L1-Penalized N-way PLS for subset of electrodes selection in BCI experiments

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2012 J. Neural Eng. 9 045010
(http://iopscience.iop.org/1741-2552/9/4/045010)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 132.168.31.62
The article was downloaded on 25/07/2012 at 16:07

Please note that terms and conditions apply.
L1-Penalized N-way PLS for subset of electrodes selection in BCI experiments

Andrey Eliseyev1,2, Cecile Moro2, Jean Faber1,2, Alexander Wyss2, Napoleon Torres2, Corinne Mestais2, Alim Louis Benabid2,3 and Tetiana Aksenova1,2,4

1 Foundation Nanosciences, Grenoble, France
2 Clinatec/LETI/CEA, Grenoble, France
3 Joseph Fourier University of Grenoble, France
E-mail: tetiana.aksenova@cea.fr

Received 11 November 2011
Accepted for publication 18 April 2012
Published 25 July 2012
Online at stacks.iop.org/JNE/9/045010

Abstract
Recently, the N-way partial least squares (NPLS) approach was reported as an effective tool for neuronal signal decoding and brain–computer interface (BCI) system calibration. This method simultaneously analyzes data in several domains. It combines the projection of a data tensor to a low dimensional space with linear regression. In this paper the L1-Penalized NPLS is proposed for sparse BCI system calibration, allowing uniting the projection technique with an effective selection of subset of features. The L1-Penalized NPLS was applied for the binary self-paced BCI system calibration, providing selection of electrodes subset. Our BCI system is designed for animal research, in particular for research in non-human primates.

(Some figures may appear in colour only in the online journal)

Introduction

The multi-way (tensor-based) analysis recently was reported as an effective tool for neuronal signal processing (Martínez-Montes et al 2004, Nazarpour et al 2006, Acar et al 2007, Mørup et al 2006, 2008, Zhao et al 2009). The advantage of this approach is the simultaneous treatment of data in several domains (modalities or ways of analysis) to improve information extraction. Spatial, frequency and temporal modalities are mostly considered in neuronal signal processing (Pfurtscheller et al 2003, Schalk et al 2007, Vidaurre et al 2009). For the multi-way data analysis, observations are represented in a form of multi-way arrays (tensors). To map the neuronal recordings to the spatial–frequency–temporal space wavelet transform is mainly used.

Recently, the multi-way analysis was reported as a tool for neuronal signal decoding in brain–computer interface (BCI) studies (Nazarpour et al 2006, Zhao et al 2009, Chao et al 2010, Phan et al 2010, Li et al 2009, Li and Zhang 2010). BCI aims to provide an alternative non-muscular communication pathway to send commands to the external world by means of analysis of recorded brain neuronal activity. Tensor-based approaches have been applied to decode electroencephalograms (EEG) (Zhao et al 2009, Phan et al 2010, Li et al 2009, Li and Zhang 2010) and electrocorticograms (ECoG) (Eliseyev et al 2011, Chao et al 2010) associated with cue-paced (Nazarpour et al 2006, Phan et al 2010) and self-paced (Eliseyev et al 2011, Chao et al 2010) BCI paradigms. The cue-paced (synchronized) control strategy uses external cues for driving the interaction between subjects and the BCI system. Thus, the users are supposed to generate commands only during specific periods. As opposed to the cue-paced systems, no stimulus is used for self-paced BCI systems. As users control them intentionally, self-paced BCI systems provide more freedom and control flexibility. However, they are based on continuous monitoring of neuronal activity and are more difficult to be realized. As a result, most reported BCIs are synchronized (for instance, see Wolpaw et al 2002, Schalk et al 2008). Even if self-paced tasks were carried out, only selected time intervals (trials), corresponding to task preparation and execution, were classified (e.g. Ball et al 2009). Nevertheless, several groups

4 Author to whom any correspondence should be addressed.
have recently designed self-paced BCI systems introducing a ‘class zero’ to discriminate mental events from basic neuronal activity (e.g. Bashashati et al 2007, Leeb et al 2007, Scherer et al 2008, Fatourechi et al 2008, Müller-Putz et al 2010, Qian et al 2010).

In general, the first stage of the BCI study consists in calibration (learning) of the BCI system. Identified on this stage model allows controlling an external effector by means of neuronal activity at the next stage of online experiments. The model estimates the effector state (output variable) depending on neuronal activity (input variables). Often, tensor input and scalar output variables are chosen in tensor-based BCIs (Zhao et al 2009, Eliseyev et al 2011). In cue-paced BCIs the binary output can be used to characterize different types of control (e.g. left- versus right-hand activity, see Zhao et al 2009). In self-paced BCIs binary output can be used to distinguish predefined events from general activity (e.g. pedal pressing versus non-pressing event, see Eliseyev et al 2011). Binary output can be predicted by classifiers (Phan et al 2010, Zhao et al 2009, Li et al 2009, Li and Zhang 2010) or by regression of binary output from observation tensors (Eliseyev et al 2011). All studies have in common that tensor factorizations have been carried out to project the data to a low dimensional space.

The particular goal of our study is to adapt self-paced BCI to real-life applications including long-term BCI experiments in a natural environment. Recently (Eliseyev et al 2011), self-paced BCI was shown to work well in an almost natural environment. A freely moving animal (rat) was trained to push a pedal to activate a food dispenser without any external stimulus. Neuronal activity was monitored and intentional control patterns were recognized by the BCI system. To identify the predictive model, the multi-way analysis, namely, N-way partial least squares (NPLS) approach (Bro 1996) was applied to project the feature tensor into a low dimensional feature space of latent variables and to estimate the regression model for the intentional control prediction. NPLS involves class information to perform the tensor decomposition that significantly increases the efficiency of modeling. As NPLS works without any prior knowledge, it can be efficiently applied to automatically generate a model predicting BCI events from recordings of neuronal brain activity. Therefore this method has been chosen as a basic approach in this study.

