Group Nonnegative Matrix Factorization with Sparse Regularization in Multi-set Data

Xiulin Wang[⋆,†], Wenya Liu[⋆,†], Fengyu Cong[⋆,†], Tapani Ristaniemi[†]
[⋆] *School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China*
[†] *Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland*
xiulin.wang@foxmail.com, wenyaliu0912@foxmail.com, cong@dlut.edu.cn, tapani.e.ristaniemi@jyu.fi

***Abstract*—Constrained joint analysis of data from multiple sources has received widespread attention for that it allows us to explore potential connections and extract meaningful hidden components. In this paper, we formulate a flexible joint source separation model termed as group nonnegative matrix factorization with sparse regularization (GNMF-SR), which aims to jointly analyze the partially coupled multi-set data. In the GNMF-SR model, common and individual patterns of particular underlying factors can be extracted simultaneously with imposing nonnegative constraint and sparse penalty. Alternating optimization and alternating direction method of multipliers (ADMM) are combined to solve the GNMF-SR model. Using the experiment of simulated fMRI-like data, we demonstrate the ADMM-based GNMF-SR algorithm can achieve the better performance.**

***Index Terms*—Alternating direction method of multipliers, coupled, group nonnegative matrix factorization, joint analysis, sparse representation**

I. Introduction

Nonnegative matrix factorization (NMF), providing a part-based representation of nonnegative data, has been widely applied in blind source separation (BSS) problems including signal processing and machine learning [1]–[4]. With increasing availability of sensor technologies, we are now facing a mass of data from multiple sources that need to be jointly separated [5]–[8], such as multi-subject/multi-modal biomedical data [6]–[8]. Although many studies have shown that conventional NMF methods are effective in a large number of single dataset applications, their inefficiency to jointly analyze multiple datasets has limited their broader usage [7]. In order to fill the gap between NMF and group analysis of multiple datasets, group nonnegative matrix factorization (GNMF) was proposed as an update to the standard NMF in multi-set problems [9], [10]. In the group model, coupling information across datasets can be fully exploited, making it possible to achieve higher performance than BSS-based algorithms originally designed for single dataset [5], [7], [9]. Moreover, it is easy to extract the underlying patterns that are common among datasets, as well as individual patterns that exhibit internal variability [8], [9]. Group analysis of multiple datasets can also automatically maintain the alignment of coupled patterns among datasets, while BSS-based algorithms need to adopt some post-aligned strategies such as correlation analysis [5], [9].

Sparse representation aims to encode the data using fewer 'active' components for better interpretation of the encoding [11], [12]. Even though NMF-based algorithms can naturally produce a sparse representation of data, the sparseness of extracted factors is not enough and uncontrollable [11]. Therefore, additional sparse regularization has been widely applied to NMF to promote sparse representation and alleviate factorization non-uniqueness [13]. Inspired by GNMF and sparse NMF, we formulate a flexible group nonnegative matrix factorization with sparse regularization (GNMF-SR) model by imposing an efficient and commonly used regularizer l_1-norm for constrained joint analysis of partially coupled datasets. Obviously, the GNMF works such as [9], [10] did not take the sparse characteristic of latent variables into consideration, and the sparse NMF works in [11]–[13] cannot utilize the coupled information across the datasets. In recent years, the alternating direction method of multipliers (ADMM) has become an effective and popular tool for constrained NMF problems [14]–[17], and in this study we employ ADMM method to optimize the GNMF-SR model. The convergence issue of NMF-based or nonconvex optimization problems about ADMM has been widely discussed in [15]–[19], which will not be discussed in this study. For more details of ADMM method, please refer to the comprehensive review in [17].

The rest of this paper is organized as follows. Section 2 introduces multi-set data model, GNMF model, GNMF-SR model, ADMM method and model optimization via ADMM method. In section 3, simulation experiment on synthetic fMRI-like data is conducted. The last section concludes this paper.

Notations: Scalars, vectors and matrices are respectively denoted by lowercase, boldface lowercase and boldface uppercase, e.g. x, $\boldsymbol{x}$ and $\boldsymbol{X}$. $\mathbb{R}_+$ denotes the nonnegative real number. Operators $(\cdot)^T$, $\|\cdot\|_1$ and $\|\cdot\|_F$ denote transpose, l_1-norm and Frobenius norm, respectively. $\langle \boldsymbol{A}, \boldsymbol{B} \rangle$ denotes the inner product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. $\langle \boldsymbol{A}, \boldsymbol{B} \rangle := \sum_{i,j} a_{ij} b_{ij}$ can be substituted by $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^T)$ for $\boldsymbol{A}$ and $\boldsymbol{B}$ with the same size $I \times J$.

