Convergence Analysis on Orthogonal Low Rank Approximation of Tensors

Guan Yu

UCLouvain

Feb 25, 2019 @ KU Louvain

Outline

Outline

Outline

Outline

Tensor Notation

- Tensor of order k:

$$T = [t_{i_1,\dots,i_k}] \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_k}$$

- Tensor of rank 1:

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

  - $\mathbf{u}^{(j)} \in \mathbb{R}^{I_j}$, $j = 1, \dots, k$.

Tensor decomposition: To rewrite the given tensor T as the summation of some rank-1 tensors.(NP-hard)

- Tucker decomposition

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

where $\lambda_{r_1, r_2, \ldots, r_k} \in \mathbb{R}$ and $\mathbf{u}_{r_\ell}^{(\ell)} \in \mathbb{R}^{I_\ell}$ are unit vectors for $\ell = 1, \ldots, k$.

- CP decomposition

$$T = \sum_r \lambda_r \mathbf{u}_r^{(1)} \circ \cdots \circ \mathbf{u}_r^{(k)},$$

where $\lambda_r \in \mathbb{R}$ and $\mathbf{u}_r^{(\ell)} \in \mathbb{R}^{I_\ell}$ are unit vectors for $\ell = 1, \ldots, k$.

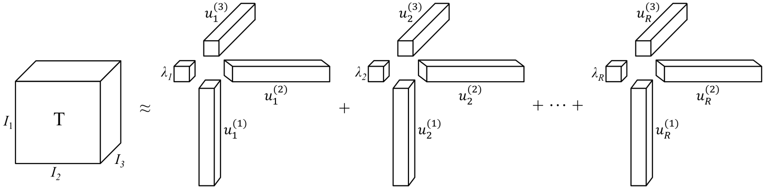
Tensor Approximation

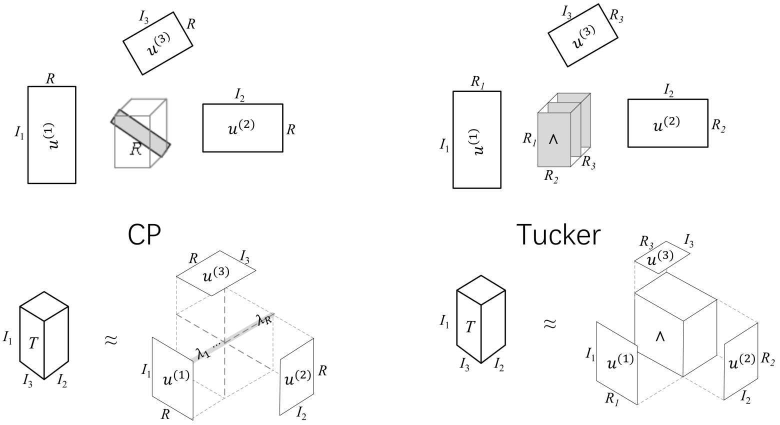
Tensor approximation: To find another tensor $\hat{T}$ with certain properties to minimize the error $\|T - \hat{T}\|_F$ for a given T

- Low rank CP approximation:
  Determine unit vectors $\mathbf{u}_r^{(\ell)} \in \mathbb{R}^{I_\ell}$, $\ell = 1, \ldots k$ and scalars λ_r to minimize

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

Low Rank CP Approximation




CP
Tucker
$u^{(1)}$
$u^{(2)}$
$u^{(3)}$
T
Λ
I_1
I_2
I_3
R
R_1
R_2
R_3
λ_1
λ_R
$\approx$

Applications of CP Approximation

- psychometrics, chemometrics, neuroscience;
- data mining, multiple access wireless communication systems, blind signal separation, image identification;
- telecommunications, independent component analysis (ICA), sensor array processing
- polarization sensitive array analysis.

Background of Algorithm

- damped Gauss-Newton (dGN) and a variant called PMF3;
- non-linear conjugate gradient approach, Levenberg-Marquardt method;
- Alternating Least Squares (ALS) algorithms, Alternating Slice-wise Diagonalization (ASD) and Self Weighted Alternating TriLinear Decomposition (SWATLD);
- Enhanced Line Search (ELS), Tikhonov regularization on the ALS.

