Toward Unsupervised Adaptation of LDA for Brain–Computer Interfaces

C. Vidaurre*, M. Kawanabe, P. von Bünau, B. Blankertz, and K. R. Müller

***Abstract*—There is a step of significant difficulty experienced by brain–computer interface (BCI) users when going from the calibration recording to the feedback application. This effect has been previously studied and a supervised adaptation solution has been proposed. In this paper, we suggest a simple unsupervised adaptation method of the linear discriminant analysis (LDA) classifier that effectively solves this problem by counteracting the harmful effect of nonclass-related nonstationarities in electroencephalography (EEG) during BCI sessions performed with motor imagery tasks. For this, we first introduce three types of adaptation procedures and investigate them in an offline study with 19 datasets. Then, we select one of the proposed methods and analyze it further. The chosen classifier is *offline* tested in data from 80 healthy users and four high spinal cord injury patients. Finally, for the first time in BCI literature, we apply this unsupervised classifier in *online* experiments. Additionally, we show that its performance is significantly better than the state-of-the-art supervised approach.**

***Index Terms*—Adaptive signal processing, brain computer interfaces, linear discriminant analysis, unsupervised learning.**

I. Introduction

THE goal of a brain–computer interface (BCI) is to classify the user's intention by observing and analyzing brain activity without relying on signals from muscles or peripheral nerves. Electroencephalography (EEG) is a very common type of signal because it is noninvasive, portable, and relatively easy to acquire in almost any environment. However, this signal has a poor signal to noise ratio and is highly nonstationary [1]–[3]. Due to this, EEG-based BCIs have to be robustified against nonstationary changes [4]–[7] or adapted to these [2], [8], [9]. BCI users, especially at early training stages, are often unable to generate stable EEG patterns. The system requires then supervised classifiers because they can "follow" unexpected class-related changes of EEG and successfully help in the learning process [1], [2], [10]; however, class information is usually not available in a practical BCI task. On the other hand, when the BCI users can generate stable patterns, task related EEG information might not change so drastically, but different electrode montages or task unrelated factors affect the signals. In this case, class information is not available for the adaptation of the system. This motivated us to study whether unsupervised adaptation based on extra assumptions works in practical BCI scenarios. In BCI experiments one of the main problems from session to session or from calibration to feedback within the same session, is the bias adaptation. Typically, the features shift in the feature space and the classifier should be re-adjusted [2], [9], [11], [12]. But even during a feedback session the bias, or even the whole classifier, is to be adjusted [10]. Shenoy *et al.* [13] suggested a method to overcome the difficulty of transition between the calibration (without feedback) and the feedback application. They adapted the bias of the linear discriminant analysis (LDA) using class labels of several on-going trials in the feedback session. In this study, we suggest a method that works without using label information. We compare our method to unsupervised covariate shift adaptation by training sample reweighting [14]. This approach has been applied successfully to BCI data [8]. However, in the covariate shift setting, it is assumed that the relationship between the input data points and the labels remains unchanged between training and testing phase, which is not always the case in BCI. Note also that the studies [11], [15], and [16] describe unsupervised adaptive classifiers. However, they are based on the expectation–maximization procedure and tested offline in motor imagery data. The works [11] and [16] used additional measures to improve the performance of the classifier when rotations in the feature space occur. None of the above publications report online results with the proposed unsupervised adaptive classifiers. In [17]–[19], they described a P300-based BCI system, which uses a different type of features that are more stationary (cf., [20] and [21]) than those obtained by motor imagery, and therefore, the system poses different challenges. Finally, Gan [22] also uses an unsupervised clustering method. However, it does not incorporate a forgetting factor of old data, which, in an adaptive framework, might not be the best approach.

In this paper, we first describe the acquisition of the data and the datasets used. Then, we explain the feature extraction method and the adaptive unsupervised classifiers based on the

Manuscript received August 27, 2010; revised October 10, 2010; accepted October 10, 2010. Date of publication November 18, 2010; date of current version February 18, 2011. This work was supported by the European Union (EU) MC-IEF-040666, IST-PASCAL2 Network of Excellence, ICT-216886 and by the Bundesministerium für Bildung und Forschung (BMBF) FKZ-01IBE01A/B. *Asterisk indicates corresponding author*.

*C. Vidaurre is with the Department of Machine Learning, Berlin Institute of Technology, 10623 Berlin, Germany.

P. von Bünau is with the Department of Machine Learning, Berlin Institute of Technology, 10623 Berlin, Germany.

M. Kawanabe is with the Department of Machine Learning, Berlin Institute of Technology, 10623 Berlin, Germany, and also with the Intelligent Data Analysis, Fraunhofer FIRST, 12489 Berlin, Germany.

B. Blankertz is with the Department of Machine Learning, Berlin Institute of Technology, 10623 Berlin, Germany, and also with the Intelligent Data Analysis, Fraunhofer FIRST, 12489 Berlin, Germany, and also with the Bernstein Focus: Neurotechnology, 10587 Berlin, Germany.

K. R. Müller is with the Department of Machine Learning, Berlin Institute of Technology, 10623 Berlin, Germany, and also with the Bernstein Focus: Neurotechnology, 10587 Berlin, Germany.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2010.2093133

simple and robust LDA. For this we exploit the fact that one can adapt parameters of a linear classifier without label information. Subsequently, we provide a description of covariate shift. Then, the improvement of the method is shown offline using several large datasets, including recorded data from high spinal cord injury (HSCI) patients and a comparison with covariate shift is included. Its feasibility is also tested in online experiments with 11 users. These results indicate that there exist underlying background activity that negatively affects the system performance, but that can be counteracted with the proposed method.

II. Material and Methods

Section II-A describes the datasets that are used to study the adaptive unsupervised classifiers. Section II-B explains how features are extracted. Section II-C gives a definition of the adaptive unsupervised classifiers used and includes as well the description of covariate shift (for the sake of comparison).

A. Datasets

1) Dataset 1: The first collection of data consists of 19 datasets recorded from ten subjects who performed motor imagery (left hand, right hand, right foot) according to visual cues without feedback. The pair of tasks with best discrimination is chosen for this study. These datasets are selected because they revealed nonstationarities in a previous study [1]. The brain activity is recorded with multichannel EEG using 55 Ag/AgCl electrodes at 1000 Hz, and then filtered and down-sampled to 100 Hz. After that, the data are filtered in a subject specific frequency band as in [23] (estimated from the training set only) to compute the features.

2) Dataset 2: This dataset was recorded jointly by the Tübingen and the BBCI groups [24]. The second type of data consists of 80 datasets from BCI inexperienced users (39 males, 41 females; age 29.94 ± 11.5 y; four left handed). Each dataset is acquired during a single BCI session with a classical motor imagery paradigm. In the calibration recording, a visual stimulus indicates the user which type of movement s/he should imagine in each trial. The visual stimulus is an arrow directed to the left, to the right, or to the bottom, corresponding to left hand, right hand, and foot movement imagination, giving rise to three classes and three possible binary class combinations. Calibration datasets consist of three concatenated runs, each run with 25 trials per class, resulting in 75 trials per class. Automatic variance-based artifact rejection is applied on the calibration data to reject trials and channels corresponding to amplitude abnormalities in the raw electroencephalogram (EEG). For each binary class combination, a semiautomatic procedure selects a specific frequency band and time interval, in which the two classes are best discriminable and common spatial patterns (CSPs) (see Section II-B) are trained on the corresponding filtered and segmented EEG. Afterward, CSPs filters are heuristically chosen and the log-variance of the CSP filtered EEG signals (in the following called CSP features) is used to train an LDA resulting in a generalized calibration performance. The best class combination of tasks is chosen depending on the best classification performance and used for the feedback session, see [23], [25], and [26] for more details on this procedure. During the feedback application, the previously trained CSP filters and LDA are used to, respectively, filter the raw data and classify the CSP features, and the classifier output is visualized at the same time as the stimulus presentation. The first 20 feedback trials were used to update the bias of the classifier (supervised) as in [13]. Feedback data consist of three concatenated runs, each run containing 50 trials per class. The stimulus duration is 5 s, while the interstimulus interval (ISI) is 9 s. EEG is recorded with 119 Ag/AgCl electrodes at positions according an extended international 10–20 system and a sample frequency of 1000 Hz. The data are filtered and down-sampled to 100 Hz during the session. Electromyogram (EMG) and electrooculogram (EOG) have been also recorded in order to assure that no muscle activity is present during the mental task and that no eye movements could influence the classification.

