Clustering-Aware Structure-Constrained Low-Rank Submodule Clustering

1st Tong Wu
Department of Electrical and Computer Engineering
Rutgers University
Piscataway, NJ, USA
Email: tong.wu.ee@rutgers.edu

***Abstract*—In this paper, a new clustering algorithm for two-dimensional data is proposed. Unlike most conventional clustering methods which are derived for dealing with matrices, the proposed algorithm performs clustering in a third-order tensor space. The images are then assumed to be drawn from a mixture of low-dimensional free submodules. The proposed method integrates spectral clustering into the optimization problem, thereby overcomes the shortcomings of existing techniques by its ability to perform optimal clustering of the submodules. An efficient algorithm via a combination of an alternating direction method of multipliers and spectral clustering is developed to find the representation tensor and the segmentation simultaneously. The effectiveness of the proposed method is demonstrated through experiments on three image datasets.**

***Index Terms*—Clustering, union of free submodules**

I. Introduction

Clustering of two-dimensional data, e.g., images, has attracted great interest in the past decade. One of the models that has shown remarkable performance is the union of subspaces (UoS) model, which assumes imaging data lie near a union of low-dimensional subspaces. Therefore, the task, known as subspace clustering, is to segment the data into their respective subspaces. This problem has numerous applications in computer vision and image processing [1]–[8].

In traditional subspace clustering literature, imaging data are always first flattened into vectors, which are then fed into a learning algorithm designed for vectorial data. Such approaches can result in poor performance because they neglect the multidimensional structure of data. Recently, there have been several algorithms proposed to improve the performance of many tasks by exploiting the tensor representation of imaging data [9]–[13]. In particular, motivated by t-product [14], [15]—a novel algebraic approach based on circular convolution, a third-order tensor can be regarded as a "matrix" whose elements are tube fibers. Subsequently, one can arrange images as lateral slices to make a third-order tensor and the data tensor can be represented in a self-expressive manner using t-product. The pairwise affinities between the images can then be built from the representation coefficients, which is pipelined into the spectral clustering [16] to obtain final clustering results. In [10], a sparsity constraint is imposed on the coefficient tensor, while [12] requires the coefficient tensor to be both structured and low rank.

Our contributions: Although the approaches proposed in [10], [12] have achieved promising results, a common shortcoming is that they divide the problem into two separate stages: affinity learning and spectral clustering. The relationship between the coefficient tensor and the segmentation of data is not explicitly captured, which can lead to sub-optimal results. In this paper, we propose a *clustering-aware structure-constrained low-rank submodule clustering* (CSLRSmC), which attempts to integrate the two separate steps, i.e., representation tensor learning and spectral clustering, into a unified optimization framework. To be specific, we add a segmentation dependent term into the optimization problem, in which the representation tensor and the submodule segmentation are learned jointly. Experimental results on real-world image datasets show that our proposed method outperforms the state-of-the-art approaches.

Organization: The rest of this paper is organized as follows. Section II gives the notations and preliminaries used throughout this paper. We mathematically formulate the CSLRSmC problem in Section III and present our algorithm in Section IV. In Section V, we describe the results of numerical experiments that validate the effectiveness of our proposed method. We finally conclude in Section VI with some remarks.

II. Technical Background

A. Notation and Definitions

We utilize calligraphy letters for tensors, for example, $\mathcal{A}$, bold lowercase letters for vectors, for example, $\mathbf{a}$, bold uppercase letters for matrices, for example, $\mathbf{A}$, and non-bold letters for scalars, for example, a. The vector of all ones and the identity matrix are denoted by $\mathbf{1}$ and $\mathbf{I}$ of appropriate dimensions, respectively. For a matrix $\mathbf{A}$, its (i, j)-th element is denoted by $a_{i,j}$. The i-th row and j-th column of a matrix $\mathbf{A}$ are denoted by $\mathbf{a}^i$ and $\mathbf{a}_j$, respectively. The ℓ_1 norm of $\mathbf{A}$ is denoted by $\|\mathbf{A}\|_1 = \sum_{i,j} |a_{i,j}|$. We use $(\cdot)^T$ and $\mathrm{tr}(\cdot)$ to denote transpose and trace operations, respectively. For a third-order tensor $\mathcal{A}$, $a_{i,j,k}$ denotes its (i, j, k)-th element. We use MATLAB notation to denote the elements in tensors. Specifically, we use $\mathcal{A}(i, :, :)$, $\mathcal{A}(:, i, :)$ and $\mathcal{A}(:, :, i)$ to denote the i-th horizontal, lateral and frontal slices, respectively; and $\mathcal{A}(:, i, j)$, $\mathcal{A}(i, :, j)$ and $\mathcal{A}(i, j, :)$ to denote the (i, j)-th mode-1, mode-2 and mode-3 fibers, respectively. In particular, $\mathbf{A}^{(i)}$ is also used to represent $\mathcal{A}(:, :, i)$. We use $\widehat{\mathcal{A}} = \mathrm{fft}(\mathcal{A}, [\ ], 3)$