This work was supported by the National Natural Science Foundation of China (Grant Nos. 91748105 & 81471742), the Fundamental Research Funds for the Central Universities [DUT2019] in Dalian University of Technology in China and the scholarships from China Scholarship Council (Nos. 201706060262 and 201706060263). Corresponding author: Fengyu Cong, cong@dlut.edu.cn

II. Methods

In this section, we first introduce the multi-set data model, then we present the GNMF and GNMF-SR models, and last give the ADMM method and the optimization solution of GNMF-SR model.

A. Multi-set data model

Given a set of nonnegative matrices $\boldsymbol{X}^{(s)} \in \mathbb{R}_+^{I^{(s)} \times J^{(s)}}$, $s = 1, 2, \cdots, S$, the multi-set data model assumes that each data $\boldsymbol{X}^{(s)}$ can be expressed by:

$$\boldsymbol{X}^{(s)} \approx \boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)} = [\boldsymbol{A}_C^{(s)} \ \boldsymbol{A}_I^{(s)}]\boldsymbol{B}^{(s)}, \tag{1}$$

where $\boldsymbol{A}^{(s)} \in \mathbb{R}_+^{I^{(s)} \times R^{(s)}}$ and $\boldsymbol{B}^{(s)} \in \mathbb{R}_+^{R^{(s)} \times J^{(s)}}$ represent the latent variable and corresponding coefficient matrix respectively. Generally, $R^{(s)} < \min(I^{(s)}, J^{(s)})$ is assumed for providing a low-rank representation of $\boldsymbol{X}^{(s)}$. Considering that the data are collected under the same condition, it can be reasonably expected that there will be some identical or highly correlated hidden information between the data. Therefore, in multi-set data model, we assume that each factor matrix $\boldsymbol{A}^{(s)} = [\boldsymbol{A}_C^{(s)} \ \boldsymbol{A}_I^{(s)}]$ includes two patterns: $\boldsymbol{A}_C^{(s)} \in \mathbb{R}_+^{I^{(s)} \times L}$, $0 \le L \le R^{(s)}$, a common matrix shared by all S matrices as $\boldsymbol{A}_C^{(1)} = \cdots \boldsymbol{A}_C^{(S)} = \boldsymbol{A}_C$, and $\boldsymbol{A}_I^{(s)} \in \mathbb{R}_+^{I^{(s)} \times (R^{(s)} - L)}$, which corresponds to the individual characteristic in each dataset.

B. Group nonnegative matrix factorization

Considering the coupling structure among latent variables $\boldsymbol{A}^{(s)}$ in multi-set data model, we need to analyze S sets of $\boldsymbol{X}^{(s)}$ simultaneously, which is different from the conventional NMF problem. Using the Euclidean divergence minimization, the GNMF of $\boldsymbol{X}^{(s)}$, $s = 1, 2, \cdots, S$, can be achieved by solving the following optimization:

$$\begin{aligned} &\underset{\boldsymbol{A}^{(s)}, \boldsymbol{B}^{(s)}}{\text{minimize}} \ \frac{1}{2}\sum_{s=1}^{S} \left\| \boldsymbol{X}^{(s)} - \boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)} \right\|_F^2 \\ &\text{subject to} \ \ \boldsymbol{A}^{(s)} \ge 0, \boldsymbol{B}^{(s)} \ge 0. \end{aligned} \tag{2}$$

In many applications, only the underlying patterns in the variable dimension need to be sparse [20]. Combing coupling constraint and sparse penalty on the factor matrix $\boldsymbol{A}^{(s)}$, we formulate a flexible group nonnegative matrix factorization with sparse regulation (GNMF-SR) model as follows:

$$\begin{aligned} &\underset{\boldsymbol{A}^{(s)}, \boldsymbol{B}^{(s)}}{\text{minimize}} \ \frac{1}{2}\sum_{s=1}^{S} \left\| \boldsymbol{X}^{(s)} - \boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)} \right\|_F^2 + \sum_{s=1}^{S} \beta^{(s)} \sum_{r=1}^{R(s)} \left\| \boldsymbol{a}_r^{(s)} \right\|_1 \\ &\text{subject to} \ \ \boldsymbol{A}^{(s)} \ge 0, \boldsymbol{B}^{(s)} \ge 0, \boldsymbol{A}_C^{(1)} = \cdots \boldsymbol{A}_C^{(S)} = \boldsymbol{A}_C, \end{aligned} \tag{3}$$

where $\boldsymbol{a}_r^{(s)}$ corresponds to the rth column of $\boldsymbol{A}^{(s)}$. The penalty term $\sum_{r=1}^{R(s)} \left\| \boldsymbol{a}_r^{(s)} \right\|_1$ is to impose the sparsity on factor matrix $\boldsymbol{A}^{(s)}$, and it can be reformed as $\langle \boldsymbol{E}, \boldsymbol{A}^{(s)} \rangle$, in which $\boldsymbol{E} \in \mathbb{R}_+^{I^{(s)} \times J^{(s)}}$ is a matrix whose entries are all ones. $\beta^{(s)} \ge 0$ is a predefined penalty parameter. For simplicity, we set $\beta^{(1)} = \beta^{(2)} = \cdots \beta^{(S)}$. Later we will give a detailed explanation of how to solve GNMF-SR model using ADMM algorithm.

