Nonnegative Matrix Factorization Algorithms using a Constraint to Increase the Discriminability of Two Classes for EEG Feature Extraction

Motoki Sakai
School of Information Environment
Tokyo Denki University
Chiba, Japan

ABSTRACT

Nonnegative matrix factorization (NMF) is an algorithm used for blind source separation. It has been reported that NMF algorithms can be utilized as an effective means to extract features from a motor-imagery related EEG spectrum, which is often used in brain-computer interfaces (BCI). BCI systems enable users to control electrical devices without moving their body parts, and are often tasked with interpreting a user's intentions through motor-imagery related EEG features. In other words, they require EEG signal classification in order to reflect user intentions. In this study, constraints are placed on NMF and kernel NMF (KNMF) algorithms to increase the discriminability between two classes by increasing the energy difference between their potential sources in a spectral EEG signal. To evaluate the proposed algorithms, the IDIAP database, which contains the motor-imagery related EEG spectrum of three subjects, was adopted to test the discrimination between two classes. Using the database, the classification accuracy of the proposed constraint was 75%, which was 7% higher than what was obtained through NMF without a constraint. Similarly, the classification accuracy of KNMF with the proposed constraint was also 4% higher than that of KNMF without a constraint, and reached 78%.

General Terms

Blind source separation

Keywords

Nonnegative matrix factorization, electroencephalogram, brain-computer Interface, feature extraction

1. INTRODUCTION

Nonnegative matrix factorization (NMF) is a type of blind-source separation method [1] that is directly applicable to nonnegative data, such as images, spectrograms, and documents [2, 3, 4], and can decompose a nonnegative input data matrix X into a nonnegative source matrix $U \geq 0$ and dictionary $V \geq 0$, i.e., $X \approx UV^T$. In doing so, NMF employs a non-subtractive and part-based representation of nonnegative data.

It was previously reported that NMF algorithms were effective in extracting features from an electroencephalogram (EEG) for use in a brain-computer interface (BCI) [5, 6]. In an EEG-based BCI, nonnegative spectral EEG features involving motor movement are often used, and a μ rhythm (8–12 Hz) and β rhythm (18–25 Hz) can be cited as motor-imagery related spectral characteristics. It has been noted that these frequency bands decrease during actual or imagined limb movements, which can be observed in the hemisphere opposite the limb. This phenomenon is called event-related desynchonization (ERD). As it stands, ERD might be seen in a different frequency band in each subject. For example, one subject may have an ERD at 8 Hz and another subject at 16–20 Hz. To summarize, the suitable spectral features for BCI applications differ for every user. Wherein, NMF was proposed as a data-driven ERD detection method that accounted for individual differences as it pertains to BCI applications.

Of the several algorithms presented, convex NMF, semi NMF, and kernel NMF (KNMF), KNMF were suggested as effective methods for EEG feature extraction [7]. While standard NMF requires a pseudo-inverse matrix or a fixation of the factor matrix, KNMF overcomes these problems because of its greater feasibility for EEG-based BCI applications. Moreover, the authors in [7] stated that KNMF was superior to other previously presented algorithms with respect to spectral EEG feature extraction.

In accordance with these previous studies, a KNMF algorithm using a constraint to improve classification accuracy was proposed with the aim of applying it to the BCI system in author's previous study [8]. This paper describes NMF and KNMF algorithms with constraints that increase the discriminability of two classes for ERD detection and compares the level of accuracy of the improved NMF algorithms.

2. NONNEGATIVE MATRIX FACTORIZATION ALGORITHMS WITH PROPOSED CONSTRAINTS

2.1 NMF

NMF decomposes an input data matrix $X \in \mathrm{R}_+^{n\times m}$ into the source matrix $U \in \mathrm{R}_+^{n\times r}$ and dictionary $V \in \mathrm{R}^{m\times r}$, and the relation of these matrices can be expressed as follows:

$$X \approx UV^T . \quad (1)$$

In this research, n indicates the number of samples, while m and r refer to the dimension of the input data and source matrices, respectively, and are both set to 96. A pre-computed power spectral density of an EEG signal ($n \times 96$) is regarded as the input data matrix X .

