Neurocomputing 470 (2022) 236–246

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Spatial filtering based on Riemannian distance to improve the generalization of ErrP classification ☆

Aniana Cruz [a,*], Gabriel Pires [a,b], Urbano J. Nunes [a,c]

[a] Institute of Systems and Robotics, University of Coimbra, Portugal
[b] Engineering Department, Polytechnic Institute of Tomar, Portugal
[c] Department of Electrical and Computer Engineering, University of Coimbra, Portugal

ARTICLE INFO

Article history:
Received 26 April 2021
Revised 1 August 2021
Accepted 24 October 2021
Available online 29 October 2021
Communicated by Zidong Wang

Keywords:
Brain-computer interface (BCI)
Error-related potentials (ErrP)
Spatial filter
Riemannian geometry
Generalization
Zero-calibration
Single-calibration

ABSTRACT

Due to the inherent non-stationarity of EEG signals, before each experimental session, BCI is usually calibrated to build the classification models, thus avoiding performance decay. This tedious re-calibration procedure is a limiting factor for real-world applications. Therefore, single-calibration or zero-calibration plays a crucial role in the use of BCIs in real contexts, outside the laboratory. Here, we propose and validate a statistical spatial filter, Riemannian Fisher criterion beamformer, based on Riemannian geometry able to use the invariance properties of Riemannian distance to handle cross-session and cross-subject generalization. The proposed method is validated with two datasets publicly available, consisting of error-related potentials. The results show that the proposed filter improves the generalization across sessions and across subjects and that it is robust to the amount of error samples used to train the classification model.

1. Introduction

BCI is being increasingly used for a variety of applications including clinical and non-clinical applications [1,2], however, it is still rarely used outside the laboratory. One of the major issues is the poor classification generalization between sessions and between subjects [3,4]. Therefore, to ensure a certain level of reliability it is usually necessary to calibrate the BCI system each time it is used, which is unpractical and takes time. This re-calibration procedure may be caused by: 1) the variations of users' mental states, such as stress and fatigue, and 2) instrumental issues, such as changes in electrode positioning and electrode impedance. As a result, the EEG signal typically changes across sessions and it changes dramatically across subjects. The classification model built with data collected from a session/subject usually works poorly or leads to a significant performance decrease in a new session/subject [5]. Currently, zero calibration has attracted a lot of attention and a variety of algorithms has been proposed to deal with EEG non-stationarity [6,3,7]. The calibration issue has been addressed using transfer learning (TL) techniques. Transfer learning can be defined as "the ability of a system to recognize and apply knowledge and skills learned from previous tasks to novel tasks" [8]. In the context of BCIs, the previous tasks can represent the previous sessions or subjects, and the novel task represents the current session or subject. Thereby, in a cross-session problem, the data of a new session is classified by learning from the past data recorded in previous sessions of the same user. In the cross-subject problem, the data of a new subject is predicted by learning from the data of other subjects.

Transfer learning methods have been used in BCIs to improve the robustness of feature extraction methods and classifiers. They can be divided into feature space learning and model space learning [9]. TL approaches based on feature space learning try to find a transformation to a data space in which the features are invariant across sessions or subjects and a single classification rule can classify all the data [10–13]. On the other hand, model space approaches attempt to learn how the decision rules differ across sessions or subjects [14,15]. Feature space learning is the approach most found in the literature, such as regularized common spatial patterns (CSP), and covariate shift adaptation [9,2]. Recently, Rie-

This work has been financially supported by Portuguese foundation for science and technology (FCT) under grant B-RELIABLE (PTDC/EEI-AUT/30935/2017) and grant ISR-UC UIDB/00048/2020. Aniana Cruz was supported by the Ph.D. FCT Scholarship SFRH/BD/111473/2015.