C. Alternating direction method of multipliers

According to [17], ADMM algorithm considers the following problem:

$$\begin{aligned} &\underset{\boldsymbol{x}, \boldsymbol{z}}{\text{minimize}} \ \ f(\boldsymbol{x}) + g(\boldsymbol{z}) \\ &\text{subject to} \ \ \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} = \boldsymbol{c}. \end{aligned} \tag{4}$$

Using the scaled from, it can be updated iteratively using the following steps:

$$\begin{cases} \boldsymbol{x} := \underset{\boldsymbol{x}}{\text{argmin}} \left(f(\boldsymbol{x}) + (\rho/2) \left\| \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} - \boldsymbol{c} + \boldsymbol{u} \right\|_2^2 \right), \\ \boldsymbol{z} := \underset{\boldsymbol{z}}{\text{argmin}} \left(g(\boldsymbol{z}) + (\rho/2) \left\| \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} - \boldsymbol{c} + \boldsymbol{u} \right\|_2^2 \right), \\ \boldsymbol{u} := \boldsymbol{u} + (\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} - \boldsymbol{c}), \end{cases} \tag{5}$$

where $\boldsymbol{u}$ denote the scaled dual variable and $\rho > 0$ denotes the preselected augmented Lagrangian parameter.

D. GNMF-SR optimization using ADMM

To solve the nonconvex optimization problem, ADMM algorithm splits it into smaller pieces so that it can be easily handled one-to-one [17]. Moreover, the problem (3) can be first converted to two sub-problems: $\boldsymbol{A}^{(s)}$ and $\boldsymbol{B}^{(s)}$ via alternating optimization strategy, and then one of sub-problems can be solved using ADMM algorithm effectively if the other is fixed [16]. Combining alternating optimization and ADMM strategies [14]–[17], [21], we introduce two auxiliary variables $\tilde{\boldsymbol{A}}^{(s)}$ and $\tilde{\boldsymbol{B}}^{(s)}$, and consider the following minimization reformation of (3) as:

$$\begin{aligned} &\frac{1}{2}\sum_{s=1}^{S} \left\| \boldsymbol{X}^{(s)} - \boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)} \right\|_F^2 + \sum_{s=1}^{S} \beta^{(s)} \sum_{r=1}^{R(s)} \left\| \tilde{\boldsymbol{a}}_r^{(s)} \right\|_1 \\ &\text{subject to} \ \ \boldsymbol{A}^{(s)} = \tilde{\boldsymbol{A}}^{(s)}, \boldsymbol{B}^{(s)} = \tilde{\boldsymbol{B}}^{(s)}, \tilde{\boldsymbol{A}}^{(s)} \ge 0, \tilde{\boldsymbol{B}}^{(s)} \ge 0. \end{aligned} \tag{6}$$

Corresponding to $\boldsymbol{A}^{(s)}$, the auxiliary variable $\tilde{\boldsymbol{A}}^{(s)}$ still consists of two parts: $\tilde{\boldsymbol{A}}_C^{(s)}$ and $\tilde{\boldsymbol{A}}_I^{(s)}$, and $\tilde{\boldsymbol{A}}_C^{(1)} = \cdots \tilde{\boldsymbol{A}}_C^{(S)} = \tilde{\boldsymbol{A}}_C$. The augmented Lagrangian function for the above problem (6) is given by:

$$\begin{aligned} &\mathcal{L}(\boldsymbol{A}^{(s)}, \boldsymbol{B}^{(s)}, \tilde{\boldsymbol{A}}^{(s)}, \tilde{\boldsymbol{B}}^{(s)}, \boldsymbol{\Lambda}^{(s)}, \boldsymbol{\Gamma}^{(s)}) \\ &= \frac{1}{2}\sum_{s=1}^{S} \left\| \boldsymbol{X}^{(s)} - \boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)} \right\|_F^2 + \sum_{s=1}^{S} \beta^{(s)} \sum_{r=1}^{R(s)} \left\| \tilde{\boldsymbol{a}}_r^{(s)} \right\|_1 \\ &+ \sum_{s=1}^{S} \frac{\rho^{(s)}}{2} \left\| \boldsymbol{A}^{(s)} - \tilde{\boldsymbol{A}}^{(s)} + \boldsymbol{\Lambda}^{(s)} \right\|_F^2 \\ &+ \sum_{s=1}^{S} \frac{\mu^{(s)}}{2} \left\| \boldsymbol{B}^{(s)} - \tilde{\boldsymbol{B}}^{(s)} + \boldsymbol{\Gamma}^{(s)} \right\|_F^2, \end{aligned} \tag{7}$$

