Current truncation is L-1 truncation: threshold on singular values.
How about using L-0 truncation: threshold on rank.
Set r_k = sqrt(d) from the beginning. Then all hat U, check U, …., are easier to handle.

Some details of high-order spectral method

Chanwoo Lee, July 12, 2021

Let us consider $\mathcal{Y} = \mathcal{X} + \mathcal{E} \in \mathbb{R}^{d_1 \times \cdots \times d_m}$ where $\mathcal{E}$ follows i.i.d. sub-Gaussian noise with $\sigma^2 = 1$ without loss of generality. For each $k = 1, ..., n$, denote

$$X_k = \mathcal{M}_k(\mathcal{X}), \quad E_k = \mathcal{M}_k(\mathcal{E}), \quad Y_k = \mathcal{M}_k(\mathcal{Y}).$$

We consider the high-order spectral method with threshold. Define the $r_k := \max_{r' \in [d_k]}\{\sigma_{r'}(Y_k) \geq c_0 d^{m/4}\}$ for all $k \in [m]$. Based on threshold r_k, for all $k \in [m]$, define

$$\tilde{U}_k = \text{SVD}_{r_k}(Y_k)$$
$$\hat{U}_k = \text{SVD}_{r_k}\left(\mathcal{M}_k\left(\mathcal{Y} \times_1 \tilde{U}_1^T \times \cdots \times_{k-1} \tilde{U}_{k-1}^T \times_k \tilde{U}_{k+1}^T \times \cdots \times_m \tilde{U}_m^T\right)\right).$$

Then we estimate true $\mathcal{X}$ by

$$\hat{\mathcal{X}} := \mathcal{Y} \times_1 (\hat{U}_1 \hat{U}_1^T) \times \cdots \times_m (\hat{U}_m \hat{U}_m^t). \tag{1}$$

Then the following (psuedo) theorem holds.

Theorem 0.1 (Tensor estimation error bound)**.** Suppose $\mathcal{Y} = \mathcal{X} + \mathcal{E} \in \mathbb{R}^{d_1 \times \cdots \times d_m}$ where $\mathcal{E}$ follows i.i.d. sub-Gaussian noise with $\sigma^2 = 1$ without loss of generality. Let $\hat{\mathcal{X}}$ defined according to (1). Then, with probability at least $1 - C\exp(-cd)$,

d^{(K-1)/2}

$$\|\hat{\mathcal{X}} - \mathcal{X}\|_F \lesssim \sqrt{\prod_k r_k} + \sqrt{\max_k d_k \max_k r_k} + \sqrt{\sum_{i=r_k+1}^{d_k} \sigma_i^2(Y_k)} + \sqrt{\sum_{i=r_k+1}^{d_k} \sigma_i^2(X_k)}.$$

A hot topic on random matrix theory.
Under spectral norm condition sigma(E_k) < d^{K/4}, the two in yellow differ by an additive error d^{m/4}.

Remark 1. We need to bound $\sqrt{\sum_{i=r_k+1}^{d_k} \sigma_i^2(Y_k)}$ with respect to $\sqrt{\sum_{i=r_k+1}^{d_k} \sigma_i^2(X_k)}$. This can be done using the relationship $Y_k = X_k + E_k$ but not sure how to bound tightly. I will think about this more.

Remark 2. I guess that $\sqrt{\sum_{i=r_k+1}^{d_k} \sigma_i^2(X_k)}$ will have the order of $\mathcal{O}\left(\frac{d^m}{r^2}\right)$ under some $\mathcal{X}$ classes such as α-smooth function classes.

Proof. We start by introducing several notations and assumptions. Denote $U_k = \text{SVD}_{r_k}((X_k)$ and $\tilde{U}_k = \text{SVD}_{r_k}(Y_k)$. For some constant $c_1 > c_0$ which will be specified later, define

$$r'_k = \max\{r' \in \{0, \ldots, d_k\} : \sigma_{r'}(X_k) \geq c_1 d^{m/4}\}.$$

We set $r' = 0$ if $\sigma_1(X_k) < c_1 d^{m/4}$ We use U_k to denote the leading r_k singular vectors of U_k and use V_k to denote the rest $r_k - r'_k$ singular vectors and thus U_k can be written as $[U_k, V_k]$. We next define

$$X'_k = X_k\left(\mathbb{P}_{U'_{k+1}} \otimes \cdots \otimes \mathbb{P}_{U'_m} \otimes \cdots \otimes \mathbb{P}_{U'_{k-1}}\right),$$

where $\mathbb{P}_U = UU^T$ for any orthonormal matrix $U \in \mathbb{R}^{d \times r}$. We also denote

$$\bar{X}_k = X_k\left(\tilde{U}_{k+1} \otimes \cdots \otimes \tilde{U}_m \otimes \cdots \otimes \tilde{U}_{k-1}\right),$$

Set r = sqrt(d) as a pre-specified rank.