* Corresponding author.
E-mail addresses: anianabrito@isr.uc.pt (A. Cruz), gpires@isr.uc.pt (G. Pires), urbano@deec.uc.pt (U.J. Nunes).

https://doi.org/10.1016/j.neucom.2021.10.078
0925-2312/

mannian geometry has also attracted the attention of the BCI research community, due to its robustness and important invariance properties. Several approaches based on Riemannian geometry that works on the Riemannian manifold of the symmetric and positive definite (SPD) matrix (e.g. covariance matrix) have been proposed to improve transfer learning issues [16], spatial filtering [17], dimensionality reduction [18], outliers detection [19], calibration time reduction [20], and adaptive classification [21] (see [22,23] for reviews). The feature space based on trial covariance matrices plays a strong role in signal processing since it provides compact and relevant information [24]. Covariance matrices are widely used in spatial filters, such as CSP, which is one of the most popular techniques for feature extraction of EEG rhythms in BCI. A covariance matrix lies in a Riemannian manifold of the SPD matrices, which has been used to improve the robustness of CSP [25,26] and obtain new spatial filters [17]. The authors in [17] proposed a novel method called the tangent space spatial filter (TSSF) that obtains the optimal filter on the tangent space and mapped it back onto the manifold. Riemannian geometry has been shown to bring improvement for cross-session and cross-subject BCI learning due to its useful invariance properties, such as invariance under congruent transformations [5,16]. The variability of EEG signals across sessions or even subjects can be considered as a domain shift problem. Thus, some authors minimize the distance between the target and source domains using the data alignment approach [27,16,28]. More precisely, in [16] authors consider that the inter-session/subject variability induces shifts of covariance matrices with respect to a reference state. The authors proposed an affine transformation approach to align the covariance matrices from target and source domains by moving them to a common reference point using the Riemannian mean of the reference state (resting period) covariance matrices. In [28], instead of aligning the covariance matrices, the authors align the EEG trials in the Euclidean space to make the data distributions more similar, using the arithmetic mean of all covariance matrices as a reference matrix.

Riemannian geometry has been mostly used in motor imagery-based BCIs [29–31,17], but still relatively little used in the context of event-related potentials (ERPs). ERP is a phase-locked signal with time-specific properties, thus temporal information is essential for its detection. In order to include temporal information in the covariance matrix, authors in [32] proposed to build an augmented trial by concatenating a single trial and the mean of the target class. Then, the covariance matrices of the augmented trial are classified with the minimum distance to Riemannian mean (MDRM) algorithm [29]. The proposed method was tested in a P300-based game. Li et al. [33] combined XDAWN spatial filter [34], an affine transformation of covariances matrice and MDRM for cross-subject transfer learning in a P300-speller paradigm.

In this paper, we propose a new spatial filtering method called the Riemannian Fisher criterion beamformer (RFCB) that extends our previous Fisher criterion beamformer (FCB) filter introduced in [35] to a Riemannian manifold of SPD matrices. While in [35] the optimum filter is obtained in the Euclidean space and tested in subject-specific data, here we explore the Riemannian space to obtain a discriminative filter that can be used in subject-independent data. As in FCB, RFCB is used to project the temporal samples of the original EEG signal into a more discriminating dimensional space, still consisting of time samples, which fits ERP classification needs. RFCB is validated in the context of Error-related potentials (ErrPs). The ErrP is an ERP generated naturally in the brain after an incorrect selection [36]. It was first observed during speeded choice reaction tasks in the early 1990's [37]. Since then, four types of ErrPs were reported, namely: the response ErrP, the feedback ErrP, the recognition/observation ErrP, and the interaction ErrP [38]. In the context of BCI, ErrPs can be used as a secondary communication channel for example trying to correct errors of the primary communication [39]. In this case, we need to calibrate the primary and secondary communication channels, requiring two calibrations which results in a long calibration process. This problem can be mitigated if it is possible to reuse the data obtained from one session/subject to build the classification model to test in other sessions/subjects without great performance loss. Although the ErrP grand average is reported as stable between sessions, there is a significant decay in classification accuracy across sessions, since the ErrP classification has to be done at a single trial level [40,39,41]. The novel proposed RFCB uses the invariance properties of Riemannian distance to obtain a more robust filter that extracts discriminative and invariant features that are not strongly affected by cross-session and cross-subject variability.

The approach was validated in two ErrP datasets publicly available (an in–house dataset associated to a P300-based speller task [42] and a dataset associated to a task that consists of observing the performance of an external agent, available in the BNCI Horizon 2020 database[1]). RFCB outperformed the original methods used for comparison, achieving an average across-session classification of around 80.4%, and good performance across-subject with an average accuracy around 74.8%, comparing favorably with the sate of the art.

2. Riemannian manifold of symmetric positive definite matrices

In this section, we introduce some basic definitions and properties of SPD manifolds. Riemannian manifolds are smooth manifolds in which the tangent space is endowed with a smoothly varying inner product called the Riemannian metric [43]. Riemannian metrics can be used directly in the native space of covariance matrices, thus some approaches have been developed to increase spatial filter robustness which is based on covariance matrices [17].

