Non-iterative clean up guarantee

Jiaxin Hu

March 22, 2022

In previous note 0306_22_proof, we do not consider the guarantee for the non-iterative clean up in Sub-Algorithm 1. This note provides the new theoretical guarantee for this non-iterative clean up procedure. This theoretical result explains why the non-iterative clean up procedure improves the performance in seeded matching, as shown in the simulation result in 0309_22_simulation. Previous condition for seeded matching may be relaxed since we do not include the clean up in the derivation of the previous theorem. We follow the proof idea of Lemma 20 in Ding et al. (2021) in this note.

To do list:

- Check how to relax previous theorem for seeded matching with the non-iterative clean up result.
- Find the literature for the large deviation inequality of Brownian motion $\mathbb{P}(\|B\|_p \geq \epsilon)$. Chanwoo and I agree that this inequality is the key to prove Conjecture 1.

For self-consistency, we write non-iterative clean up procedure as a separate Algorithm 1 here.

Algorithm 1 Non-iterative clean up

Input: Gaussian tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n^{\otimes m}}$ and permutation π_1.

1: For each pair $(i,k) \in [n]^2$, calculate $W_{ik} = \sum_{\omega \in [n]^{m-1}} \mathcal{A}_{i,\omega} \mathcal{B}_{k,\pi_1(\omega)}$.
2: Sort $\{W_{ik} : (i,k) \in [n]^2\}$ and let $\hat{S}$ denote the set of indices of largest n elements.
3: **if** there exists a permutation $\hat{\pi}$ such that $\hat{S} = \{(i, \hat{\pi}(i)) : i \in [n]\}$ **then**
4: Output $\hat{\pi}$.
5: **else**
6: Output error.
7: **end if**

Output: Estimated permutations $\hat{\pi}$ or error.

Theorem 0.1 (Guarantee for clean up)**.** *Suppose the input permutation π_1 has at most r fake pairs such that $(n-r)^{(m-1)/2} \gtrsim n^{(m-1)/4} \log^{1/4} n - \log^{1/2} n$. Then, the output of non-iterative clean up Algorithm 1 is equal to the true permutation with a high probability; i.e., $\hat{\pi} = \pi^*$ with a high probability as $n \to \infty$.*

Proof of Lemma 0.1. Without loss of generality, we assume π^* is the identity mapping. Let L denote the set of indices of the true pairs in π_1; i.e., $\pi(i) = i$ for all $i \in L$ and $|L| = \ell = n - r$. To show the Algorithm 1 picks π^* with a high probability, it suffices to show the following event holds with a high probability tends to 1 as $n \to \infty$:

$$\min_{i\in[n]} W_{ii} \geq \max_{i\neq k} W_{ik},$$

recalling that

$$W_{ik} = \sum_{\omega\in[n]^{m-1}} \mathcal{A}_{i,\omega}\mathcal{B}_{k,\pi_1(\omega)}.$$

Note that for an arbitrary $i \in [n]$, we have

$$W_{ii} = \sum_{\omega\in L^{m-1}} \mathcal{A}_{i,\omega}\mathcal{B}_{i,\pi_1(\omega)} + \sum_{\omega\in[n]^{m-1}/L^{m-1}} \mathcal{A}_{i,\omega}\mathcal{B}_{i,\pi_1(\omega)} =: W_1 + W_2,$$

where the variables $\mathcal{A}_{i,\omega}$ and $\mathcal{B}_{i,\pi_1(\omega)}$ are correlated with parameter ρ in the first term while $\mathcal{A}_{i,\omega}$ and $\mathcal{B}_{i,\pi_1(\omega)}$ are independent with each other in the second term. Hence, we have

