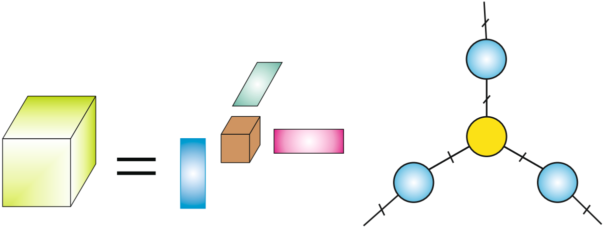
Tucker Decomposition

February 15, 2022

Part I

Tucker Decomposition - TKD

Tucker model I



- Tucker decomposition

$$\mathcal{Y} = \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = [\![\mathcal{G}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

  where $\mathcal{G}$ is a core tensor of size $R_1 \times R_2 \times R_3$, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are of size $I \times R_1$, $J \times R_2$ and $K \times R_3$, respectively.
  $R_1 \leq I$, $R_2 \leq J$ and $R_3 \leq K$.
- **Multilinear rank**-R_1, R_2, R_3

Tucker model II

- Unfolding and vectorization

$$
\begin{aligned}
\mathbf{Y}_{(1)} &= \mathbf{A}\,\mathbf{G}_{(1)}(\mathbf{C}\otimes\mathbf{B})^T \\
\mathbf{Y}_{(2)} &= \mathbf{B}\,\mathbf{G}_{(2)}(\mathbf{C}\otimes\mathbf{A})^T \\
\mathbf{Y}_{(3)} &= \mathbf{C}\,\mathbf{G}_{(3)}(\mathbf{B}\otimes\mathbf{A})^T \\
\mathrm{vec}(\mathcal{Y}) &= (\mathbf{C}\otimes\mathbf{B}\otimes\mathbf{A})\,\mathrm{vec}(\mathcal{G})
\end{aligned}
$$

- Multiplication and norm

$$
\begin{aligned}
\mathcal{Y}\times_1\mathbf{U} &= [\![\mathcal{G};\mathbf{UA},\mathbf{B},\mathbf{C}]\!] \\
\|\mathcal{Y}\|_F^2 &= \mathrm{vec}(\mathcal{G})^T\,(\mathbf{C}^T\mathbf{C}\otimes\mathbf{B}^T\mathbf{B}\otimes\mathbf{A}^T\mathbf{A})\,\mathrm{vec}(\mathcal{G}) \qquad (1)
\end{aligned}
$$

When **A**, **B** and **C** are orthogonal matrices,

$$
\|\mathcal{Y}\|_F^2 = \|\mathcal{G}\|_F^2
$$

Tucker model III

- **Uniqueness** subspaces of the factor matrices but not them since

$$
\begin{aligned}
[\![\mathcal{G}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] &= [\![\mathcal{G}; \mathbf{A}\mathbf{Q}\mathbf{Q}^T, \mathbf{B}, \mathbf{C}]\!] \\
&= [\![\mathcal{G} \times_1 \mathbf{Q}^T; \mathbf{A}\mathbf{Q}, \mathbf{B}, \mathbf{C}]\!]
\end{aligned}
$$

  where $\mathbf{Q}$ is an arbitrary orthonormal matrix
- **Orthogonality constraints.** Because of rotational ambiguity, without loss of generality, the matrices **A**, **B** and **C** can be assumed to have orthonormal columns, that is

$$
\mathbf{A}^T\mathbf{A} = \mathbf{I}_{R_1}, \quad \mathbf{B}^T\mathbf{B} = \mathbf{I}_{R_2}, \quad \mathbf{C}^T\mathbf{C} = \mathbf{I}_{R_3}
$$

  .

Higher Order SVD - HOSVD I

- Tucker model with $R_1 = I$, $R_2 = J$ and $R_3 = K$
- HOSVD: SVD in each mode of the tensor
- Compute SVD of mode-n unfoldings of $\mathcal{Y}$

$$\mathbf{Y}_{(n)} = \mathbf{U}_n \Sigma_n \mathbf{V}_n^T \tag{2}$$

where $\mathbf{U}_n$ are orthonormal matrices.

- Then, the core tensor is given by

$$\mathcal{G} = \mathcal{Y} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \cdots \times_N \mathbf{U}_N^T$$

- Slices of $\mathcal{G}$ are orthogonal

Proof.

$$\begin{aligned} \mathbf{G}_{(n)}\mathbf{G}_{(n)}^T &= \\ &= \\ &= \Sigma_n \end{aligned}$$

Higher Order SVD - HOSVD II

- **Truncated HOSVD.** De Lathauwer et al. (2000)
  Find a truncated model which holds

$$\|\mathcal{Y} - \hat{\mathcal{Y}}\|_F^2 \leq \varepsilon^2 \tag{3}$$

R_1, R_2 and R_3 are determined such that

$$\begin{aligned} \min \quad & I_1 R_1 + I_2 R_2 + I_3 R_3 + R1 R_2 R_3 \qquad (4) \\ \text{s.t} \quad & \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3}^{R_3} g_{r_1,r_2,r_3}^2 \leq \|\mathcal{Y}\|_F^2 - \varepsilon^2 \,, \end{aligned}$$

This initialization method usually provides models with lower approximation error and smaller number of parameters than the sequential projectin based method.

Sequential Projection for Truncated HOSVD

- let $\mathcal{T}_0 = \mathcal{Y}$. Compute SVD of mode-1 unfolding of $\mathcal{T}_0$

$$\mathbf{Y}_{(1)} \approx \mathbf{U}_1 \mathbf{S}_1 \mathbf{V}_1^T \tag{5}$$

- Define $\mathcal{T}_1 = \mathcal{Y} \times_1 \mathbf{U}_1^T$, i.e., mode-1 unfolding of $\mathcal{T}_1$ is $\mathbf{S}_1 \mathbf{V}_1^T$. Then compute SVD of mode-2 unfolding of $\mathcal{T}_1$

$$[\mathbf{T}_1]_{(2)} \approx \mathbf{U}_2 \mathbf{S}_2 \mathbf{V}_2^T \tag{6}$$

- Define $\mathcal{T}_2 = \mathcal{T}_1 \times_2 \mathbf{U}_2^T$, and compute SVD of $\mathcal{T}_2$ to give

$$[\mathbf{T}_2]_{(3)} \approx \mathbf{U}_3 \mathbf{S}_3 \mathbf{V}_4^T \tag{7}$$

- The core tensor finally $\mathcal{G} = \mathcal{T}_2 \times_3 \mathbf{U}_3^T$.

