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

Brain-Computer Interfacing in Discriminative and Stationary Subspaces

Wojciech Samek[1,2], Klaus-Robert Müller[1,4], Motoaki Kawanabe[3] and Carmen Vidaurre[1]

Abstract— **The non-stationary nature of neurophysiological measurements, e.g. EEG, makes classification of motion intentions a demanding task. Variations in the underlying brain processes often lead to significant and unexpected changes in the feature distribution resulting in decreased classification accuracy in Brain Computer Interfacing (BCI). Several methods were developed to tackle this problem by either adapting to these changes or extracting features that are invariant. Recently, a method called Stationary Subspace Analysis (SSA) was proposed and applied to BCI data. It diminishes the influence of non-stationary changes as learning and classification is performed in a stationary subspace of the data which can be extracted by SSA. In this paper we extend this method in two ways. First we propose a variant of SSA that allows to extract stationary subspaces from labeled data without disregarding class-related variations or treating class-differences as non-stationarities. Second we propose a discriminant variant of SSA that trades-off stationarity and discriminativity, thus it allows to extract stationary subspaces without losing relevant information. We show that learning in a discriminative and stationary subspace is advantageous for BCI application and outperforms the standard SSA method.**

I. Introduction

In Brain-Computer Interfacing (BCI) [1] one major challenge is to understand the non-stationarities in the signal of interest e.g. EEG and to develop methods that are invariant to those of them that decrease the signal to noise ratio. The sources and time scales of non-stationarities in the signal can be very different e.g. changes in electrode impedance may occur when an electrode gets loose or the skin prepping gel dries out, muscular activity or eye movements lead to artefacts in the signal and we often observe changes of task involvement and attention over the course of an experiment. Additionally, differences between sessions may exist, e.g. the way the stimulus is presented or feedback is provided to the user may be different or the positions of the electrodes may vary slightly.

Several methods were proposed to reduce the impact of non-stationarities in BCI applications. The approaches can be divided into two main groups, namely methods extracting robust or invariant features and approaches adapting to changes in the data. One of the first approaches to extract invariant features was the invariantCSP method [2]. A very recent work addressing the non-stationarity problem on the feature extraction level is [3]. A lot of work on adaptation has been published in the past, e.g. [4] uses techniques for co-adaptive learning of user and machine, [5] applies covariate shift adaptation to account for changes of the features and [6], [7], [8] use other unsupervised adaptation approaches.

Recently, Bünau et al. [9] proposed a novel technique called Stationary Subspace Analysis (SSA) that finds low-dimensional projections having stationary distributions from high-dimensional observations. This method can be applied to EEG data as a preprocessing step in order to extract the stationary part of the signal as done in [10]. The authors showed that restricting the BCI to the stationary sources found by SSA can significantly increase the classification accuracy. However, SSA is a general purpose method and its usage is limited when applying it to multi-class data. The distinctive different tasks can be considered as non-stationary components of the signal (it is expected that the statistical properties of the data change with the task) and therefore disregarded by SSA. Furthermore SSA is an unsupervised method and thus does not differentiate between discriminant and non task-relevant directions. In other words SSA may remove information that is essential for classification in subsequent steps.

In this paper we extend the work of Bünau et al. [10] and propose a method that allows to compute the stationary subspace from multi-class data (groupSSA) without disregarding class-related variations or treating class-differences as non-stationarities. Furthermore we propose a discriminant variant of SSA (dSSA) that trades-off stationarity and discriminativity, thus it allows to extract stationary subspaces without losing relevant information. We analyse the emerging stationarity and non-stationarity patterns obtained from five subjects and show that our method is better suited for BCI data and consequently outperforms SSA.

This paper is organized as follows. In the next section we present SSA and introduce the two extensions. After that in Section III we evaluate the methods on a dataset of five subjects performing motor imagery and analyse the results. We conclude in Section IV with a discussion.

