

# MLE phase transition of Gaussian tensor matching (Positive part of non-symmetric observations, Q&A)

Jiaxin Hu

May 17, 2022

## 1 Preliminary

### Non-symmetric correlated Gaussian observations.

Consider two order- $m$  random tensor observations  $\mathcal{A}, \mathcal{B}' \in \mathbb{R}^{n^{\otimes m}}$  and use  $\omega \in [n]^m$  to index the entries in  $\mathcal{A}$  and  $\mathcal{B}$ . Suppose that for all  $\omega \in [n]^m$  and some  $\rho \in (0, 1)$

$$\begin{pmatrix} \mathcal{A}_\omega \\ \mathcal{B}'_\omega \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \quad \text{and} \quad \begin{pmatrix} \mathcal{A}_\omega \\ \mathcal{B}'_\omega \end{pmatrix} \text{ is independent with } \begin{pmatrix} \mathcal{A}_{\omega'} \\ \mathcal{B}'_{\omega'} \end{pmatrix} \text{ for all } \omega \neq \omega'. \quad (1)$$

Let  $\pi^*$  be a permutation on  $[n]$  with corresponding permutation matrix  $\Pi^* \in \{0, 1\}^{n \times n}$ , and consider the permuted observation  $\mathcal{B}$  such that for all  $\omega \in [n]^m$

$$\mathcal{B}_\omega = \mathcal{B}'_{\pi^* \circ \omega}, \quad \text{or equivalently} \quad \mathcal{B} = \mathcal{B}' \times_1 \Pi^* \times_2 \cdots \times_m \Pi^*.$$

Our goal is to recover  $\pi^*$  (or equivalently  $\Pi^*$ ) observing  $\mathcal{A}, \mathcal{B}$ . **Note that  $\mathcal{A}, \mathcal{B}$  are not super-symmetric tensors while the permutation on every mode is the same!**

### MLE

By Theorem 1 in note 0402, the MLE of  $\pi^*$ , denoted  $\hat{\pi}_{MLE}$ , satisfies

$$\hat{\Pi}_{MLE} = \arg \max_{\Pi \in \mathcal{P}_n} \langle \mathcal{A} \times_1 \Pi \times_2 \cdots \times_m \Pi, \mathcal{B} \rangle,$$

where  $\hat{\Pi}_{MLE}$  is the permutation matrix corresponding to  $\hat{\pi}_{MLE}$ , and  $\mathcal{P}_n$  is the collection for all possible permutation matrices on  $[n]$ .

## 2 Theorem

**Theorem 1** (Achivability of MLE with non-symmetric observations). *Consider the observations  $(\mathcal{A}, \mathcal{B})$  from model (1) with true permutation  $\pi^*$ . Assume  $n$  is large enough and*

$$\rho^2 \geq \frac{C_0 \log n}{n^{m-1}},$$

for some  $C_0 > 0$ . Then, the MLE  $\hat{\pi}_{MLE}$  exactly recovers true permutation  $\pi^*$ ; i.e.,  $\hat{\pi}_{MLE} = \pi^*$  with probability tends to 1.

*Proof of Theorem 1.* Without the loss of generality, assume the true permutation  $\pi^*$  is the identity mapping. With observations  $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n \otimes m}$ , consider the loss function

$$\mathcal{L}(\pi, \mathcal{A}, \mathcal{B}) = \langle \mathcal{A} \times_1 \Pi \times_2 \cdots \times_m \Pi, \mathcal{B} \rangle,$$

where  $\Pi \in \{0, 1\}^{n \times n}$  is the permutation matrix corresponding to  $\pi$ . We define the difference

$$\begin{aligned} \Delta(\pi) &:= \mathcal{L}(\pi, \mathcal{A}, \mathcal{B}) - \mathcal{L}(\pi^*, \mathcal{A}, \mathcal{B}) \\ &= \rho \sum_{\omega \in [n]^m} (\mathcal{A}_{\pi \circ \omega} - \mathcal{A}_\omega) \mathcal{A}_\omega + \sqrt{1 - \rho^2} \sum_{\omega \in [n]^m} (\mathcal{A}_{\pi \circ \omega} - \mathcal{A}_\omega) \mathcal{Z}_\omega, \end{aligned}$$

where the second equality follows from the fact that  $\mathcal{B} = \rho \mathcal{A} + \sqrt{1 - \rho^2} \mathcal{Z}$ , where  $\mathcal{Z}_\omega \sim N(0, 1)$  for all  $\omega \in [n]^m$  independently and  $\mathcal{Z}$  is independent with  $\mathcal{A}$ . Hence, to show the exact recovery of MLE  $\hat{\pi}$  with high probability, it suffices to show that

$$\mathbb{P}(\hat{\pi}_{MLE} \neq \pi^*) = \mathbb{P}(\text{exists a } \pi \neq \pi^* \text{ such that } \Delta(\pi) \geq 0) = o(1).$$