Note that in BCI experiments neuronal signals of the brain are processed in real time. Thus, the computational efficiency of BCI systems is of crucial importance. Selecting the appropriate features subset allows optimization of computational efficiency and improves quality of control. In BCIs in particular, the selection of electrode subsets is crucial (see, for example, Lal et al 2004). To solve this problem approaches based on classification accuracy (Sannelli et al 2010), spatial filters (Wang et al 2005, Lv and Liu 2008), mutual information (Lan et al 2005), correlation (Barachant et al 2008) and Riemannian geometry (Barachant and Bonnet 2011) were proposed. For all of them, the procedure of selecting electrodes and the calibration of the model are done apart. The procedure of selecting electrodes is carried out either during pre-processing (e.g. Wang et al 2005), or done according to classification accuracy of the models, based on different subsets of electrodes (e.g. Lal et al 2004). Applying the first method has a constraint (loss of information), whereas the second one is time and labor consuming.

Generic NPLS involves all variables to generate the final model. In this paper, we propose the L1-Penalized NPLS algorithm to directly include feature selection in the modeling process. Generic NPLS leads to a linear combination of all features. The L1-Penalized NPLS provides a sparse solution for different directions of analyses (e.g. spatial, frequency or temporal modalities). In this study, the L1-Penalized NPLS was used for binary self-paced BCI system calibration and the selection of the subset of electrodes. Real datasets of one non-human primate collected in a series of experiments lasting for up to 30 min were run with the L1-Penalized version as well as with the generic NPLS. Subsequently results were compared. The animal was trained to push a pedal to activate a food dispenser without any external stimulus. Calibrations based on recordings from training sessions for different positions of the pedal were used to establish predictive models. The models were tested for their generalization abilities. For this purpose recordings from training sessions were evaluated offline playing back the corresponding datasets. Computational experiments confirmed the high performance of the system.

Methods

Generic NPLS

The N-way PLS algorithm combines the data projection to a low dimensional feature space (the space of latent variables), with estimation of a linear regression model. This method was introduced by Bro (1996) as a generalization of the ordinary partial least squares (PLS) (Geladi and Kowalski 1986) for multi-way data sets (tensors). The PLS regression models a linear relationship between a vector of output variables and a vector of input variables on the basis of observation matrices X and Y. To build the model, the observations are projected into the low dimensional spaces in such a way that the maximum variances of X and Y are explained simultaneously. At the first iteration, the matrices X and Y are represented as $X = t_1 p_1^T + E_1$, $Y = u_1 q_1^T + F_1$, where $t_1$ and $u_1$ are the latent variables (score vectors), whereas $p_1$ and $q_1$ are the loading vectors. $E_1$ and $F_1$ are the residuals matrices. The score vectors are calculated to maximize the covariance between $t_1$ and $u_1$ (Geladi and Kowalski 1986). The coefficient $b_1$ of a regression $u_1 = b_1 t_1 + r_1$ is calculated to minimize the norm of the residuals $r_1$. Then, the procedure is applied iteratively to the residual matrices.

Similar to PLS, NPLS projects the tensor of data into the space of latent variables. Tensors (multi-way arrays) are a higher-order generalization of vectors and matrices. Elements of a tensor $X \in R^{h_1 \times h_2 \times \ldots \times h_N}$ are denoted as $X_{i_1,i_2,\ldots,i_N}$. Here, $N$ is the order of the tensor, i.e. the number of dimensions (ways or modes). The number of the variables $I_i$ in the mode $i$ shows the dimensionality of this mode (Kolda and Bader 2007). For example in a fourth-order tensor of observations $X \in R^{h_1 \times h_2 \times h_3 \times h_4}$ which contains $n$ samples $X_i \in R^{h_1 \times h_2 \times h_3 \times h_4}$.
First, the NPLS method decomposes the tensor $\mathbf{X}$ as indicated below.

$$\mathbf{X} = \mathbf{t}_1 \circ \mathbf{w}_1 \circ \mathbf{w}_2 \circ \mathbf{w}_3 + \mathbf{E},$$

where the operation ‘$\circ$’ is called the outer product (see Kolda and Bader 2007). The latent variable $\mathbf{t}_1 \in \mathbb{R}^r$ is extracted from the first mode of the tensor $\mathbf{X}$ providing a maximum of covariance between $\mathbf{t}_1$ and $\mathbf{y}$. In parallel, the algorithm forms the factor, i.e. the set of projectors $\{\mathbf{w}_1 \in \mathbb{R}^{d_1}, \mathbf{w}_2 \in \mathbb{R}^{d_2}, \mathbf{w}_3 \in \mathbb{R}^{d_3}\}$, $||\mathbf{w}_i|| = 1$, $i = 1, 2, 3$ related to the second, the third and the fourth mode of $\mathbf{X}$, respectively, in such a way that the projection of the tensor $\mathbf{X}$ on these vectors results in $\mathbf{t}_1$. The projectors correspond to each modality of analyses (e.g. spatial, frequency and temporal). To build the projectors, a tensor of correlation $\mathbf{Z} = \mathbf{X} \times_1 \mathbf{y}$ is calculated ($x_1$ is the first-mode vector product of the tensor $\mathbf{X}$ and the vector $\mathbf{y}$). Then the vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ are estimated decomposing the tensor $\mathbf{Z}$: $\mathbf{Z} = \mathbf{w}_1 \circ \mathbf{w}_2 \circ \mathbf{w}_3 + \mathbf{E}$.

$$||\mathbf{Z} - \mathbf{w}_1 \circ \mathbf{w}_2 \circ \mathbf{w}_3||_F \rightarrow \min,$$

where $||.||_F$ is the Frobenius norm, which is the generalization of the Euclidean norm for tensors (Kolda and Bader 2007).