II. Stationary and Discriminative Subspaces

A. Stationary Subspace Analysis Method

Stationary Subspace Analysis (SSA) [9] is a novel method to factorize a high-dimensional multivariate time-series into

[1]W. Samek (wojciech.samek@campus.tu-berlin.de), K.-R. Müller (klaus-robert.mueller@tu-berlin.de), C. Vidaurre (carmen.vidaurre@tu-berlin.de) are with the Berlin Institute of Technology, Franklinstr. 28 / 29, 10587 Berlin, Germany.

[2]W. Samek is with the Bernstein Center for Computational Neuroscience, Philippstr. 13, 10115 Berlin, Germany

[3]M. Kawanabe (kawanabe@atr.jp) is with the Advanced Telecommunications Research Institute International. 2-2-2 Hikaridai, Keihanna Science City, Kyoto 619-0288, Japan.

[4] K.-R. Müller is with the Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea.

*We thank Paul von Bünau for valuable discussions. This work was supported by the German Research Foundation (GRK 1589/1) and the World Class University Program through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology, under Grant R31-10008.

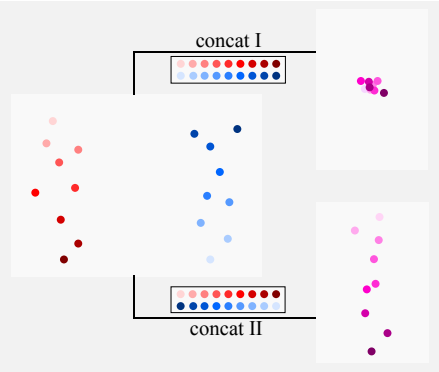

Fig. 1. The red and blue points represent trial-wise covariance matrices from both classes. Time flow is color-coded from bright to dark and both classes have opposing trends in the vertical direction and a stable discriminative horizontal direction. Two concatenation strategies are presented. The concat I method combines trials according to time flow which results in covariance matrices (upper pink points) without a prominent non-stationary direction. The other strategy concat II does exactly the opposite, thus preserves the non-stationarities in the data.

its stationary and non-stationary components. Its underlying assumption is that the observed signal $\mathbf{x}(t)$ is a linear superposition of stationary $\mathbf{s}^{\mathfrak{s}}(t)$ and non-stationary $\mathbf{s}^{\mathfrak{n}}(t)$ sources

$$\mathbf{x}(t) = A\,\mathbf{s}(t) = \begin{bmatrix} A^{\mathfrak{s}} & A^{\mathfrak{n}} \end{bmatrix} \begin{bmatrix} \mathbf{s}^{\mathfrak{s}}(t) \\ \mathbf{s}^{\mathfrak{n}}(t) \end{bmatrix}, \tag{1}$$

and A is an invertible matrix. The goal of SSA is to find a linear transformation $\hat{A}^{-1}$ that separates the $\mathfrak{s}$-sources from the $\mathfrak{n}$-sources. For that the signal $\mathbf{x}(t)$ is divided into epochs and an optimization criterion is employed to recover the sources. More precisely, SSA minimizes the distance measured as Kullback-Leibler Divergence D_{KL}, between the distribution of the estimated $\mathfrak{s}$-sources in each epoch (described by first two moments) and the standard normal distribution.

B. Limitations of SSA for BCI Application

Stationary Subspace Analysis is a general purpose method that is not optimized for application in a BCI setting. Since it is expected that the statistical properties of the data change with the task, one should be very careful when applying SSA to BCI data as it may treat discriminant variations between classes as non-stationarity which needs to be removed. Such preprocessing will negatively affect the classification performance. On the other hand when applying SSA to each class separately, one obtains two different projections that can not be combined in a straight forward manner.

The authors in [10] introduced a different approach to circumvent this problem, they simply concatenate trials (SSAconcat) from opposing classes in order to cancel out the differences between both classes before applying SSA. Note that concatenation of trials is equal to summation of the corresponding covariance matrices (assuming zero means)[1]. However, concatenation may lead to suboptimal results as it ignores task-specific non-stationarities, e.g. when trends in both classes propagate in opposite directions. Figure 1 shows covariance matrices from two classes (red and blue points) with opposing trends in the vertical direction and two different trial combination scheme concat I and concat II. We see that concat I cancels out the prominent non-stationary direction in the data, whereas concat II preserves it.

