Mathematical Aspects of Tensor Subspace Method

Hayato Itoh[1], Atsushi Imiya[2(✉)], and Tomoya Sakai[3]

[1] School of Advanced Integration Science, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan
hayato-itoh@graduate.chiba-u.jp

[2] Institute of Media and Information Technology, Chiba University, Yayoi-cho 1-33, Inage-ku, Chiba 263-8522, Japan
imiya@faculty.chiba-u.jp

[3] Graduate School of Engineering, Nagasaki University, Bunkyo-cho 1-14, Nagasaki 852-8521, Japan
tsakai@cis.nagasaki-u.ac.jp

Abstract. The mathematical and computational backgrounds of pattern recognition are the geometries in Hilbert space for functional analysis and applied linear algebra for numerical analysis, respectively. Organs, cells and microstructures in cells dealt with in biomedical image analysis are volumetric data. We are required to process and analyse these data as volumetric data without embedding vector space from the viewpoints of object oriented data analysis. Therefore, sampled values of volumetric data are expressed as three-way array data. These three-way array data are expressed as the third order tensor. This embedding of the data leads us to the construction of subspace method for higher-order tensors expressing multi-way array data.

1 Introduction

The aim of the paper is to clarify mathematical properties of pattern recognition of tensor data from the viewpoints of object oriented data analysis for volumetric data, which dealt with in biomedical image analysis, retrievals and recognition. In traditional pattern recognition, sampled patterns for numerical computation are embedded in an appropriate-dimensional Euclidean space as vectors. The other way is to deal with sampled patterns as three-way array data. These three-way array data are expressed as tensors [6–10] to preserve the linearity of the original pattern space.

The subspace method based on Karhunen-Loève transform is a fundamental technique in pattern recognition. Modern pattern recognition techniques for sampled value of patterns are described using linear algebra for sampled value embedded in vector space. Organs, cells and microstructures in cells dealt with in biomedical image analysis are volumetric data. We are required to process and analyse these data as volumetric data without embedding sampled values in vector space from the viewpoints of object oriented data analysis [5]. We express sampled values of volumetric data as three-way array data. These three-way

A. Robles-Kelly et al. (Eds.): S+SSPR 2016, LNCS 10029, pp. 37–48, 2016.
DOI: 10.1007/978-3-319-49055-7_4

array data are processed as the third order tensor. This expression of data requires to develop subspace method for tensor data.

We derive and clarify the mutual subspace and constrained mutual subspace methods for tensors using tensor PCA based on the Tucker-3 tensor decomposition. The mutual subspace method is stable against geometric perturbation of queries for pattern recognition, since the method assumes that a query is an element of a low-dimensional subspace. Furthermore, since the constrained subspace method eliminates the common parts of subspaces among categories, the method confirms robust recognition against global deformation of queries in Hilbert space. Since the constrained mutual subspace method is a combination of the mutual subspace method and constrained subspace method the method is stable and robust both against geometric perturbations and global deformations of queries for pattern recognition.

2 Pattern Recognition in Vector Space

A volumetric pattern is assumed to be a square integrable function in a linear space and to be defined on a finite support in three-dimensional Euclidean space [1–3] such that $\int_{\Omega} |f|^2 d\boldsymbol{x} < \infty$ for $\Omega \subset \mathbb{R}^3$. Furthermore, we assume $\int_{\Omega} |\nabla f|^2 d\boldsymbol{x} < \infty$ and $\int_{\Omega} tr\{(\nabla\nabla^\top f)^\top (\nabla\nabla^\top f)\} d\boldsymbol{x} < \infty$, where $\nabla\nabla^\top f$ is the Hessian matrix of f. For an orthogonal projection $\boldsymbol{P}_\perp = \boldsymbol{I} - \boldsymbol{P}$, $f^\parallel = \boldsymbol{P} f$ and $f^\perp = \boldsymbol{P}_\perp f$ are the canonical element and canonical form of f with respect to $\boldsymbol{P}$ and $\boldsymbol{P}_\perp$, respectively. If $\boldsymbol{P}$ is the projection to the space spanned by the constant element, the operation $\boldsymbol{P}_\perp f$ is called the constant canonicalisation. Let $\boldsymbol{P}_i$ be the orthogonal projection to the linear subspace corresponding to the category C_i. For a pattern f, if $|\boldsymbol{P}_{i^*}(f/|f|)| \leq \delta$ for an appropriately small positive number δ, we conclude that $f \in C_{i^*}$.

Setting $\boldsymbol{\delta}$ and ε to be a small vector and a small positive number, we have the relation

$$|f(\boldsymbol{x} + \boldsymbol{\delta}) - (f(\boldsymbol{x}) + \boldsymbol{\delta}^\top \nabla f + \frac{1}{2} \boldsymbol{\delta}^\top (\nabla\nabla^\top f) \boldsymbol{\delta})| < \varepsilon, \tag{1}$$

for local geometric perturbations. All f, f_x, f_y, f_z, f_{xx}, f_{yy}, f_{zz}, f_{xy}, f_{yz} and f_{zx} are independent, if f is not sinusoidal in each direction. Therefore, Eq. (1) implies that, for a pattern defined on three-dimensional Euclidean space, the local dimensions of a pattern are four and ten, if local geometric perturbations and local bending deformation of the pattern are assumed as local transformations to the pattern. This property of the local dimensionality allows us to establish the mutual subspace method, which deals with a query as a pattern in a subspace.