3) Dataset 3: The original study describing the acquisition of these data can be found in [27]. Seven tetraplegic, Patients (6 males, 1 females; mean age 36 years; BCI session carried out on average 11 month after SCI (spinal cord injury); all right-handed) participated in this one-session study after providing informed consent according to the guidelines of the local Ethics Commission. They had complete motor and sensory loss below a cervical spinal cord lesion (Grade A or B ASIA Impairment Scale; [28].) due to traumatic (5) or ischemic (2) cervical spinal cord injury (level C4–C7). According to the level of motor impairment and the patient's individual characteristic of kinetic experience (e.g., professional or sportive) prior to SCI, a movement simple and quick to perform is chosen (e.g., clench fist). All movements are accomplished as attempted, not imagined movements. Before measurement a detailed supervised training takes place. During the measurement patients are seated in a relaxed manner in their wheelchair with arms positioned on their lap. Brain activity is recorded from the scalp with 64-channel EEG amplifiers in an extended 10–20 system. The brain activity is recorded using 1000 Hz, and then filtered and down-sampled to 100 Hz. Using data from an initial 30 min "calibration" recording the classifier is trained in a short break, which includes the selection of the best pair of classes, and a subject-selected frequency band and a time interval. Subsequently, one to three runs of classic feedback modus take place where users perform a 1-D feedback task moving a cursor from the center of a monitor to a randomly indicated horizontal direction. In a second part, one to three runs "cursor off" are accomplished where the feedback signal is provided only at the end of each trial. As in the previous study (dataset 2), the first 20 feedback trials are used to update the bias of the classifier (supervised) following the method presented in [13].

4) Dataset 4: Data from 11 users are collected in a study with a similar design to that described in Section II-A2. However, the feedback is performed using an unsupervised classifier that we introduce in this paper (in particular PMean; see Section II-C2). The session is also divided between calibration and feedback, and the feedback application is the same of Section II-A2. We record signals from 64 Ag/AgCl electrodes at positions according an extended international 10–20 system. The data are recorded at 1000 Hz, filtered and down-sampled to

100 Hz. The calibration data are used to estimate the subject-specific frequency band and time interval, needed for computing the features as in [23]. During the feedback runs, instead of using the labels and data from the first 20 trials to update the bias, this bias is updated in an unsupervised manner after every trial.

B. Preprocessing

With all datasets, we use CSPs to extract band power features. CSP is a technique to analyze multichannel data based on recordings from two classes (tasks). It yields a data-driven supervised decomposition of the signal $\boldsymbol{x}(t)$ parameterized by a matrix $\boldsymbol{W}$ that projects the signal in the original sensor space to a surrogate sensor space $\boldsymbol{x}_{\mathrm{CSP}}(t)$ (cf. [23]): $\boldsymbol{x}_{\mathrm{CSP}}(t) = \boldsymbol{W} \cdot \boldsymbol{x}(t)$. Each row vector of a $\boldsymbol{W}$ is a spatial filter. CSP filters maximize the variance of the spatially filtered signal under one task while minimizing it for the other task. Since the variance of a bandpass-filtered signal is equal to band power, CSP analysis is applied to bandpass-filtered signals to obtain an effective discrimination of mental states that are characterized by ERD/ERS effects. Detailed information about this technique can be found in [23].

C. Classifiers

We concentrate on a binary classification problem with linear classifiers (LDA), which are specified by discriminant functions. We assume the covariance matrices of both classes to be equal and denote it by Σ. Furthermore, we denote the classwise means by $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and arbitrary feature vector by $\boldsymbol{x}$ and define

$$D(\boldsymbol{x}) = [b, \boldsymbol{w}^\top] \begin{bmatrix} 1 \\ \boldsymbol{x} \end{bmatrix} \tag{1}$$

$$\boldsymbol{w} = \Sigma^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \tag{2}$$

$$b = -\boldsymbol{w}^\top \boldsymbol{\mu} \tag{3}$$

$$\boldsymbol{\mu} = \frac{1}{2} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2). \tag{4}$$

Then, $D(\boldsymbol{x})$ is the difference in the distance of the feature vector $\boldsymbol{x}$ to the separating hyperplane described by its normal vector $\boldsymbol{w}$ and the bias b. If $D(\boldsymbol{x})$ is greater than 0, the observation $\boldsymbol{x}$ is classified as class 2, and otherwise, as class 1.

The parameters specifying the model are $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma)$. In the supervised setting, where the label y is also presented in online sessions, all parameters can be estimated adaptively. On the other hand, in the unsupervised setting, the label y is unavailable and only the input $\boldsymbol{x}$ is provided as feedback information. Therefore, we can just infer the global mean (i.e., the mean of the entire samples) $\boldsymbol{\mu} := \boldsymbol{\mu}_1/2 + \boldsymbol{\mu}_2/2$ and the global covariance $\widetilde{\Sigma} := \Sigma + \boldsymbol{\delta}\,\boldsymbol{\delta}^\top/4$, where the class probabilities $\pi_1 = \pi_2 = 1/2$ are balanced and $\boldsymbol{\delta}$ denotes the mean difference, i.e., $\boldsymbol{\delta} := \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$. The notation ":=" means "the variable in the left hand side is *defined by* the equation in the right hand side." We remark that we can use the global covariance $\widetilde{\Sigma}$ for calculating the LDA direction $\boldsymbol{w}$ in (2), because

$$\widetilde{\Sigma}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) = \left(1 + \boldsymbol{\delta}^\top \Sigma^{-1} \boldsymbol{\delta}/4\right)^{-1} \boldsymbol{w}.$$

Thus, we redefine the weight vector as $\boldsymbol{w} := \widetilde{\Sigma}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$ and also consider online estimation of the global covariance in the adaptation schemes below. Although the unknown difference $\boldsymbol{\delta}$ between the class means is assumed to be constant over online sessions, we will see in Section III that this assumption is indeed useful in practice to design unsupervised adaptation methods for LDA.

We consider five online adaptation schemes: two of them require label information (supervised LDA and Mean) and the other three can update the classifier without knowing which the performed tasks are. Note that using a global sample covariance matrix (calculated using all features) instead of an averaged one (average covariance of the class covariances) does not affect the classification result [29]. In the following adaptation schemes, the quantities with the time indices denote the estimates of the corresponding variables at the time points. For instance, $\widetilde{\Sigma}(t)$ is the estimate of the global covariance $\widetilde{\Sigma}$ at time t.

1) Supervised Adaptive LDA: The supervised classifier is used in the first dataset to help studying the performance of different unsupervised classifiers. In the supervised scenario, we can update the class means $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and the global covariance matrix $\widetilde{\Sigma}$ in an online manner. LDA relies on the inverse $\widetilde{\Sigma}^{-1}$ of the covariance matrix $\widetilde{\Sigma}$ see (2) and (3), which can be recursively estimated applying the matrix inversion lemma [30], where η is the update coefficient and $\boldsymbol{x}(t)$ is the current sample vector without the mean

$$I(t) = \widetilde{\Sigma}(t-1)^{-1} - \frac{\boldsymbol{v}(t)\,\boldsymbol{v}(t)^\top}{\frac{1-\eta}{\eta} + \boldsymbol{x}(t)^\top \boldsymbol{v}(t)} \tag{5}$$

$$\widetilde{\Sigma}(t)^{-1} = I(t)/(1-\eta) \tag{6}$$

with $\boldsymbol{v}(t) = \widetilde{\Sigma}(t-1)^{-1}\,\boldsymbol{x}(t)$. Note, the term $\boldsymbol{x}(t)^\top \boldsymbol{v}(t)$ is a scalar, and no costly matrix inversion is needed. To estimate the class-specific adaptive mean $\boldsymbol{\mu}_1(t)$ and $\boldsymbol{\mu}_2(t)$ one can use

$$\begin{aligned} \boldsymbol{\mu}_i(t) &= (1-\eta)\,\boldsymbol{\mu}_i(t-1) + \eta\,\boldsymbol{x}(t) \\ i &:= \text{class of } \boldsymbol{x}(t). \end{aligned} \tag{7}$$

The previous algorithms describe an exponential update rule, in which the current sample has the greatest weight.