where $\boldsymbol{\Lambda}^{(s)} \in \mathbb{R}_+^{I^{(s)} \times R^{(s)}}$ and $\boldsymbol{\Gamma}^{(s)} \in \mathbb{R}_+^{R^{(s)} \times J^{(s)}}$ are termed as dual variables. $\rho^{(s)}$ and $\mu^{(s)}$ are the penalty parameters predefined by the user, and here we set $\rho^{(s)} = \left\| \boldsymbol{B}^{(s)} \right\|_F^2 / R^{(s)}$ and $\mu^{(s)} = \left\| \boldsymbol{A}^{(s)} \right\|_F^2 / R^{(s)}$ as suggested in [16].

$$
\begin{cases}
\boldsymbol{A}_C &= \left[\sum_{s=1}^{S} \boldsymbol{X}^{(s)} \big(\boldsymbol{B}_C^{(s)}\big)^T - \sum_{s=1}^{S} \boldsymbol{A}_I^{(s)} \boldsymbol{B}_I^{(s)} \big(\boldsymbol{B}_C^{(s)}\big)^T - \sum_{s=1}^{S} \rho^{(s)} \boldsymbol{\Lambda}_C^{(s)} + \sum_{s=1}^{S} \rho^{(s)} \tilde{\boldsymbol{A}}_C^{(s)} \right] \left[\sum_{s=1}^{S} \boldsymbol{B}_C^{(s)} \big(\boldsymbol{B}_C^{(s)}\big)^T + \sum_{s=1}^{S} \rho^{(s)} \boldsymbol{I} \right]^{-1} \\
\boldsymbol{A}_I^{(s)} &= \left[\boldsymbol{X}^{(s)} \big(\boldsymbol{B}_I^{(s)}\big)^T - \boldsymbol{A}_C^{(s)} \boldsymbol{B}_C^{(s)} \big(\boldsymbol{B}_I^{(s)}\big)^T - \rho^{(s)} \boldsymbol{\Lambda}_I^{(s)} + \rho^{(s)} \tilde{\boldsymbol{A}}_I^{(s)} \right] \left[\boldsymbol{B}_I^{(s)} \big(\boldsymbol{B}_I^{(s)}\big)^T + \rho^{(s)} \boldsymbol{I} \right]^{-1} \\
\boldsymbol{B}^{(s)} &= \left[\big(\boldsymbol{X}^{(s)}\big)^T \boldsymbol{A}^{(s)} - \mu^s \boldsymbol{\Gamma}^{(s)} + \mu^{(s)} \tilde{\boldsymbol{B}}^{(s)} \right] \left[\big(\boldsymbol{A}^{(s)}\big)^T \boldsymbol{A}^{(s)} + \mu^{(s)} \boldsymbol{I} \right]^{-1} \\
\tilde{\boldsymbol{A}}_C &= \left[\boldsymbol{A}_C + \frac{\sum_{s=1}^{S} \rho^{(s)} \boldsymbol{\Lambda}_C^{(s)}}{\sum_{s=1}^{S} \rho^{(s)}} - \frac{\sum_{s=1}^{S} \beta^{(s)} \boldsymbol{E}_C}{\sum_{s=1}^{S} \rho^{(s)}} \right]_+, \quad \tilde{\boldsymbol{A}}_I^{(s)} = \left[\boldsymbol{A}_I^{(s)} + \boldsymbol{\Lambda}_I^{(s)} - \frac{\beta^{(s)} \boldsymbol{E}_I}{\rho^{(s)}} \right]_+ \\
\tilde{\boldsymbol{B}}^{(s)} &= \left[\boldsymbol{B}^{(s)} + \boldsymbol{\Gamma}^{(s)} \right]_+, \quad \boldsymbol{\Lambda}^{(s)} = \boldsymbol{\Lambda}^{(s)} + \boldsymbol{A}^{(s)} - \tilde{\boldsymbol{A}}^{(s)}, \quad \boldsymbol{\Gamma}^{(s)} = \boldsymbol{\Gamma}^{(s)} + \boldsymbol{B}^{(s)} - \tilde{\boldsymbol{B}}^{(s)}
\end{cases}
\tag{8}
$$