Note that

$$\sum_{\omega \in [n]^m} (\mathcal{A}_{\pi \circ \omega} - \mathcal{A}_\omega) \mathcal{A}_\omega = -\frac{1}{2} (\|\mathcal{A}\|_F^2 + \|\pi \circ \mathcal{A}\|_F^2 - 2 \langle \mathcal{A}, \pi \circ \mathcal{A} \rangle) = -\frac{1}{2} \|\pi \circ \mathcal{A} - \mathcal{A}\|_F^2,$$

and conditional on  $\mathcal{A}$  we have  $\sum_{\omega \in [n]^m} (\mathcal{A}_{\pi \circ \omega} - \mathcal{A}_\omega) \mathcal{Z}_\omega \sim N(0, \|\pi \circ \mathcal{A} - \mathcal{A}\|_F^2)$ .

Then, given a permutation  $\pi$ , we have

$$\begin{aligned} \mathbb{P}(\Delta(\pi) \geq 0) &= \mathbb{E}[\mathbb{E}[\mathbf{1}\{\Delta(\pi) \geq 0\} | \mathcal{A}]] \\ &\leq \mathbb{E} \left[ \mathbb{P} \left( -\frac{\rho}{2} \|\pi \circ \mathcal{A} - \mathcal{A}\|_F^2 + \sqrt{1 - \rho^2} N(0, \|\pi \circ \mathcal{A} - \mathcal{A}\|_F^2) \geq 0 \middle| \mathcal{A} \right) \right] \\ &\leq \mathbb{E} \left[ \mathbb{P} \left( N(0, 1) \geq \frac{\rho \|\pi \circ \mathcal{A} - \mathcal{A}\|_F}{2\sqrt{1 - \rho^2}} \middle| \mathcal{A} \right) \right] \\ &\leq \mathbb{E} \left[ \exp \left( -\frac{\rho^2}{8} \|\pi \circ \mathcal{A} - \mathcal{A}\|_F^2 \right) \right], \end{aligned} \tag{2}$$

where the last inequality follows from the inequality that  $\mathbb{P}(N(0, 1) \geq t) \leq \exp(-t^2/2)$  for all  $t \geq 0$  and  $1 - \rho^2 \leq 1$ .

**Above statements also work for symmetric observations. Following statements only works for non-symmetric observations since we may have a different lower bound for  $\|\pi \circ \mathcal{A} - \mathcal{A}\|_F$ .**

Now, to bound the probability  $\mathbb{P}(\Delta(\pi) \geq 0)$ , we need to find the lower bound of  $\|\pi \circ \mathcal{A} - \mathcal{A}\|_F$  which represents edge permutation error brought by  $\pi$ . Intuitively, more node disagreements in  $\pi$  lead to more edges disagreements between  $\pi \circ \mathcal{A}$  and  $\mathcal{A}$ . Propositions 1 and 2 provide the relationship between the node disagreements and edge disagreement with non-symmetric observations.

Let  $D_\pi = \{i \in [n] : \pi(i) \neq i\}$  denote the set of unfixed points of  $\pi$  and  $D_\pi^{m,E} = \{\omega \in [n]^m : \pi \circ \omega \neq \omega\}$  denote the set of unfixed order- $m$  edges of  $\pi$ . Consider the event

$$E(\mathcal{A}) := \{\text{for all } \pi \neq \pi^*, \quad \|\pi \circ \mathcal{A} - \mathcal{A}\|_F^2 \geq 2|D_\pi^{m,E}|(1 - \epsilon_n)\}, \quad (3)$$

where  $\epsilon_n = \frac{C}{2} \sqrt{\log n / n^{m-1}}$ . By Proposition 2, we have  $\mathbb{P}(E^c(\mathcal{A})) = o(1)$ .