We also consider a simpler adaptive classifier (mean classifier), which only updates the class means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ by (7), while the covariance matrix $\widetilde{\Sigma}$ is kept constant.

2) Unsupervised Adaptive LDA I (Changes in the Pooled Mean): As shown in [31], there are changes which affect the mean of the features. One can modify part of the bias given in (3) by adapting the common mean $\boldsymbol{\mu}(t)$ (the average of the two class means). We update the global mean $\boldsymbol{\mu}(t)$ by the same rule as (7) except that all trials from both tasks are used. This classifier is called **PMean** in this paper.

$$b(t) = -\boldsymbol{w}^\top \boldsymbol{\mu}(t) \tag{8}$$

3) Unsupervised Adaptive LDA II (Pooled Mean Changes and Global Sample Covariance Changes): Here, we update not only the global mean but also the global covariance matrix (**PMean-GCov classifier**), while keeping the difference between the two class means constant. We estimate the inverse

of the "global sample covariance matrix" for which no class information is needed, as in (6). For this estimation, we only need to subtract the common mean estimator to the current feature vector $\boldsymbol{x}(t)$. The LDA bias and weights are modified

$$\boldsymbol{w}(t) = \widetilde{\Sigma}(t)^{-1}\,(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1),\ b(t) = -\,\boldsymbol{w}(t)^{\top}\,\boldsymbol{\mu}(t). \quad (9)$$

4) Unsupervised Adaptive LDA III: A scaling happening in the feature space can be accounted for by using a parallel formula to that explained in [32] for the case of adaptive CSP filters

$$\boldsymbol{w}(t) = \widetilde{\Sigma}(t)^{-1/2}\,\widetilde{\Sigma}^{-1/2}\,(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \quad (10)$$

$$b(t) = -\,\boldsymbol{w}^{\top}\,\boldsymbol{\mu}(t). \quad (11)$$

In which one should use the "normalization assumption" of [32]

$$\widetilde{\Sigma}(t)^{-1/2}\,(\boldsymbol{\mu}_2(t) - \boldsymbol{\mu}_1(t)) = \widetilde{\Sigma}^{-1/2}\,(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1). \quad (12)$$

This classifier is also referred as **Scaling**.

D. Theoretical Consideration

Because we do not use the label information in our unsupervised adaptation procedures, they cannot follow any kinds of nonstationary changes of data clouds. In this section, we discuss under which conditions the nonadaptive and unsupervised adaptive LDA can work properly, or in other words, when they give the same decision boundaries as those of supervised LDAs in online sessions.

Let $(\boldsymbol{\mu}_0, \boldsymbol{\delta}_0, \Sigma_0)$ and $(\boldsymbol{\mu}_*, \boldsymbol{\delta}_*, \Sigma_*)$ be the parameters in the initial calibration session and an online session, respectively, where the matrices Σ_0 and Σ_* here are the common classwise covariances, and $\boldsymbol{\delta}_0$ and $\boldsymbol{\delta}_*$ denote the classwise mean differences. For theoretical discussions, we consider an ideal situation in which the initial and online estimators achieve the true parameters almost perfectly and the class probabilities are balanced. The true LDA classifier in the online session can be expressed as

$$D_*(\boldsymbol{x}) = \boldsymbol{w}_*^{\top}\,(\boldsymbol{x} - \boldsymbol{\mu}_*)$$

$$\boldsymbol{w}_* = \left(1 + \boldsymbol{\delta}_*^{\top}\Sigma_*^{-1}\boldsymbol{\delta}_*/4\right)^{-1}\Sigma_*^{-1}\boldsymbol{\delta}_*.$$

On the other hand, the classifiers obtained from the no-adaptation and unsupervised adaptation procedures are summarized as

$$D(\boldsymbol{x}) = \boldsymbol{w}^{\top}\,(\boldsymbol{x} - \boldsymbol{\mu})$$

$$\text{nonadaptive}: \boldsymbol{w} = \boldsymbol{w}_0, \qquad \boldsymbol{\mu} = \boldsymbol{\mu}_0$$

$$\text{unsup. LDA I}: \boldsymbol{w} = \boldsymbol{w}_0, \qquad \boldsymbol{\mu} = \boldsymbol{\mu}_*$$

$$\text{unsup. LDA II}: \boldsymbol{w} = \boldsymbol{w}_1, \qquad \boldsymbol{\mu} = \boldsymbol{\mu}_*$$

$$\text{unsup. LDA III}: \boldsymbol{w} = \boldsymbol{w}_2, \qquad \boldsymbol{\mu} = \boldsymbol{\mu}_*$$

where

$$\boldsymbol{w}_0 := \left(1 + \boldsymbol{\delta}_0^{\top}\Sigma_0^{-1}\boldsymbol{\delta}_0/4\right)^{-1}\Sigma_0^{-1}\boldsymbol{\delta}_0$$

$$\boldsymbol{w}_1 := \left(\Sigma_* + \boldsymbol{\delta}_*\boldsymbol{\delta}_*^{\top}/4\right)^{-1}\boldsymbol{\delta}_0$$

$$= \Sigma_*^{-1}\boldsymbol{\delta}_0 - \frac{\boldsymbol{\delta}_*^{\top}\Sigma_*^{-1}\boldsymbol{\delta}_0}{4 + \boldsymbol{\delta}_*^{\top}\Sigma_*^{-1}\boldsymbol{\delta}_*}\,\Sigma_*^{-1}\boldsymbol{\delta}_*$$

$$\boldsymbol{w}_2 := \left(\Sigma_* + \boldsymbol{\delta}_*\boldsymbol{\delta}_*^{\top}/4\right)^{-1/2} \times \left(\Sigma_0 + \boldsymbol{\delta}_0\,\boldsymbol{\delta}_0^{\top}/4\right)^{-1/2}\boldsymbol{\delta}_0.$$

All unsupervised adaptive LDAs can follow the true global mean $\boldsymbol{\mu}_*$ in the online session. LDA I does not change the weight vector $\boldsymbol{w}$ as the nonadaptive case, while LDA II and III update it in different ways. When do these procedures work properly in the online session? The actual classifiers coincide with the true one, if $\boldsymbol{w} \propto \boldsymbol{w}_*$ and $\boldsymbol{w}_*^{\top}\boldsymbol{\mu} = \boldsymbol{w}_*^{\top}\boldsymbol{\mu}_*$, where $\boldsymbol{w} \propto \boldsymbol{w}_*$ means "$\boldsymbol{w}$ is proportional to $\boldsymbol{w}_*$," i.e., $\boldsymbol{w} = a\,\boldsymbol{w}_*$ for a constant $a \neq 0$. They lead to the following conditions:

$$\text{nonadaptive}: \Sigma_*^{-1}\boldsymbol{\delta}_* \propto \Sigma_0^{-1}\boldsymbol{\delta}_0, \quad \boldsymbol{\mu}_* = \boldsymbol{\mu}_0 + \boldsymbol{\eta} \text{ where } \boldsymbol{\eta}^{\top}\Sigma_0^{-1}\boldsymbol{\delta}_0 = 0.$$

$$\text{unsup. LDA I}: \Sigma_*^{-1}\boldsymbol{\delta}_* \propto \Sigma_0^{-1}\boldsymbol{\delta}_0 \quad (13)$$

$$\text{unsup. LDA II}: \boldsymbol{\delta}_* \propto \boldsymbol{\delta}_0$$

$$\text{unsup. LDA III}: \left(\Sigma_* + \boldsymbol{\delta}_*\boldsymbol{\delta}_*^{\top}/4\right)^{-1/2}\boldsymbol{\delta}_* \propto \left(\Sigma_0 + \boldsymbol{\delta}_0\boldsymbol{\delta}_0^{\top}/4\right)^{-1/2}\boldsymbol{\delta}_0. \quad (14)$$