Setting (f, g) to be the inner product in Hilbert space $\mathfrak{H}$, the relation $|f|^2 = (f, f)$ is satisfied. Let θ be the canonical angle between a pair of linear subspaces L_1 and L_2. Setting $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ to be the orthogonal projections to L_1 and L_2, respectively, $\cos^2\theta$ is the maximiser of $(\boldsymbol{P}_1 f, \boldsymbol{P}_2 g)^2$ with respect to the conditions

$|f| = 1$, $|g| = 1$ $\boldsymbol{P}_1 f = f$ and $\boldsymbol{P}_2 g = g$. The relation $\cos^2\theta = \lambda_{\max}^2$ is satisfied, where $\lambda_{\max}$ is the maximal singular value of $\boldsymbol{P}_2\boldsymbol{P}_1$.

Since, in mutual subspace method, a query f is expressed by using a set of local bases, we set that $\boldsymbol{Q}_f$ is the orthogonal projection to linear subspace expressing the query f. Then, if the canonical angle between $\boldsymbol{Q}_f$ and $\boldsymbol{P}_i$ satisfies the relation $\angle(\boldsymbol{Q}_f, \boldsymbol{P}_i) < \angle(\boldsymbol{Q}_f, \boldsymbol{P}_i^*)$ for all C_i, we conclude that $f \in C_{i^*}$.

Setting $\boldsymbol{P}_i$ to be the orthogonal projection to linear subspace $\mathfrak{L}_i$ corresponding to the category C_i, the orthogonal projection which maximises the criterion $J = \sum_{i=1}^n |\boldsymbol{Q}\boldsymbol{P}_i|_2^2$ with respect to the condition $\boldsymbol{Q}^*\boldsymbol{Q} = \boldsymbol{I}$ where $\boldsymbol{Q}^*$ is the conjugate of $\boldsymbol{Q}$ and $|\boldsymbol{A}|$ is the trace norm of the operator $\boldsymbol{A}$ in Hilbert space $\mathfrak{H}$. Though operation $\boldsymbol{Q}f$ removes common part for all categories from f, $(\boldsymbol{I} - \boldsymbol{Q})f$ preserves essentially significant parts for pattern recognition of f.

For f and g in $\mathfrak{H}$, we define the metric d for $\mu(f)$ and $\mu(g)$, such that $d(\mu(f), \mu(g))$, using an appropriate transform μ from $\mathfrak{H}$ to its subset. Furthermore, using an appropriate mapping Φ, we define a measure

$$s(f, g) = \Phi(d(\mu(f), \mu(g))). \tag{2}$$

If we set $\mu(f) = \frac{f}{|f|}$ and set d and Φ the geodesic distance on the unit sphere in $\mathfrak{H}$ and $\Phi(x) = \cos x$, respectively, $s(f, g)$ becomes the similarity measure based on the angle between f and g. For $f' = f + \delta_f$ and $g' = g + \delta_g$, setting

$$\min(|f|, |g|) = \Lambda, \;\; \max(\delta_f, \delta_g) = \Delta, \tag{3}$$

we have the relation

$$\left| \left(\frac{f'}{|f'|}, \frac{g'}{|g'|} \right) - \left(\frac{f}{|f|}, \frac{g}{|g|} \right) \right| = c\frac{\Delta}{\Lambda}, \tag{4}$$

for a positive constant c. Therefore, $s(f, g)$ is stable and robust against perturbations and noises for f and g.

For patterns in $\mathfrak{H}$, we have the following property.

Property 1. For $|f| = 1$ and $|g| = 1$, assuming $|f - g| \leq \frac{1}{3} \cdot \frac{\pi}{2}$ the geodesic distance $\theta = d_S(f, g)$ between f and g satisfies the relation $|\theta - |f - g|| < \varepsilon$ for a positive small number ε.

In traditional pattern recognition, these sampled patterns are embedded in an appropriate-dimensional Euclidean space as vectors. For $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{R}^{n \times n}$, $|\boldsymbol{x}|_2$ and $|\boldsymbol{X}|_F$ are the vector norm and Frobenius norm of $\boldsymbol{x}$ and $\boldsymbol{X}$, respectively.

Setting the data matrix $\boldsymbol{X}$ to be $\boldsymbol{X} = (\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_m)$ for data vectors $\{\boldsymbol{f}_i\}_{i=1}^m$ in $\mathbb{R}^N$, whose mean is zero, the Karhunen-Loève transform is established by computing $\hat{\boldsymbol{f}}_i = \boldsymbol{U}\boldsymbol{f}_i$ for $\boldsymbol{U}$ which minimises $J_1 = |\boldsymbol{U}\boldsymbol{X}|_F^2$ with the condition $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_N$. The orthogonal matrix $\boldsymbol{U}$ is the minimiser of

$$J_{11} = |\boldsymbol{U}\boldsymbol{X}|_F^2 + \langle(\boldsymbol{U}^\top\boldsymbol{U} - \boldsymbol{I})\boldsymbol{\Lambda}\rangle \tag{5}$$

where

$$\boldsymbol{\Lambda} = Diag(\lambda_1, \lambda_2, \cdots, \lambda_N) \tag{6}$$

for

$$\lambda_1 \geq \lambda_2 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0. \tag{7}$$

The minimiser of Eq. (5) is the solution of the eigenmatrix problem

$$\boldsymbol{M}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{\Lambda},\ \boldsymbol{M} = \boldsymbol{X}\boldsymbol{X}^\top \tag{8}$$

The row vectors of $\boldsymbol{U}$ are the principal components.