TKD

We consider the two types of approximation of a tensor $\mathcal{Y}$ by a tensor $\hat{\mathcal{Y}}$ in the Tucker format

- **TKD with given -rank** based on a minimisation in the form

$$\min \quad D = \|\mathcal{Y} - \hat{\mathcal{Y}}\|_F^2 \ . \tag{8}$$

- **TKD with given approximation accuracy** which is typically used in the presence of noise or when the multilinear rank is not specified

$$\min \quad \sum_{n=1}^{N-1} R_n \quad \text{s.t.} \ \|\mathcal{Y} - \hat{\mathcal{Y}}\|_F^2 \leq \varepsilon^2 \,, \tag{9}$$

such that the estimated TKD-tensor $\hat{\mathcal{Y}}$ should have minimum total multilinear-rank or minimum number of parameters

$$\min \quad \sum_{n=1}^{N} I_n R_{n-1} R_n \quad \text{s.t.} \ \|\mathcal{Y} - \hat{\mathcal{Y}}\|_F^2 \leq \varepsilon^2 \,, \tag{10}$$

where ε^2 represents approximation accuracy.

ALS for TKD with given rank

When ranks are given, the decomposition minimizes Frobenius norm of the error

$$\min \ D_F(\mathcal{Y} \,||\hat{\mathcal{Y}}) = ||\mathcal{Y} - \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}||_F^2 \tag{11}$$

Since the factor matrices $\mathbf{A}^{(n)}$ are orthogonal

$$||\hat{\mathcal{Y}}||_F^2 = ||\mathcal{G}||_F^2$$

$$\begin{aligned} D_F(\mathcal{Y} \,||\hat{\mathcal{Y}}) &= ||\mathcal{Y}||_F^2 + ||\mathcal{G}||_F^2 - 2\,\langle \mathcal{Y}, \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \rangle \\ &= ||\mathcal{Y}||_F^2 + ||\mathcal{G}||_F^2 - 2\,\langle \mathcal{Y} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T, \mathcal{G} \rangle \end{aligned}$$

where $\langle \mathcal{Y}, \hat{\mathcal{Y}} \rangle$ denotes the inner product.

ALS for TKD with given rank

Optimal core tensor.

$\min \; \|\mathcal{Y}\|_F^2 + \|\mathcal{G} - \mathcal{Y} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T\|_F^2 - \|\mathcal{Y} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T\|_F^2$

give $\mathcal{G}^\star = \mathcal{Y} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T$

Optimization for update A. The optimization problem becomes

$$\max \quad \|\mathcal{Y} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T\|_F^2 \tag{11}$$

Let $\mathcal{Z} = \mathcal{Y} \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T$, then **A** is found in the EVD problem

$$\max \quad \operatorname{tr}(\mathbf{A}^T(\mathbf{Z}_{(1)}\mathbf{Z}_{(1)}^T)\mathbf{A}) \tag{12}$$

i.e., **A** comprises leading singular vectors of $\mathcal{Y} \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T$

ALS for TKD with given rank

Similarly, $\mathbf{B}$ comprises leading singular vectors of $\mathcal{Y} \times_1 \mathbf{A}^T \times_3 \mathbf{C}^T$
and
$\mathbf{C}$ comprises leading singular vectors of $\mathcal{Y} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T$

Approximation Error.

$$D(\mathcal{Y}|\hat{\mathcal{Y}}) = \|\mathcal{Y}\|_F^2 - \|\mathcal{G}\|_F^2 \tag{11}$$

HOOI (ALS for TKD

Algorithm 1: Higher Order Orthogonal Iteration)

Input: $\mathcal{Y}$: input data of size $I_1 \times I_2 \times I_3$, R_1, R_2, R_3 are multilinear rank

Output: $\mathcal{Y} \approx [\![\mathcal{G}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$

begin
  Initialize **B** and **C** using truncatedHOSVD or Seq-projection;
  repeat
    $\mathcal{Z} = \mathcal{Y} \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T$;
    Update **A** as R_1 left leading singular vectors of $\mathbf{Z}_{(1)}$;
    $\mathcal{Z} = \mathcal{Y} \times_1 \mathbf{A}^T \times_3 \mathbf{C}^T$;
    Update **B** as R_2 left leading singular vectors of $\mathbf{Z}_{(2)}$;
    $\mathcal{Z} = \mathcal{Y} \times_1 \mathbf{B}^T \times_2 \mathbf{A}^T$;
    Update **C** as R_3 left leading singular vectors of $\mathbf{Z}_{(3)}$;
  until *a stopping criterion is met*;
  $\mathcal{G} = \mathcal{Z} \times_3 \mathbf{C}^T$;
end

HOOI for Best Rank-1 Tensor Approximation I

Algorithm 2: HOOI = ALS for CPD

Input: $\mathcal{Y}$: input data of size $I_1 \times I_2 \times I_3$

Output: $\mathcal{Y} \approx [\![\lambda; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}]\!]$

```
begin
    Initialize b and c using truncated HOSVD or
      sequential-projection;
    repeat
        a ← 𝒴 ×_2 b^T ×_3 c^T;
        a ← a / ||a||;
        b ← 𝒴 ×_1 a^T ×_3 c^T;
        b ← b / ||b||;
        c ← 𝒴 ×_1 a^T ×_2 b^T;
        λ = ||c||, c ← c / λ;
    until a stopping criterion is met;
end
```

where the updates are:

$\boldsymbol{a} \leftarrow \mathcal{Y} \times_2 \boldsymbol{b}^T \times_3 \boldsymbol{c}^T$;

$\boldsymbol{a} \leftarrow \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}$;

$\boldsymbol{b} \leftarrow \mathcal{Y} \times_1 \boldsymbol{a}^T \times_3 \boldsymbol{c}^T$;

$\boldsymbol{b} \leftarrow \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|}$;

$\boldsymbol{c} \leftarrow \mathcal{Y} \times_1 \boldsymbol{a}^T \times_2 \boldsymbol{b}^T$;

