ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/220320293

# Group Nonnegative Matrix Factorization for EEG Classification.

**Article** · January 2009

Source: DBLP

CITATIONS
46

READS
55

**2 authors:**

Lee Hyekyoung
Seoul National University
**28** PUBLICATIONS **622** CITATIONS
SEE PROFILE

Seungjin Choi
Pohang University of Science and Technology
**278** PUBLICATIONS **4,037** CITATIONS
SEE PROFILE

**Some of the authors of this publication are also working on these related projects:**


Anomaly Detection Project for SKT View project

All content following this page was uploaded by Lee Hyekyoung on 24 June 2014.

The user has requested enhancement of the downloaded file.

# Group Nonnegative Matrix Factorization for EEG Classification

**Hyekyoung Lee**
Department of Computer Science
Pohang University of Science and Technology
San 31 Hyoja-dong, Nam-gu
Pohang 790-784, Korea

**Seungjin Choi**
Department of Computer Science
Pohang University of Science and Technology
San 31 Hyoja-dong, Nam-gu
Pohang 790-784, Korea

## Abstract

Given electroencephalogram (EEG) data measured from several subjects under the same conditions, our goal is to estimate common task-related bases in a linear model that capture intra-subject variations as well as inter-subject variations. Such bases capture the common phenomenon in group data, which is a core of *group analysis*. In this paper we present a method of nonnegative matrix factorization (NMF) that is well suited to analyzing EEG data of multiple subjects. The method is referred to as *group nonnegative matrix factorization* (GNMF) where we seek task-related common bases reflecting both intra-subject and inter-subject variations, as well as bases involving individual characteristics. We compare GNMF with NMF and some modified NMFs, in the task of learning spectral features from EEG data. Experiments on brain computer interface (BCI) competition data indicate that GNMF improves the EEG classification performance. In addition, we also show that GNMF is useful in the task of subject-to-subject transfer where the prediction for an unseen subject is performed based on a linear model learned from different subjects in the same group.

## 1 INTRODUCTION

Electroencephalogram (EEG) is multivariate time series data measured at multiple sensors placed on scalp, which reflects electrical potentials induced by brain activities. EEG classification is an important task in brain computer interface (BCI) which provides a new dimension in human computer interface, directly connecting a computer with human thinking (Ebrahimi et al., 2003).

In the task of motor imagery, spectral characteristics of EEG have been used as features, including $\mu$ rhythm (8-12 Hz) (Wolpaw et al., 2002) and $\beta$ rhythm (18-25 Hz) on sensori-motor cortex. Those rhythms decrease during movement or in preparation for movement, known as *event-related desynchronization* (ERD) and increase after movement or in relaxation, referred to as *event-related synchronization* (ERS). These phenomena, however, can happen in different frequency bands or different regions, depending on subjects. For instance, they might occur in 16-20 Hz, but not in 8-12 Hz (Lal et al., 2003).

Nonnegative matrix factorization (NMF) is a linear data model which is useful in handling nonnegative data (Lee & Seung, 1999). NMF allows only nonsubtractive combinations of nonnegative basis vectors, leading to (possibly) a parts-based representation. NMF was shown to be useful in determining discriminative basis vectors which well reflect meaningful spectral characteristics without cross-validation in motor imagery EEG tasks (Lee et al., 2006). However, a direct application of NMF to EEG data measured from multiple subjects takes only intra-subject variations into account.

Given data measured from several subjects, group analysis seeks task-specific patterns which are common in two or more subjects in a group. Such commonly-appearing patterns are more reliable than individual ones in a single subject. Two types of variations are considered in group analysis, including: (1) *intra-subject variations* since a subject's response varies from trial to trial; (2) *inter-subject variations* since responses vary from subject to subject.

Appearing in Proceedings of the $12^{th}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2009, Clearwater Beach, Florida, USA. Volume 5 of JMLR: W&CP 5. Copyright 2009 by the authors.

Approaches to group analysis of multiple subjects include: (1) *fixed-effects analysis* (FFX); (2) *random-effects analysis* (RFX) (Frackowiak et al., 2003). FFX assumes that all subjects in a group have the same activation patterns with noise, allowing only intra-subject variability. On the other hand, RFX assumes that each subject has activation patterns with noise which are different across subjects, taking both intra-subject and inter-subject variability into account. These two approaches have been embodied in generalized linear model (GLM) or independent component analysis (ICA) for a group fMRI study (Holmes & Friston, 1998; Lee et al., 2008; Park et al., 2003).

Group analysis is widely used in fMRI study but is a pre-mature technique for EEG analysis. In this paper we develop group NMF (GNMF) which is more suitable to EEG analysis of multiple subjects. GNMF seeks task-related common bases which capture both intra-subject and inter-subject variations, as well as task-independent bases involving individual characteristics. Several naive modifications of NMF are also considered and compared to GNMF, in a task of EEG classification using IDIAP data in BCI competition III (Blankertz et al., 2006):

- **GNMF:** GNMF (the main contribution of this paper) seeks common bases which capture both intra-subject and inter-subject variations as well as bases involving individual characteristics, as in RFX.
- **Hierarchical-NMF:** NMFs are applied to each subject's data individually, followed by a hierarchical clustering to determine commonly-appearing bases across subjects. Such common bases reflect both intra-subject and inter-subject variations. Bases other than common bases are considered as individual task-independent characteristics.
- **One-NMF:** A single NMF is applied to a concatenated data constructed from all subjects in a group at hand, assuming that the common bases capture only intra-subject variations.
- **FFX-NMF:** FFX-NMF seeks common bases reflecting only intra-subject variations as well as bases involving individual characteristics, as in FFX.