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

Let $\mathbf{u}_1, \mathbf{v}_1 \in \mathbb{R}^{I_1}$, $\mathbf{u}_2, \mathbf{v}_2 \in \mathbb{R}^{I_2}$, and $\mathbf{u}_3, \mathbf{v}_3 \in \mathbb{R}^{I_3}$ be vectors such that each pair $\mathbf{u}_i, \mathbf{v}_i$ is linearly independent. Define tensor

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

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. But $\|T_n - T\| \to 0$ as $n \to \infty$. Therefore, T does not have a best rank-2 approximation.

Solution

- Orthogonality requirement ensures the existence. For $i = 1, \ldots, k$ and $1 \le r \le R$, $\mathbf{u}_r^{(i)}$ are unit vectors.
  1. Complete orthogonality:
     For all $i = 1, \ldots, k$,
     $$\langle \mathbf{u}_{r_1}^{(i)}, \mathbf{u}_{r_2}^{(i)} \rangle = 0, \quad \forall 1 \le r_1 \ne r_2 \le R.$$
  2. Semi-orthogonality:
     There is one i such that
     $$\langle \mathbf{u}_{r_1}^{(i)}, \mathbf{u}_{r_2}^{(i)} \rangle = 0, \quad \forall 1 \le r_1 \ne r_2 \le R.$$
  3. μ-orthogonality:
     For some $1 \le i_1 < \cdots < i_\mu \le k$,
     $$\left\langle \mathbf{u}_{r_1}^{(i_1)}, \mathbf{u}_{r_2}^{(i_1)} \right\rangle = 0, \cdots, \left\langle \mathbf{u}_{r_1}^{(i_\mu)}, \mathbf{u}_{r_2}^{(i_\mu)} \right\rangle = 0, \quad \forall 1 \le r_1 \ne r_2 \le R.$$

Orthogonal Low Rank Approximation

- Given $T \in \mathbb{R}^{I_1 \times \ldots \times I_k}$, determine
  - unit vectors $\mathbf{u}_r^{(i)} \in \mathbb{R}^{I_i}$, $i = 1, \ldots k$,
  - scalars $\lambda_r \in \mathbb{R}$,

  such that

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

  is minimized subject to the mutual orthogonality condition that

$$\langle H_{r_1}, H_{r_2} \rangle = \prod_{i=1}^{k} \left\langle \mathbf{u}_{r_1}^{(i)}, \mathbf{u}_{r_2}^{(i)} \right\rangle = \delta_{r_1 r_2}, \quad \text{for all} \quad 1 \leq r_1, r_2 \leq R,$$

Our Problem

Orthogonal low rank approximation:

$$\begin{cases} \min \left\| T - \sum_{r=1}^{R} \lambda_r \bigotimes_{i=1}^{k} \mathbf{u}_r^{(i)} \right\|_F^2, \\ \text{subject to } \mu - \text{orthogonality constraint.} \end{cases} \quad (2)$$

Open Question

- Complete orthogonal low rank approximation are studied in [1].
- Semi-orthogonal low rank approximation of tensors are studied in [2].
- It is interesting to impose orthogonality to more than one factor matrix.
  - [2] pointed that "More study is needed".
  - [2] addressed that "The question of more than one semi-orthogonal factor matrix, except for the case of complete orthogonality, remains open".

Linear Mapping

Given a fixed partitioning $[\![k]\!] = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$, we shall regard an order-k tensor $T \in \mathbb{R}^{I_1 \times \dots \times I_k}$ as a "matrix representation" of a linear operator mapping order-s tensors to order-t tensors. Specifically, we identify T with the linear map

$$\mathscr{T}_{\boldsymbol{\beta}} : \mathbb{R}^{I_{\alpha_1} \times \dots \times I_{\alpha_s}} \to \mathbb{R}^{I_{\beta_1} \times \dots \times I_{\beta_t}},$$

such that for any $S \in \mathbb{R}^{I_{\alpha_1} \times \dots \times I_{\alpha_s}}$,

Linear Mapping

we have

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

where

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

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

An Equivalent Formulation

- The optimal scales λ_r can also be interpreted as the length of the projection of the "vector" T onto the "unit vector" H_r under the Frobenius inner product,

$$\lambda_r = \left\langle T, \bigotimes_{i=1}^{k} \mathbf{u}_r^{(i)} \right\rangle = \left\langle T \circledast_\ell \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_r^{(i)} \circ \bigotimes_{i=\ell+1}^{k} \mathbf{u}_r^{(i)} \right), \mathbf{u}_r^{(\ell)} \right\rangle . \tag{3}$$

- The orthogonal low rank approximation problem (2) can be reformulated as

$$\begin{cases} \max \sum_{r=1}^{R} \lambda_r^2, \\ \text{subject to the } \mu - \text{orthogonality constraint.} \end{cases} \tag{4}$$

Alternating Least Squares Algorithm

- 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.
  - [2] proved convergence globally.
  - Numerical computation of the completely orthogonal in [1].