Note that variations may also occur along discriminative directions e.g. due to learning effects. A method that does not differentiate between different kinds of non-stationarities (task-relevant / not task-relevant) will perform poorly in a classification setting, especially when many directions are removed. Therefore we not only propose a principled approach to treat multi-class data in SSA, but also present a supervised variant that prevents that discriminative directions are removed. We show that learning in a discriminative and stationary subspace is advantageous in a BCI setting.

C. Extension: groupSSA

The idea behind groupSSA is to consider groups of epochs in order to find projections that are as stationary as possible within each group. This does not necessarily imply stationarity across all epochs, however, it allows to combine data from many subjects to conduct group studies and to apply SSA to multi-class data in a principled way. The objective function of groupSSA measures the divergence between epochs and their group averages, thus it can be written as

$$L(R) = \sum_{i=1}^{M} \sum_{j=1}^{N_i} D_{\mathrm{KL}} \left[\mathcal{N}(\hat{\boldsymbol{\mu}}_{ij}^{\mathfrak{s}}, \hat{\Sigma}_{ij}^{\mathfrak{s}}) \,||\, \mathcal{N}(\overline{\boldsymbol{\mu}}_{j}^{\mathfrak{s}}, \overline{\Sigma}_{j}^{\mathfrak{s}}) \right], \tag{2}$$

where M is the number of groups, N_i is the number of epochs in group i, $\mathcal{N}(\hat{\boldsymbol{\mu}}_{ij}^{\mathfrak{s}}, \hat{\Sigma}_{ij}^{\mathfrak{s}})$ is the distribution of epoch j in group i, $\mathcal{N}(\overline{\boldsymbol{\mu}}_{j}^{\mathfrak{s}}, \overline{\Sigma}_{j}^{\mathfrak{s}})$ is the average distribution in group i and R is a rotation matrix. Note that the divergence is computed in the projected stationary subspace. In summary, the goal is to find a rotation (projection), so that the divergence between the distribution of the estimated $\mathfrak{s}$-sources in each epoch $\mathcal{N}(\hat{\boldsymbol{\mu}}_{ij}^{\mathfrak{s}}, \hat{\Sigma}_{ij}^{\mathfrak{s}})$ and the corresponding mean distribution for group j $\mathcal{N}(\overline{\boldsymbol{\mu}}_{j}^{\mathfrak{s}}, \overline{\Sigma}_{j}^{\mathfrak{s}})$ is minimized.

D. Extension: dSSA

In order to extract subspaces that are not only stationary, but also contain discriminative information, one needs to add a discriminativity term to the groupSSA objective function. Since we measure intra-class variations as Kullback-Leibler Divergence between the epochs and the group mean, it is straight forward to measure discriminativity, or inter-class differences, as divergence between the average distributions of both classes, namely

$$D_{\mathrm{KL}} \left[\mathcal{N}(\overline{\boldsymbol{\mu}}_{1}^{\mathfrak{s}}, \overline{\Sigma}_{1}^{\mathfrak{s}}) \,||\, \mathcal{N}(\overline{\boldsymbol{\mu}}_{2}^{\mathfrak{s}}, \overline{\Sigma}_{2}^{\mathfrak{s}}) \right]. \tag{3}$$

We subtract this term from the groupSSA objective function and use a trade-off parameter $\lambda \in [0\ 1]$ to control the amount of discriminativity and stationarity. A small λ pushes the groupSSA objective function towards zero, so that the discriminativity aspect is overemphasized. On the other hand

[1] Assume two trials $\mathbf{x}$ and $\mathbf{y}$ are concatenated resulting in $\mathbf{z} = [\mathbf{x}\ \mathbf{y}]$. The covariance matrix of $\mathbf{z}$ is $\mathbf{C}_z = \mathbf{z}\mathbf{z}^T = [\mathbf{x}\ \mathbf{y}][\mathbf{x}\ \mathbf{y}]^T = \mathbf{x}\mathbf{x}^T + \mathbf{y}\mathbf{y}^T$.

if λ is equal to one, we obtain the same solution as with groupSSA. We minimize the objective function by conjugate gradient descend in the space of antisymmetric matrices [9].