The rest of this paper is presented as follows. We begin with a detailed explanation on how to make use of NMF to learn spectral features for EEG classification in Section 2. The main contribution, GNMF, is described in Section 3 with a toy example. Other modifications of NMF such as Hierarchical-NMF, One-NMF, FFX-NMF are also explained in Section 3. In Section 4, numerical experiments on IDIAP dataset in BCI competition III indicate that GNMF improves the EEG classification performance. In addition, we also emphasize that GNMF is useful in the task of subject-to-subject transfer where the prediction for an unseen subject is performed based on a linear model learned from different subjects in the same group. Finally conclusions are drawn in Section 5.

## 2 NMF FOR SPECTRAL EEG FEATURE EXTRACTION

We construct the data matrix $\boldsymbol{X} \in \mathbb{R}_+^{m \times n}$ by applying a wavelet transform or short-time Fourier transform to the time-domain EEG signal such that each row in $\boldsymbol{X}$ is associated with a frequency profile across trials. A detailed description of the data matrix construction is given in Section 4.1.

NMF seeks a rank-$r$ decomposition of $\boldsymbol{X} \in \mathbb{R}_+^{m \times n}$ that is of the form:

$$\boldsymbol{X} \approx \boldsymbol{A}\boldsymbol{S}, \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}_+^{m \times r}$ contains *bases* in its columns and $\boldsymbol{S} \in \mathbb{R}_+^{r \times n}$ is the *encoding matrix* where each row represents the extent to which each basis is used to reconstruct the data vector. An exemplary application of NMF to the time-frequency representation of EEG data is shown in Fig. 1.

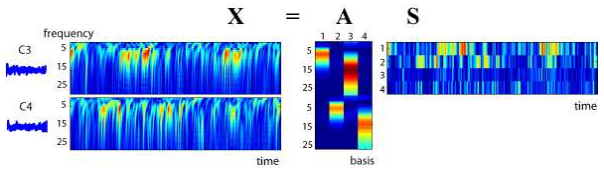



Figure 1: The data matrix $\boldsymbol{X}$ (constructed by collecting frequency profiles from two channels $C_3$ and $C_4$) is decomposed into a product of basis matrix $\boldsymbol{A}$ and encoding matrix $\boldsymbol{S}$ with both of them restricted to be nonnegative. The first two columns of $\boldsymbol{A}$ are related to the characteristics of $\mu$ rhythm (8-12 Hz) for $C_3$ and $C_4$ channels. The last two columns of $\boldsymbol{A}$ represent $\beta$ rhythm around 15 Hz for both channels. Rows of $\boldsymbol{S}$ are corresponding feature profiles.

In the case where the squared Euclidean distance is used as a discrepancy measure between the data $\boldsymbol{X}$ and the model $\boldsymbol{A}\boldsymbol{S}$, NMF involves the following optimization:

$$\underset{A \geq 0, S \geq 0}{\arg\min}\, \mathcal{J} = \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{A}\boldsymbol{S}\|^2. \tag{2}$$

Gradient descent learning (which is additive update) can be applied to determine a solution to (2), however, nonnegativity for $\boldsymbol{A}$ and $\boldsymbol{S}$ is not preserved without further operations at iterations.

On the other hand, a multiplicative method developed in (Lee & Seung, 1999) provides a simple iterative algorithm to solve (2). We apply a slightly different approach to derive the same multiplicative algorithm. Suppose that the gradient of an error function has a decomposition that is of the form

$$\nabla \mathcal{J} = [\nabla \mathcal{J}]^+ - [\nabla \mathcal{J}]^-, \tag{3}$$

where $[\nabla \mathcal{J}]^+ > 0$ and $[\nabla \mathcal{J}]^- > 0$. Then multiplicative update for parameters $\Theta$ has the form

$$\Theta \leftarrow \Theta \odot \left( \frac{[\nabla \mathcal{J}]^-}{[\nabla \mathcal{J}]^+} \right)^{\cdot \eta}, \tag{4}$$

where $\odot$ represents Hadamard product (element-wise product) and $(\cdot)^{\cdot \eta}$ denotes the element-wise power and $\eta$ is a learning rate ($0 < \eta \leq 1$). It can be easily seen that the multiplicative update (4) preserves the nonnegativity of the parameter $\Theta$, while $\nabla \mathcal{J} = 0$ when the convergence is achieved.

Derivatives of the error function (2) with respect to $\boldsymbol{A}$ with $\boldsymbol{S}$ fixed and with respect to $\boldsymbol{S}$ with $\boldsymbol{A}$ fixed, are given by

$$\begin{aligned} \nabla_A \mathcal{J} &= [\nabla_A \mathcal{J}]^+ - [\nabla_A \mathcal{J}]^- \\ &= \boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^\top - \boldsymbol{X}\boldsymbol{S}^\top, \end{aligned} \tag{5}$$

$$\begin{aligned} \nabla_S \mathcal{J} &= [\nabla_S \mathcal{J}]^+ - [\nabla_S \mathcal{J}]^- \\ &= \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{S} - \boldsymbol{A}^\top \boldsymbol{X}. \end{aligned} \tag{6}$$

With these gradient calculations, the rule (4) with $\eta = 1$ yields the well-known Lee and Seung's multiplicative updates (Lee & Seung, 1999)

$$\boldsymbol{A} \leftarrow \boldsymbol{A} \odot \frac{\boldsymbol{X}\boldsymbol{S}^\top}{\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^\top}, \tag{7}$$

$$\boldsymbol{S} \leftarrow \boldsymbol{S} \odot \frac{\boldsymbol{A}^\top \boldsymbol{X}}{\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{S}}, \tag{8}$$

where $\odot$ is an element-wise product (Hadamard product) and all divisions are element-wise divisions.