to denote the Fourier Transform along the third dimension of a tensor $\mathcal{A}$. Similarly, one can compute $\mathcal{A}$ from $\widehat{\mathcal{A}}$ via ifft($\widehat{\mathcal{A}}, [\,], 3$) using the inverse Fourier transform along mode-3 of $\widehat{\mathcal{A}}$. The inner product of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i,j,k} a_{i,j,k} b_{i,j,k}$. The Frobenius and infinity norms of a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ are defined as $\|\mathcal{A}\|_F = \sqrt{\sum_{i,j,k} a_{i,j,k}^2}$ and $\|\mathcal{A}\|_\infty = \max_{i,j,k} |a_{i,j,k}|$, respectively.

Definition 1 (t-product [15])**.** *The t-product between $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ is an $n_1 \times n_4 \times n_3$ tensor $\mathcal{C}$ whose (i,p)-th tube $\mathcal{C}(i,p,:)$ is given by*

$$\mathcal{C}(i,p,:) = \sum_{j=1}^{n_2} \mathcal{A}(i,j,:) \circ \mathcal{B}(j,p,:), \tag{1}$$

where $i = 1, \ldots, n_1$, $p = 1, \ldots, n_4$, and $\circ$ denotes circular convolution between two tubes of the same size. The t-product in the original domain corresponds to the matrix multiplication of the frontal slices in the Fourier domain.

Definition 2 (Identity tensor)**.** *The identity tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is defined as follows,*

$$\mathcal{I}(:,:,0) = \mathbf{I}_n, \quad \mathcal{I}(:,:,k) = \mathbf{0}_n, \quad k = 2, 3, \ldots, n_3, \tag{2}$$

where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix and $\mathbf{0}_n$ denotes the zero matrix of size $n \times n$.

Definition 3 (Tensor transpose)**.** *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the transpose tensor $\mathcal{A}^T$ is an $n_2 \times n_1 \times n_3$ tensor obtained by transposing each frontal slice of $\mathcal{A}$ and then reversing the order of the transposed frontal slices 2 through n_3.*

Definition 4 (Orthogonal tensor)**.** *A tensor $\mathcal{Q} \in \mathbb{R}^{n \times n \times n_3}$ is orthogonal if it satisfies*

$$\mathcal{Q} * \mathcal{Q}^T = \mathcal{Q}^T * \mathcal{Q} = \mathcal{I}. \tag{3}$$

Definition 5 (f-diagonal tensor)**.** *A tensor $\mathcal{A}$ is called f-diagonal if each frontal slice $\mathbf{A}^{(k)}$ is a diagonal matrix.*

Definition 6 (t-SVD [14])**.** *The t-SVD of a third-order tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is given by*

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T, \tag{4}$$

where $*$ denotes the t-product, $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a rectangular f-diagonal tensor.

One can obtain this decomposition by performing matrix SVDs in the Fourier domain, see Algorithm 1 for details.

Definition 7 (Tensor nuclear norm [9])**.** *The tensor nuclear norm $\|\mathcal{A}\|_\circledast$ of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the sum of singular values of all the frontal slices of $\widehat{\mathcal{A}}$.*

Algorithm 1 t-SVD

Input: $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$.

1: $\widehat{\mathcal{A}} = \text{fft}(\mathcal{A}, [\,], 3)$.
2: **for** $k = 1, \ldots, n_3$ **do**
3: $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{SVD}(\widehat{\mathcal{A}}(:,:,k))$.
4: $\widehat{\mathcal{U}}(:,:,k) = \mathbf{U}$, $\widehat{\mathcal{S}}(:,:,k) = \mathbf{S}$, $\widehat{\mathcal{V}}(:,:,k) = \mathbf{V}$.
5: **end for**
6: $\mathcal{U} = \text{ifft}(\widehat{\mathcal{U}}, [\,], 3)$, $\mathcal{S} = \text{ifft}(\widehat{\mathcal{S}}, [\,], 3)$, $\mathcal{V} = \text{ifft}(\widehat{\mathcal{V}}, [\,], 3)$.

Output: $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ such that $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$.

