Numerical Computation for Orthogonal Low Rank Approximation of Tensors

Yu Guan [1] Delin Chu [2]

[1]UCLouvain

[2]National University of Singapore

July 17, 2019 @ ICIAM

Outline

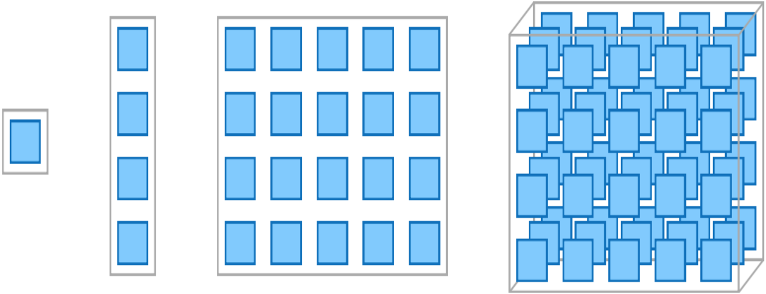


Figure : $x \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^4, X \in \mathbb{R}^{4\times 5}, \mathcal{X} \in \mathbb{R}^{4\times 5\times 3}$

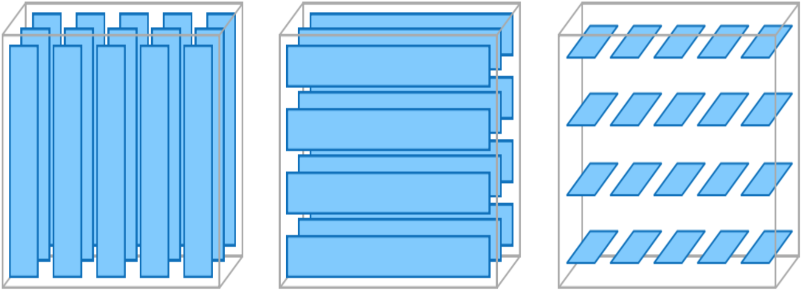


Figure : Column, row, and tube fibers of a order-3 tensor

Multiway Data

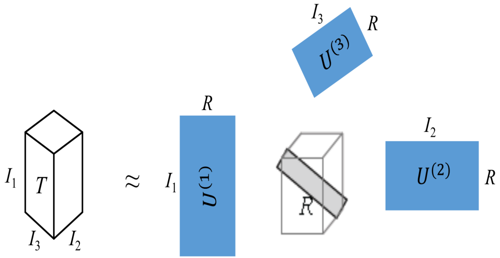
- Psychometrics: individual $\times$ variable $\times$ time
- Time-series analysis: time $\times$ variable $\times$ lag
- Facial image: people $\times$ view $\times$ illumination $\times$ expression $\times$ pixels
- Atmospheric science: location $\times$ variable $\times$ time $\times$ observation

Tensor of rank 1

$$\bigotimes_{i=1}^{k} \mathbf{u}^{(i)} = \mathbf{u}^{(1)} \circ \cdots \circ \mathbf{u}^{(k)} := \left[u_{i_1}^{(1)} \cdots u_{i_k}^{(k)} \right],$$

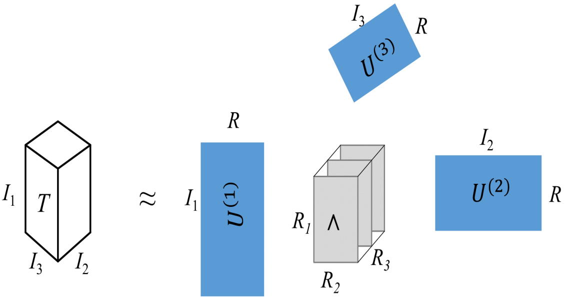
where $\mathbf{u}^{(j)} \in \mathbb{R}^{l_j}$, $j = 1, \ldots, k$.

