Tensor Decomposition with Smoothness (ICML2017)

Masaaki Imaizumi\(^1\) Kohei Hayashi\(^2,3\)

\(^1\)Institute of Statistical Mathematics
\(^2\)National Institute of Advanced Industrial Science and Technology
\(^3\)RIKEN Center for Advanced Intelligence Project

August 4, 2017
### Tensor Data

- **Tensor data**
  - Data as a multi-dimensional array.
  - $X \in \mathbb{R}^{I_1 \times \cdots \times I_K}$
    - $K$ : mode, $I_k$ : # elements

#### 3D fMRI image

$(X$-axis $\times$ $Y$-axis $\times$ $Z$-axis)

#### Recommendation system

$(item \times time \times user)$
Tensor Decomposition : Low Rank

- **Tensor data are high-dimensional.**
  - E.g., 3D image: 1000 pixels for each axis $\Rightarrow 10^9$ pixels
- **Dimension Reduction** is important
Tensor Decomposition : Low Rank

- **Tensor data are high-dimensional.**
  - E.g., 3D image: 1000 pixels for each axis ⇒ $10^9$ pixels

- **Dimension Reduction** is important
  - Tensors are low-rank ⇒ **Tucker Decomposition** (Tucker (1966))

\[
X = \sum_{r_1, \ldots, r_K = 1}^{R_1, \ldots, R_K} g_{r_1, \ldots, r_K} x^{(k)}_{r_1} \otimes x^{(k)}_{r_2} \otimes \ldots \otimes x^{(k)}_{r_K}
\]

$(R^X_1, \ldots, R^X_K)$: rank of $X$, $g_{r_1, \ldots, r_K}$: coefficients

![Tensor Decomposition Diagram](image)
**New Approach for Dimension Reduction**

- **Smoothness appears in real data.**
  - A pair of adjacent elements are similar.

Time Series data (Smooth in Time)

Image data (Smooth in Location)
Idea of Smoothness

Generating Process for Smooth Tensor

- Tensor data are generated by smooth functions.

Tensor (matrix) with Smoothness
Idea of Smoothness

- **Generating Process for Smooth Tensor**
  - Tensor data are generated by smooth functions.

Tensor (matrix) with Smoothness

Generating Process (Smooth Function)

⇒ Generate
Idea of Smoothness

Generating Process for Smooth Tensor

- Tensor data are generated by smooth functions.

Our Idea

Introduce the smooth generating function into the tensor data.
Let $X \in \mathbb{R}^{I_1 \times \cdots \times I_K}$ be a $K$-mode tensor.

- Function: $f_X : [0, 1]^K \rightarrow \mathbb{R}$, Grids: $\{(g_{j_1}, \ldots, g_{j_K}) \in [0, 1]^K\}$
- $X$ is represented by observed value of $f_X$ on the grids.

$$[X]_{j_1 \ldots j_K} = f_X(g_{j_1}, \ldots, g_{j_K}).$$

**Main Assumption**

$f_X$ is smooth (differentiable).
Idea of Smoothness

Functional Representation of Tensors

**Representation Theorem**

\[\{\phi_j(\cdot) : [0, 1] \rightarrow \mathbb{R}\}_{j=1}^{\infty} : \text{orthonormal basis (given)}\]

\[f_X(g_1, \ldots, g_K) = \sum_{m_1} \cdots \sum_{m_K} w_{m_1\ldots m_K} \phi_{m_1}(g_1) \cdots \phi_{m_K}(g_K).\]

\[\Downarrow \text{Discretize} \left( [X]_{j_1\ldots j_K} = f_X(g_{j_1}, \ldots, g_{j_K}) \right)\]

**Smooth Tensor Formation**

\[W_X \in \mathbb{R}^{M(1) \times \cdots \times M(K)} : \text{coefficient tensor (} M^{(k)} : \text{# basis, } M^{(k)} \leq I_k \text{)}\]

\[[X]_{j_1\ldots j_K} = \sum_{m_1=1}^{M(1)} \cdots \sum_{m_K=1}^{M(K)} [W_X]_{m_1\ldots m_K} \phi_{m_1}(g_{j_1}) \cdots \phi_{m_K}(g_{j_K}).\]
Idea of Smoothness

Image of the formation

- \((\phi_1, \ldots, \phi_M)\) is given.
- \(f_X\) is smooth \(\Rightarrow\) small \(M^{(k)}\) can represent \(f_X\) \(\Rightarrow\) \(W_X\) is small