The compression of $\boldsymbol{f}_i$ to a low-dimensional linear subspace is achieved by computing the transform $\boldsymbol{P}_n\boldsymbol{U}\boldsymbol{f}$, where $\boldsymbol{P}_n$, for $n < N$, is the orthogonal projection such that

$$\boldsymbol{P}_n = \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{O} \\ \boldsymbol{O}^\top & \boldsymbol{O} \end{pmatrix}. \tag{9}$$

3 Pattern Recognition in Multi-linear Forms

For the triplet of positive integers I_1, I_2 and I_3, the third-order tensor $\mathbb{R}^{I_1 \times I_2 \times I_3}$ is expressed as $\mathcal{X} = ((x_{ijk}))$ Indices i, j and k are called the 1-mode, 2-mode and 3-mode of $\mathcal{X}$, respectively. The tensor space $\mathbb{R}^{I_1 \times I_2 \times I_3}$ is interpreted as the Kronecker product of three vector spaces $\mathbb{R}^{I_1}$, $\mathbb{R}^{I_2}$ and $\mathbb{R}^{I_3}$ such that $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \otimes \mathbb{R}^{I_3}$. We set $I = \max(I_1, I_2, I_3)$.

For a square integrable function $f(\boldsymbol{x})$, which is zero outside of a finite support Ω in three-dimensional Euclidean space, the sample $Sf(\Delta\mathbf{z})$ for $\mathbf{z} \in \mathbf{Z}^3$ and $|\mathbf{z}|_\infty \leq I$ defines an $I \times I \times I$ three-way array $\mathbf{F}$. To preserve the multi-linearity of the function $f(\boldsymbol{x})$, we deal with the array $\mathbf{F}$ as a third-order tensor $\mathcal{F}$. The operation $vec\mathcal{F}$ derives a vector $\boldsymbol{f} \in \mathbb{R}^{I_{123}}$ for $I_{123} = I_2 \cdot I_2 \cdot I_3$. We can reconstruct f from $\mathcal{F}$ using an interpolation procedure.

For $\mathcal{X}$, the n-mode vectors, $n = 1, 2, 3$, are defined as the I_n-dimensional vectors obtained from $\mathcal{X}$ by varying this index i_n while fixing all the other indices.

The unfolding of $\mathcal{X}$ along the n-mode vectors of $\mathcal{X}$ is defined as matrices such that

$$\mathcal{X}_{(1)} \in \mathbb{R}^{I_1 \times I_{23}},\ \mathcal{X}_{(2)} \in \mathbb{R}^{I_2 \times I_{13}},\ \mathcal{X}_{(3)} \in \mathbb{R}^{I_3 \times I_{12}} \tag{10}$$

for $I_{12} = I_1 \cdot I_2$, $I_{23} = I.I_3$ and $I_{13} = I_1 \cdot I_3$, where the column vectors of $\mathcal{X}_{(j)}$ are the j-mode vectors of $\mathcal{X}$ for $i = 1, 2, 3$. We express the j-mode unfolding of $\mathcal{X}_i$ as $\mathcal{X}_{i,(j)}$.

For matrices

$$\boldsymbol{U} = ((u_{ii'})) \in \mathbb{R}^{I_1 \times I_1},\ \boldsymbol{V} = ((v_{jj'})) \in \mathbb{R}^{I_2 \times I_2},\ \boldsymbol{W} = ((w_{kk'})) \in \mathbb{R}^{I_3 \times I_3}, \tag{11}$$

the n-mode products for $n = 1, 2, 3$ of a tensor $\mathcal{X}$ are the tensors with entries

$$(\mathcal{X} \times_1 \boldsymbol{U})_{ijk} = \sum_{i'=1}^{I_1} x_{i'jk} u_{i'i},$$
$$(\mathcal{X} \times_2 \boldsymbol{V})_{ijk} = \sum_{j'=1}^{I_2} x_{ij'k} v_{j'j}, \tag{12}$$
$$(\mathcal{X} \times_3 \boldsymbol{W})_{ijk} = \sum_{k'=1}^{I_3} x_{ijk'} w_{k'k},$$

where $(\mathcal{X})_{ijk} = x_{ijk}$ is the ijk-th element of the tensor $\mathcal{X}$. The inner product of two tensors $\mathcal{X}$ and $\mathcal{Y}$ in $\mathbb{R}^{I_1 \times I_2 \times I_3}$ is

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i=1}^{I_1} \sum_{j=1}^{I_2} \sum_{k=1}^{I_3} x_{ijk} y_{ijk}. \tag{13}$$

Using this inner product, we have the Frobenius norm of a tensor $\mathcal{X}$ as $|\mathcal{X}|_F = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$. The Frobenius norm $|\mathcal{X}|_F$ of the tensor $\mathcal{X}$ satisfies the relation $|\mathcal{X}|_F = |\boldsymbol{f}|_2$, where $|\boldsymbol{f}|_2$ is the Euclidean norm of the vector $\boldsymbol{f}$.

To project a tensor $\mathcal{X}$ in $\mathbb{R}^{I_1 \times I_2 \times I_3}$ to the tensor $\mathcal{Y}$ in a lower-dimensional tensor space $\mathbb{R}^{P_1 \times P_2 \times P_3}$, where $P_n \leq I_n$, three projection matrices $\{\boldsymbol{U}^{(n)}\}_{n=1}^{3}$, for $\boldsymbol{U}^{(n)} \in \mathbb{R}^{I_n \times P_n}$ are required for $n = 1, 2, 3$. Using these three projection matrices, we have the tensor orthogonal projection such that

$$\mathcal{Y} = \mathcal{X} \times_1 \boldsymbol{U}^{(1)\top} \times_2 \boldsymbol{U}^{(2)\top} \times_3 \boldsymbol{U}^{(3)\top}. \tag{14}$$

This projection is established in three steps, where in each step, each n-mode vector is projected to a P_n-dimensional space by $\boldsymbol{U}^{(n)}$ for $n = 1, 2, 3$.

