Infinitely smooth single index model for tensors

Miaoyan Wang, Nov 09, 2021

1 Model

We use $f^{(\ell)}(x)$ to denote the ℓ-th derivative of f, evaluated at x.

Definition 1 (Analytic function class)**.** Let $C > 0$ be a positive constant. The analytic function class $\mathscr{F}(C)$ on $[-1, 1]$ is defined as the set of functions $f \colon [-1, 1] \to \mathbb{R}$ whose derivatives satisfy

$$\sup_{x\in[-1,1]} |f^{(\ell)}(x)| \leq C^{\ell+1}\ell! \quad \text{for all } \ell \in \mathbb{N}_+.$$

Equivalently, f is infinitely differentiable and its Taylor expansion around any point in its domain converges to the function.

A higher-order tensor $\mathcal{T}$ can be unfolded into a matrix. We now introduce several quantities that controls the complexity of matrix unfolding. We use $\text{Mat}(\mathcal{T})$ to denote the matrix unfolding. We use $\text{Trank}(\cdot)$ and $\lambda(\mathcal{T})$ the rank and spectral norm of the matrix $\text{Mat}(\mathcal{T})$,

$$\text{Trank}(\mathcal{T}) := \text{rank}(\text{Mat}(\mathcal{T})), \quad \lambda(\mathcal{T}) := \|\text{Mat}(\mathcal{T})\|_{\text{sp}}.$$

The following family consists of d-dimensional order-m tensor with Tucker rank bounded by r.

Definition 2 (Low-rank tensor class)**.** The family of d-dimensional rank-s tensor $\mathscr{T}(d, s, m)$ is defined as the set of tensors with Tucker rank bounded by r:

$$\mathscr{T}(d, s, m) = \{\mathcal{T} \in (\mathbb{R}^d)^{\otimes m} \colon \text{Trank}(\mathcal{T}) \leq s \text{ and } \lambda(\mathcal{T}) \leq 1\}$$

Equivalently, the tensor in class $\mathscr{T}(d, s, m)$ admits the rank-s Tucker decomposition:

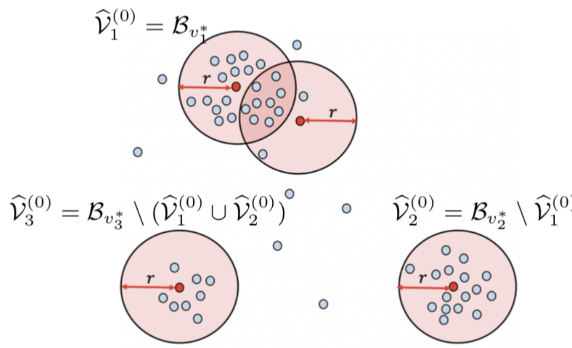
$$\mathcal{T} = \mathcal{C} \times_1 \boldsymbol{X} \times \cdots \times_k \boldsymbol{X}.$$

where $\mathcal{C} \in \mathbb{R}^{s\times\cdots\times s}$ is core tensor and $\boldsymbol{X}$ are factor matrices with orthornormal columns. The condition $\lambda(\mathcal{T}) \leq 1$ is imposed without loss of generality. The scale between $\mathcal{T}$ and the f is undetermined; the tensor $f(\mathcal{T}) = f'(\mathcal{T}')$ are the same by setting $\mathcal{T}' = c\mathcal{T}$ and $f' = f/c$. We also note that, the rank-s CP tensor is automatically included in $\mathscr{T}(d, s, m)$.

Now, we are ready to describe the main model. Let $\mathcal{Y}$ be the data tensor. We propose the following observation model

$$\mathcal{Y} = f(\mathcal{T}) + \mathcal{E}, \quad \text{for some unknown } \mathcal{T} \in \mathscr{T}(d, s, m) \text{ and } f \in \mathscr{F}(C), \tag{1}$$

where we assume the noise tensor $\mathcal{E}$ consists of i.i.d. entries with zero-mean and sub-Gaussian



parameter σ^2. We call (1) the *single index tensor model*. The name of single index model comes from the observation that

$$\mathbb{E}\left[\mathcal{Y}(\omega)|\mathcal{X}(\omega)\right] = f(\langle \mathcal{T},\ \mathcal{X}(\omega)\rangle), \quad \text{for all } \omega \in [d]^m,$$

where, for every index ω, the predictor $\mathcal{X}(\omega) \in (\mathbb{R}^d)^{\otimes m}$ is a dummy-variable tensor with 1 at the ω-th position, and zero everywhere else. Note that the signal tensor $f(\mathcal{T})$ is often high rank. Our goal is to address the following two questions:

- What are the *statistical* and *computational* limits for signal estimation in single index model?
- Are there any intrinsic distinctions for matrices $m = 2$ vs. tensors $m \geq 3$ for high-rank estimation based on model (1)?

2 Smooth single index models are of log-rank

Let $\mathcal{T} \in \mathscr{T}(d, s, m)$, and $\mathcal{T}^{\circ \ell}$ be the tensor of the same size, with entrywise polynomial transformation $a \to a^\ell$. We have the following rank bound.

Proposition 1 (Monomial tensor)**.** *For every tensor $\mathcal{T} \in \mathscr{T}(d, s, m)$ and every natural number $\ell \in \mathbb{N}_+$*

$$\text{Trank}(\mathcal{T}^{\circ \ell}) \leq \binom{\ell + s - 1}{s - 1}.$$

This proposition shows the polynomial rank growth with respect to ℓ. The exponent depends on the original rank s, which is assumed small. When $s = 1$, then $\text{Trank}(\mathcal{T}^{\circ \ell}) = 1$ for all $\ell \in \mathbb{N}_+$. The bound is nontrivial when $\ell \leq d$. In fact, later we will set $\ell \lesssim \log d$.

Proposition 2 (Uniform approximation for single index tensor)**.** *Consider $\mathcal{T} \in \mathscr{T}(d, s, m)$ and $f \in \mathscr{F}(C)$. There exists a set of basis tensors $\{\mathcal{B}_{\ell,k} \colon (k, \ell) \in [d] \times \mathbb{N}\}$, such that, for every number* *of pieces $k \in [d]$ and degree $\ell \in \mathbb{N}$, we have*

$$\|f(\mathcal{T}) - \mathcal{B}_{k,\ell}\|_\infty \leq C \left(\frac{k}{sm}\right)^{-(\ell+1)} \quad \textit{and} \quad \operatorname{Trank}(\mathcal{B}_{k,\ell}) \leq k^s \ell^s.$$

The following is the key property of single index matrix by taking $k \gtrsim \Omega(sm)$ and $\ell = r^{1/s}k^{-1}$ in Proposition 2.

Theorem 2.1 (Dimension-free approximation error)**.** *Consider a single index tensor $\Theta = f(\mathcal{T})$, where $\mathcal{T} \in \mathscr{T}(d,s,m)$ and $f \in \mathscr{F}(C)$. For every rank $r \in \mathbb{N}_+$, we have*

$$\frac{1}{d^m}\|\Theta - \operatorname{Proj}_r(\Theta)\|_F^2 \lesssim C^2 \exp\left(-\frac{r^{1/s}}{sm}\right).$$

The bound is uniform over $\mathcal{T} \in \mathscr{T}(d,s,m)$ and $f \in \mathscr{F}(C)$.

Corollary 1 (Nice smooth tensor model is of log rank)**.** *For any fixed $\varepsilon > 0$, we have*

$$\operatorname{Trank}_\varepsilon(\Theta) := \min\{\operatorname{Trank}(\mathcal{A})\colon \|\mathcal{A} - \Theta\|_\infty \leq \varepsilon\} \lesssim \log^s d.$$

3 Estimation algorithm

3.1 Non-convex double spectral algorithm

Theorem 3.1 (Polynomial algorithm)**.** *low-rank approximation is substantially better.*

$$\mathcal{R}(\hat{\mathbf{\Theta}}, \mathbf{\Theta}) \leq \begin{cases} r^m + d^{m/2}r + d^m r^{-2\alpha} & r^{-2\alpha} \leq d^{-(m-1)}, \\ \min(d^{-\frac{m}{2}+\frac{(m-\alpha-1)m}{2\alpha}}, 1) & \alpha \leq \frac{(m-1)^2}{m},\ r \asymp \min(d^{(m-1)/2\alpha}, d) \\ d^{-\frac{m}{2}+\frac{m-1}{2\alpha}} & \alpha \geq \frac{(m-1)^2}{m},\ r \asymp d^{\frac{m-1}{2\alpha}} \leq d^{\frac{m}{2(m-1)}} \\ d^{-\frac{m}{2}}\log d & \alpha = \infty,\ r \asymp \log d \end{cases}$$