B. Linear Algebra with t-product

Given an image with a size of $n_1 \times n_3$, one can *twist* it into a third-order tensor of size $n_1 \times 1 \times n_3$. The set of all tubes in $\mathbb{R}^{1 \times 1 \times n_3}$ forms a ring $\mathbb{K}_{n_3}$ under standard tensor addition and the t-product [14]. Let $\mathbb{K}_{n_3}^{n_1}$ be the set of all $n_1 \times 1 \times n_3$ lateral slices. In fact, the tensor with a size of $n_1 \times 1 \times n_3$ can be regarded as a vector of length n_1, where each element is a $1 \times 1 \times n_3$ tube fiber. It then follows that $\mathbb{K}_{n_3}^{n_1}$ forms a *free module* of dimension n_1 over the ring $\mathbb{K}_{n_3}$ [12], [17]. As discussed in [10], data may be generated from shifted copies of the generating set using t-product, and this is what distinguishes t-linear combinations from linear combinations. Now suppose the image samples are drawn from a union of low-dimensional free submodules, where a free submodule is a subset of $\mathbb{K}_{n_3}^{n_1}$ with a generating set of dimension $s < n_1$ [10], [12]. Our goal is to identify these free submodules and group the data into their respective clusters.

III. Problem Formulation

In this section, we give a brief review of structure-constrained low-rank submodule clustering (SCLRSmC) [12], and mathematically formulate the problem studied in this paper. Consider a collection of images $\mathbb{X} = \{\mathbf{X}_j\}_{j=1}^N$ that belong to L different clusters, where each $\mathbf{X}_j \in \mathbb{R}^{n_1 \times n_3}$ and N represents the number of image samples. Different from typical subspace clustering approaches that first vectorize $\mathbf{X}_j$'s into $\mathbf{x}_j \in \mathbb{R}^m$, $m = n_1 n_3$, and then assume all these vectors are drawn from a union of L subspaces in $\mathbb{R}^m$, SCLRSmC keeps the samples as matrices and arranges $\mathbf{X}_j$'s as lateral slices of a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times N \times n_3}$. It is then assumed that the images belong to a union of L free submodules $\{{}^\ell\mathcal{S}_{n_3}^{n_1}\}_{\ell=1}^L$ of dimensions $\{s_\ell < n_1\}_{\ell=1}^L$. The task of *submodule clustering* is to segment the sample set $\mathbb{X}$ according to the underlying submodules. SCLRSmC exploits the fact that images under the union-of-free-submodules (UoFS) model are self-expressive. In other words, an image belonging to a union of free submodules can be expressed as a t-linear combination of other images. SCLRSmC seeks a low-rank representation tensor $\mathcal{Z}$ by solving the following optimization problem:

$$\min_{\mathcal{Z}} \|\mathcal{Z}\|_\circledast + \lambda_1 \sum_{k=1}^{n_3} \|\mathbf{B} \odot \mathbf{Z}^{(k)}\|_1 + \lambda_2 \|\mathcal{X} - \mathcal{X} * \mathcal{Z}\|_F^2, \tag{5}$$

where λ_1 and λ_2 are penalty parameters, $\mathcal{Z} \in \mathbb{R}^{N \times N \times n_3}$ is the representation tensor, $\mathbf{B} \in \mathbb{R}^{N \times N}$ is a predefined weight matrix associated with the data, and $\odot$ denotes the Hadamard product. In general, the matrix $\mathbf{B}$ imposes the block-diagonal structure on each frontal slice of the representation tensor by penalizing affinities between images from different clusters, while rewarding affinities between images from the same cluster.

After obtaining an optimal representation tensor $\mathcal{Z}^*$, one can define an affinity matrix $\mathbf{W}$ as $w_{i,j} = \|\mathcal{Z}^*(i,j,:)\|_F + \|\mathcal{Z}^*(j,i,:)\|_F$ and use spectral clustering [16] to obtain final clustering, which solves the following problem:

$$\min_{\mathbf{F}} \operatorname{tr}(\mathbf{F}^T(\mathbf{H} - \mathbf{W})\mathbf{F}) \quad \text{s.t.} \quad \mathbf{F} \in \mathbb{P}, \tag{6}$$

where $\mathbf{F} \in \mathbb{R}^{N \times L}$ is a binary matrix indicating the membership of the data points to each submodule. That is, $f_{i,\ell} = 1$ if $\mathcal{X}(:,i,:)$, i.e., $\mathbf{X}_i$, belongs to submodule ${}^{\ell}\mathcal{S}_{n_3}^{n_1}$ and $f_{i,\ell} = 0$ otherwise. Here, $\mathbf{H} \in \mathbb{R}^{N \times N}$ is a diagonal matrix with $h_{i,i} = \sum_j w_{i,j}$ and $\mathbb{P} = \{\mathbf{F} \in \{0,1\}^{N \times L} : \mathbf{F}\mathbf{1} = \mathbf{1}, \operatorname{rank}(\mathbf{F}) = L\}$ is the space of valid segmentation matrices with L clusters.