Fig. 1 illustrates the cases in which the adaptation procedures are expected to follow the nonstationary changes properly. Since the classwise covariance is unchanged ($\Sigma_* = \Sigma_0$), both (13) and $\boldsymbol{\delta}_* \propto \boldsymbol{\delta}_0$ hold in the cases (a) and (b). Therefore, the unsupervised adaptive LDA I and II are consistent with the true one in the online session, while LDA III has a tiny deviation. In the models (c) and (d), the mean difference direction $\boldsymbol{\delta}_*$ is tilted from $\boldsymbol{\delta}_0$ in order to compensate the change of the classwise covariance such that $\Sigma_*^{-1}\boldsymbol{\delta}_* \propto \Sigma_0^{-1}\boldsymbol{\delta}_0$ holds. Thus, LDA I coincide with the truth, while LDA II and III are completely wrong. In the case (e), only LDA II works properly, since the mean difference direction $\boldsymbol{\delta}_*$ does not change ($\boldsymbol{\delta}_* \propto \boldsymbol{\delta}_0$). On the other hand, only LDA III is correct in the model (f), where (14) holds. In the last two situations, LDA I has bigger mismatches due to the nonstationary change of the classwise covariance.

E. Covariate Shift Adaptation

Covariate shift is the scenario where the distribution of the input data $\boldsymbol{x}$ changes between training and testing phase, while the conditional distribution of the label (or target) given the data $p(y|\boldsymbol{x})$ remains the same (cf., [14] and [33]). This means that the relationship between the input data $\boldsymbol{x}$ and the variable y that we aim to predict is constant, but that the distribution of the training data differs from the distribution of the prediction tasks posed after calibration. In the context of BCI, it is very likely that the distribution of the input data $\boldsymbol{x}$ changes between training and testing phase because the training data are collected in a controlled environment, whereas the distribution of the test data depends on the user [8], [34]. Since many learning algorithms, including LDA, hinge on the assumption that the training and test distribution are equal, generalization performance can deteriorate significantly if a covariate shift occurs. Therefore, unsupervised covariate shift adaptation techniques have been

	Σ_*	$\boldsymbol{\delta}_*$	$\boldsymbol{\mu}_*$	Non	u. LDA I (PMean)	u. LDA II (PMean-GGov)	u. LDA III (Scaling)
(a)	Σ_0	$\propto \boldsymbol{\delta}_0$	inside	◯	◯	◯	×
(b)	Σ_0	$\propto \boldsymbol{\delta}_0$	outside	×	◯	◯	×
(c)	change	(13)	inside	◯	◯	×	×
(d)	change	(13)	outside	×	◯	×	×
(e)	change	$\propto \boldsymbol{\delta}_0$	outside	×	×	◯	×
(f)	change	(14)	outside	×	×	×	◯

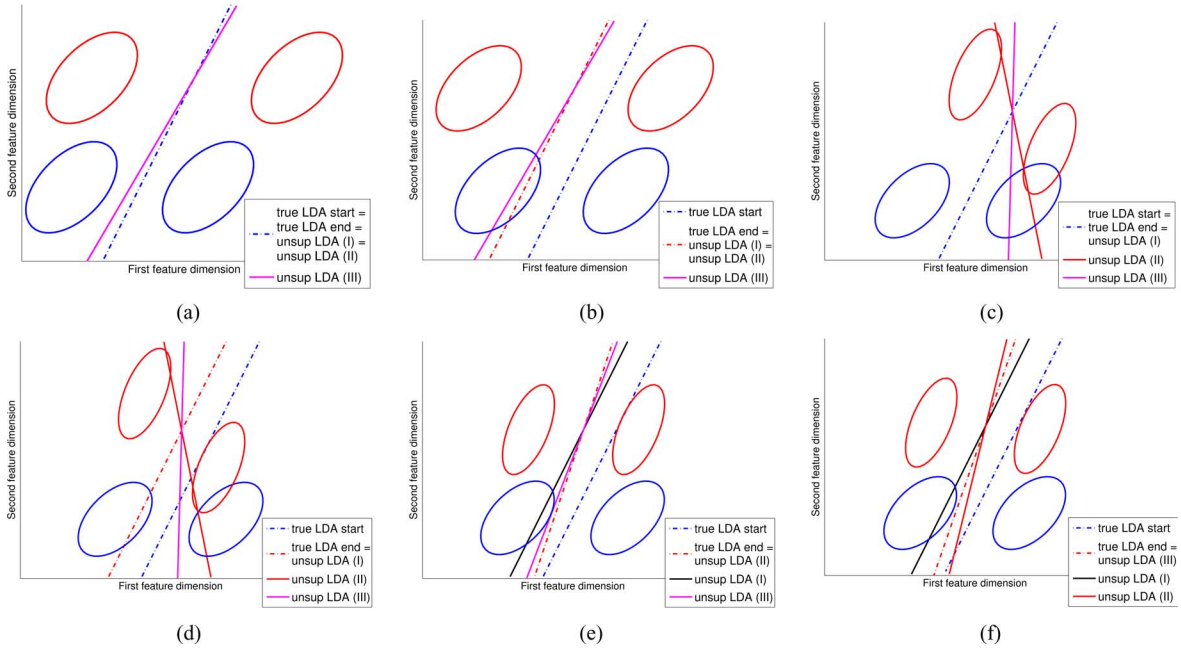

Fig. 1. (Top) Table with the summary of the illustrative examples. The cases in which the original and adaptive classifiers coincide with the optimal linear classifier (LDA) for the test set are marked with "o." ("inside" = mean shift within the decision boundary, "outside" = arbitrary mean shift). (Bottom) Figures depicting the examples that illustrate the abilities of the unsupervised adaptation procedures to follow nonstationary changes. The blue ellipsoids are the initial distributions, while the red ones show the situations in online sessions. The blue and red dashed lines are the true decision boundaries in the initial and online sessions. The nonadaptive classifier is correct, if there is no shift outside of the decision boundary marked with the blue dashed line (a and c). The decision boundaries by the unsupervised LDA's are plotted with black (I), red (II), and magenta (III).

developed [35], [36] which have also been successfully applied to BCI [8], [34].

The idea of covariate shift adaptation is to reweight the individual samples in the training data to reflect the test distribution more closely. This requires an estimate of the ratio of test and training density (also called the *importance*) at the training samples. Importance estimation is a separate task for which several methods have been proposed [35], [36], all of which require an unlabeled sample from the test distribution. In many circumstances this is much easier to obtain than a labeled sample. In BCI, one could for example periodically update the classifier using sets of unlabeled samples collected in the online phase.

