Tensor block model and graphon estimation

Chanwoo Lee

February 15, 2021

1 Close relation between the graphon estimation and nonparametric regression without design information

This part is mostly based on Gao et al. [2015].

Consider the one-dimensional regression problem

$$y_i = f(\xi_i) + \epsilon_i, \quad i \in [n],$$

where $\{\xi_i\}$ are sampled from some $\mathbb{P}_\xi$, and ϵ_i are i.i.d. $N(0,1)$ variables. For Holder class with smoothness α, the minimax rate under the square loss is at order of $n^{-2\alpha/(2\alpha+1)}$ when the design $\{\xi_i\}$ is given [Tsybakov, 2008]. However, when the design $\{\xi_i\}$ is not observed, the minimax rate is at a constant order. For example, consider a problem $y_i = \theta_i + \epsilon_i$, for $i \in [n]$, where we assume θ_i can take only two possible values q_1 and q_2. Then we can show that

$$\inf_{\hat\theta} \sup_{\theta} \mathbb{E}\left\{ \frac{1}{n} \sum_{i\in[n]} (\hat\theta_i - \theta_i)^2 \right\} \asymp 1.$$

In contrast to the one-dimensional problem, a two dimensional nonparametric regression without knowing design is more informative. Consider

$$y_{ij} = f(\xi_i, \xi_j) + \epsilon_{ij}, \quad i,j \in [n], \tag{1}$$

where $\{\xi_i\}$ are sampled from some $\mathbb{P}_\xi$, and ϵ_i are i.i.d. $N(0,1)$. Consider the Holder class $\mathcal{H}_\alpha(M) = \{f\colon \|f\|_{\mathcal{H}_\alpha} \le M\}$ with Holder norm defined as

$$\|f\|_{\mathcal{H}_\alpha} = \max_{j+k\le\lfloor\alpha\rfloor} \sup_{x,y\in\mathcal{D}} |\Delta_{jk} f(x,y)| + \max_{j+k=\lfloor\alpha\rfloor} \sup_{(x,y)\ne(x',y')\in\mathcal{D}} \frac{|\Delta_{jk} f(x,y) - \Delta_{jk} f(x',y')|}{(|x-x'|+|y-y'|)^{\alpha-\lfloor\alpha\rfloor}},$$

where $\Delta_{jk} f(x,y) = \partial^{j+k} f(x,y)/(\partial x)^j (\partial y)^k$.

$$\mathcal{Y}_{ijk} = f(\xi_i, \xi_j, \xi_k) + \mathcal{E}_{ijk},$$

When the design $\{\xi_i\}$ is known, the minimax rate under the loss $\frac{1}{n^2}\sum_{i,j\in[n]}(\hat f(\xi_i,\xi_j) - f(\xi_i, xi_j))^2$ is at the order of $n^{-2\alpha/(\alpha+1)}$. When design is unknown, Gao et al. [2015] shows

$$\inf_{\hat f} \sup_{f\in\mathcal{H}_\alpha(M)} \sup_{\mathbb{P}_\xi} \frac{1}{n^2} \sum_{i,j\in[n]} (\hat f(\xi_i,\xi_j) - f(\xi_i,\xi_j))^2 \asymp \begin{cases} n^{-2\alpha/(\alpha+1)}, & 0<\alpha<1 \\ \frac{\log n}{n} & \alpha \ge 1. \end{cases}$$

The minimax rate is identical to that of graphon estimation which demonstrates the close relation between nonparametric regression and graphon estimation. The main reason for the difference between one-dimensional and the two dimensional problems is that the form (1) imposes more structure on the model so that the the lack of identifiability caused by the ignorance of design is only resulted from row and column permutation.

Conjecture on order-K tensors: the typical nonparametric rate is $N^{-2\alpha/(2\alpha+d)}$ where N is the number of observations and d is the function dimension. In order-K tensors, the number of observations $N \asymp d^K$ and the function dimension is K. So my conjecture of the rate for order-K is $d^{-K\alpha/(\alpha+K)}$ when $0<\alpha<1$. When $\alpha \ge 1$, my guess is $K\log d/d^K$ but I need to check the proof to see if the settings in order-K are different

d^(K-1) alpha + 1

from that in the paper.

