34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

A Bayes Optimal Matrix-variate LDA for Extraction of Spatio-Spectral Features from EEG Signals

Mohammad Shahin Mahanta, Amirhossein S. Aghaei, Konstantinos N. Plataniotis

***Abstract*— Classification of mental states from electroencephalogram (EEG) signals is used for many applications in areas such as brain-computer interfacing (BCI). When represented in the frequency domain, the multichannel EEG signal can be considered as a two-directional spatio-spectral data of high dimensionality. Extraction of salient features using feature extractors such as the commonly used linear discriminant analysis (LDA) is an essential step for the classification of these signals. However, multichannel EEG is naturally in matrix-variate format, while LDA and other traditional feature extractors are designed for vector-variate input. Consequently, these methods require a prior vectorization of the EEG signals, which ignores the inherent matrix-variate structure in the data and leads to high computational complexity. A matrix-variate formulation of LDA have previously been proposed. However, this heuristic formulation does not provide the Bayes optimality benefits of LDA. The current paper proposes a Bayes optimal matrix-variate formulation of LDA based on a matrix-variate model for the spatio-spectral EEG patterns. The proposed formulation also provides a strategy to select the most significant features among the different rows and columns.**

I. Introduction

Electroencephalogram (EEG) signals are extensively used for development of brain computer interface (BCI) systems. A BCI is an interface to control external devices using electrical activities of the brain. BCI systems are mainly used to help disabled individuals, but can also be used to assist healthy individuals in performing highly demanding tasks or navigation in virtual environments. This paper focuses on spontaneous BCI systems that are based on decoding motor imagery tasks using noninvasive EEG signals.

Spatial and spectral characteristics of EEG signals are widely used in BCI systems to classify motor imagery tasks [1], [2]. These systems operate on the power spectrum of multichannel EEG which can be represented as a matrix-variate[1] spatio-spectral pattern $\mathbf{X}\in\mathbb{R}^{m\times n}$, with each element X_{ij} corresponding to the i^{th} frequency component of the j^{th} EEG channel (electrode). However, there are significant spatial and spectral correlations in the EEG signal. The correlated components result in redundant data dimensions which pose a challenge for parameter estimation and classification tasks. Therefore, a feature extractor is needed to extract a set of uncorrelated features along the channels and frequencies. Conventional feature extractors assume *vector-variate* data, while the spatio-spectral EEG patterns are inherently a *matrix-variate* data: $\mathbf{X}_{m\times n}$. In this paper, we propose a new algorithm for extraction of discriminant features from the matrix-variate spatio-spectral EEG patterns.

One of the most commonly used feature extraction algorithms is the linear discriminant analysis (LDA) which provides a simple non-iterative linear solution applicable to the general multiclass case. Yet, LDA also assumes vector-variate input. The most trivial approach to apply LDA on matrix-variate EEG data is to vectorize the data through concatenation of the columns (or rows) of $\mathbf{X}_{m\times n}$ matrix [3]. However, breaking the EEG matrix along the rows or columns ignores the inherent structure along that dimension, and hence introduces unnecessary degrees of freedom in the design of the feature extractor.

There have been many attempts to avoid the challenges of the vectorial approach by introducing a heuristic matrix-based variant of LDA. A basic two-directional matrix-based LDA applies LDA sequentially on the columns and rows of the matrix [4]. However, it has been indicated that such sequential approach leads to unnecessary information loss [5]. To solve this problem, [6] has introduced an intuitive two-sided matrix-variate LDA, called 2DLDA hereby, which is widely-used in the context of image processing. The 2DLDA method alternates iteratively between a row-wise and a column-wise LDA step.

The 2D approach can utilize the inherent matrix structure of the data. Furthermore, it provides computational efficiency by breaking down the computations into the dimensionality of columns m and rows n; whereas, the vectorial approach suffers from a high computational complexity by operating on the space with the high dimensionality mn for the concatenated matrix. However, it has been shown that the existing 2DLDA method provides a generally sub-optimal solution compared to the Bayes-optimal LDA [7].