Incorporating assumption (1), the Euclidean distance-based objective function can be written as follows:

$$D_N(W,V) = \frac{1}{2}\left\| X - UV^T \right\|^2 . \quad (2)$$

Typically, (2) is rewritten using a different approach for a slightly more general objective function by taking additional constraints into consideration:

$$D_N(W,V) = \frac{1}{2}\left\| X - UV^T \right\|^2 + \alpha J_V(V) + \beta J_U(U), \quad (3)$$

where α and β are nonnegative regularization parameters and penalty terms $J_V(V)$ and $J_U(U)$ are determined in order to enforce certain application-dependent characteristics.

In this paper, $J_V(V)$ is set to increase the discriminability of two classes. Note that input data matrix X is divided into matrices of class 1, $X_1 \in R_+^{n1\times m}$, and class 2, $X_2 \in R_+^{n2\times m}$ ($n = n1 + n2$). The constraint for NMF is as follows:

$$J_V(V) = -[tr\{(U_2)^T(U_2)\} - tr\{(U_1)^T(U_1)\}]. \quad (4)$$

Given that the source matrix U can be expressed as $X(V^T)^{-1}$, Eq. (4) is rewritten as follows:

$$J_V(V) = -[tr\{(X_2(V^T)^{-1})^T(X_2(V^T)^{-1})\} - tr\{(X_1(V^T)^{-1})^T(X_1(V^T)^{-1})\}]. \quad (5)$$

Eq. (5) indicates the difference between the energy summations of the class 1 source matrix, $U_1 = X_1(V^T)^{-1}$, and class 2 source matrix, $U_2 = X_2(V^T)^{-1}$. Hence, the objective function of the proposed NMF using a constraint to increase the discriminability (NMF_ID) can be rewritten as follows:

$$D_{Nr}(W,V) = \frac{1}{2}\left\| X - UV^T \right\|^2 - \frac{\alpha}{2}[tr\{(X_2(V^T)^{-1})^T(X_2(V^T)^{-1})\} - tr\{(X_1(V^T)^{-1})^T(X_1(V^T)^{-1})\}] + \beta\|U\|. \quad (6)$$

The constraint for source matrix *U*, l_1-norm was selected, as it is typically used to increase sparsity. The extremization of (6) containing the constraint written in (5) solves *V*, which maximizes the ratio of energy between U_1 and U_2. This idea was inspired by the common spatial pattern (CSP). The CSP learns spatial filters that can maximize the variance of a dataset from one class while minimizing the variance of datasets from other classes [9, 10], and the proposed method is expected to produce similar results while guaranteeing the sparsity of source matrices U_1 and U_2.

The partial derivations of Eq. (6) with respect to dictionary *V* and source matrix *U* are as follows:

$$\frac{\partial D_{Nr}}{\partial V} = VU^TU - X^TU - \alpha(X_2^TX_2 - X_1^TX_1)(V^T)^{-1} \quad (7)$$

$$\frac{\partial D_{Nr}}{\partial U} = UV^TV - XV + \beta. \quad (8)$$

To optimize dictionary V, a multiplicative update rule of NMF_ID is generated using the standard gradient (SG),

$$V^{new} = V \bullet \frac{X^TU + \alpha(K_2 - K_1)(V^T)^{-1}}{VU^TU + \delta} \quad (9)$$

where $\bullet$ represents component-wise multiplication, δ is a small positive constant used to avoid numerical instabilities, and $K_C = X_C^TX_C$ ($c = 1,2$). In a similar way, source matrix U is updated using the SG,

$$U^{new} = U \bullet \frac{XV - \beta}{UV^TV + \delta}. \quad (10)$$

2.2 KNMF

KNMF is based on a convex NMF, which constrains dictionary vectors into a convex combination of the input data matrix, $X \in R_+^{n\times m}$ [11]. Source matrix $U \in R_+^{n\times r}$ can be described as follows:

$$U = XW \quad (11)$$

where each column in factor matrix $W \in R_+^{m\times r}$ satisfies a sum-to-one constraint. Incorporating assumption (11), the objective function for the KNF algorithm can be written as follows:

$$D_K(W,V) = \frac{1}{2}\left\| X - XWV^T \right\|^2. \quad (12)$$

The column vectors in dictionary $V \in R^{m\times r}$ are called bias vectors, which indicate that r representative spectral features are generated from training the EEG data samples. (In this research, $r = 96$.)