When a test data matrix $\boldsymbol{X}_{test}$ is given, its associated feature matrix $\boldsymbol{A}_{test}$ can be computed in two different ways:

- The feature matrix $\boldsymbol{A}_{test}$ is determined by LS projection,

$$\boldsymbol{S}_{test} = [\boldsymbol{A}]^\dagger \boldsymbol{X}_{test} \tag{9}$$

  where $\dagger$ represents the pseudo-inverse. In such a case, $\boldsymbol{S}_{test}$ might have negative elements but works well in the viewpoint of feature extraction.

- We iterate the update rule (7) until convergence, with $\boldsymbol{S}$ (learned in the training phase) fixed.

## 3 NMF FOR GROUP ANALYSIS

Suppose that we are given $L$ sets of data (measured from $L$ different subjects), $\boldsymbol{X} = \{\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(L)}\}$, each of which $\boldsymbol{X}^{(l)} \in \mathbb{R}_+^{m \times n_l}$ contains frequency profiles of EEG data measured from subject $l$. Assume that EEG measurements across subjects were obtained under the same conditions (i.e., subjects share the same task).

The linear model in GNMF assumes that each data matrix $\boldsymbol{X}^{(l)}$ is generated by

$$\boldsymbol{X}^{(l)} = \boldsymbol{A}^{(l)} \boldsymbol{S}^{(l)} = \left[ \boldsymbol{A}_C^{(l)} \;\; \boldsymbol{A}_I^{(l)} \right] \boldsymbol{S}^{(l)}, \tag{10}$$

where the basis matrix $\boldsymbol{A}^{(l)} \in \mathbb{R}_+^{m \times r}$ is composed of two types of bases, where $\boldsymbol{A}_C^{(l)} \in \mathbb{R}_+^{m \times r_C}$ consists of task-related common bases capturing intra-subject and inter-subject variations, and $\boldsymbol{A}_I^{(l)} \in \mathbb{R}_+^{m \times r_I}$ ($r = r_C + r_I$) consists of bases which reflect task-independent individual characteristics. The composition of the basis matrix in GNMF is shown in Fig. 2

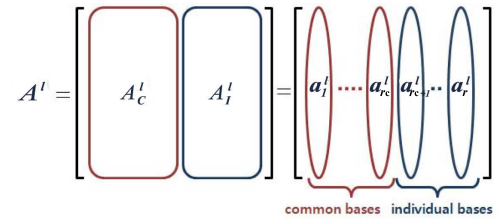



Figure 2: The composition of basis matrix $\boldsymbol{A}^{(l)} \in \mathbb{R}^{m \times r}$ is shown, consisting of $r_C$ (task-related) common bases and $r_I$ (task-independent) individual bases.

### 3.1 GNMF

As mentioned earlier, GNMF seeks task-related common bases $\boldsymbol{A}_C^{(l)}$ that are as close as possible across subjects and task-independent individual bases $\boldsymbol{A}_I^{(l)}$ are as far apart as possible. At the same time, GNMF also seeks a rank-$r$ approximation, as in the standard NMF. To this end, we consider the following objective function:

$$\begin{aligned} \mathcal{J}_{GNMF} &= \lambda \sum_{l=1}^{L} \left\| \boldsymbol{X}^{(l)} - \boldsymbol{A}^{(l)} \boldsymbol{S}^{(l)} \right\|^2 + \gamma \sum_l \left\| \boldsymbol{A}^{(l)} \right\|^2 \\ &\quad + \frac{\alpha}{2} \sum_l \sum_{j \neq l} \left\| \boldsymbol{A}_C^{(j)} - \boldsymbol{A}_C^{(l)} \right\|^2 \\ &\quad - \frac{\beta}{2} \sum_l \sum_{j \neq l} \left\| \boldsymbol{A}_I^{(j)} - \boldsymbol{A}_I^{(l)} \right\|^2, \end{aligned} \tag{11}$$

where the first two terms involve the rank-$r$ approximation with regularization and the third term enforces common bases $\boldsymbol{A}_C^{(j)}$ and $\boldsymbol{A}_C^{(l)}$ to be as close as possible, while the last term keeps individual bases $\boldsymbol{A}_I^{(j)}$ and $\boldsymbol{A}_I^{(l)}$ as far apart as possible. The minimization of (11) is done subject to constraints $\boldsymbol{A}_C^{(l)} \geq 0$, $\boldsymbol{A}_I^{(l)} \geq 0$, and $\boldsymbol{S}^{(l)} \geq 0$ for $l = 1, 2, \ldots, L$. We derive multiplicative algorithms for GNMF using the same technique as introduced in Section 2. Note that $\boldsymbol{S}^{(l)}$ involves only $\sum_{l=1}^{L} \left\| \boldsymbol{X}^{(l)} - \boldsymbol{A}^{(l)} \boldsymbol{S}^{(l)} \right\|^2$ in (11), multiplicative updates are the same as (8), i.e.,

$$\boldsymbol{S}^{(l)} \leftarrow \boldsymbol{S}^{(l)} \odot \frac{{\boldsymbol{A}^{(l)}}^\top \boldsymbol{X}^{(l)}}{{\boldsymbol{A}^{(l)}}^\top \boldsymbol{A}^{(l)} \boldsymbol{S}^{(l)}}, \tag{12}$$