In this work, we explore the congruence (transformation of the form $A \rightarrow XAX^T$) invariance of the Riemannian distance to extract "invariant" features that might improve the generalization across sessions/subjects.

2.1. Preliminaries and Notation

Let us consider a spatio-temporal matrix $X \in \Re^{N_c \times L}$ representing an EEG epoch $X = [x_1\ x_2\ \dots\ x_L]$ of N_c channels with L time samples. The spatial covariance matrix $P_i \in \Re^{N_c \times N_c}$ of the i-th trial is computed from

$$P_i = X_i X_i^T \tag{1}$$

The matrix P is said to be SPD if $P = P^T, x^T P x > 0, \forall x \neq 0$ and its eigenvalues are positives [44]. A domain $\mathcal{D} = \{\mathcal{X}, Q(x)\}$ comprises a d-dimensional feature space $\mathcal{X}$ and a marginal probability distribution $Q(x)$. Usually, different domains may have different feature spaces or different marginal probability distributions [8].

We consider the EEG data from previous sessions or subjects in the source domain $\mathcal{D}_S = \{(x_{S1}, y_{S1}), \dots (x_{SN}, y_{SN})\}$, where $x_{Si} \in \mathcal{X}_S$ is the data instance and $y_{Si} \in \mathcal{Y}_S$ is the associated class label. Similarly, the current sessions or subjects are in the target domain $\mathcal{D}_T = \{(x_{T1}, y_{T1}), \dots (x_{TN}, y_{TN})\}$, where $x_{Ti} \in \mathcal{X}_T$ and $y_{Ti} \in \mathcal{Y}_T$.

Let us consider

- $M(n) = \{M \in \Re^{n \times n}\}$ the space of $n \times n$ square matrices;
- $S(n) = \left\{S \in M(n), S = S^T\right\}$ the set of all $n \times n$ symmetric matrices in the space of $M(n)$;
- $P(n) = \{P \in S(n), P > 0\}$ the space of all symmetric positive-definite (SPD) matrices;

[1] http://bnci-horizon-2020.eu/database/data-sets.

- $\mathscr{S}^n_{++}$ the SPD manifold;
- $GL(n)$ the set of real invertible $n \times n$ matrices;
- $I_n \in \Re^{n\times n}$ the identity, and
- ^{T} the transpose operator.

2.2. Riemannian Distance

Riemannian manifolds are smooth differentiable manifolds equipped with metrics for estimation of similarity or dissimilarity between points. The Affine Invariant Riemannian Metric (AIRM) is the Riemannian distance between two SPD matrices $P1, P2 \in P(n)$ which is defined as the minimum length of the curves between them, called geodesic [45]

$$\delta_R(P_1, P_2) = \|log\left(P_1^{-1}P_2\right)\|_F = \left[\sum_i log^2 \lambda_i\right]^{\frac{1}{2}} \tag{2}$$

where $log(.)$ denotes the matrix logarithm, $\|.\|_F$is the Frobenius norm of a matrix, and λ_i are the real eigenvalues of $P_1^{-1}P_2$.

The AIRM $\delta_R(.,.)$ has many attractive invariance properties. For example the invariance to affine transformations by any invertible matrix $G \in GL(n)$ can be particularly relevant in BCI [16]: $\delta\left(GP_1G^T, GP_2G^T\right) = \delta(P_1, P_2)$.

Congruence invariance property is very important across sessions/subjects BCI learning because it implies that the distance between two SPD matrices remains unchanged after a linear transformation in the data. For example, the change in set up configurations between sessions (e.g., electrode positions and electrode impedance) produces a different EEG signal that consequently produces different covariance matrices, however, the AIRM is very similar between these sessions.