$$\begin{aligned}
\mathbb{P}(W_{ii} < t_1) &\leq \mathbb{P}(W_1 < t_1 + t') + \mathbb{P}(W_2 < -t') \\
&= \mathbb{P}\left(\frac{W_1}{\ell^{m-1}} < \frac{t_1+t'}{\ell^{m-1}}\right) + \mathbb{P}\left(\frac{W_2}{n^{m-1}-\ell^{m-1}} \leq -\frac{t'}{n^{m-1}-\ell^{m-1}}\right) \\
&\leq 2\exp\left(-\min\left\{\frac{1}{32\rho^2}, \frac{1}{16(1-\rho^2)}\right\}\ell^{m-1}\left(\rho - \frac{t_1+t'}{\ell^{m-1}}\right)^2\right) \\
&\quad + \exp\left(-\frac{(t')^2}{4(n^{m-1}-\ell^{m-1})}\right),
\end{aligned}$$

for $\rho - \frac{t_1+t'}{\ell^{m-1}} \in [0, \min\{2\rho, 2\sqrt{2}\sqrt{1-\rho^2}\}]$ and $\frac{t'}{n^{m-1}-\ell^{m-1}} \in [0, \sqrt{2}]$, where the inequality follows from the Lemma 1.

Note that for all $i \neq k$

$$\mathbb{P}\left(W_{ik} > t_2\right) = \mathbb{P}\left(\frac{1}{n^{m-1}}W_{ik} > \frac{t_2}{n^{m-1}}\right) \leq \exp\left(-\frac{t_2^2}{4n^{m-1}}\right)$$

for $t_2/n^{m-1} \in [0, \sqrt{2}]$, where the inequality follows from the Lemma 1.

By union bound, we have

$$\begin{aligned}
&\mathbb{P}\left(\min_{i\in[n]} W_{ii} < t_1\right) \\
&\leq n\left[2\exp\left(-\min\left\{\frac{1}{32\rho^2}, \frac{1}{16(1-\rho^2)}\right\}\ell^{m-1}\left(\rho - \frac{t_1+t'}{\ell^{m-1}}\right)^2\right) + \exp\left(-\frac{(t')^2}{4(n^{m-1}-\ell^{m-1})}\right)\right]
\end{aligned} \tag{1}$$

and

$$\mathbb{P}\left(\max_{i\neq k} W_{ik} > t_2\right) \leq n^2\exp\left(-\frac{t_2^2}{4n^{m-1}}\right). \tag{2}$$

Now, we only need to verify that there exist proper $t_1 > t_2$ such that probabilities (1) and (2) tend to 0 as $n \to \infty$. We check the constraints for t', t_1, t_2, respectively.

For t', we have

$$\begin{cases} \frac{t'}{n^{m-1}-\ell^{m-1}} \in [0, \sqrt{2}] \\ \frac{(t')^2}{4(n^{m-1}-\ell^{m-1})} \geq \log n \end{cases} \quad \Rightarrow \quad 2\log^{1/2} n \left(n^{m-1} - \ell^{m-1}\right)^{1/2} \leq t' \leq \sqrt{2}\left(n^{m-1} - \ell^{m-1}\right), \quad (3)$$

where upper bound follows from Lemma 1 (first constraint), and the lower bound follows from the decay of probability (1) (second constraint).

For t_1, note that $\min\left\{\frac{1}{32\rho^2}, \frac{1}{16(1-\rho^2)}\right\} \geq \frac{1}{32}$, we have

$$\begin{cases} \rho - \frac{t_1+t'}{\ell^{m-1}} \in [0, \min\{2\rho, 2\sqrt{2}\sqrt{1-\rho^2}\}] \\ \frac{1}{32}\ell^{m-1}\left(\rho - \frac{t_1+t'}{\ell^{m-1}}\right)^2 \geq \log n \end{cases} \quad \Rightarrow f(\rho)\ell^{m-1} - t' \leq t_1 \leq \rho\ell^{m-1} - \sqrt{32}\log^{1/2} n\ell^{(m-1)/2} - t', \tag{4}$$

where $f(\rho) = \rho - \min\{2\rho, 2\sqrt{2}\sqrt{1-\rho^2}\}$, the upper bound follows from the decay of probability (1) (second constraint), and the lower bound follows from Lemma 1 (first constraint).

