Another polynomial-time estimation algorithm

Miaoyan Wang, July 8, 2021

Assumption 1 ($\sqrt{d}$-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 $\|\mathcal{M}(\cdot)\|_{\mathrm{sp}}$, between Θ and its rank-r projection,

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

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

Equivalently, Θ admits the decomposition

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

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

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

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

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

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

$$\mathcal{R}(\hat{\Theta}, \Theta) \lesssim d^{-1}.$$

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

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

More careful analysis is needed though, e.g. additive Gaussian vs. Bernoulli models, non-uniqueness of $\mathcal{A}$ and its singular space, etc. Also, 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 (HOC)
$d^{-6/5}$	$d^{-6/5}$ (for restricted model)	$d^{-2/3}$	d^{-1} (for restricted model)

Table 1: Convergence rate for order-3 tensor.

Remark 2. Based on the proof of [1, Proposition 1], the $\sqrt{d}$-approximatable assumption (1) can be replaced to the following two assumptions:

1. $\mathcal{A}$ is a block tensor;

2. $\mathcal{A}^{\perp}$ has controlled spectral complexity in that

$$\|\mathcal{A}^{\perp}\|_{\mathrm{sp}} \le \sqrt{d}, \quad \text{and} \quad \|\mathrm{Unfold}(\mathcal{A}^{\perp})\|_{\mathrm{sp}} \le d^{3/4}. \tag{3}$$

Correspondingly, the second term in (2) needs replaced by d^3/r^2. We did not present this alternative form because it has less intuitive interpretation. Note that (3) implies (1) because $\|\mathcal{A}^{\perp}\|_{\mathrm{sp}} \le \|\mathrm{Unfold}(\mathcal{A}^{\perp})\|_{\mathrm{sp}}$ [2].

Unlike matrices, not every order-3 smooth tensor is $\sqrt{d}$-approximatable. How large is the order-3 tensor family that satisfy (1) and (3)? How to generate to order-m tensor?

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).

[2] Miaoyan Wang, Khanh Dao Duc, Jonathan Fischer, and Yun S Song, *Operator norm inequalities between tensor unfoldings on the partition lattice*, Linear Algebra and Its Applications **520** (2017), 44–66.