To improve the performance of KNMF, the objective function described in Eq. (12) can also be rewritten to consider additional constraints:

$$D_K(W,V) = \frac{1}{2}\left\| X - XWV^T \right\|^2 + \alpha J_W(W) + \beta J_V(V), \quad (13)$$

where α and β are nonnegative regularization parameters and constraints $J_W(W)$ and $J_V(V)$ are added for the factor matrix W and the dictionary V, respectively.

To increase the discriminability of two classes using KNMF, $J_W(W)$ should be devised, because factor matrix W is related more directly to an estimation of the source matrix than dictionary V. Note that input data matrix *X* is divided into matrices of class 1, X_1, and class 2, X_2

($n = n1 + n2$), and the source matrices can be expressed as $U_1 = X_1 W$ and $U_2 = X_2 W$ using the common factor matrix W . The proposed constraint for KNMF is as follows:

$$J_W(W) = -[tr\{(X_2W)^T(X_2W)\} - tr\{(X_1W)^T(X_1W)\}] \quad . \quad (14)$$

Eq. (14) also refers to the difference of energy summations between the class 1 source matrix $U_1 = X_1 W$, and class 2 source matrix $U_2 = X_2 W$. Hence, KNMF with a proposed constraint to increase the discriminability (KNMF_ID) models can be rewritten as follows:

$$D_{Kr}(W,V) = \frac{1}{2}\left\|X - XWV^T\right\|^2 - \frac{\alpha}{2}[tr\{(X_2W)^T(X_2W)\} - tr\{(X_1W)^T(X_1W)\}] + \beta\|V\| \cdot \quad (15)$$

In the same way as NMF, l_1 -norm is selected as a constraint for the dictionary V .

To minimize the objective function (15), D_{Kr} is partially differentiated with respect to the factor matrix W and dictionary V as follows in (16) and (17).

$$\frac{\partial D_{Kr}}{\partial W} = X^T X \; WV^T V - X^T X \; V - \alpha(X_2^T X_2 - X_1^T X_1)W \quad . \quad (16)$$

$$\frac{\partial D_{Kr}}{\partial V} = VW^T X^T X \; W - X^T X \; W + \beta \quad . \quad (17)$$

In the KNMF algorithm, $K_C = X_C^T X_C$ ($c = 1,2$) is called linear kernel matrix ($K = K_1 + K_2$), and eqs. (16) and (17) are rewritten as follows:

$$\frac{\partial D_{Kr}}{\partial W} = KWV^T V - KV - \alpha \cdot (K_2 - K_1)W \quad (18)$$

$$\frac{\partial D_{Kr}}{\partial V} = VW^T KW - KW + \beta \cdot \quad (19)$$

To optimize the factor matrix W and dictionary V , the SG leads to the following multiplicative update rules for KNMF_ID,

$$W^{new} = W \bullet \frac{KV + \alpha(K_2 - K_1)W}{KWV^T V + \delta} \quad (20)$$

$$V^{new} = V \bullet \frac{KW - \beta}{VW^T KW + \delta} \quad . \quad (21)$$

In this research, a linear kernel $K = X^T X$ is adopted. Generally, we can select kernels such as $K = \Phi^T \Phi$, which are expressed as an inner product of Φ that maps data matrix X into a feature space.

3. EVALUATION AND RESULTS

3.1 Evaluation of IDIAP database

To evaluate the KNMF_ID-based EEG feature extraction, a database from the BCI competition III provided by the IDIAP Research Institute (Silvia Chiappa, José del R. Millán) was used [12].

In this database, EEG data from three subjects was recorded during four non-feedback sessions. Each subject sat in a normal chair, rested and relaxed their arms on their legs, and performed the following three tasks.

i. Imagine repetitive self-paced left-hand movements

ii. Imagine repetitive self-paced right-hand movements

iii. Generate words beginning with the same random letter

Four 4-min sessions were conducted on the same day, with 5-10 min breaks between each session. The subjects performed each task for approximately 15 s. The tasks were switched randomly between one another at the request of the operator.