To solve the optimization problem (1) the alternating least squares (ALS) (Yates 1933) algorithm can be applied (see appendix A). ALS is an iterative procedure. It fixes all the projectors except one, which is estimated in a least-square sense. After the projectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ and the latent variable $\mathbf{t}_1$ are defined, a coefficient $b_1$ of a regression $\mathbf{y} = b_1 \mathbf{t}_1 + \mathbf{e}_1$ is estimated using the minimal least squares (MLS) approach. The next factors are calculated decomposing the residuals. After $F$ iterations, all the particular regressions $\hat{\mathbf{y}}_f = \mathbf{T}_f \mathbf{b}_f, f = 1, \ldots, F$ are summarized to a final model $\hat{\mathbf{y}} = \sum_{f=1}^F \mathbf{T}_f \mathbf{b}_f = \mathbf{Tb}$. A vector $\mathbf{b}$ represents the regression coefficients for the whole set of latent variables $\mathbf{T} = [\mathbf{t}_1 | \ldots | \mathbf{t}_F]$.

**L1-Penalized NPLS algorithm**

The NPLS approach can be generalized including additional feature selection opportunities. For this purpose, the ALS algorithm can be substituted by its penalized version for decomposition of the tensor $\mathbf{Z} = \mathbf{X} \times_1 \mathbf{y}$:

$$||\mathbf{Z} - \mathbf{w}_1 \circ \mathbf{w}_2 \circ \mathbf{w}_3||_F^2 + P(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3) \rightarrow \min.$$

Here, $P(\cdot)$ is a penalization term. Penalization is widely used for solving various optimization tasks (Tychonoff and Arsenin 1977). For different applications various penalization operators were considered: the least absolute shrinkage selection operator (LASSO), $P(\mathbf{w}) = \lambda \|\mathbf{w}\|_1$ (Tibshirani 1996); the fusion lasso $P(\mathbf{w}) = \lambda \|D\mathbf{w}\|_1$, where $D$ is a difference operator (Land and Friedman 1996); the elastic net (Enet) (Zou and Hastie 2005) (weighted $\ell_1$-norm and $\ell_2$-norm); etc. Here $\lambda$ is a non-negative penalization parameter. To obtain a sparse solution, often the $\ell_1$-norm penalty (LASSO) is applied. LASSO can be implemented easily providing a sufficient level of selectivity. Alternating penalized least squares proposed in Martínez-Montes et al (2008) combines L1 and L2 penalties. In this study, the $\ell_1$-norm penalty was integrated into the ALS algorithm considering $P(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3) = \lambda_1 \|\mathbf{w}_1\|_1 + \lambda_2 \|\mathbf{w}_2\|_1 + \lambda_3 \|\mathbf{w}_3\|_1$. At each step of the algorithm, all projectors are fixed except one. That leads to the optimization task:

$$\mathbf{w}_i = \arg \min_{\mathbf{w}_i} \left( ||\mathbf{Z} - \mathbf{w}_1 \circ \mathbf{w}_2 \circ \mathbf{w}_3||_F^2 + \lambda_1 \|\mathbf{w}_1\|_1 \right),$$

$i = 1, 2, 3$.

Penalized decomposition of tensor $\mathbf{Z} = \mathbf{X} \times_1 \mathbf{y}$ results in factor $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ (see appendix B).

To select the optimal value of parameters of penalization, different approaches can be used: e.g. cross-validation (Devijver and Kittler 1982), Akaike’s information criterion (Akaike 1974), Schwartz’s Bayesian information criterion (Schwartz 1978), etc.

The algorithm results in a regression model predicting output $\mathbf{y}$ from the observations $\mathbf{x} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$.

The L1-Penalized NPLS algorithm combines computational simplicity and moderate memory consumption with sufficient selectivity. This method was applied for the binary self-paced BCI system calibration and for the selection of a subset of electrodes in the context of BCI experiments in non-human primates. Estimated at the calibration stage the predictive model is intended to send commands to the effector during execution in real-time applications.

**Influence analysis**

The elements of the input data have an implicit impact on the NPLS regression model through the latent variables. The modality influence (MI) analysis (Cook and Weisberg 1982) allows estimating the relative importance of the elements of each mode for the final model.

The MI analysis can be applied to estimate the importance of electrodes, frequency bands and time intervals related to control events (Eliseyev 2011).

**Application**

**Data description**

Data were collected in behavioral experiments in non-human primates based on a simple reward-oriented task. During the experiment the monkey is sitting in a custom-made primate chair minimally restrained. Its neck is collar hooked to the chair. The monkey has to push a pedal which can be mounted in four different positions (denoted below as ‘left’, ‘right’, ‘up’ and ‘down’) on a vertical panel facing the monkey. Every correct pushing event activates a food dispenser. We did not use any cue or conditioning stimulus to trigger the pressing intention of the monkey. A set of ECoG recordings was gathered by recording 16 bipolar electrodes placed in four different positions (denoted below as ‘left’, ‘right’, ‘up’ and ‘down’) on a vertical panel facing the monkey.
collected from 32 surface electrodes, chronically implanted on the cortex of the monkey (figure 1(C)). Simultaneously, information about the state of the pedal was stored. The ECoG signals were obtained at a sampling rate of 1 kHz, using the Micromed system (Micromed SD64, Micromed, Italy). The signals were band-pass filtered between 0.5 and 500 Hz.

**BCI system calibration**

One recording of each position was used to calibrate the BCI system for this position (figure 1(A)). Training data sets included all event-related epochs (at least 50 trials) and randomly selected 1000 ‘non-event’ epochs over all the experiment.