This work proposes a novel matrix-variate LDA solution with two contributions: First, we adopt a theoretical approach based on the Bayes optimality or minimal sufficiency of the extracted features. Thus, motivated by the sufficiency of LDA features and using a matrix-variate data model, a Bayes-optimal matrix-variate version of LDA is developed. Second, the proposed formulation can determine the best set of extracted features to be selected as a feature *vector* of arbitrarily reduced dimension. In contrast, the existing 2DLDA method was restricted to extraction of a feature *matrix*, and moreover, failed to specify any priority between the row or column dimensionality of that feature matrix.

The authors are with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Canada. Emails: {mahanta, aghaei, kostas}@comm.utoronto.ca

[1]In this paper, scalars, vectors, and matrices are respectively shown in regular lowercase/uppercase (e.g. a or A), boldface lowercase (e.g. $\mathbf{a}$), and boldface uppercase (e.g. $\mathbf{A}$). Trace of $\mathbf{A}$ is denoted by $\mathrm{tr}(\mathbf{A})$. Also, the Kronecker product of the matrices $\mathbf{A}$ and $\mathbf{B}$ is denoted as $\mathbf{A}\otimes\mathbf{B}$.

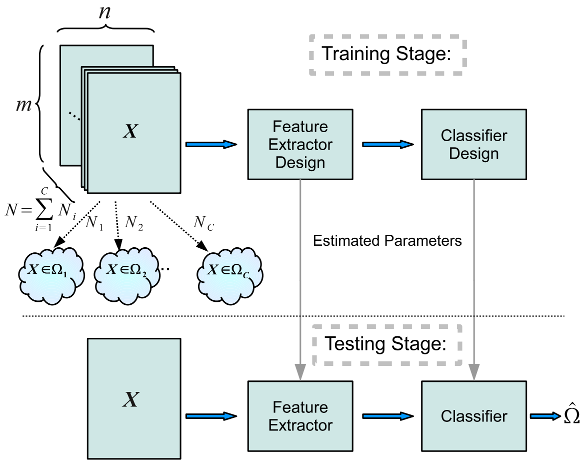

Fig. 1: Outline of the training and testing stages of the EEG classification system.

II. Matrix-Variate Gaussian Model for EEG Data

The desired EEG classification system is demonstrated in Fig. 1. The BCI system classifies each spatio-spectral pattern $\mathbf{X}$ into one of the classes Ω_i, $1 \leq i \leq C$, corresponding to different BCI tasks. The design target is to minimize the probability of classification error, i.e. maximize the probability of correct classification. Each class Ω_i is characterized using a likelihood density $f(\mathbf{X}|\Omega_i)$ and non-zero prior probability $P(\Omega_i)$ which are estimated based on the training samples. There are N_i training samples $\mathbf{X}_{ij}, 1 \leq j \leq N_i$, available for class Ω_i.

A. Homoscedastic Matrix-variate Gaussian Model

The likelihood of each class Ω_i, $f(\mathbf{X}|\Omega_i)$, is modeled as a matrix-variate homoscedastic Gaussian distribution [8]:

$$f(\mathbf{X}|\Omega_i) = \mathcal{N}(\mathbf{M}_i, \mathbf{\Phi}, \mathbf{\Psi}) \tag{1}$$

where the matrices $\mathbf{M}_i, 1 \leq i \leq C$, denote the class means. Matrix $\mathbf{\Phi}$ is the spectral covariance, also called column-wise or left covariance, and matrix $\mathbf{\Psi}$ is the spatial covariance, also called row-wise or right covariance. Therefore, knowledge of the parameters $\mathbf{M}_i$, $\mathbf{\Phi}$, and $\mathbf{\Psi}$ will suffice to determine the likelihood functions. Note that here the spectral or spatial covariances of different classes are assumed to be the same for all the classes: $\mathbf{\Phi}_i = \mathbf{\Phi}$ and $\mathbf{\Psi}_i = \mathbf{\Psi}$, with the per-class quantities defined as:

$$\mathbf{\Phi}_i = \mathrm{tr}^{-1}(\mathbf{\Psi}_i) * \mathrm{E}_{\mathbf{X}|\Omega_i}((\mathbf{X}-\mathbf{M}_i)(\mathbf{X}-\mathbf{M}_i)^T), \tag{2}$$

$$\mathbf{\Psi}_i = \mathrm{tr}^{-1}(\mathbf{\Phi}_i) * \mathrm{E}_{\mathbf{X}|\Omega_i}((\mathbf{X}-\mathbf{M}_i)^T(\mathbf{X}-\mathbf{M}_i)), \tag{3}$$