CP Decomposition



$$T = \sum_r \lambda_r \mathbf{u}_r^{(1)} \circ \cdots \circ \mathbf{u}_r^{(k)} = [\![\Lambda; U^{(1)}, \ldots, U^{(k)}]\!]$$

Tucker Decomposition



$$T = \sum_{r_1, r_2, \ldots, r_k} \lambda_{r_1, r_2, \ldots, r_k} \mathbf{u}^{(1)}_{r_1} \circ \cdots \circ \mathbf{u}^{(k)}_{r_k}$$

Applications

- Decomposition into Directional Components (Dedicom):

$$\sum_{i=1}^{m} \|A_{(i)} - Q^\top R_i Q\|_F^2,$$

where $A_{(i)} \in \mathbb{R}^{n\times n}$ is the unfolding of tensor $\mathcal{A}$. Optimize over $R_i \in \mathbb{R}^{r\times r}$ and $Q \in \mathbb{R}^{r\times n}$ with orthonormal columns.

- Simultaneous Components Analysis (SCA):

$$\sum_{i=1}^{m} \|A_{(i)} - A_{(i)} B P_i\|_F^2,$$

where $A_{(i)} \in \mathbb{R}^{m_i\times n}$ is the unfolding of tensor $\mathcal{A}$. Optimize over $B \in \mathbb{R}^{n\times r}$ full-rank and $P_i \in \mathbb{R}^{r\times n}$ patterned matrix.

Low rank CP approximation

Determine unit vectors $\mathbf{u}_r^{(\ell)} \in \mathbb{R}^{I_\ell}$, $\ell = 1, \dots k$ and scalars λ_r to minimize

$$\left\| T - \sum_{r=1}^{R} \lambda_r \mathbf{u}_r^{(1)} \circ \cdots \circ \mathbf{u}_r^{(k)} \right\|_F^2 .$$

Challenges and ill-posedness

- Best low rank approximation of a matrix ($k = 2$) always exists. (Eckart-Young Theorem)
- The rank-1 approximation is theoretically guaranteed to have a global optimum.
- Best rank-R ($R > 1$) approximation for high-order tensors may not exist.

Example 1

Let $\mathbf{u}_i, \mathbf{v}_i$ be linearly independent. Define

$$T := \mathbf{u}_1 \circ \mathbf{u}_2 \circ \mathbf{v}_3 + \mathbf{u}_1 \circ \mathbf{v}_2 \circ \mathbf{u}_3 + \mathbf{v}_1 \circ \mathbf{u}_2 \circ \mathbf{u}_3 \in \mathbb{R}^{I_1 \times I_2 \times I_3},$$

and for each $n \in \mathbb{N}$,

$$T_n := n\left(\mathbf{u}_1 + \frac{1}{n}\mathbf{v}_1\right) \circ \left(\mathbf{u}_2 + \frac{1}{n}\mathbf{v}_2\right) \circ \left(\mathbf{u}_3 + \frac{1}{n}\mathbf{v}_3\right) - n\,\mathbf{u}_1 \circ \mathbf{u}_2 \circ \mathbf{u}_3.$$

Then T has rank 3 and rank of T_n is at most 2.

Solution

- μ-orthogonality:

$$U^{(i_1)\top} U^{(i_1)} = I, \ldots, U^{(i_\mu)\top} U^{(i_\mu)} = I$$

- Includes
  - complete orthogonality $\mu = k$
  - semi-orthogonality $\mu = 1$

- Complete orthogonal low rank approximation are studied in (Chen et al. 08').
- Semi-orthogonal low rank approximation of tensors are studied in (Wang et al. 14').
- It is interesting to impose orthogonality to more than one factor matrix.
  - (Wang et al. 14') pointed that "More study is needed".
  - (Wang et al. 14') addressed that "The question of more than one semi-orthogonal factor matrix, except for the case of complete orthogonality, remains open".

Orthogonal low rank approximation

$$\min \left\| T - \sum_{r=1}^{R} \lambda_r \bigotimes_{i=1}^{k} \mathbf{u}_r^{(i)} \right\|_F^2,$$

subject to μ-orthogonality.