Given a fixed labeled training sample and an unlabeled sample from the test distribution obtained at time window t, the parameter $\boldsymbol{w}(t)$ of adaptive importance weighted LDA (AIWLDA) [8] is obtained as follows. Let X be a $d \times n$ matrix composed of the n training samples $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathrm{R}^d$ and let $\boldsymbol{y}$ be a row vector with the corresponding labels, $y_1, \ldots, y_n \in \{-1, +1\}$. Moreover, let $D(t)$ be a diagonal matrix of estimated importance weights

$$(D(t))_{ii} = \left(\frac{\widehat{p_{\mathrm{te}}(\boldsymbol{x}_i)}}{p_{\mathrm{tr}}(\boldsymbol{x}_i)} \right)^{\lambda}$$

where p_{te} and p_{tr} are test and training densities, respectively, and $0 < \lambda \leq 1$ is a regularization parameter which flattens the importance weights in order to reduce the variance due to finite samples [14]. The value of the regularization parameter λ can be determined using importance weighted cross validation (IWCV) [8], a variant of cross validation that is consistent under covariate shift. Thus, the parameters of the AIWLDA classifier are given by

$$\boldsymbol{w} = (X^\top D(t) X)^{-1} X^\top D(t) \boldsymbol{y}.$$

III. Results

We will now outline the results obtained from the introduced methods. First, we present the offline studies. With the first dataset, we compare the unsupervised classifiers introduced

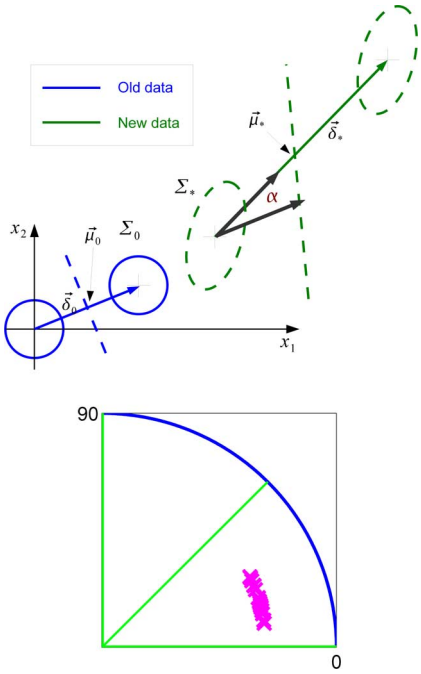

Fig. 2. (Top) Schema where it can be seen which angle α is computed to compare the change between the training and the test sets. (Bottom) α angles obtained using dataset 1.

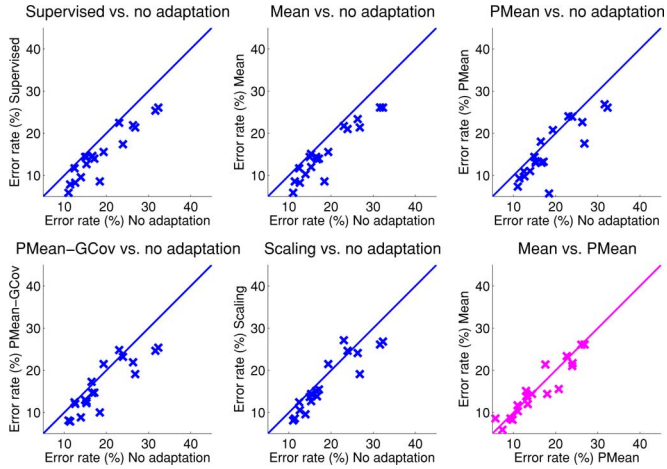

Fig. 3. Comparison of classifiers using error rates. All of them are compared to no adaptation except the bottom-right corner, in which adaptation with and without class labels are compared.

before to finally select one of them and perform the selection of the parameter η. With datasets 2 and 3, we compare the selected unsupervised classifier against the supervised state-of-the-art and a noncausal approach. With dataset 2, we also explore the limits of the unsupervised classifier. Then, we perform the same experiment with the covariate shift approach. Finally, we show the results of the online experiment.

A. *Offline Analyses*

1) Dataset 1 (Classifier and Parameter Selection): The sessions (only calibration) are divided into two halves. The parameters are optimized in the first half and applied to the second one. The proposed unsupervised adaptive LDAs (see Section II-C2–II-C4) are based on the assumption that the means of the two feature distributions drift in a similar way, i.e., the difference between the two means is nearly constant and no big changes are expected in the covariance matrices. In order to quantify the validity of this assumption, the angle formed between the vectors connecting the mean values of each class of the first half and the corresponding vector of the second half is determined for each dataset. Fig. 2 summarizes the angles computed using the first half against the second. Fig. 3 depicts scatter plots of error rates of each of the classifiers versus no adaptation except in the bottom-right corner, where the mean and pooled mean classifiers are compared. The time in which the performance is calculated is fixed beforehand to assure causality of the results. As error rates are depicted, the values below the diagonal mean that the classifier of the *y*-axis performs better than the one of the *x*-axis. The results show that the supervised options perform best against no adaptation, however, no significant difference is observed between adapting the mean with and without class information (Mean and PMean, respectively). After performing a Wilcoxon test with a *p*-value corrected for multicomparison, the unsupervised methods studied do not exhibit significant differences in performance between each other. However, unsupervised LDA I, from now on denoted as PMean, has the lowest number of parameters to be selected and estimated, and it performs as good as the true LDA in more settings (see table given in Fig. 1). Thus, we choose PMean for the subsequent studies. The update coefficient for the rest of the analysis is fixed to $\eta = 0.05$ resulting from the optimization using dataset 1.

2) Dataset 2 (Calibration to Feedback Transfer and Effect of Changing the Class Probabilities of the Test Set): In this section, we study the transition from calibration to feedback and the effect that changing the class probabilities (in the test set) has in the performance of the selected method (unsupervised LDA I, PMean). In PMean, the class probabilities are set *a priori*; therefore, we study here how robust the method is against their change. The update coefficient is fixed (from results of dataset 1) and equal to 0.05. The results of these analyses are depicted in Fig. 4. On the left, we observe that when the assumed class probabilities actually change in the test set, the performance slowly decreases. After removing approximately 75% of the trials of one class, the performance is equal to the one obtained during the feedback session with the original method described in [13], in which they use the first 20 trials of the feedback session to reestimate the bias of the LDA and after that maintain the classifier unchanged (we will call this classifier **supervised 20 trials** or **original classifier** throughout the next sections). Before reaching that point, the unsupervised PMean classifier outperforms the original method. The scatter plot comparing the online results with the unsupervised classification performance is on the right, showing that our method PMean is significantly better than the original. Fig. 5 depicts the accuracies reached using the original classifier (supervised 20 trials), PMean, and a classifier with an optimal bias. This optimal bias is found offline, using all the test data, so it is a noncausal method (unlike

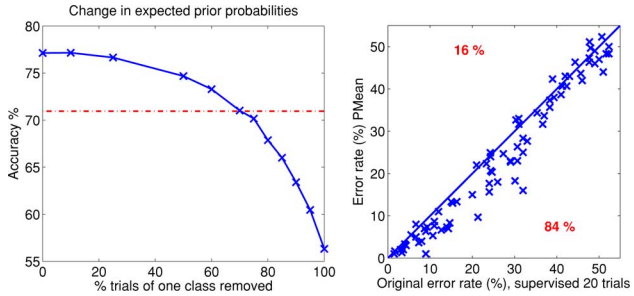

Fig. 4. (Left) Effect of removing trials of one class (test dataset) in the performance. The horizontal red line is the original accuracy obtained during the experiments. The blue crosses are obtained applying PMean. (Right) Scatter plot comparing the performance obtained during the online experiments with the performance obtained with the method PMean. PMean outperforms the original method used to provide feedback during the experimental session when the crosses are below the diagonal.

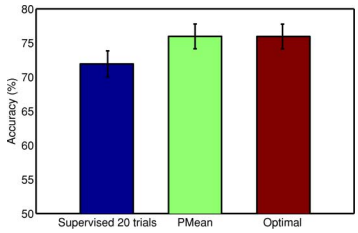

Fig. 5. Comparison of accuracy between original method (used in the experiments to provide feedback), PMean and optimal bias (noncausal).

PMean). The bias is found using brute force, choosing the maximum performance reached using a classifier where the weights are constant and the bias slowly changes from the mean of one class to the mean of the other. One can observe that there is no significant difference between the performance of PMean and optimal bias, and both are significantly better than the original method. To check this, we apply a one-sided signed-rank test, obtaining a p-value under 0.01.