For a collection of matrices $\{\boldsymbol{F}_i\}_{i=1}^{N} \in \mathbb{R}^{m \times n}$ satisfying $\mathrm{E}_i(\boldsymbol{F}_i) = 0$, the orthogonal-projection-based data reduction $\hat{F}_i = \boldsymbol{U}^\top \boldsymbol{F}_i \boldsymbol{V}$ is performed by maximising

$$J_2(\boldsymbol{U}, \boldsymbol{V}) = \mathrm{E}_i \left(|\boldsymbol{U} \hat{\boldsymbol{F}}_i \boldsymbol{V}^\top|_{\mathrm{F}}^2 \right) \tag{15}$$

with respect to the conditions $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_m$ and $\boldsymbol{V}^\top \boldsymbol{V} = \boldsymbol{I}_n$. The solutions are the minimiser of the Euler-Lagrange equation

$$J_{22}(\boldsymbol{U}, \boldsymbol{V}) = \mathrm{E} \left(|\boldsymbol{U} \hat{\boldsymbol{F}}_i \boldsymbol{V}^\top|_{\mathrm{F}}^2 \right) + \langle (\boldsymbol{I}_m - \boldsymbol{U}^\top \boldsymbol{U}), \boldsymbol{\Sigma} \rangle + \langle (\boldsymbol{I}_n - \boldsymbol{V}^\top \boldsymbol{V}), \boldsymbol{\Lambda} \rangle \tag{16}$$

For diagonal matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$.

Setting

$$\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{F}_i^\top \boldsymbol{F}_i = \boldsymbol{M}, \ \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{F}_i \boldsymbol{F}_i^\top = \boldsymbol{N}, \tag{17}$$

$\boldsymbol{U}$ and $\boldsymbol{V}$ are the solutions of the eigendecomposition problems

$$\boldsymbol{M}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Lambda}, \ \boldsymbol{N}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{\Sigma}, \tag{18}$$

where $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$ are diagonal matrices satisfying the relationships $\lambda_i = \sigma_i$ for

$$\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_K, 0, \ldots, 0), \tag{19}$$
$$\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_K, 0, \ldots, 0). \tag{20}$$

The equation

$$(\boldsymbol{P}_1\boldsymbol{U})^\top \boldsymbol{X}(\boldsymbol{P}_2\boldsymbol{V}) = \boldsymbol{Y} \tag{21}$$

is rewritten as

$$(\boldsymbol{P}_2\boldsymbol{V} \otimes \boldsymbol{P}_1\boldsymbol{U})\mathrm{vec}\boldsymbol{X} = (\boldsymbol{P_2} \otimes \boldsymbol{P_1})(\boldsymbol{V} \otimes \boldsymbol{U})\boldsymbol{X} = \boldsymbol{P}\mathrm{vec}\boldsymbol{X} = \mathrm{vec}\boldsymbol{Y}. \tag{22}$$

Using three projection matrices $\boldsymbol{U}^{(i)}$ for $i = 1, 2, 3$, we have the tensor orthogonal projection for a third-order tensor as

$$\mathcal{Y} = \mathcal{X} \times_1 \boldsymbol{U}^{(1)\top} \times_2 \boldsymbol{U}^{(2)\top} \times_3 \boldsymbol{U}^{(3)\top}. \tag{23}$$

For a collection $\{\mathcal{X}_k\}_{k=1}^m$ of the third-order tensors, the orthogonal-projection-based dimension reduction procedure is achieved by maximising the criterion

$$J_3 = E_k(|\mathcal{X}_k \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \times_3 \boldsymbol{U}^{(3)}|_F^2) \tag{24}$$

with respect to the conditions $\boldsymbol{U}^{(i)^\top}\boldsymbol{U}^{(i)} = \boldsymbol{I}$ for $i = 1, 2, 3$. The Euler-Lagrange equation of this conditional optimisation problem is

$$\begin{aligned} J_{33}(\boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \boldsymbol{U}^{(3)}) &= E_k(|\mathcal{X}_k \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \times_3 \boldsymbol{U}^{(3)}|_F^2) \\ &\quad + \sum_{i=1}^{3} \langle (\boldsymbol{I} - \boldsymbol{U}^{(i)^\top}\boldsymbol{U}^{(i)}), \boldsymbol{\Lambda}^{(i)} \rangle. \end{aligned} \tag{25}$$

This minimisation problem is solved by the iteration procedure [11].