To form a tensor of observation the brain activity signals were mapped to the temporal–frequency–spatial space. For each epoch $j$ (determined by its final moment $t_j$), electrode $c$, frequency $f$ and time shift $\tau$, elements $x_{j,\tau,f,c}$ of the tensor $X$ were calculated as norm of CWT of ECoG signal (see figures 1(B) and (D)). Frequency bands of 10–300 Hz with step $\delta f = 2$ Hz and sliding windows $[t - \Delta \tau, t]$, $\Delta \tau = 0.5$ s with step $\delta \tau = 0.01$ s were considered for all electrodes $c = 1, 2, \ldots, 32$. The resulting dimension of a point is $146 \times 51 \times 32$.  

![Figure 1](https://example.com/figure1.png)
After comparing the set of mother wavelets, the Meyer wavelet was chosen for signal decomposition (see appendix C).

The binary-dependent variable was set to one, $y_j = 1$, if the pedal was pressed at the moment $t$, otherwise the binary-dependent variable was set to zero, $y_j = 0$. The resulting tensors of observations and the binary vector, indicating the pedal position, were used for modeling.

Five factors and the corresponding latent variables $l_i$, $i = 1, 2, \ldots, 5$ were extracted by the Penalized NPLS algorithm for each pedal position. To find a subset of electrodes, most impacting the final model, the L1-penalized version of the NPLS algorithm was applied to the spatial modality $(\lambda_1 = 0.9(\lambda_1)_{\text{max}}, \lambda_2 = \lambda_3 = 0)$, where $(\lambda_1)_{\text{max}}$ is defined by the L1-penalized algorithm, see appendix B, figure A1). The penalization parameter $\lambda_1$ was chosen taking into account real-time computational restrictions imposed by the system. Figure 2 illustrates time consumption for online prediction (Intel Dual Core, 3.16 GHz) as a function of the number of electrodes. The calculation time is estimated for a signal acquisition system using a 0.5 s buffer. Taking into account data collection, transfer, visualization of results, hard drive access time, operation system needs, etc, with our BCI system recordings of about ten electrodes can be processed in real-time. The parameter $\lambda_1$ was chosen to provide a high level of selectivity, i.e. less than ten electrodes with non-zero weights in the final model for all positions of the pedal.

To transform the prediction of output variable $\hat{y}$ to the BCI system decision (‘event’–‘non-event’), a simple decision rule, based on binarization of the predicted value, was applied. A scalar threshold was chosen in such a way to maximize the overall performance (OP) criterion (Eliseyev 2011) of the offline playbacks of the recordings of BCI experiments (see details below).

**BCI evaluation**

The variety of experimental paradigms and evaluation criteria (Leeb et al. 2007, Scherer et al. 2008, Fatourechi et al. 2008, Müller-Putz et al. 2010, Qian et al. 2010) makes BCI system comparison complicated. The accuracy of classification (ACC) referred to also as decoding accuracy (DA) (Ball et al. 2009) or decoding power (DP) (Rickert et al. 2005) is a commonly used evaluation criterion in BCI research. It shows the percentage of correctly classified samples. However, being efficient for the cue-paced BCIs, ACC as well as error rate (ERR = $1 - \text{ACC}$), fails to characterize the performance of the self-paced BCIs, due to highly unbalanced classes (Schl{"o}gl et al. 2007). ACC depends on class ratio, which varies essentially even within the same series of experiments. For example, in the given set of experiments, the class ‘zero’ is represented by about 90% of buffers. That is why other criteria are applied to characterize the efficiency of the self-paced BCIs. The true positive rate (TPR = $\text{TP}/(\text{TP} + \text{FN})$), and the false positives rate (FPR = $\text{FP}/(\text{FP} + \text{TN})$) are widely used to evaluate the self-paced BCIs performance. They show the percentages of errors for individual classes, namely, the relative amount of successfully detected events (TPR) and the relative amount of false activations (FPR). Nevertheless, FPR is also influenced by the decision rate and the ratio of the classes. Additional criteria characterizing false activations of the self-paced BCIs were proposed: the number of false activations per minute (Mason and Birch 2000) and the positive predictive value (PPV = $\text{TP}/(\text{TP} + \text{FP})$) (Müller-Putz et al. 2010). While TPR shows the percentage of successfully detected events, PPV corresponds to the percentage of correct detections. However, simultaneous comparison of several criteria is complicated. Since standard ACC (DA) fails to evaluate the performance of self-paced BCIs, numerous attempts were made to introduce a single metrics: weighted ACC (Zhu and Yao 2004), HF-difference (Huggins et al. 1999), $F_1$-criterion (Rijssbergen 1979), TPR at a fixed false positive rate (Borisoff et al. 2004), ratio TPR/FPR (Fatourechi et al. 2007) and others. In our study, the overall performance characteristic, balancing FP and FN types of errors, $\text{OP} = (\text{TPR} + \text{PPV})/2$ was chosen.

Introducing the chance level (CL) can help to correlate and compare the systems of different paradigms. In synchronized BCIs the Bernoulli scheme is applied explicitly or implicitly to estimate the level of decoding by chance, according to the probabilities of classes. It results in 50% of CL for classification accuracy in the case of two classes of equal probabilities, 25% for four equiprobable classes, etc. The self-paced BCI proceeds continuous monitoring of neuronal activity. In this case, the assumption of the Bernoulli scheme for the sequence of independent and identically distributed random variables can lead to misestimates. That is why we decided to estimate the CL with computational experiments for given datasets. Namely, for each recording the random detection was made with probabilities of classes estimated from recordings. The computational experiments are explained below.