Contribution

- 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 algorithm is analyzed for both objective function and iterates themselves.
- Numerical performance is demonstrated.

Algorithm Description

- How can we update $\mathbf{u}_r^{(\ell)}, \mathbf{u}_r^{(\ell+1)}$ to obtain "better" ones?
  - For any $1 \leq \ell \leq k - \mu - 1$ and $r = 1, 2, \ldots, R$, 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 $C_r^{(\ell)}$.
  - By Eckart-Young theorem, update $\mathbf{u}_r^{(\ell)}$ by $\tilde{\mathbf{u}}_r^{(\ell)}$ and $\mathbf{u}_r^{(\ell+1)}$ by $\tilde{\mathbf{u}}_r^{(\ell+1)}$.

- The update of first $k - \mu$ factors has been provided.
- To address the orthogonality constraint, how to update $\mathbf{u}_r^{(\ell)}$ for $k - \mu + 1 \leq \ell \leq k$?
  - Check the optimality condition to ensure the monotone of the objective value.

Lagrangian

- The Lagrangian for the optimization problem (4) (i.e., (2)) is

$$\mathcal{L} := \sum_{r=1}^{R} \lambda_r^2 - \sum_{\ell=1}^{k} \sum_{r=1}^{R} \rho_r^{(\ell)} \left(\left\langle \mathbf{u}_r^{(\ell)}, \mathbf{u}_r^{(\ell)} \right\rangle - 1 \right)$$

$$- \sum_{1 \le r_1 < r_2 \le R} \sum_{i=k-\mu+1}^{k} \alpha_{r_1 r_2}^{(\ell)} \left\langle \mathbf{u}_{r_1}^{(\ell)}, \mathbf{u}_{r_2}^{(\ell)} \right\rangle,$$

where λ_r is given by (3) and $\rho_r^{(\ell)}, \alpha_{r_1 r_2}^{(\ell)}$ are Lagrange multipliers.

Optimality Condition

The first order optimality condition for a stationary point is to satisfy for $r = 1, \ldots, R$,

$$\lambda_r T \circledast_\ell \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_r^{(i)} \circ \bigotimes_{i=\ell+1}^{k} \mathbf{u}_r^{(i)} \right) = \rho_r^{(\ell)} \mathbf{u}_r^{(\ell)}, \ell = 1, \ldots, k - \mu.$$

and

$$\lambda_r T \circledast_\ell \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_r^{(\ell)} \circ \bigotimes_{i=\ell+1}^{k} \mathbf{u}_r^{(i)} \right) = \rho_r^{(\ell)} \mathbf{u}_r^{(\ell)} + \sum_{r_1 < r} \frac{\alpha_{r_1 r}^{(\ell)}}{2} \mathbf{u}_{r_1}^{(\ell)} + \sum_{r < r_2} \frac{\alpha_{r r_2}^{(\ell)}}{2} \mathbf{u}_{r_2}^{(\ell)}$$

$$\ell = k - \mu + 1, \ldots, k.$$

It follows from the orthogonality condition that

$$V^{(\ell)}\Lambda^{(\ell)} = U^{(\ell)}S^{(\ell)}, \; S^{(\ell)} \text{ is symmetric}, \quad \ell = k-\mu+1, \cdots, k,$$

where

$$\mathbf{v}_r^{(\ell)} = T_{\circledast_\ell}\left(\bigotimes_{i=1}^{\ell-1}\mathbf{u}_r^{(\ell)} \otimes \bigotimes_{i=\ell+1}^{k}\mathbf{u}_r^{(i)}\right), \quad \ell = 1,\ldots,k-\mu+1, r = 1,\ldots,R,$$

