Permuted α-smooth tensors and two main interpretations of the sparsity parameter

Chanwoo Lee

April 15, 2021

1 Permuted α-smooth tensors

permutation-equivariance

A tensor $\Theta \in (\mathbb{R}^d)^{\otimes m}$ is called (α, L)-smooth, if for all $\omega, \omega' \in [d]^m$,

$$|\Theta(\omega) - \Theta(\omega')| \leq \frac{L\|\omega - \omega'\|_1^\alpha}{d^\alpha},$$

where $\|\cdot\|_1$ denotes the l_1 norm in $\mathbb{R}^m$. We use $\mathcal{P}$ to denote the family of permuted $\alpha-$smooth tensors,

$$\mathcal{P}(\alpha, L) = \{\Theta \circ \sigma \colon \Theta \text{ is an } (\alpha, L)\text{-smooth tensor}, \sigma \in S_d\}.$$

Then, for every integer $k \leq n$, there exists $z\colon [n] \to [k]$, satisfying

$$\frac{1}{|E|} \sum_{a \in [k]^m} \sum_{\omega \in E_{(z)^{-1}(a)}} (\Theta_\omega^{\text{true}} - \bar{\Theta}_a(z))^2 \leq L^2 \left(\frac{m}{k}\right)^{2\alpha}. \tag{1}$$

Proof. Define $z(i) = \ell$ for all $i \in [d]$ if $i \in \left[(\ell-1)\left\lceil\frac{d}{k}\right\rceil, \ell\left\lceil\frac{d}{k}\right\rceil\right]$. Notice that for any $\omega, \omega' \in z^{-1}(a) \subset [n]^d$ where $a = (a_1, \ldots, a_m) \in [k]^m$, we have $\|\omega - \omega'\|_1 \leq \|(\lceil d/k\rceil, \ldots, \lceil d/k\rceil)\|_1 \leq m\lceil d/k\rceil$ by definition. Therefore, for any $\omega \in z^{-1}(a)$, given $a \in [k]^m$,

approximation error
+ estimation error

1) block tensor to smooth tensor

2) estimation error:

E(Y) = block - tensor
error (hat theta - block-tensor);

E(Y)= smooth tensor
= (smooth tensor - block tensor) + block tensor

$$\begin{aligned}
|\Theta(\omega) - \bar{\Theta}_a(z)| &= \left|\Theta(\omega) - \frac{1}{|E_{z^{-1}(a)}|}\sum_{\omega' \in z^{-1}(a)} \Theta(\omega')\right| \\
&= \frac{1}{|E_{z^{-1}(a)}|}\sum_{\omega' \in z^{-1}(a)} |\Theta(\omega) - \Theta(\omega')| \\
&\leq \frac{L\|\omega - \omega'\|_1^\alpha}{d^\alpha} \\
&\leq \frac{Lm^\alpha \lceil d/k\rceil^\alpha}{d^\alpha} \\
&\lesssim L\left(\frac{m}{k}\right)^\alpha.
\end{aligned}$$

This entry-wise bound completes the proof. $\square$

Remark 1. If we assume that $|\Theta(\omega) - \Theta(\omega')| \leq \frac{L\|\omega - \omega'\|_\alpha^\alpha}{d^\alpha}$, the bound of (1) becomes $L^2 m^2 \left(\frac{1}{k^2}\right)^\alpha$

Remark 2. By the similar proof technique for α-smooth Hölder function, we show that

$$\frac{1}{n^m}\|\hat{\Theta} - \Theta\|_F^2 \leq C\left(m^\alpha L^2 n^{\frac{-2m\alpha}{m+2\alpha}} + \frac{\log n}{n^{m-1}}\right),$$

which is the exactly the same bound for $f \in \mathcal{H}(\alpha, L)$. Remember that we did not use the whole smoothness property of $f \in \mathcal{H}(\alpha, L)$ in the proof but that of observed entries. Therefore, all the results of α-smooth tensors are actually the same as hypergraphon model with α-Hölder class. When we consider the whole estimation such as $\mathbb{E}(\delta^2(\hat{f}_\Theta, f))$ including outside of tensor entries, we have the distinction between permuted α-smooth tensors and α -Hölder hypergraphon. In this case, we need to define what function value would be like outside of tensor entries for permuted α-smooth tensors.