Write X = X(good)+X(bad). Assume
(1) X(good) has rank r; (2) [X(bad) + E] satisfies the 5 conditions in Lemma 0.1.
This is the intent of beta-approximatable assumption in my note.

$$\bar{Y}_k = Y_k \left(\tilde{U}_{k+1} \otimes \cdots \otimes \tilde{U}_m \otimes \cdots \otimes \tilde{U}_{k-1} \right),$$
$$\bar{E}_k = E_k \left(\tilde{U}_{k+1} \otimes \cdots \otimes \tilde{U}_m \otimes \cdots \otimes \tilde{U}_{k-1} \right).$$

Now we bound

$$\|\hat{\mathcal{X}} - \mathcal{X}\|_F \le \underbrace{\|\mathcal{X} \times_1 (\hat{U}_1\hat{U}_1^T) \times \cdots \times_m (\hat{U}_m\hat{U}_m^T) - \mathcal{X}\|_F}_{(*)} + \underbrace{\|\mathcal{E} \times_1 \hat{U}_1^T \times \cdots \times_m \hat{U}_m^T\|_F}_{(**)}.$$

Notice that $(**)$ term is bounded by $C(\sqrt{\prod_k r_k} + \sum_{\ell\in[m]} \sqrt{d_\ell r_\ell})$. For $(*)$, we have

$$\begin{aligned}(*) &\le \sum_{k\in[m]} \|(I - \hat{U}_k\hat{U}_k^T)X_k\|_F \\ &= \|\hat{U}_{k,\perp}^T X_k\|_F \\ &\le \|\hat{U}_{k,\perp}^T X_K'\|_F + \|X_k - X_k'\|_F.\end{aligned} \tag{2}$$

geometric interpretation of the proof?

Therefore, it suffices to bound $\|\hat{U}_{k,\perp}^T X_k'\|_F$ and $\|X_k - X_k'\|_F$.

1. Bound of $\|\hat{U}_{k,\perp}^T X_k'\|_F$: Combining equation (44) and (45) in [1] gives us

$$\|\hat{U}_{k,\perp}^T \bar{X}_k\|_F + \|X_k - X_k'\|_F \ge \|\hat{U}_{k,\perp}^T X_k'\|_F \prod_{\ell\ne k} \sqrt{1 - \|\tilde{U}_{k,\perp} U_k'\|_{\text{sp}}^2} \tag{3}$$

For $\|\hat{U}_{k,\perp}^T \bar{X}_k\|_F$, Combination of Lemma 0.2 and the fact that $\bar{Y}_k = \bar{X}_k + \bar{E}_k$ yields,

$$\begin{aligned}\|\hat{U}_{k,\perp}^T \bar{X}_k\|_F &\le \sqrt{\sum_{i=r_k+1}^{d_k} \sigma_i^2(\bar{Y}_k)} + \|\bar{E}_k\|_F \\ &\le \sqrt{\sum_{i=r_k+1}^{d_k} \sigma_i^2(Y_k)} + C(\sqrt{d_k r_k}),\end{aligned} \tag{4}$$

where the last line uses the definition of $\bar{E}_k$ and Lemma 0.1.

For $\|\tilde{U}_{k,\perp} U_k'\|_{\text{sp}}^2$, we use the proof of Lemma 2 in [1]. The main difference is that we did not use rank of X_k but use the new definition of $r_k := \max_{r'\in[d_k]}\{\sigma_{r'}(Y_k) \ge c_0 d^{m/4}\}$. Adapting this fact, we can show that with high probability

$$\|\tilde{U}_{k,\perp} U_k'\|_{\text{sp}}^2 \le \frac{1}{\sqrt{2}}, \tag{5}$$

by carefully choosing constants $c_0, c_1 > 0$ (I will fill the details later).

Finally combining (3),(4), (5), we have

$$\|\hat{U}_{k,\perp}^T X_k'\|_F \lesssim 2^{(m-1)/2} \left(\sqrt{\sum_{i=r_k+1}^{d_k} \sigma_i^2(Y_k)} + \sqrt{d_k r_k} + \|X_k - X_k'\|_F \right). \tag{6}$$

2. Bound of $\|X_k - X_k'\|_F$: For notation simplicity, we focus on $k=1$, while the analysis for other modes can be similarly carried on.