Theorem 3.2 (Minimax optimality for tensors)**.** *Order-m with known design (bottleneck: truncate at $d^{-(m-1)}$ with unknown design)*

$$\mathcal{R}(\hat{\mathbf{\Theta}}, \mathbf{\Theta}) \leq \begin{cases} d^{-\frac{2\alpha m}{m+2\alpha}} & \alpha \leq \frac{m(m-1)}{2} \\ d^{-(m-1)}\log d & \alpha \geq \frac{m(m-1)}{2} \end{cases}$$

A Proof of Proposition 1

Proof. By definition $\mathcal{T} \in \mathscr{T}(d,s,m)$, $\mathrm{Trank}(\mathrm{Mat}(\mathcal{T})) \le r$. Therefore, there exists matrix SVD such that

$$\mathrm{Mat}(\mathcal{T}) = \sum_{i\in[s]} \lambda_i \boldsymbol{a}_i \otimes \boldsymbol{b}_i,$$

where $\boldsymbol{a}_i \in \mathbb{R}^d$, $\boldsymbol{b} \in \mathbb{R}^{d^{m-1}}$, and $\lambda_1 \ge \cdots \lambda_s \ge 0$. By definition

$$\begin{aligned}\mathrm{Mat}(\mathcal{T}^{\circ\ell}) = [\mathrm{Mat}(\mathcal{T})]^{\circ\ell} &= \left(\sum_{i\in[s]} \lambda_i \boldsymbol{a}_i \otimes \boldsymbol{b}_i\right)^{\circ\ell} \\ &= \sum_{\substack{\kappa_1+\cdots+\kappa_s=\ell,\\ (\kappa_1,\ldots,\kappa_s)\in\mathbb{N}_+^s}} \lambda_1^{\kappa_1}\cdots\lambda_s^{\kappa_s}(\boldsymbol{a}_1^{\circ\kappa_1}\circ\cdots\circ\boldsymbol{a}_s^{\circ\kappa_s})\otimes(\boldsymbol{b}_1^{\circ\kappa_1}\circ\cdots\circ\boldsymbol{b}_s^{\circ\kappa_s}).\end{aligned} \tag{2}$$

Here $(\boldsymbol{a}_1^{\circ\kappa_1}\circ\cdots\circ\boldsymbol{a}_s^{\circ\kappa_s}) \in \mathbb{R}^d$ and $(\boldsymbol{b}_1^{\circ\kappa_1}\circ\cdots\circ\boldsymbol{b}_s^{\circ\kappa_s}) \in \mathbb{R}^{d-1}$. Now notice that by counting argument,

$$\#\{(\kappa_1,\ldots,\kappa_s)\in\mathbb{N}_+^s \colon \kappa_1+\cdots+\kappa_s=\ell\} = \binom{\ell+s-1}{s-1}.$$

Therefore, the summation (2) consists of no more than $\binom{\ell+s-1}{s-1}$ rank-1 terms. We conclude that

$$\mathrm{Trank}(\mathcal{T}^{\circ\ell}) = \mathrm{rank}(\mathrm{Mat}(\mathcal{T}^{\circ\ell})) \le \binom{\ell+s-1}{s-1}.$$

□

B Proof of Lemma 2

Lemma 1. *we have*

$$\|f(\mathcal{T}) - \mathcal{B}_{k,\ell}\|_\infty \le \frac{\max_{|\eta|\le 1}|f^{(\ell+1)}|(\eta)}{(\ell+1)!}\|(\mathcal{T}-\mathcal{O})^{\circ\ell+1}\|_\infty \le \left(\frac{k}{sm}\right)^{-(\ell+1)}.$$