Algorithm 1. Iterative method for the tensor PCA

Data: $0 < \varepsilon \ll 1$
Data: $\alpha := 0$
Data: For $i = 1, 2, 3$, $\boldsymbol{U}^i_{(0)} := \boldsymbol{Q}^{(i)}$, where $\boldsymbol{Q}^{(i)\top}\boldsymbol{Q}^{(i)} = \boldsymbol{I}$
Result: Orthogonal matrices $\boldsymbol{U}^{(i)}$ for $i = 1, 2, 3$
while $|\boldsymbol{U}^{(i)}_{\alpha+1} - \boldsymbol{U}^{(i)}_{\alpha}| \gg \varepsilon$ **do**
  maximise $E_i|\boldsymbol{U}^{(1)\top}_{(\alpha+1)}\mathcal{X}_{i1}^\top\mathcal{X}_{i1}\boldsymbol{U}^{(1)}_{(\alpha+1)}|_F^2 + \langle(\boldsymbol{U}^{(1)}_{\alpha+1}\boldsymbol{U}^{(1)}_{(\alpha+1)} - \boldsymbol{I})\boldsymbol{\Lambda}^{(1)}_{(\alpha)}\rangle$;
    for $\mathcal{X}_{i1} = \mathcal{X}_i \times_2 \boldsymbol{U}^{(2)\top}_{(\alpha)} \times_3 \boldsymbol{U}^{(3)\top}_{(\alpha)}$;
  maximise $E_i|\boldsymbol{U}^{(2)\top}_{(\alpha+1)}\mathcal{X}_{i2}^\top\mathcal{X}_{i2}\boldsymbol{U}^{(2)}_{(\alpha+1)}|_F^2 + \langle(\boldsymbol{U}^{(2)}_{(\alpha+1)}\boldsymbol{U}^{(2)}_{(\alpha+1)} - \boldsymbol{I})\boldsymbol{\Lambda}^{(2)}_{(\alpha)}\rangle$;
    for $\mathcal{X}_{i2} = \mathcal{X}_i \times_1 \boldsymbol{U}^{(1)\top}_{(\alpha+1)} \times_3 \boldsymbol{U}^{(3)\top}_{(\alpha)}$;
  maximise $E_i|\boldsymbol{U}^{(3)\top}_{(\alpha+1)}\mathcal{X}_{i3}^\top\mathcal{X}_{i3}\boldsymbol{U}^{(3)}_{(\alpha+1)}|_F^2 + \langle(\boldsymbol{U}^{(3)}_{(\alpha+1)}\boldsymbol{U}^{(3)}_{(\alpha+1)} - \boldsymbol{I})\boldsymbol{\Lambda}^{(3)}_{(\alpha)}\rangle$;
    for $\mathcal{X}_{i3} = \mathcal{X}_i \times_1 \boldsymbol{U}^{(1)\top}_{(\alpha+1)} \times_2 \boldsymbol{U}^{(2)\top}_{(\alpha+1)}$;
  $\alpha := \alpha + 1$
end

As an relaxation of the iteration algorithm, we define the system of optimisation problems

$$J_j = E(|\boldsymbol{U}^{(j)\top}\mathcal{X}_{i,(j)}\boldsymbol{U}^{(j)}|_F^2) + \langle(\boldsymbol{U}^{(j)\top}\boldsymbol{U}^{(j)} - \boldsymbol{I}_j), \boldsymbol{\Lambda}^{(j)}\rangle \tag{26}$$

for $i = 1, 2, 3$, where $\mathcal{X}_{i,(j)}$ is the ith column vector of the unfolding matrix $\mathcal{X}_{(j)}$. These optimisation problems derive a system of eigenmatrix problems

$$\boldsymbol{M}^{(j)}\boldsymbol{U}^{(j)} = \boldsymbol{U}^{(j)}\boldsymbol{\Lambda}^{(j)}, \ \ \boldsymbol{M}^{(j)} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{X}_{i,(j)}\mathcal{X}_{i,(j)}^{\top} \tag{27}$$

for $j = 1, 2, 3$.

Setting $\boldsymbol{P}^{(j)}$ to be an orthogonal projection in the linear space $\mathcal{L}(\{\boldsymbol{u}_i^{(j)}\}_{i=1}^{I_j})$ spanned by the column vectors od $\boldsymbol{U}^{(j)}$, data reduction is computed by

$$\mathcal{Y} = \mathcal{X} \times_1 \boldsymbol{P}^{(1)}\boldsymbol{U}^{(1)} \times_2 \boldsymbol{P}^{(2)}\boldsymbol{U}^{(2)} \times_3 \boldsymbol{P}^{(3)}\boldsymbol{U}^{(3)}. \tag{28}$$

This expression is equivalent to the vector form as

$$vec\mathcal{Y} = (\boldsymbol{P}^{(3)} \otimes \boldsymbol{P}^{(2)} \otimes \boldsymbol{P}^{(1)})(\boldsymbol{U}^{(3)} \otimes \boldsymbol{U}^{(2)} \otimes \boldsymbol{U}^{(1)})vec\mathcal{X}, \tag{29}$$

The eigenvalues of eigenmatrices of Tucker-3 orthogonal decomposition satisfy the following theorem.

Theorem 1. *The eigenvalues of* $\boldsymbol{U} = \boldsymbol{U}^{(1)} \otimes \boldsymbol{U}^{(2)} \otimes \boldsymbol{U}^{(3)}$ *define a semi-order.*

(*Proof*). For the eigenvalues $\lambda_i^{(1)}, \lambda_j^{(2)}, \lambda_k^{(3)}$ of the 1-, 2- and 3-modes of tensors, the inequalities $\lambda_i^{(1)}\lambda_j^{(2)}\lambda_k^{(3)} \geq \lambda_i^{(1)}\lambda_j^{(2)}\lambda_{k+1}^{(3)}$, $\lambda_i^{(1)}\lambda_j^{(2)}\lambda_k^{(3)} \geq \lambda_i^{(1)}\lambda_{j+1}^{(2)}\lambda_k^{(3)}$, $\lambda_i^{(1)}\lambda_j^{(2)}\lambda_k^{(3)} \geq \lambda_{i+1}^{(1)}\lambda_j^{(2)}\lambda_k^{(3)}$ define the semi-orders among eigenvalues as

$$\lambda_i^{(1)}\lambda_j^{(2)}\lambda_k^{(3)} \succeq \left\langle \lambda_i^{(1)}\lambda_j^{(2)}\lambda_{k+1}^{(3)}, \lambda_i^{(1)}\lambda_{j+1}^{(2)}\lambda_k^{(3)}, \lambda_{i+1}^{(1)}\lambda_j^{(2)}\lambda_k^{(3)} \right\rangle$$

□

Regarding the selection of the dimension of tensor subspace, Theorem 1 implies the following theorem.

