Polynomial-time estimation of permutation equivarant tensors

Miaoyan Wang, June 27, 2021

Consider an order-m permutation equivariant tensor

$$\mathcal{Y} = \Theta \circ \sigma + \mathcal{E}, \tag{1}$$

where $\mathcal{E} \in \mathbb{R}^{d\times\cdots\times d}$ is a symmetric mean-zero sub-Gaussian noise tensor, $\sigma\colon [d] \to [d]$ is an unknown permutation, and Θ is an unknown signal tensor sampled from a Lipschitz symmetric function with fixed design

$$\Theta(i_1,\ldots,i_m) = f\left(\frac{i_1}{d},\ldots,\frac{i_m}{d}\right), \quad \text{for all } (i_1,\ldots,i_m).$$

1 Subclass I: monotonic degree

Assumption 1 (β-monotonic degree)**.** We call a smooth tensor $\Theta \in \mathcal{P}(L)$ is degree-identifiable, if there exists a constant $\beta \in [0,1]$ and a small tolerance $\varepsilon_d \lesssim d^{-(m-1)/2}$ such that

$$\deg(i) - \deg(j) \gtrsim \left(\frac{i-j}{d}\right)^{1/\beta} - \varepsilon_d, \quad \text{for all } i \geq j \in [d]. \tag{2}$$

Remark 1. The condition (2) assumes the polynomial growth of population degree function up to a small error. The tolerance $\mathcal{O}(d^{-(m-1)/2})$ allows for small fluctuations within statistical accuracy. We call β the signal level, because it quantifies the identifiability of permutation from the degree. A lower value of β implies flatness of the function. We make the convention that a constant degree function is 0-monotonic.

Theorem 1.1 (Sorting-and-blocking under β-monotonicity)**.** Consider model 1 under Assumption 1. Consider the sorting-and-blocking algorithm with number of blocks $k = d^{\frac{m}{2+m}}$. When $\beta \geq \frac{2m}{(m-1)(m+2)}$, the algorithm output attains the optimal estimation rate

$$\mathcal{R}(\hat{\Theta}_{\text{LS}}, \Theta) \leq d^{-\frac{2m}{2+m}}.$$

2 Subclass II: non-constant fluctuation

Assumption 2 (β-detectable functions)**.** A function f is called weakly β-detectable, if the function f has at least a local $(1/\beta)$-polynomial fluctuation in each coordinate,

$$\max_{y\in[0,1]} \left| f\left(y,\ \boldsymbol{x}_{-1}\right) - f\left(y + d^{-1}, \boldsymbol{x}_{-1}\right) \right| \geq d^{-1/\beta} \quad \text{for all } \boldsymbol{x}_{-1} \in [0,1]^{m-1}, \tag{3}$$

where we use the shorthand $\boldsymbol{x}_{-1} = (x_2,\ldots,x_m)$ to denote the $(m-1)$ coordinates except the first one.

Remark 2. The exponent β quantifies the signal level of the function. A lower value of β implies global flatness of the function (low signal), whereas a high value of β implies polynomial fluctuation (high signal). By convention, a constant function has $\beta = 0$. We view the condition (3) as a mild non-degeneracy condition because it precludes nearly constant function in certain coordinates. In the latter case one may reduce the m-order tensor to the problem of $(m-1)$-order tensor.

Theorem 2.1 (Iterative blocking under β-detectability)**.** Consider model 1 under Assumption 2. Consider the iterative tensor block algorithm [1] with number of blocks $k = d^{\frac{m}{2+m}}$. When $\beta \geq 4/m$, the algorithm output attains the optimal estimation rate

$$\mathcal{R}(\hat{\Theta}_{\text{LS}}, \Theta) \leq d^{-\frac{2m}{2+m}}.$$

Corollary 2.1 (Blessing of orders for bi-lipschitz tensors)**.** Consider the smooth tensor model (1) with order $m \geq 4$. Furthermore, suppose f is bi-Lipschitz in that

$$0 < l \leq \frac{|f(\omega) - f(\omega')|}{\|\omega - \omega'\|_1} \leq L < \infty,$$

for two positive constants $l, L > 0$. Then, the algorithm output from [1] attains the optimal estimation rate.

3 Summary

Table 1 shows that the required signal level threshold β vanishes to zero as $m \to \infty$. Recall that a lower value of β implies less constrained function. Therefore, the required signal condition on β becomes weaker as the tensor order m increases.

Model class	MLE (theory)	Algorithm I	Algorithm II	NN smoothing
Assumption	-	local fluctuation	monotonic degree	-
Signal level	-	require $\beta \geq \frac{4}{m}$	require $\beta \geq \frac{2m}{(m-1)(m+2)}$	-
Rate	$d^{-2m/(2+m)}$	$d^{-2m/(2+m)}$	$d^{-2m/(2+m)}$	$d^{-\min(2m/(2+m),\ 2(m-1)/3)*}$

Table 1: Polynomial algorithms for smooth tensor estimation. *conjecture. $-$ none.

Questions:

1. We have described two polynomial algorithms and their successful regimes. Which assumptions are more relaxed?
2. Assumption 2 is derived based on the signal level requirement in [1]. See Figure 2 and Equations (11) and (12) in [1]. Please verify the sufficiency. Any better reformulation in our context?

References

[1] 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).