Note that the covering number of s-dimensional bounded set $\mathcal{X}$ is $\mathcal{N}(1/k,\mathcal{X},\|\cdot\|_\infty) \le k^s$. Let $\mathcal{E}_k$ denote the corresponding covering set. Then, $\mathcal{E}_k$ satisfies

1. $|\mathcal{E}_k| \le k^s$;
2. For every $\Delta \in \mathcal{E}_k$, we have
$$\max_{\boldsymbol{x}_i,\boldsymbol{x}_j\in\Delta}\|\boldsymbol{x}_i-\boldsymbol{x}_j\|_\infty \lesssim \frac{1}{k}.$$
3. $\Delta\cap\Delta' = \emptyset$ for all $\Delta \ne \Delta' \in \mathcal{E}_k$.

We label the center of the covering set by $\{\boldsymbol{o}_1, \ldots, \boldsymbol{o}_{k^s}\} \subset \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d\}$. We use $z\colon [d] \to [k^s]$ to denote the membership of rows of $\boldsymbol{X}$

$$z(i) = \arg\min_{j \in [k^s]} \|\boldsymbol{x}_i - \boldsymbol{o}_j\|_\infty, \quad \text{such that} \quad \|\boldsymbol{x}_i - \boldsymbol{o}_{z(i)}\|_\infty \leq \frac{1}{k}.$$

We define a block matrix $\mathcal{O} \in (\mathbb{R}^d)^{\otimes m}$ with entries

$$\mathcal{O}(i_1, \ldots, i_m) = \mathcal{C} \times_1 \boldsymbol{o}_{z(i_1)} \times_2 \cdots \times_m \boldsymbol{o}_{z(i_m)}$$

Therefore

$$\|\mathcal{T} - \mathcal{O}\|_\infty \leq \max_{(i_1, \ldots, i_m) \in [d]^m} |\mathcal{C} \times_1 \boldsymbol{x}_{i_1} \times_2 \cdots \times_m \boldsymbol{x}_{i_m} - \mathcal{C} \times_1 \boldsymbol{o}_{z(i_1)} \times_2 \cdots \times_m \boldsymbol{o}_{z(i_m)}| \lesssim mk^{-1}s^{1/2}.$$

Define

$$\mathcal{B}_{k,\ell} := f^{(1)}(\mathcal{O}) \circ (\mathcal{T} - \mathcal{O}) + \frac{f^{(2)}(\mathcal{O})}{2} \circ (\mathcal{T} - \mathcal{O})^{\circ 2} + \cdots \frac{f^{(\ell)}(\mathcal{O})}{\ell!} \circ (\mathcal{T} - \mathcal{O})^{\circ \ell}.$$

The membership partition $[d]^m$ into k^{sm} blocks. We use $[d]^m = \cup_{n=1}^{k^{sm}} \Delta_n$ to denote these blocks. Within each block, the tensor $\mathcal{O}$ takes the same value,

$$\mathcal{B}_{k,\ell}(\omega) = \underbrace{(a_{0,\Delta} + a_{1,\Delta}\mathcal{T} + a_{2,\Delta}\mathcal{T}^{\circ 2} + \cdots a_{\ell,\Delta}\mathcal{T}^{\circ \ell})}_{:=\mathcal{I}} \mathbb{1}\{\omega \in \Delta\}.$$

Because there are k^s blocks along mode 1, we conclude that

$$\text{Trank}(\mathcal{B}_{k,\ell}) \leq k^s \sum_{n=1}^{\ell} \binom{n+s-1}{s-1} \leq k^s(\ell + s - 1)^s.$$

Theorem B.1 (Dimension-free approximation)**.** *Let $\Theta \in \mathcal{P}(d, s, L)$. For every rank $r \in \mathbb{N}_+$,*

$$\frac{1}{d^m}\|\Theta,\ \textit{Proj}_r(\Theta)\|_F^2 \lesssim L^2 \exp(-c_0 r^{1/s}).$$

Proof. Let $k = \Omega(2a)$ and $\ell = r^{1/s}k^{-1}$. Then

$$\|\Theta - \mathcal{A}_{k,\ell}\|_\infty \leq L2^{-\ell}\ell^{ms} \lesssim L2^{-\ell} \asymp L\exp(-r^{1/s})$$

□