2 Extension to tensor

Consider the three dimensional regression problem (or higher dimension),

$$y_{j_1,j_2,j_3} = f(\xi_{j_1}, \xi_{j_2}, \xi_{j_3}) + \epsilon_{j_1,j_2,j_3}, \quad j_k \in [d_k] \text{ for } k = 1,2,3, \tag{2}$$

where $\{\xi_i\}$ are sampled from some $\mathbb{P}_\xi$, and ϵ_i are i.i.d. $N(0,1)$. We cannot estimate f directly without observing the design. Instead, we estimate $\Theta_{j_1,j_2,j_3} = f(\xi_{j_1}, \xi_{j_2}, \xi_{j_3})$. We consider the following loss function:

$$\frac{1}{n^2} \sum_{j_1,j_2,j_3} (\hat{\Theta}_{j_1,j_2,j_3} - \Theta_{j_1,j_2,j_3})^2.$$

We consider block structure of $\{\Theta_{j_1,j_2,j_3}\}$, where we have

$$\Theta_{j_1,j_2,j_3} = \mathcal{S}_{(z_1)_{j_1},(z_2)_{j_2},(z_3)_{j_3}},$$

where $S \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ is the core tensor with collected block means, $z_k \in [r_k]^{d_k}$ are membership vectors of locations. Define the objective function

$$L(\mathcal{S}, z_1, z_2, z_3) = \sum_{j_1,j_2,j_3} \left(y_{j_1,j_2,j_3} - \mathcal{S}_{(z_1)_{j_1},(z_2)_{j_2},(z_3)_{j_3}} \right)^2.$$

1. MSE. conditinoal on xi
\hat f - f

For any optimizer of the objective function,

$$(\hat{\mathcal{S}}, \hat{z}_1, \hat{z}_2, \hat{z}_3) \in \underset{\mathcal{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}, z_k \in [r_k]^{d_k}}{\arg\min} L(\mathcal{S}, z_1, z_2, z_3) \tag{3}$$

2. integrated loss. integral with respect to xi.

Is the integrated loss the average?

the estimator of $f(\xi_{j_1}, \xi_{j_2}, \xi_{j_3})$ is defined as $\hat{f}(\xi_{j_1}, \xi_{j_2}, \xi_{j_3}) = \hat{\mathcal{S}}_{(\hat{z}_1)_{j_1},(\hat{z}_2)_{j_2},(\hat{z}_3)_{j_3}}$.

Estimation procedure of (3) follows the alternating optimization approach [Han et al., 2020]

Extreme case:
f = hat f at every discrete points.

1. update $\mathcal{S}$

$$\mathcal{S}_{i_1,i_2,i_3} = \text{Average}\left(\{y_{j_1,j_2,j_3} \colon (z_k)_{j_k} = i_k, \forall k \in [3]\}\right)$$

hat f(x) = hat S(x), for x = 0, 1, 2, …
hat f(x) = hat S(x_i, x_{i+1}) for x not in observation

2. update z_k for $k = 1,2,3$.

$$(z_k)_j = \underset{a \in [r_k]}{\arg\min} \|(\mathcal{M}_k(\mathcal{Y}_k))_{j:} - (\mathcal{M}_k(\mathcal{S}))_{a:}\|_F^2.$$

integrated loss between hat f and f
\int_x l\hat f (x) - f(x)l ~ alpha, P(x)

where $(\mathcal{Y}_1)_{j,i_2,i_3} = \text{Average}\left(\{y_{j,j_2,j_3} \colon (z_l)_{j_l} = i_l, l = 2,3\}\right)$, $(\mathcal{Y}_2)_{i_1,j,i_3} = \text{Average}\left(\{y_{j_1,j,j_3} \colon (z_l)_{j_l} = i_l, l = 1,3\}\right)$, and $(\mathcal{Y}_3)_{i_1,i_2,j} = \text{Average}\left(\{y_{j_1,j_2,j} \colon (z_l)_{j_l} = i_l, l = 1,2\}\right)$.

2.1 What to do

1. Checked the proof of (1) results and extend minimax upper bound of graphon estimation to the tensor case.
2. Find good examples where tensor extension (2) is advantageous.

References

Chao Gao, Yu Lu, Harrison H Zhou, et al. Rate-optimal graphon estimation. *Annals of Statistics*, 43(6): 2624–2652, 2015.

Rungang Han, Yuetian Luo, Miaoyan Wang, and Anru R Zhang. Exact clustering in tensor block model: Statistical optimality and computational limit. *arXiv preprint arXiv:2012.09996*, 2020.

Alexandre B Tsybakov. *Introduction to nonparametric estimation*. Springer Science & Business Media, 2008.