Existing submodule clustering methods use spectral clustering as a post-processing step, which may lead to sub-optimal results. Our challenge is to define a metric that can quantify the disagreement between the coefficient tensor $\mathcal{Z}$ and the segmentation matrix $\mathbf{F}$. First, notice that SCLRSmC imposes a weighted sparsity constraint on each frontal slice of the coefficient tensor $\mathcal{Z}$. In order to make our final algorithm tractable, it is reasonable to redefine the affinity matrix $\mathbf{W}$ whose (i,j)-th entry is given by

$$w_{i,j} = \frac{1}{2}(\|\mathcal{Z}(i,j,:)\|_1 + \|\mathcal{Z}(j,i,:)\|_1), \tag{7}$$

where $\|\mathcal{Z}(i,j,:)\|_1 = \sum_{k=1}^{n_3} |z_{i,j,k}|$. For the purpose of clustering, we expect the coefficient tensor to be *submodule-preserving*, i.e., $\mathcal{Z}(i,j,:) = 0$ if image samples $\mathbf{X}_i$ and $\mathbf{X}_j$ lie in different submodules. The interaction between $\mathcal{Z}$ and $\mathbf{F}$ can then be quantified as

$$\begin{aligned} &\sum_{i,j} \|\mathcal{Z}(i,j,:)\|_1 (\frac{1}{2}\|\mathbf{f}^i - \mathbf{f}^j\|_2^2) = \sum_{k=1}^{n_3} \|\mathbf{\Theta} \odot \mathbf{Z}^{(k)}\|_1 \\ &= \frac{1}{2}\sum_{i,j} w_{i,j}\|\mathbf{f}^i - \mathbf{f}^j\|_2^2 = \operatorname{tr}(\mathbf{F}^T(\mathbf{H} - \mathbf{W})\mathbf{F}), \end{aligned} \tag{8}$$

where $\theta_{i,j} = \frac{1}{2}\|\mathbf{f}^i - \mathbf{f}^j\|_2^2$. By adding (6) into the objective of SCLRSmC, we finally pose the problem of *clustering-aware structure-constrained low-rank submodule clustering* (CSLRSmC) in terms of the following optimization problem:

$$\begin{aligned} \min_{\mathcal{Z},\mathbf{F}} \ & \|\mathcal{Z}\|_{\circledast} + \lambda_1 \sum_{k=1}^{n_3} \|\mathbf{B} \odot \mathbf{Z}^{(k)}\|_1 + \lambda_2 \sum_{k=1}^{n_3} \|\mathbf{\Theta} \odot \mathbf{Z}^{(k)}\|_1 \\ & + \lambda_3 \|\mathcal{X} - \mathcal{X} * \mathcal{Z}\|_F^2 \\ \text{s.t.} \ & \mathbf{F} \in \mathbb{P}, \end{aligned} \tag{9}$$

where λ_1, λ_2 and λ_3 are penalty parameters. Similar to [12], the (i,j)-th entry of $\mathbf{B}$ is defined as $b_{i,j} = 1 - \exp\left(-\frac{1 - |\langle \widetilde{\mathcal{X}}(:,i,:), \widetilde{\mathcal{X}}(:,j,:)\rangle|}{\sigma}\right)$, where $\widetilde{\mathcal{X}}(:,i,:)$ and $\widetilde{\mathcal{X}}(:,j,:)$ are the normalized data points of $\mathcal{X}(:,i,:)$ and $\mathcal{X}(:,j,:)$, respectively, and σ is empirically set as the mean of all $1 - |\langle \widetilde{\mathcal{X}}(:,i,:), \widetilde{\mathcal{X}}(:,j,:)\rangle|$'s. The CSLRSmC encourages consistency between the representation coefficients and the submodule segmentation by making each frontal slice of the coefficient tensor more block-diagonal, which can help spectral clustering achieve the best results.

Algorithm 2 Solving problem (11) using ADMM

Input: Data $\mathcal{X}$, matrix $\mathbf{B}$ and $\mathbf{\Theta}$, and parameters λ_1, λ_2 and λ_3.

