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 centrods are rank-1 matrices: $\mathbf{u}^{(r)}\mathbf{v}^{(r)T}$
  - HOSVD does simultaneous compression and K-means clustering
    - Cluster centroids are of the type: $US^{(r)}V^T$.

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

- Extensive experiments

ParaFac Objective Function

- ParaFac is the simplest and most widely used model

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

$$\min_{U,V,W,W^T W = I} \ J_{\text{ParaFac}} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} ||X_{ijk} - \sum_{r=1}^{R} U_{ir} V_{jr} W_{kr}||^2$$

Bounds on ParaFac Reconstruction Error

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

$$B_0 \leq J_{ParaFac}^{opt} \leq B_1 \leq B_2. \qquad (2.8)$$

Eckart-Young type Error bounds:

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

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

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

$$A^{(r)} = \sum_{kk'} W_{kr} W_{k'r} X^{(k)} X^{(k')T} \qquad A = \sum_{r=1}^{R} A^{(r)}$$

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.

$$J_{\text{ParaFac}}^{\text{opt}} \le J_{\text{ParaFac1}}^{\text{opt}} \le B_1 \le B_2.$$

- 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*

$$\min_W J_{ParaFac1}(W) \tag{3.16}$$

where

$$J_{ParaFac1}(W) = ||X||^2 - Tr\, W^T G W + \sum_{r=1}^{R} \sum_{l=2}^{n_3} \lambda_l(A^{(r)})$$

Outline of the Lower Error Bound

- Increasing the domain of variables ➔ more accurate approximation ➔ lower bound

In ParaFac decomposition:

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

We replace $U_{ir}V_{jr} \Longleftarrow C_{ijr}$ or $u^{(r)}v^{(r)T} \Longleftarrow C^{(r)}$

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:

$$\min_{C,W} J_{\mathrm{T1}} = \sum_{k=1}^{n_3} ||X^{(k)} - \sum_{r=1}^{R} C^{(r)} W_{kr}||^2$$

Experiments on ParaFac Error Bounds

	MNIST	size 28x28x1000	in unit of 10^9
R	B_0	$J^{opt}_{ParaFac}$	B_1
1	3.2498	3.3317	3.3351
4	2.4685	2.6458	2.8729
8	1.9013	2.1356	2.5790
	BinAlpha	size 20x16x1014	in unit of 10^4
R	B_0	$J^{opt}_{ParaFac}$	B_1
1	7.1770	7.2835	7.2875
4	5.7091	5.8993	6.1862
8	4.6031	4.8507	5.4960
	AT&T Face	size 56x46x400	in unit of 10^9
R	B_0	$J^{opt}_{ParaFac}$	B_1
1	1.3025	1.4279	1.4282
4	0.8534	0.9583	1.1053
8	0.6201	0.7356	0.9768

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

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

$$\text{s.t.} \qquad U^T U = I, V^T V = I, W^T W = I.$$

$$J_1 = \sum_{ijk} \left(X_{ijk} - \sum_{pqr} U_{ip} V_{jq} W_{kr} S_{pqr} \right)^2$$

HOSVD Error Bounds

Theorem 1 HOSVD *has the upper and lower error bounds*

$$\sum_{m=k_1+1}^{n_1} \lambda_m^F \le J_1^{opt} \le \sum_{m=k_1+1}^{n_1} \lambda_m^F + \sum_{m=k_2+1}^{n_2} \lambda_m^G + \sum_{m=k_3+1}^{n_3} \lambda_m^H$$

where $(\lambda_1^F, \cdots, \lambda_{n_1}^F)$ *are eigenvalues of matrix* F, $(\lambda_1^G, \cdots, \lambda_{n_2}^G)$ *are eigenvalues of matrix* G, $(\lambda_1^H, \cdots, \lambda_{n_3}^H)$ *are eigenvalues of matrix* H. F, G, H *are appropriate co-variance matrices*

$$F_{ii'} = \sum_{jk} X_{ijk} X_{i'jk} \qquad G_{jj'} = \sum_{ii'k} X_{ijk} (UU^T)_{ii'} X_{i'j'k}$$