for $l = 1, \ldots, L$. On the other hand, introducing ${\boldsymbol{S}^{(l)}}^\top = \left[ {\boldsymbol{S}_C^{(l)}}^\top \; {\boldsymbol{S}_I^{(l)}}^\top \right]$, multiplicative updates for common bases and individual bases (for $l = 1, \ldots, L$) are given by

$$\boldsymbol{A}_C^{(l)} \leftarrow \boldsymbol{A}_C^{(l)} \odot \left( \frac{[\nabla_{A_C^{(l)}} \mathcal{J}_{GNMF}]^-}{[\nabla_{A_C^{(l)}} \mathcal{J}_{GNMF}]^+} \right), \tag{13}$$

$$\boldsymbol{A}_I^{(l)} \leftarrow \boldsymbol{A}_I^{(l)} \odot \left( \frac{[\nabla_{A_I^{(l)}} \mathcal{J}_{GNMF}]^-}{[\nabla_{A_I^{(l)}} \mathcal{J}_{GNMF}]^+} \right), \tag{14}$$

where

$$\left( \frac{[\nabla_{A_C^{(l)}} \mathcal{J}_{GNMF}]^-}{[\nabla_{A_C^{(l)}} \mathcal{J}_{GNMF}]^+} \right) = \frac{\boldsymbol{X}^{(l)} {\boldsymbol{S}_C^{(l)}}^\top + (\alpha/\lambda) \sum_{j \neq l} \boldsymbol{A}_C^{(j)}}{\left[ \boldsymbol{A}^{(l)} \boldsymbol{S}^{(l)} \right] {\boldsymbol{S}_C^{(l)}}^\top + (\alpha/\lambda)(L-1) \boldsymbol{A}_C^{(l)} + (\gamma/\lambda) \boldsymbol{A}_C^{(l)}},$$

and

$$\left( \frac{[\nabla_{A_I^{(l)}} \mathcal{J}_{GNMF}]^-}{[\nabla_{A_I^{(l)}} \mathcal{J}_{GNMF}]^+} \right) = \frac{\boldsymbol{X}^{(l)} {\boldsymbol{S}_I^{(l)}}^\top + (\beta/\lambda)(L-1) \boldsymbol{A}_I^{(l)}}{\left[ \boldsymbol{A}^{(l)} \boldsymbol{S}^{(l)} \right] {\boldsymbol{S}_I^{(l)}}^\top + (\beta/\lambda) \sum_{j \neq l} \boldsymbol{A}_I^{(j)} + (\gamma/\lambda) \boldsymbol{A}_I^{(l)}}.$$

### 3.2 OTHER MODIFICATIONS OF NMF

In addition to GNMF, three more (naive) modifications of NMF for group analysis are provided and compared with GNMF: including Hierarchical-NMF, One-NMF, and FFX-NMF.

#### 3.2.1 Hierarchical-NMF

First, NMF is applied to each data set individually, i.e.,

$$\boldsymbol{X}^{(l)} = \boldsymbol{A}^{(l)} \boldsymbol{S}^{(l)},$$

for $l = 1, \ldots, L$. Task-related common bases are determined by applying an agglomerative hierarchical clustering to column vectors of $\boldsymbol{A}^{(l)}$. Hierarchical-NMF can be viewed as a multi-level approach to RFX, in the sense that bases are first estimated individually by assuming only intra-subject variability, then common bases are selected by a clustering method by taking inter-subject variability into account. Hierarchical-NMF works fine for the toy example that is explained in Section 3.3. As will be shown, however, for real-world data, it is not easy to compute clusters associated with common bases.

#### 3.2.2 One-NMF

One-NMF considers a concatenated data matrix $\boldsymbol{X} = [\boldsymbol{X}^{(1)} \cdots \boldsymbol{X}^{(L)}]$ to find a decomposition of $\boldsymbol{X} = \boldsymbol{A}[\boldsymbol{S}^{(1)} \cdots \boldsymbol{S}^{(L)}]$. That is One-NMF is nothing but an one-time application of NMF to the concatenated data matrix $\boldsymbol{X}$.

#### 3.2.3 FFX-NMF

FFX-NMF assumes that common bases are exactly the same ($\boldsymbol{A}_C^{(l)} = \boldsymbol{A}_C$ for $l = 1, \ldots, L$) across subjects, allowing individual bases to be different. Only intra-subject variations are implicitly considered in FFX-NMF, in addition to individual characteristics, just as in FFX. FFX-NMF seeks a decomposition of the form

$$\boldsymbol{X}^{(l)} = \left[ \boldsymbol{A}_C \;\; \boldsymbol{A}_I^{(l)} \right] \boldsymbol{S}^{(l)}, \tag{15}$$

for l=1,...,L. The objective function for FFX-NMF is given by

$$\begin{aligned} \mathcal{J}_{FFX-NMF} &= \sum_{l=1}^{L} \left\| \boldsymbol{X}^{(l)} - \boldsymbol{A}_C \boldsymbol{S}_C^{(l)} - \boldsymbol{A}_I^{(l)} \boldsymbol{S}_I^{(l)} \right\|^2 \\ &+ \gamma \left\{ L \left\| \boldsymbol{A}_C \right\|^2 + \sum_l \left\| \boldsymbol{A}_I^{(l)} \right\|^2 \right\}, \end{aligned} \tag{16}$$