III. Experimental Results

A. Data

The data used in this paper consists of two calibration (i.e. without feedback) recordings from five healthy participants. The volunteers performed motor imagery of two limbs, specifically 'left hand' and 'foot'. The cues were presented either visually (with an arrow appearing in the center of the screen) or auditory (a voice announcing the task to be performed), resulting in two different datasets for each user. In this experiment, the training data (132 trials) was the calibration with visual stimuli and the testing data (132 trials), the calibration with auditory stimuli. The preprocessing parameters (frequency band and time interval) were subject-optimized in the training set. The data was recorded with a multichannel system 85 electrodes densely covering the motor cortex. After filtering, it was down-sampled to 100 Hz. Subsequently, Common Spatial Patterns (CSP) were computed and features were extracted using log-band power on CSP filtered channels (three filters per class). Finally, the classifier was Linear Discriminant Analysis (LDA). In the case of SSA or one of its variants the band-pass filtered training data was used to feed the algorithm. The data was projected in the resulting stationary dimensions and after that the same feature extraction method as explained above was applied. The SSA methods were restarted 50 times in order to avoid local minima and the dimensionality of the stationary subspace was selected via 5-fold cross-validation on the training set using classification accuracy.

B. Results

In order to study the different SSA variants we created two artificial data sets, one with a task-unrelated non-stationary direction and one where the dominant direction of variation is discriminative. The upper row of Figure 2 shows the distribution of data points (9 epochs per class) for both data sets. As in Figure 1 we color-code the classes (red and blue) and the time-flow (bright to dark). For each epoch we extract the covariance matrix and plot the variances in the middle row of Figure 2. The data in the left panel have a discriminative horizontal and a non-stationary vertical direction (but trends have opposite sign). In contrast, in the second data set the prominent non-stationary change coincides with the discriminative direction. In the last row of Figure 2 we visualize the non-stationary direction obtained by different methods. We see that not all SSA variants ensure that the discriminative information remains in the stationary part. For instance a direct application of SSA would discard discriminative information in both data sets, whereas the performance of SSAconcat highly depends on the concatenation scheme used. The groupSSA method performs well in the first example, but fails when the discriminative and non-stationary direction coincide. Only dSSA finds a subspace that is both discriminative and stationary.

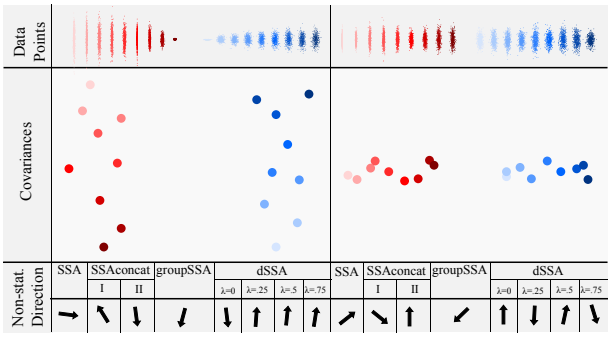

Fig. 2. The first row shows the distribution of points for each epoch of the two data sets. The classes are represented by red and blue color and the time-flow goes from bright to dark. The middle row visualizes the variances of the point distributions, e.g. the red points in the left panel show a trend downwards which corresponds to the shrinking vertical variance in the point distributions. The last row shows the non-stationary direction found by different methods. We see that it often contains discriminative information, i.e. contributions from the horizontal direction. Only dSSA finds the correct non-stationary direction, thus avoids that important information is removed.

TABLE I

Comparison of classification accuracies for five subjects performing motor imagery. The SSA-based methods are applied as preprocessing step and dimensionality is selected via 5-fold CV. We see that learning in stationary and discriminative subspaces performs best ($\text{dSSA}_{\lambda=0.75}$).