Theorem 2. *The dimension of the subspace of the tensor space for data compression is* $\frac{1}{6}n(n+1)(n+2)$ *if we select* n *principal components in each mode of three-way array data.*

(*Proof*). For a positive integer n, the number s_n of eigenvalues $\lambda_i^{(1)}\lambda_j^{(2)}\lambda_k^{(3)}$ is

$$\begin{aligned} s_n &= \sum_{i+j+k=0, i,j,k \geq 0}^{n-1} (i+j+k) = \sum_{l=1}^{n}\sum_{m=1}^{l} m \\ &= \frac{1}{2}\left(\frac{1}{6}n(n+1)(2n+1) + \frac{1}{2}n(n+1)\right) = \frac{1}{6}n(n+1)(n+2) \end{aligned} \tag{30}$$

(*Q.E.D*)

If $n = 1, 2, 3, 4$, we have $N = 1, 4, 10, 20$, respectively, for tensors $mathcalX = ((x_{ijk}))$ in $\mathbb{R}^{I\times I\times I}$.

Since the discrete cosine transform (DCT) [4] is asymptotically equivalent to the matrix of K-L transform [13] for data observed from a first-order Markov model [12,13], the dimension reduction by PCA is performed using DCT as

$$f^n_{ijk} = \sum_{i'j'k'=0}^{n-1} g_{i'j'k'}\varphi_{i'i}\varphi_{j'j}\varphi_{k'k}, \;\; g_{ijk} = \sum_{i'j'k'=0}^{N-1} f_{i'j'k'}\varphi_{ii'}\varphi_{jj'}\varphi_{kk'} \tag{31}$$

for $n \leq N$, where

$$\boldsymbol{\Phi}_{(N)} = ((\epsilon \cos\frac{(2j+1)i}{2\pi N})) = ((\varphi_{ij})), \;\; \epsilon = \begin{cases} 1 & \text{if } j = 0 \\ \frac{1}{\sqrt{2}} & \text{otherwise} \end{cases} \tag{32}$$

is the DCT-II matrix of the order N. If we apply the fast cosine transform to the computation of the 3D-DCT-II matrix, the computational complexity is $\mathcal{O}(3n\log n)$.

Since

$$vec(\boldsymbol{u}\circ\boldsymbol{v}\circ\boldsymbol{w}) = \boldsymbol{u}\otimes\boldsymbol{v}\otimes\boldsymbol{w} \tag{33}$$

outer products of vectors redescribes the DCT-based transform as

$$\mathcal{F} = \sum_{i,j,k=0}^{N-1} a_{ijk}\boldsymbol{\varphi}_i\circ\boldsymbol{\varphi}_j\circ\boldsymbol{\varphi}_k, \;\; a_{ijk} = \langle\mathcal{F}, (\boldsymbol{\varphi}_i\circ\boldsymbol{\varphi}_j\circ\boldsymbol{\varphi}_k)\rangle \tag{34}$$

where

$$\boldsymbol{\Phi}^{(N)} = (\boldsymbol{\varphi}_0, \boldsymbol{\varphi}_1, \cdots, \boldsymbol{\varphi}_{N-1}). \tag{35}$$

4 Mutual Tensor Subspace Method

The mutual subspace method is stable against geometric perturbation of queries for pattern recognition. Furthermore, the constrained subspace method confirms robust recognition against global perturbations of queries. Therefore, as the combination of two methods, the constrained mutual subspace method is stable and robust both against geometric and global perturbations of queries.

The angle $\theta = \angle(\mathcal{A}, \mathcal{B})$ between two tensor $\mathcal{A}$ and $\mathcal{B}$ is computed as $\cos\theta = \frac{\langle\mathcal{A},\mathcal{B}\rangle}{|\mathcal{A}|_F|\mathcal{B}|_F}$. The angle between two spaces defined by $\{\boldsymbol{U}^{(i)}, \boldsymbol{V}^{(i)}\}_{i=1}^3$ is the extremal of the criterion

$$\cos^2\theta = |\langle\mathcal{A}\times_1\boldsymbol{U}^{(1)}\times_2\boldsymbol{U}^{(2)}\times_3\boldsymbol{U}^{(3)}, \mathcal{B}\times_1\boldsymbol{V}^{(1)}\times_2\boldsymbol{V}^{(2)}\times_3\boldsymbol{V}^{(3)}\rangle|^2 \tag{36}$$

with respect to the conditions $|\mathcal{A}|_F = 1$ and $|\mathcal{B}|_F = 1$. Therefore, the minimiser of Eq. (36) is the solution of the Euler-Lagrange equation

$$T_1 = \cos^2\theta + \lambda(1 - |\mathcal{A}|^2_F) + \mu(1 - |\mathcal{B}|^2_F) \tag{37}$$

Since, we have the system of equations

$$\mathcal{A} \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \times_3 \boldsymbol{U}^{(3)} = \mu\mathcal{B} \tag{38}$$

$$\mathcal{B} \times_1 \boldsymbol{V}^{(1)} \times_2 \boldsymbol{V}^{(2)} \times_3 \boldsymbol{V}^{(3)} = \lambda\mathcal{A}, \tag{39}$$

for the tensor singular value problem

$$\mathcal{A} \times_1 \boldsymbol{P}_1 \times_2 \boldsymbol{P}_2 \times_3 \boldsymbol{P}_3 = \lambda\mu\mathcal{A} \tag{40}$$

$$\mathcal{B} \times_1 \boldsymbol{Q}_1 \times_2 \boldsymbol{Q}_2 \times_3 \boldsymbol{Q}_3 = \lambda\mu\mathcal{B}, \tag{41}$$

where $\boldsymbol{P}_i = \boldsymbol{U}^{(1)}\boldsymbol{V}^{(i)}$ and $\boldsymbol{Q}_i = \boldsymbol{V}^{(i)}\boldsymbol{U}^{(i)}$, the maximiser of Eq. (36) is

$$T_{\max} = \lambda_{\max}\mu_{\max} \tag{42}$$

with the condition $\lambda = \mu$. This mathematical property implies the following theorem.