**Simulation of BCI experiments**

To study the generalization ability of the predictive model, a set of offline playbacks of the BCI experiments were carried out for each position of the pedal. The decision (‘event’ or ‘non-event’) was made analyzing buffer by buffer the entire recordings. Buffers were of 0.5 s length in accordance with the
real-time data acquisition system of the CLINATEC/CEA BCI experimental platform. The predictors were calculated every 0.125 s, i.e., four times per buffer. The buffer was considered as containing the ‘event’ if at least one of these predictors passed the threshold of binarization. After any detection, the system was blocked for 1.5 s to prevent multiple activations. The real event was considered as detected (TP) if the time interval between the real event and its detection did not exceed the TP interval. To optimize the TP interval the delay distribution was analyzed for all training recordings taking into account the detection events within a 1.5 s interval around the real event (Fatourechi et al. 2008). Then 95% confidential interval was calculated and rounded to the nearest entire buffer. The resulted TP interval was ±0.5 s (figure 3).

The same experiments were carried out to estimate CL. The detections were randomly generated with probability of class ‘events’ in the given recording. The same true positive and silent intervals (0.5 s and 1.5 s) were applied.

Results

Results of calibration

The BCI system was calibrated for each pedal position: ‘left’, ‘right’, ‘up’ and ‘down’. Namely, the signal of the corresponding training dataset was mapped to the frequency–spatial–temporal space. Then, five factors and the corresponding latent variables were extracted by L1-Penalized NPLS. The number of factors was determined by the cross-validation procedure. The coefficients $b_i^*$ of the normalized predictive model $\hat{y} = \sum_{i=1}^{5} t_i^* b^*_i$ correspond to weights of the related factors in the final decomposition:

- ‘left’: 0.346, 0.273, 0.232, 0.111, 0.038;
- ‘right’: 0.346, 0.217, 0.195, 0.138, 0.104;
- ‘up’: 0.383, 0.263, 0.158, 0.151, 0.045;
- ‘down’: 0.278, 0.210, 0.194, 0.182, 0.138.

The resulting predictive models are based on subsets of few electrodes: 6, 6, 7 and 9 for respective positions ‘left’, ‘right’, ‘up’ and ‘down’ of the pedal. The corresponding most influenced factors are shown in figure A2. MI analysis revealed the summarized influence of elements of each modality (figure A3).

Applied to the spatial modality, the MI analysis revealed that the electrode #22, located in the primary motor cortex, has the highest impact on the decision rule (46%, 71%, 56% and 37% of extracted information for respective positions ‘left’, ‘right’, ‘up’ and ‘down’ of the pedal). High frequencies ($\geq 100$ Hz) are crucial for decisions. However, the influence of the lower frequencies ($< 100$ Hz) can be important too, especially for the position ‘left’ of the pedal. In the time domain the interval (−0.2, 0) s before the event is the most a significant for all positions of the lever.

Comparison of L1-Penalized NPLS to generic NPLS

To compare the L1-Penalized NPLS method with the generic algorithm, the BCI system was calibrated with ordinary NPLS (figure A4) using the same training tensors. Figure 4 shows the weights of the different electrodes in the prediction rules obtained by the generic NPLS and L1-Penalized NPLS approaches for different positions of the pedal. The models were compared on the test recordings for each pedal position and for a different number of factors. The computational experiment revealed that the L1-Penalized NPLS provides comparable results (RMSE) or outperforms the generic NPLS (figure 5). At the same time, the penalized algorithm uses a significantly restricted subset of electrodes.

Single electrode calibration

The single electrode calibration was the next step. Five factors were extracted by NPLS for the electrode (#22) which is the best in all the cases (figure 4). The relative weights of these factors in the final decomposition are as follows:

- ‘left’: 0.419, 0.234, 0.116, 0.092, 0.089;
- ‘right’: 0.371, 0.228, 0.137, 0.122, 0.079;
- ‘up’: 0.313, 0.201, 0.182, 0.137, 0.092;
- ‘down’: 0.232, 0.212, 0.167, 0.121, 0.106.

The leverages of elements of each modality for the respective best electrode according to the MI analyses are shown in figure A5. The results of the calibration procedure are the single-electrode predictor of the pedal-pressing events and the threshold-based decision rule.

Validation of the generalization ability of predictive models

To study the generalization ability, the NPLS, the L1-Penalized NPLS and the single-electrode predictive models were tested in a set of offline playbacks of previously collected data. For this purpose, the recordings of four series of behavioral experiments, $6 \times 4 = 24$ experiments in total were used. The experiments lasted 4 to 20 min, about 8 min on average. Thresholds of binarization were adjusted to each recording. Playbacks of BCI experiments were carried out offline to estimate the BCI performance. Part of one experiment (about 1.5 min) and corresponding time-delay histograms of detections for the single-electrode calibration are represented.
Figure 4. The weights of the different electrodes in the prediction models, obtained by the Generic NPLS and L1-Penalized NPLS for different positions of the pedal.

Figure 5. Comparison of prediction errors (root mean squared error, RMSE) of the NPLS and the L1-Penalized NPLS algorithms on the test sets of different numbers of factors and different positions of the pedal.

Figure 6. 1.5 min length fragment of the experiment and its time-delay histograms of detection.
The self-paced BCI requires a high level of selectivity for identification and discrimination of specific neuronal activity against background brain activity, during continuous monitoring. To achieve the necessary level of selectivity, the multi-way analysis was chosen since it provides simultaneous signal processing in several domains. To extract knowledge from the experimental data, NPLS was applied as a basic approach for the BCI system calibration. This method requires neither exhaustive search of the model, nor regularization of the task. The disadvantage of this projection-based method is that the final model includes all available variables. While the main goal of the study was to discriminate the specific neuronal pattern, related to the control action, an additional goal was to make the decision using a few electrodes or even only one electrode. To overcome the problem, the L1-penalization was incorporated into the NPLS method, providing the sparse capacity by means of feature selection integration to the generic algorithm.