$\lambda = \|\boldsymbol{c}\|$, $\boldsymbol{c} \leftarrow \frac{\boldsymbol{c}}{\lambda}$;

ALS for TKD with Error bound constraint I

- When the approximation error is bounded, we seek a model with a minimum number of parameters

$$\begin{aligned} \min \quad & I_1 R_1 + I_2 R_2 + I_3 R_3 + R_1 R_2 R_3 \\ \text{s.t.} \quad & \|\mathcal{Y} - \hat{\mathcal{Y}}\|_F^2 \leq \varepsilon^2, \;\; \mathbf{A}^T \mathbf{A} = \mathbf{I}_{R_1}, \; \mathbf{B}^T \mathbf{B} = \mathbf{I}_{R_2}, \; \mathbf{C}^T \mathbf{C} = \mathbf{I}_{R_3} \,. \end{aligned} \tag{12}$$

- We can show that the factor matrices **A**, **B** and **C** can be updated in a similar fashion to the case when ranks are given.

Lemma

In the optimisation problem (12) the core tensor $\mathcal{G}$ can be eliminated using $\mathcal{G}^\star = \mathcal{Y} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T$ without changing the minimal achievable number of parameters.

Proof.

□

Denote $\mathcal{G}^\star = \mathcal{Y} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T$.

ALS for TKD with Error bound constraint II

The error bound condition

$$\begin{aligned} \varepsilon^2 &\geq \|\mathcal{Y} - \hat{\mathcal{Y}}\|_F^2 \\ &= \|\mathcal{Y}\|_F^2 - \|\mathcal{G}^\star\|_F^2 + \|\mathcal{G} - \mathcal{G}^\star\|_F^2 \end{aligned}$$

indicates that we can represent $\mathcal{G} = \mathcal{G}^\star + \mathcal{E}$, where $\mathcal{E}$ is a tensor such that its norm

$$\gamma^2 = \|\mathcal{E}\|_F^2 \leq \varepsilon^2 + \|\mathcal{G}^\star\|_F^2 - \|\mathcal{Y}\|_F^2$$

Assume that $\mathbf{B}^\star$ and $\mathbf{C}^\star$ are the optimal factor matrices with minimal values $R_2^\star$ and $R_3^\star$. From the optimization problem for TKD

$$\begin{aligned} \min \quad & I_1 R_1 + I_2 R_2 + I_3 R_3 + R_1 R_2 R_3 \\ \text{s.t.} \quad & \|\mathcal{Y} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T\|_F^2 \geq \|\mathcal{Y}\|_F^2 + \gamma^2 - \varepsilon^2 \,, \quad \mathbf{A}^T\mathbf{A} = \mathbf{I}_{R_1} , \end{aligned}$$

ALS for TKD with Error bound constraint III

Optimization problem to update $\mathbf{A}$

$$\min \quad R_1 \quad \text{s.t.} \quad \operatorname{tr}(\mathbf{A}^T \mathbf{Q}_1 \mathbf{A}) \geq \|\mathcal{Y}\|_F^2 + \gamma^2 - \varepsilon^2\,, \quad \mathbf{A}^T \mathbf{A} = \mathbf{I}_{R_1}\,,$$

where $\mathbf{Z}_1 = \mathcal{Y} \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T$, $\mathbf{Q}_1 = \mathbf{Z}_1 \mathbf{Z}_1^T$.
The optimal matrix $\mathbf{A}^\star$ comprises R_1 principal eigenvectors of $\mathbf{Q}_1$, where R_1 is the smallest number of eigenvalues, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{R_1}$ such that their norm exceeds the bound $\delta^2 = \|\mathcal{Y}\|_F^2 - \varepsilon^2 + \gamma^2$, that is

$$\sum_{r=1}^{R_1} \lambda_r \geq \delta^2 > \sum_{r=1}^{R_1-1} \lambda_r\,. \tag{13}$$

It is obvious that a minimal number of columns R_1 is achieved, when the bound $\|\mathcal{Y}\|_F^2 + \gamma^2 - \varepsilon^2$ is smallest, i.e., $\gamma = 0$. Implying that the optimal $\mathcal{G}$ in the problem (12) is also $\mathcal{G}^\star$.

ALS for TKD with Error bound constraint IV

Update A

$$
\begin{aligned}
\min \quad & \operatorname{rank}(\mathbf{A}) \\
\text{s.t.} \quad & \operatorname{tr}(\mathbf{A}^T \mathbf{Q}_1 \mathbf{A}) \geq \|\mathcal{Y}\|_F^2 - \varepsilon^2 \\
& \mathbf{A}^T \mathbf{A} = \mathbf{I}.
\end{aligned}
$$

- The matrices **B** and **C** are updated similarly. The algorithm sequentially updates **A**, **B** and **C**.

TKD with Error bound constraint I

Example

TK2-decomposition of random tensors of size $100 \times 10 \times 100$ with relative error bound of $\varepsilon = 0.8$

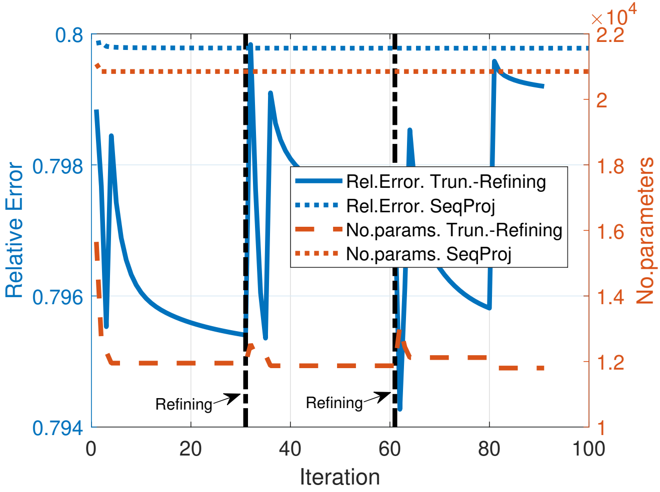
$$\|\mathcal{Y} - \hat{\mathcal{Y}}\|_F \leq \varepsilon \|\mathcal{Y}\|_F$$

- The SeqProj-like initialization yields a rank-$(13, 92)$ model which contains 22460 parameters within 2 iterations and remains badly balanced after 200 iterations.
- Using the truncated HOSVD, we obtain an initial rank-$(35, 27)$ model having only 15650 parameters and a lower approximation error.