Methods	S1	S2	S3	S4	S5	Mean	Std
No SSA	90.9	80.0	73.3	70.8	94.2	81.8	10.4
SSA	90.9	60.0	**82.5**	70.8	82.5	77.3	12.0
SSAconcat I	87.8	75.8	77.5	74.1	93.3	81.7	8.3
SSAconcat II	88.7	71.7	75.0	70.8	78.3	76.9	7.2
$\text{dSSA}_{\lambda=0}$	90.9	81.7	75.0	76.7	95.0	83.9	8.7
$\text{dSSA}_{\lambda=0.25}$	90.2	78.3	74.2	69.2	94.2	81.2	10.6
$\text{dSSA}_{\lambda=0.5}$	90.2	**83.3**	79.2	70.8	95.8	83.9	9.7
$\text{dSSA}_{\lambda=0.75}$	90.9	82.5	80.8	**78.3**	**97.5**	**86.0**	8.0
$\text{dSSA}_{\lambda=1}$	**91.7**	78.3	80.0	77.5	**97.5**	85.0	9.0

In order to study how useful it is for a BCI application to perform learning and classification in a stationary subspace we apply the SSA variants to the data set described in the previous subsection and summarize the results in Table I. As can be seen dSSA outperforms the baseline methods (except for subject 3) and best performance is achieved for $\lambda = 0.75$. This indicates that preserving discriminativity is important when removing non-stationarities from data. We further see that different concatenation schemes for SSAconcat lead to different results. This shows that grouping is important as it preserves class-specific non-stationarities which may be cancelled out when applying averaging of opposing trials. Note that the methods have different sensitivity with respect to the target dimensionality. Removing few directions may lead to a significant performances drop in the case of SSA (e.g. subject 2), whereas dSSA is much more stable in this respect as it actively prevents that discriminative information is removed. The question remains why learning in a discriminative and stationary subspace is advantageous. Clearly, discriminitivity is necessary for subsequent classification, but why is stationarity favourable ?

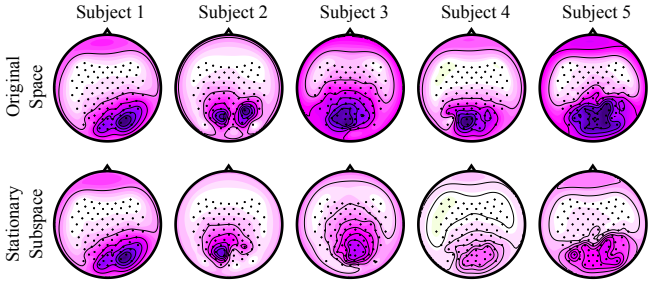

Fig. 3. Scalp plots showing the mean difference in power between the training and the test data. When no preprocessing is applied there is a significant change between training and test features, especially in occipital regions. This is probably due to different cues that are being presented in training and testing stage. Projecting the data to a stationary subspace reduces this shift, thus makes the signal more stationary. The effect is especially large for the subjects 3, 4 and 5.

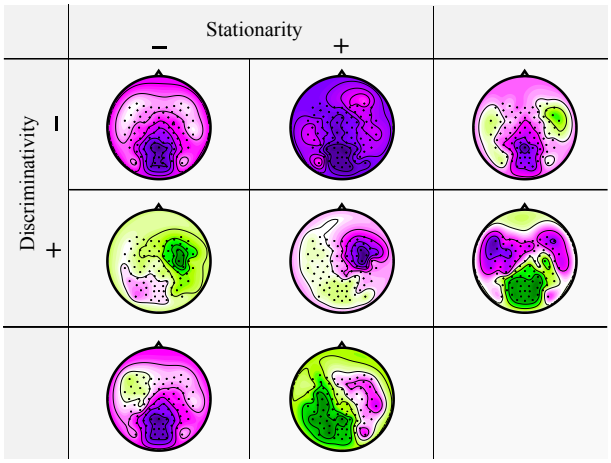

Fig. 4. Scalp plots showing the most (non-)discriminative / (non-)stationary directions extracted by dSSA of subject 5. Note that $-$ (or $+$) stands for low (or high) discriminativity / stationarity. We see that the occipital area is not only non-discriminative (first row) but also non-stationary (first column). In contrast the region around the right motor cortex is discriminative (second row). This interpretation is in line with the fact that left hand motor imagery is performed by the subject.

