A new truncation strategy for the higher-order singular value decomposition

Nick Vannieuwenhoven
KU Leuven, Belgium

OPTEC Seminar on Tensors, Computing, Optimization, and Signal Processing,
KU Leuven, Leuven, Belgium

May 2, 2011

Joint work with Raf Vandebril and Karl Meerbergen
1 Introduction

2 Orthogonal Tucker approximation

3 Truncated HOSVD

4 Sequentially truncated HOSVD
   • Definition
   • Properties

5 Numerical examples
   • Handwritten digit classification
   • Compression of simulation results
   • Rank-reduction reconstruction

6 Conclusions
Overview

1. Introduction
2. Orthogonal Tucker approximation
3. Truncated HOSVD
4. Sequentially truncated HOSVD
   - Definition
   - Properties
5. Numerical examples
   - Handwritten digit classification
   - Compression of simulation results
   - Rank-reduction reconstruction
6. Conclusions
Notation 1 — Mode $k$ vector space

A tensor $\mathcal{A}$ of order $d$ is an object in the tensor product of $d$ vector spaces:

$$
\mathcal{A} \in \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2} \otimes \ldots \otimes \mathbb{R}^{n_d} \sim \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_d}
$$

A 3$^{\text{rd}}$ order tensor has 3 associated vector spaces:

- **Mode 1 vectors** ($\mathbb{R}^{n_1}$)
- **Mode 2 vectors** ($\mathbb{R}^{n_2}$)
- **Mode 3 vectors** ($\mathbb{R}^{n_3}$)
Notation II — Unfolding

\[ A = \begin{array}{c}
\in \mathbb{R}^{n_1 \times n_2 \times n_3} \\
\text{Mode 2 unfolding}
\end{array} \]

\[ A_{(2)} = \begin{array}{c}
\in \mathbb{R}^{n_2 \times n_1 n_3}
\end{array} \]
Frobenius norm:

\[ \| \mathcal{A} \|^2 := \sum_{i_1, i_2, i_3} A_{i_1, i_2, i_3}^2. \]

Multilinear multiplication:

\[ [(I, M_2, I) \cdot \mathcal{A}]_{(2)} := M_2 A_{(2)}. \]
\[ (M_1, M_2, M_3) \cdot \mathcal{A} := (M_1, I, I) \cdot (I, M_2, I) \cdot (I, I, M_3) \cdot \mathcal{A}. \]

Projection of mode 2 vectors on span of \( U_2 \) (orthogonal columns):

\[ \pi_2 \mathcal{A} := (I, U_2 U_2^T, I) \cdot \mathcal{A} \]

Projection of mode 2 vectors on complement of \( U_2 \):

\[ \pi_{2 \perp} \mathcal{A} := \mathcal{A} - \pi_2 \mathcal{A} \]
Overview

1 Introduction
2 Orthogonal Tucker approximation
3 Truncated HOSVD
4 Sequentially truncated HOSVD
  • Definition
  • Properties
5 Numerical examples
  • Handwritten digit classification
  • Compression of simulation results
  • Rank-reduction reconstruction
6 Conclusions
Best rank-\((r_1, r_2, r_3)\) approximation problem:

\[
\begin{align*}
\min_{\text{rank}(\mathcal{B}) \leq (r_1, r_2, r_3)} \| \mathcal{A} - \mathcal{B} \|_F &= \\
\min_{U_i \in O(n_i, r_i)} \| \mathcal{A} - (U_1 U_1^T, U_2 U_2^T, \ldots, U_d U_d^T) \cdot \mathcal{A} \|_F.
\end{align*}
\]

with \(O(n_i, r_i)\) the group of \(n_i \times r_i\) matrices with orthonormal columns.