TKD with Error bound constraint II

- HOOI updates and yields the model with 12650, 12300 and 11950 parameters in the first three iterations, and remains unchaged and well-balanced, while the approximation error converges after 30 iterations.

TKD with Error bound constraint III

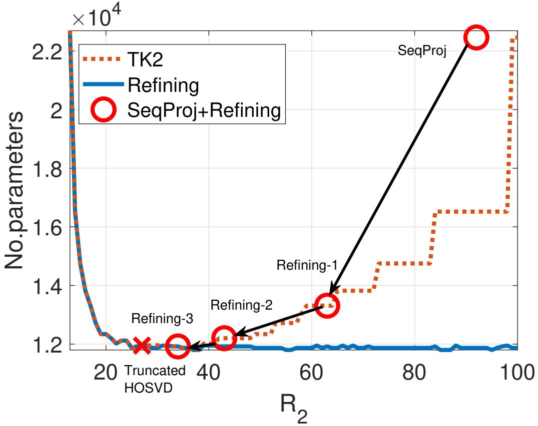


TKD with Error bound constraint IV

Model Refining.

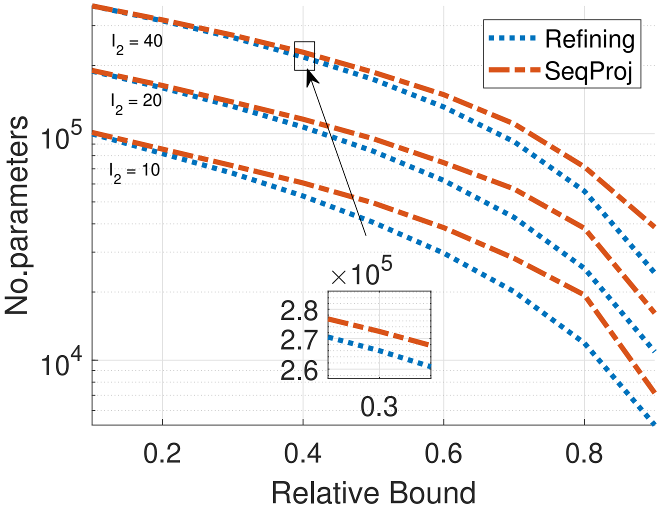
- Truncated HOSVD initialization demonstrates efficient to TKD with bound constraint.
- In practice, the estimated model can often be squeezed to a smaller one. For a feasible solution, our approach is to generate a truncated HOSVD model with the same rank, then use it to initialize another TKD decomposition.
- We continue the model refining to the estimated model, if it is more compact. Otherwise we terminate the procedure.

TKD with Error bound constraint V



- When the rank is relatively small, $R_2 < 15$, there is no feasible solution. When $R_2 \geq 37$, refining always yields estimated model close to the optimal.
- For the model having 22460 parameters initialized by SeqProj, we execute the refining three times, and obtain a final model with 11920 parameters.

TKD with Error bound constraint VI



TKD as Prior Compression for CPD

For CPD with the rank R smaller than the tensor dimensions, the tensor can be first compressed using TKD

$$\mathcal{Y} \approx [\![\mathcal{G}; \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3]\!]$$

then decompose the core tensor $\mathcal{G}$ by a rank-R CPD

$$\mathcal{G} \approx [\![\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3]\!]$$

CPD of $\mathcal{Y}$ is finally given by

$$\mathcal{Y} \approx [\![\mathbf{U}_1\mathbf{V}_1, \mathbf{U}_2\mathbf{V}_2, \mathbf{U}_3\mathbf{V}_3]\!]$$

Tucker-CPD I

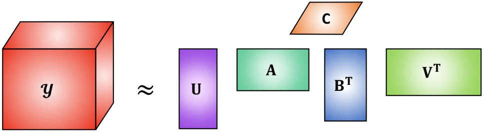


Figure: Illustration of the CPD with low-rank constraints.

- We consider the CPD with high rank, in which factor matrices are modelled in a low-rank format

$$\mathcal{Y} \approx [\![\mathbf{UA}, \mathbf{VB}, \mathbf{WC}]\!] \tag{14}$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{I}_{R_1}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}_{R_2}$ and $\mathbf{W}^T\mathbf{W} = \mathbf{I}_{R_3}$.

Tucker-CPD II

Relation with other models

- When $R \leq I$, we can apply the two-stages TKD+CPD, which first performs the TKD then decomposes the core tensor Bro and Andersson (1998).
- When $R \gg I$, the subspaces sought by TKD may not be optimal, and CPD of the compressed core tensor does not give the best result of the model in (14).
- When $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ are design matrices, i.e., fixed and known in advance, LrCPD becomes the CANDELINCCarroll et al. (1970); Bro and Andersson (1998).
- When $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are dependence matrices consisting of zeros and ones, we obtain the parallel profiles with linear dependencies (PARALIND) Bro et al. (2009) or CONFAC **?**.

LrCPD - Factor Matrices with Given Ranks I

- Assume that the ranks of $\mathbf{A}$ and $\mathbf{B}$ are given, and the model need not factorize $\mathbf{C}$.

$$\begin{aligned} \min_{\mathbf{U},\mathbf{A},\mathbf{V},\mathbf{B},\mathbf{C}} \quad & f = \|\mathcal{Y} - [\![\mathbf{UA}, \mathbf{VB}, \mathbf{C}]\!]\|_F^2 \\ \text{s.t.} \quad & \mathbf{U}^T\mathbf{U} = \mathbf{I}_{R_1}, \ \mathbf{V}^T\mathbf{V} = \mathbf{I}_{R_2} \end{aligned} \tag{15}$$

  where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices of size $I \times R_1$ and $J \times R_2$, respectively, $R_1 < I_1$, $R_2 < I_2$ and $R > \max(R_1, R_2)$. The matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ comprise R columns.
- **Update rules for A.**
  - Rewrite the objective function in form of mode-1 unfolding of $\mathcal{Y}$, that is

$$\begin{aligned} f &= \|\mathcal{Y} - [\![\mathbf{UA}, \mathbf{VB}, \mathbf{C}]\!]\|_F^2 = \|\mathbf{Y}_{(1)} - \mathbf{UA}(\mathbf{C} \odot \mathbf{VB})^T)\|_F^2 \\ &= \|\mathcal{Y}\|_F^2 + \|\mathbf{A}(\mathbf{C} \odot \mathbf{VB})^T)\|_F^2 - 2\,\mathrm{tr}(\mathbf{A}(\mathbf{C} \odot \mathbf{VB})^T(\mathbf{U}^T\mathbf{Y}_{(1)})^T) \\ &= \|\mathcal{Y}\|_F^2 - \|\mathbf{U}^T\mathbf{Y}_{(1)}\|_F^2 + \|\mathbf{U}^T\mathbf{Y}_{(1)} - \mathbf{A}(\mathbf{C} \odot \mathbf{VB})^T)\|_F^2, \end{aligned}$$