In this research, pre-computed signals from recorded EEG data were used, which were obtained as the power spectral density (PSD) in the 8–30 Hz band every 62.5 ms with a 2 Hz frequency resolution for eight EEG electrodes: C3, Cz, C4, CP1, CP2, P3, Pz, and P4. As a result, the number EEG sample dimensions is (8 channels) × (12 frequency components) = 96.

For the BCI competition III, a discrimination of three classes, i, ii, and iii, is required. However, in our study we utilized this database for the discrimination of only two classes. Henceforth, class i and classes ii and iii will be called the target (class 1, described in section II) and non-target (class 2) classes, respectively.

3.2 Preprocessing

The training data matrix $X \in \mathrm{R}_+^{(n1+n2)\times 96}$ can be expressed as $X = [P^{(1)}, P^{(2)}, ..., P^{(k)}]$. $P^{(k)} = [p_{t,f}^{(k)}] \in \mathrm{R}_+^{(n1+n2)\times 12}$ is a spectral matrix in the k-th electrode (k = 1, 2,…, 8). Time and frequency indices run over t = 1,...,(n1 + n2 = 10528), and $f \in \{8,10,12,...,30\}$. Before multiplicative update rules of (9) and (11) (or those of (20) and (21)) are performed, the spectral matrix is normalized as follows:

$$\overline{P}_{t,f}^{(k)} = \frac{P_{t,f}^{(k)}}{\sum_f P_{t,f}^{(k)}} \cdot \quad (22)$$

Test data matrix $X_{test} \in \mathrm{R}_+^{3504\times 96}$ is also normalized by (22) before decomposition.

3.3 Features and classifier

To discriminate the target class from the non-target class, the liner discriminant (LD) classifier was adopted. The LD classifier was generated using features calculated as a maximum value every 0.5 s in each dimension of the source matrix $U_{training}$, as other features, such as the mean, standard deviation, median, or minimum value, did not yield better results in comparison to the maximum value. For the test dataset, the features were calculated in the same manner as the training dataset, and the correct classification rate was then evaluated for each subject.

3.4 Results

Table 1 shows the classification accuracies obtained through two methods: the NMF without an applied constraint (NMF), the proposed NMF_ID, the KNMF without an applied constraint (KNMF) and the proposed KNMF_ID.

Figures 1 and 2 illustrate the $U^T U$ (96×96) matrices of subject 1 calculated by KNMF and KNMF_ID, respectively. Figures 1 (a) and 2 (a) show the matrices of the target classes, and Figures 1 (b) and 2 (b) depict those of the non-target classes. Figures 3 and 4 present the $U^T U$ (96×96) matrices of subject 1 computed by NMF and NMF_ID, respectively. The colors shown in Figures 1 through 4 represent the energy intensity; brown indicates higher energy, whereas blue indicates lower energy.