- Optimum is found by **orthogonal projection** onto a new, optimal tensor basis, but
- no closed solution known.
Orthogonal Tucker approximation I — Definition

\[ \mathcal{A} \approx (\hat{U}_1, \hat{U}_2, \hat{U}_3) \cdot \hat{S} \]

Rank \((r_1, r_2, r_3)\) orthogonal Tucker approximation to \(\mathcal{A}\)

Columns of \(\begin{cases} \hat{U}_1 \in \mathbb{R}^{n_1 \times r_1} \\ \hat{U}_2 \in \mathbb{R}^{n_2 \times r_2} \\ \hat{U}_3 \in \mathbb{R}^{n_3 \times r_3} \end{cases} \)

can be extended to a basis of \(\begin{cases} \mathbb{R}^{n_1} \\ \mathbb{R}^{n_2} \\ \mathbb{R}^{n_3} \end{cases} \)
If

\[ A \approx \hat{A} := \pi_1 \pi_2 \pi_3 A = (U_1 U_1^T, U_2 U_2^T, U_3 U_3^T) \cdot A. \]

Then an error expression is

\[
\| A - \pi_1 \pi_2 \pi_3 A \|^2 = \| \pi_2^\perp A \|^2 + \| \pi_1 \pi_2 A \|^2 + \| \pi_3 \pi_1 \pi_2 A \|^2
\]

with upper bound

\[
\| A - \pi_1 \pi_2 \pi_3 A \|^2 \leq \| \pi_2^\perp A \|^2 + \| \pi_1 \pi_2 A \|^2 + \| \pi_3 \pi_1 \pi_2 A \|^2
\]
Overview

1. Introduction
2. Orthogonal Tucker approximation
3. **Truncated HOSVD**
4. Sequentially truncated HOSVD
   - Definition
   - Properties
5. Numerical examples
   - Handwritten digit classification
   - Compression of simulation results
   - Rank-reduction reconstruction
6. Conclusions
**Truncated Higher-Order SVD (T-HOSVD)**

Recall upper bound?

\[
\| \mathbf{A} - \pi_1 \pi_2 \pi_3 \mathbf{A} \|^2 \leq \| \pi_2 \mathbf{A} \|^2 + \| \pi_1 \mathbf{A} \|^2 + \| \pi_3 \mathbf{A} \|^2
\]

Minimize it!

\[
\min_{\pi_1, \pi_2, \pi_3} \| \mathbf{A} - \pi_1 \pi_2 \pi_3 \mathbf{A} \|^2 \leq \min_{\pi_1, \pi_2, \pi_3} \sum_{k=1}^{3} \| \pi_k \mathbf{A} \|^2,
\]

\[
= \sum_{k=1}^{3} \min_{\pi_k} \| \pi_k \mathbf{A} \|^2.
\]

Minimum given by \( r_k \) first singular vectors in every mode \( k \)!
Algorithm

Rank \((r_1, r_2, r_3)\) T-HOSVD:

\begin{verbatim}
for every mode \(k\) do
    Compute rank \(r_k\) truncated SVD:
    \[
    A(k) = \begin{bmatrix}
    U_k & U_k^\perp \\
    \end{bmatrix}
    \begin{bmatrix}
    \bar{S}_k & S_k^\perp \\
    \end{bmatrix}
    \begin{bmatrix}
    V_k & V_k^\perp \\
    \end{bmatrix}
\end{bmatrix}
    
end for

Project:
\[
\bar{S} = (U_1^T, U_2^T, U_3^T) \cdot A
\]
\end{verbatim}
Overview

1. Introduction
2. Orthogonal Tucker approximation
3. Truncated HOSVD
4. Sequentially truncated HOSVD
   - Definition
   - Properties
5. Numerical examples
   - Handwritten digit classification
   - Compression of simulation results
   - Rank-reduction reconstruction
6. Conclusions
Sequentially truncated HOSVD (ST-HOSVD)

Recall error expression?

\[
\| \mathcal{A} - \pi_1 \pi_2 \pi_3 \mathcal{A} \|^2 = \| \pi_2 \mathcal{A} \|^2 + \| \pi_1 \pi_2 \mathcal{A} \|^2 + \| \pi_3 \pi_1 \pi_2 \mathcal{A} \|^2
\]

(Try to) minimize it!

\[
\begin{align*}
\min_{\pi_1, \pi_2, \pi_3} & \quad \| \mathcal{A} - \pi_1 \pi_2 \pi_3 \mathcal{A} \|^2 \\
= & \quad \min_{\pi_1, \pi_2, \pi_3} \left[ \| \pi_1 \mathcal{A} \|^2 + \| \pi_2 \pi_1 \mathcal{A} \|^2 + \| \pi_3 \pi_1 \pi_2 \mathcal{A} \|^2 \right] \\
= & \quad \min_{\pi_1} \left[ \| \pi_1 \mathcal{A} \|^2 + \min_{\pi_2} \left[ \| \pi_2 \pi_1 \mathcal{A} \|^2 + \min_{\pi_3} \| \pi_3 \pi_1 \pi_2 \mathcal{A} \|^2 \right] \right]
\end{align*}
\]
The sequentially truncated HOSVD computes solution to

