Matching Wishart matrices via Umeyama algorithm

Shuyang Gong

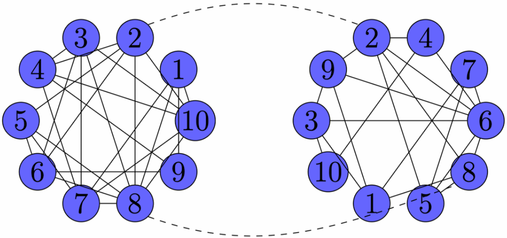
School of Mathematical Sciences, Peking University

September 9, 2024

Joint work with Zhangsong Li (PKU)

Motivations: random graph matching

- **Random Graph Matching** is an extensively studied topic in recent years, which lies in the intersection of **probability, statistics and theoretical computer science**.
- Goal: find a bijection between two vertex sets which maximizes the number of common edges (i.e. minimize the adjacency disagreements)
- $\max_{\pi \in \mathcal{S}_n} A_{i,j} B_{\pi(i),\pi(j)}$ and $\arg\max_{\pi \in \mathcal{S}_n} A_{i,j} B_{\pi(i),\pi(j)}$.



Model

- Sample n i.i.d Gaussian vectors $X_1, X_2, \ldots, X_n \sim \mathcal{N}(0, \mathrm{I}_d)$.
- Denote σ the noise parameter. Let $Z_i \sim N(0, \mathrm{I}_d)$ be independent of X_i's.
- Plant a latent permutation π^* (unknown). Define $Y_i = X_{\pi^*(i)} + \sigma Z_i$.
- The observations: $A = XX^\top$ and $B = YY^\top$.
- **Goal:** Is it possible to recover the latent permutation based on the observations? (i.e., for which values of σ is this possible?)

Informational upper bound

Define the estimator

$$\widehat{\Pi} := \arg\max_{\Pi \in \mathcal{S}_n} \max_{Q \in O(d)} \langle B^{1/2}, \Pi A^{1/2} Q \rangle , \tag{1}$$

where $A = U^\top \Lambda U$ and $A^{1/2} := U^\top \Lambda^{1/2}$.

Theorem (H. Wang, Y. Wu, J. Xu, I. Yolou, 22)

For $d = o(\log n)$, the following holds:

- *For $\sigma = o(n^{-1/d})$, we have*

$$\mathbb{P}\left(\frac{\text{dist}(\hat{\pi}, \pi^*)}{n} = o(1) \right) = 1 - o(1) .$$

- *For $\sigma = o(n^{-2/d})$, we have*

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

- Computation of $\widehat{\pi}$ requires $n^{O(d^2)}$ time, which is not efficient when $d = \omega(1)$. ($d = \omega(1)$ means $d = d_n \to \infty$)

Informational lower bound

For the "easier" model (linear assignment model), where the observations are X and $Y = \Pi^* X + \sigma Z$.

Theorem (H. Wang, Y. Wu, J. Xu, I. Yolou, 22)

- *For any* $\epsilon \in (0,1)$*, if there exists an estimator* $\widehat{\pi} = \widehat{\pi}(X, Y)$ *such that* $\mathbb{E}\text{dist}(\pi^*, \widehat{\pi}) \le \epsilon n$*, then*

$$\frac{d}{2} \log\left(1 + \frac{1}{\sigma^2}\right) - (1-\epsilon)\log n + 1 + \frac{\log(n+1)}{n} \ge 0\,.$$

- *Suppose that* $\sigma^2 \le d/40$ *and*

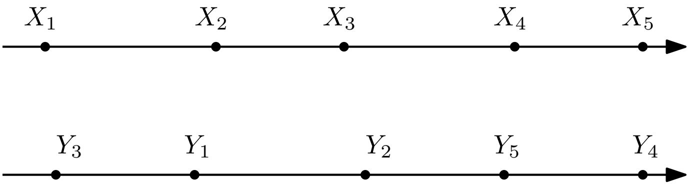
$$\frac{d}{4} \log\left(1 + \frac{1}{\sigma^2}\right) - \log n + \log d \le C, \tag{2}$$

for a constant $C > 0$*. Then there exists a constant* c *that only depend on* C *such that for any estimator* $\widehat{\pi}$*,* $\mathbb{P}\left(\widehat{\pi} \ne \pi^*\right) \ge c$*.*

- When $d = o(\log n)$, the necessary conditions become $\sigma = O\left(n^{-(1-\epsilon)/d}\right)$ and $\sigma \le n^{-2/d}$. When $d = O(1)$, there is **no sharp phase transition** in σ.