Multivariate Gaussianity and homoscedasticity are fairly common practical assumptions for EEG signals [9] as implied by utilization of relevant methods such as LDA. Furthermore, the matrix-variate model in (1) corresponds to a specific structure for the covariance of the vectorized data, as follows. Assuming a column concatenation operation $\mathrm{vec}(.)$, denote the vectorized data as $\mathbf{x}_{mn\times 1} = \mathrm{vec}(\mathbf{X})$. Then, the mean of $\mathbf{x}$ in Ω_i equals $\boldsymbol{\mu}_i = \mathrm{vec}(\mathbf{M}_i)$, and assuming that (1) holds, the class-conditional covariance of $\mathbf{x}$ equals

$$\mathbf{\Sigma}_{mn\times mn} = \mathbf{\Psi}_{n\times n} \otimes \mathbf{\Phi}_{m\times m}. \tag{4}$$

Therefore, the matrix-variate Gaussianity implies a *separable* structure for the covariance matrix of the vectorized data as defined by (4).

The above separability property intuitively means that the variance between two elements of the matrix-variate data $\mathbf{X}_{m\times n}$ can be decomposed into an inter-row and an inter-column component. The simulation results for our method suggest that this property is a reasonable practical approximation for the spatio-spectral EEG patterns.

B. Separable Between-Class Scatter Matrix

The vectorial between-class scatter matrix is also modeled to be separable into matrix-variate counterparts:

$$\mathbf{S}_B = \mathbf{S}_{BR} \otimes \mathbf{S}_{BL}, \tag{5}$$

where $\mathbf{S}_B = \sum_{i=1}^{C} P(\Omega_i)(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T$, and

$$\mathbf{S}_{BL} = \sum_{i=1}^{C} P(\Omega_i)(\mathbf{M}_i - \mathbf{M})(\mathbf{M}_i - \mathbf{M})^T, \tag{6}$$

$$\mathbf{S}_{BR} = \sum_{i=1}^{C} P(\Omega_i)(\mathbf{M}_i - \mathbf{M})^T(\mathbf{M}_i - \mathbf{M}). \tag{7}$$

The assumption of (5) enables us to derive a Bayes optimal matrix-variate LDA solution. In general, as demonstrated by the experimental results, this assumption can be considered as a reasonable simplification that results in more computational efficiency and robust parameter estimation.

III. Current Methods

A *Bayes optimal* feature extractor yields a minimal set of features that in theory can be used to classify the data with the optimal accuracy determined by the Bayes error [10]. Such features contain all the discriminatory information of the original data. Mathematically, the optimal Bayes error performance can be achieved if and only if the feature extractor provides a sufficient statistic for Ω_i [10], [11]. Among linear feature extractors, LDA is a Bayes optimal solution for homoscedastic vector-variate Gaussian data [11].

For data with restricted matrix-variate structure, a heuristic two-directional LDA (2DLDA) has been proposed [6] that can deploy the matrix structure of the data. This method can reduce computational complexity, and results in more stable parameter estimates in small-sample-size scenarios. However, 2DLDA is not a Bayes optimal solution as LDA, and therefore theoretically yields higher Bayes error values. Furthermore, the number of extracted features in 2DLDA is always a product of two integers, qr, where both the number of rows q and columns r of the feature matrix need to be decided by the user. It should also be noted that 2DLDA separately deals with *spectral/spatial* covariances, both for within-class covariance and between-class scatter matrices, and hence implicitly assumes separable $\mathbf{\Sigma}$ and $\mathbf{S}_B$. In fact, these conditions are a generalization of the conditions asserted in [7] for the Bayes optimality of 2DLDA in the two-class case.

