MLE phase transition of Gaussian tensor matching

Jiaxin Hu

May 10, 2022

This note investigates the positive part in the MLE phase transition in Gaussian tensor matching. The key idea is to use the likelihood property of MLE and the concentration of Gaussian entries.

The following statements and proofs are rigorous for $m \leq 3$. The technical difficulty for higher order $m > 3$ cases lies in counting the number of exchange edges with unfixed points $e_m(\pi) = |\{\boldsymbol{\omega} \in D_\pi^m : \pi \circ \boldsymbol{\omega} = \text{swap}(\pi^* \circ \boldsymbol{\omega})\}|$, where $D_\pi = \{i \in [n] : \pi(i) \neq \pi^*(i)\}$ with number of fixed points $d_\pi = D_\pi$. We need $e_m(\pi) = o(d_\pi n^{m-1})$ to show the sharp threshold for ρ^2 is $\mathcal{O}(\log n / n^{m-1})$.

1 Preliminaries

Higher-order Correlated Winger Model Consider two random super-symmetric tensors $\mathcal{A}, \mathcal{B}' \in \mathbb{R}^{n^{\otimes m}}$. Assume that all the pairs $\{(\mathcal{A}_{\boldsymbol{\omega}}, \mathcal{B}'_{\boldsymbol{\omega}}) \colon \boldsymbol{\omega} \in [n]^m \cap \{\boldsymbol{\omega} \colon \omega_1 \leq \cdots \leq \omega_m\}\}$ follow the i.i.d. correlated multivariate zero-mean Gaussian distribution with variance 1 and correlation $\rho \in (0,1)$; i.e.,

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

for all $\boldsymbol{\omega}' \neq \boldsymbol{\omega}$ and $\boldsymbol{\omega}' \in [n]^m \cap \{\boldsymbol{\omega} \colon \omega_1 \leq \cdots \leq \omega_m\}$. The tensors $\mathcal{A}, \mathcal{B}'$ are two correlated Winger tensors. Let π^* be a permutation on $[n]$, and $\Pi^* \in \{0,1\}^{n \times n}$ denote the corresponding permutation matrix with entries $\Pi^*_{ij} = 1$ if $j = \pi^*(i)$ and $\Pi^*_{ij} = 0$, otherwise. Consider the permuted tensor $\mathcal{B}$ such that for all $\boldsymbol{\omega} \in [n]^m$

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

We call the pair $(i,k) \in [n]^2$ as a *true pair* if $k = \pi^*(i)$, and (i,k) is a *fake pair*, otherwise. We also call the observation $(\mathcal{A}, \mathcal{B})$ follow the *permuted higher-order correlated Winger model (pHCWM) with parameter* π^* *and* ρ, denoted as $(\mathcal{A}, \mathcal{B}) \sim pHCWM_{n,m}(\pi^*, \rho)$.

Our goal is to recover π^* (or equivalently Π^*) observing $\mathcal{A}, \mathcal{B}$.

Theorem 1 (MLE for Higher-order correlated Winger model)**.** *Suppose that the order-m tensor observation* $(\mathcal{A}, \mathcal{B}) \sim pHCWM_{n,m}(\pi^*, \rho^*)$. *The MLE of the true permutation* π^**, denoted* $\hat{\pi}_{MLE}$*, satisfies*

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

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 Positive Phase transition result

Theorem 2 (Achievability (Conjecture when $m > 3$))**.** *Consider the observations* $(\mathcal{A}, \mathcal{B}) \sim pHCWM_{n,m}(\pi^*, \rho)$. *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 $\hat{\pi}$ in Theorem 1 exactly recovers true permutation π^*; i.e., $\hat{\pi}_{MLE} = \pi^*$ with probability tends to 1.

Proof of Theorem 2. Without the loss of generality, assume the true permutation 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 π. We define the difference