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

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

Trace Maximizing Property

Lemma

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

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

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

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

$$V^{(\ell)}\Lambda^{(\ell)} = \tilde{U}^{(\ell)}\tilde{S}^{(\ell)},$$

    where $\tilde{U}^{(\ell)}$ is column orthogonal and $\tilde{S}^{(\ell)}$ is symmetric and positive semi-definite.
  -

$$\tilde{\lambda}_r^{(\ell)} = \left\langle \mathbf{v}_r^{(\ell)},\ \tilde{\mathbf{u}}_r^{(\ell)} \right\rangle, \quad \ell = 1,\ldots,k-\mu+1, \quad r = 1,\ldots,R.$$

$$\lambda_r^{(\ell)} = \left\langle \mathbf{v}_r^{(\ell)},\ \mathbf{u}_r^{(\ell)} \right\rangle, \quad \ell = 1,\ldots,k-\mu+1, \quad r = 1,\ldots,R.$$

- By trace maximizing property,

$$\sum_{r=1}^{R}(\lambda_r^{(\ell)})^2 = \mathrm{Trace}\left((U^{(\ell)})^T V^{(\ell)} \Lambda^{(\ell)}\right)$$

$$\leq \mathrm{Trace}\left((\tilde{U}^{(\ell)})^T V^{(\ell)} \Lambda^{(\ell)}\right) = \sum_{r=1}^{R} \tilde{\lambda}_r^{(\ell)} \lambda_r^{(\ell)}.$$

- By Cauchy-Schwarz inequality,

$$\sum_{r=1}^{R}(\lambda_r^{(\ell)})^2 \leq \sum_{r=1}^{R}(\tilde{\lambda}_r^{(\ell)})^2, \quad \ell = 1, \ldots, k - \mu + 1,$$

and the equality holds if and only if

$$\lambda_r^{(\ell)} = \tilde{\lambda}_r^{(\ell)}, \quad \ell = 1, \ldots, k - \mu + 1, \quad r = 1, \ldots, R.$$

Require: Starting unit vectors $\mathbf{u}_{r.[0]}^{(\ell)} \in \mathbb{R}^{I_\ell}$ and $\mathbf{u}_{i,[0]}^{(\ell)} \perp \mathbf{u}_{j,[0]}^{(\ell)}$ for $\ell = k-\mu+1, \cdots, k$

$\tau := k-\mu-1$
if $k-\mu$ is odd **then**
  $\tau := k-\mu-2$
end if
for $p = 0, 1, \ldots,$ **do**
  for $\ell = 1, 3, \cdots, \tau$ **do**
   $\beta_\ell = (\ell, \ell+1)$ **do**
   for $r = 1, 2, \ldots, R$,
    $C_{r,[p+1]}^{(\ell)} = T \circledast_{\beta_\ell} \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_{r,[p+1]}^{(i)} \circ \bigotimes_{i=\ell+2}^{k} \mathbf{u}_{r,[p]}^{(i)} \right)$ {A matrix of size $I_\ell \times I_{\ell+1}$}
    $[\mathbf{u}, s, \mathbf{v}] = \text{svds}(C_{r,[p+1]}^{(\ell)}, 1)$ {Dominant singular value triplet via Matlab routine svds;assume uniqueness}
    if $\mathbf{u}_1 < 0$ **then**
     $\mathbf{u} = -\mathbf{u}, \mathbf{v} = -\mathbf{v}$
    end if
    $\mathbf{u}_{r,[p+1]}^{(\ell)} := \mathbf{u}$
    $\mathbf{u}_{r,[p+1]}^{(\ell+1)} := \mathbf{v}$ {if $k-\mu$ is even, use $\hat{\mathbf{u}}_{r,[p+1]}^{(k-\mu-2)} := \mathbf{v}$}
    $\lambda_{r,[p+1]}^{(\ell)} := s, \quad \lambda_{r,[p+1]}^{(\ell+1)} := s$ {if $k-\mu$ is odd, use $\hat{\lambda}_{r,[p+1]}^{(k-\mu-2)} := s$}
   end for
  end for

for $\ell = k-\mu+1, \cdots, k$ **do**

 for $r = 1, 2, \ldots, R$, **do**

$$\mathbf{v}_{r,[p+1]}^{(\ell)} = T \circledast_{\ell} \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_{r,[p+1]}^{(i)} \circ \bigotimes_{i=\ell+1}^{k} \mathbf{u}_{r,[p]}^{(i)} \right) \{\text{define columns of } V_{[p+1]}^{(\ell)}\}$$

$$\hat{\lambda}_{r,[p+1]}^{(\ell)} := \langle \mathbf{v}_{r,[p+1]}^{(\ell)}, \mathbf{u}_{r,[p]}^{(\ell)} \rangle \{\text{define diagonals of } \Lambda_{[p+1]}^{(\ell)}\}$$