IV. Proposed Matrix-variate LDA (MLDA)

We will use the matrix-variate model of Section II to find a matrix-variate formulation for the LDA solution. The parameters of this model are estimated using the training data as follows. The prior probabilities $P(\Omega_i)$ can be estimated as the fraction of the training samples belonging to Ω_i, i.e. $\frac{N_i}{N}$. For mean and covariance parameters, maximum-likelihood (ML) estimates are used. The ML estimate of $\mathbf{M}_i$ is the average of the samples in that class. The ML estimates of the spectral and spatial covariances can be calculated using an iterative procedure similar to that of [12]. In each iteration, the pooled estimate of the average covariance in either row or column direction is updated:

$$\mathbf{\Psi} = \frac{1}{mN}\sum_{i=1}^{C}\sum_{j=1}^{N_i}(\mathbf{X}_{ij}-\mathbf{M}_i)^T\mathbf{\Phi}^{-1}(\mathbf{X}_{ij}-\mathbf{M}_i), \quad (8)$$

$$\mathbf{\Phi} = \frac{1}{nN}\sum_{i=1}^{C}\sum_{j=1}^{N_i}(\mathbf{X}_{ij}-\mathbf{M}_i)\mathbf{\Psi}^{-1}(\mathbf{X}_{ij}-\mathbf{M}_i)^T. \quad (9)$$

The two steps are iterated until the Frobenius distance of two consecutive estimates is less than a selected threshold. In the experiment of Section V, an order of 10 iterations were enough to reach convergence with a threshold of 10^{-5}.

The Bayes optimal LDA features for this problem are $\mathbf{y} = \mathbf{T}^T\mathrm{vec}(X)$, where the d columns of $\mathbf{T}_{mn\times d}$ are selected as the most significant eigenvectors of

$$\mathbf{\Sigma}^{-1}\mathbf{S}_B = (\mathbf{\Psi}\otimes\mathbf{\Phi})^{-1}(\mathbf{S}_{BR}\otimes\mathbf{S}_{BL}) \quad (10)$$

In order to calculate these Bayes optimal LDA features using matrix-variate operations, denote the eigenvalues and eigenvectors of $\mathbf{\Phi}^{-1}\mathbf{S}_{BL}$ or $\mathbf{\Psi}^{-1}\mathbf{S}_{BR}$ by λ_i or γ_j and $\mathbf{u}_i$ or $\mathbf{v}_j$ respectively, where $1 \le i \le m$ and $1 \le j \le n$. Let λ_i and γ_j be ordered in descending order. It can be shown that instead of the LDA linear operator $\mathbf{T}$, we can calculate the same features using a bilinear operation consisting of the spectral and spatial linear operators $\mathbf{U}$ and $\mathbf{V}$:

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m], \quad \mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]. \quad (11)$$

The proposed matrix-variate LDA (MLDA) procedure projects $\mathbf{X}$ onto columns of $\mathbf{U}$ and $\mathbf{V}$ to get $\mathbf{Y} = \mathbf{U}^T\mathbf{X}\mathbf{V}$. Then the elements y_{ij} of $\mathbf{Y}$ which correspond to the d largest $\gamma_i\lambda_j$ values are selected and stacked in a $\mathbf{y}$ feature vector. Therefore, this proposed method belongs to the overall category of tensor-to-vector projection methods [13]. Tab. I outlines the pseudo-code for training the MLDA method.

The MLDA solution relies only on the m- or n-dimensional operations. Therefore, the computational complexity of the eigendecomposition step for MLDA is broken down into $O(m^3 + n^3)$, compared to LDA's complexity of $O((mn)^3)$. In MLDA, the two eigendecompositions of order $O(m^3)$ and $O(n^3)$ can be implemented in parallel. Furthermore, the lower-dimensional parameters can be estimated more reliably than the higher-dimensional parameters required by LDA. Finally, unlike similar methods such as 2DLDA, MLDA provides Bayes optimal features for the underlying data model, and also specifies the relative priority of all the extracted features in terms of the corresponding $\gamma_i\lambda_j$ values.

Inputs:
- N_i training samples $\mathbf{X}_{ij}, 1 \le j \le N_i$ for each class Ω_i, $1 \le i \le C$. The total number of samples is N.
- The number of desired extracted features d.

Outputs:
- The feature extraction operators $\mathbf{U}_{m\times m}$ and $\mathbf{V}_{n\times n}$.
- The corresponding γ_i and λ_j values which determine the priority in selecting the elements of the resulting feature matrix.