2.3. Robustness of Riemannian structure using EEG signal

The EEG signal $X(t) \in \Re^{N_c\times L}$ generated by the brain sources $s(t) \in \Re^M$ can be represented by

$$X(t) = As(t) \tag{3}$$

where $M \leqslant N_c$ is the number of EEG dipolar fields and $A \in \Re^{N_c\times M}$ is a mixing matrix [5]. The mixing matrix varies according to the physical properties of the head, the impedance of the electrodes, and electrode positioning. Thus, it strongly depends on the conditions of each session. Therefore, it is not surprising that cross-session BCI learning is a very challenging task and cross-subject learning is even more difficult. In order to analyze the variability between sessions let us consider two trials (i and j) of source domain (session 1) with the mixing matrix A_S and with the source covariance matrices C_i and C_j, and two trials of target domain (session 2) with the mixing matrix A_T, and the same source covariance matrices C_i and C_j. The sensor covariance matrices of the two trials from session 1 can be expressed as $P1_i = A_SC_iA_S^T$ and $P1_j = A_SC_jA_S^T$, and the sensor covariance matrices of the two trials from session 2 are given by $P2_i = A_TC_iA_T^T$ and $P2_j = A_TC_jA_T^T$. The difference between A_S and A_T results in covariance matrices that are different from each other, however the congruence invariance of the AIRM guarantees that

$$\delta(P1_i, P1_j) = \delta(C_i, C_j) \quad and \quad \delta(P2_i, P2_j) = \delta(C_i, C_j) \tag{4}$$

and hence, it yields

$$\delta(P1_i, P1_j) = \delta(P2_i, P2_j) \tag{5}$$

Therefore, the Riemannian distance between covariance matrices of session 1 is similar to the Riemannian distance between covariance matrices of session 2, thereby it is expected that a spatial filter that maximizes the Riemannian distance between classes will be identical either calculated using the data recorded in session 1 or the data from session 2. Thus, we propose to extend the FCB to the Riemannian manifold of the SPD matrix taking advantage of the congruence invariance property of the AIRM to minimize the effect of data variability in cross-sessions and cross-subjects.

3. Methods

3.1. Statistical spatial filter: Fischer Criterion Beamformer

A spatial filter is defined as a weighting vector, that combines the data of N_c channels at each time instant to obtain information of the true source of interest s(t)

$$Y = V^TX = V^TAs(t) \tag{6}$$

where Y denotes the output projections using the spatial filters $V^{N_c\times N_c}$.

FCB is one of the spatial filters proposed in our research lab [35], which extends the concept of Fisher linear discriminant (FLD) to the spatial domain. FLD aims to optimize class separability based on the within and between class spatial scatter matrices, in the Euclidean space.

Let S_w and S_b denote the spatial within-class scatter matrix and the between-class scatter matrix defined by

$$\mathbf{S}_w = \sum_i \sum_{k\in C_i} \left(\mathbf{X}_{i,k} - \mathbf{m}_i\right)\left(\mathbf{X}_{i,k} - \mathbf{m}_i\right)^T \tag{7}$$

$$\mathbf{S}_b = \frac{1}{K}\sum_i K_c(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \tag{8}$$

where K is the number of all trials, K_c is the number of trials in class c, m_i and m are the mean of the epochs in class c and the mean of all epochs respectively defined as

$$m_i = \frac{1}{K_c}\sum_{k=1}^{K_c}\mathbf{X}_{i,k} \qquad \text{and} \qquad m = \frac{1}{K}\sum_{k=1}^{K}\mathbf{X}_k \tag{9}$$

Using S_w and S_b, the Fisher's criterion (FC) is defined as the Rayleigh quotient

$$J(V) = \frac{V^TS_bV}{V^TS_wV} \tag{10}$$

where the optimal filter V simultaneously maximizes the between-class scatter matrix, and minimizes the within-class scatter matrix. The selected filter is the eigenvector associated with the largest eigenvalue and it is obtained by solving the generalized eigenvalue problem

$$\mathbf{S}_bV = [(I-\theta)\mathbf{S}_w + \theta I]V\Lambda \tag{11}$$

where I is the identity matrix, Λ is the eigenvalue matrix, and θ denotes the parameter obtained from training data, which increases class discrimination.

3.2. Riemannian Fisher Criterion Beamformer

The proposed RFCB algorithm computes the spatial filters $W^{N_c\times m}$ that project the EEG signal to a lower m-dimensional subspace ($m < N_c$) more discriminative, which simultaneously maximize the Riemannian distance between classes and minimizes the Riemannian distance within classes. The proposed spatial filter is inspired in [18] that uses a discriminant analysis approach on Riemannian manifold for dimensionality reduction in image and video classification. Consider $f(W,P) : \mathscr{S}^{N_c}_{++} \mapsto \mathscr{S}^m_{++}$ a generic mapping that maps a SPD matrix P from the original dimension N_c to