\[
\begin{align*}
\pi_1^* &= \arg\min_{\pi_1} \| \pi_1^\perp A \|^2 \\
\pi_2^* &= \arg\min_{\pi_2} \| \pi_2^\perp \pi_1^* A \|^2 \\
\pi_3^* &= \arg\min_{\pi_3} \| \pi_3^\perp \pi_1^* \pi_2^* A \|^2
\end{align*}
\]

\( \pi_k^* \) given by \( r_k \) first singular vectors of \([\pi_1^* \cdots \pi_{k-1}^* A](k)\)!
Algorithm

Rank \((r_1, r_2, r_3)\) ST-HOSVD:

\[
\hat{S} = A
\]

\textbf{for} every mode \(k\) \textbf{do}

Compute rank \(r_k\) truncated SVD:

\[
\hat{S}_{(k)} = [\hat{U}_k \quad \hat{U}_k^\perp] [\hat{S}_k \quad \hat{S}_k^\perp] [\hat{V}_k \quad \hat{V}_k^\perp]^T
\]

Project:

\[
\hat{S}_{(k)} = \hat{U}_k^T \hat{S}_{(k)}
\]

\textbf{end for}

\[
\hat{S} = A
\]

\[
\hat{S}_{(1)} = \hat{U}_1^T \hat{S}_{(1)}
\]

\[
\hat{S}_{(2)} = \hat{U}_2^T \hat{S}_{(2)}
\]

\[
\hat{S}_{(3)} = \hat{U}_3^T \hat{S}_{(3)}
\]
Property 1: Equivalence

**Theorem**

Let \( A \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) be of rank \((r_1, r_2, r_3)\). Let

\[
A = (U_1 U_1^T, U_2 U_2^T, U_3 U_3^T) \cdot A, \quad \text{and}
\]

\[
A = (U_1 U_1^T, U_2 U_2^T, U_3 U_3^T) \cdot A
\]

be the rank-\((r_1, r_2, r_3)\) T-HOSVD and ST-HOSVD, respectively. Then

\[
U_1 = U_1 \quad \quad U_2 = U_2 \quad \quad U_3 = U_3
\]

(Holds for any order)
Property 2: Error bounds

**ST-HOSVD** error bound:

\[
\| A - \pi_1 \pi_2 \pi_3 A \|^2 = \| \pi_{2} A \|^2 + \| \pi_{1} \pi_{2} A \|^2 + \| \pi_{3} \pi_{1} \pi_{2} A \|^2
\]

**T-HOSVD** error bound:

\[
\| A - \pi_1 \pi_2 \pi_3 A \|^2 \leq \| \pi_{2} A \|^2 + \| \pi_{1} A \|^2 + \| \pi_{3} A \|^2
\]

Both are quasi-optimal:

\[
\| A - \pi_1 \pi_2 \pi_3 A \|_F \leq \sqrt{d} \| A - A^{\text{opt}} \|_F.
\]
Overview

1. Introduction

2. Orthogonal Tucker approximation

3. Truncated HOSVD

4. Sequentially truncated HOSVD
   - Definition
   - Properties

5. Numerical examples
   - Handwritten digit classification
   - Compression of simulation results
   - Rank-reduction reconstruction

6. Conclusions
Introduction

Orthogonal Tucker

T-HOSVD

ST-HOSVD

Numerical examples

Conclusions

Handwritten digit classification I

Classification of handwritten digits by T-HOSVD (Savas and Eldén, 2007).

0123456789

Tensor of size $786 \times 5421 \times 10$ (Texel $\times$ Example $\times$ Digit).

- Unstructured, 42.6 million non-zeros.
- T-HOSVD and ST-HOSVD truncated to relative error of 10%.

