Estimating smooth tensors with unknown permutations
Chanwoo Lee \(^1\)  Miaoyan Wang \(^1\)
\(^1\)Department of Statistics, University of Wisconsin-Madison

**Main Problem**

We consider permuted tensor model,

\[ Y = \Theta \circ \sigma + \xi, \]

where \( Y \in \mathbb{R}^{d_1 \times \ldots \times d_k} \) is an observed data tensor, \( \Theta \) is an unknown smooth signal tensor, \( \sigma \) is an unknown permutation, and \( \xi \) is a noise tensor consisting of zero-mean sub-Gaussian entries.

**Main problem:** how to estimate \( \Theta \circ \sigma \) from the observed data tensor \( Y \)?

**Limitations of low-rank assumption**

Low-rank models assume

\[ \Theta \circ \sigma = \sum_{\ell=1}^{r} \lambda_{\ell} a_{\ell}^{(1)} \otimes \cdots \otimes a_{\ell}^{(d)}, \]

where \( \lambda_{\ell} \) is a scalar and \( a_{\ell}^{(d)} \in \mathbb{R}^d \) for all \( (k, \ell) \in [m] \times [p] \).

However, low rank models are

- **Sensitive** to order-preserving transformation

\[ \theta = \frac{1}{1 + \exp(-c(2Z))}, \]

where

\[ Z = a^{(1)} + b^{(2)} + c^{(3)}. \]

- **Inadequate** for special structures.

\[ \Theta = \log(1 + 2), \]

where

\[ Z(i, j, k) = \frac{1}{d} \max(i, j, k). \]

**Our main assumption**

Instead, we assume there exists \( f : [0, 1]^m \rightarrow \mathbb{R} \) that satisfies

- **Representation:** \( \Theta(\omega) = f(\omega/d) \) for all \( \omega \in [d^m] \),

\[ \| f(x) - f(y) \| \leq 1 \| x - y \|_m \]

for all \( x, y \in [0, 1]^m \), where the norm \( \| x \|_m = \sum_i |x_i|^m \) for \( x \in \mathbb{R}^m \).

**Stochastic block approximation (SBA) to smooth tensor**

**Lemma** (Block approximation). For any \( k < d \), there exists \( k \)-membership function \( z : [d] \rightarrow [k] \), and \( G \in \mathbb{R}^{d \times \cdots \times d} \) such that

\[ \frac{1}{d^m} \sum_{\omega \in [d^m]} |(\Theta \circ \sigma)(\omega) - G(z(\omega))|^2 \leq \frac{m^2 L^2}{kn}. \]

Based on this lemma and algorithms in [2], we find the optimizer,

\[ (\hat{z}, \hat{G}) = \arg \min_z \sum_{\omega \in [d^m]} |Y(\omega) - \hat{G}(z(\omega))|^2. \]

We estimate the \( \Theta \circ \sigma \) by

\[ (\hat{\Theta} \circ \hat{\sigma})(\omega) = \hat{G}(\hat{z}(\omega)), \]

for all \( \omega \in [d^m] \).

**Theoretical guarantees**

**Theorem** (Mean square error). Let \( \hat{\Theta} \) be the estimator from (2) with the choice of \( k = \lceil d^m \rceil \). Then,

\[ \frac{1}{d^m} \| (\hat{\Theta} \circ \hat{\sigma} - \Theta \circ \sigma) \|^2 \leq \frac{d^m}{n} \text{ Nonparametric rate} + \frac{\log d}{d^m n} \text{ Clustering rate}, \]

with high probability.

**Remark:** Depending on constants \( m \) and \( \alpha \), convergence rate becomes

\[
\text{RHS of (3)} \propto \begin{cases} 
\frac{d^m}{m^2} & m = 2, \alpha \in (0, 1), \\
\frac{d^m}{m^2} & m = 2, \alpha = 1 \\
\frac{d^m}{m^2} & m > 2
\end{cases}
\]

Though SBA guarantees fast convergence rate, polynomial complexity algorithms for (1) are unknown.

**Ongoing work:** polynomial algorithms

**Sort-And-Smooth (SAS) method extended from [1]**

Under the monotonically increasing degree assumption on signal \( \Theta \),

\[ \frac{1}{d^m-1} \sum_{\ell=1}^{d^m-1} \Theta(i, j, \ldots, m) > \frac{1}{d^m-1} \sum_{\ell=1}^{d^m-1} \Theta(j, \ell, \ldots, m), \]

for all \( i > j \).

**Step 1** (sorting): Find \( \sigma \) so that the degree of \( \Theta \circ \sigma^{-1} \) is increasing.

**Step 2** (smoothing): Estimate signal matrix \( \Theta = \text{Block} \{Y \circ \sigma^{-1}\} \),
where

**Spectral method extended from [3]**

**Step 1** (Unfolding): Unfold \( Y \) into \( \text{Mat}(Y) \in \mathbb{R}^{m_1 \times \cdots \times m_d \times m'}. \)

**Step 2** (SVD): Obtain SVD of \( \text{Mat}(Y) = \sum_{i=1}^{m'} \lambda_i v_i u_i^T \).

**Step 3** (Thresholding): Obtain \( \hat{\Theta} = \sum_{i=1}^{m_k} \lambda_i v_i u_i^T \) for all \( \lambda_i \geq d/m \) and fold back to tensor \( \Theta \).

**Figure 1:** Right triangular matrices show the true signal and left ones show the estimated ones. Simulation 1 follows monotonic degree assumption while Simulation 2 does not.

<table>
<thead>
<tr>
<th>Convergence rate (power of ( d^{-1} ))</th>
<th>Optimal complexity</th>
<th>SBA</th>
<th>SAS</th>
<th>Spectral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynomial complexity</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Table 1:** Comparison of SBA, SAS, and Spectral method for \( m = 1 \) and \( m > 2 \).

**References**


IFDS Summer School 2021

The research was supported in part by NSF DMS-1915978, NSF DMS-2023239, and Wisconsin Alumni Research Foundation.