Discussion note: spectral method for smooth hypergraphon estimation

Chanwoo Lee (and Miaoyan Wang)

1 Rank-$\sqrt{d}$ approximatable

Definition 1 (Rank-$\sqrt{d}$ approximatable tensor)**.** Let Θ be an order-3 tensor. We use $f\colon [d] \to \mathbb{R}$ to denote the distance function, in the sense of matrix spectral norm $\|\mathcal{M}(\cdot)\|_{\text{sp}}$, between Θ and its rank-r projection,

$$f(r) = \inf\{\|\mathcal{M}(\Theta - \mathcal{A})\|_{\text{sp}}\colon \text{Rank}(\mathcal{A}) \leq (r,r,r)\}.$$

The tensor Θ is called rank-$\sqrt{d}$ approximatable, if $f(\sqrt{d}) \leq \sqrt{d}$. Geometrically, the intersection point between two curves $f(r)$ and $g(r) = r$ is smaller than $\sqrt{d}$.

Equivalently, Θ admits the decomposition

$$\Theta = \mathcal{A} + \mathcal{A}^{\perp}, \quad \text{s.t.} \quad \text{Rank}(\mathcal{A}) \leq (\sqrt{d}, \sqrt{d}, \sqrt{d}), \quad \text{and} \quad \|\text{Unfold}(\mathcal{A}^{\perp})\|_{\text{sp}} \leq \sqrt{d}. \tag{1}$$

Proposition 1 (Smooth matrix)**.** Every Lipschitz smooth matrix is rank-$\sqrt{d}$ approximatable.

Proof of Proposition 1. Let Θ be a Lipschitz smooth matrix. Set $\mathcal{A} = \text{Block}(\Theta, \sqrt{d})$ and $\mathcal{A}^{\perp} = \Theta - \text{Block}(\Theta, \sqrt{d})$. Then, by approximation theorem,

$$\|\text{Unfold}(\mathcal{A}^{\perp})\|_{\text{sp}} \leq \|\mathcal{A}^{\perp}\|_F \leq \sqrt{\frac{d^2}{d}} = \sqrt{d}.$$

Since $\mathcal{A}$ is of rank at most $\sqrt{d}$, the decomposition satisfies the condition (1). □

Conjecture 1 (Higher-order spectral algorithm)**.** Suppose Θ is an order-3, rank-$\sqrt{d}$ approximatable tensor. Then, the rank-$\sqrt{d}$ higher-order spectral algorithm [1] yields the estimate $\hat{\Theta}$ with error bound

$$\mathcal{R}(\hat{\Theta}, \Theta) := \frac{\|\hat{\Theta} - \Theta\|_F^2}{d^3} \lesssim d^{-1}.$$

Intuition: We decompose the error into estimation error and approximation bias

$$\begin{aligned}
\|\hat{\Theta} - \Theta\|_F^2 &\leq \|\hat{\Theta} - \mathcal{A}\|_F^2 + \|\mathcal{A}^{\perp}\|_F^2 \\
&\lesssim \underbrace{(d^{3/2}r + dr^2 + r^3)}_{\text{by Proposition 1 in [1]}} + \underbrace{d[f(r)]^2}_{\leq d^2 \text{ by Assumption 1}} \\
&\lesssim d^2 \text{ if } r \asymp \sqrt{d}.
\end{aligned}$$

The rank choice $\asymp \sqrt{d}$ is meaningful only in asymptotical sense. In practice, we should choose rank $C\sqrt{d}$ where the constant C may depend on actual Θ, noise, etc.

SBM (HOS+iteration)	sort-and-smoothing	square spectral	higher-order spectral (HOS)
$d^{-6/5}$	$d^{-6/5}$ (restricted model)	$d^{-2/3}$	d^{-1} (restricted model)

Table 1: Convergence rate for order-3 smooth tensors.

Questions 1. Unlike matrices, not every order-3 smooth tensor is $\sqrt{d}$-approximatable. How large is the order-3 tensor family that satisfy (1)? Does the signal tensor in our simulations satisfy (1)? How about general order-m tensors?

Remark 1. The HOS algorithm is applicable to approximate rank-$\sqrt{d}$ tensor. Unfortunately, the notion of "approximately" seems very restricted. We need the stringent assumption

$$\|\text{Unfold}(\mathcal{A}^{\perp})\|_{\text{sp}} \leq \sqrt{d}$$

to ensure bias is negligible compared to noise. For matrix problem, this term is unavoidable because noise matrix has $\sqrt{d}$ spectral norm. For tensor case, this assumption stems from the following lemma.

