Tensor Clustering and Error Bounds

Chris Ding

Department of Computer Science and Engineering
University of Texas, Arlington

Joint work with Heng Huang and Dijun Luo

Work Supported by NSF CISE/DMS
Tensors

- The word tensor is used in 1900 (time of A. Einstein) in physics
  - General relativity is entirely written in tensor format
  - Physicists see tensor and think of coordinate transformation properties
  - Computer scientists see tensor and wants to compute them faster
Tensor Decompositions: Main new results

• Two main tensor decompositions
  - ParaFac (CanDecomp, rank-1)
  - HOSVD (Tucker-3)

• Data clustering
  - ParaFac does simultaneous compression and K-means clustering
    • Cluster centroids are rank-1 matrices:
      \[ u^{(r)} v^{(r)T} \]
  - HOSVD does simultaneous compression and K-means clustering
    • Cluster centroids are of the type:
      \[ U S^{(r)} V^T \]

• Eckart-Young type lower and upper error bounds
  - ParaFac
  - HOSVD

• Extensive experiments
ParaFac Objective Function

- ParaFac is the simplest and most widely used model

\[ X \approx \sum_{r=1}^{R} u^{(r)} \otimes r^{(r)} \otimes w^{(r)}, \text{ or } X_{ijk} \approx \sum_{r=1}^{R} U_{ir} V_{jr} W_{kr} \]

\[
\min_{U, V, W, W^T W = I} \quad J_{\text{ParaFac}} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \left| \left| X_{ijk} - \sum_{r=1}^{R} U_{ir} V_{jr} W_{kr} \right| \right|^2
\]
Bounds on ParaFac Reconstruction Error

**Theorem 2.2.** ParaFac decomposition has the following lower and upper bounds:

\[ B_0 \leq J_{ParaFac}^{opt} \leq B_1 \leq B_2. \]  

(2.8)

Eckart-Young type Error bounds:

\[ B_0 = \|X\|^2 - \sum_{r=1}^{R} \lambda_r(G) = \sum_{r=R+1}^{\text{rank}(R)} \lambda_r(G). \]

\[ B_1 = B_0 + \sum_{r=1}^{R} \sum_{l=2}^{n_1} \lambda_l(A^{(r)}), \quad B_2 = B_0 + \sum_{l=2}^{n_1} \lambda_l(A) \]

\[ G_{kk'} = \text{Tr}(X^{(k)} T X^{(k')}) = \sum_{ij} X_{ijk} X_{ijk'} \quad W= \text{eigenvectors of } G \]

\[ A^{(r)} = \sum_{kk'} W_{kr} W_{k'r} X^{(k)} X^{(k')T} \quad A = \sum_{r=1}^{R} A^{(r)} \]
Outline of the Upper Error Bounds

- In standard ParaFac, columns of $W$ is only required to be linearly independent
  - We study $W$-orthogonal ParaFac where $W$ is required to be orthogonal.
  - Upper bound is obtained because the domain is further restricted.
  - Any feasible solution of $W$-orthogonal ParaFac gives an upper bound.

\[ J_{\text{ParaFac}}^{\text{opt}} \leq J_{\text{ParaFac1}}^{\text{opt}} \leq B_1 \leq B_2. \]

- $W$-orthogonal ParaFac can be reduction [$(U,V,W)$ to $W$-only]

**Theorem 3.1.** The $W$-orthonormal ParaFac is equivalent to the following optimization

\[ \min_W J_{\text{ParaFac1}}(W) \quad (3.16) \]

where

\[ J_{\text{ParaFac1}}(W) = \|X\|^2 - \text{Tr} \ W^T GW + \sum_{r=1}^{R} \sum_{l=2}^{n_3} \lambda_l (A^{(r)}). \]
Outline of the Lower Error Bound

- Increasing the domain of variables \( \rightarrow \) more accurate approximation \( \rightarrow \) lower bound

In ParaFac decomposition:

\[
X \approx \sum_{r=1}^{R} u^{(r)} \otimes r^{(r)} \otimes w^{(r)}, \text{ or } X_{ijk} \approx \sum_{r=1}^{R} U_{ir} V_{jr} W_{kr}
\]

We replace

\[
U_{ir} V_{jr} \leftrightarrow C_{ijr} \text{ or } u^{(r)} v^{(r)T} \leftrightarrow C^{(r)}
\]

that is, we decompose \( X \) as

\[
X_{ijk} \approx \sum_{r=1}^{R} C_{ijr} W_{kr}, \text{ or } X_{ij}^{(k)} \approx \sum_{r=1}^{R} C_{ij}^{(r)} W_{kr}.
\]

Lower bound:

\[
\min_{C,W} J_{\text{T1}} = \sum_{k=1}^{n_3} \| X^{(k)} - \sum_{r=1}^{R} C^{(r)} W_{kr} \|^2
\]
## Experiments on ParaFac Error Bounds

<table>
<thead>
<tr>
<th></th>
<th>MNIST</th>
<th>BinAlpha</th>
<th>AT&amp;T Face</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>size 28x28x1000</td>
<td>size 20x16x1014</td>
<td>size 56x46x400</td>
</tr>
<tr>
<td></td>
<td>in unit of $10^9$</td>
<td>in unit of $10^4$</td>
<td>in unit of $10^9$</td>
</tr>
<tr>
<td>R</td>
<td>$B_0$</td>
<td>$J_{ParaFac}^{opt}$</td>
<td>$B_1$</td>
</tr>
<tr>
<td>1</td>
<td>3.2498</td>
<td>3.3317</td>
<td>3.3351</td>
</tr>
<tr>
<td>4</td>
<td>2.4685</td>
<td>2.6458</td>
<td>2.8729</td>
</tr>
<tr>
<td>8</td>
<td>1.9013</td>
<td>2.1356</td>
<td>2.5790</td>
</tr>
</tbody>
</table>