which is to be minimized with respect to $\boldsymbol{A}_C \geq 0$, $\boldsymbol{A}_I^{(l)} \geq 0$, $\boldsymbol{S}_C^{(l)} \geq 0$, and $\boldsymbol{S}_I^{(l)} \geq 0$ for $l = 1, \ldots, L$. The regularization term

$$\sum_l \left\| \boldsymbol{A}^{(l)} \right\|^2 = L \left\| \boldsymbol{A}_C \right\|^2 + \sum_l \left\| \boldsymbol{A}_I^{(l)} \right\|^2$$

enforces the information learned to be evenly distributed on column vectors of $\boldsymbol{A}^{(l)}$, avoiding that the information concentrated on $\boldsymbol{A}_C$ since $\boldsymbol{A}_C$ is estimated using the whole data. Assuming $\boldsymbol{A}_I = \boldsymbol{A}_I^{(l)}$ for $l = 1, \ldots, L$ in FFX-NMF leads to One-NMF.

Multiplicative algorithms to learn $\boldsymbol{S}^{(l)}$ have the same form as (12), except that $\boldsymbol{A}^{(l)} = [\boldsymbol{A}_C \;\; \boldsymbol{A}_I^{(l)}]$ are different to the ones in GNMF. Multiplicative updates for

$\boldsymbol{A}_C$ and $\boldsymbol{A}_I^{(l)}$ are given by

$$\boldsymbol{A}_C \leftarrow \boldsymbol{A}_C \odot \frac{\sum_l \boldsymbol{X}^{(l)} \boldsymbol{S}_C^{(l)\top}}{\sum_l \boldsymbol{A}^{(l)} \boldsymbol{S}^{(l)} \boldsymbol{S}_C^{(l)\top} + \gamma L \boldsymbol{A}_C}, \quad (17)$$

$$\boldsymbol{A}_I^{(l)} \leftarrow \boldsymbol{A}_I^{(l)} \odot \frac{\boldsymbol{X}^{(l)} \boldsymbol{S}_I^{(l)\top}}{\boldsymbol{A}^{(l)} \boldsymbol{S}^{(l)} \boldsymbol{S}_I^{(l)\top} + \gamma \boldsymbol{A}_I^{(l)}}. \quad (18)$$

FFX-NMF does not consider task-related inter-subject variations, as in FFX, leading to the same common bases as One-NMF, which is also demonstrated in the toy example explained in Section 3.3.

### 3.3 A TOY EXAMPLE

We present a toy example where we synthetically generate common bases and individual bases as well as associated encodings, mimicking brain activities of three different subjects. Through this toy example, we stress the useful behavior of GNMF, compared to Hierarchical-NMF, One-NMF, and FFX-NMF.

Fig. 3 shows three common basis images (each of which is of size $23 \times 23$) and one individual basis image for each of three subjects. Each basis image is converted to a vector, leading to $\boldsymbol{A}^{(l)} \in \mathbb{R}^{529\times 4}$ for $l = 1, 2, 3$. Three common bases look similar to each other, imitating left/right/bottom activations in the spatial domain. Individual bases differ depending on subjects.

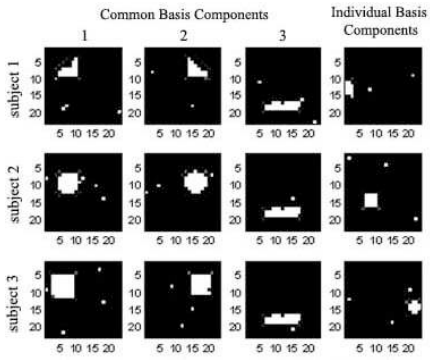



Figure 3: From top to bottom, bases for three subjects $\{\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}, \boldsymbol{A}^{(3)}\}$ are shown. From left to right, three common bases and one individual basis are displayed for each subject. Each image of size $23\times 23$ is converted to a 529-dimensional vector, leading to $\boldsymbol{A}^{(l)} \in \mathbb{R}^{529\times 4}$. The first three bases involve common patterns that are activated in a similar region (reflecting inter-subject variations). The basis images in the last column are individual components that are randomly activated, depending on subjects.

Synthetically-generated on-off encodings (1 and 0 represent 'activation' and 'no activation', respectively) are shown in Fig. 4 where activation time courses are displayed, including $\{\boldsymbol{S}^{(1)}, \boldsymbol{S}^{(2)}, \boldsymbol{S}^{(3)}\}$. Data matrices $\boldsymbol{X}^{(l)}$ are generated by $\boldsymbol{A}^{(l)}\boldsymbol{S}^{(l)} + \mathcal{N}(0, 0.1)$ where $\mathcal{N}(0, 0.1)$ denotes Gaussian noise with mean 0 and variance 0.1. Exemplary observed data are shown in Fig. 5.

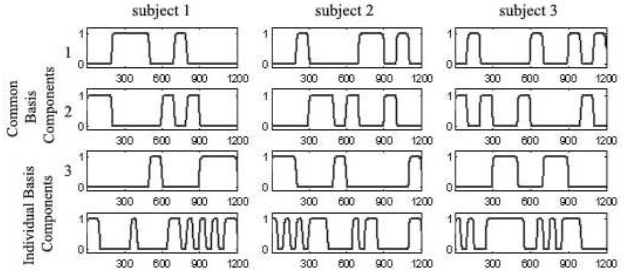



Figure 4: From left to right, encoding matrices $\boldsymbol{S}^{(1)}$, $\boldsymbol{S}^{(2)}$, and $\boldsymbol{S}^{(3)}$ are displayed, each of which is in $\mathbb{R}^{4\times 1200}$. Time courses involving on-off activations constitute the rows of $\boldsymbol{S}^{(l)}$. The first three rows are encodings associated with common bases, which are orthogonal each other since a single subject cannot perform two tasks simultaneously. The last row is encodings corresponding to individual bases, which are randomly activated over time.