LrCPD - Factor Matrices with Given Ranks II

- Let $\mathcal{G} = \mathcal{Y} \times_1 \mathbf{U}^T \times_2 \mathbf{V}^T$.
  We keep the other factor matrices fixed, and exploit the identity

$$\mathbf{C} \odot \mathbf{V}\mathbf{B} = (\mathbf{I} \otimes \mathbf{V})(\mathbf{C} \odot \mathbf{B})$$

  $f()$ achieves its minimum at the optimal $\mathbf{A}^\star$ given by

$$\begin{aligned} \mathbf{A}^\star &= \mathbf{U}^T \mathbf{Y}_{(1)} (\mathbf{C} \odot \mathbf{V}\mathbf{B})((\mathbf{C}^T\mathbf{C}) \circledast (\mathbf{B}^T\mathbf{B}))^{-1} \\ &= \mathbf{G}_{(1)} (\mathbf{C} \odot \mathbf{B})((\mathbf{C}^T\mathbf{C}) \circledast (\mathbf{B}^T\mathbf{B}))^{-1} \,, \end{aligned} \tag{16}$$

  provided that $((\mathbf{C}^T\mathbf{C}) \circledast (\mathbf{B}^T\mathbf{B}))$ is invertible.

LrCPD - Factor Matrices with Given Ranks III

- **Update U.** Substitute $\mathbf{A}^\star$ into the objective function $f()$, we obtain

$$
\begin{aligned}
f(\mathbf{A}^\star) &= \|\mathcal{Y}\|_F^2 - \|\mathbf{U}^T\mathbf{Y}_{(1)}\|_F^2 + \|\mathbf{U}^T\mathbf{Y}_{(1)} - \mathbf{A}^\star(\mathbf{C}\odot\mathbf{V}\mathbf{B})^T)\|_F^2 \\
&= \|\mathcal{Y}\|_F^2 + \operatorname{tr}(\mathbf{A}^\star(\mathbf{C}\odot\mathbf{V}\mathbf{B})^T(\mathbf{C}\odot\mathbf{V}\mathbf{B})\mathbf{A}^{\star T}) \\
&\quad -2\operatorname{tr}(\mathbf{A}^\star(\mathbf{C}\odot\mathbf{V}\mathbf{B})^T\mathbf{Y}_{(1)}^T\mathbf{U}) \\
&= \|\mathcal{Y}\|_F^2 - \operatorname{tr}(\mathbf{U}^T\,\mathbf{Q}\,\mathbf{U})
\end{aligned}
$$

  where $\mathbf{Q} = \mathbf{Y}_{(1)}(\mathbf{C}\odot\mathbf{V}\mathbf{B})^T((\mathbf{C}^T\mathbf{C})\circledast(\mathbf{B}^T\mathbf{B}))^{-1}(\mathbf{C}\odot\mathbf{V}\mathbf{B})\mathbf{Y}_{(1)}^T$.
- The optimal $\mathbf{U}$, which minimizes f, comprises R_1 principle eigenvectors of $\mathbf{Q}$

$$
\max \quad \operatorname{tr}(\mathbf{U}^T\mathbf{Q}\mathbf{U}) \quad \text{s.t.} \quad \mathbf{U}^T\mathbf{U} = \mathbf{I}_{R_1} \tag{17}
$$

- The factor matrices, $\mathbf{V}$ and $\mathbf{B}$, can be updated similarly.

LrCPD - Factor Matrices with Given Ranks IV

- **ALS update rule for C**

$$\begin{aligned}\mathbf{C} &= \mathbf{Y}_{(3)}(\mathbf{VB}\odot\mathbf{UA})(((\mathbf{UA})^T(\mathbf{UA})\circledast((\mathbf{VB})^T(\mathbf{VB})))^{-1} \\ &= \mathbf{G}_{(3)}(\mathbf{B}\odot\mathbf{A})((\mathbf{A}^T\mathbf{A})\circledast(\mathbf{B}^T\mathbf{B}))^{-1}. \end{aligned} \tag{18}$$

- In summary, the update rules for **A**, **B**, and **C** are similar to the ALS updates for CPD of $\mathcal{G}$ of size $R_1 \times R_2 \times I_3$, i.e., have lower complexity than those in the ordinary CPD. However, the updates for **U** and **V** demand an extra cost for the EVD of **Q**.

The Case with Optimal Ranks for Factor Matrices I

- We seek a model with the smallest number of parameters

$$\begin{aligned} \min \quad & (I+R)R_1 + (J+R_2)R_2 + KR \\ \text{s.t.} \quad & \|\mathcal{Y} - [\![\mathbf{UA}, \mathbf{VB}, \mathbf{C}]\!]\|_F^2 \leq \varepsilon \|\mathcal{Y}\|_F^2 \qquad (0 \leq \varepsilon < 1) \end{aligned}$$

- Define $\mathbf{F} = \mathbf{UA}$. Then $\text{rank}(\mathbf{F}) = R_1$ and $\mathbf{F}$ is solution of a rank minimization problem

$$\min \quad \text{rank}(\mathbf{F}) \quad \text{s.t.} \ \|\mathbf{Y}_{(1)} - \mathbf{F}(\mathbf{C} \odot \mathbf{VB})^T\|_F^2 \leq \varepsilon \|\mathcal{Y}\|_F^2.$$