a lower dimension m defined as $f(W,P) = W^T PW$. The spatial filters W are computed to preserve as much as possible the intra-class compactness (intra-class distances) while maximizing the inter-class distances, by solving the following optimization problem

$$W = \arg\min_{W^T W = I_m} \mathscr{L}(W) \text{ where } \mathscr{L}(W) = \frac{\widetilde{S}_w(W)}{\widetilde{S}_b(W)} \tag{12}$$

where $\widetilde{S}_w(W)$ and $\widetilde{S}_b(W)$ are the within-class and between-class cost functions formulated as follows [31]

$$\widetilde{S}_w(W) = \frac{1}{2}\sum_{i,j\neq i}^{K} g_w(P_i,P_j)\delta^2\left(f(W,P_i),f(W,P_j)\right) \tag{13}$$

$$\widetilde{S}_b(W) = \frac{1}{2}\sum_{i,j\neq i}^{K} g_b(P_i,P_j)\delta^2\left(f(W,P_i),f(W,P_j)\right) \tag{14}$$

where $\delta(f(W,P_i),f(W,P_j))$ denotes the Riemannian distance, i and j are the trials, $g_w(P_i,P_j)$ and $g_b(P_i,P_j)$ are the within and between class similarity to compute the affinity between SPD matrices. g_w and g_b can be defined using several methods such as Fisher Discriminant Analysis (FDA), Local Scaling (LS), and k-Nearest Neighbor. Here we used the LS approach [46], given by

$$g_w(P_i,P_j) = \begin{cases} \frac{A(P_i,P_j)}{K_c}, & \text{if } P_i \text{ and } P_j \in \text{same class} \\ 0, & \text{if } P_i \text{ and } P_j \notin \text{same class} \end{cases} \tag{15}$$

$$g_b(P_i,P_j) = \begin{cases} A(P_i,P_j)\times\left(\frac{1}{K}-\frac{1}{K_c}\right), & \text{if } P_i \text{ and } P_j \in \text{same class} \\ \frac{1}{K}, & \text{if } P_i \text{ and } P_j \notin \text{same class} \end{cases} \tag{16}$$

where $A(P_i,P_j)$ is the affinity matrix defined as

$$A(P_i,P_j) = exp\left(-\frac{\delta^2(P_i,P_j)}{\rho_i\times\rho_j}\right) \quad with \quad \rho_i = \delta\left(P_i,P_i^{(k)}\right) \tag{17}$$

where $P_i^{(k)}$ is the k-th nearest neighbor of P_i, and $k = 7$ as recommended in [46]. The AIRM is used to compute the similarity between SPD matrices and to define the nearest neighbors.

The minimization problem (12) is an optimization problem on a Grassmann manifold, and as proposed in [18], it can be solved by the conjugate gradient descent method. We need to compute the gradient $\nabla_W \mathscr{L}(W)$ on the manifold

$$\nabla_W \mathscr{L}(W) = \left(I_{Nc} - WW^T\right)\sum_{i,j\neq i}^{K}\frac{\partial \mathscr{L}(W)}{\partial \delta^2(P_i,P_j)}\frac{\partial \delta^2(P_i,P_j)}{\partial W} \tag{18}$$

From [18] the Jacobian $\frac{\partial \delta^2}{\partial W}$ of the squared distance of the AIRM metric is given by

$$\begin{aligned}\frac{\partial \delta^2\left(W^TP_iW,W^TP_jW\right)}{\partial W} &= 4\left(P_iW\left(W^TP_iW\right)^{-1} - P_jW\left(W^TP_jW\right)^{-1}\right) \\ &\times log\left(W^TP_iWP_jW\left(W^TP_jW\right)^{-1}\right)\end{aligned} \tag{19}$$

This optimization problem was solved through the Manopt toolbox [47], which provides several techniques to solve the optimization over various Riemannian manifolds. The solution W finds the lower-dimensional subspace, but it does not identify the individual directions of each spatial filter. Thus, to obtain the directions of each spatial filter, in the last step of the algorithm, we solve the generalized eigenvalue decomposition problem within the detected subspace. That is, we calculate the $S'_b = W^T S_b W$ and $S'_w = W^T S_w W$ and compute the eigenvectors $V\prime = eig(S'_b, S'_w)$ sorted according to the largest eigenvalue. The final optimum spatial filters are defined as

$$W_*^{Nc\times m} = W \times V\prime \tag{20}$$

The main steps of RFCB method are described in Algorithm 1.