Theorem 3. *The canonical angle between a pair of linear subspaces spanned by triples of tensors $\{\boldsymbol{U}^{(i)}\}_{i=1}^3$ and $\{\boldsymbol{V}^{(i)}\}_{i=1}^3$ is $\cos^{-1}\sigma$ where σ is the maximum eigenvalue of tensor $\boldsymbol{P}_1 \times_2 \boldsymbol{P}_2 \times_3 \boldsymbol{P}_3$.*

To shrink the size of linear problem, we evaluate the difference between two subspaces using the perpendicular length of the normalised tensors.

For a collection of tensors $\{\mathcal{X}_i\}_{i=1}^M$ with the condition and $\mathrm{E}(\mathcal{X}_i) = 0$, we define a collection of categories of volumetric data. We assume that we have N_{C} categories For the kth category, we compute a system of orthogonal projections $\{\boldsymbol{U}_{k,j}\}_{j=1,k=1}^{3,N_c}$. to define tensor subspaces, that is, a tensor subspace corresponds to a category of volumetric data.

Setting $\{\mathcal{G}_{i'}\}_{i'=1}^{M'}$ to be queries, we compute projected tensor

$$\mathcal{A}_{i'} = \mathcal{G}_{i'} \times_1 \boldsymbol{U}_k^{(1)\top} \times_2 \boldsymbol{U}_k^{(2)\top} \times_3 \boldsymbol{U}_k^{(3)\top}. \tag{43}$$

Furthermore, assuming that queries belong a category from N_{C} categories, we have orthogonal projection matrices $\{\boldsymbol{V}_j\}_{j=1}^3$, from a tensor subspace of queries. This system of orthogonal projections yields the projected tensor

$$\mathcal{B}_{i'} = \mathcal{G}_{i'} \times_1 \boldsymbol{V}^{(1)\top} \times_2 \boldsymbol{V}^{(2)\top} \times_3 \boldsymbol{V}^{(3)\top}. \tag{44}$$

Using a tensor subspace $\mathcal{C}_k$ corresponding to a category and a tensor subspace $\mathcal{C}_{\mathrm{q}}$ computed from queries, we define the dissimilarity of subspaces $d(\mathcal{C}_k, \mathcal{C}_{\mathrm{q}})$ by

$$\mathrm{E}\left(|\mathcal{A}_{i'} \times_1 \boldsymbol{P}\boldsymbol{U}_k^{(1)} \times_2 \boldsymbol{P}\boldsymbol{U}_k^{(2)} \times_3 \boldsymbol{P}\boldsymbol{U}_k^{(3)} - \mathcal{B}_{i'} \times_1 \boldsymbol{P}\boldsymbol{V}^{(1)} \times_2 \boldsymbol{P}\boldsymbol{V}^{(2)} \times_3 \boldsymbol{P}\boldsymbol{V}^{(3)}|_{\mathrm{F}}^2\right), \tag{45}$$

for $\mathcal{A}_{i'} := \mathcal{A}_i/|\mathcal{A}_i|_F$ and $\mathcal{B}_{i'} := \mathcal{B}_i/|\mathcal{B}_i|_F$, where a unitary matrix $\boldsymbol{P}$ selects bases for each mode of tensors. For the dissimilarity of (45), the condition

$$\arg\left(\min_l d(\mathcal{C}_l, \mathcal{C}_{\mathrm{q}})\right) = \mathcal{C}_k, \tag{46}$$

leads to the property that $\{\mathcal{G}_{i'}\}_{i'=1}^{M'} \in \mathcal{C}_k(\delta)$ for $k, l = 1, 2, \ldots, N_{\mathrm{C}}$.

For the construction of the common space spanned by $\{\boldsymbol{P}_1^{(j)} \otimes \boldsymbol{P}_2^{(j)} \otimes \boldsymbol{P}_2^{(j)}\}_{j=1}^{N_C}$, the orthogonal projection $\boldsymbol{Q} = \boldsymbol{Q}^{(1)} \otimes \boldsymbol{Q}^{(2)} \otimes \boldsymbol{Q}^{(3)}$ which minimises criterion

$$J_{\mathrm{CMS}_1} = \sum_{i=1}^{N_C} |\boldsymbol{Q}\boldsymbol{P}_i|_F^2 \tag{47}$$

with the conditions

$$\boldsymbol{P}_i = \boldsymbol{P}_i^{(1)} \otimes \boldsymbol{P}_i^{(2)} \otimes \boldsymbol{P}_i^{(3)} \tag{48}$$

and

$$\boldsymbol{Q}_i^{(j)\top} \boldsymbol{Q}_i^{(j)} = \boldsymbol{I}_{n_i}, \ \ j = 1, 2, 3. \tag{49}$$

Therefore, we minimise the Euler-Lagrange equation

$$J_{\mathrm{CMS}_2} = \sum_{i=1}^{N_C} |\boldsymbol{Q}\boldsymbol{P}_i|_F^2 + \sum_{j=1}^{3} \langle \boldsymbol{\Lambda}^{(j)} (\boldsymbol{I}_{m_j} - \boldsymbol{Q}^{(j)^\top} \boldsymbol{Q}^{(j)}) \rangle \tag{50}$$

Same with the minimisation for the tensor PCA, we solve the system of optimisation problems.

In each mode, the orthogonal projection $\boldsymbol{Q}^{(j)}$, which maximises the criterion

$$J_{\mathrm{CMSM}} = \sum_{i=1}^{N_j} |\boldsymbol{Q}^{(j)} \boldsymbol{U}_i^{(j)}|_F^2, \tag{51}$$

for the collection of orthogonal projections $\{\boldsymbol{U}_i^{(j)}\}_{i=1}^{N_j}$, approximates the common space of spaces spanned by $\{\boldsymbol{U}_i^{(j)}\}_{i=1}^{N_j}$.