Procedure:
1) Estimate the parameters $\mathbf{\Phi}$ and $\mathbf{\Psi}$ using a modified version of iterative algorithm of [12], and the parameters $\mathbf{M}_i$, $1 \le i \le C$, and $\mathbf{M}$ as sample means of the classes and their average.
2) Calculate $\mathbf{S}_{BL}$ and $\mathbf{S}_{BR}$ according to (6) and (7).
3) Calculate the eigenvalues λ_i and γ_j and the corresponding eigenvectors $\mathbf{u}_i$, $1 \le i \le m$, and $\mathbf{v}_j$, $1 \le j \le n$, for $\mathbf{\Phi}^{-1}\mathbf{S}_{BL}$ and $\mathbf{\Psi}^{-1}\mathbf{S}_{BR}$ respectively.
4) Construct $\mathbf{U}$ and $\mathbf{V}$ according to (11)

TABLE I: Pseudocode for training the proposed MLDA feature extractor.

V. Simulations

We will compare our proposed feature extraction method, MLDA, against the LDA and 2DLDA methods based on the corresponding classification performance on a BCI data set. To have a fair comparison, a simple linear Gaussian classifier is utilized with these feature extractors.

A. Data Set and Preprocessing

Data set V from BCI competition III [14] is selected which contains EEGs of three normal subjects recorded in four sessions, using 32-electrode Biosemi system at 512Hz sampling rate. Each record consists of sequential 15-second trials of three possible mental imagery tasks: left-hand movement, right-hand movement, and generation of words beginning with a random letter. The last session is used for testing and the rest for training. The goal of competition is to classify the mental task every 0.5 second using only the last second of data. The highest correct classification rate (CCR) for this competition without post-processing is %62.72 [15]. For comparison purposes, three different versions of preprocessed data have been studied in this section:

- 32-Channel Spectrum: A short-time Fourier transform (STFT) with a Hamming window of length one second, with overlapping factor of $\frac{15}{16}$, is applied to the raw data available in the dataset. Then, 12 power spectral features from 8-30Hz (α and β band) with 2Hz resolution are retained. This results in spatio-spectral patterns $\mathbf{X}_{12\times 32}$.
- 8-Channel Spectrum: The 8 centro-parietal channels (C3, Cz, C4, CP1, CP2, P3, Pz, and P4) are recommended by the competition organizers to be correlated to the motor imagery tasks. This prior information is

TABLE II: Classification results for different feature extractors.

Data ($m \times n$)	Method	Subject a			Subject b			Subject c			Avg.
		%CCR	d	$\#Iter$	%CCR	d	$\#Iter$	%CCR	d	$\#Iter$	%CCR
32-Channel Spectrum (12×32)	LDA	56.04	2	-	50.53	2	-	47.45	2	-	51.34
	2DLDA	39.81	22 (2×11)	10	40.56	18 (3×6)	10	38.24	30 (5×6)	10	39.54
	MLDA	65.88	33	13	56.07	5	14	52.55	14	12	58.17
8-Channel Spectrum (12×8)	LDA	61.14	2	-	51.79	2	-	46.56	2	-	53.16
	2DLDA	61.54	88 (11×8)	10	51.79	96 (12×8)	10	46.56	96 (12×8)	10	53.30
	MLDA	64.23	5	11	53.32	42	9	49.68	23	10	55.74
8-Channel Precomputed Spectrum (12×8)	LDA	75.8	2	-	61.52	2	-	52.06	1	-	63.13
	2DLDA	76.48	72 (12×6)	10	61.52	96 (12×8)	10	53.21	11 (11×1)	10	63.74
	MLDA	79.68	9	8	66.82	17	6	54.59	1	7	67.03

used to retain only 8 channels from the 32-Channel spectrum. This results in spatio-spectral patterns $\mathbf{X}_{12\times 8}$.

- 8-Channel Precomputed Spectrum: As part of this data set, a set of precomputed 8-channel spectrum for centro-parietal channels is provided, in which a spatial filtering is applied to the raw data prior to the STFT.

B. Results

Tab. II outlines the correct classification rate (CCR) of our proposed MLDA method for all three versions of the preprocessed data, and compares it to the performance of LDA and 2DLDA methods. From this table, the proposed MLDA method consistently outperforms LDA and the existing 2DLDA method. This result demonstrates the fact that MLDA combines the Bayes optimality advantage of LDA and the reliable matrix-variate estimation of 2DLDA. The improvement with respect to LDA is most evident for the 32-channel data. In this case, the original dimensionality of the data is relatively large, and hence the reliable matrix-variate parameter estimates used by MLDA result in a significant advantage over LDA. Furthermore, the better performance of MLDA in 32-channel data compared to 8-channel data indicates that MLDA performs a better task of channel filtering than the traditional *channel selection* strategy.