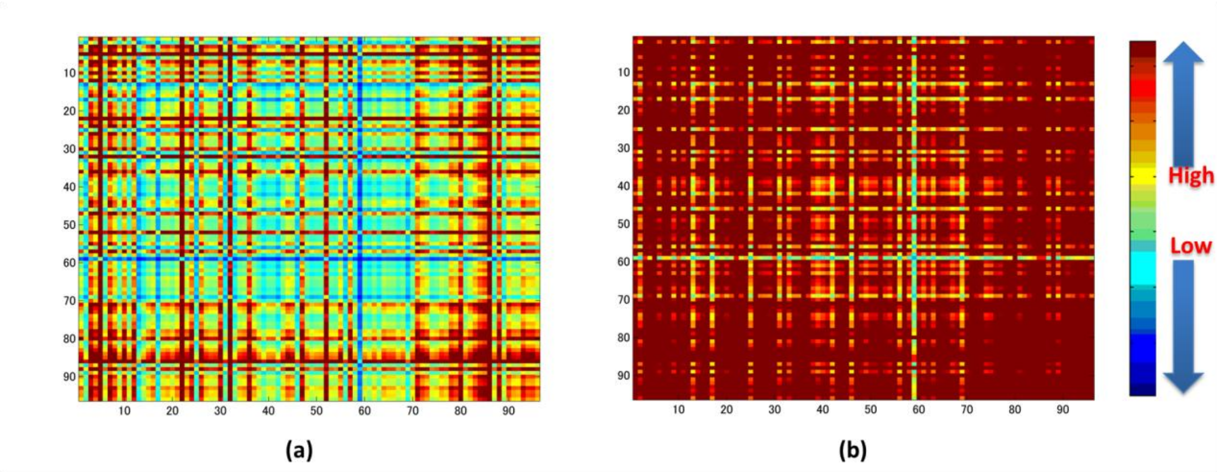

Figure 1: $U^T U$ (96×96) **matrices of subject 1 obtained by KNMF: (a) target; (b) non-target,** $tr(U_1^T U_1)/tr(U_2^T U_2) = 0.913$

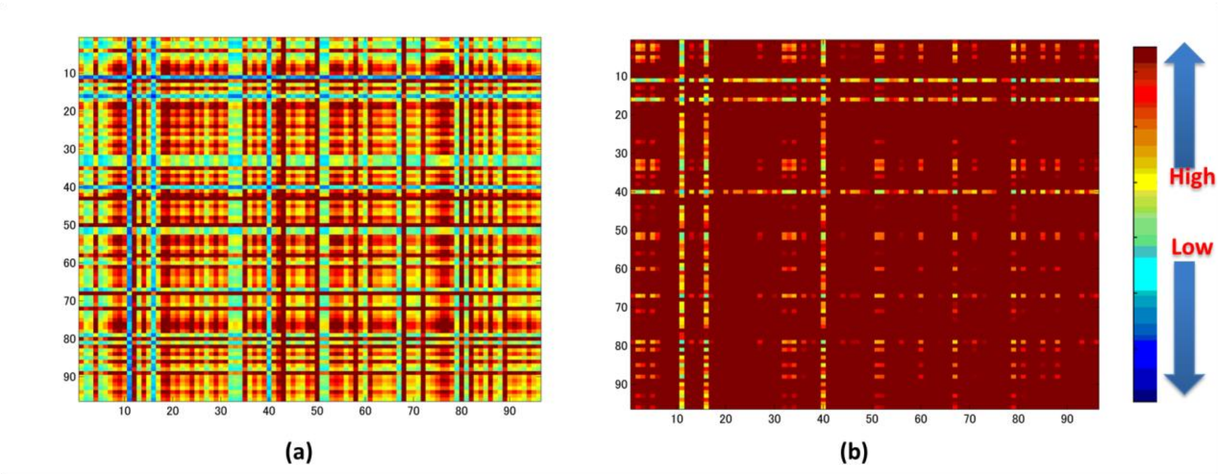

Figure 2: $U^T U$ (96×96) **matrices of subject 1 obtained by KNMF_ID: (a) target; (b) non-target,** $tr(U_1^T U_1)/tr(U_2^T U_2) = 0.788$