Lemma 1. (Perturbation Bound on Subspaces of Different Dimensions with deterministic bias). Consider the signal plus noise model,

$$Y = X + X_\perp + E \in \mathbb{R}^{d_1 \times d_2},$$

where X is a signal tensor such that $\mathrm{rank}(X) = r$, $X_\perp$ is a (deterministic) perturbation, and E is a noise matrix with i.i.d. standard sub-Gaussian entries. Define

$$r' = \max\{r' \in \{0, 1, \dots, r\} : \sigma_{r'}(X) \geq \max(\sqrt{d_1 d_2},\ (8 + 6\sqrt{2})\|X_\perp\|_{\mathrm{sp}})\}.$$

We denote

$$\hat{U}_r = \mathrm{SVD}_r(Y), \quad U_{r'} = \mathrm{SVD}_{r'}(X).$$

Then with probability at least $1 - \exp(-cd)$, we have

$$\|\hat{U}_{r,\perp} U_{r'}\|_{\mathrm{sp}} \leq \frac{\sqrt{d_1} + \|X_\perp\|_{\mathrm{sp}}}{\sigma_{r'}(X)} + \frac{(\sqrt{d_1} + \|X_\perp\|_{\mathrm{sp}})(\sqrt{d_2} + \|X_\perp\|_{\mathrm{sp}})}{\sigma^2_{r'}(X)}.$$

Bad news is that we have to require $\|X_\perp\|_{\mathrm{sp}} \leq \sqrt{d_1}$ in order to alleviate the error in the left singular space.

2 Block approximatable

Based on the proof of [1, Proposition 1], Conjecture 1 also applies to the block approximatable tensor. More generally, the following tensor family is the regime for which HOS algorithm works.

Definition 2 (Block-d^β approximatable tensor)**.** An order-m tensor Θ is called block approximatable with index $\beta \in [0, 1]$ if it admits the decomposition $\Theta = \mathcal{A} + \mathcal{A}^\perp$ satisfying the following two constraints:

1. $\mathcal{A}$ is a d^β-block tensor;
2. $\mathcal{A}^\perp$ has controlled spectral complexity in that

$$\|\mathrm{Unfold}(\mathcal{A}^\perp)\|_{\mathrm{sp}} \leq \sqrt{d}. \tag{2}$$

By definition, every tensor is block approximatable with trivial $\beta = 1$. We make the convention that β denotes the minimal block complexity in the decomposation for which the residual tensor satisfy (2).

Proposition 2 (Examples)**.**

- Every Lipschitz smooth matrix is block approximatable with $\beta = 1/2$;
- Every low-rank tensor with bounded factors has $\beta = 0$ (conjecture).
- Gaussian random tensor has $\beta \to 1$ for every $m \geq 2$ (conjecture).

Conjecture 2. Suppose Θ is a block approximatable tensor with $\beta \leq \beta_*$, where

$$\beta_* = \begin{cases} \frac{1}{3}, & \text{when } m = 2; \\ \frac{1}{2}, & \text{when } m = 3; \\ \frac{m}{m+2}, & \text{when } m \geq 4, \end{cases} \tag{3}$$

Then the HOS algorithm in [1] with rank $r = d^{\beta_*}$ gives the estimator $\hat{\Theta}$ with error rate

$$\begin{aligned} \mathcal{R}(\Theta, \hat{\Theta}) &\leq d^{-m}\left\{d^{\frac{m}{2}+\beta} + d^{\beta m} + \min(d^{m-2\beta}, d^{\frac{m}{2}+1})\right\} \\ &\leq \begin{cases} d^{-\frac{2}{3}}, & \text{when } m = 2; \\ d^{-1}, & \text{when } m = 3; \\ d^{-\frac{2m}{m+2}}, & \text{when } m \geq 4. \end{cases} \end{aligned}$$

If $\beta \geq \beta_*$, then we should take $r = d^\beta$, so

$$\mathcal{R}(\Theta, \hat{\Theta}) \leq \begin{cases} d^{-1+\beta}, & \text{when } m = 2; \\ d^{-3/2+\beta}, & \text{when } m = 3; \\ d^{-m(1-\beta)}, & \text{when } m \geq 4; \end{cases}$$

3 Intuition

- Oracle risk:

$$\underbrace{r^m}_{\text{block mean}} + \underbrace{d\log r}_{\text{block position}} \asymp \underbrace{\frac{d^m}{r^2}}_{m\text{-way approximation}}$$

  Therefore, the best $r \asymp d^{\frac{m}{m+2}}$. When $m = 2$ (matrix), $r = \sqrt{d}$.

- Oracle Spectral risk:

$$\underbrace{dr + r^m}_{\text{d.f. in spectral method}} \asymp \underbrace{\frac{d^m}{r^2}}_{m\text{-way approximation}}.$$

  When $m = 2$, the left hand side is computable by matrix SVD. The best $r = d^{1/3}$.

  When $m \geq 3$, no polynomial time algorithm is able to solve exact SVD. The best-so-far polynomial algorithm increases the risk to

$$\underbrace{d^{m/2}r + r^m}_{\text{d.f. in spectral method}} \asymp \underbrace{\frac{d^m}{r^2}}_{m\text{-way approximation}}.$$

  Notice the extra cost one has to pay on d when $m \geq 3$. The best r is solved in (3).

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).