Equivalent formulation

The orthogonal low rank approximation problem can be reformulated as

$$\max \quad \sum_{r=1}^{R} \lambda_r^2 = \sum_{r=1}^{R} \langle T, \bigotimes_{i=1}^{k} \mathbf{u}_r^{(i)} \rangle^2,$$

subject to the μ-orthogonality.

Existing Algorithms

- For matrices ($k = 2$), the best low rank approximation is TSVD (Eckart-Young theorem).
- For general tensors ($k > 2$), the "workhorse" algorithm for orthogonal low rank approximation of tensor has been alternating least squares (ALS) method.
  - (Wang et al. 14') proved convergence globally.
  - Numerical computation of the completely orthogonal in (Chen et al. 08').

Contributions

- We develop an SVD-based algorithm which updates two factors simultaneously.
- To address the orthogonality, we apply polar decomposition for μ factors.
- The convergence of our SVD-based algorithm is analyzed for both objective function and iterates themselves.

Linear Mapping

β product

$$\mathscr{T}_{\boldsymbol{\beta}}(S) := T \circledast_{\boldsymbol{\beta}} S = [\langle \tau_{[:|\ell_1,\ldots,\ell_t]}, S \rangle] \in \mathbb{R}^{I_{\beta_1} \times \ldots \times I_{\beta_t}}$$

where

$$\langle \tau_{[:|\ell_1,\ldots,\ell_t]}, S \rangle := \sum_{i_1=1}^{I_{\alpha_1}} \cdots \sum_{i_s=1}^{I_{\alpha_s}} \tau_{[i_1,\ldots,i_s|\ell_1,\ldots,\ell_t]} s_{i_1,\ldots,i_s}$$

is the Frobenius inner product generalized to multi-dimensional arrays.

Algorithm Description

How can we update $\mathbf{u}_r^{(\ell)}$ for $\ell = 1, \ldots, k - \mu$?

- For any $1 \leq \ell \leq k - \mu - 1$ and let $\beta_\ell = (\ell, \ell + 1)$,

$$C_r^{(\ell)} = T \circledast_{\beta_\ell} \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_r^{(i)} \circ \bigotimes_{i=\ell+2}^{k} \mathbf{u}_r^{(i)} \right).$$

- $\tilde{\mathbf{u}}_r^{(\ell)}$ and $\tilde{\mathbf{u}}_r^{(\ell+1)}$ be the dominant left and right singular vectors of $C_r^{(\ell)}$.

Update $U^{(\ell)}$ by $\tilde{U}^{(\ell)}$ which is from the orthogonal polar factor of the matrix $V^{(\ell)}\Lambda^{(\ell)}$ for $\ell = k - \mu + 1, \dots, k$, where

$$\mathbf{v}_r^{(\ell)} = \mathcal{T}\circledast_\ell \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_r^{(\ell)} \circ \bigotimes_{i=\ell+1}^{k} \mathbf{u}_r^{(i)} \right), \quad \ell = 1, \dots, k-\mu+1, r = 1, \dots, R,$$

$$V^{(\ell)} = \left[\mathbf{v}_1^{(\ell)}, \dots, \mathbf{v}_R^{(\ell)} \right], \quad U^{(\ell)} = \left[\mathbf{u}_1^{(\ell)}, \dots, \mathbf{u}_R^{(\ell)} \right],$$

$$\Lambda^{(\ell)} = \begin{bmatrix} \lambda_1^{(\ell)} & & \\ & \ddots & \\ & & \lambda_R^{(\ell)} \end{bmatrix}.$$

Trace Maximizing Property

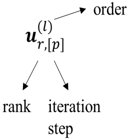
Let matrix $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ have polar decomposition

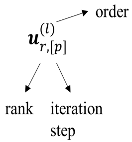
$$A = QS,$$

where $Q \in \mathbb{R}^{m \times n}$ is the column orthogonal polar factor and $S \in \mathbb{R}^{n \times n}$ is the symmetric positive semi-definite factor. Then

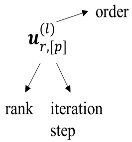
$$Q = \arg \max_{P \in \mathbb{R}^{m \times n},\ P^T P = I} \operatorname{Trace}\left(P^T A\right).$$

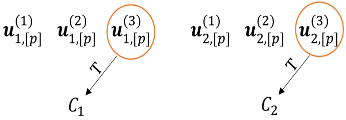
Moreover, if A is of full column rank, then Q above is unique.