For the solutions of $\{\boldsymbol{A}^{(s)}, \tilde{\boldsymbol{A}}^{(s)}, \boldsymbol{\Lambda}^{(s)}\}$, $\{\boldsymbol{B}^{(s)}, \tilde{\boldsymbol{B}}^{(s)}, \boldsymbol{\Gamma}^{(s)}\}$ in (7), we can calculate them successively via minimizing $\mathcal{L}$ with respect to one of them while fixing the others. Note that the primal variable $\boldsymbol{A}^{(s)}$ and auxiliary variable $\tilde{\boldsymbol{A}}^{(s)}$ both include the common and individual patterns, we need to calculate these two patterns separately. Furthermore, since the common pattern $\boldsymbol{A}_C$ (or $\tilde{\boldsymbol{A}}_C$) is shared by $\boldsymbol{A}^{(s)}$ (or $\tilde{\boldsymbol{A}}^{(s)}$), $s = 1, 2, \cdots, S$, we need to combine the information from all matrices from 1 to S to calculate their solutions. Different from the common pattern, the individual pattern $\boldsymbol{A}_I^{(s)}$ or $\tilde{\boldsymbol{A}}_I^{(s)}$ just needs to be calculated separately by the corresponding sth set data. Moreover, we also divide $\boldsymbol{B}^{(s)}$ into two parts $\boldsymbol{B}_C^{(s)} \in \mathbb{R}_+^{L \times J^{(s)}}$ and $\boldsymbol{B}_I^{(s)} \in \mathbb{R}_+^{(R^{(s)}-L) \times J^{(s)}}$ row-wisely. The specific solutions of primal, auxiliary and dual variables are given in (8), in which $\boldsymbol{E}_C \in \mathbb{R}_+^{I^{(l)} \times L}$ and $\boldsymbol{E}_I \in \mathbb{R}_+^{I^{(l)} \times (R^{(l)}-L)}$ are the matrices whose elements are all equal to one. We summarize the GNMF-SR algorithm based on ADMM update (termed as GNMF-SR-ADMM) in **Algorithm 1**.

Algorithm 1: GNMF-SR-ADMM algorithm

Input: $\boldsymbol{X}^{(s)}$, L, and $R^{(s)}$, $s = 1, \cdots, S$

Initialization:

$\boldsymbol{A}^{(s)}$, $\boldsymbol{B}^{(s)}$, $\tilde{\boldsymbol{A}}^{(s)}$, $\tilde{\boldsymbol{B}}^{(s)}$, $\boldsymbol{\Lambda}^{(s)}$, $\boldsymbol{\Gamma}^{(s)}$, $s = 1, \cdots, S$

for $k = 1, \cdots, MAX_k$ **do**

  According to (8);

  Update $\boldsymbol{A}_C$ and $\tilde{\boldsymbol{A}}_C$;

  for $s = 1, \cdots, S$ **do**

    Update $\boldsymbol{A}_I^{(s)}$, $\tilde{\boldsymbol{A}}_I^{(s)}$ and $\boldsymbol{\Lambda}^{(s)}$;

    Let $\boldsymbol{A}^{(s)} = \tilde{\boldsymbol{A}}^{(s)}$;

    Update $\boldsymbol{B}^{(s)}$, $\tilde{\boldsymbol{B}}^{(s)}$ and $\boldsymbol{\Gamma}^{(s)}$;

    Let $\boldsymbol{B}^{(s)} = \tilde{\boldsymbol{B}}^{(s)}$;

  end

  if *stopping criterion is satisfied* **then**

    return

  end

end

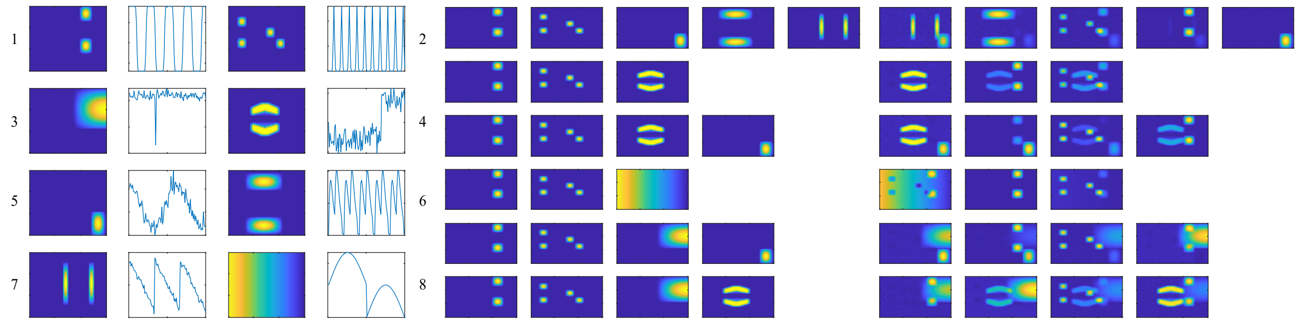
Output: $\boldsymbol{A}^{(s)}$, $\boldsymbol{B}^{(s)}$, $s = 1, 2, \cdots, S$