Dimension reduction, in general, and selection of effective feature subsets, in particular, are the important problem in BCI studies. Traditional approaches use either projections, or sorting procedures. L1-penalization, integrated to the projection-based NPLS algorithm, provides sparse projections and allows integrated feature selection. Applied to the spatial modality, it results in selection of electrodes subset. Let us stress that tensor data representation and processing allow introducing different penalties in different directions, according to the particular problem (feature selection with L1 penalty in one modality, robust solution with L2 penalty in another, etc). Moreover, additional restrictions can be introduced to the optimization in particular directions (e.g. non-negativity, unimodality, etc).

### Discussion

The performance of offline playbacks of BCI experiments is shown in Table 1 and Figure 7. A summary of the results of the performance evaluation is shown in Table 1 and Figure 7.

To compare the results to the CL the random detection procedure was applied to the test recordings.

<table>
<thead>
<tr>
<th>Method</th>
<th>No. of events</th>
<th>Time (s)</th>
<th>TPR (%)</th>
<th>PPV (%)</th>
<th>FPR (%)</th>
<th>ERR (%)</th>
<th>OP (%)</th>
<th>FP (min⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left NPLS</td>
<td>81 ± 33</td>
<td>493 ± 197</td>
<td>61.0 ± 7.1</td>
<td>62.5 ± 6.7</td>
<td>3.24 ± 1.12</td>
<td>6.10 ± 2.06</td>
<td>61.7 ± 6.8</td>
<td>3.53 ± 1.15</td>
</tr>
<tr>
<td>L1-PNPLS</td>
<td>63.7 ± 10.2</td>
<td>64.3 ± 5.6</td>
<td>3.16 ± 1.23</td>
<td>5.70 ± 1.84</td>
<td>64.0 ± 7.8</td>
<td>3.44 ± 1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPLS(1el)</td>
<td>60.8 ± 5.7</td>
<td>61.9 ± 6.5</td>
<td>3.28 ± 1.07</td>
<td>6.16 ± 1.97</td>
<td>61.3 ± 1.87</td>
<td>3.57 ± 1.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random</td>
<td>6.46 ± 2.13</td>
<td>8.03 ± 2.45</td>
<td>6.53 ± 2.07</td>
<td>13.71 ± 4.40</td>
<td>7.25 ± 2.24</td>
<td>7.13 ± 2.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right NPLS</td>
<td>71 ± 25</td>
<td>437 ± 137</td>
<td>59.5 ± 23.3</td>
<td>58.0 ± 21.0</td>
<td>4.50 ± 1.39</td>
<td>7.61 ± 1.38</td>
<td>58.8 ± 21.9</td>
<td>4.78 ± 1.42</td>
</tr>
<tr>
<td>L1-PNPLS</td>
<td>62.2 ± 21.4</td>
<td>64.2 ± 19.0</td>
<td>3.67 ± 1.85</td>
<td>6.68 ± 2.14</td>
<td>63.2 ± 20.0</td>
<td>3.90 ± 1.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPLS(1el)</td>
<td>62.7 ± 6.0</td>
<td>63.2 ± 9.3</td>
<td>3.19 ± 1.00</td>
<td>5.97 ± 1.43</td>
<td>62.7 ± 7.3</td>
<td>3.49 ± 1.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up NPLS</td>
<td>52 ± 11</td>
<td>462 ± 329</td>
<td>63.1 ± 12.4</td>
<td>62.1 ± 11.9</td>
<td>3.41 ± 1.34</td>
<td>5.92 ± 1.53</td>
<td>62.6 ± 11.8</td>
<td>3.71 ± 1.35</td>
</tr>
<tr>
<td>L1-PNPLS</td>
<td>72.5 ± 6.8</td>
<td>75.4 ± 8.3</td>
<td>2.21 ± 0.94</td>
<td>4.20 ± 1.26</td>
<td>73.9 ± 6.4</td>
<td>2.40 ± 0.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPLS(1el)</td>
<td>63.7 ± 20.6</td>
<td>63.0 ± 20.9</td>
<td>3.90 ± 1.40</td>
<td>6.75 ± 2.17</td>
<td>63.4 ± 20.7</td>
<td>4.15 ± 1.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random</td>
<td>5.15 ± 3.42</td>
<td>6.67 ± 4.57</td>
<td>6.80 ± 2.04</td>
<td>14.15 ± 4.35</td>
<td>5.91 ± 3.98</td>
<td>7.41 ± 1.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Down NPLS</td>
<td>59 ± 18</td>
<td>486 ± 186</td>
<td>47.8 ± 15.2</td>
<td>48.0 ± 15.8</td>
<td>3.31 ± 0.44</td>
<td>6.20 ± 0.90</td>
<td>47.9 ± 15.5</td>
<td>3.71 ± 0.43</td>
</tr>
<tr>
<td>L1-PNPLS</td>
<td>55.6 ± 15.0</td>
<td>52.3 ± 16.1</td>
<td>3.20 ± 0.46</td>
<td>5.59 ± 0.49</td>
<td>53.9 ± 15.5</td>
<td>3.59 ± 0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPLS(1el)</td>
<td>52.6 ± 18.9</td>
<td>51.9 ± 18.6</td>
<td>3.04 ± 0.59</td>
<td>5.58 ± 0.78</td>
<td>52.3 ± 18.7</td>
<td>3.41 ± 0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random</td>
<td>3.10 ± 5.13</td>
<td>3.90 ± 6.06</td>
<td>5.21 ± 1.82</td>
<td>11.04 ± 3.45</td>
<td>3.50 ± 5.59</td>
<td>5.81 ± 1.89</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 7.** Comparison of the overall performance in the set of computational experiments for the Single-Electrode NPLS, the Generic NPLS, the L1-Penalized NPLS, as well as the Random Detection procedure (chance level).
To achieve the main goal, namely, the fully autonomous self-paced BCI functioning in a natural environment, we have analyzed the recordings of the series of behavioral experiments in the non-human primate, controlling the food dispenser by pushing the lever. The duration of experiments varied from 4 to 20 min. The factors extracted by Penalized NPLS can be interpreted taking into account their influence on the final model with the aid of the MI analysis. In our research, we found out that electrodes located in motor primary cortex (especially, the electrode #22) have the strongest influence on the decision. Additional experiments will allow better study of the stability of signal’s localization and evolution with time, brain’s plasticity, etc. In all series of experiments, 6–9 electrodes out of 32 were involved in the final models. In other modalities, the most significant influence on the decisions has the increase of signal energy in the high frequencies (>100 Hz) and the time interval 0.2 s length before the pushing event. Frequency and temporal modalities of the most influenced factor show high gamma synchronization and beta desynchronization in the time interval 0.1 s before the pushing event (figure A2).

