Proof for the estimation accuracy of square spectral algorithm

Chanwoo Lee, July 6, 2021

1   Theorem 2.1 and the proof

Theorem 2.1 (Estimation accuracy of square spectral algorithm)**.** For an order-m dimensional-d data tensor, we perform SVD on the unfolded matrix $\text{Mat}(\mathcal{Y}) \in d^{\lfloor m/2 \rfloor}$-by-$d^{\lceil m/2 \rceil}$ with singular value truncation threshold $\hat{\lambda}_i \geq d^{\frac{\lceil m/2 \rceil}{2}}$. Then, the algorithm output satisfies the error bound

$$\mathcal{R}(\Theta, \hat{\Theta}) := \frac{1}{d^m} \|\hat{\Theta} - \Theta\|_F^2 \leq \begin{cases} d^{-\frac{2m}{m+4}}, & \text{even } m, \\ d^{-\frac{2(m-1)}{m+3}}, & \text{odd } m. \end{cases}$$

Proof. By definition of permuted smooth tensor model, $\text{Mat}(\mathcal{Y})$ is from the permuted smooth matrix model

$$\text{Mat}(\mathcal{Y}) = \boldsymbol{M} + \text{Mat}(\mathcal{E}),$$

where $\boldsymbol{M} := \text{Mat}(\Theta) \in \mathbb{R}^{\lfloor m/2 \rfloor \times \lceil m/2 \rceil}$ is the square unfolding of the signal tensor Θ, and $\text{Mat}(\mathcal{E}) \in \mathbb{R}^{\lfloor m/2 \rfloor \times \lceil m/2 \rceil}$ is a noise matrix with i.i.d. zero-mean subGuassian entries. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{d^{\lfloor m/2 \rfloor}}$ and $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_{d^{\lfloor m/2 \rfloor}}$ denote the singular values in descending order of $\boldsymbol{M}$ and $\text{Mat}(\mathcal{Y})$, respectively. By Weyl's inequality, we have

$$|\lambda_i - \hat{\lambda}_i| \leq \|\text{Mat}(\mathcal{E})\|_{\text{sp}} \leq 2d^{\frac{\lceil m/2 \rceil}{2}}, \quad \text{for all }, i = 1, \ldots, d^{\lfloor m/2 \rfloor}. \tag{1}$$

Notice that the last inequality is from the fact that $\|\text{Mat}(\mathcal{E})\|_{\text{sp}} \leq 2d^{\frac{\lceil m/2 \rceil}{2}}$ with probability at least $1 - 2\exp\left(-d^{\lfloor m/2 \rfloor}/2\right)$ [1, Theorem 2.6].

Let ℓ count the number of singular values of $\boldsymbol{M}$ that are above d, i.e.,

$$\ell = \sum_{i \in [d^{\lfloor m/2 \rfloor}]} \mathbb{1}\{\lambda_i \geq d^{\frac{\lceil m/2 \rceil}{2}}\}. \tag{2}$$

Now we decompose the error as

$$\|\hat{\boldsymbol{M}} - \boldsymbol{M}\|_F^2 \leq \underbrace{\|\hat{\boldsymbol{M}} - \text{Rank}(\boldsymbol{M}, \ell)\|_F^2}_{\text{variance}} + \underbrace{\|\text{Rank}(\boldsymbol{M}, \ell) - \boldsymbol{M}\|_F^2}_{\text{bias}}, \tag{3}$$

where $\text{Rank}(\boldsymbol{M}, \ell)$ denotes the best rank-ℓ approximation of the matrix $\boldsymbol{M}$ in least square sense. We claim that, both $\hat{\boldsymbol{M}}$ and $\text{Rank}(\boldsymbol{M}, \ell)$ have rank bounded by ℓ. By definition, $\text{Rank}(\boldsymbol{M}, \ell)$ has rank ℓ. To see the rank of $\hat{\boldsymbol{M}}$, notice that by Weyl's inequality (1) and definition of ℓ in (2),

$$\hat{\lambda}_i \leq \lambda_i + 2d^{\frac{\lceil m/2 \rceil}{2}} \leq 3d^{\frac{\lceil m/2 \rceil}{2}}, \quad \text{for all } i = \ell+1, \ell+2, \ldots, d^{\lfloor m/2 \rfloor}, \tag{4}$$

Therefore, $\text{Mat}(\mathcal{Y})$ has at most ℓ singular values above $3d^{\frac{\lceil m/2 \rceil}{2}}$. By the construction of $\hat{\boldsymbol{M}}$, $\text{Rank}(\hat{\boldsymbol{M}}) \leq \ell$.