Table 2: Lower and upper error bounds for ParaFac
High Order SVD (HOSVD)

• Initially called Tucker-3 Decomposition
• HOSVD uses 3 factors and a core tensor $S$:

$$X \approx U \otimes_1 V \otimes_2 W \otimes_3 S, \text{ or } X_{ijk} \approx \sum_{p=1}^{m_1} \sum_{q=1}^{m_2} \sum_{r=1}^{m_3} U_{ip} V_{jq} W_{kr} S_{pqr}$$

$U, V, W, S$ are obtained by minimizing the reconstruction error

$$\min_{U,V,W,S} \quad J_1 = \| X - U \otimes_1 V \otimes_2 W \otimes_3 S \|^2$$

s.t. \hspace{1cm} U^T U = I, V^T V = I, W^T W = I.$$

$$J_1 = \sum_{i,j,k} \left( X_{ijk} - \sum_{p,q,r} U_{ip} V_{jq} W_{kr} S_{pqr} \right)^2$$
HOSVD Error Bounds

Theorem 1. HOSVD has the upper and lower error bounds

\[
\sum_{m=k_1+1}^{n_1} \lambda^F_m \leq J^{\text{opt}}_1 \leq \sum_{m=k_1+1}^{n_1} \lambda^F_m + \sum_{m=k_2+1}^{n_2} \lambda^G_m + \sum_{m=k_3+1}^{n_3} \lambda^H_m
\]

where \((\lambda^F_1, \cdots, \lambda^F_{n_1})\) are eigenvalues of matrix \(F\), \((\lambda^G_1, \cdots, \lambda^G_{n_2})\) are eigenvalues of matrix \(G\), \((\lambda^H_1, \cdots, \lambda^H_{n_3})\) are eigenvalues of matrix \(H\). \(F, G, H\) are appropriate co-variance matrices

\[
F_{i'i'} = \sum_{ik} X_{ijk} X_{i'jk}
\]

\[
G_{jj'} = \sum_{ii'k} X_{ijk} (UU^T)_{ii'} X_{i'j'k}
\]

\[
H_{kk'} = \sum_{ii'jj'} X_{ijk} (UU^T)_{ii'} X_{i'j'k'} (VV^T)_{jj'}
\]

\[
U = \text{eigenvectors}(F), \quad V = \text{eigenvectors}(G)
\]
Outline of the Upper Error Bound

We need to find a feasible solution, which gives an upper bound

**A) 3-step up-bounding strategy.**

Using the following inequality

\[
|a - b| = |a - a_1 + a_1 - a_2 + a_2 - b| 
\leq |a - a_1| + |a_1 - a_2| + |a_2 - b|,
\]

we obtain

\[
||Y - U \otimes_1 V \otimes_2 W \otimes_3 M|| 
\leq ||Y - U \otimes_1 \tilde{M}|| 
+ ||U \otimes_1 \tilde{M} - U \otimes_1 V \otimes_2 \tilde{M}|| 
+ ||U \otimes_1 V \otimes_2 \tilde{M} - U \otimes_1 V \otimes_2 W \otimes_3 M|| 
= ||Y - U \otimes_1 \tilde{M}|| 
+ ||\tilde{M} - V \otimes_2 \tilde{M}|| 
+ ||\tilde{M} - W \otimes_3 M||.
\]
Outline of the Upper Error Bound

The above inequality suggests a 3-step optimization procedure to obtain a good feasible solution for $J_1$ of Eq.(3).

**Step-1:**

$$
\min_{U \in \mathbb{R}^{n_1 \times k_1}} \min_{\tilde{M} \in \mathbb{R}^{k_1 \times n_2 \times n_3}} J_u = \|Y - U \otimes_1 \tilde{M}\| \quad (11)
$$

**Step-2:** we fix $\tilde{M}$ to the values obtained Step-1 and minimize

$$
\min_{V \in \mathbb{R}^{n_2 \times k_2}} \min_{\tilde{M} \in \mathbb{R}^{k_1 \times k_2 \times n_3}} \|\tilde{M} - V \otimes_2 \tilde{M}\| \quad (12)
$$

**Step-3:** we fix $\tilde{M}$ to the values obtained in Step-2, and minimize

$$
\min_{W \in \mathbb{R}^{n_3 \times k_3}} \min_{M \in \mathbb{R}^{k_1 \times k_2 \times k_3}} \|\tilde{M} - W \otimes_3 M\| \quad (13)
$$

All these are T1 decompositions and trivially solved.
Outline of the Upper Error Bound

**Proposition 3** Using the solutions \((U^*, V^*, W^*)\) provided by the 3-step procedure, the objective function value

\[
J_1(U^*, V^*, W^*) = \sum_{m=k_1+1}^{n_1} \lambda^F_m + \sum_{m=k_2+1}^{n_2} \lambda^G_m + \sum_{m=k_3+1}^{n_3} \lambda^H_m.
\]
4.1. Upper bound experiment

The upper bound in our theorem 1 is astonishing tight. We reconstruct 4D tensor images from AT&T face database [1], YALE face database B [5], and PIE face database [11] with $k_1 = k_2 = 20$, $k_3 =$ the length of image sequence. The error ratio is defined as $(J_{upper bound} - J_{opt}^{1})/\|X\|_F^2$. The error ratios of all three datasets are less than $10^{-6}$.

Compute eigenvalues and use the error bounds to determine HOSVD/ParaFac parameters