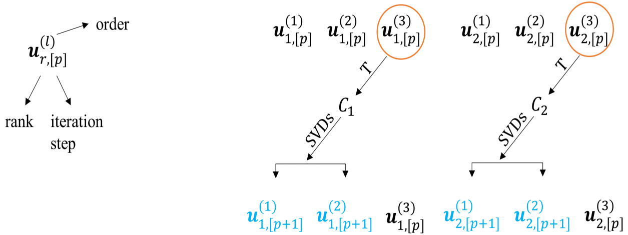

$\boldsymbol{u}_{r,[p]}^{(l)}$
order
rank
iteration step



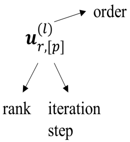
$$\boldsymbol{u}^{(1)}_{1,[p]} \quad \boldsymbol{u}^{(2)}_{1,[p]} \quad \boldsymbol{u}^{(3)}_{1,[p]} \qquad \boldsymbol{u}^{(1)}_{2,[p]} \quad \boldsymbol{u}^{(2)}_{2,[p]} \quad \boldsymbol{u}^{(3)}_{2,[p]}$$

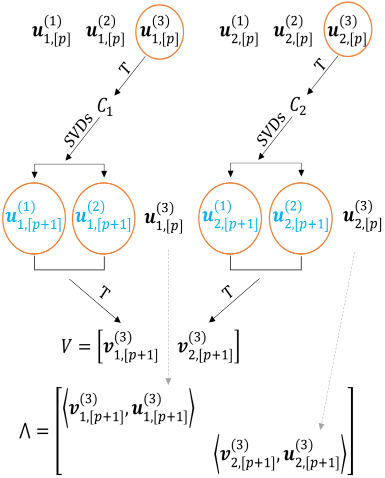

order
rank
iteration step

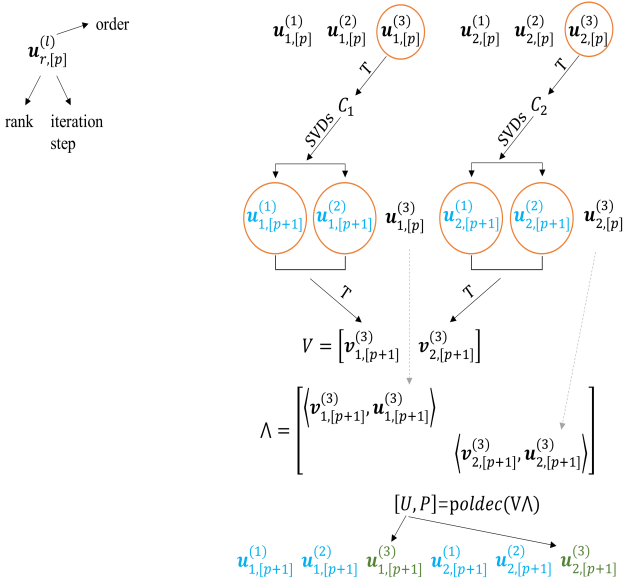

T
T



order
rank
iteration step
T
C_1
SVDs
T
C_2
SVDs


$u^{(l)}_{r,[p]}$
order
rank
iteration step


$u^{(1)}_{1,[p]}$ $u^{(2)}_{1,[p]}$ $u^{(3)}_{1,[p]}$ $u^{(1)}_{2,[p]}$ $u^{(2)}_{2,[p]}$ $u^{(3)}_{2,[p]}$
T
C_1
C_2
SVDs
$u^{(1)}_{1,[p+1]}$ $u^{(2)}_{1,[p+1]}$ $u^{(3)}_{1,[p]}$ $u^{(1)}_{2,[p+1]}$ $u^{(2)}_{2,[p+1]}$ $u^{(3)}_{2,[p]}$
$V = [v^{(3)}_{1,[p+1]} \; v^{(3)}_{2,[p+1]}]$
$\Lambda = \begin{bmatrix} \langle v^{(3)}_{1,[p+1]}, u^{(3)}_{1,[p+1]} \rangle & \\ & \langle v^{(3)}_{2,[p+1]}, u^{(3)}_{2,[p+1]} \rangle \end{bmatrix}$