- Assume that $\mathbf{C} \odot \mathbf{VB}$ is of full column rank. Denote SVD of $\mathbf{C} \odot \mathbf{VB} = \mathbf{ZDK}^T$, where $\mathbf{Z}$ is of size $I_2 I_3 \times R$, $\mathbf{K}$ of size $R \times R$. SVD of $\mathbf{C} \odot \mathbf{VB}$ is computed through a smaller matrix $(\mathbf{C} \odot \mathbf{B})$.
- Rewrite the bound constraint as

$$\begin{aligned} \|\mathbf{Y}_{(1)} - \mathbf{F}(\mathbf{C} \odot \mathbf{VB})^T\|_F^2 &= \|\mathcal{Y}\|_F^2 - \|\mathbf{Y}_{(1)}\mathbf{Z}\|_F^2 + \|\mathbf{Y}_{(1)}\mathbf{Z} - \mathbf{FKD}\|_F^2 \\ &\leq \varepsilon \|\mathcal{Y}\|_F^2 \end{aligned}$$

The Case with Optimal Ranks for Factor Matrices II

- Since $\operatorname{rank}(\mathbf{F}) = \operatorname{rank}(\mathbf{FKD})$, we solve an equivalent rank-minimization problem for $\tilde{\mathbf{F}} = \mathbf{FKD}$

$$\begin{aligned} \min \quad & \operatorname{rank}(\tilde{\mathbf{F}}) \\ \text{s.t.} \quad & \|\mathbf{Y}_{(1)}\mathbf{Z} - \tilde{\mathbf{F}}\|_F^2 \le \|\mathbf{Y}_{(1)}\mathbf{Z}\|_F^2 - (1-\varepsilon)\|\mathcal{Y}\|_F^2 . \end{aligned} \tag{19}$$

- **Optimal F** is achieved through the truncated SVD of the matrix, $\mathbf{Y}_{(1)}\mathbf{Z}$

$$\tilde{\mathbf{F}}^{\star} = \mathbf{US}\tilde{\mathbf{V}}^T \approx \mathbf{Y}_{(1)}\mathbf{Z} \tag{20}$$

which obeys the bound

$$\|\mathbf{Y}_{(1)}\mathbf{Z} - \tilde{\mathbf{F}}\|_F^2 \le \|\mathbf{Y}_1\mathbf{Z}\|_F^2 - (1-\varepsilon)\|\mathcal{Y}\|_F^2$$

$\mathbf{S} = \operatorname{diag}(\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_{R_1})$ and the rank R_1 is the smallest number of singular values such that $\sum_{r=1}^{R_1} \sigma_r^2 \ge (1-\varepsilon)\|\mathcal{Y}\|_F^2 > \sum_{r=1}^{R_1-1} \sigma_r^2$.

The Case with Optimal Ranks for Factor Matrices III

- In other words, $\mathbf{U}$ comprises the left leading singular vectors of $\mathbf{Y}_{(1)}\mathbf{Z}$ and $\mathbf{A} = \mathbf{S}\tilde{\mathbf{V}}^T\mathbf{D}^{-1}\mathbf{K}^T$. Similar update rule can be derived for $\mathbf{V}$ and $\mathbf{B}$.

Nested network of TK2 I

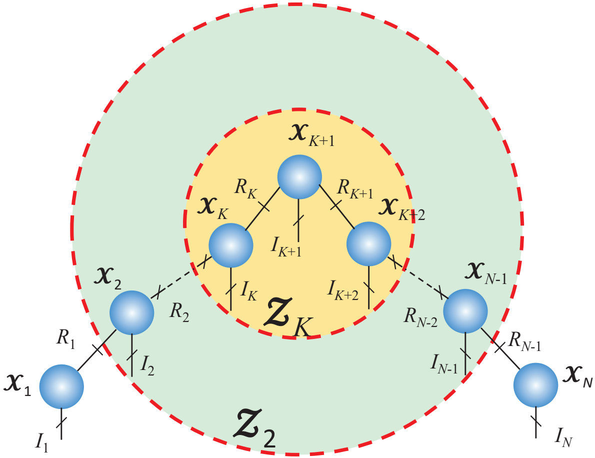


Figure: A nested network of TK2 decompositions forms a TT network.

Nested network of TK2 II

- Given a tensor $\mathcal{Y}$ of order-N and size $I_1 \times I_2 \times \cdots \times I_N$
- TK2 decomposition of $\mathcal{Y}$ which gives two core tensors, $\mathcal{X}_1$ and $\mathcal{X}_N$, for the first and last modes

$$\mathcal{Y} \approx \mathcal{X}_1 \bullet \mathcal{Z}_2 \bullet \mathcal{X}_N \,.$$

where $\mathcal{Z}_2 = \mathcal{X}_1^T \bullet \mathcal{Y} \bullet \mathcal{X}_N^T$ is of size $R_1 \times I_2 \times I_3 \times \cdots \times I_{N-2} \times I_{N-1} \times R_{N-1}$.

- Next, reshape the tensor $\mathcal{Z}_2$ to size $R_1 I_2 \times I_3 \times \cdots \times I_{N-2} \times I_{N-1} R_{N-1}$
- Apply TKD-2 to $\mathcal{Z}_2$

$$\mathcal{Z}_2 \approx \mathcal{X}_2 \bullet \mathcal{Z}_3 \bullet \mathcal{X}_{N-1}$$

to give two core tensors $\mathcal{X}_2$ of size $R_1 I_2 \times R_2$ and $\mathcal{X}_{N-1}$ of size $R_{N-2} \times I_{N-1} R_{N-1}$, and $\mathcal{Z}_3$ is a tensor of size $R_2 \times I_3 \times I_4 \times \cdots \times I_{N-2} \times R_{N-2}$.