The variety of experimental paradigms and evaluation criteria complicates BCI systems comparison. In this paper we use the OP criterion, due to its efficiency for self-paced BCIs. OP for the respective pedal positions (64.0% for ‘left’, 63.2% for ‘right’, 73.9% for ‘up’ and 53.9% for ‘down’) is achieved for the restricted set of electrodes (from 6 to 9). For the single-electrode case, OP equals 61.3% for the ‘left’, 62.7% for the ‘right’, 63.4% for the ‘up’ and 52.3% for the ‘down’ positions of the pedal, correspondingly. The relatively low level of correct detections for the position ‘down’ can be explained by the experiment setup (the monkey did not make a sufficient effort to reach this pedal). At the same time L1-Penalized NPLS outperforms the generic NPLS. It can be explained by excluding non-informative electrodes from the predictive model by the penalized version of algorithm.

The CL can be considered as the basis for the comparison of BCIs of different paradigms and using different criteria. The CL for the OP criterion is about 6.5% for our experiments, which corresponds to CL of the synchronized BCI with 15 equiprobable classes. For comparison, in previous studies, for the case of the BCI with eight equiprobable classes (CL = 12.5%) decoding accuracy (DA) ≈ 25% was achieved using one electrode, whereas DA ≈ 50% using eight electrodes (Rickert et al 2005). DA ≈ 50% was demonstrated for ten electrodes for the local field potential recordings (Mehring et al 2003), and up to 60% for ECoG (Ball et al 2009). Finally, only the intervals preceding the events were used to create the predictive model, while in most studies, the intervals before and after the movement are analyzed as well (e.g. Ball et al 2009).

Acknowledgments
This work was partially supported by project CE ICoBI, Nanosciences Foundation RTRA, Edmond J Safra Philanthropic Foundation, Fondation de l’Avenir, France.

Appendix A. Alternating least squares
ALS is an iterative procedure. Vectors $\mathbf{w}^1, \mathbf{w}^2, \mathbf{w}^3$ are initialized by $\mathbf{I}$. Then, ALS fixes all the projectors except one, which is estimated in a least-square sense: $\hat{\mathbf{w}}^i = \arg\min_{\mathbf{w}^i} (\|\mathbf{Z} - \mathbf{w}^1 \circ \mathbf{w}^2 \circ \mathbf{w}^3\|_F^2)$, where $\mathbf{w}^2, \mathbf{w}^3$ are fixed; $\mathbf{w}^1$ is updated.
Figure A2. The first factor: frequency, temporal and spatial projections for every position of the pedal. The values of elements of the spatial projectors are shown in colors according to the color bar; the electrode positions are indicated by their numbers.

\[ \hat{w}^2 = \arg \min_{w^2} (\|Z - w^1 \circ w^2 \circ w^3 \|^2) \], \quad w^1, w^3 \text{ are fixed;}
\[ \hat{w}^3 = \arg \min_{w^3} (\|Z - w^1 \circ w^2 \circ w^3 \|^2) \], \quad w^1, w^2 \text{ are fixed.} \]

As a matrix:
\[ \hat{w}^1 = \arg \min_{w^1} (\|Z_{(1)} - w^1 w_{1,3}^T \|^2) \],
\[ \hat{w}^2 = \arg \min_{w^2} (\|Z_{(2)} - w^2 w_{2,3}^T \|^2) \],
\[ \hat{w}^3 = \arg \min_{w^3} (\|Z_{(3)} - w^3 w_{3,2}^T \|^2) \],

where matrix \( Z_{(i)} \) is the unfolding of the tensor \( Z \) along the modality \( i \), \( w_{i,j} = \text{vect}(w_i \circ w_j) \). The solutions of the optimization problems are:
\[ \hat{w}^1 = Z_{(1)} w_{2,3} (w_{2,3}^T w_{2,3})^{-1}, \]
\[ \hat{w}^2 = Z_{(2)} w_{1,3} (w_{1,3}^T w_{1,3})^{-1}, \]
\[ \hat{w}^3 = Z_{(3)} w_{1,2} (w_{1,2}^T w_{1,2})^{-1}. \]

The ALS procedure is repeated until convergence.

Appendix B. L1-Penalized alternating least squares

L1-Penalized alternating least squares is an iterative procedure. Vectors \( w^1, w^2, w^3 \) are initialized by \( 1 \). Then, all projectors except one are fixed, that leads to the task:
\[ \hat{w}^i = \arg \min_{w^i} (\|Z - w^1 \circ w^2 \circ w^3 \|^2_F + \lambda_i \|w^i\|_1), \]
\[ i = 1, 2, 3. \]
Figure A3. L1-Penalized NPLS calibration: impact (weights) of components of different modalities on the predictive model for each pedal position (according to the MI analysis). Spatial modalities are represented by the graphs and corresponding color maps.

