\right) + C_2\gamma_{\max} O\left(\sqrt{\frac{\alpha}{p}} \vee \alpha\sqrt{\frac{\alpha}{p}} \vee \sqrt{\frac{1}{\alpha p}}\right). \tag{30}$$

□