Initialize: $\mathcal{C}^{(0)} = \mathcal{Q}^{(0)} = \mathcal{Z}^{(0)} = \mathcal{H}_1^{(0)} = \mathcal{H}_2^{(0)} = 0 \in \mathbb{R}^{N \times N \times n_3}$, $\rho = 1.9$, $\mu^{(0)} = 0.1$, $\mu_{\max} = 10^{10}$, $\epsilon = 10^{-5}$, and $t = 0$.

while not converged **do**
  Update $\mathcal{C}$, $\mathcal{Q}$, and $\mathcal{Z}$.
  $\mathcal{H}_1^{(t+1)} = \mathcal{H}_1^{(t)} + \mu^{(t)}(\mathcal{Z}^{(t+1)} - \mathcal{C}^{(t+1)})$,
  $\mathcal{H}_2^{(t+1)} = \mathcal{H}_2^{(t)} + \mu^{(t)}(\mathcal{Z}^{(t+1)} - \mathcal{Q}^{(t+1)})$.
  Update $\mu^{(t+1)}$ as $\mu^{(t+1)} = \min(\mu_{\max}, \rho\mu^{(t)})$.
  Check convergence conditions
  $$\max \left\{ \begin{matrix} \|\mathcal{Z}^{(t+1)} - \mathcal{C}^{(t+1)}\|_\infty, \|\mathcal{Z}^{(t+1)} - \mathcal{Q}^{(t+1)}\|_\infty \\ \|\mathcal{Z}^{(t+1)} - \mathcal{Z}^{(t)}\|_\infty, \|\mathcal{C}^{(t+1)} - \mathcal{C}^{(t)}\|_\infty \\ \|\mathcal{Q}^{(t+1)} - \mathcal{Q}^{(t)}\|_\infty \end{matrix} \right\} < \epsilon.$$
  Update t by $t = t + 1$.
end while

Output: Representation tensor $\mathcal{Z}^* = \mathcal{Z}^{(t+1)}$.

IV. Optimization

In this section, we present an efficient algorithm for our model (9) by alternatively solving the following two subproblems:

- Fix $\mathbf{F}$, find the representation tensor $\mathcal{Z}$.
- Fix $\mathcal{Z}$, find the clustering indicator matrix $\mathbf{F}$ by spectral clustering.

A. Solution of the Representation Tensor

Given the segmentation matrix $\mathbf{F}$, we first compute the matrix $\mathbf{\Theta}$ with $\theta_{i,j} = \frac{1}{2}\|\mathbf{f}^i - \mathbf{f}^j\|_2^2$. Then we compute $\mathcal{Z}$ by solving the following problem:

$$\begin{aligned} \min_{\mathcal{Z}} \ & \|\mathcal{Z}\|_{\circledast} + \lambda_1 \sum_{k=1}^{n_3} \|\mathbf{B} \odot \mathbf{Z}^{(k)}\|_1 + \lambda_2 \sum_{k=1}^{n_3} \|\mathbf{\Theta} \odot \mathbf{Z}^{(k)}\|_1 \\ & + \lambda_3 \|\mathcal{X} - \mathcal{X} * \mathcal{Z}\|_F^2. \end{aligned} \tag{10}$$

To solve (10), we first introduce auxiliary variables $\mathcal{C}$ and $\mathcal{Q}$ to make (10) separable and convert (10) to the following problem:

$$\begin{aligned} \min_{\mathcal{C},\mathcal{Q},\mathcal{Z}} \ & \|\mathcal{C}\|_{\circledast} + \lambda_1 \sum_{k=1}^{n_3} \|\mathbf{B} \odot \mathbf{Q}^{(k)}\|_1 + \lambda_2 \sum_{k=1}^{n_3} \|\mathbf{\Theta} \odot \mathbf{Q}^{(k)}\|_1 \\ & + \lambda_3 \|\mathcal{X} - \mathcal{X} * \mathcal{Z}\|_F^2 \\ \text{s.t} \ & \mathcal{Z} = \mathcal{C}, \ \mathcal{Z} = \mathcal{Q}. \end{aligned} \tag{11}$$

The constrained problem (11) can now be solved using the Alternating Direction Method of Multipliers (ADMM) [18]. The augmented Lagrangian function of (11) is

$$\begin{aligned}\mathcal{L}(\mathcal{C},\mathcal{Q},\mathcal{Z},\mathcal{H}_1,\mathcal{H}_2,\mu) &= \|\mathcal{C}\|_{\circledast} + \lambda_1\sum_{k=1}^{n_3}\|\mathbf{B}\odot\mathbf{Q}^{(k)}\|_1 \\ &+ \lambda_2\sum_{k=1}^{n_3}\|\mathbf{\Theta}\odot\mathbf{Q}^{(k)}\|_1 + \lambda_3\|\mathcal{X}-\mathcal{X}*\mathcal{Z}\|_F^2 + \langle\mathcal{H}_1,\mathcal{Z}-\mathcal{C}\rangle \\ &+ \langle\mathcal{H}_2,\mathcal{Z}-\mathcal{Q}\rangle + \frac{\mu}{2}(\|\mathcal{Z}-\mathcal{C}\|_F^2 + \|\mathcal{Z}-\mathcal{Q}\|_F^2), \qquad (12)\end{aligned}$$

where the tensors $\mathcal{H}_1$ and $\mathcal{H}_2$ comprise Lagrange multipliers and $\mu > 0$ is a penalty parameter. The ADMM optimizes (12) iteratively by updating $\mathcal{C}$, $\mathcal{Q}$ and $\mathcal{Z}$ one at a time while fixing the others.