Intuition

- The "typical" minimal distance for a given vector X_i, $\min_{j:j\neq i} \|X_j - X_i\| \sim n^{-1/d}$.
- The minimal distance of all the vectors, $\min_{(i,j):i\neq j} \|X_i - X_j\| \sim n^{-2/d}$.
- To recover the permutation, the noise σ should not exceed the minimal distances (the correspondence is not "blurred").



$$\pi^* = (132)(45)$$

Computation: the Umeyama algorithm

- In [WWXY22], the authors proposed the Umeyama algorithm and conducted numerical experiments.
- Heuristic I: consider

$$\widehat{\Pi}_{\mathrm{Diag}(\{\pm 1\}^d)} := \arg\max_{\Pi \in \mathcal{S}_n} \max_{Q \in \mathrm{Diag}(\{\pm 1\}^d)} \langle B^{1/2}, \Pi A^{1/2} Q \rangle$$
$$= \arg\max_{\Pi \in \mathcal{S}_n} \max_{Q \in \mathrm{Diag}(\{\pm 1\}^d)} \langle V \Sigma^{1/2}, \Pi U \Lambda^{1/2} Q \rangle \,.$$

  By concentration of eigenvalues, we can define

$$\widehat{\Pi}_{Ume} := \arg\max_{\Pi \in \mathcal{S}_n} \max_{Q \in \mathrm{Diag}(\{\pm 1\}^d)} \langle V, \Pi U Q \rangle$$
$$= \arg\max_{\Pi \in \mathcal{S}_n} \max_{q \in \{\pm 1\}^d} \left\langle \Pi, \sum_{i=1}^{d} q_i v_i u_i^\top \right\rangle .$$

- Running time: $O(2^d n^3)$.

Computation: the Umeyama algorithm

Theorem (G. and Li 24+)

When $d = O(\log n)$, for $\widehat{\Pi}_{Ume}$ (output of the Umeyama algorithm) we have

- *if* $\sigma = o(d^{-3}n^{-2/d})$, $\mathbb{P}(\widehat{\Pi}_{Ume} = \Pi^*) = 1 - o(1)$.
- *if* $\sigma = o(d^{-3}n^{-1/d})$, $\mathbb{P}(\mathrm{dist}(\widehat{\Pi}_{Ume}, \Pi^*) = o(n)) = 1 - o(1)$.

- This result approaches the informational threshold up to a d^3 factor.
- For $d = O(1)$, there is **no information computation gap**. For $d = \omega(1)$, it remains an open question.

Heuristic II

(Assume $\Pi^* = \mathrm{Id}$, thus $Y = X + \sigma Z$)

- Expect $\langle X, Y \rangle$ to be the maximizer of all $\langle X, \Pi Y \rangle$'s.
- **Singular value decomposition:** $X = U\Lambda^{1/2}Q_1^\top$ and $Y = V\Sigma^{1/2}Q_2^\top$.
- $\langle X, Y \rangle = \langle U\Lambda^{1/2}Q_1^\top, V\Sigma^{1/2}Q_2^\top \rangle = \langle U\Lambda^{1/2}, V\Sigma^{1/2}Q_2^\top Q_1 \rangle$.
- Then $\widehat{\Pi}_{Ume}$ works if $Q_2^\top Q_1$ is approximately an element in $\mathrm{Diag}(\{\pm 1\}^d)$.

Proposition (G. and Li 24+)

There exists a $\Psi_0 \in \mathrm{Diag}(\{\pm 1\}^d)$ *such that*

$$\|Q_1\Psi_0 - Q_2\|_F \leq d^3\sigma\,.$$

Proof of proposition

- Write $Q_1 = [q_1^{(1)}, \dots, q_d^{(1)}]$ and $Q_2 = [q_1^{(2)}, \dots, q_d^{(2)}]$.
- By **Davis-Kahan** theorem

$$\sin \angle(q_i^{(1)}, q_i^{(2)}) \leq \frac{2\|X^\top X - Y^\top Y\|_{op}}{\min_{j:j\neq i} |\lambda_j - \lambda_i|}.$$

- By standard result, it can be shown $\|X^\top X - Y^\top Y\|_{op} \leq \sigma d\sqrt{n}$.
- For minimal spacing term, we compare the eigenvalues of a $d * d$ Wishart matrices to standard GOE. For GOE, the limiting distribution of the minimal spacing has been derived in [**Feng-Tian-Wei19**].
- We have $\min_{(i,j):i\neq j} |\lambda_i - \lambda_j| \geq \frac{1}{2d}\sqrt{n}$.

Summary

Take-home messages

- When $d = O(1)$, there is no information-computation gap for this problem.
- When $d = \omega(1)$, the Umeyama algorithm approaches the informational thresholds up to a poly(d) factor in the low-dimensional regime.
- When $d = \omega(1)$, it remains an intriguing question to answer if I-C gap phenomenon occurs.

Thank you!