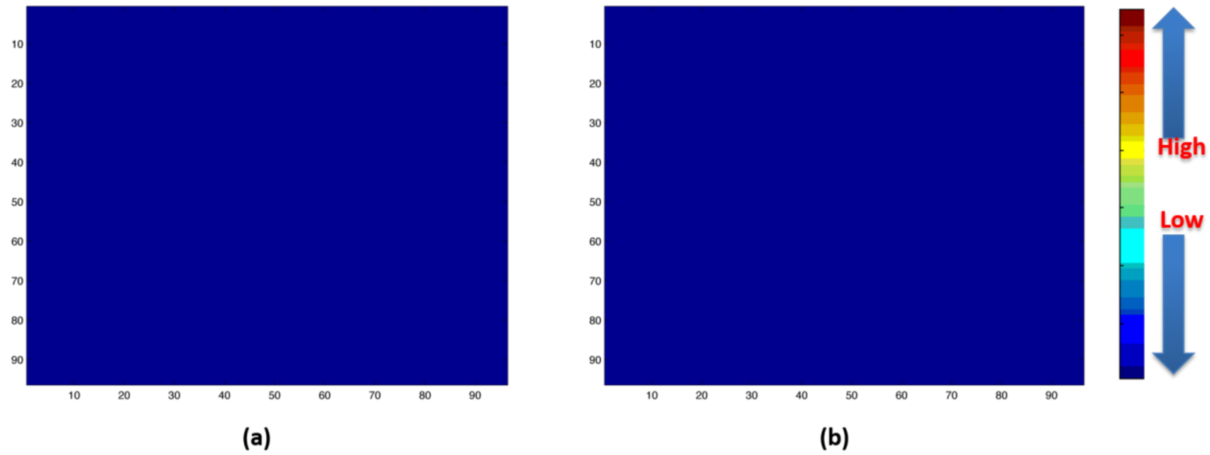

Figure 3: U^TU (96×96) **matrices of subject 1 obtained by NMF: (a) target; (b) non-target,** $tr(U_1^TU_1)/tr(U_2^TU_2)=0.939$

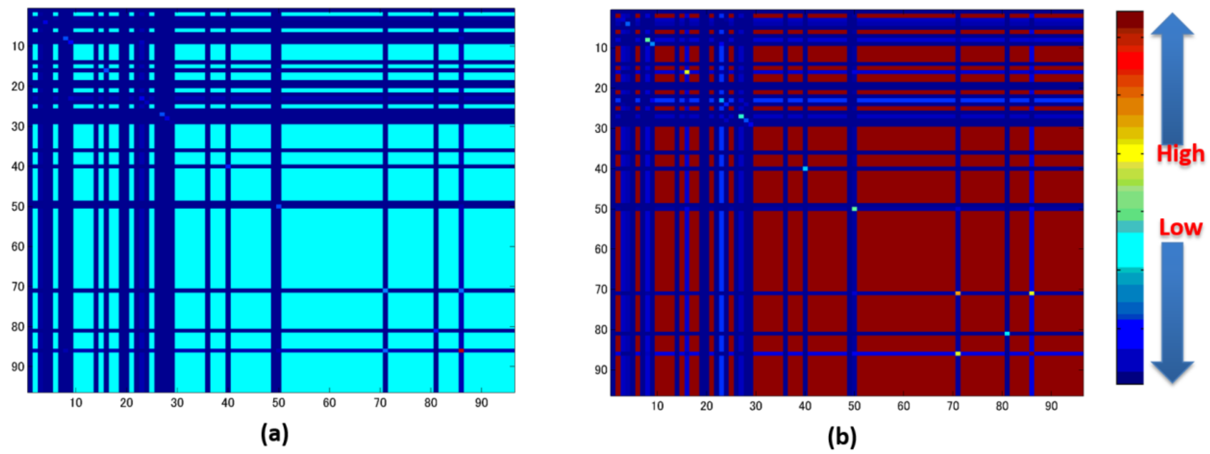

Figure 4: U^TU (96×96) **matrices of subject 1 obtained by NMF_ID: (a) target; (b) non-target,** $tr(U_1^TU_1)/tr(U_2^TU_2)=0.7895$

4. DISCUSSION

NMF algorithms are a more effective than algorithms that have been previously proposed for the extraction and classification of spectral EEG features. The goal of this research was to effectively discriminate two classes characterized by motor imagery EEG spectrums, and to improve conventional NMF algorithms using a constraint that maximizes the difference between the spectral energies of both classes. Of the several NMF algorithms, NMF and KNMF were the focus of this research.

As shown in Table 1, the classification accuracy yielded through NMF_ID was approximately 75% on average, which was 7% higher than that of normal NMF. It should be noted that this accuracy was also higher than that of normal KNMF. This result indicates the effectiveness of the proposed constraint, as previous research [7] shows that KNMF is superior to NMF for EEG feature extraction. Additionally, the classification accuracy obtained by KNMF_ID was approximately 78% on average, which was 4% higher than those obtained by KNMF without a constraint.

Figures 1 and 2 show that KNMF_ID results in higher energies for the decomposed source matrices of both the target and non-target classes compared to KNMF without a constraint. However, the ratio between the trace of source matrices $U_1^TU_1$ and $U_2^TU_2$ show that KNMF_ID can increase the difference between the energies of the decomposed source matrices of the two classes compared to KNMF without a constraint. While, in Figures 3 and 4, though NMF_ID also increases energies for the decomposed source matrices of both the target and non-target classes compared to NMF without a constraint, the effectiveness of the proposed constraint in NMF is visually clearer than that of KNMF. NMF_ID also descended the energy ratio between the trace of source matrices $U_1^TU_1$ and $U_2^TU_2$ at the same level as KNMF_ID.