Nested network of TK2 III

- Reshape $\mathcal{X}_2$ to tensor of size $R_1 \times I_2 \times R_2$ and $\mathcal{X}_{N-1}$ to tensor of size $R_{N-2} \times I_{N-1} \times R_{N-1}$
  We obtain

$$\mathcal{Y} \approx \mathcal{X}_1 \bullet \mathcal{X}_2 \bullet \mathcal{Z}_3 \bullet \mathcal{X}_{N-1} \bullet \mathcal{X}_N \tag{21}$$

- It can be verified that estimation of the three core tensors $\mathcal{X}_2$, $\mathcal{Z}_3$, $\mathcal{X}_{N-2}$ within the TT-tensor $\mathcal{X}_1 \bullet \mathcal{X}_2 \bullet \mathcal{Z}_3 \bullet \mathcal{X}_{N-1} \bullet \mathcal{X}_N$, while fixing the two orthogonal matrices, $\mathcal{X}_1$ and $\mathcal{X}_N$, becomes the estimation of a TK2 decomposition of $\mathcal{Z}_2$

$$\begin{aligned} &\|\mathcal{Y} - \mathcal{X}_1 \bullet \mathcal{X}_2 \bullet \mathcal{Z}_3 \bullet \mathcal{X}_{N-1} \bullet \mathcal{X}_N\|_F^2 \\ &\quad = \|\mathcal{Y}\|_F^2 - \|\mathcal{Z}_2\|_F^2 + \|\mathcal{Z}_2 - \mathcal{X}_2 \bullet \mathcal{Z}_3 \bullet \mathcal{X}_{N-1}\|_F^2 . \end{aligned}$$

- Similarly, we perform TK2 decomposition of $\mathcal{Z}_3$ to get the core tensors $\mathcal{X}_3$ and $\mathcal{X}_{N-2}$.

Nested network of TK2 IV

- TK2 with bound constraint can be used to construct a TT-model with bounded approximation error.
- In the first layer, $\mathcal{X}_1$ and $\mathcal{X}_N$ are estimated within a smallest TK2 model such that

$$\|\mathcal{Y} - \mathcal{X}_1 \bullet \mathcal{Z}_2 \bullet \mathcal{X}_N\|_F^2 = \|\mathcal{Y}\|_F^2 - \|\mathcal{Z}_2\|_F^2 \leq \varepsilon^2 .$$

  This is achieved when $\|\mathcal{Y}\|_F^2 - \|\mathcal{Z}_2\|_F^2$ is close to or attains the bound $\varepsilon_1^2 = \varepsilon^2$ so that $\mathcal{X}_1$ and $\mathcal{X}_N^T$ have small ranks.
- In the second layer, we solve a TK2 with a much smaller bound

$$\|\mathcal{Z}_2 - \mathcal{X}_2 \bullet \mathcal{Z}_3 \bullet \mathcal{X}_{N-1}\|_F^2 \leq \varepsilon_2^2 = \varepsilon^2 - \|\mathcal{Y}\|_F^2 + \|\mathcal{Z}_2\|_F^2 \ll \varepsilon_1^2.$$

- A similar procedure is applied to the core tensors $\mathcal{Z}_3, \mathcal{Z}_4, \ldots$, but the approximation errors are decreasing significantly.

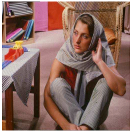
Nested network of TK2 V

- If the bound is attained in the first layer, i.e., $\varepsilon_2^2 = \varepsilon^2 - \|\mathcal{Y}\|_F^2 + \|\mathcal{Z}_2\|_F^2 = 0$, then higher layers solve exact TK2 models.
  Implying that the factor matrices within TK2 will have full rank or very high rank, i.e., $R_3 \approx R_2 I_2$, $R_{N-1} \approx I_{N-1} R_N$, $R_4 \approx R_3 I_3 \approx R_2 I_2 I_3$.
  In this case, dimensions of the core tensors, especially the central cores, grow dramatically, and as a result the final TT-model is not very compact. This behavior is also observed in the TT-SVD.

Nested network of TK2 VI

In order to deal with the large rank issue, we suggest to scale the error bounds in some first layers to smaller than the required bounds, e.g., by a factor of $\exp(-1 + n/\lfloor \frac{N}{2} \rfloor)$, where $n = 1, 2, \ldots, \lfloor \frac{N}{2} \rfloor$ is the layer index, $\lfloor \frac{N}{2} \rfloor$ greatest integer less than or equal to $N/2$.

Image denoising I

- Color images $\mathcal{T}$ degraded by additive Gaussian noise at SNR = 10 dB
- Constructed tensors, $\mathcal{Y}_{r,c}$, of a size $w \times w \times 3 \times (2d+1) \times (2d+1)$

$$\mathcal{Y}_{r,c}(:,:,:,d+1+i,d+1+j) = \mathcal{T}_{r+i,c+j}$$

comprising $(2d+1)^2$ blocks, around the block $\mathcal{T}_{r,c} = \mathcal{T}(r:r+w-1, c:c+w-1,:)$, where $i,j = -d,\dots,0,\dots,d$, and d represents the neighbour width.

- Each tensor $\mathcal{Y}_{r,c}$ was then approximated with bounded approximation error, where $\delta^2 = 3\sigma^2 w^2 (2d+1)^2$, and σ the noise level.
- For a color image $\mathbf{Y}$ of size $I \times J \times 3$, degraded by additive Gaussian noise, the basic idea behind the proposed method is that for each block of pixels of size $h \times w \times 3$, given by $\mathbf{Y}_{r,c} = \mathbf{Y}(r:r+h-1, c:c+w-1,:)$, a small tensor $\mathcal{Y}_{r,c}$ of size $h \times w \times 3 \times (2d+1) \times (2d+1)$, comprising $(2d+1)^2$ blocks around $\mathbf{Y}_{r,c}$ is constructed, in the form

$$\mathcal{Y}_{r,c}(:,:,:,d+1+i,d+1+j) = \mathbf{Y}_{r+i,c+j}$$

, where $i,j = -d,\dots,0,\dots,d$, and d represents the neighbourhood width.

Image denoising II

- Every (r, c)-block $\mathbf{Y}_{r,c}$ is then approximated
$$\|\mathcal{Y}_{r,c} - \mathcal{X}_{r,c}\|_F^2 \leq \varepsilon^2$$
where ε^2 is the noise level.
- A pixel is then reconstructed as an average of all its approximations by approximated tensors which cover that pixel.
- DCT spatial filtering used as preprocessing.

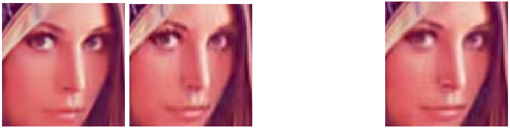
Table: The performance comparison of algorithms considered in terms of PSNR (dB) and SSIM for image denoising when SNR = 10 dB.