The projection $\boldsymbol{Q}^{(j)}$ is the solution of the variational problem

$$J_{\mathrm{CMSM\text{-}EL}} = \sum_{i=1}^{N} |\boldsymbol{Q}\boldsymbol{U}_i^{(j)}|_F^2, + \langle (\boldsymbol{I}_{m_j} - \boldsymbol{Q}^{(j)\top} \boldsymbol{Q}^{(j)}), \boldsymbol{\Sigma} \rangle. \tag{52}$$

Therefore, the $\boldsymbol{Q}^{(j)}$ is the eigenmatrix of

$$\sum_{i=1}^{N_j} \boldsymbol{U}_i^{(j)} \boldsymbol{Q}^{(j)} = \boldsymbol{Q}^{(j)} \boldsymbol{\Sigma}^{(j)}. \tag{53}$$

Let $\{\mathcal{X}_i^j\}_{i=1}^{j(n)}$ be elements of category $\mathcal{C}_j$ and For $\Pi_i \in 2^N$, where $N = \{0, 1, 2, \cdots, N-1\}$, we define the minimisation criterion

$$J(\Pi_k, \Pi_m, \Pi_n) = E_i |\mathcal{X}_i^j - \mathcal{X}_i^j \times_1 \boldsymbol{P}_{\Pi_k} \Phi_{(N)} \times_2 \boldsymbol{P}_{\Pi_m} \Phi_{(N)} \times_3 \boldsymbol{P}_{\Pi_n} \Phi_{(N)}|_F^2 \tag{54}$$

for the selection of the ranges of orthogonal projections $\boldsymbol{P}_{\Pi_k}$, $\boldsymbol{P}_{\Pi_m}$ and $\boldsymbol{P}_{\Pi_n}$. The minimisation of Eq. (54) defines the base of the category of $\mathcal{C}_i$ as

$$\mathcal{L}(\mathcal{C}_i) = \{\boldsymbol{\varphi}_p\}_{p \in \Pi_k} \otimes \{\boldsymbol{\varphi}_q\}_{q \in \Pi_m} \otimes \{\boldsymbol{\varphi}_r\}_{r \in \Pi_n} \tag{55}$$

using row vectors of DCT-II matrix.

Setting

$$\bar{F}^j_{pqr} = E_i(F^j_{ipqr}), \tag{56}$$

where F^j_{pqr} is the DCT coefficient of f^j_{ipqr}, the ranges are selected

$$\Pi_k \times \Pi_m \times \Pi_n = \{(p, q, r) \mid \bar{F}^j_{pqr} \geq T\} \tag{57}$$

for an appropriate threshold T. Since the low frequency parts are common for almost all data, the canonicalisation of patterns is achieved by separating data to low and high frequency parts. The separation of the constrain subspace $\cap_{i=1}^n \mathcal{C}_i$ for $\{\mathcal{C}_i\}_{i=1}^n$ is achieved by estimating common low frequency parts to categories.

5 Conclusions

We have reviewed several fundamental and well-established methodologies in pattern recognition in vector space to unify data procession for higher-dimensional space using tensor expressions and multi-way principal component analysis.

In traditional method in medical image analysis outline shapes of objects such as organs and statistical properties of interior textures are independently extracted using separate methods. However, the tensor PCA for volumetric data allows us to simultaneously extract both outline shapes of volumetric objects and statistical properties of interior textures of volumetric data from data projected onto a low-dimensional linear subspace spanned by tensors. We also showed frequency-based pattern recognition methodologies for tensor subspace method.

This research was supported by the “Multidisciplinary Computational Anatomy and Its Application to Highly Intelligent Diagnosis and Therapy” project funded by a Grant-in-Aid for Scientific Research on Innovative Areas from MEXT, Japan, and by Grants-in-Aid for Scientific Research funded by the Japan Society for the Promotion of Science.

References

1. Iijima, T.: Pattern Recognition. Corona-sha, Tokyo (1974). (in Japanese)
2. Otsu, N.: Mathematical studies on feature extraction in pattern recognition. Res. Electrotech. Lab. **818**, 1–220 (1981). (in Japanese)
3. Grenander, U., Miller, M.: Pattern theory: from representation to inference. In: OUP (2007)
4. Strang, G., Nguyen, T.: Wavelets and Filter Banks. Wellesley-Cambridge Press, Wellesley (1996)
5. Marron, J.M., Alonso, A.M.: Overview of object oriented data analysis. Biom. J. **56**, 732–753 (2014)
6. Cichocki, A., Zdunek, R., Phan, A.-H., Amari, S.: Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation. Wiley, Hoboken (2009)
7. Itskov, M.: Tensor Algebra and Tensor Analysis for Engineers. Springer, Berlin (2013)

8. Mørup, M.: Applications of tensor (multiway array) factorizations and decompositions in data mining. Wiley Interdiscipl. Rev.: Data Min. Knowl. Disc. **1**, 24–40 (2011)
9. Malcev, A.: Foundations of Linear Algebra, in Russian edition 1948, (English translation W.H, Freeman and Company (1963)
10. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**, 455–500 (2008)
11. Kroonenberg, P.M.: Applied Multiway Data Analysis. Wiley, Hoboken (2008)
12. Hamidi, M., Pearl, J.: Comparison of the cosine Fourier transform of Markov-1 signals. IEEE ASSP **24**, 428–429 (1976)
13. Oja, E.: Subspace Methods of Pattern Recognition. Research Studies Press, Letchworth (1983)