For t_2,

$$\begin{cases} \frac{t_2}{n^{m-1}} \in [0, \sqrt{2}] \\ \frac{t_2^2}{4n^{m-1}} \geq 2\log n \end{cases} \quad \Rightarrow \quad 4n^{(m-1)/2}\log^{1/2} n \leq t_2 \leq \sqrt{2}n^{m-1}, \tag{5}$$

where upper bound follows from Lemma 1 (first constraint), and the lower bound follows from the decay of probability (2) (second constraint).

Note that $\ell \in [n]$. When $\ell = n$, we have $n^{m-1} - \ell^{m-1} = 0$. When $\ell \neq n$, $\min\{n^{m-1} - \ell^{m-1}\} = n^{m-1} - (n-1)^{m-1} = \mathcal{O}(n^{m-1}) \gtrsim \log n$. Therefore, there always exists a t' satisfies (3).

Take $t' = 3\log^{1/2} n\left(n^{(m-1)/2} - \frac{\ell^{m-1}}{2n^{(m-1)/2}}\right)$, which satisfies constraint (3) since

$$\left(n^{m-1} - \ell^{m-1}\right)^{1/2} = n^{(m-1)/2}\left(1 - \frac{\ell^{m-1}}{n^{m-1}}\right)^{1/2} \leq n^{(m-1)/2}\left(1 - \frac{\ell^{m-1}}{2n^{m-1}}\right),$$

where the inequality follows by $(1-x)^{1/2} \leq 1 - x/2$ for $x \in [0,1]$. Then, to verify the existence of $t_1 > t_2$ under the constraints (4) and (5), we need to show the upper bound of t_1 is larger than the lower bound of t_2, which requires

$$4n^{(m-1)/2}\log^{1/2} n \leq \rho\ell^{m-1} - \sqrt{32}\log^{1/2} n\ell^{(m-1)/2} - 3\log^{1/2} n\left(n^{(m-1)/2} - \frac{\ell^{m-1}}{2n^{m-1}}\right).$$

This indicates

$$0 \leq \left(\rho + \frac{3\log^{1/2} n}{2n^{m-1}}\right)\ell^{m-1} - \sqrt{32}\log^{1/2} n\ell^{(m-1)/2} - 7n^{(m-1)/2}\log^{1/2} n.$$

Note that the coefficient for ℓ^{m-1} is dominated by ρ for ρ near to 1 when n is large enough. By the root formula of quadratic equation, when

$$\ell^{(m-1/2)} \geq \frac{-b + \sqrt{b^2 - 4ac}}{2a} = \mathcal{O}\left(n^{(m-1)/4}\log^{1/4} n - \log^{1/2} n\right),$$

there exist $t_1 > t_2$ to let the probabilities (1) and (2) tend to 0 as $n \to \infty$.

□

Lemma 1 (Tail bounds for the product of normal variables)**.** *Consider the correlated pairs of normal variables* (X_i, Y_i) *for* $i \in [n]$*, where* $X_i, Y_i \sim N(0,1)$*. Let* $H = \frac{1}{n}\sum_{i\in[n]} X_i Y_i$*. If* $cov(X_i, Y_i) = \rho > 0$*, then we have*

$$\mathbb{P}\left(|H - \rho| \geq t\right) \leq 4\exp\left(-\min\left\{\frac{1}{32\rho^2}, \frac{1}{16(1-\rho^2)}\right\} nt^2\right),$$

for constant $t \in [0, \min\{2\rho, 2\sqrt{2}\sqrt{1-\rho^2}\}]$*. If* $cov(X_i, Y_i) = 0$*, then, we have*

$$\mathbb{P}\left(|H| \geq t\right) \leq 2\exp\left(-\frac{nt^2}{4}\right),$$

for constant $t \in [0, \sqrt{2}]$.

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.