Tensor Decomposition with Smoothness (ICML2017)

Masaaki Imaizumi[1] Kohei Hayashi[2,3]

[1]Institute of Statistical Mathematics

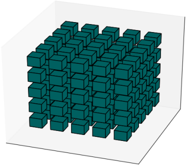
[2]National Institute of Advanced Industrial Science and Technology

[3]RIKEN Center for Advanced Intelligence Project

August 4, 2017

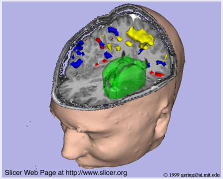
Tensor Data

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



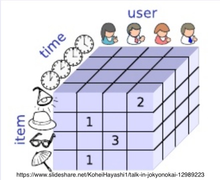
3D fMRI image

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



Recommendation system

(item $\times$ time $\times$ user)



Tensor Decomposition : Low Rank

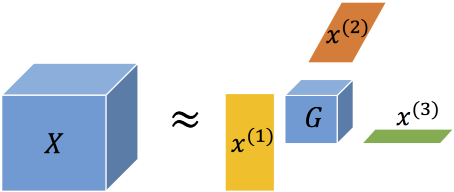
- **Tensor data are high-dimensional.**
  - E.g., 3D image : 1000 pixels for each axis $\Rightarrow 10^9$ pixels
- **Dimension Reduction** is important

Tensor Decomposition : Low Rank

- **Tensor data are high-dimensional.**
  - E.g., 3D image : 1000 pixels for each axis $\Rightarrow 10^9$ pixels
- **Dimension Reduction** is important
  - Tensors are low-rank $\Rightarrow$ **Tucker Decomposition** (Tucker (1966))

$$X = \sum_{r_1,\dots,r_K=1}^{R_1,\dots,R_K} g_{r_1,\cdots,r_K} x_{r_1}^{(k)} \otimes x_{r_2}^{(k)} \otimes \dots x_{r_K}^{(k)}$$

$(R_1^X, \dots, R_K^X)$: rank of X, $g_{r_1,\dots,r_K}$: coefficients



New Approach for Dimension Reduction

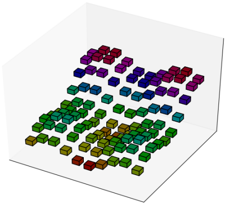
- **Smoothness appears in real data**.
  - A pair of adjacent elements are similar.

Time Series data
(Smooth in Time)

Image data
(Smooth in Location)

Idea of Smoothness

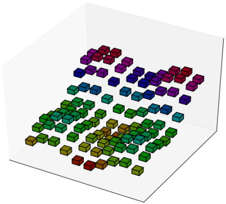
- **Generating Process for Smooth Tensor**
  - Tensor data are generated by smooth functions.



Tensor (matrix)
with Smoothness

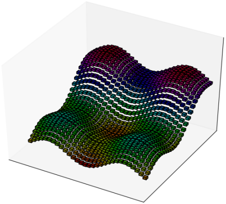
Idea of Smoothness

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



Tensor (matrix)
with Smoothness

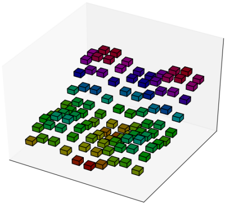
$\Leftarrow$
Generate



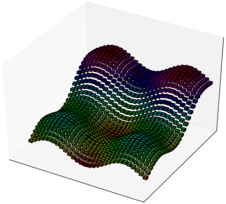
Generating Process
(Smooth Function)

Idea of Smoothness

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



$\Leftarrow$
Generate



Tensor (matrix)
with Smoothness

Generating Process
(Smooth Function)

Our Idea

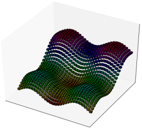
Introduce **the smooth generating function** into the tensor data.

Smoothness Formation of Tensors

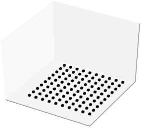
- Let $X \in \mathbb{R}^{I_1 \times \cdots \times I_K}$ be K-mode tensor.
  - Function : $f_X : [0,1]^K \to \mathbb{R}$, Grids : $\{(g_{j_1}, \ldots, g_{j_K}) \in [0,1]^K\}$
  - X is represented by observed value of f_X on the grids.

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

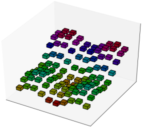
$K = 2$ (matrix) case



$f_X(\cdot,\cdot)$



$\{(g_{j_1}, g_{j_2}) \in [0,1]^2\}$



$f_X(g_{j_1}, g_{j_2})$

Main Assumption

f_X is smooth (differentiable).

Functional Representation of Tensors

Representation Theorem

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

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

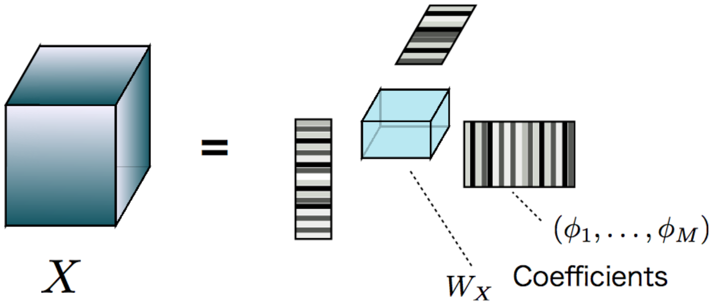
$\Downarrow$ **Discretize** $([X]_{j_1 \ldots j_K} = f_X(g_{j_1}, \ldots, g_{j_K}))$

Smooth Tensor Formation

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

$$[X]_{j_1 \ldots j_K} = \sum_{m_1=1}^{M^{(1)}} \cdots \sum_{m_K=1}^{M^{(K)}} \underbrace{[W_X]_{m_1 \ldots m_K}}_{\text{coefficient}} \underbrace{\phi_{m_1}(g_{j_1}) \cdots \phi_{m_K}(g_{j_K})}_{\text{given}}.$$

Image of the formation



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

Dimension Reduction by Smoothness.