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

U = eigenvectors(F), V=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

$$\begin{aligned} |a-b| &= |a-a_1+a_1-a_2+a_2-b| \\ &\leq |a-a_1|+|a_1-a_2|+|a_2-b|, \end{aligned}$$

we obtain

$$\begin{aligned} & ||Y-U\otimes_1 V\otimes_2 W\otimes_3 M|| \\ \leq\ & ||Y-U\otimes_1 \bar{M}|| \\ +\ & ||U\otimes_1 \bar{M}-U\otimes_1 V\otimes_2 \tilde{M}|| \\ +\ & ||U\otimes_1 V\otimes_2 \tilde{M}-U\otimes_1 V\otimes_2 W\otimes_3 M|| \\ =\ & ||Y-U\otimes_1 \bar{M}|| \\ +\ & ||\bar{M}-V\otimes_2 \tilde{M}|| \\ +\ & ||\tilde{M}-W\otimes_3 M||. \end{aligned}$$

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:

$$\min_{\substack{U \in \Re^{n_1 \times k_1} \\ \bar{M} \in \Re^{k_1 \times n_2 \times n_3}}} \quad J_u = ||Y - U \otimes_1 \bar{M}|| \qquad (11)$$

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

$$\min_{\substack{V \in \Re^{n_2 \times k_2} \\ \tilde{M} \in \Re^{k_1 \times k_2 \times n_3} \\ \bar{M} \text{fixed}}} \quad ||\bar{M} - V \otimes_2 \tilde{M}|| \qquad (12)$$

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

$$\min_{\substack{W \in \Re^{n_3 \times k_3} \\ M \in \Re^{k_1 \times k_2 \times k_3} \\ \tilde{M} \text{fixed}}} \quad ||\tilde{M} - W \otimes_3 M|| \qquad (13)$$

All these are T1 decompositions and trivially solved.

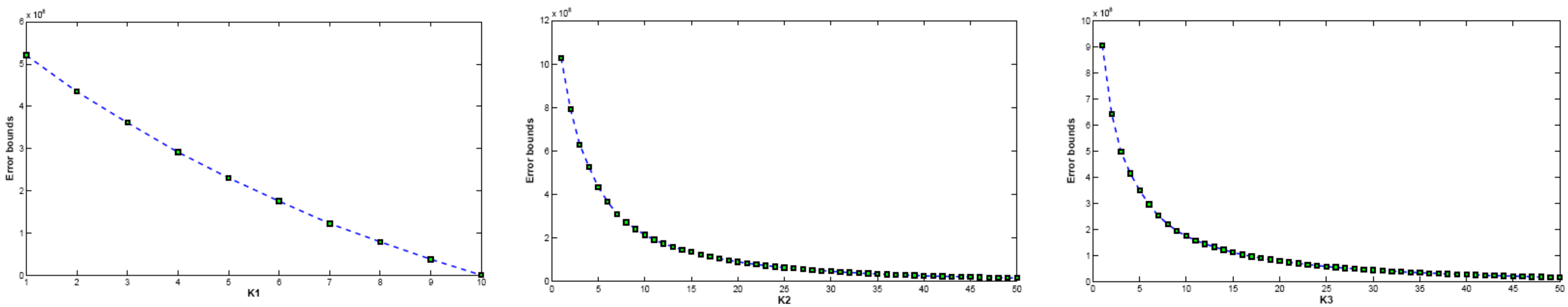
Outline of the Upper Error Bound

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

$$J_1(U^*, V^*, W^*) = \sum_{m=k_1+1}^{n_1} \lambda_m^F + \sum_{m=k_2+1}^{n_2} \lambda_m^G + \sum_{m=k_3+1}^{n_3} \lambda_m^H.$$

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 $(J_{upperbound} - J_1^{opt})/||X||_F^2$. The error ratios of all three datasets are less than 10^{-6}.



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