$$\begin{aligned}
\Delta(\pi) &\coloneqq \mathcal{L}(\pi, \mathcal{A}, \mathcal{B}) - \mathcal{L}(\pi^*, \mathcal{A}, \mathcal{B}) \\
&= \langle \mathcal{A} \times_1 (\Pi - \boldsymbol{I}_n) \times_2 \cdots \times_m (\Pi - \boldsymbol{I}_n), \mathcal{B} \rangle \\
&= \left\langle \mathcal{A} \times_1 (\Pi - \boldsymbol{I}_n) \times_2 \cdots \times_m (\Pi - \boldsymbol{I}_n), \rho\mathcal{A} + \sqrt{1-\rho^2}\mathcal{Z} \right\rangle \\
&= \rho \sum_{\boldsymbol{\omega} \in [n]^m} (\mathcal{A}_{\pi \circ \boldsymbol{\omega}} - \mathcal{A}_{\boldsymbol{\omega}}) \mathcal{A}_{\boldsymbol{\omega}} + \sqrt{1-\rho^2} \sum_{\boldsymbol{\omega} \in [n]^m} (\mathcal{A}_{\pi \circ \boldsymbol{\omega}} - \mathcal{A}_{\boldsymbol{\omega}}) \mathcal{Z}_{\boldsymbol{\omega}},
\end{aligned}$$

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

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

To bound above probability, we firstly introduce a few notions related to the correctness of π.

1. $\text{swap}(\boldsymbol{\omega})$: any possible vector after swapping the elements in the vector $\boldsymbol{\omega}$;
2. $D_\pi \coloneqq \{i \in [n] : \pi(i) \neq i\}$: the set of unfixed points of π;
3. $d_\pi \coloneqq |D_\pi|$: the number of unfixed points of π;
4. $D_\pi^{m,E} \coloneqq \{\boldsymbol{\omega} \in [n]^m : \pi \circ \boldsymbol{\omega} \neq \text{swap}(\boldsymbol{\omega})\}$: the set of unfixed order-m edges of π;
5. $d_\pi^{m,E} \coloneqq |D_\pi^{m,E}|$: the number of unfixed order-m edges of π;
6. $E_\pi(k) \coloneqq \{\boldsymbol{\omega} \in D_\pi^k : \pi \circ \boldsymbol{\omega} = \text{swap}(\boldsymbol{\omega})\}$: the set of *exchangeable* dimension-k vectors of π ;
7. $e_\pi(k) \coloneqq |E_\pi(k)|$ for all $k \in [m]$: the number of exchangeable dimension-k vectors of π;
8. $v_{\mathcal{A}}(\pi) \coloneqq \sum_{\boldsymbol{\omega} \in [n]^m} (\mathcal{A}_{\pi \circ \boldsymbol{\omega}} - \mathcal{A}_{\boldsymbol{\omega}})^2$: the Frobenius norm between the permuted and original observation $\mathcal{A}$.

Note that

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

where the first equation follows from the fact that $\|\mathcal{A}\|_F^2 = \|\pi\circ\mathcal{A}\|_F^2$. Also, note that conditional on $\mathcal{A}$, the term $\sum_{\boldsymbol{\omega}\in[n]^m}(\mathcal{A}_{\pi\circ\boldsymbol{\omega}}-\mathcal{A}_{\boldsymbol{\omega}})\mathcal{Z}_{\boldsymbol{\omega}} \sim N(0, v_{\mathcal{A}}(\pi))$.

Then, given a permutation π, we have

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

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

Take $m=3$. Consider the event

$$Evt(\mathcal{A}) \coloneqq \{\text{for all } \pi \text{ with } d_\pi\geq 2, \quad v_{\mathcal{A}}(\pi)\geq 2d_\pi^{3,E}-C\sqrt{d_\pi^{3,E}d_\pi\log n}\geq d_\pi^{3,E}(2-2\epsilon_n)\},$$

where $\epsilon_n = C_1\sqrt{\log n/n^2}$ for some positive constant C_1 and the last inequality in $Evt(\mathcal{A})$ holds with the Proposition 2. Then, by Proposition 3, we have $\mathbb{P}(Evt^c(\mathcal{A})) = 1-\mathbb{P}(Evt(\mathcal{A})) = o(1)$.

Hence, we have

$$\begin{aligned}
\mathbb{P}(\hat{\pi}_{MLE}\neq\pi^*) &= \mathbb{P}(\text{exists a } \pi\neq\pi^* \text{ such that } \Delta(\pi)\geq 0) \\
&\leq \mathbb{P}(Evt^c(\mathcal{A}))+\mathbb{P}(Evt(\mathcal{A}), \text{exists a } \pi\neq\pi^* \text{ such that } \Delta(\pi)\geq 0) \\
&\leq o(1)+\sum_{\pi\neq\pi^*}\mathbb{E}\left[\exp\left(-\frac{\rho^2}{8(1-\rho^2)}v_{\mathcal{A}}(\pi)\right)\mathbb{1}\{Evt(\mathcal{A})\}\right] \\
&\leq o(1)+\sum_{d=2,\dots,n}\sum_{\pi:d_\pi=d}\exp\left(-\frac{\rho^2}{4(1-\rho^2)}(1-\epsilon_n)d_\pi^{3,E}\right) \\
&\leq o(1)+\sum_{d=2,\dots,n}n^d\exp\left(-\frac{\rho^2}{4}(1-\epsilon_n)n^2(d-2^{1/3}\sqrt{d})\right) \\
&\leq o(1),
\end{aligned}$$