Now we bound the estimation error (3). For the variance term,

$$\|\hat{\boldsymbol{M}} - \text{Rank}(\boldsymbol{M}, \ell)\|_F \leq \sqrt{\ell}\|\hat{\boldsymbol{M}} - \text{Rank}(\boldsymbol{M}, \ell)\|_{\text{sp}}$$

$$\leq \sqrt{\ell}\left(\underbrace{\|\hat{\boldsymbol{M}} - \text{Mat}(\mathcal{Y})\|_{\text{sp}}}_{\text{goodness-of-fit}} + \underbrace{\|\text{Mat}(\mathcal{Y}) - \boldsymbol{M}\|_{\text{sp}}}_{\text{noise}} + \underbrace{\|\boldsymbol{M} - \text{Rank}(\boldsymbol{M}, \ell)\|_{\text{sp}}}_{\text{bias}}\right)$$

$$\leq \sqrt{\ell}(\hat{\lambda}_{\ell+1} + 2d^{\frac{\lceil m/2\rceil}{2}} + \lambda_{\ell+1})$$

$$\lesssim \sqrt{\ell}d^{\frac{\lceil m/2\rceil}{2}},$$

where the third inequality uses the fact that $\|\text{Mat}(\mathcal{Y}) - \boldsymbol{M}\|_{\text{sp}} = \|\text{Mat}(\mathcal{E})\|_{\text{sp}} \leq 2d^{\frac{\lceil m/2\rceil}{2}}$ with probability at least $1 - 2\exp\left(-d^{\lfloor m/2\rfloor}/2\right)$ [1, Theorem 2.6], and the last inequality is from (4). Therefore, (3) has the upper bound,

$$\|\hat{\boldsymbol{M}} - \boldsymbol{M}\|_F^2 \lesssim \ell d^{\lceil m/2\rceil} + \|\text{Rank}(\boldsymbol{M}, \ell) - \boldsymbol{M}\|_F^2 \tag{5}$$

$$\leq r d^{\lceil m/2\rceil} + \|\text{Rank}(\boldsymbol{M}, r) - \boldsymbol{M}\|_F^2, \quad \text{for all } r = 1, 2, \ldots, d^{\lfloor m/2\rfloor},$$

where the last line uses the fact that $\ell = \sum_{i\in[d^{\lfloor m/2\rfloor}]} \mathbb{1}\{\lambda_i \geq d^{\frac{\lceil m/2\rceil}{2}}\}$ is the global optimizer of the function

$$g(r) = r d^{\lceil m/2\rceil} + \sum_{i\geq r+1} \lambda_i^2.$$

Finally, by Lemma 4, for every integer k, there exists a $(k, \ldots, k)$-block tensor such that

$$\|\text{Block}(\Theta; k) - \Theta\|_F^2 \lesssim \frac{d^m}{k^2},$$

where $\text{Block}(\Theta; k)$ denotes the blcok tensor with k blocks on each of the modes. Based on the relationship $\boldsymbol{M} = \text{Mat}(\Theta)$ and the fact that $\text{Mat}(\text{Block}(\Theta; k))$ is of rank at most $k^{\lfloor m/2\rfloor}$, we conclude from (5) that

$$\|\hat{\boldsymbol{M}} - \boldsymbol{M}\|_F^2 \lesssim r d^{\lceil m/2\rceil} + \|\Theta - \text{Block}(\Theta, r^{1/\lfloor m/2\rfloor})\|_F \leq r d^{\lceil m/2\rceil} + \frac{d^m}{r^{2/\lfloor m/2\rfloor}}, \quad \text{for all } r = 1, \ldots, d^{\lfloor m/2\rfloor}.$$