One argument is that since different cues are used in the training (visual) and testing (auditory) stage, one should remove all cue-related information in order to obtain similar feature distributions in both stages. In other words we assume that the visual processing of the stimuli is a non task-related non-stationarity, thus it is advantageous to remove it as it will not be present (or will have different form) in the testing stage which uses auditory stimuli. In fact when comparing the difference in power between the training and the test data for the five subjects, we clearly see that the changes become smaller when learning is performed in a stationary subspace (see Figure 3). This is especially prominent in occipital areas which are known for visual processing.

As mentioned before our dSSA method distinguishes non-stationarities which are discriminative from those that do not contain class-relevant information. In Figure 4 we visualize the most (non-)discriminative / (non-)stationary direction extracted from subject 5. Note that one can extract the discriminative directions by flipping the sign in the objective function of dSSA, e.g. adding the discriminativity term to the objective function of groupSSA instead of subtracting it. From Figure 4 we see that non-discriminative areas (first row) are mainly located in the occipital regions and on border electrodes. The occipital areas are also very non-stationary (first column), probably because a lot of visual processing occurs there. Regions of high discriminativity (middle row) are located over the right motor cortex which is not surprising as left hand motor imagery is performed by the subject.

IV. Discussion

In this paper we analysed limitations of SSA with respect to classification of motor imagery data and presented two extensions of the algorithm, which allow to (1) identify stationary brain sources from different conditions in a principled way and (2) trade-off discriminativity and stationarity which is crucial when classification is performed in a subsequent step. Note that the grouping of the data is not limited to class labels, but can be applied to sessions or even multiple users, thus allowing to extract common non-stationarities even in the case of opposing trends. In the future we want to use this tool to show how to analyze EEG data in group studies. Furthermore we want to interpret the non-stationary subspace neurophysiologically and analyze its stability across subjects and sessions in order to find out whether it is possible to perform learning in a discriminative and stationary subspace extracted from other subjects.

References

[1] G. Dornhege, J. del R. Millán, T. Hinterberger, D. McFarland, and K-R. Müller, editors. *Toward Brain-Computer Interfacing*. MIT Press, Cambridge, MA, 2007.

[2] B. Blankertz, M. Kawanabe R. Tomioka, F. U. Hohlefeld, V. Nikulin, and K-R. Müller. Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. In *Ad. in NIPS 20*, pages 113–120, 2008.

[3] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe. Stationary common spatial patterns for brain-computer interfacing. *Journal of Neural Engineering*, 9:026013, 2012.

[4] C. Vidaurre, C. Sannelli, K-R. Müller, and B. Blankertz. Machine-learning based co-adaptive calibration. *Neural Comp.*, 23(3):791-816, 2011.

[5] Y. Li, H. Kambara, Y. Koike, and M. Sugiyama. Application of covariate shift adaptation techniques in brain-computer interfaces. *IEEE Trans. Biomed. Eng.*, 57(6):1318–24, 2010.

[6] P. W. Ferrez A. Buttfield and J. del R. Millán. Online classifier adaptation in high frequency eeg. In *Proceedings of the 3rd International Brain-Computer Interface Workshop*, 2006.

[7] J. Q. Gan. Self-adapting BCI based on unsupervised learning. In *3rd Int. Workshop on Brain-Computer Interfaces*, pages 50–51, 2006.

[8] S. Lu, C. Guan, and H. Zhang. Unsupervised brain computer interface based on intersubject information and online adaptation. *IEEE Trans Neural Syst Rehabil Eng*, 17(2):135–145, 2009.

[9] P. von Bünau, F. C. Meinecke, F. Király, and K.-R. Müller. Finding stationary subspaces in multivariate time series. *Physical Review Letters*, 103:214101, 2009.

[10] P. von Bünau, F. C Meinecke, S. Scholler, and K.-R. Müller. Finding stationary brain sources in EEG data. In *Proceedings of 32nd Conference of EMBS*, 2010.