where the last second inequality follows from the fact that $|\pi: d_\pi = p|\leq n^d$ and Proposition 2, and the last inequality holds under the assumption that $\rho\geq\frac{C_1\log n}{n^2}$ for some positive constant C_1. $\square$

3 Useful Propositions

Following are useful propositions of the new notions with a given permutation π on $[n]$.

Proposition 1 (Upper bound of the number of exchangeable vectors (Need to extend the upper bound for $k > 4$)). *Suppose that we have a given permutation π on $[n]$ with the number of unfixed points $d_\pi \geq 2$. Then, we have*

$$e_\pi(1) = 0; \quad e_\pi(2) \leq d_\pi; \quad e_\pi(3) \begin{cases} = 0 & d_\pi = 2 \\ \leq 2d & d_\pi \geq 3 \end{cases},$$

and thus

$$e_\pi(k) \leq (2^{1/3}\sqrt{d_\pi})^k, \quad \textit{for } k = \textit{1, 2,3.}$$

Proof of Proposition 1. When $k = 1$, by definition, we have $\pi(\omega) \neq \omega$ for all $\omega \in D_\pi$ and thus $e_\pi(1) = 0$.

When $k = 2$, consider an arbitrary $\boldsymbol{\omega} = (\omega_1, \omega_2) \in E_\pi(2)$. Since $\pi \circ \boldsymbol{\omega} = \mathrm{swap}(\boldsymbol{\omega})$ and ω_1, ω_2 are unfixed points, then we have $\omega_2 = \pi(\omega_1)$ and $\pi(\omega_2) = \omega_1$. Therefore, we have at most swap-wise different $\lfloor d_\pi/2 \rfloor$ vectors $\boldsymbol{\omega} \in D_\pi^2$ such that $\mathrm{swap}(\boldsymbol{\omega}) \in E_\pi(2)$, and thus $e_\pi(2) \leq \lfloor d_\pi/2 \rfloor \times 2 \leq d_\pi$, where $\times 2$ comes from the number of possible $\mathrm{swap}(\boldsymbol{\omega})$ given $\boldsymbol{\omega} \in E_\pi(2)$.

When $k = 3$ and $d_\pi = 2$, we have $D_\pi = \{\omega_1, \omega_2\}$ such that $\omega_2 = \pi(\omega_1)$ and $\pi(\omega_2) = \omega_1$. Then, all the vectors in D_π^3 are not in $E_\pi(3)$ due to the odd number of elements.

When $k = 3$ and $d_\pi \geq 3$, consider an arbitrary vectors $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3) \in E_\pi(3)$. Since $\pi \circ \boldsymbol{\omega} = \mathrm{swap}(\boldsymbol{\omega})$ and $\omega_1, \omega_2, \omega_3$ are unfixed points, we must have $\{\omega_2, \omega_3\} = \{\pi(\omega_1), \pi^{-1}(\omega)\}$ where $\pi(\omega_1) \neq \pi^{-1}(\omega)$ and $\pi(\pi(\omega_1)) = \pi^{-1}(\omega)$. Otherwise, if $\omega_2 = \pi(\omega_1) = \pi^{-1}(\omega)$, the element ω_3 should be a fixed point.

Consider another vector $\boldsymbol{v} = (\omega_1, v_2, v_3) \in E_\pi(3)$ such that $\boldsymbol{v} \neq \boldsymbol{\omega}$. Similarly, we must have $\{v_2, v_3\} = \{\pi(\omega_1), \pi^{-1}(\omega)\}$ and thus $\boldsymbol{v} = \mathrm{swap}(\boldsymbol{\omega})$, which implies that all the vectors in $E_\pi(3)$ with element ω_1 is the swap of $\boldsymbol{\omega}$. The same conclusion holds when considering all the vectors in $E_\pi(3)$ with element ω_2 or ω_3.