Updating $\mathcal{C}$: By fixing other variables in (12), we can solve $\mathcal{C}$ through

$$\mathcal{C}^{(t+1)} = \arg\min_{\mathcal{C}} \|\mathcal{C}\|_{\circledast} + \frac{\mu^{(t)}}{2}\|\mathcal{C} - (\mathcal{Z}^{(t)} + \frac{\mathcal{H}_1^{(t)}}{\mu^{(t)}})\|_F^2. \quad (13)$$

This subproblem can be solved by using [13, Theorem 4.2].

Updating $\mathcal{Q}$: Keeping other tensors invariant, minimization of (12) with respect to $\mathcal{Q}$ yields the following update:

$$\begin{aligned}\mathcal{Q}^{(t+1)} &= \arg\min_{\mathcal{Q}} \lambda_1\sum_{k=1}^{n_3}\|\mathbf{B}\odot\mathbf{Q}^{(k)}\|_1 + \lambda_2\sum_{k=1}^{n_3}\|\mathbf{\Theta}\odot\mathbf{Q}^{(k)}\|_1 \\ &+ \langle\mathcal{H}_2^{(t)},\mathcal{Z}^{(t)}-\mathcal{Q}\rangle + \frac{\mu^{(t)}}{2}\|\mathcal{Z}^{(t)}-\mathcal{Q}\|_F^2. \qquad (14)\end{aligned}$$

We break (14) into n_3 independent subproblems. The update of the k-th frontal slice ${\mathbf{Q}^{(k)}}^{(t+1)}$ of $\mathcal{Q}$ can be written as

$$\begin{aligned}{\mathbf{Q}^{(k)}}^{(t+1)} &= \arg\min_{\mathbf{Q}} \lambda_1\|(\mathbf{B}+\frac{\lambda_2}{\lambda_1}\mathbf{\Theta})\odot\mathbf{Q}\|_1 \\ &+ \frac{\mu^{(t)}}{2}\|\mathbf{Q} - ({\mathbf{Z}^{(k)}}^{(t)} + \frac{{\mathbf{H}_2^{(k)}}^{(t)}}{\mu^{(t)}})\|_F^2, \qquad (15)\end{aligned}$$

which has a closed-form solution given in [4, Proposition 3].

Updating $\mathcal{Z}$: When other tensors are fixed in (12), the following problem of updating $\mathcal{Z}$ need to be solved:

$$\begin{aligned}\mathcal{Z}^{(t+1)} &= \arg\min_{\mathcal{Z}} \lambda_3\|\mathcal{X}-\mathcal{X}*\mathcal{Z}\|_F^2 \\ &+ \frac{\mu^{(t)}}{2}\left(\|\mathcal{Z}-\mathcal{V}_1^{(t+1)}\|_F^2 + \|\mathcal{Z}-\mathcal{V}_2^{(t+1)}\|_F^2\right), \qquad (16)\end{aligned}$$

where $\mathcal{V}_1^{(t+1)} = \mathcal{C}^{(t+1)} - \frac{\mathcal{H}_1^{(t)}}{\mu^{(t)}}$ and $\mathcal{V}_2^{(t+1)} = \mathcal{Q}^{(t+1)} - \frac{\mathcal{H}_2^{(t)}}{\mu^{(t)}}$. As described in [12], this problem can be transformed into the Fourier domain and again decomposed into n_3 subproblems, with the k-th frontal slice of $\widehat{\mathcal{Z}}$ given by

$$\begin{aligned}{\widehat{\mathbf{Z}}^{(k)}}^{(t+1)} &= (2\lambda_3{\widehat{\mathbf{X}}^{(k)}}^T\widehat{\mathbf{X}}^{(k)} + 2\mu^{(t)}\mathbf{I})^{-1} \\ &\left(2\lambda_3{\widehat{\mathbf{X}}^{(k)}}^T\widehat{\mathbf{X}}^{(k)} + \mu^{(t)}({\widehat{\mathbf{V}}_1^{(k)}}^{(t+1)} + {\widehat{\mathbf{V}}_2^{(k)}}^{(t+1)})\right). \qquad (17)\end{aligned}$$

After updating $\widehat{\mathcal{Z}}^{(t+1)}$, one can obtain $\mathcal{Z}^{(t+1)}$ by setting $\mathcal{Z}^{(t+1)} = \text{ifft}(\widehat{\mathcal{Z}}^{(t+1)},[\ ],3)$. The ADMM algorithm for solving problem (11) is summarized in Algorithm 2.