After computing the optimal spatial filters (W_*), the spatial filters projections are obtained from

$$Y = W_*^T X \tag{21}$$

The projections of the two most discriminative spatial filters are concatenated. Each projection corresponds to 256 time samples, which concatenated result in a vector of 512 features. Then, using the R-square correlation method the 200 most discriminative features are selected. The number of features was set experimentally during pilot experiments and was the same for all participants. The features are classified with a Naïve Bayes classifier [48], which was selected based on our previous results (see section 5.5.1 in [49] and [39]).

Algorithm 1:Riemannian Fisher Criterion Beamformer (RFCB) algorithm.

Segment EEG into epochs, perform the pre-processing and define the label of epochs
1: **Begin**
2: X_i defines segmented epoch $i = 1 \ \ldots \ K$, where K is the number of all trials
3: Compute the covariance matrices P_i of each trials using Eq. (1)
4: Compute the affinity matrix $A(P_i,P_j)$ using Eq. (17)
5: Compute the within and between class similarity ($g_w(P_i,P_j)$ and $g_b(P_i,P_j)$) using Eq. (15 and 16)
6: Compute the within-class and between-class cost functions ($\widetilde{S}_w(W)$ and $\widetilde{S}_b(W)$) using Eq. (13 and 14)
7: Solve the minimization problem $W = \arg\min_{W^TW=I_d} \mathscr{L}(W) = \frac{\widetilde{S}_w(W)}{\widetilde{S}_b(W)}$ using the conjugate gradient descent method
8: Compute the $S'_b = W^T S_b W$ and $S'_w = W^T S_w W$
9: Solve the generalized eigenvalue decomposition (GED) problem $V\prime^{m\times m} = eig(S'_b, S'_w)$
10: Compute the final spatial filters $W_* = WV\prime$
11: **End**

4. Experimental design

We evaluate the proposed approach using the BCI-Double-ErrP-Dataset called hereinafter dataset I [42] and the monitoring error-related potentials dataset [50], referred to as dataset II.

4.1. In-house dataset (dataset I)

The dataset was acquired during a P300-based speller task performed by seven able-bodied participants, and one tetraplegic participant (P1) with medullar injury (C4/C5 level). This dataset is available online in [42], for task details see [39]. The experiment comprised two sessions in three phases. Session 1 included P300 calibration (phase 1) and ErrP calibration (phase 2), and session 2 was the final online operation of the P300-ErrP speller (phase 3). These two sessions were held on different days, except for participant S9, who made the two sessions on the same day. Here, we

only use the ErrP calibration data (session 1) and the final P300-ErrP data (session 2). Session 1 has on average 304 trials and the number of errors varied across participants with a minimum of 31 and a maximum of 86. Session 2 has 176 trials with the number of errors ranging from 8 to 40. Twelve EEG channels (Fz, Cz, C3, C4, CPz, Pz, P3, P4, PO7, PO8, POz, and Oz) were recorded at a sampling rate of 256 Hz with a gUSBamp bioamplifier. The EEG data were preprocessed with the same framework used in [39]. More precisely, signals were filtered using a 50 Hz notch filter and a band-pass filter with a lower cutoff frequency of 1 Hz and a higher cutoff frequency of 10 Hz. A one-second time window after the feedback (letter detected by the BCI speller) is used to classify the event as ErrP or correct ERP.

4.2. Benchmark dataset (dataset II)

This dataset was gathered while users were observing the performance of an external agent, but without having any control over the agent [50] (dataset available from the BNCI Horizon 2020 database). The EEG signal was recorded from 6 subjects with 64 electrodes at a sampling rate of 512 Hz. The experiment contains two sessions conducted several weeks apart. Participants monitor a moving cursor to reach a target location. Each session consisted of 10 blocks of 3 min. We downsampled the data at 256 Hz and used the same twelve EEG channels of our in–house dataset, which are Fz, Cz, C3, C4, CPz, Pz, P3, P4, PO7, PO8, POz, and Oz. A one-second time window after the cursor movements is used to classify the event as ErrP or correct ERP.