Taking $r = d^{\frac{m-\lceil m/2\rceil}{1+2/\lfloor m/2\rfloor}}$ yields,

$$\|\hat{\boldsymbol{M}} - \boldsymbol{M}\|_F^2 \lesssim \begin{cases} d^{\frac{m^2+2m}{m+4}}, & \text{even } m, \\ d^{\frac{m^2+m+2}{m+3}}, & \text{odd } m. \end{cases}$$

Therefore, we have the final upper bound as

$$\mathcal{R}(\Theta, \hat{\Theta}) := \frac{1}{d^m}\|\hat{\Theta} - \Theta\|_F^2 = \frac{1}{d^m}\|\hat{\boldsymbol{M}} - \boldsymbol{M}\|_F^2 \leq \begin{cases} d^{-\frac{2m}{m+4}}, & \text{even } m, \\ d^{-\frac{2(m-1)}{m+3}}, & \text{odd } m. \end{cases}$$

$\square$

2

Remark 1 (Connection between ℓ and r). In the proof, we set rank $r = d^{\frac{m-\lceil m/2 \rceil}{1+2/\lfloor m/2 \rfloor}}$ and obtain the upper bound.

$$\|\hat{\boldsymbol{M}} - \boldsymbol{M}\|_F^2 \lesssim \ell d^{\lceil m/2 \rceil} + \|\text{Rank}(\boldsymbol{M}, \ell) - \boldsymbol{M}\|_F^2$$
$$\leq \min_{r \in [d^{\lfloor m/2 \rfloor}]} \left(r d^{\lceil m/2 \rceil} + \|\text{Rank}(\boldsymbol{M}, r) - \boldsymbol{M}\|_F^2 \right)$$

Originally, $\|\text{Rank}(\boldsymbol{M}, r) - \boldsymbol{M}\|_F^2$ depends on the signal matrix $\boldsymbol{M} := \text{Mat}(\Theta)$. To be specific, Theorem 1 in [2] has the error bound consisting of tails of eigenvalues of $\boldsymbol{M}$. Since $\|\text{Rank}(\boldsymbol{M}, r) - \boldsymbol{M}\|_F^2$ has closely related to tails of eigenvalues of $\boldsymbol{M}$, it is hard to obtain the general bound. This is why the author assume the distribution of tails of eigenvalues of $\boldsymbol{M}$ in the Corollary 1.

Instead of assuming the distribution of eigenvalues of $\boldsymbol{M}$, we assume that the signal tensor Θ is from the Lipchitz smooth tensor. With block approximation technique, this assumption allows us to obtain the upper bound of $\|\text{Rank}(\boldsymbol{M}, r) - \boldsymbol{M}\|_F^2$ independent of $\boldsymbol{M}$ but dependent on r. Therefore, the major difference between r and ℓ is that the rank $r = d^{\frac{m-\lceil m/2 \rceil}{1+2/\lfloor m/2 \rfloor}}$ is independent of $\boldsymbol{M}$ while ℓ is directly connected to the eigenvalues of $\boldsymbol{M}$.

Notice that Proposition 1 in [2] uses the similar technique assuming Hölder condition or Soblev condition on graphon to bound $\|\text{Rank}(\boldsymbol{M}, r) - \boldsymbol{M}\|_F^2$ independently of $\boldsymbol{M}$.

Lemma 4 (Approximation error). For every fixed integer $k \leq d$, we have

$$\|\text{Block}(\Theta; k) - \Theta\|_F^2 \lesssim \frac{d^m}{k^2}.$$

Proof. Notice that for any $\omega \in [d]^m$,

$$([\text{Block}(\Theta; k)]_\omega - \Theta_\omega)^2 = \left(\frac{1}{h^m} \sum_{\omega' \in \left[\lfloor \frac{\omega-1}{h} \rfloor h + 1, \lfloor \frac{\omega-1}{h} \rfloor h + h \right]} (\Theta_{\omega'} - \Theta_\omega) \right)^2. \tag{6}$$

