

MLE of matching

Jiaxin Hu

May 5, 2022

This note investigates the MLE issues in the Gaussian matching problems. For self-consistency, we first recall the Higher-order Correlated Winger Model and the MLE of the unknown permutation.

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 $\{(\mathcal{A}_\omega, \mathcal{B}'_\omega) : \omega \in [n]^m \cap \{\omega : \omega_1 \leq \dots \leq \omega_m\}\}$ follow the i.i.d. correlated multivariate zero-mean Gaussian distribution with variance 1 and correlation $\rho \in (0, 1)$; i.e.,

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

for all $\omega' \neq \omega$ and $\omega' \in [n]^m \cap \{\omega : \omega_1 \leq \dots \leq \omega_m\}$. 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 $\omega \in [n]^m$

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

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 $(\mathcal{A}, \mathcal{B}) \sim \text{pHCWM}_{n,m}(\pi^*, \rho)$.

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 $(\mathcal{A}, \mathcal{B}) \sim \text{pHCWM}_{n,m}(\pi^*, \rho^*)$. The MLE of the true permutation π^* , denoted $\hat{\pi}$ satisfies*

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

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

1 Q & A

1. Is the MLE estimate of correlated Gaussian matching known in the literature?

The MLE result for the correlated Winger/Gaussian matrices are known in the literature. See [Wu et al. \(2022\)](#) and [Ganassali \(2020\)](#).

2. What's the state-of-art algorithm for matrix Gaussian matching?

The state-of-art matrix Gaussian matching algorithms are [Ding et al. \(2021\)](#) based on the degree profile and the spectral method in [Fan et al. \(2019\)](#). Both of the methods achieves exact recovery under the noise regime $\sigma = \sqrt{1 - \rho^2} \leq \mathcal{O}(1/\text{polylog}(n))$ with complexity $\mathcal{O}(n^3)$. [Ganassali et al. \(2019\)](#) proposes another spectral method with less complexity $\mathcal{O}(n^2)$ that achieves exact recovery under a stricter noise condition $\sigma \leq \mathcal{O}(1/\text{poly}(n))$.

For matrix Bernoulli matching (with correlated Erdős-Rényi graphs), the state-of-art algorithm is proposed by [Mao et al. \(2021\)](#). The method in [Mao et al. \(2021\)](#) achieves exact recovery under the regime $np \geq \text{polylog}(n)$ and $\sigma \leq \mathcal{O}(1/\text{polylog}(\log n))$, where p is the parameter in the Bernoulli entry. The degree profile method in [Ding et al. \(2021\)](#) shows the exact recovery in a smaller regime $np \geq \text{polylog}(n)$ and $\sigma \leq \mathcal{O}(1/\text{polylog}(n))$. Unlike [Ding et al. \(2021\)](#), [Mao et al. \(2021\)](#) does not match the vertices directly but matches the partitions of the vertices and then refine the mapping to a full permutation.

3. Is the matching problem getting harder or easier as n increases?

Note that estimating a permutation on n vertices is not harder than estimating an n -dimensional vector. The degree of freedom, or the number of parameters, to estimate the permutation on $[n]$ is upper bounded by n . The sample size in the matching problem is $\frac{n(n-1)}{2}$ noticing the observations are symmetric matrices. Therefore, the ratio of ($\#$ parameters)/($\#$ samples) goes to 0 as n increases, and thus the problem to should be easier as n increases.

4. What's the sharp thresholds to exact recover the Gaussian matching?

We first formally define the exact recovery. Consider the metric measuring the agreement between to permutations on $[n]$

$$OV_n(\pi_1, \pi_2) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}\{\pi_1(i) \neq \pi_2(i)\}.$$

We call the estimate $\hat{\pi}$ exact recovers the true permutation π^* if

$$\mathbb{P}(OV_n(\hat{\pi}, \pi^*) = 0) = 1 - o(1).$$

Both [Wu et al. \(2022\)](#) and [Ganassali \(2020\)](#) show that:

- (1) When $\rho^2 > \frac{4 \log n}{n}$, the MLE $\hat{\pi}_{MLE}$ exact recovers the true permutation π^* ;
- (2) When $\rho^2 < \frac{4 \log n}{n}$, for any estimator $\hat{\pi}$ can not exact recovers the true permutation π^* .

The threshold agrees with the ratio

$$\frac{\log(\# \text{ complexity of the models})}{(\# \text{ sample size})} = \frac{\log(n!)}{n^2} = \mathcal{O}\left(\frac{\log n}{n}\right),$$

where the $n!$ is the number of possible permutations on $[n]$, and the last inequality follows the Stirling's approximation $\log(n!) = \mathcal{O}(n \log n)$.

Two papers [Ganassali \(2020\)](#) and [Mao et al. \(2021\)](#) comments a little bit on the gap between the sharp threshold of MLE and current polynomial-time algorithm, but none of them provide an informative reason for the big statistical-computational gap.

2 MLE phase transition of Gaussian tensor matching

References

- Ding, J., Ma, Z., Wu, Y., and Xu, J. (2021). Efficient random graph matching via degree profiles. *Probability Theory and Related Fields*, 179(1):29–115.
- Fan, Z., Mao, C., Wu, Y., and Xu, J. (2019). Spectral graph matching and regularized quadratic relaxations i: The gaussian model. *arXiv preprint arXiv:1907.08880*.
- Ganassali, L. (2020). Sharp threshold for alignment of graph databases with gaussian weights. *arXiv preprint arXiv:2010.16295*.
- Ganassali, L., Lelarge, M., and Massoulié, L. (2019). Spectral alignment of correlated gaussian random matrices. arxiv e-prints, page. *arXiv preprint arXiv:1912.00231*.
- Mao, C., Rudelson, M., and Tikhomirov, K. (2021). Random graph matching with improved noise robustness. In *Conference on Learning Theory*, pages 3296–3329. PMLR.
- Wu, Y., Xu, J., and Sophie, H. Y. (2022). Settling the sharp reconstruction thresholds of random graph matching. *IEEE Transactions on Information Theory*.