3) Dataset 3, HSCI Patients: Dataset 3 is recorded from patients suffering from HSCI. We want to check whether the effect of counteracting the change from calibration to feedback works similarly well for healthy users and patients. Previously published work reported poor performance of HSCI patients as BCI users [37] using motor imagery tasks. In this study, only the data of HSCI patients who performed feedback is analyzed. To see the description of the original study, see [27]. Fig. 6 shows that the unsupervised version is again better than or similar to the original supervised approach. Although the number of users is too small to study how significant this result is, it clearly indicates the potential of the unsupervised method. Fig. 7 depicts the mean and standard error of the original (supervised in 20 trials), PMean, and optimal bias. The number of users is too small to study the significance of the results, but again, Fig. 7 shows the encouraging potential of PMean.

4) Covariate Shift Adaptation: We employ this method to dataset 2 (see Section II-A2). In order to apply the generic covariate shift adaptation technique, we need an unlabeled sample from the distribution that we want our classifier to adapt to. Therefore, we set aside the first half of the online data for importance estimation and evaluate the classifier on the second half. For importance estimation we use uLSIF [36]; KLIEP [35] yields very similar results at a slightly higher computational cost. The regularization parameter for the importance weights are determined using tenfold IWCV on the training set [8] with squared loss and chronological partitioning due to the time structure.[1]

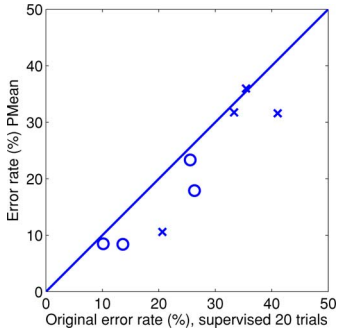

Fig. 6. Scatter plot to study the calibration to feedback transfer performance. The crosses indicate feedback runs recorded with discrete visual feedback only at the end of each trial. The circles indicate feedback data recorded with continuous visual feedback. For more details on the paradigm, see [27].

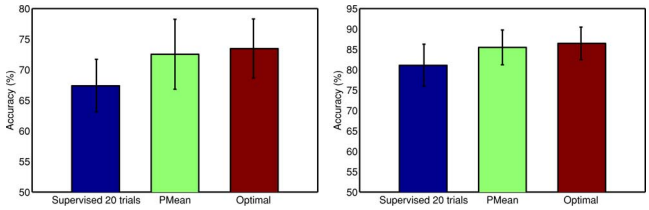

Fig. 7. (Left) Accuracy reached with discrete feedback at the end of the trial for the original method (supervised 20 trials), PMean, and optimal bias. (Right) Accuracy reached with continuous visual feedback for the original method (supervised 20 trials), PMean, and optimal bias.

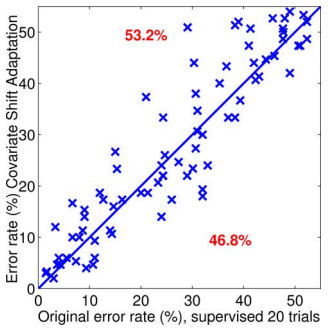

Fig. 8. Results of unsupervised covariate shift adaptation. Values under the diagonal indicate that the covariate shift adaptation is better than the original method.

The results are shown in Fig. 8. Overall, covariate shift adaptation does not improve the performance on this dataset (compare to the results obtained with the unsupervised classifier PMean on the right of Fig. 4). We conjecture that in BCI, nonstationary brain processes do not only affect the distribution of the covariates, but can also change the relationship between the data and the class label.

[1]This is a similar setting as in [8].

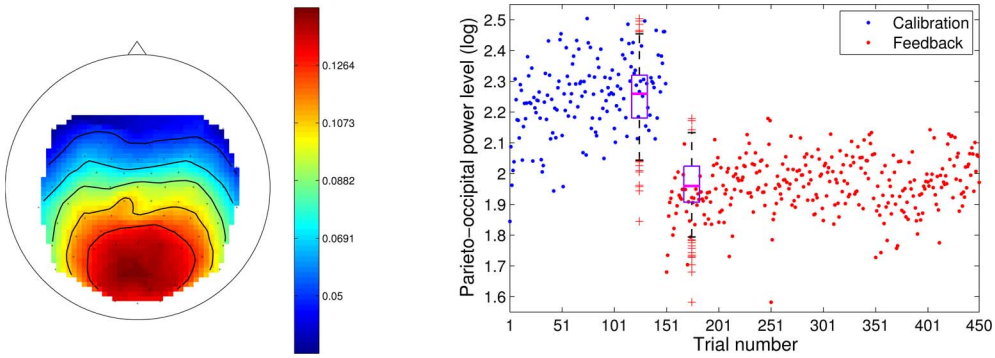

Fig. 9. (Left) Averaged principal component of the transfer from calibration to feedback (calibration–feedback). It has a strong focus in the parietooccipital area. (Right) Time course of the averaged parietooccipital power in time. The median of the users is depicted. The boxplots refer to the complete calibration and feedback, respectively, but they are located at the end and beginning of the two measurements to ease the comparison between them.

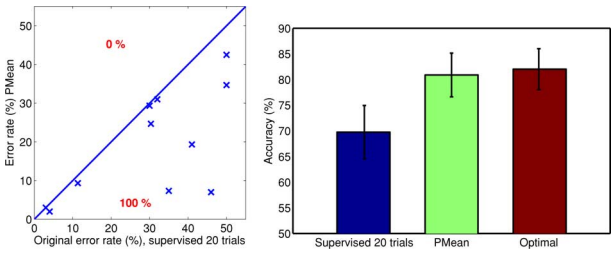

Fig. 10. (Left) Scatter plot comparing the unsupervised adaptive classifier with the state-of-the-art approach. (Right) Mean accuracy and standard error of the state-of-the art classifier, the unsupervised classifier, and the classifier with the optimal bias (noncausal).

B. Toward Understanding the Merits of the Unsupervised Approach

We study the changes observed between calibration and feedback using dataset 2. For that, we need a way to observe the main change over all BCI users. The following analysis is inspired by [1]. First, the data of the calibration and feedback runs are filtered in the frequency band used during the feedback period. Then, the data are epoched and one covariance matrix is calculated for each trial. All covariance matrices belonging to the calibration are averaged, the same is done for the feedback data. After that, the difference between these two matrices is computed. This difference is the averaged change observed in the EEG of each user between the two conditions (calibration and feedback) and contains the variation between calibration and feedback. We want to be able to plot in the electrode space the principal component of such variation. For that we perform an eigenvalue decomposition of the matrix and choose the eigenvector corresponding to the most extreme eigenvalue (positive or negative). This eigenvector can be interpreted as the source of the main variance of the transfer from calibration to feedback conditions. The individual contribution is averaged over all users. The result is depicted in Fig. 9. It exhibits a strong focus on parietooccipital regions of the scalp, which means that the differences in the activity occurring in the feedback band are responsible for a shift in the sample covariance matrices from calibration to feedback. Indeed, the median power change in the parietooccipital area between calibration and feedback is significantly different from zero, as observable in the right side of Fig. 9 and computable using a sign test (p-value $= 7.01 \cdot 10^{-5}$). However, the question whether the actual CSP-based features are actually affected by this change still remains. In order to assure that, we can find whether the correlation between the averaged bias change (between calibration and feedback) and the average parietooccipital power change is significant, which again, is the case (p-value $= 0.016$ with a Spearman correlation test). Summarizing, the CSP features are affected by a power change in the parietooccipital areas of the scalp in the frequency band of interest, and this change can be counteracted using PMean, which significantly improves the classification performance.