It should be noted that 2DLDA has not significantly superseded LDA in any data set, since the matrix-variate advantage of 2DLDA has not compensated its deviation from Bayes optimality. Furthermore, although the optimal number of features for MLDA is larger than that for LDA, it is generally fewer than that for 2DLDA. In cases where only a very low number of features is desirable, a two-stage MLDA and LDA operation can also be an alternative.

The number of iterations of MLDA for the ML estimation of $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is also reported in Tab. II. The convergence threshold for the estimation algorithm is selected as 10^{-5}. It can be observed that consistently for all data sets, convergence is achieved with a moderate number of iterations.

VI. Conclusions

A theoretically Bayes optimal matrix-variate formulation of LDA called MLDA was introduced based on a matrix-variate Gaussian model. The assumed model provides a reasonable approximation for EEG data. And if the model is satisfied, the MLDA solution provides Bayes optimality. Compared to LDA, MLDA provides a reduced computational complexity, allows for possibility of parallel training of spatial/spectral operators, and most importantly, utilizes more reliable parameter estimates. Furthermore, compared to the existing 2DLDA method, the proposed MLDA method can determine the most discriminant features, according to Fisher's criterion, for an arbitrary reduction in the dimension. In particular, the ambiguity of 2DLDA in the selection of the number of extracted rows and columns is obviated in the new formulation.

References

[1] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Mller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Sig. Proc. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[2] C. Neuper and G. Pfurtscheller, "Event-related dynamics of cortical rhythms: Frequency-specific features and functional correlates," *Int. J. Psychology*, vol. 43, no. 1, pp. 41–58, 2001.

[3] A. Subasi, "Selection of optimal AR spectral estimation method for EEG signals using Cramer-Rao bound," *Computers in Biology and Medicine*, vol. 37, no. 2, pp. 183 – 194, 2007.

[4] J. Yang, D. Zhang, X. Yong, and J. yu Yang, "Two-dimensional discriminant transform for face recognition," *Pattern Recognition*, vol. 38, no. 7, pp. 1125 – 1129, 2005.

[5] S. Noushath, G. H. Kumar, and P. Shivakumara, "(2D)2LDA: An efficient approach for face recognition," *Pattern Recognition*, vol. 39, no. 7, pp. 1396 – 1400, 2006.

[6] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis." in *NIPS'04*, 2004.

[7] W.-S. Zheng, J.-H. Lai, and S. Z. Li, "1D-LDA vs. 2D-LDA: When is vector-based linear discriminant analysis better than matrix-based?" *Pattern Recognition*, pp. 2156–2172, 2008.

[8] A. Gupta and D. Nagar, *Matrix variate distributions*. Chapman & Hall/CRC, 1999.

[9] A. Subasi and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8659 – 8666, 2010.

[10] L. Devroye, *A Probabilistic Theory of Pattern Recognition*. Springer, 1996.

[11] M. S. Mahanta, A. S. Aghaei, K. N. Plataniotis, and S. Pasupathy, "Heteroscedastic linear feature extraction based on sufficiency conditions," *Pattern Recognition*, vol. 45, no. 2, pp. 821 – 830, 2012.

[12] P. Dutilleul, "The MLE algorithm for the matrix normal distribution," *Journal of Statistical Computation and Simulation*, vol. 64, no. 2, pp. 105–123, 1999.

[13] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, *A Taxonomy of Emerging Multilinear Discriminant Analysis Solutions for Biometric Signal Recognition*. John Wiley & Sons, Inc.

[14] J. Millan, "On the need for on-line learning in brain-computer interfaces," in *Proc. Int. Joint Conf. on Neural Networks*, vol. 4, July 2004, pp. 2877–2882.

[15] F. Galán, F. Oliva, and J. Guàrdia, "Using mental tasks transitions detection to improve spontaneous mental activity classification," *Med. Bio. Eng. Comput.*, vol. 45, no. 6, pp. 603–609, 2007.