$u^{(l)}_{r,[p]}$
order
rank
iteration step
$u^{(1)}_{1,[p]}$ $u^{(2)}_{1,[p]}$ $u^{(3)}_{1,[p]}$ $u^{(1)}_{2,[p]}$ $u^{(2)}_{2,[p]}$ $u^{(3)}_{2,[p]}$
T
C_1
SVDs
T
C_2
SVDs
$u^{(1)}_{1,[p+1]}$ $u^{(2)}_{1,[p+1]}$ $u^{(3)}_{1,[p]}$ $u^{(1)}_{2,[p+1]}$ $u^{(2)}_{2,[p+1]}$ $u^{(3)}_{2,[p]}$
T
T
$V = \left[v^{(3)}_{1,[p+1]} \quad v^{(3)}_{2,[p+1]} \right]$
$\Lambda = \begin{bmatrix} \langle v^{(3)}_{1,[p+1]}, u^{(3)}_{1,[p+1]} \rangle & \\ & \langle v^{(3)}_{2,[p+1]}, u^{(3)}_{2,[p+1]} \rangle \end{bmatrix}$
$[U, P] = poldec(V\Lambda)$
$u^{(1)}_{1,[p+1]}$ $u^{(2)}_{1,[p+1]}$ $u^{(3)}_{1,[p+1]}$ $u^{(1)}_{2,[p+1]}$ $u^{(2)}_{2,[p+1]}$ $u^{(3)}_{2,[p+1]}$

Convergence of Objective Values

- As SVD is involved for the first $k-\mu$ factors,

$$\sum_{r=1}^{R}(\lambda_{r,[p]})^2 \leq \sum_{r=1}^{R}(\lambda^{(1)}_{r,[p+1]})^2 \leq \cdots \leq \sum_{r=1}^{R}(\lambda^{(k-\mu)}_{r,[p+1]})^2.$$

- For last μ factors, by trace maximizing property and Cauchy-Schwarz inequality,

$$\sum_{r=1}^{R}(\lambda^{(k-\mu)}_{r,[p+1]})^2 \leq \sum_{r=1}^{R}\lambda^{(k-\mu)}_{r,[p+1]}\lambda^{(k-\mu+1)}_{r,[p+1]} \leq \cdots$$

$$\leq \sum_{r=1}^{R}\lambda^{(k-1)}_{r,[p+1]}\lambda^{(k)}_{r,[p+1]} \leq \sum_{r=1}^{R}(\lambda^{(k)}_{r,[p+1]})^2 = \sum_{r=1}^{R}(\lambda_{r,[p+1]})^2.$$

Convergence of Objective Values

- As SVD is involved for the first $k-\mu$ factors,

$$\sum_{r=1}^{R}(\lambda_{r,[p]})^2 \leq \sum_{r=1}^{R}(\lambda_{r,[p+1]}^{(1)})^2 \leq \cdots \leq \sum_{r=1}^{R}(\lambda_{r,[p+1]}^{(k-\mu)})^2.$$

- For last μ factors, by trace maximizing property and Cauchy-Schwarz inequality,

$$\sum_{r=1}^{R}(\lambda_{r,[p+1]}^{(k-\mu)})^2 \leq \sum_{r=1}^{R}\lambda_{r,[p+1]}^{(k-\mu)}\lambda_{r,[p+1]}^{(k-\mu+1)} \leq \cdots$$

$$\leq \sum_{r=1}^{R}\lambda_{r,[p+1]}^{(k-1)}\lambda_{r,[p+1]}^{(k)} \leq \sum_{r=1}^{R}(\lambda_{r,[p+1]}^{(k)})^2 = \sum_{r=1}^{R}(\lambda_{r,[p+1]})^2.$$

Assumption A.