**Dimension Reduction by Smoothness.**
Idea of Smoothness

**Decomposition Method**

- A model for tensor completion ($n$ elements are observed).

\[ Y = \mathcal{X}(X^*) + \mathcal{E}. \]

- $X^* \in \mathbb{R}^{I_1 \times \cdots \times I_K}$: true tensor (unknown)
- $Y \in \mathbb{R}^n$: observed vector
- $\mathcal{X}: \mathbb{R}^{I_1 \times \cdots \times I_K} \to \mathbb{R}^n$: rearranging operator (known)
- $\mathcal{E} \in \mathbb{R}^n$: noise vector (each element is i.i.d. Gaussian)

- Method for Tucker decomposition with **low-rank** $X^*$ (Liu et al.(2009))

\[
\min_X \left[ \frac{1}{2n} \|Y - \mathcal{X}(X)\|^2 + \lambda_n \|X\|_s \right],
\]

where $\|\cdot\|_s$ is the Schatten-1 norm (rank regularization for $X$).
Our Decomposition Method

- Decomposition method for smooth $X$
  - Regularize the coefficient tensor $W_X$.
- Smooth Tucker Decomposition (STD)

$$\min_{W_X} \left[ \frac{1}{2n} \| Y - \mathcal{X}(X) \|_F^2 + \lambda_n \| W_X \|_s + \mu_n \| W_X \|_F^2 \right]$$

- Empirical Loss
- Rank Penalty
- Volume Penalty

$\| \cdot \|_F$: the Frobenius norm, $\| \cdot \|_s$: Schatten-1 norm

- Solved by the alternating direction method of multipliers (ADMM).
Error Bound

- **Decomposition Accuracy**
  - $X^*$: true tensor, $\hat{X}$: estimated by STD
  - $(R^W_1, \ldots, R^W_K)$: the Tucker rank of $W_X$.

**Theorem 1**

Suppose the smoothness and some assumptions hold, then with high probability,

$$
\|\| \hat{X} - X^* \|\|_F^2 \leq C \left( K^{-1} \sum_{k=1}^K \sqrt{I_k} + \sqrt{I_{\setminus k}} \right)^2 \left( K^{-1} \sum_{k=1}^K \sqrt{R^W_k} \right)^2.
$$

From Noise

$$
\text{From Noise} := A
$$

- By the original Tucker decomposition (Tomioka et al.(2011)),

$$
A = \left( K^{-1} \sum_{k=1}^K \sqrt{R^X_k} \right)^2. \quad (R^X_k \text{ is rank of } X, \ R^W_k \leq R^X_k \text{ in general}.)
$$
Experiments for Accuracy

- **Experiments**: Recovery $X^*$ with noise
  - Compare the squared error with a smooth $X^*$ and a nonsmooth $X^*$.
  - The proposed method outperforms when $X^*$ is smooth.

![Graphs showing MSE for smooth and non-smooth matrices with varying noise levels.](image-url)
Function Estimation

- We can estimate $f_X$, not only $X^*$
  - $\hat{W}_X$: the minimizer of the problem of STD.
  - Define the estimator for $f_X$ as
    $\hat{f} := \sum_{m_1=1}^{M^{(1)}} \cdots \sum_{m_K=1}^{M^{(K)}} [\hat{W}_X]_{m_1 \cdots m_K} \phi_{m_1} \cdots \phi_{m_K}$

Theorem 2
Suppose the conditions for Theorem 1 hold. Then, we have

$$\sup_{g \in [0,1]^K} |\hat{f}(g) - f_X(g)| \leq \text{Same Bound}$$
Experiments (image interpolation)

- Interpolate amino acid data (Kier et al. (1998)).
- Shed light to amino acids and measure the volume of absorbed and reflected light with each wavelength.

**Figure:** Completion of missing elements of the acid data.
Experiments (motion interpolation)

- Interpolate video data (Schuldt et al. (2004)).

  0.14 sec
  (50% missing)
  Tucker STD (proposed)

  0.18 sec
  Tucker
  STD (proposed)

  (100% missing)
  Tucker
  STD (proposed)

  0.22 sec
Topics
- High-dimensional tensor data.

Idea
- Using information of real data improves analysis.
- Smoothness is a key factor for dimension reduction.

Result
- Accurate and good analysis.


Images

- Irasuto-ya http://www.irasutoya.com
- Slicer Web http://slicer.org