From these results, an increase in the energy difference between the two classes can be expected to contribute to a more accurate classification of motor imagery EEG spectrums. The influence of the constraint was verified in both NMF and KNMF, and it can be concluded that KNMF_ID was superior to NMF_ID in respect to classification accuracy. In future work, the contribution of proposed constraint will be evaluated further using simulated data and other databases.

Table 1. Classification accuracies

Method	Subject1	Subject2	Subject3	Average
NMF	71.00	66.82	66.06	67.96
NMF_ID	78.54	75.58	69.50	74.54
KNMF	83.10	74.42	65.60	74.37
KNMF_ID	85.37	77.19	71.33	77.96

5. CONCLUSION

This study presented NMF and KNMF algorithms with effective constraints used to classify motor imagery EEG data. The constraint was added to improve the discriminability between two classes during motor imagery tasks. A linear discriminant classifier was used to discriminate the two classes, and the IDIAP database was used to evaluate the proposed constraints,. After the source matrices were computed using NMF_ID and KNMF_ID, the energy differences between the source matrices of the target and non-target classes were actually increased. As a result, the classification accuracies obtained by the proposed methods were higher than those of NMF and KNMF without a constraint. This indicated that the constraint effectively increased the discriminability between two classes.

6. ACKNOWLEDGMENTS

The author would like to thank Dr. Shuxue Ding and Dr. Yuichi Okuyama from the University of Aizu for their recommendations regarding research into the NMF algorithm as it pertains to BCI applications.

7. REFERENCES

[1] D. D. Lee and H. S. Seung, "Algorithms for Non-negative Matrix Factorization," Advances in Neural Information Processing Systems, pp. 556–562, 2000

[2] L. Jing and C. Zhang, "Supervised NMF-Based Image Classification and Annotation," IEEE Transactions on Image Processing, vol. 21, issue 11, pp. 4508–4521, 2012

[3] T. Virtanen, "Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria," IEEE Transactions on Audio, Speech, and Language Processing, vol. 15 , issue 3, pp. 1066–1074, 2007

[4] S. Essid and C. Fevotte, "Smooth Nonnegative Matrix Factorization for Unsupervised Audiovisual Document Structuring," IEEE Transactions on Multimedia, vol. 15 , issue 2, pp. 415–425, 2013

[5] H. Lee, Y. Cichocki, and S. Choi, "Nonnegative Matrix Factorization for Motor Imagery EEG Classification," ICANN'06 Proceedings of the 16th international conference on Artificial Neural Networks, vol. 2, pp. 250–259, 2006

[6] H. Lee, A. Cichocki, and S. Choi, "CUR+NMF for learning spectral features from large data matrix," Proceedings of IEEE International Joint Conference on Neural Networks 2008, 1592–1597, 2008

[7] H. Lee, A. Cichocki, and S. Choi, "Kernel nonnegative matrix factorization for spectral EEG feature extraction," Neurocomputing, vol. 72, pp. 3182–3190, 2009

[8] M. Sakai, "Kernel Nonnegative Matrix Factorization with Constraint Increasing the Discriminability of Two Classes for the EEG Feature Extraction," The proceedings of The 9th International Conference on SIGNAL IMAGE TECHNOLOGY & INTERNET BASED SYSTEMS, Dec. 2-5, 2013

[9] Q. Novi and T.H. Dat, "Sub-band Common Spatial Pattern (SBCSP) for Brain-Computer Interface," Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering," pp. 204–207, 2007

[10] F. Lotte and G. Cuntai, "Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms," IEEE Transactions on Biomedical Engineering, vol. 58, issue 2, pp. 355–362, 2011

[11] C. Ding, T. Li, and J. Michael, "Convex and Semi-Nonnegative Matrix Factorizations," IEEE Transactions on Pattern Analysis and Machine Intelligence archive, vol. 32, issue 1, pp. 45–55, 2010

[12] J. Millan, "On the need for on-line learning in brain-computer interfaces," Proceedings of IEEE International Joint Conference on Neural Networks 2004, pp. 2877–2882, 2004