Figure 5: Exemplary generated data.

Bases determined by Hierarchical-NMF, One-NMF, FFX-NMF, and GNMF are shown in Figs. 6, 8, and 9, respectively. Both Hierarchical-NMF and GNMF successfully restore common bases as well as individual bases. However, determining common bases by a hierarchical clustering is not an easy job and for real-world data it does not work well.

## 4 NUMERICAL EXPERIMENTS

Numerical experiments for EEG classification are conducted using the dataset V in BCI competition III, which was provided by the IDIAP Research Institute (Blankertz et al., 2006). We compare Hierarchical-NMF, One-NMF, FFX-NMF, and GNMF as spectral feature extraction methods and evaluate the classification performance using the Viterbi algorithm as in (Lee et al., 2007).

### 4.1 IDIAP DATASET

The IDIAP dataset contains EEG data recorded from 3 normal subjects and involves three tasks, including the imagination of repetitive self-paced left/right

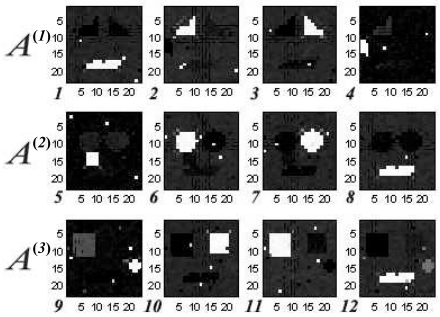


Figure 6: Bases are first determined by applying NMF to $\boldsymbol{X}^{(l)}$ individually in Hierarchical-NMF. Bases are not yet sorted to determine common bases. Each basis is numbered from 1 to 12 on its lower-left corner. Similarities between these bases are shown in Fig. 7, where a agglomerative hierarchical clustering is applied to find groups of common bases.

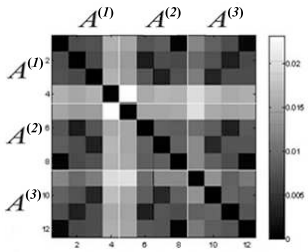
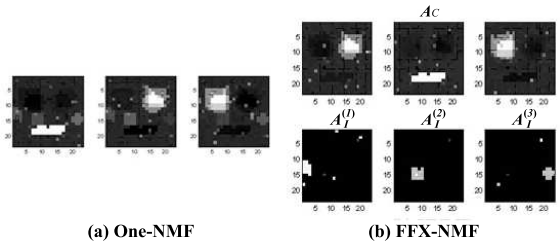


Figure 7: A similarity matrix between 12 bases learned in Hierarchical-NMF is pseudo-colored in such a way that the darker color means a smaller distance (the more similar). A hierarchical clustering yields (1, 8, 12), (2, 6, 11) and (3, 7, 10) as groups. The 4th, 5th, and 9th bases do not have similar bases, which mean they are individual bases.



(a) One-NMF (b) FFX-NMF

Figure 8: Bases learned by One-NMF and FFX-NMF are shown: (a) One-NMF does not capture inter-subject variations, yielding bases averaged over subjects; (b) FFX-NMF also determines the common bases which do not capture inter-subject variations just like One-NMF, while individual bases are well estimated by FFX-NMF.

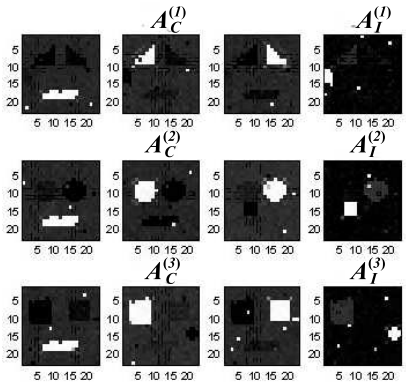


Figure 9: Bases learned by GNMF are shown, where for each subject three common bases and one individual basis are almost identical to the original ones in Fig. 3, implying that GNMF successfully captures task-related intra- and inter-subject variations as well as individual characteristics.

hand movements and the generation of words beginning with the same random letter. EEG data is not splitted in trials, since the subjects are continuously performing any of the mental tasks.

We use the precomputed features which were obtained by the power spectral density (PSD) in the band 8-30 Hz every 62.5 ms, (i.e., 16 times per second) over the last second of data with a frequency resolution of 2 Hz for the eight centro-parietal channels $C_3$, $C_z$, $C_4$, $CP_1$, $CP_2$, $P_3$, $P_z$, and $P_4$.

Spectral components $\bar{P}_{t,f}^{(k)} \in \mathbb{R}^{12\times 10528}$ is the normalized precomputed features satisfying $\sum_f \bar{P}_{f,t}^{(k)} = 1$ for $f \in \{8, 10, \ldots, 28, 30\}$ Hz, $k = 1, \ldots, 8$ (8 different channels), and $t = 1, \ldots, 10528$ where 10528 is the number of data points in the training set.

Then we construct the training data matrix $\boldsymbol{X}_{train} \in \mathbb{R}^{96\times 10528}$ by collecting $12 \times 10528$ spectral matrices computed at 8 different channels,

$$\boldsymbol{X} = \left[\bar{\boldsymbol{P}}^{(1)}; \bar{\boldsymbol{P}}^{(2)}; \ldots; \bar{\boldsymbol{P}}^{(8)}\right] \in \mathbb{R}^{m\times n}, \tag{19}$$

where $m = 12 \times 8$ as shown in Fig. 10. In the same way, we make the test data matrix, $\boldsymbol{X}_{test} \in \mathbb{R}^{96\times 3504}$.