 end for

$$[U_{[p+1]}^{(\ell)}, S_{[p+1]}^{(\ell)}] = \text{poldec}(V_{[p+1]}^{(\ell)} \Lambda_{[p+1]}^{(\ell)})$$

 for $r = 1, 2, \ldots, R$, **do**

$$\mathbf{u}_{r,[p+1]}^{(\ell)} := U_{[p+1]}^{(\ell)}(:, r)$$

$$\lambda_{r,[p+1]}^{(\ell)} := S_{[p+1]}^{(\ell)}(r, r) (= \langle \mathbf{v}_{r,[p+1]}^{(\ell)}, \mathbf{u}_{r,[p+1]}^{(\ell)} \rangle)$$

 end for

end for

end for

Convergence of Objective Values

- As SVD is involved for the first $k - \mu$ factors, the generalized Rayleigh quotients are bounded and monotone increasing.
- Polar decomposition is applied for last μ factors, by trace maximizing property.

Lemma

Assume that a^* *is an isolated accumulation point of a sequence* $\{a_k\}$ *such that for every subsequence* $\{a_{k_j}\}$ *converging to* a^**, there is an infinite subsequence* $\{a_{k_{j_i}}\}$ *such that* $|a_{k_{j_i}+1} - a_{k_{j_i}}| \to 0$*. Then the whole sequence* $\{a_k\}$ *converges to* a^**.*

Accumulation Points

For $r = 1, \ldots, R$,

$$\begin{cases} T \circledast_\ell \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_r^{(i)} \circ \bigotimes_{i=\ell+1}^{k} \mathbf{u}_r^{(i)} \right) = \left\langle T,\ \bigotimes_{i=1}^{k} \mathbf{u}_r^{(i)} \right\rangle \mathbf{u}_r^{(\ell)}, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \ell = 1, \ldots, k-\mu, \\ T \circledast_\ell \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_r^{(i)} \circ \bigotimes_{i=\ell+1}^{k} \mathbf{u}_r^{(i)} \right) \\ = \sum_{t=1}^{R} \left\langle T, \bigotimes_{i=1}^{\ell-1} \mathbf{u}_r^{(i)} \circ \mathbf{u}_t^{(\ell)} \circ \bigotimes_{i=\ell+1}^{k} \mathbf{u}_r^{(i)} \right\rangle \mathbf{u}_t^{(\ell)}, \quad \ell = k-\mu+1, \cdots, k. \end{cases}$$

Isolation

Lemma

For almost all tensors $T \in \mathbb{R}^{I_1 \times \cdots \times I_k}$, the accumulation points of any sequence generated by Algorithm 1 are necessarily isolated.

- A polynomial system with leading coefficients from entries of T.
- By the theory of parameter continuation.

Assumption A

We say that a given tensor $T \in \mathbb{R}^{I_1 \times \cdots \times I_k}$ satisfies Assumption A if for every convergent subsequence $\left\{\mathbf{u}_{r,[p_j]}^{(\ell)}\right\}$ generated by Algorithm 1, the dominant singular value of the limiting point $C_r^{(\ell)}$ of the corresponding subsequence $\left\{C_{r,[p_j]}^{(\ell)}\right\}$ are simple for all $\ell = 1, \ldots, k - \mu$, $r = 1, \ldots, R$. Moreover, the limiting point $V^{(\ell)}\Lambda^{(\ell)}$ of the matrix $V_{[p_j]}^{(\ell)}\Lambda_{[p_j]}^{(\ell)}$ for $\ell = k - \mu + 1, \cdots, k$ are of full column rank.

Lemma

For all $\ell = 1, \ldots, k$, $r = 1, \ldots, R$, 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. Furthermore, 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.

- Subsequence $\left\{C_{r,[p_j+1]}^{(\ell)}\right\}$ and $\left\{\mathbf{u}_{r,[p_j+1]}^{(\ell)}\right\}$ converge.
- Converge to the same limiting point