2 Sparsity parameter versus sampling probability

2.1 Interpretation of sparsity parameter ρ

We assume that all probability of m-hyper edge being connected is represented as

$$f(\xi_{\omega_1}, \ldots, \xi_{\omega_m}) = \mathbb{P}\left(\mathcal{A}_\omega = 1 | \xi_{\omega_1}, \ldots, \xi_{\omega_m}\right) \text{ for all } \omega = (\omega_1, \ldots, \omega_m) \in E,$$

where $\xi_1, \ldots, \xi_n \overset{\text{i.i.d}}{\sim} \mathrm{U}[0,1]$. Let

$$\rho = \mathbb{P}(\text{Edge}) = \int_{[0,1]^m} f(u_1, \ldots, u_m) d\mu(u_1, \ldots, u_m),$$

where μ is the Lebesgue measure. Then the conditional density of $(\xi_{\omega_1}, \ldots, \xi_{\omega_m})$ given there is an edge among $\{\omega_1, \ldots, \omega_m\}$ is

$$W(\xi_{\omega_1}, \ldots, \xi_{\omega_m}) := \mathbb{P}(\xi_{\omega_1}, \ldots, \xi_{\omega_m} | A_\omega = 1) = \frac{f(\xi_{\omega_1}, \ldots \xi_{\omega_m})}{\rho},$$

by Bayes theorem. This reparametrization permits us to decouple ρ of the graph from the inhomogeneity structure. We let ρ depend on n and $w(\cdot, \cdots, \cdot)$ to be fixed. In addition, this reparametrization naturally impose the condition that $\int_{[0,1]^m} W(u_1, \ldots, u_m) d\mu(u_1, \ldots, u_m) = 1$. Therefore, it is more natural to set the probability tensor Θ^{true} as

$$\Theta^{\text{true}}_\omega := f(\xi_{\omega_1}, \ldots, \xi_{\omega_m}) = \rho W(\xi_{\omega_1}, \ldots, \xi_{\omega_m}).$$

This explicit interpretation of ρ is used in Bickel and Chen [2009], Bickel et al. [2011], Wolfe and Olhede [2013]. In this setting, ρ is estimated by

$$\hat{\rho} = \frac{1}{|E|} \sum_{\omega \in E} \mathbb{1}\{\mathcal{A}_\omega = 1\}.$$

Then, we estimate Θ^{true} as

$$\hat{\Theta} = \text{cut}(\hat{\rho}\tilde{\Theta}), \qquad \text{where } \tilde{\Theta} = \underset{\Theta \in \mathcal{P}_k}{\arg\min} \sum_{\omega \in E} |\mathcal{A}_\omega - \hat{\rho}\Theta_\omega|^2.$$

There is subtle distinction from ρ in Klopp et al. [2017]. Their ρ is not from internal characteristic of graphon f but external sampling distribution as in their interpretation saying *The model has been sparsified in the sense that its edges have been independently removed with probability 1-ρ and kept with probability ρ.* In this sense, their ρ is more like sampling probability ρ in the next subsection. However, ρ is not estimable in their interpretation (see next subsection).

2.2 Interpretation of sampling probabilitly ρ

Here we denote ρ sampling probability and interpret as,

$$\mathbb{P}\left[\mathcal{A}_\omega \text{ is observed } |\Theta^{\text{true}}_\omega\right] = \rho.$$

If we assign missing entries as 0 in $\mathcal{A}$, the marginal probability of observed network being connected has

$$\mathbb{P}(\mathcal{A}_\omega = 1 | \Theta^{\text{true}}_\omega) = \rho \Theta^{\text{true}}_\omega,$$

for all $\omega \in E$. In incomplete setting, we can easily estimate ρ as

$$\hat{\rho} = \frac{1}{|E|} \sum_{\omega \in E} \mathbb{1}\{\mathcal{A}_\omega = NA\}.$$

However, ρ is not estimable in Klopp et al. [2017] because we cannot distinguish 0 entries sparsified by sampling from entries that represents disconnection among nodes. What only we can do in this setting is to set ρ as known parameter.

In incomplete setting with estimated $\hat{\rho}$, we estimate Θ^{true} by

$$\hat{\Theta} = \text{cut}(\tilde{\Theta}), \quad \text{where } \tilde{\Theta} = \arg\min_{\Theta \in \mathcal{P}_k} \sum_{\omega \in E} |\mathcal{A}_\omega - \hat{\rho}\Theta_\omega|^2.$$

With adaptation of the new parameter ρ, we modify previous theorems incorporating ρ.

References

Peter J Bickel and Aiyou Chen. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073, 2009.

Peter J Bickel, Aiyou Chen, Elizaveta Levina, et al. The method of moments and degree distributions for network models. *The Annals of Statistics*, 39(5):2280–2301, 2011.

Olga Klopp, Alexandre B Tsybakov, Nicolas Verzelen, et al. Oracle inequalities for network models and sparse graphon estimation. *Annals of Statistics*, 45(1):316–354, 2017.

Patrick J Wolfe and Sofia C Olhede. Nonparametric graphon estimation. *arXiv preprint arXiv:1309.5936*, 2013.

Nest step: hyperparameter tuning:

1. choice of unknown K for alpha-smooth tensors (two-fold cross validation)
2. sub-roution — missing alpha-smooth tensor (—> application tensor completion)