$$\begin{aligned}
\|X_1 - X_1'\|_F &\leq \|X_1 \left((\mathbb{P}_{U_2'} + \mathbb{P}_{V_2'}) \otimes \cdots \otimes (\mathbb{P}_{U_m'} + \mathbb{P}_{U_m'}) - \mathbb{P}_{U_2'} \otimes \cdots \otimes \mathbb{P}_{U_m'} \right) \|_F \qquad (7)\\
&\leq \sum_{k=2}^{m} \|V_k'^T X_k\|_F \\
&\leq \sum_{k=2}^{m} \sqrt{\sum_{i=r_k+1}^{d_k} \sigma_i^2(X_k)} \\
&\leq m \sqrt{\sum_{i=r_k+1}^{d_k} \sigma_i^2(X_k)}
\end{aligned}$$

Finally, plugging (6) and (7) into (2) yields,

$$\|\hat{\mathcal{X}} - \mathcal{X}\|_F \lesssim \sqrt{\prod_k r_k} + \sqrt{\max_k d_k \max_k r_k} + \sqrt{\sum_{i=r_k+1}^{d_k} \sigma_i^2(Y_k)} + \sqrt{\sum_{i=r_k+1}^{d_k} \sigma_i^2(X_k)}.$$

$\square$

I like your current proof layout. Separating lemma 0.1 from main proof is good.

Lemma 0.1 (Lemma 8 in [1]). Let $E \in \mathbb{R}^{d_1 \times \cdots \times d_m}$ be a noise tensor whose each entry has independent mean-zero sub-Gaussian distribution with $\sigma = 1$ without loss of generality. Fix $U_k^* \in \mathbb{O}_{d_k, r_k}$. Then with probability at least $1 - \exp(-cp)$, the following holds.

Assume II res II_sp \leq d^{3/4}

r~\sqrt{d}

$$\|E_k \left(U_{k+1}^* \otimes \cdots \otimes U_m^* \otimes U_1^* \otimes \cdots \otimes U_{k-1}^*\right)\|_{\text{sp}} \leq C(\sqrt{d_k} + \sqrt{r_{-k}}),$$

$$\|E_k \left(U_{k+1}^* \otimes \cdots \otimes U_m^* \otimes U_1^* \otimes \cdots \otimes U_{k-1}^*\right)\|_F \leq C\sqrt{d_k r_{-k}},$$

$$\sup_{\substack{U_\ell \in \mathbb{O}_{d_\ell, r_\ell} \\ \ell \neq [m]}} \|E_k \left(U_{k+1} \otimes \cdots \otimes U_m \otimes U_1 \otimes \cdots \otimes U_{k-1}\right)\|_{\text{sp}} \leq C(\sqrt{d_k} + \sqrt{r_{-k}} + \sum_{\ell \neq k} \sqrt{d_\ell r_\ell}),$$

E_k + residual tensor

d^{3/4}

$$\sup_{\substack{U_\ell \in \mathbb{O}_{d_\ell, r_\ell} \\ \ell \neq [m]}} \|E_k \left(U_{k+1} \otimes \cdots \otimes U_m \otimes U_1 \otimes \cdots \otimes U_{k-1}\right)\|_F \leq C(\sqrt{d_k r_{-k}} + \sum_{\ell \neq k} \sqrt{d_\ell r_\ell}),$$

$$\sup_{\substack{U_\ell \in \mathbb{O}_{d_\ell, r_\ell} \\ \ell \neq [m]}} \|\mathcal{E} \times_1 U_1^T \times \cdots \times_m U_m^T\|_F \leq C(\sqrt{r_*} + \sum_{\ell \in [m]} \sqrt{d_\ell r_\ell})$$

Lemma 0.2 (Projection bound of perturbation). Suppose $X, E \in \mathbb{R}^{m \times n}$. Let $U \in \mathbb{O}_{m,r}$ be the leading r singular vectors of $Y = X + E$. Then,

$$\|(I - UU^T)X\|_{\text{sp}} \leq \sigma_{r+1}(Y) + \|E\|_{\text{sp}}$$

$$\|(I - UU^T)X\|_F \leq \sqrt{\sum_{i=r+1}^{m \wedge n} \sigma_i^2(Y)} + \|E\|_F.$$

key step where ``low-rankness'' of X helps.

Proof. For matrix norm bound we have,

$$\|(I - UU^T)X\|_{\mathrm{sp}} \leq \|(I - UU^T)Y\|_{\mathrm{sp}} + \|E\|_{\mathrm{sp}} \leq \sigma_{r+1}(Y) + \|E\|_{\mathrm{sp}}.$$

Similarly we bound Frobenius norm,

$$\|(I - UU^T)X\|_F \leq \|(I - UU^T)Y\|_F + \|E\|_F \leq \sqrt{\sum_{i=r+1}^{m \wedge n} \sigma_i^2(Y)} + \|E\|_F.$$

□

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.