Let  $\mathcal{P}_{n,d}$  be the collection of all the permutations  $\pi$  with  $|D_\pi| = d$ . We have

$$\begin{aligned} \mathbb{P}(\hat{\pi}_{MLE} \neq \pi^*) &= \mathbb{P}(\text{exists a } \pi \neq \pi^* \text{ such that } \Delta(\pi) \geq 0) \\ &\leq \mathbb{P}(E^c(\mathcal{A})) + \sum_{\pi \neq \pi^*} \mathbb{P}(E(\mathcal{A}), \Delta(\pi) \geq 0) \\ &\leq o(1) + \sum_{d=2, \dots, n} \sum_{\pi \in \mathcal{P}_{n,d}} \mathbb{E} \left[ \mathbb{E}[\mathbb{1}\{\Delta(\pi) \geq 0\} | \mathcal{A}] \mathbb{1}\{E(\mathcal{A})\} \right] \\ &\leq o(1) + \sum_{d=2, \dots, n} \sum_{\pi \in \mathcal{P}_{n,d}} \exp \left( -\frac{\rho^2}{4} (1 - \epsilon_n) |D_\pi^{m,E}| \right) \\ &\leq o(1) + \sum_{d=2, \dots, n} n^d \exp \left( -\frac{\rho^2}{4} (1 - \epsilon_n) d n^{m-1} \right) \\ &= o(1), \end{aligned}$$

where the last second inequality follows from inequality (2) and the definition of  $E(\mathcal{A})$  (3), the last inequality follows from Proposition 1 and the fact that  $|\mathcal{P}_{n,d}| \leq n^d$ , and the last equality holds under the assumption that  $\rho^2 \geq \frac{C_0 \log n}{n^{m-1}}$  for some positive constant  $C_0$ . □

**Proposition 1** (Relationship between unfixed points and unfixed edges). *Suppose we have a permutation  $\pi$  on  $[n]$ . Let  $D_\pi = \{i \in [n] : \pi(i) \neq i\}$  denote the set of unfixed points of  $\pi$  and  $D_\pi^{m,E} = \{\omega \in [n]^m : \pi \circ \omega \neq \omega\}$  denote the set of unfixed order- $m$  edges. Then, we have*

$$n^{m-1}|D_\pi| \leq |D_\pi^{m,E}| \leq m n^{m-1}|D_\pi|.$$

*Proof of Proposition 1.* For simplicity, let  $d = |D_\pi|$  and  $d^E = |D_\pi^{m,E}|$ . Note that

$$d^E = \sum_{k=0}^{m-1} (n-d)^k d^{m-k} \binom{m}{k},$$

where  $k$  refers to the number of fixed points (i.e.,  $\pi(i) = i$ ) and the unfixed order- $m$  edge at most have  $m-1$  fixed points,  $(n-d)^k$  refers to all the combinations of fixed points and  $d^{m-k}$  refers to all the combinations of unfixed points in the edge, and  $\binom{m}{k}$  is the number of all position positions for fixed points. By Binomial Identity, we have

$$n^{m-1}d \leq d^E = n^m - (n-d)^m = n^m \left[ 1 - \left( 1 - \frac{d}{n} \right)^m \right] \leq m n^{m-1}d$$

where the first inequality follows from the fact that  $(1-x)^m \leq 1-x$  for  $x \in (0,1)$  and the second inequality follows from the inequality that  $(1-x)^m \geq 1-mx$  for  $x \geq -1$ . □

**Proposition 2** (Edge disagreement with permutation  $\pi$ ). *Suppose we have an order- $m$  observation  $\mathcal{A} \in \mathbb{R}^{n^{\otimes m}}$  with i.i.d. standard Gaussian entries. Let  $D_\pi^{m,E} = \{\omega \in [n]^m : \pi \circ \omega \neq \omega\}$  denote the set of unfixed order- $m$  edges. We have the expectation*

$$\mathbb{E} [\|\pi \circ \mathcal{A} - \mathcal{A}\|_F^2] = 2|D_\pi^{m,E}|,$$

and there exists a positive constant  $C$  such that

$$\left| \|\pi \circ \mathcal{A} - \mathcal{A}\|_F^2 - 2|D_\pi^{m,E}| \right| \leq C|D_\pi^{m,E}| \sqrt{\frac{\log n}{n^{m-1}}},$$

with high probability.

*Proof of Proposition 2.* Note that

$$\mathbb{E} [\|\pi \circ \mathcal{A} - \mathcal{A}\|_F^2] = \sum_{\omega \in [n]^m} \mathbb{E}[(\mathcal{A}_{\pi \circ \omega} - \mathcal{A}_\omega)^2] = \sum_{\omega \in D_\pi^{m,E}} \mathbb{E}[(\mathcal{A}_{\pi \circ \omega} - \mathcal{A}_\omega)^2] = 2|D_\pi^{m,E}|,$$

where the last equation follows from the fact that  $\mathcal{A}_{\pi \circ \omega} - \mathcal{A}_\omega \sim N(0, 2)$  for all  $\omega \in D_\pi^{m,E}$ .