Algorithm 3 CSLRSmC

Input: Data $\mathcal{X}$, matrix $\mathbf{B}$ and $\mathbf{\Theta}$, and parameters L, λ_1, λ_2 and λ_3.

Initialize: $\mathbf{\Theta} = \mathbf{0}$.

1: **while** not converged **do**
2: Given $\mathbf{F}$, solve problem (10) via Algorithm 2 to obtain $\mathcal{Z}$;
3: Given $\mathcal{Z}$, solve problem (19) via spectral clustering to obtain $\mathbf{F}$;
4: **end while**

Output: Segmentation matrix $\mathbf{F}$.

B. Spectral Clustering

When the representation tensor $\mathcal{Z}$ is available, problem (9) reduces to the following problem:

$$\min_{\mathbf{F}} \sum_{i,j}\|\mathcal{Z}(i,j,:)\|_1(\frac{1}{2}\|\mathbf{f}^i - \mathbf{f}^j\|_2^2) \quad \text{s.t.} \quad \mathbf{F}\in\mathbb{P}, \qquad (18)$$

Based on (8), this problem is equivalent to the following problem:

$$\widehat{\mathbf{F}} = \arg\min_{\mathbf{F}} \text{tr}(\mathbf{F}^T\mathbf{\Phi}\mathbf{F}) \quad \text{s.t.} \quad \mathbf{F}\in\mathbb{P}, \qquad (19)$$

where $\mathbf{\Phi} = \mathbf{H} - \mathbf{W}$ is the graph Laplacian of the matrix $\mathbf{W}$ with $w_{i,j} = \frac{1}{2}(\|\mathcal{Z}(i,j,:)\|_1 + \|\mathcal{Z}(j,i,:)\|_1)$, and $\mathbf{H}$ is a diagonal matrix whose diagonal entries are $h_{i,i} = \sum_j w_{i,j}$. To make the problem tractable, we relax the constraint $\mathbf{F}\in\mathbb{P}$ to $\mathbf{F}^T\mathbf{H}\mathbf{F} = \mathbf{I}$ and solve

$$\widehat{\mathbf{F}} = \arg\min_{\mathbf{F}} \text{tr}(\mathbf{F}^T\mathbf{\Phi}\mathbf{F}) \quad \text{s.t.} \quad \mathbf{F}^T\mathbf{H}\mathbf{F} = \mathbf{I}, \qquad (20)$$

which can be solved efficiently by eigendecomposition. Specifically, by letting $\widetilde{\mathbf{F}} = \mathbf{H}^{\frac{1}{2}}\mathbf{F}$, instead of solving $\mathbf{F}$, we solve for $\widetilde{\mathbf{F}}$, whose optimal solution is given by the eigenvectors of the matrix $\mathbf{H}^{-\frac{1}{2}}\mathbf{\Phi}\mathbf{H}^{-\frac{1}{2}}$ associated with its smallest L eigenvalues. The rows of $\widetilde{\mathbf{F}}$ are then used as input to the k-means algorithm and the clustering result is used to generate the binary matrix $\mathbf{F}\in\{0,1\}^{N\times L}$ such that $\mathbf{F}\mathbf{1} = \mathbf{1}$. The complete algorithm is outlined in Algorithm 3.

V. Experimental Results

In this section, we evaluate the performance of our proposed method on MNIST [19], USPS [20] handwritten datasets and Weizmann[1] face dataset. We compare the performance of CSLRSmC with those obtained using several state-of-the-art clustering methods, namely, SCLRSmC [12], SSmC [10], SSC [3], LRR [2], SC-LRR [4], S3C [8], and LSR [1]. Note that in SCLRSmC, the affinity matrix $\mathbf{W}$ is defined in [12], i.e., $w_{i,j} = \|\mathcal{Z}(i,j,:)\|_F + \|\mathcal{Z}(j,i,:)\|_F$, whereas in SCLRSmC† we use the similarity defined in (7). For all these methods, we tune the parameters to achieve their best performance. Since the first iteration of CSLRSmC is equivalent to SCLRSmC†, the parameters λ_1 and λ_3 in CSLRSmC are set to be the same as λ_1 and λ_2 for SCLRSmC†, respectively. Finally, we fix $\lambda_2 = 0.001$ for all experiments.