5. Results

The effectiveness of the proposed method is evaluated in two datasets. For a comparative analysis of generalization, three methods are evaluated: 1) the proposed RFCB; 2) FCB used in [39], and 3) TSSF proposed in [17]. We used the TSSF method as a feature extraction method with a Fisher discriminator on the tangent space (see implementation details in appendix A). For the three methods under analysis, the spatially filtered data (the two most discrimintative projections) are then classified with a Naïve Bayes classifier. In order to ensure that the results are not biased by imbalanced classes (as the probability of "correct" class is much higher than "error" class) the performance was assessed through balanced accuracy defined as

$$bAcc = 0.5 \times \frac{TP}{TP + FN} + 0.5 \times \frac{TN}{TN + FP} \tag{22}$$

where TP, TN, FN, and FP correspond to the number of true positives, true negatives, false negatives, and false positives, respectively.

5.1. Dataset I

In this section, we present the results obtained with dataset I. We evaluate the methods in three scenarios:

- **Intra-session cross-validation.** For each session the classification is done through cross-validation. During the BCI spelling experiment, each participant had to write several sentences. Using the EEG data recorded for each sentence, we followed a leave-one-sentence-out cross validation.
- **Cross-session generalization.** Session 1 was used as training data and session 2 as test data referred to TrS1-TeS2, and session 2 was used as training data, and session 1 as test data referred to (TrS2-TeS1).
- **Cross-subject generalization.** The empirical results showed that cross-subject generalization can be affected by both variabilities between subjects and bad performance of the training subject. Thus, to avoid the effect of bad performance we used for training the data of subjects with great cross-validation classification, and good generalization between sessions. We selected two subjects with these properties, S3 and S9. Cross-session results showed that the best results are obtained when session 1 was used as the training set and session 2 as the test set, therefore, we used the data of session 1 of subject S3 or subject S9 as a training set and the data of session 2 of the other subjects as a test set.

5.1.1. Dimensionality selection of the lower SPD manifold

The RFCB algorithm finds the optimal filter $W_*^{N_c \times m}$, which will have a dimension lower than the original manifold ($N_c = 12$), that aims to preserve the discriminative information of the original SPD, called the lower SPD. To analyze the effect of the dimensionality of the lower SPD manifold (associated to parameter m), we performed the classification of session 1 using cross-validation with m varying within the interval [2: $N_c - 1$]. For intra-session cross-validation, the parameter m is selected individually for each participant. For example, from Table 1, we see that participant S2 achieved the best result with m = 2. If setting the same m value for all participants, m = 9 and m = 11 would have the best generalization (classification accuracy of about 83% as shown in Table 1). We observe that the performance is very sensitive to parameter m. For example, in participants S5, S6, and P1 the difference between the maximum accuracy and minimum accuracy is greater than 10%.

As verified, the performance of the RFCB algorithm is affected by the SPD manifold dimension. Therefore, to better understand its influence, for cross-session and cross-subject we select m following two approaches: 1) for each participant, we select the dimension m that produces the best performance, referred to as $RFCB_{Ref}$, and 2) the parameter m is set automatically as explained below, referred to as $RFCB_{Aut}$. It should be noted that $RFCB_{Ref}$ is a

Table 1
Balanced intra-session cross-validation classification accuracy using session 1 of dataset I for different values of m.

	Lower SPD Dimension (parameter m)										
Subjects	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Best m**
S1	86.8	86.5	85.8	86.2	88.5	87.1	88.5	89.1	88.5	88.5	89.1
S2	81.4	73.9	75.8	77.4	78.4	78.2	80.5	76.8	76.2	77.5	81.4
S3	93.1	90.8	91.4	92.7	92.7	93.9	93.0	94.6	94.4	94.0	94.6
S4	80.7	80.0	81.7	82.2	80.2	80.3	80.0	82.8	82.3	84.1	84.1
S5	67.0	69.3	74.6	73.6	71.8	78.1	77.9	80.4	79.9	79.9	80.4
S6	68.0	78.5	74.4	80.0	75.7	76.8	78.8	76.4	76.2	76.4	80.0
S9	92.3	94.2	91.2	87.7	93.2	92.7	93.6	95.2	94.5	97.3	97.3
P1	52.8	72.5	59.0	67.2	62.5	68.1	68.0	70.6	68.2	68.7	72.5
Mean	**77.8**	**80.7**	**79.2**	**80.9**	**80.4**	**81.9**	**82.5**	**83.2**	**82.5**	**83.3**	**84.9**

biased approach as it uses the test results to tune the parameter. It was used as a reference to evaluate the performance of $RFCB_{Aut}$.