Following the proof of Corollary 1.1 in [Ganassali \(2020\)](#), with high probability, we have

$$\left| \|\pi \circ \mathcal{A} - \mathcal{A}\|_F^2 - 2|D_\pi^{m,E}| \right| \leq C' \sqrt{|D_\pi^{m,E}| d_\pi \log n} \leq C|D_\pi^{m,E}| \sqrt{\frac{\log n}{n^{m-1}}},$$

where  $d_\pi = |\{i \in [n] : \pi(i) \neq i\}|$  is the number of unfixed points in  $\pi$ , and the second inequality follows from the Proposition 1.  $\square$

### 3 Q&A

#### 1. How the symmetry of observations affects the MLE phase transition theorem?

Maximizing likelihood function is equivalent to minimizing the loss  $\sum_{\omega \in [n]^m} (\mathcal{A}_{\pi \circ \omega} - \mathcal{B}_\omega)^2$ , which measures the edge difference between the permuted  $\mathcal{A}$  and  $\mathcal{B}$  with given permutation  $\pi$  and correlation  $\rho$ . Note that  $\mathcal{B}_\omega = \rho \mathcal{A}_{\pi^* \circ \omega} + \sqrt{1 - \rho^2} Z$ , where  $Z$  is a standard normal variable independent with  $\mathcal{A}, \mathcal{B}$ . The edge difference comes from two aspects: 1) the noise  $\sqrt{1 - \rho^2} Z$  and 2) the error in edge permutation  $\mathcal{A}_{\pi \circ \omega} - \mathcal{A}_{\pi^* \circ \omega}$ , which is control by  $\rho$  and  $\pi$ , respectively.

Intuitively, if the likelihood is dominated by the error in edge permutation, then true permutation  $\pi^*$  can be easily recovered by MLE since  $\pi^*$  leads to 0 edge permutation error. Hence, we need to find the lower bound of the edge permutation error with given  $\pi$  and find the noise condition to make the edge permutation error dominant in likelihood.

With a given permutation  $\pi$ , symmetry or non-symmetry of the observations lead to different relationships between the node disagreements and the edge permutation error. Without loss of generality, we assume  $\pi^*$  is identity mapping. Then, the edge permutation error  $\mathcal{A}_{\pi \circ \omega} - \mathcal{A}_\omega \neq 0$  if

- (1)  $\mathcal{A}$  is non-symmetric and  $\pi \circ \omega \neq \omega$  ;
- (2)  $\mathcal{A}$  is super-symmetric and  $\pi \circ \omega \notin \{v = (v_1, \dots, v_m) \in [n]^m : \{v_1, \dots, v_m\} = \{\omega_1, \dots, \omega_m\}\}$ .

For example,  $\mathcal{A}_{1,2,3} \neq \mathcal{A}_{2,1,3}$  in non-symmetric case but  $\mathcal{A}_{1,2,3} = \mathcal{A}_{2,1,3}$  in super-symmetric case. Therefore, we have

$$\{\boldsymbol{\omega} \in [n]^m : \mathcal{A}_{\pi \circ \boldsymbol{\omega}} - \mathcal{A}_{\boldsymbol{\omega}} \neq 0, \mathcal{A} \text{ is super-symmetric}\} \subset \{\boldsymbol{\omega} \in [n]^m : \mathcal{A}_{\pi \circ \boldsymbol{\omega}} - \mathcal{A}_{\boldsymbol{\omega}} \neq 0, \mathcal{A} \text{ is non-symmetric}\}.$$

This implies that **under the same permutation  $\pi$ , the non-symmetric observation has a larger edge permutation error than that for symmetric observations**. So, it is more easier for MLE to recover true permutation in non-symmetric observations, and thus we will have a looser noise condition for  $\rho^2$ .

Since we need to find the lower bound of the permutation error, in symmetric case, we need to find the lower bound of  $|\{\boldsymbol{\omega} \in [n]^m : \mathcal{A}_{\pi \circ \boldsymbol{\omega}} - \mathcal{A}_{\boldsymbol{\omega}} \neq 0, \mathcal{A} \text{ is super-symmetric}\}|$  which is difficult when  $m > 3$ . Because you need to exclude the edges  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_m)$  such that  $\{\omega_1, \dots, \omega_m\} = \{\pi(\omega_1), \dots, \pi(\omega_m)\}$  and  $\pi(\omega_i) \neq \omega_i$  for  $i \in [m]$  (for example  $\boldsymbol{\omega} = (1, 2, 3, 4)$  with  $\pi(\boldsymbol{\omega}) = (3, 1, 4, 2)$ ).

## References

Ganassali, L. (2020). Sharp threshold for alignment of graph databases with gaussian weights. *arXiv preprint arXiv:2010.16295*.