[1]http://www.wisdom.weizmann.ac.il/~/vision/FaceBase/

TABLE I
CLUSTERING ACCURACY (%) ON DIFFERENT DATASETS

	CSLRSmC	SCLRSmC†	SCLRSmC	SSmC	SSC	LRR	SC-LRR	S3C	LSR
MNIST	**81.33**	79.33	79.87	69.20	51.20	49.33	52.13	50.27	47.73
USPS	**62.20**	59.80	60.20	60.00	38.80	47.40	52.20	27.40	40.80
Weizmann	**89.17**	84.17	58.33	75.00	45.00	56.67	49.17	47.50	35.00

The MNIST dataset [19] comprises a collection of 28×28 images of handwritten digits. Similar to [12], we consider clustering of digits $\{2, 4, 8\}$ and randomly select 100 images of each of these digits; hence $N = 300$. We add a random (left or right) 3-pixel horizontal shift to each image. The results, averaged over 20 random trials, are listed in the second row of Table I, from which we make two observations. First, all the submodule clustering approaches perform better than subspace clustering methods. This is because the UoFS model is robust against spatial shifts. Second, although SCLRSmC† performs slightly worse than SCLRSmC, CSLRSmC performs better than all other approaches.

The USPS dataset [20] consists of 9298 images of 10 subjects, corresponding to 10 handwritten digits. Each image has 16×16 pixels. We use the first 50 images of each digit in our experiment; resulting $N = 500$. We again randomly shift the digits horizontally with respect to the center by 3 pixels either side (left or right). Experimental results are presented in the third row of Table I. It can be inferred that CSLRSmC obtains the best performance compared to other approaches on this dataset.

Our last set of results corresponds to the Weizmann face dataset. We select images of the first 4 individuals and the total number of images is $N = 120$. For each image, we reduce the original size of the image by first downsampling by factor of 4, followed by cropping it into 120×80 pixels. Note that due to variations in pose, the data can be well modeled using UoFS. The clustering results are reported in the last row of Table I. We can see that our proposed method again outperforms all other methods.

VI. CONCLUSION

In this paper, we proposed a novel approach for submodule clustering of third-order tensor data. By using t-product based on circular convolution, the data tensor is represented as the t-product of the tensor itself and a low-rank tensor. The proposed method integrates spectral clustering into a unified framework, and by solving a joint optimization problem, our method is able to find the optimal coefficient tensor which helps spectral clustering achieve the best clustering results. Experimental results on three real-world datasets demonstrated the effectiveness of the proposed method.

REFERENCES

[1] C. Lu, H. Min, Z. Zhao, L. Zhu, D. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," in *Proc. Eur. Conf. Computer Vision (ECCV)*, 2012, pp. 347–360.

[2] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[3] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

[4] K. Tang, R. Liu, Z. Su, and J. Zhang, "Structure-constrained low-rank representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2167–2179, Dec. 2014.

[5] T. Wu and W. U. Bajwa, "Learning the nonlinear geometry of high-dimensional data: Models and algorithms," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6229–6244, Dec. 2015.

[6] T. Wu, P. Gurram, R. M. Rao, and W. U. Bajwa, "Hierarchical union-of-subspaces model for human activity summarization," in *Proc. IEEE Int. Conf. Computer Vision Workshops (ICCVW)*, 2015, pp. 1053–1061.

[7] ——, "Clustering-aware structure-constrained low-rank representation model for learning human action attributes," in *Proc. IEEE Image Video and Multidimensional Signal Process. (IVMSP) Workshop*, 2016, pp. 1–5.

[8] C.-G. Li, C. You, and R. Vidal, "Structured sparse subspace clustering: A joint affinity learning and subspace clustering framework," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2988–3001, Jun. 2017.

[9] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-SVD," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 3842–3849.

[10] E. Kernfeld, S. Aeron, and M. Kilmer, "Clustering multi-way data: A novel algebraic approach," *arXiv preprint*, 2014. [Online]. Available: https://arxiv.org/abs/1412.7056

[11] S. Soltani, M. E. Kilmer, and P. C. Hansen, "A tensor-based dictionary learning approach to tomographic image reconstruction," *BIT Numerical Mathematics*, vol. 56, no. 4, pp. 1425–1454, Dec. 2016.

[12] T. Wu and W. U. Bajwa, "A low tensor-rank representation approach for clustering of imaging data," *IEEE Signal Process. Lett.*, vol. 25, no. 8, pp. 1196–1200, Aug. 2018.

[13] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019.

[14] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 641–658, Aug. 2011.

[15] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, Feb. 2013.

[16] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[17] K. Braman, "Third-order tensors as linear operators on a space of matrices," *Linear Algebra Appl.*, vol. 433, no. 7, pp. 1241–1253, Dec. 2010.

[18] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Advances in Neural Information Processing Systems (NIPS)*, 2011, pp. 612–620.

[19] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[20] J. J. Hull, "A database for handwritten text recognition research," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 550–554, May 1994.