Taking limit,

$$
\begin{aligned}
& \lambda_r = \langle \mathbf{u}_r^{(1)},\ \tilde{C}_r^{(1)} \mathbf{u}_r^{(2)} \rangle \\
\leq\ & \tilde{\lambda}_r^{(1)} = \langle \tilde{\mathbf{u}}_r^{(1)},\ \tilde{C}_r^{(1)} \tilde{\mathbf{u}}_r^{(2)} \rangle = \langle \mathbf{u}_r^{(3)},\ \tilde{C}_r^{(3)} \mathbf{u}_r^{(4)} \rangle = \tilde{\lambda}_r^{(2)} \\
\leq\ & \tilde{\lambda}_r^{(3)} = \langle \tilde{\mathbf{u}}_r^{(3)},\ \tilde{C}_r^{(3)} \tilde{\mathbf{u}}_r^{(4)} \rangle = \langle \mathbf{u}_r^{(5)},\ \tilde{C}_r^{(5)} \mathbf{u}_r^{(6)} \rangle = \tilde{\lambda}_r^{(4)} \\
\leq\ & \cdots \\
\leq\ & \tilde{\lambda}_r^{(\ell)} = \langle \tilde{\mathbf{u}}_r^{(\ell)},\ \tilde{C}_r^{(\ell)} \tilde{\mathbf{u}}_r^{(\ell+1)} \rangle = \langle \mathbf{u}_r^{(\ell+2)},\ \tilde{C}_r^{(\ell+2)} \mathbf{u}_r^{(\ell+3)} \rangle = \tilde{\lambda}_r^{(\ell+1)} \\
\leq\ & \cdots \\
\leq\ & \begin{cases} \tilde{\lambda}_r^{(k-\mu-3)} = \langle \tilde{\mathbf{u}}_r^{(k-\mu-3)},\ \tilde{C}_r^{(k-\mu-3)} \tilde{\mathbf{u}}_r^{(k-\mu-2)} \rangle & \\ = \langle \mathbf{u}_r^{(k-\mu-1)},\ \tilde{C}_r^{(k-\mu-1)} \mathbf{u}_r^{(k-\mu)} \rangle = \tilde{\lambda}_r^{(k-\mu-2)} & \text{if } k-\mu \text{ is even} \\ \tilde{\lambda}_r^{(k-\mu-2)} = \langle \tilde{\mathbf{u}}_r^{(k-\mu-2)},\ \tilde{C}_r^{(k-\mu-2)} \hat{\mathbf{u}}_r^{(k-\mu-1)} \rangle & \\ = \langle \hat{\mathbf{u}}_r^{(k-\mu-1)},\ \tilde{C}_r^{(k-\mu-1)} \mathbf{u}_r^{(k-\mu)} \rangle = \hat{\lambda}_r^{(k-\mu-1)} & \text{if } k-\mu \text{ is odd} \end{cases} \\
\leq\ & \tilde{\lambda}_r^{(k-\mu-1)} = \langle \tilde{\mathbf{u}}_r^{(k-\mu-1)},\ \tilde{C}_r^{(k-\mu-1)} \tilde{\mathbf{u}}_r^{(k-\mu)} \rangle = \tilde{\lambda}_r^{(k-\mu)}.
\end{aligned}
$$

For $r = 1, \ldots, R$, we obtain

$$\begin{cases} \lambda_r = \tilde{\lambda}_r^{(1)} = \cdots = \tilde{\lambda}_r^{(k-\mu-1)} = \tilde{\lambda}_r^{(k-\mu)}, & \text{if } k-\mu \text{ is even} \\ \lambda_r = \tilde{\lambda}_r^{(1)} = \cdots = \tilde{\lambda}_r^{(k-\mu-1)} = \hat{\lambda}_r^{(k-\mu-1)} = \tilde{\lambda}_r^{(k-\mu)}, & \text{if } k-\mu \text{ is odd} \end{cases}$$

- Monotone increasing of objective value.
- $\sum_{r=1}^{R}(\lambda_r)^2 = \sum_{r=1}^{R}(\tilde{\lambda}_r)^2$.

- $\mathbf{u}_r^{(\ell)} = \tilde{\mathbf{u}}_r^{(\ell)}$ hold for $\ell = 1, \ldots, k-\mu$, $r = 1, \ldots, R$.
-
$$\tilde{V}^{(\ell)} = \lim V_{[p_j+1]}^{(\ell)} = \lim V_{[p_j]}^{(\ell)} = V^{(\ell)},$$
$$\tilde{\Lambda}^{(\ell)} = \lim \Lambda_{[p_j+1]}^{(\ell)} = \lim \Lambda_{[p_j]}^{(\ell)} = \Lambda^{(\ell)},$$
- For $\ell = k-\mu+1, \ldots, k$, combined with Assumption A.
$$\tilde{U}^{(\ell)} = \lim U_{[p_j+1]}^{(\ell)} = \lim U_{[p_j]}^{(\ell)} = U^{(\ell)}.$$

- **Theorem**

  For almost all tensors T satisfying Assumption A, the sequence $\left\{\mathbf{u}_{r,[p]}^{(\ell)}\right\}$ generated in Algorithm 1 converges for $\ell = 1, \cdots, k$, $r = 1, \cdots, R$.