III. Experiment and Results

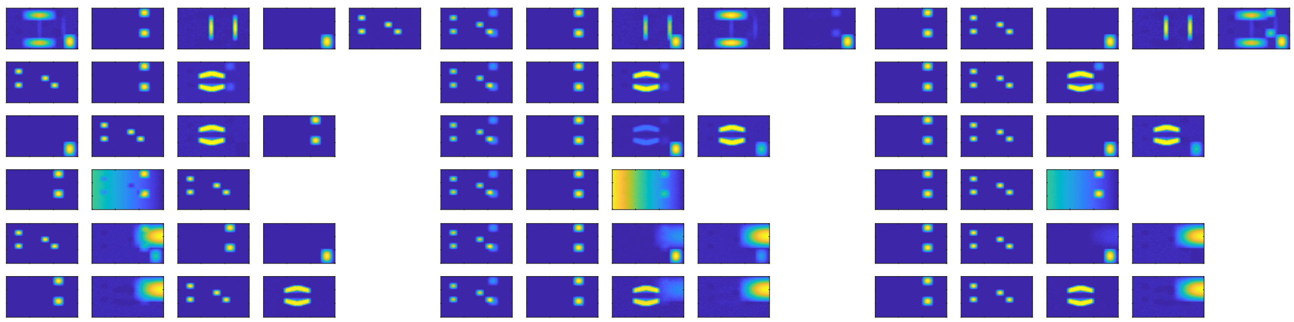
In this section, we provide an experiment of synthetic nonnegative fMRI-like data to demonstrate the performance of GNMF-SR-ADMM algorithm. Multiplicative update (MU, [1], [9]), alternating proximal gradient (APG, [4], [22]), alternative least squares (ALS, [3]) and fast hierarchical alternative least squares (fHALS, [2]) are also extended to solve the GNMF-SR model for comparison. In addition, by controlling the values of β and L, three other models including NMF ($\beta = 0$, $L = 0$), NMF-SR ($L = 0$) and GNMF ($\beta = 0$) are also considered in this experiment.

All experiments are carried out with the following computer configurations: CPU: Intel Core i5-7500 @ 3.40Hz 3.41Hz; Memory: 16Gb; System:64-bit Windows 10; Matlab R2016b.
Initialization. For the initialization of factor matrices, we use the uniformly distributed pseudorandom numbers generated by Matlab function `rand`.
Termination criterion. We use the change of relative error [22] (the threshold is set by 10^{-8}), and fix the maximum number of iterations to 1000.
Evaluation index. We adopt peak signal-to-noise ratio (PSNR, [3]) and inter-symbol-interference (ISI, [23]) to evaluate the accuracy of the estimated factor matrices. Meanwhile, we use the values of objective function (Obj), relative error (RelErr) and running time to assess the data fittings.
Data construction. We apply the GNMF-SR model to the joint analysis of multi-subject nonnegative fMRI-like data, which are constructed from the benchmark simulated complex fMRI dataset[1]. The amplitude of spatial maps (SM) and corresponding time courses (TC) are shown in **Fig. 1(a)** and they are adopted to generate the nonnegative fMRI-like data for 6 subjects according to the source index set {1,2,5,6,7}, {1,2,4}, {1,2,4,5}, {1,2,8}, {1,2,3,5} and {1,2,3,4} designed in [23], and more information about data construction can be found in [23]. The SM images of all subjects are shown in **Fig. 1(b)**. Each row corresponds to one subject, and the first two columns are shared by all the subjects, which are considered as the common patterns and the remains are the individual ones.

[1] http://mlsp.umbc.edu/simulated_complex_fmri_data.html



(a) Original SMs and TCs (b) Regroup SMs (c) NMF-ADMM, PSNR=45.52dB



(d) NMF-SR-ADMM, PSNR=61.55dB (e) GNMF-ADMM, PSNR=52.07dB (f) GNMF-SR-ADMM, PSNR=64.31dB

Fig. 1. (a) Amplitude images of 1-8 simulated fMRI-like spatial maps (1st and 3rd columns) and corresponding time courses (2nd and 4th columns). (b-f) SM images of constructed data and that of estimated ones via NMF-ADMM ($\beta = 0$, $L = 0$), NMF-SR-ADMM ($\beta = 3e-4$, $L = 0$), GNMF-ADMM ($\beta = 0$, $L = 2$) and GNMF-SR-ADMM ($\beta = 3e-4$, $L = 2$) under SNR=20dB.

We fix SNR=20dB, and select 25 values for β ranging from 0 to 5. With varying βs, the PSNR curves of SM estimates averaged from 30 Monte Carlo runs in the GNMF-SR model ($L = 0$ & $L = 2$) via MU, ALS, APG, fHALS and ADMM algorithms are shown in **Fig. 2**. Note that when $L = 0$ and $\beta = 0$, the GNMF-SR will degenerate into the NMF problem. From **Fig. 2**, we can see that the PSNR values of all algorithms will increase and reach the highest at some point when the sparse penalty parameter β increases, except that MU-based algorithms show the insensitivity to the settings of β between 0 and 5. The sparse penalty will have a negative effect on the algorithm performance when β increases to a certain point.