C. Online Experiment

Finally, using the PMean strategy, we validate, for the first time, an unsupervised classifier during online experiments with 11 users. Fig. 10 depicts the scatter plot between PMean and the method described in [13], previously in this paper referred as original or state of the art. All users perform better with PMean. However, PMean is the method used to provide feedback and the differences between the methods are biased toward PMean. Please note that for datasets 2 and 3, the results are biased to the original supervised classifier. On the right of Fig. 10, one can observe the average accuracies of the two methods plus the one obtained using the optimal bias. There are no significant differences between PMean and optimal bias, but PMean is better than the original supervised method in 20 trials. This dataset is interesting to observe what occurs during the feedback session. For that we study the Kullback–Leibler (KL) divergence and observe whether the class distributions diverge from the ones during the calibration. The left pane of Fig. 11 shows the KL divergence of local (≈ 40 trials) density estimates for each class during the feedback from the overall density estimate of that class during the calibration. Clearly, the data distribution for the right class changes over the course of the experiment with respect to the calibration, and the KL-divergence curve reflects these changes. However, KL divergence does not provide information about the separability of the data; therefore, information about the location of the classifiers is necessary. The right pane of Fig. 11 shows local estimates of the distributions

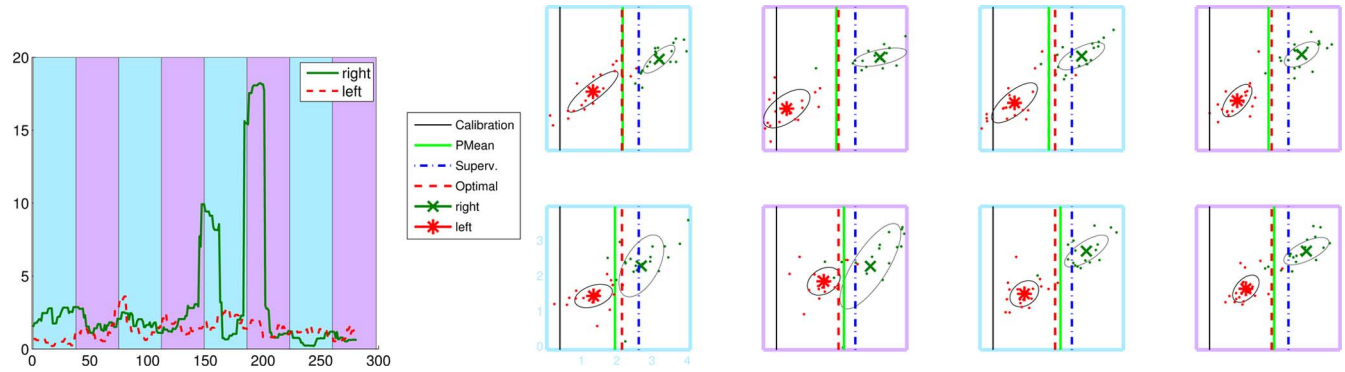

Fig. 11. (Left) Ongoing change of the KL divergence of the feedback data from the calibration data. (Right) Corresponding feature distributions for the shaded intervals. The data are projected on the plane spanned by the normal vector of the PMean hyperplane and the largest PCA component of the feedback data. This plane is the same for the calibration, supervised, optimal, and the PMean classifiers. Only bias changes occur. The black line corresponds to the classifier trained only in Calibration data. The green line is the PMean classifier in the corresponding interval. The dashed-red line is the optimal classifier for the complete feedback session. The blue line corresponds to the supervised classifier in 20 trials.

of both classes during one feedback session. We compute the weights of the PMean classifier (the weights are the same for all classifiers) and the largest PCA component of the features. In this way, the projection shows the dimension with the largest variance. The x-axis shows the projection of the data on the normal vector of that hyperplane of the feature space corresponding to the decision boundary of the classifier. The other dimension is chosen orthogonally to that normal vector, such that largest PCA component is contained in this 2-D subspace. The projection preserves angles. The location of the hyperplanes depends on the corresponding bias. Each plot corresponds to the time intervals shaded with colors on the left pane of the same figure. The Calibration classifier (black line) is obviously a bad selection to perform the feedback measurement. The Supervised classifier is a better option, but PMean and Optimal (noncausal) outperform it (as expected from Fig. 10).

IV. Discussion

The results of dataset 1 presented in Fig. 2 show that angles between the distance of the class means just vary from 8° to 25°. Some of the users were naive, but did not present bigger angle-differences than experienced ones. This finding suggests that PMean is a valid approach to perform unsupervised learning of the classifier, as we are able to demonstrate with the rest of the datasets studied. It is interesting to note that adapting means with and without class labels is not found significantly different, which is explained by the small difference between the vectors connecting the two means (small angles) found in most of the users. The Rotation classifier exhibits the worst average error rate of the unsupervised classifiers, and is found to be significantly worse than PMean. This might be caused by its too restrictive assumptions.

The results using dataset 2 strongly support the earlier findings in dataset 1 and our hypothesis that using the PMean classifier is a good approach to adapt the LDA in an unsupervised way. One limitation of the method is that the prior class probabilities have to be known. However, in Fig. 4, we show that even if the class probabilities are different from the ones assumed by the method, the results will be better than using the state-of-the-art method also when 75% of the trials of one class disappear. The analysis on 80 users, with the results biased to the state-of-the-art method (used for feedback) shows that PMean is significantly better than the classifier used during feedback runs and that its performance is not significantly different than the performance obtained using an idealized noncausal classifier with an optimal bias. These findings are also supported by the results on HSCI patients (dataset 3) and in online experiments (dataset 4).

In this study, the generic covariate shift adaptation approach did not improve the performance significantly. This is mainly due to the fact that in BCI not only does the distribution of the data change between training and test phase, but also the decision boundary may shift over time. Moreover, in practical BCI systems, the datasets available for adaptation are fairly small which makes it difficult to accurately estimate the ratio between test and training density, which is used to reweight the training data points.

Finally, the results presented in this study allow to infer that, in general, the spatial distribution of the activation patterns is quite stable in the transfer from offline to online (feedback), and therefore, the PMean strategy is successful. However, the unsupervised approach might not be sufficient for users whose brain patterns are hardly separable. For them, the expected accuracy (cross-validation error in the calibration phase) is below 70% in a two-class system. In that case, we suggest to perform coadaptive calibration as presented in [10].

V. Conclusion

When operating a BCI, there is a considerable fluctuation in the underlying statistics [7]–[9], [13], [38]. This observation is subject- and even task-dependent. Compensating such nonstationary effects and investigating their underlying cause is an important milestone on the way to more robust BCI systems. Our focus in this paper is to study nontask-related fluctuations, for which—as we find–unsupervised data analysis methods can contribute to compensating such nonstationarity.

We consider three unsupervised classifiers that are shown to successfully counteract the effect of nonclass-related EEG changes. We have shown that one of them, specifically PMean which updates the bias of the LDA by adapting the global feature average, consistently outperforms the state-of-the-art method on different data: offline analysis with many users, offline analysis with HSCI users and online experiments.

We also compare our method to another state-of-the-art unsupervised framework (covariate shift), and show that for this setting PMean can provide significantly better results (class conditional changes). This is in line with the small fluctuations that can be found when analyzing the change in the vectors that connect the class means at different time points. In other words, in the studied setting the majority of BCI users suffer a considerable signal variation which is task unrelated and can thus be tackled in an unsupervised manner.

Finally, the methods presented in this paper are applied in a synchronous paradigm. The extension to asynchronous is possible by introducing another level of control in charge of the detection of motor imagery (as done, for example, in [39], a winning algorithm of the BCI-Competition III). If motor imagery is being performed, the PMean classifier can be adapted during that period.

ACKNOWLEDGMENT

The authors thank A. Schlögl valuable discussions. This publication only reflects the authors' views.

REFERENCES

[1] M. Krauledat, "Analysis of nonstationarities in EEG signals for improving BCI performance," Ph.D. dissertation, TU-Berlin, Fakult. IV- Elektrotech. und Inf., 2008.