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

Test tensors $R^{20\times 16\times 10\times 32}$:

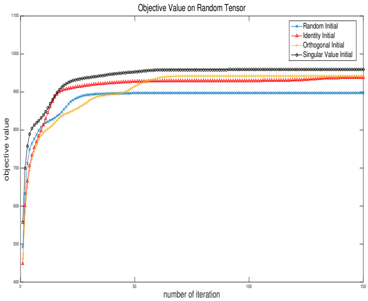
- Random tensor: randomly generate.
- Tensor1: randomly generate a rank-5 order-4 tensor, add a noise tensor which is generated by $10^{-4} * randn(20, 16, 10, 32)$.
- Tensor2: randomly generate a rank-5 order-4 tensor, add a noise tensor which is generated by $10^{-2} * randn(20, 16, 10, 32)$.
- Stochastic tensor:
$$t_{i_1,i_2,i_3,i_4} = \begin{cases} c & i_1 \neq i_2, i_2 \neq i_3, i_3 \neq i_4 \\ 0 & i_1 = i_2, i_2 \neq i_3, i_3 \neq i_4 \\ 1/20 & \text{otherwise} \end{cases}$$
, where c is randomly in $(0, 1)$ by the homogenous distribution such as $\sum_{i_1\in[\![20]\!]} t_{i_1,i_2,i_3,i_4} = 1$ with $i_j \neq i_{j+1}, j = 1, 2, 3$.

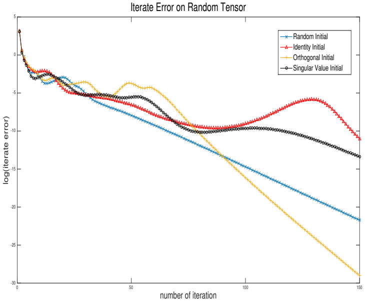
- Cauchy tensor: $t_{i_1,i_2,i_3,i_4} = \frac{1}{c(i_1)+c(i_2)+c(i_3)+c(i_4)}$, where c is a random vector with size 32.
- Hilbert tensor: $t_{i_1,i_2,i_3,i_4} = \frac{1}{i_1+i_2+i_3+i_4-3}$.
- Toeplitz tensor: $t_{i_1+j,i_2+j,i_3+j,i_4+j} = t_{i_1,i_2,i_3,i_4}$ for $j \in [\![min(20-i_1, 16-i_2, 10-i_3, 32-i_4)]\!]$.

Initial vectors:

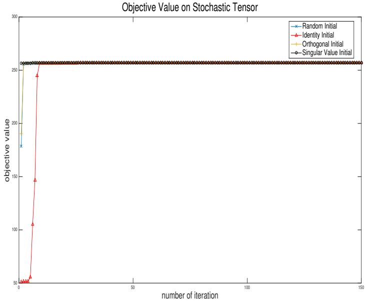
- 'Random Initial'–unit vectors $\mathbf{u}_r^{(\ell)}$ for $\ell = 1, \ldots, k$ and $r = 1, \ldots, R$ are generated randomly to satisfy orthogonality constrain with $\mu = 2$.
- 'Identity Initial'–each $[\mathbf{u}_1^{(\ell)}, \ldots, \mathbf{u}_R^{(\ell)}]$ for $\ell = 1, \ldots, k$ are taken as 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.