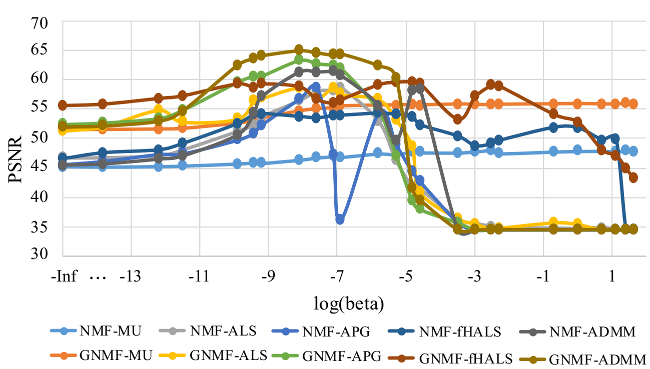


Fig. 2. Mean PSNR of SM estimates for 6 subjects under NMF-SR ($L = 0$) and GNMF-SR ($L = 2$) models with the βs of 25 values varying from 0 to 5, SNR=20dB.

We also present the specific values of PSNR, ISI, Obj, RelErr and running time for each algorithm under $\beta = 0$ and a post-selected β (which corresponds to the best performance) in **Table I**. The performance of the GNMF-based methods is superior to that of NMF-based ones. With sparse regularization, the performance of NMF-based and GNMF-based methods can be both significantly improved. Interestingly, sparse penalty yields better performance improvements than group constraint for NMF-based methods. GNMF-SR-ADMM algorithm achieves the best performance, followed by GNMF-SR-APG, NMF-SR-ADMM and GNMF-SR-fHALS algorithms. However, from **Table I**, we can see that ADMM-based methods seem are time consuming and will be improved in our future work.

Furthermore, the SM images estimated via NMF-ADMM, NMF-SR-ADMM, GNMF-ADMM and GNMF-SR-ADMM at $\beta = 0,\ 3e-4$ and $L = 0,\ 2$ are shown in **Fig.1(c-f)**. It can be clearly seen that some of SM images obtained by NMF-ADMM and GNMF-ADMM algorithms are blurred with shadows or small outliers. By imposing adequate sparse regularization, those blurs are basically eliminated in the results of NMF-SR-ADMM and GNMF-SR-ADMM algorithms. Moreover, from **Fig. 1(e-f)**, we can denote that two group analysis methods including GNMF-ADMM and GNMF-SR-ADMM can extract both the common and individual patterns for all the datasets, and also successfully correct the disorder scenario of common patterns in the results of two NMF-based algorithms as shown in **Fig. 1(c-d)**.

TABLE I
PERFORMANCE COMPARISON ON FMRI_LIKE DATA BASED ON GNMF-SR MODEL ($L = 0, 2$, SNR=20DB)

	Method	β	PSNR	ISI	Obj	RelErr	Time/s
$L = 0$	ALS	0	46.68	0.0545	0.0103	0.3332	3.6941
		1e-3	58.70	0.0117	0.0103	0.3335	6.2381
	MU	0	45.23	0.0741	0.0102	0.3312	5.3236
		8e-2	48.02	0.0645	0.0102	0.3318	5.2624
	APG	0	45.56	0.0859	0.0101	0.3303	6.8339
		5e-4	58.61	0.0303	0.0105	0.3367	3.7993
	fHALS	0	46.50	0.0753	0.0101	0.3298	6.1396
		3e-3	54.24	0.0281	0.0101	0.3308	6.1959
	ADMM	0	45.47	0.0876	0.0101	0.3303	7.0635
		3e-4	61.27	0.0128	0.0103	0.3330	7.1136
$L = 2$	ALS	0	54.33	0.0393	0.0197	0.4245	2.9618
		3e-4	58.69	0.0152	0.0149	0.3789	4.3242
	MU	0	51.43	0.0330	0.0103	0.3340	5.2181
		4	55.91	0.0344	0.0104	0.3352	5.2229
	APG	0	52.37	0.0258	0.0103	0.3336	3.6041
		5e-4	62.70	0.0215	0.0108	0.3412	5.0651
	fHALS	0	55.58	0.0376	0.0109	0.3387	4.9561
		8e-3	59.65	0.0083	0.0103	0.3335	5.7812
	ADMM	0	51.95	0.0302	0.0103	0.3335	7.2444
		3e-4	64.94	0.0062	0.0104	0.3359	7.1910

IV. CONCLUSION

In this paper, we formulated a flexible group nonnegative matrix factorization with sparse regularization (GNMF-SR) model for the group analysis of data from multiple sources. Alternating optimization and alternating direction method of multipliers (ADMM) strategies were combined to optimize the GNMF-SR model, in which the common and individual patterns can be simultaneously extracted while aligning the common patterns. The experiment of simulated fMRI-like data demonstrates that the proposed GNMF-SR-ADMM algorithm has better performance than its counterparts in terms of high PSNRs and factorization accuracy. Imposing group constraint and sparse penalty can greatly improve the performance of NMF-based algorithms.