Algorithms	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
	Lena		Tiffany		Barbara	
EPC	**33.34**	**0.924**	**36.01**	**0.932**	**33.73**	**0.927**
TT-SVD	32.68	0.892	35.47	0.913	33.04	0.901
Tucker	32.74	0.919	35.47	0.926	32.92	0.919
BRTF	32.07	0.840	35.07	0.888	33.10	0.899
K-SVD	32.72	0.908	35.61	0.928	32.64	0.908
	Pepper		Pens		House	
EPC	**33.15**	**0.926**	**32.20**	**0.896**	**35.41**	**0.895**
TT-SVD	32.07	0.861	31.61	0.884	34.38	0.877
Tucker	32.23	0.917	31.27	0.884	34.40	0.885
BRTF	31.42	0.825	31.82	0.877	33.67	0.823
K-SVD	32.60	0.918	31.14	0.862	34.67	0.881

Image denoising III

(a) Noisy image at SNR = 10 dB

(b) TKD, SSIM = 0.919

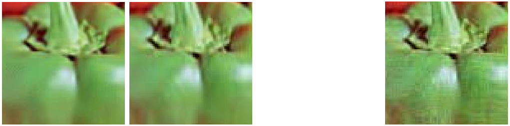


(c) From left to right, EPC(SSIM = 0.924), Tucker(0.919), TT-SVD(0.892), BRTF(0.840) and K-SVD(0.908)

Image denoising IV

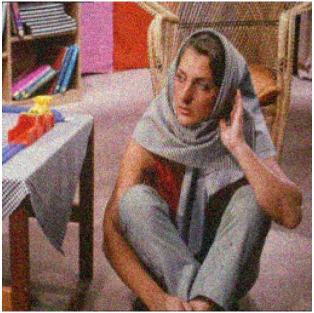
(d) Noisy image at SNR = 10 dB

(e) TKD, SSIM = 0.917



(f) From left to right, EPC(SSIM = 0.926), Tucker(0.917), TT-SVD(0.861), BRTF(0.825) and K-SVD(0.918)

Image denoising V



(g) Noisy image at SNR = 10 dB

(h) TKD, SSIM = 0.927

(i) From left to right, EPC(SSIM = 0.927), Tucker(0.919), TT-SVD(0.901), BRTF(0.899) and K-SVD(0.908)

Decomposition of Pepper color image I

- This example compares the proposed algorithm to TKD+CPD which first performs TKD then CPD.
- We decompose the Pepper image of size $128 \times 128 \times 3$ into three factor matrices with fixed ranks $R_1 = R_2 = 6$ and $R_3 = 2$, but different number of columns $R = 7, 8, \ldots$.
- Approximation errors show that LrCPD gives a better approximation than TKD+CPD.
- CPD is sensitive to the initial values and unstable because the decomposition is with rank exceeding the tensor dimensions. Its approximation error does not always decrease with increasing the rank. The fixed subspaces obtained by TKD are not optimal for the low-rank constrained CPD.

Decomposition of Pepper color image II

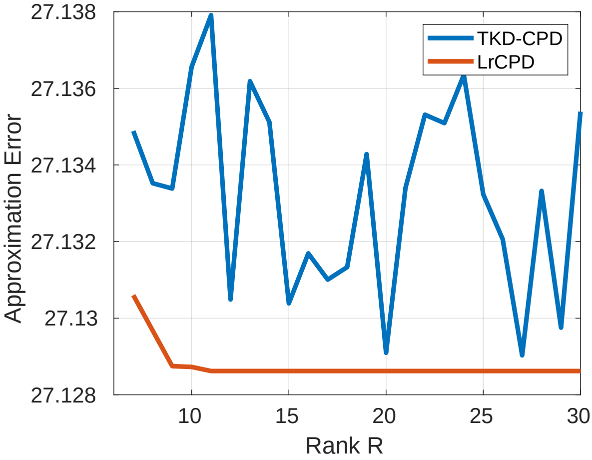
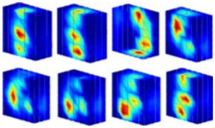
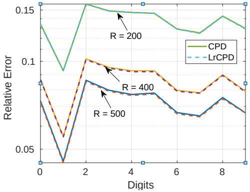


Figure: Decomposition of the Pepper image using LrCPD and TKD+CPD with $R_1 = R_2 = 6$. The parameters in CPD are initialized randomly.

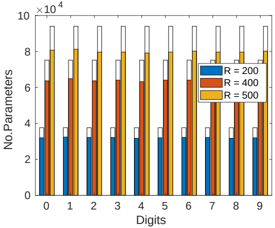
Decomposition of Gabor tensor of MNIST handwritten digit images I



(a) Gabor tensor of a handwritten digit image



(b)



(c) CPD (white (blank) boxes) vs LrCPD (shading boxes)

- The CPD and the LrCPD decompose Gabor tensors of size $28 \times 28 \times 32 \times 100$ computed from 100 images for each digit.
- ALS+line search is used to initialize the constrained CPD, and $\varepsilon = \|\mathcal{Y} - \mathcal{Y}_{cpd}\|_F^2 / \|\mathcal{Y}\|_F^2$

Decomposition of Gabor tensor of MNIST handwritten digit images II

- The new models have a fewer number of parameters and yield lower approximation errors than the CPD results.
  For the same CP-ranks $R = 200, 400, 500$, the constrained CPD yields models with less 5547, 11376, and 14002 parameters on average for all digits images than the ordinary CPD, respectively.

References I

Bro, R. and Andersson, C. (1998). Improving the speed of multiway algorithms - Part II: Compression. *Chemometrics and Intelligent Laboratory Systems*, 42:105–113.

Bro, R., Harshman, R. A., Sidiropoulos, N. D., and Lundy, M. E. (2009). Modeling multi-way data with linearly dependent loadings. *Journal of Chemometrics*, 23(7-8):324–340.

Carroll, J., Pruzansky, S., and Kruskal, J. (1970). Candelinc: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, 45(1):3–24.

De Lathauwer, L., Moor, B. D., and Vandewalle, J. (2000). On the best rank-1 and rank-(R1,R2,. . . ,RN) approximation of higher-order tensors. *SIAM Journal of Matrix Analysis and Applications*, 21(4):1324–1342.