Therefore, we at most have $\lfloor d_\pi/3 \rfloor$ swap-wise different vectors $\boldsymbol{\omega} \in D_\pi^3$ such that $\mathrm{swap}(\boldsymbol{\omega}) \in E_\pi(3)$, and thus $e_\pi(3) \leq \lfloor d_\pi/3 \rfloor \times 3! \leq 2d_\pi$, where $\times 3!$ the number of possible $\mathrm{swap}(\boldsymbol{\omega})$ given $\boldsymbol{\omega} \in E_\pi(3)$.

$\square$

Proposition 2 (The number of unfixed edges). *The number of unfixed order-m edges of π with $d_\pi \geq 2$ is lower bounded as*

$$d_\pi^{m,E} = \sum_{k=0}^{m-1} (n - d_\pi)^k (d_\pi^{m-k} - E_\pi(m-k)) \binom{m}{k} \leq d_\pi n^{m-1}. \tag{2}$$

When $m = 3$, we have

$$n^2(d_\pi - 2^{1/3}\sqrt{d_\pi}) \leq d_\pi^{m,E} \leq d_\pi n^2. \tag{3}$$

Proof of Proposition 2. Consider the unfixed edges $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m) \in D_\pi^{m,E}$ with k fixed points in $\{\omega_1, \ldots, \omega_m\}$. Then, the term $\binom{m}{k}$ counts the number of possible positions for fixed points, and $(n - d_\pi)^k$ counts all possible combinations of k fixed points. And the term $(d_\pi^{m-k} - E_\pi(m-k))$

counts all possible combinations of $m-k$ unfixed points excluding the exchangeable dimension-$(m-k)$ vectors. Noticing that we at most have $m-1$ fixed points in the unfixed edges, we obtain the equation in (2). Noticing that there is at least 1 unfixed point for the edges in $D_\pi^{m,E}$ and n^{m-1} is the all possible combinations of other points, the inequality in (3) holds

Take $m=3$. For the first inequality in (3), we have

$$\begin{aligned}
d_\pi^{3,E} &= \sum_{k=0}^{2} (n-d_\pi)^k (d_\pi^{3-k} - E_\pi(3-k)) \binom{3}{k} \\
&\geq \sum_{k=0}^{2} (n-d_\pi)^k d_\pi^{3-k} \binom{3}{k} - \sum_{k=0}^{2} (n-d_\pi)^k (2^{1/3}\sqrt{d_\pi})^{3-k} \binom{3}{k} \\
&= n^3 - (n-d_\pi)^3 - (n-d_\pi+2^{1/3}\sqrt{d_\pi})^3 + (n-d_\pi)^3 \\
&= n^3 \left(1 - \left(1 - \frac{d_\pi - 2^{1/3}\sqrt{d_\pi}}{n}\right)^3\right) \\
&\geq n^2 (d_\pi - 2^{1/3}\sqrt{d_\pi}),
\end{aligned}$$

where the inequality follows from Proposition 1, the second equation follows from the Binomial Identity, and the last inequality follows from the fact that $0 < d_\pi - 2^{1/3}\sqrt{d_\pi} < n$ for $d_\pi \geq 2$ and $(1-x)^3 \leq 1-x$ for $x \in (0,1)$. □

Proposition 3 (Concentration of $v_{\mathcal{A}}(\pi)$)**.** *Suppose $\mathcal{A} \in \mathbb{R}^{n^{\otimes m}}$ follows model (1). Then, we have*

$$\mathbb{E}[v_{\mathcal{A}}(\pi)] = 2d_\pi^{m,E},$$

and there exists a positive constant C such that with high probability

$$|v_{\mathcal{A}}(\pi) - 2d_\pi^{m,E}| \leq C\sqrt{d_\pi^{m,E} d_\pi \log n}.$$

Proof of Proposition 3. Note that

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

where the last inequality follows from the that that $\mathcal{A}_{\pi\circ\boldsymbol{\omega}} - \mathcal{A}_{\boldsymbol{\omega}} \sim N(0,2)$ for all $\boldsymbol{\omega}$ such that $\pi\circ\boldsymbol{\omega} \neq \mathrm{swap}(\boldsymbol{\omega})$.

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

$$|v_{\mathcal{A}}(\pi) - 2d_\pi^{m,E}| \leq C\sqrt{d_\pi^{m,E} d_\pi \log n}.$$

□

References

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