We bound each summand by

$$|\Theta_{\omega'} - \Theta_\omega|^2 \leq \frac{L^2 |\omega' - \omega|_1^2}{d^2} \tag{7}$$
$$\lesssim \frac{1}{k^2},$$

where the last inequality uses the fact that $\omega'_i \in \left[\lfloor \frac{\omega_i - 1}{h} \rfloor h + 1, \lfloor \frac{\omega_i - 1}{h} \rfloor h + h \right]$ for all $i \in [m]$. Therefore, plugging (7) into (6) yields

$$\|\text{Block}(\Theta; k) - \Theta\|_F^2 \lesssim \frac{d^m}{k^2}.$$

$\square$

2 Illustrative figure of $\mathcal{R}$ vs. m.

Figure 1 shows the error rate of MLE and spectral method according to different tensor order m. We check that our error bound is very close to gold-criteria MLE bound $d^{\frac{-2m}{m+2}}$.
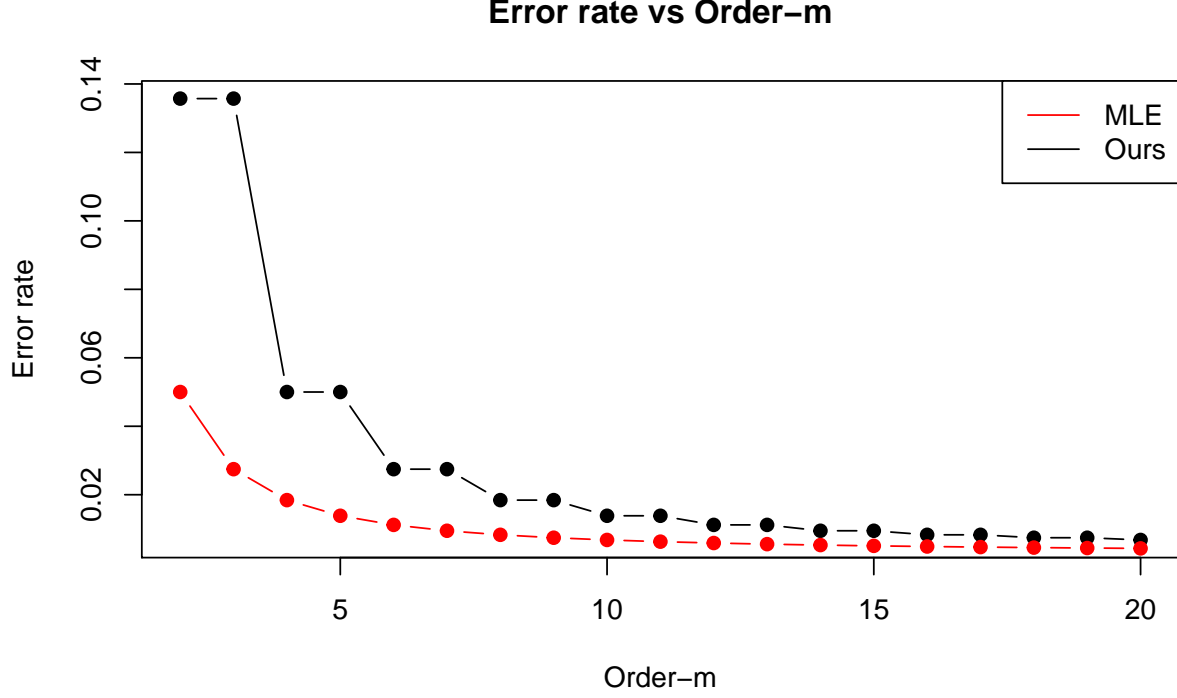

Figure 1: An illustration figure of error rate $\mathcal{R}$ and tensor order-m. Red colored line is MLE error bound while black one is spectral estimation error bound. We set tensor dimension $d = 20$.

References

[1] Mark Rudelson and Roman Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures*, pages 1576–1602. World Scientific, 2010.

[2] Jiaming Xu. Rates of convergence of spectral methods for graphon estimation. In *International Conference on Machine Learning*, pages 5433–5442. PMLR, 2018.