[2] C. Vidaurre, A. Schlögl, R. Scherer, R. Cabeza, and G. Pfurtscheller, "Study of on-line adaptive discriminant analysis for EEG-based brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 54, pp. 550–556, 2007.

[3] C. Vidaurre, R. Scherer, R. Cabeza, A. Schlögl, and G. Pfurtscheller, "Study of discriminant analysis applied to motor imagery bipolar data," *Med. Bio. Eng. Comput.*, vol. 45, pp. 61–68, 2007.

[4] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K-R Müller, "Invariant common spatial patterns: Alleviating nonstationarities in brain–computer interfacing," in *Advances in Neural Information Processing Systems 20*. Cambridge, MA: MIT Press, 2008, pp. 113–120.

[5] M. Kawanabe, C. Vidaurre, B. Blankertz, and K. R. Müller, "A maxmin approach to optimize spatial filters for eeg single-trial classification," in *Proc. IWANN'09 Part I, LNCS*, 2009, pp. 674–682.

[6] M. Kawanabe and C. Vidaurre, "Improving bci performance by modified common spatial patterns with robustly averaged covariance matrices," in *Proc. World Congress*, 2009, pp. 297–282, Munich, Germany.

[7] P. von Bünau, F. C. Meinecke, F. C. Király, and K.-R. Müller, "Finding stationary subspaces in multivariate time series," *Phys. Rev. Lett.*, vol. 103, no. 21, p. 214101, Nov. 2009..

[8] M. Sugiyama, M. Krauledat, and K.-R. Müller, "Covariate shift adaptation by importance weighted cross validation," *J. Mach. Learning Res.*, vol. 8, pp. 985–1005, May. 2007.

[9] J. del R. Millán, A. Buttfield, C. Vidaurre, M. Krauledat, A. Schlögl, P. Shenoy, B. Blankertz, P. N. Rao, R. Cabeza, G. Pfurtscheller, and K-R. Müller, "Adaptation in brain–computer interfaces," in *Toward Brain–Computer Interfacing*. Cambridge, MA: MIT Press, 2007, pp. 303–326.

[10] C. Vidaurre, C. Sannelli, K-R. Müller, and B. Blankertz, "Machine learning co-adaptive calibration," *Neural Comput.*, vol. 23, no. 3. to appear.

[11] J. Blumberg, J. Rickert, S. Waldert, A. Schulze-Bonhage, A. Aersten, and C. Mehring, "Adaptive classification for brain–computer interfaces," in *Proc. 29th Ann. Int. Conf. of the IEEE EMBS*, 2007, pp. 2536–2539.

[12] Y. Wang, B. Hong, X. Gao, and S. Gao, "Implementation of a brain–computer interface based on three states of motor imagery," pp. 5059–5062, 2007.

[13] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller, "Towards adaptive classification for BCI," *J. Neural Eng.*, vol. 3, pp. R13–R23, 2006.

[14] H. Shimodaira, "Improving predictive inference under covariate shift by weighting the log-likelihood function," *J. Stat. Planning Inference*, vol. 90, no. 2, pp. 227–244, 2000.

[15] L. Yuanqing and G. Cuntai, "An extended EM algorithm for joint feature extraction and classification in brain–computer interfaces," *Neural Comput.*, vol. 18, no. 11, pp. 2730–2761, 2006.

[16] B. A. S. Hasan and J. Q. Gan, "Unsupervised adaptive gmm for bci," in *Proc. Int. IEEE EMBS Conf. Neural Eng.*, Antalya, Turkey, 2009, pp. 295–298.

[17] A. Lenhardt, M. Kaper, and H. J. Ritter, "An adaptive P300-based online brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 2, pp. 121–130, Apr. 2008.

[18] L. Shijian, G. Cuntai, and Z. Haihong, "Unsupervised brain–computer interface based on inter-subject information and online adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 2, pp. 1–11, 2009.

[19] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K-R. Müller, "Single trial analysis and classification of ERP components—A tutorial," *NeuroImage*, to be published.

[20] C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller, "How many people are able to operate an EEG-based brain–computer interface (BCI)?" *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 145–147, Jun. 2003.

[21] C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabalona, F. Gramatica, and G. Edlinger, "How many people are able to control a P300-based brain–computer interface (BCI)?" *Neurosci. Lett.*, vol. 462, pp. 94–98, Oct. 2009.

[22] J. Q. Gan, "Elf-adapting BCI based on unsupervised learning," in *Proc. 3rd Int. Workshop Brain–Comput. Interfaces*, Graz, Austria, 2006, pp. 50–51.

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

[24] B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K. R. Müller, G. Curio, and T. Dickhaus, "Neurophysiological predictor of SMR-based BCI performance," *NeuroImage*, vol. 51, pp. 1303–1309, 2010.

[25] B. Blankertz, G. Dornhege, M. Krauledat, K. R. Müller, and G. Curio, "The non-invasive berlin brain–computer interface: Fast acquisition of effective performance in untrained subjects," *Neuroimage*, vol. 37, pp. 539–550, 2007.

[26] R. Tomioka and K. R. Müller, "A regularized discriminative framework for EEG based communication," *Neuroimage*, vol. 49, pp. 415–432, 2010.

[27] J. Conradi, B. Blankertz, M. Tangermann, V. Kunzmann, and G. Curio, "Brain-computer interfacing in tetraplegic patients with high spinal cord injury," *Int. J. Bioelectromag.*, vol. 11, no. 2, pp. 65–68, 2009.

[28] F. M. Maynard, M. B. Bracken, G. Creasey, J. F. Jr. Ditunno, W. H. Donovan, T. B. Ducker, S. L. Garber, R. J. Marino, S. L. Stover, C. H. Tator, R. L. Waters, J. E. Wilberger, and W. Young, "International standards for neurological and functional classification of spinal cord injury," *Amer. Spinal Injury Assoc. Spinal Cord*, vol. 35, pp. 266–274, 1997.

[29] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic Press, 1990.

[30] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed. Philadelphia, PA: Soc. Ind. Appl. Math., 2002.

[31] M. Kawanabe, M. Krauledat, and B. Blankertz, "A Bayesian approach for adaptive BCI classification," in *Proc. 3rd Int. BCI Workshop 2006*, Verlag TU Graz, 2006, pp. 54–55.

[32] R. Tomioka, J. Hill, B. Blankertz, and K. Aihara, "Adapting spatial filtering methods for nonstationary bcis," in *Proc. IBIS*, 2006, pp. 65–70.

[33] M. Sugiyama and K. R. Müller, "Input-dependent estimation of generalization error under covariate shift," *Stat. Decis.*, vol. 23, no. 4, pp. 249–279, 2005.

[34] Y. Li, H. Kambara, Y. Koike, and M. Sugiyama, "Application of covariate shift adaptation techniques in brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 6, pp. 1318–1324, Jun. 2010.

[35] M. Sugiyama, T. Suzuki, S. Nakajima, H. Kashima, P. von Bünau, and M. Kawanabe, "Direct importance estimation for covariate shift adaptation," *Ann. Inst. Stat. Math.*, vol. 60, no. 4, pp. 699–746, 2008.

[36] T. Kanamori, S. Hido, and M Sugiyama, "A least-squares approach to direct importance estimation," *J. Mach. Learning Res.*, vol. 10, pp. 1391–1445, Jul. 2009.

[37] L. Kauhanen, P. Jylänki, J. Lehtonen, P. Rantanen, H. Alaranta, and M. Sams, "EEG-based brain–computer interface for tetraplegics," *Comput. Intell. Neurosci.*, pp. 23864, 2007, 2011.

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

[39] D. Zhang, Y. Wang, X. Gao, B. Hong, and S. Gao, "An algorithm for idle state detection in motor imagery based brain–computer interface," *Comput. Intell. Neurosci.*, pp. ID-39714, 2007.

Authors' photographs and biographies not available at the time of publication.