### 4.2 FEATURE EXTRACTION AND CLASSIFICATION

After Hierarchical-NMF, One-NMF, FFX-NMF or GNMF, we obtain the feature matrix $\boldsymbol{S}^{(l)\top} = \left[\boldsymbol{S}_C^{(l)\top} \;\; \boldsymbol{S}_I^{(l)\top}\right]$ for $l = 1, \ldots, L$. Then, we classify $\boldsymbol{S}^{(l)}$ (common and individual features) or $\boldsymbol{S}_C^{(l)}$ (common features only) using the Viterbi algorithm. Total classification accuracies of IDIAP dataset are shown in Table 1.

The classification results of common features (the

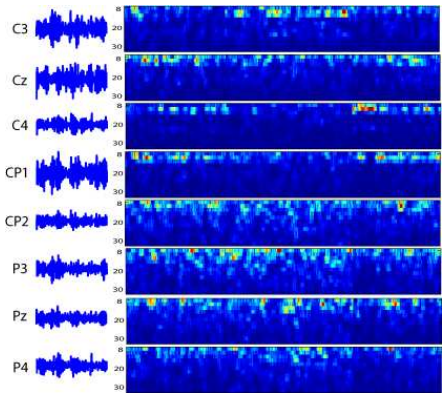



Figure 10: IDIAP dataset are shown in the time-domain (left panel) and in the time-frequency domain (right panel). Waveforms of EEG in the time-domain are shown in the left panel, each of which is measured at 8 different channels. Corresponding time-frequency representations are shown in the right panel, where frequency (horizontal axis in each plot) ranges over $[8, 10, \ldots, 28, 30]$ (i.e., the number of frequency bands is 12). In this case, the data matrix $\boldsymbol{X} \in \mathbb{R}^{96\times n}$ is constructed by collecting 12 frequency profiles at each channel ($96 = 12 \times 8$).

lower half) is better than the results of common and individual features (the upper half). Thus, we can interpret that common bases are really related to tasks. Fig. 11 shows that the classification results are affected by $\alpha$ and $\beta$ which are the degrees of similarity between common bases and dissimilarity between individual bases in Eq. (10). The performance of $\alpha > 0, \beta > 0$ (the solid lines) are better than the performance of $\alpha = \beta = 0$ (the dotted lines) where there is no relation between bases.

By comparing GNMF with One-NMF and FFX-NMF, we can confirm that RFX works better than FFX. However, the results of Hierarchical-NMF which considers RFX is the worst among the proposed methods because the algorithm finding common bases (agglomerative clustering algorithm) is failed for the real data.

We also apply our methods to subject-to-subject transfer, i.e., after calculating basis vectors using only two subjects, we use them to predict for the other subject who is considered as a unseen subject. The results are shown in Table 2. The classification accuracy is somewhat weakened than the case of using all datasets except for the results of subject 3. The result of subject 3 is originally not good. We can infer that the subject 3's data is too noisy to degrade the performance.

Fig. 12 shows the accuracy of subject-to-subject transfer varying $\alpha$ and $\beta$ when $\lambda = 1$. For all transferred cases, the results of GNMF (the solid lines) are better than the results of One-NMF (the dotted lines). Thus,

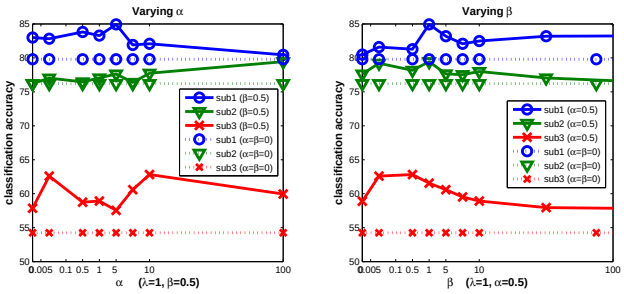



Figure 11: Classification accuracy of GNMF varying $\alpha$ ($\lambda = 1, \beta = 0.5$) (left panel) and $\beta$ ($\lambda = 1, \alpha = 0.5$) (right panel) in (13) and (14). The dotted lines are the results of GNMF when $\alpha = 0$ and $\beta = 0$ which have no assumption between bases of subjects, while the solid lines are the results of GNMF. Lines with circle, triangle and x-mark correspond to the classification accuracy of subjects 1, 2 and 3, respectively.

Table 1: Classification accuracy of Hierarchical-NMF, One-NMF, FFX-NMF, and GNMF. The upper half summarizes classification results when both common and individual features are used and the lower half summarizes results when only common features are used. (Classification accuracy is computed for every single data point (every 0.0625 sec), while the results of BCI competition winner are obtained for every 8 data points (every 0.5 sec).

| | Common + Individual | | | |
|---|---|---|---|---|
| [%] | Sub1 | Sub2 | Sub3 | Avg |
| Hierarchical-NMF | 77.23 | 71.49 | 48.80 | 65.84 |
| One-NMF | 79.48 | **75.49** | 48.51 | 67.83 |
| FFX-NMF | 78.25 | 71.43 | 51.58 | 67.09 |
| GNMF | **79.91** | 70.77 | **54.01** | **68.23** |
| | Common | | | |
| | Sub1 | Sub2 | Sub3 | Avg |
| Hierarchical-NMF | 77.45 | 70.59 | 51.12 | 66.39 |
| One-NMF | - | - | - | - |
| FFX-NMF | 81.76 | **78.17** | 51.09 | 70.34 |
| GNMF | **84.39** | 77.56 | **61.58** | **74.51** |
| BCI comp. winner | 79.60 | 70.31 | 56.02 | 68.65 |