- The dominant singular value of the limiting point $C_r^{(\ell)}$ of the corresponding subsequence $\left\{ C_{r,[p_j]}^{(\ell)} \right\}$ are simple .
- The limiting point $V^{(\ell)}\Lambda^{(\ell)}$ of the matrix $V_{[p_j]}^{(\ell)}\Lambda_{[p_j]}^{(\ell)}$ are of full column rank.

Theorem

Under the Assumption A, the sequence $\left\{\mathbf{u}_{r,[p]}^{(\ell)}\right\}$ *generated in Algorithm 1 converges for* $\ell = 1, \ldots, k$, $r = 1, \ldots, R$.

Proof.

- Accumulation points are isolated.
- If subsequences $\left\{\mathbf{u}_{r,[p_j]}^{(\ell)}\right\}$ generated by Algorithm 1 converge simultaneously, then subsequences $\left\{\mathbf{u}_{r,[p_j+1]}^{(\ell)}\right\}$ also converge simultaneously.
- Under Assumption A, $\left\{\mathbf{u}_{r,[p_j]}^{(\ell)}\right\}$ and $\left\{\mathbf{u}_{r,[p_j+1]}^{(\ell)}\right\}$ converge to the same limiting point.

□

Numerical Example

Test Algorithm 1

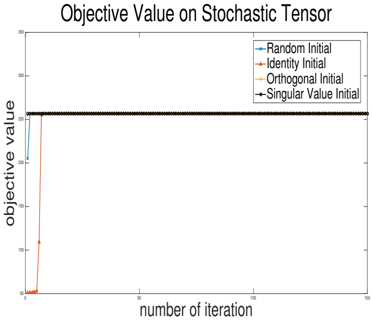
- $\mu = 2$ and $R = 5$;
- First 150 steps.

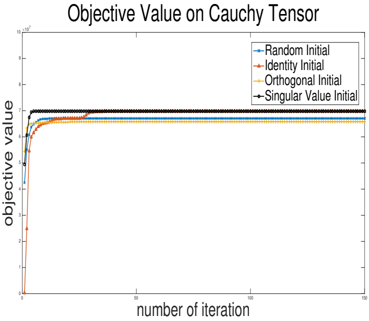
Comparison: by measuring

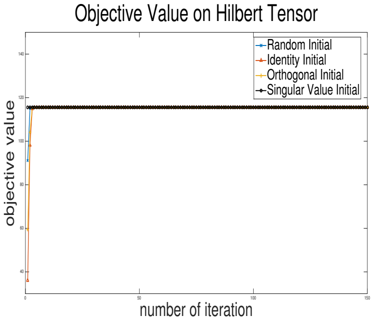
- objective value $\sum_{r=1}^{R} \lambda_r^2$;
- iterate error $\sum_{\ell=1}^{k} \sum_{r=1}^{R} \|\mathbf{u}_{r,[p+1]}^{(\ell)} - \mathbf{u}_{r,[p]}^{(\ell)}\|_2^2$.

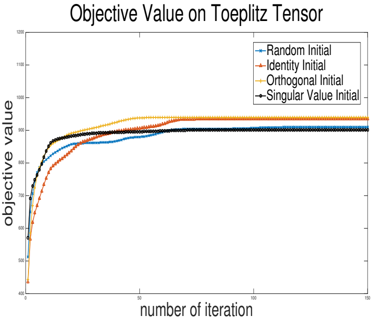
Initial vectors:

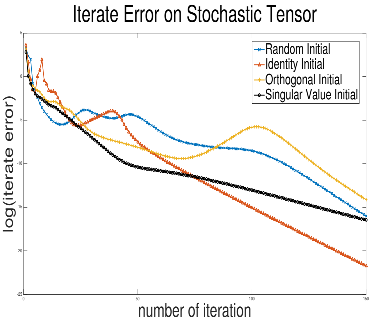
- 'Random Initial'–unit vectors $\mathbf{u}_r^{(\ell)}$ for $\ell = 1, \ldots, k$ and $r = 1, \ldots, R$ are generated randomly to satisfy orthogonality constrain with $\mu = 2$.
- 'Identity Initial'–each $[\mathbf{u}_1^{(\ell)}, \ldots, \mathbf{u}_R^{(\ell)}]$ for $\ell = 1, \ldots, k$ are taken as the first R columns of identity matrices.
- 'Orthogonal Initial'–each $[\mathbf{u}_1^{(\ell)}, \ldots, \mathbf{u}_R^{(\ell)}]$ for $\ell = 1, \ldots, k$ are taken as the first R columns of random orthonormal matrices.
- 'Singular Value Initial'–the major left singular vectors of the unfoldings of the tensors are used as initials.

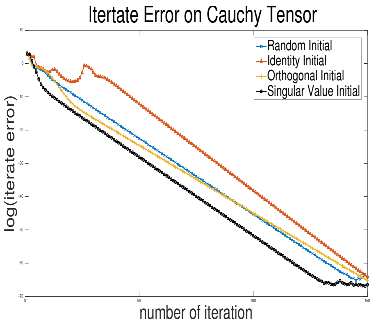

Objective Value on Stochastic Tensor
Random Initial
Identity Initial
Orthogonal Initial
Singular Value Initial
objective value
number of iteration

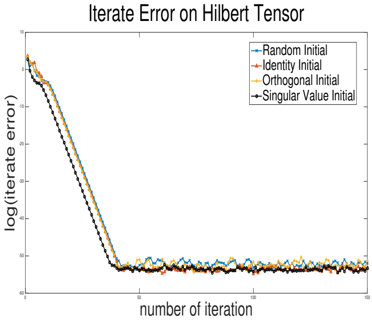

Objective Value on Cauchy Tensor
Random Initial
Identity Initial
Orthogonal Initial
Singular Value Initial
objective value
number of iteration

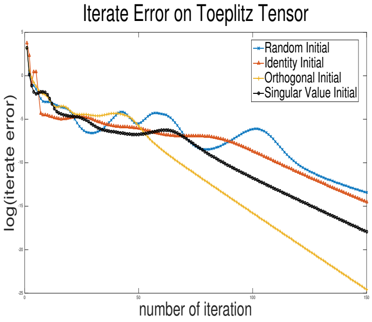

Objective Value on Hilbert Tensor
objective value
number of iteration
Random Initial
Identity Initial
Orthogonal Initial
Singular Value Initial


Objective Value on Toeplitz Tensor
objective value
number of iteration
Random Initial
Identity Initial
Orthogonal Initial
Singular Value Initial


Iterate Error on Stochastic Tensor
Random Initial
Identity Initial
Orthogonal Initial
Singular Value Initial
log(iterate error)
number of iteration
5
0
-5
-10
-15
-20
-25
0
50
100
150


Itertate Error on Cauchy Tensor
Random Initial
Identity Initial
Orthogonal Initial
Singular Value Initial
log(iterate error)
number of iteration
10
0
-10
-20
-30
-40
-50
-60
-70
0
50
100
150


Iterate Error on Hilbert Tensor
Random Initial
Identity Initial
Orthogonal Initial
Singular Value Initial
log(iterate error)
number of iteration


Iterate Error on Toeplitz Tensor
Random Initial
Identity Initial
Orthogonal Initial
Singular Value Initial
log(iterate error)
number of iteration

Conclusions

- An SVD-based algorithm has been presented including
  - Completely orthogonal low rank approximation (Chen et al. 08').
  - Semi-orthogonal low rank approximation (Wang et al. 14').
- The convergence of the proposed algorithm has been analyzed.
- Numerical examples have been provided to illustrate the convergence behavior of proposed algorithm.

Reference

[1] J. CHEN AND Y. SAAD, *On the tensor SVD and the optimal low rank orthogonal approximation of tensors*, SIAM J. Matrix Anal. Appl., 30 (2008/09), pp. 1709–1734.

[2] L. WANG, M. T. CHU AND B. YU, *Orthogonal low rank tensor approximation: Alternating least squares method and its global convergence*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1058–1072.

Questions?

Thank you!