<table>
<thead>
<tr>
<th></th>
<th>T-HOSVD</th>
<th>ST-HOSVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rel. model error</td>
<td>9.90%</td>
<td>9.68%</td>
</tr>
<tr>
<td>Model rank</td>
<td>(94, 511, 10)</td>
<td>(94, 511, 10)</td>
</tr>
</tbody>
</table>
Handwritten digit classification II

<table>
<thead>
<tr>
<th>Method</th>
<th>Classification error</th>
<th>Factorization time</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-HOSVD</td>
<td>4.94%</td>
<td>49m 26.0s</td>
</tr>
<tr>
<td>ST-HOSVD</td>
<td>4.94%</td>
<td>1m 8.7s</td>
</tr>
</tbody>
</table>

43x speedup!
Handwritten digit classification III — Why?

Recall truncation rank? (94, 511, 10).

**T-HOSVD** requires:
- SVD of $786 \times 54210$ matrix,
- SVD of $5421 \times 7860$ matrix, and
- SVD of $10 \times 4260906$ matrix.

**ST-HOSVD** (only) requires:
1. SVD of $786 \times 54210$ matrix,
2. SVD of $5421 \times 940$ matrix, and
3. SVD of $10 \times 48034$ matrix.
Compression of simulation results I

Compression of a numerical solution of a heat equation on a square domain with explicit Euler. Inspired by Lorente et al., 2011.

Tensor of size $101 \times 101 \times 10001$ ($x \times y \times t$).

- Partially symmetric, 102.0 million non-zeros.
- T-HOSVD and ST-HOSVD truncated to absolute error of $10^{-4}$ (discretization accuracy).
## Compression of simulation results II

<table>
<thead>
<tr>
<th></th>
<th>T-HOSVD</th>
<th>ST-HOSVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abs. error</td>
<td>$8.512 \cdot 10^{-5}$</td>
<td>$9.587 \cdot 10^{-5}$</td>
</tr>
<tr>
<td>Rank</td>
<td>(22, 22, 20)</td>
<td>(22, 21, 19)</td>
</tr>
</tbody>
</table>

### Factorization time (Compact SVD)

<table>
<thead>
<tr>
<th></th>
<th>T-HOSVD</th>
<th>ST-HOSVD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2h 46m</td>
<td>1m 14.7s</td>
</tr>
</tbody>
</table>

133x speedup!

### Factorization time (EIGS)

<table>
<thead>
<tr>
<th></th>
<th>T-HOSVD</th>
<th>ST-HOSVD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1m 50.8s</td>
<td>5.4s</td>
</tr>
</tbody>
</table>

20x speedup!
Compression of simulation results III

1st basis vector (T-HOSVD)  
2nd basis vector (T-HOSVD)  
3rd basis vector (T-HOSVD)  
1st basis vector (ST-HOSVD)  
2nd basis vector (ST-HOSVD)  
3rd basis vector (ST-HOSVD)
Tensor unfolding principles and applications to rank-reduction reconstruction and denoising

Nadia Kreimer* and Mauricio Sacchi,
Seismic Analysis and Imaging Group, University of Alberta
GeoConvention 2012: Vision
May 2012
Rank-reduction reconstruction II - (Kreimer and Sacchi)

Seismic acquisition rarely yields fully sampled wave fields, but tensor techniques allow reconstruction from incomplete noisy data.

5D reconstruction methods yield 5th order tensor (1 frequency, 4 spatial modes). Rank reduction and reconstruction is applied to 4th order spatial tensor for every frequency as follows:

\[ D^{(k+1)} \leftarrow \alpha D_{\text{obs}} + (1 - \alpha) \mathcal{T} \tilde{D}^{(k)} + (1 - \mathcal{T}) \tilde{D}^{(k)} \]

where

- \( D_{\text{obs}} \) are original incomplete data.
- \( \tilde{D}^{(k)} \) is ST-HOSVD approximation.
- \( \mathcal{T} \) is binary “selection” tensor: \( \mathcal{T} D_{\text{obs}} = D_{\text{obs}} \) and \( (1 - \mathcal{T}) D_{\text{obs}} = 0 \).
- \( \alpha \in [0, 1] \) is a relaxation parameter.
Synthetic $12 \times 12 \times 12 \times 12$ tensor.
Overview

1. Introduction
2. Orthogonal Tucker approximation
3. Truncated HOSVD
4. Sequentially truncated HOSVD
   - Definition
   - Properties
5. Numerical examples
   - Handwritten digit classification
   - Compression of simulation results
   - Rank-reduction reconstruction
6. Conclusions
Early projection can greatly improve the performance of T-HOSVD.
Thank you for your attention.
References

References

- N. Kreimer, personal communication.