Decomposition Method

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

$$Y = \mathfrak{X}(X^*) + \mathcal{E}.$$

  - $X^* \in \mathbb{R}^{I_1 \times \cdots \times I_K}$: true tensor (unknown)
  - $Y \in \mathbb{R}^n$: observed vector
  - $\mathfrak{X} : \mathbb{R}^{I_1 \times \cdots \times I_K} \to \mathbb{R}^n$: 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))

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

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)**

$$\min_{W_X}\left[\underbrace{\frac{1}{2n}\|Y-\mathfrak{X}(X)\|_F^2}_{\text{Empirical Loss}}+\underbrace{\lambda_n |||W_X|||_s}_{\text{Rank Penalty}}+\underbrace{\mu_n |||W_X|||_F^2}_{\text{Volume Penalty}}\right],$$

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

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

Error Bound

- **Decomposition Accuracy**
  - X^* : true tensor, $\widehat{X}$: estimated by STD
  - $(R_1^W, \dots, R_K^W)$: the Tucker rank of W_X.

Theorem 1

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

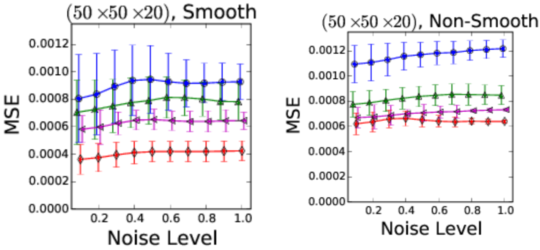
$$\|\widehat{X} - X^*\|_F^2 \leq C \underbrace{\left(K^{-1} \sum_{k=1}^{K} \sqrt{I_k} + \sqrt{I_{\backslash k}} \right)^2}_{\text{From Noise}} \underbrace{\left(K^{-1} \sum_{k=1}^{K} \sqrt{R_k^W} \right)^2}_{:=\mathbb{A}}.$$

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

$$\mathbb{A} = \left(K^{-1} \sum_{k=1}^{K} \sqrt{R_k^X} \right)^2. \quad (R_k^X \text{ is rank of } X, \; R_k^W \leq R_k^X \text{ in general.})$$

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



Proposed
Tucker 1
Tucker 2
Matrix

Function Estimation

- **We can estimate f_X, not only X^***
  - $\widehat{W}_X$: the minimizer of the problem of STD.
  - Define the estimator for f_X as

$$\widehat{f} := \sum_{m_1=1}^{M^{(1)}} \cdots \sum_{m_K=1}^{M^{(K)}} [\widehat{W}_X]_{m_1 \dots m_K} \phi_{m_1} \cdots \phi_{m_K}$$

Theorem 2

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

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

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.

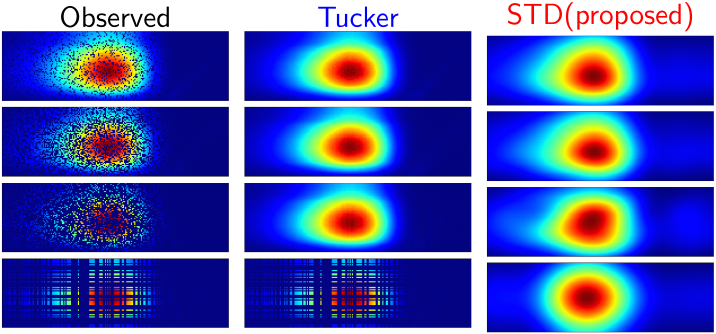
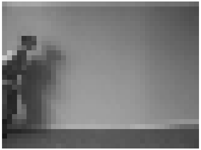


Figure: Completion of missing elements of the acid data.

Experiments (motion interpolation)

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

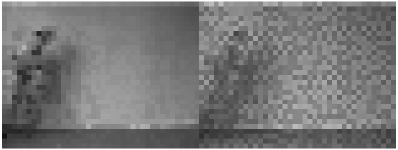
0.14 sec



0.18 sec

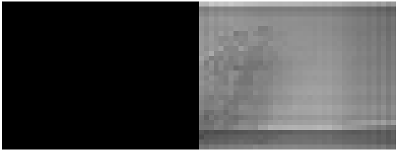
(50% missing)

Tucker STD(proposed)

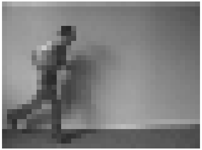


(100% missing)

Tucker STD(proposed)



0.22 sec



Summary

- **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.

Reference

- Liu, Ji, Musialski, Przemyslaw, Wonka, Peter, and Ye, Jieping. Tensor completion for estimating missing values in visual data. In ICCV, 2009.
- Tomioka, Ryota, Suzuki, Taiji, Hayashi, Kohei, and Kashima, Hisashi. Statistical performance of convex tensor decomposition. In NIPS, 2011.
- Tucker, Ledyard R. Some mathematical notes on three-mode factor analysis. Psychometrika, 31(3):279–311, 1966.
- Kiers, Henk AL. A three-step algorithm for candecomp/parafac analysis of large data sets with multi- collinearity. Journal of Chemometrics, 12(3):155–171, 1998.
- Schuldt, Christian, Laptev, Ivan, and Caputo, Barbara. Recognizing human actions: A local svm approach. In Pat- tern Recognition, 2004. In ICPR 2004.

Images

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