REFERENCES

[1] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788, 1999.

[2] Andrzej Cichocki and Anh-Huy Phan, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 92, no. 3, pp. 708–721, 2009.

[3] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*, John Wiley & Sons, 2009.

[4] Naiyang Guan, Dacheng Tao, Zhigang Luo, and Bo Yuan, "Nenmf: An optimal gradient method for nonnegative matrix factorization," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2882–2898, 2012.

[5] Xiao-Feng Gong, Xiu-Lin Wang, and Qiu-Hua Lin, "Generalized non-orthogonal joint diagonalization with lu decomposition and successive rotations," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1322–1334, 2015.

[6] Guoxu Zhou, Qibin Zhao, Yu Zhang, Tülay Adalı, Shengli Xie, and Andrzej Cichocki, "Linked component analysis from matrices to high-order tensors: Applications to biomedical data," *Proceedings of the IEEE*, vol. 104, no. 2, pp. 310–331, 2016.

[7] Xun Chen, Z Jane Wang, and Martin McKeown, "Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods," *IEEE Signal Processing Magazine*, vol. 33, no. 3, pp. 86–107, 2016.

[8] Xiulin Wang, Wenya Liu, Petri Toiviainen, Tapani Ristaniemi, and Fengyu Cong, "Group analysis of ongoing eeg data based on fast double-coupled nonnegative tensor decomposition," *Journal of neuroscience methods*, vol. 330, pp. 108502, 2020.

[9] Hyekyoung Lee and Seungjin Choi, "Group nonnegative matrix factorization for EEG classification," in *Artificial Intelligence and Statistics*, 2009, pp. 320–327.

[10] Bonggun Shin and Alice Oh, "Bayesian group nonnegative matrix factorization for EEG analysis," *arXiv preprint arXiv:1212.4347*, 2012.

[11] Patrik O Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of machine learning research*, vol. 5, no. Nov, pp. 1457–1469, 2004.

[12] Jun Xu, Lei Xiang, Guanhao Wang, Shridar Ganesan, Michael Feldman, Natalie NC Shih, Hannah Gilmore, and Anant Madabhushi, "Sparse non-negative matrix factorization (snmf) based color unmixing for breast histopathological image analysis," *Computerized Medical Imaging and Graphics*, vol. 46, pp. 20–29, 2015.

[13] Morten Mørup and Lars Kai Hansen, "Tuning pruning in sparse non-negative matrix factorization," in *2009 17th European Signal Processing Conference*. IEEE, 2009, pp. 1923–1927.

[14] Dennis L Sun and Cedric Fevotte, "Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence," in *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2014, pp. 6201–6205.

[15] Davood Hajinezhad, Tsung-Hui Chang, Xiangfeng Wang, Qingjiang Shi, and Mingyi Hong, "Nonnegative matrix factorization using admm: Algorithm and convergence analysis," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 4742–4746.

[16] Kejun Huang, Nicholas D Sidiropoulos, and Athanasios P Liavas, "A flexible and efficient algorithmic framework for constrained matrix and tensor factorization," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5052–5065, 2016.

[17] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[18] Davood Hajinezhad and Qingjiang Shi, "Alternating direction method of multipliers for a class of nonconvex bilinear optimization: convergence analysis and applications," *Journal of Global Optimization*, vol. 70, no. 1, pp. 261–288, 2018.

[19] Yu Wang, Wotao Yin, and Jinshan Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, vol. 78, no. 1, pp. 29–63, 2019.

[20] Evrim Acar, Gözde Gürdeniz, Morten A Rasmussen, Daniela Rago, Lars O Dragsted, and Rasmus Bro, "Coupled matrix factorization with sparse factors to identify potential biomarkers in metabolomics," in *2012 IEEE 12th International Conference on Data Mining Workshops*. IEEE, 2012, pp. 1–8.

[21] Deqing Wang, Fengyu Cong, and Tapani Ristaniemi, "Sparse non-negative candecomp/parafac decomposition in block coordinate descent framework: A comparison study," *arXiv preprint arXiv:1812.10637*, 2018.

[22] Yangyang Xu and Wotao Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM Journal on imaging sciences*, vol. 6, no. 3, pp. 1758–1789, 2013.

[23] Xiao-Feng Gong, Lei Mao, Ying-Liang Liu, and Qiu-Hua Lin, "A jacobi generalized orthogonal joint diagonalization algorithm for joint blind source separation," *IEEE Access*, vol. 6, pp. 38464–38474, 2018.