Non-iterative clean up guarantee v2

Jiaxin Hu

March 30, 2022

This note cleans the proof in 0321_22_cleanup. The theoretical result for the clean up theorem does not change.

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 $(n-r)^{(m-1)/2} \gtrsim n^{(m-1)/4}\log^{1/4} n$. 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 Theorem 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

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

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 $i \neq k$

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

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

By union bound, we have

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

and

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

Now, we only need to verify that there exist proper $t_1 > t_2$ such that probabilities (1) and (2) tend to 0 as $n \to \infty$. For simplicity, let $t_n = n^{(m-1)/2}\log^{1/2} n$ and by assumption $\ell^{m-1} \gtrsim t_n$.

Take

$$t_1 = \frac{\rho}{2}\ell^{m-1}, \quad t' = \begin{cases} 1 & \text{if } \ell = n \\ \min\{n^{m-1} - \ell^{m-1}, \frac{\rho}{4}t_n\} & \text{if } \ell \le n-1 \end{cases}.$$

Note that $t_1 \gtrsim t_n \gtrsim t'$ and thus t_1 dominates t'. Then, we have

$$\rho - \frac{t_1 + t'}{\ell^{m-1}} \in [0, \min\{2\rho, 2\sqrt{2}\sqrt{1-\rho^2}\}] \quad \text{and} \quad \frac{t'}{n^{m-1} - \ell^{m-1}} \in [0, \sqrt{2}].$$

Hence, we have

$$\min\left\{\frac{1}{32\rho^2}, \frac{1}{16(1-\rho^2)}\right\}\ell^{m-1}\left(\rho - \frac{t_1+t'}{\ell^{m-1}}\right)^2 \ge \frac{1}{32 \times 16}\ell^{m-1} \gtrsim \log n, \quad (3)$$

and

$$\frac{(t')^2}{4(n^{m-1}-\ell^{m-1})} \geq \begin{cases} \infty & \text{if } \ell = n \\ \min\{\frac{1}{4}(n^{m-1}-\ell^{m-1}), \frac{\rho^2}{16}\frac{t_n^2}{(n^{m-1}-\ell^{m-1})}\} & \text{if } \ell \leq n-1 \end{cases}. \tag{4}$$

Note that when $\ell \neq n-1$, $n^{m-1} - \ell^{m-1} \asymp n^{m-1} \gtrsim \log n$, and $\frac{t_n^2}{(n^{m-1}-\ell^{m-1})} \geq\gtrsim \frac{t_n^2}{n^{m-1}} \gtrsim \log n$. Plugging the lower bounds (3) and (4) into probability (1), we have probability (1) tends to 0 as $n \to \infty$.

Take $t_2 = 3t_n$. Then, we have

$$\frac{t_2}{n^{m-1}} \in [0, \sqrt{2}], \quad \text{and} \quad \frac{t_2^2}{4n^{m-1}} - 2\log n \geq \frac{1}{4}\log n,$$

which implies the probability (2) tends to 0 as $n \to \infty$.

Notice that by assumption $\ell^{m-1} \gtrsim t_n$, which implies $t_1 > t_2$ when n is large enough. We finished the proof of Theorem 0.1.

□

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 $M = \frac{1}{n}\sum_{i\in[n]} X_iY_i$. If cov$(X_i, Y_i) = \rho > 0$, then we have*

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

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

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

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

References