Consider the particular case where $i = 1$:

$$\hat{w}^1 = \arg \min_{w^1} (\|Z - w^1 \circ w^2 \circ w^3\|_F^2 + \lambda_1 \|w^1\|_1),$$

or as a matrix:

$$\hat{w}^1 = \arg \min_{w^1} (\|Z^{(1)} - w^1 w^T_{2,3}\|_2^2 + \lambda_1 \|w^1\|_1) \quad (B.1)$$

The optimization problem related to the $\ell_1$-norm penalization can be solved using various approaches (see Schmidt 2005). In this study we applied an algorithm proposed in Shavade and Keerthi (2003); Schmidt (2005). The advantages of this algorithm are its simplicity and its low iteration cost, as well as its low memory consumption. We have applied this algorithm to solve the optimization task (B.1). Namely, the anti-gradient of a functional $\left(\|Z^{(1)} - w^1 w^T_{2,3}\|_2^2 + \lambda_1 \|w^1\|_1\right)$ was calculated:

$$-G(w^1) = 2w^T_{2,3} (Z_{(1)} - w^1 w^T_{2,3}) - \lambda_1 \text{sign}(w^1).$$

For the first iteration, $w^1$ is set equal to zero: $-G(0^+) = 2w^T_{2,3} Z_{(1)} - \lambda_1 1$. Then, the elements of $w^1$ with the largest magnitude of the anti-gradient are added to a set of ‘free’ variables. These ‘free’ variables are optimized in a ‘one at a
time’ way. For details see Shevade and Keerthi (2003). Note, that if $\lambda_1 \geq \lambda_{\max} = \max (2w_1^T Z^T(t))$, the method returns as a solution $\hat{w}_1 = 0$.

Appendix C. Mother wavelet selection

Several mother wavelets $\psi$ (Meyer, Morlet, Symlet ‘7’ and ‘8’, 2nd and 10th order Debauchies, Coiflets ‘5’, and Haar) were compared to be used for signal decomposition. Results were evaluated with respect to the maximum level of Pearson correlation between the absolute value of the wavelet’s coefficients $C_\psi(t - \tau, s)$ and the signal of the pedal $y(t)$:

$$R_\psi = \max \{| \text{corr} (C_\psi(t - \tau, s), y(t)) | \}$$

using all recordings representing all series of experiments. Pearson correlation coefficients were calculated for scale factors $s$, corresponding to the frequencies of the band $[10, 300]$ Hz and time shifts $\tau \in [0, 0.5]$ s. $y(t) \in \{0, 1\}$ represents the position of the pedal at the moment $t$. Comparison shows that the Haar mother wavelet leads to an unstable and relatively low level of correlation, whereas the performances of all other wavelets are comparable (figure A1). Considering computational efficiency (Sherwood and Derakhshani 2009) the Meyer wavelet was chosen as the mother wavelet to form the tensor of observation $X$. 

Figure A4. Generic NPLS calibration: impact (weights) of components of different modalities on the predictive model for each pedal position (according to the MI analysis). Spatial modalities are represented by the graphs and corresponding color maps.
Figure A5. Single-electrode calibration: impact on the predictive model of the components of the different modalities according to the MI analysis for each pedal position.

References

Acar E, Aykut-Bingol C, Bingol H, Bro R and Yener B 2007 Multiway analysis of epilepsy tensors Bioinformatics 23 i10–i8
Ball T, Schulze-Bonhage A, Aertsen A and Mehring C 2009 Differential representation of arm movement direction in relation to cortical anatomy and function J. Neural Eng. 6 016006
Barachant A, Aksenova T and Bonnet S 2008 Filtrage spatial robuste à partir d’un sous-ensemble optimal d’ électrodes en BCI EEG 22nd Coll. GRETSI (Dijon, France) 2009
Barachant A and Bonnet S 2011 Channel selection procedure using Riemannian distance for BCI applications 5th Int. IEEE/EMBS Conf. on Neural Engineering (NER), 2011 pp 348–51
Chao Z C, Nagasaka Y and Fujii N 2010 Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkeys Front. Neuroeng. 3 1–10
Cook R D and Weisberg S 1982 Residuals and Influence in Regression (London: Chapman and Hall)
Kolda T G and Bader B W 2007 Tensor decompositions and applications Sandia Report SAND2007-6702
Li J and Zhang L 2010 Regularized tensor discriminant analysis for single trial EEG classification in BCI Pattern Recognit. Lett. 31 619–28
Lv J and Liu M 2008 Common spatial pattern and particle swarm optimization for channel selection in BCI 3rd Int. Conf. on Innovative Computing Information and Control (ICICIC ’08) p 457
Mørup M, Hansen L K, Herrmann C S, Parnas J and Arnfred S M 2006 Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG Neuroimage 29 938–47


Rijsbergen C J 1979 Information retrieval. Available at http://www.dcs.gla.ac.uk/~iain/keith/


Schmidt M 2005 Least squares optimization with L1-norm regularization Project Report University of British Columbia


Vidaurre C, Krämer N, Blankertz B and Schlögl A 2009 Time domain parameters as a feature for EEG-based brain computer interfaces Neural Netw. 22 1313–9


Yates F 1933 The analysis of replicated experiments when the field results are incomplete Empire J. Exp. Agric. 1 129

Zhao Q, Caiafia C F, Cichocki A, Zhang L and Phan A H 2009 Slice oriented tensor decomposition of EEG data for feature extraction in space, frequency and time domains Lecture Notes in Computer Science vol 5863 (Berlin: Springer) pp 221–8