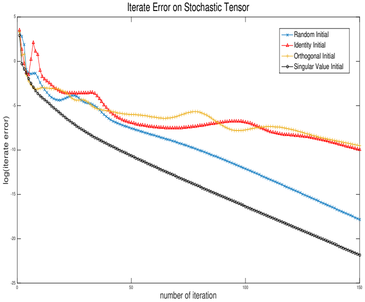
Comparison on Random Tensor



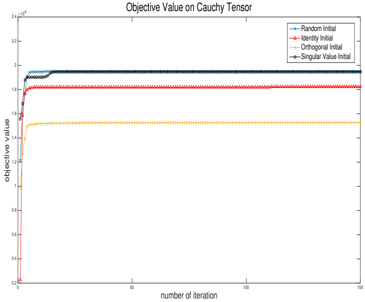


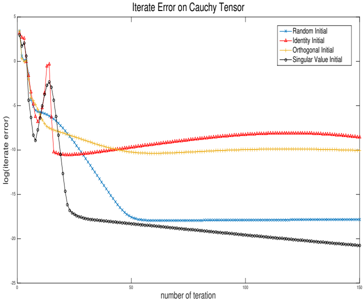
Comparison on Stochastic Tensor



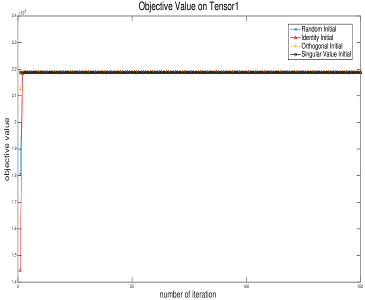


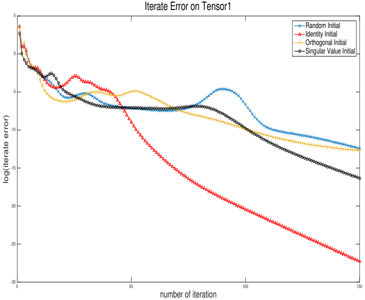
Comparison on Cauchy Tensor



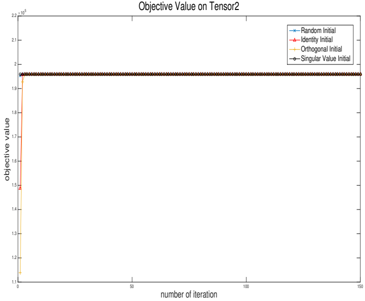


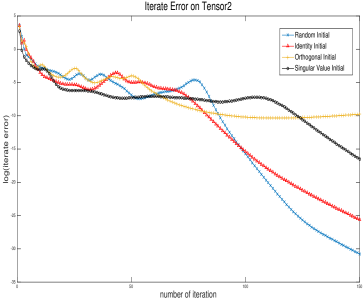
Comparison on Tensor1



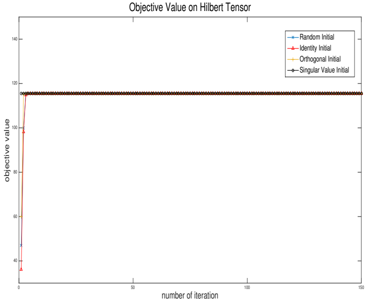


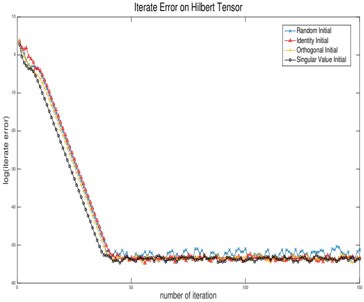
Comparison on Tensor2



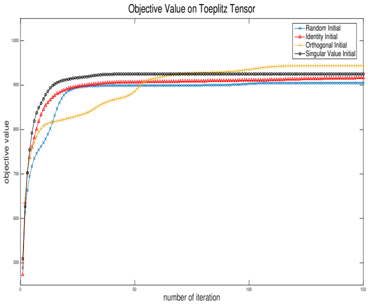


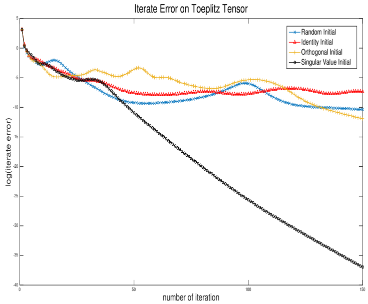
Comparison on Hilbert Tensor





Comparison on Toeplitz Tensor





Observation

Objective value:

- Objective value satisfies the monotone increasing property for each iteration;
- Algorithm 1 is more effective on structured tensor than random tensor;
- For different initial vectors, the objective values may be different for the same test tensor, that is, iterates may converge to different limit points.
  - It is interesting to study for what tensors or what initial guesses Algorithm 1 converges to the global optimum [1].

Observation

Iterates error:

- Iterates converge, but they are not monotone in each step.
- Iterates converge but slower than that of objective values.
- When it comes to the qualities of the final approximation, among 4 different initial vectors, no any one does offer obvious advantage.

Conclusion

- An SVD-based algorithm has been presented including
  - Completely orthogonal low rank approximation [1].
  - Semi-orthogonal low rank approximation [2].
- 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.

Thank you!