Table 2: Classification accuracy by subject-to-subject transfer. After learning the bases using two subjects' data, we obtain two common basis matrices, $\boldsymbol{A}_C^{(l_1)}$ and $\boldsymbol{A}_C^{(l_2)}$. Then, we choose the common basis matrix which have higher averaged classification results for two subjects, and apply it to a new test subject.

| Learning<br>Test | Sub1,Sub2<br>Sub3 | Sub2,Sub3<br>Sub1 | Sub1,Sub3<br>Sub2 |
|---|---|---|---|
| One-NMF | 47.71 | 72.06 | 67.19 |
| FFX-NMF | 51.00 | 76.83 | 73.56 |
| GNMF | **69.12** | **82.68** | **77.22** |

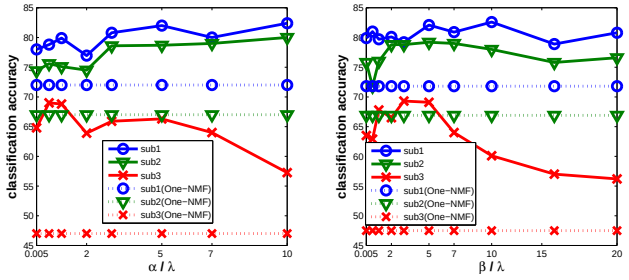



Figure 12: Classifiation accuracy by subject-to- subject transfer varying $\alpha$ ($\lambda = 1$) (left panel) and $\beta$ ($\lambda = 1$) (right panel). Lines with circle, triangle and x-mark correspond to the classification accuracy of subjects 1, 2 and 3 using features learned by the other subjects. The dotted lines are the results of One-NMF, while the solid lines are the results of GNMF.

we conclude that GNMF is not highly sensitive to parameters and is suitable to subject-to-subject transfer.

## 5 CONCLUSIONS

We have presented GNMF which can effectively analyze EEG data of multiple subjects, capturing task-related intra- and inter-subject variations as well as individual characteristics. We have developed multiplicative updates for GNMF, seeking task-related common bases that are as close as possible while keeping individual bases as far apart as possible, in addition to the rank-$r$ approximation as in the standard NMF. Several naive extensions of NMF have also been presented for group analysis. Numerical experiments on IDIAP dataset in BCI competition III confirmed that GNMF improved the EEG classification performance. In addition, we also demonstated that GNMF is useful in subject-to-subject transfer tasks where the prediction for an unseen subject is performed based on a linear model learned from different subjects in the same group. GNMF involves several tuning parameters such as $\lambda$, $\gamma$, $\alpha$ and $\beta$. A Bayesian method (Salakhutdinov & Mnih, 2008) is useful to automatically determine such parameters and the application of it is our future work.

**Acknowledgments:** This work was supported by KOSEF Basic Research Program (grant R01-2006-000-11142-0), National Core Research Center for Systems Bio-Dynamics, and KOSEF WCU Program (Project No. R31-2008-000-10100-0).

## References

Blankertz, B., Müller, K. R., Krusierski, D. J., Schalk, G., Wolpaw, J. R., Schlögl, A., Pfurtscheller, G., & Birbaumer, N. (2006). The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *14*, 153–159.

Ebrahimi, T., Vesin, J. F., & Garcia, G. (2003). Brain-computer interface in multimedia communication. *IEEE Signal Processing Magazine*, *20*, 14–24.

Frackowiak, R., Ashburner, J. T., Penny, W. D., & Zeki, S. (2003). *Human brain function*. Academic Press. 2 edition.

Holmes, A. P., & Friston, K. J. (1998). Generalisability, random effects and population inference. *NeuroImage*, *7*, S754.

Lal, T. N., Schroder, M., Hinterberger, T., Weston, J., Bogdan, M., Birbaumer, N., & Schölkopf, B. (2003). *Support vector channel selection in BCI* (Technical Report 120). Max Planck Institute for Biological Cybernetics.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*, 788–791.

Lee, H., Cichocki, A., & Choi, S. (2006). Nonnegative matrix factorization for motor imagery EEG classification. *Proceedings of the International Conference on Artificial Neural Networks (ICANN)*. Athens, Greece: Springer.

Lee, H., Kim, Y. D., Cichocki, A., & Choi, S. (2007). Nonnegative tensor factorization for continuous EEG classification. *International Journal of Neural Systems*, *17*, 305–317.

Lee, J. H., Lee, T. W., Jolesz, F. A., & Yoo, S. S. (2008). Independent vector analysis (IVA): Multivariate approach for fMRI group study. *NeuroImage*, *40*, 86–109.

Park, H. J., Kim, J. J., Youn, T., Lee, D. S., Lee, M. C., & Kwon, J. S. (2003). Independent component model for cognitive functions of multiple subjects using $[^{15}O]H_2O$ PET images. *Human Brain Mapping*, *18*, 284–295.

Salakhutdinov, R., & Mnih, A. (2008). Bayesian probablistic matrix factorization using MCMC. *Proceedings of the International Conference on Machine Learning (ICML)*. Helsinki, Finland.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, *113*, 767–791.



View publication stats