For cross-session generalization, the dimensionality of the lower SPD manifold ($\mathcal{S}^{m}_{++}$) for $RFCB_{Aut}$ approach was obtained by cross-validation individually for each subject. For cross-subject generalization, we selected the parameter m that produced the best results using a leave-one-subject-out cross validation.

5.1.2. Intra-session performance

In this section, we present the results of intra-session validation. The classification performance of the RFCB, FCB, and TSSF, using intra-session data are presented in Table 2. The results showed that the mean classification accuracy was respectively 84.9%, 83.6%, and 72.3% for session 1, and 79.2%, 77.2%, and 67.6% for session 2. The mean classification accuracy obtained for data of session 2 was lower by 5.7%, 6.4%, and 4.7% respectively. This drop is likely due to the difference in the number of error samples. When session 1 was used as a training set the RFCB and FCB had similar performance and TSSF had lower performance. When session 2 was used as the training set the mean accuracy of RFCB was 2.0% and 11.6% higher compared with FCB, and TSSF respectively (one-tailed paired t-test, $p = 0.019$, and $p = 0.006$), i.e. RFCB was the less sensitive to the amount of training samples. These intra-session results were used as a baseline for comparison with cross-session and cross-subject scenarios.

5.1.3. Cross-session performance

The RFCB method performed better than the other two methods as shown in Table 3. The mean classification accuracy of $RFCB_{Ref}$ is very similar for both sessions with a value around 85%, meaning that the variability across sessions and the lack of error samples affected less the performance. The $RFCB_{Aut}$ approach achieved an average accuracy of 82.3%, and 78.6% for session 2 and session 1, respectively. The accuracy of session 2 is about 4% lower than session 1, however, this value is less than the decreases obtained using FCB and TSSF, namely, 8% and 5%, respectively. Comparing with baseline results the decrease was 2.6% and 0.6% for session 1 and session 2 respectively. The FCB method presented better results than TSSF with a mean classification accuracy of 80.6% for session 2 and 72.7% for session 1, with a reduction of 3.0% and 4.5% when compared to baseline. The average classification accuracies of intra-session condition (mean of session 1 and session 2) were 82.1%, 80.4%, and 69.9% for $RFCB_{Ref}$, FCB, and TSSF respectively. Comparing these results with cross-session data we observe an increase of 2.6% (one-tailed paired t-test, $p = 0.205$) for $RFCB_{Ref}$, and a decrease of about 3.8% and 4.9% (one-tailed paired t-test, $p = 0.012$, and $p = 0.018$) for FCB, and TSSF respectively. This means that the proposed $RFCB_{Ref}$ approach has better generalization across sessions. The $RFCB_{Ref}$ also presented better classification accuracy in cross-session analysis with statistically significance in session 2 (one-tailed paired t-test, $p = 0.029$, and $p = 0.032$) and session 1 (one-tailed paired t-test, $p = 0.020$, and $p = 0.019$). The $RFCB_{Aut}$ showed better classification accuracy than FCB but without statistical significance (one-tailed paired t-test, $p = 0.218$, and $p = 0.100$), and better classification accuracy than TSSF with statistical significance (one-tailed paired t-test, $p = 0.00005$, and $p = 0.004$).

5.1.4. Cross-subject performance

The results reported in Table 4 show that for $RFCB_{Ref}$ and $RFCB_{Aut}$ approaches the mean accuracies are almost the same when data from subject S3 or subject S9 are used as a training set, attaining values of 76.7%, 75.6%, 75.0%, and 74.7% respectively. For FCB, using the data of subject S9 for training, leads to an accuracy of about 5% higher than subject S3. TSSF method has poor generalization across subjects. Using the data of subject S9 or subject S3 as training set $RFCB_{Aut}$ presented statistically better results than FCB and TSSF methods (one-tailed paired t-test, $p = 0.033$, $p = 0.025$, p

Table 2
Balanced classification accuracy for intra-session cross-validation data (using dataset I).

	RFCB		FCB		TSSF	
Subjects	**Session1**	**Session2**	**Session1**	**Session2**	**Session1**	**Session2**
S1	89.1	94.8	88.5	93.8	80.2	78.3
S2	81.4	72.2	77.1	68.4	69.1	75.5
S3	94.6	95.6	93.0	95.9	81.3	76.8
S4	84.1	80.5	84.9	78.8	69.7	60.8
S5	80.4	72.9	82.1	68.2	72.6	56.8
S6	80.0	54.3	78.9	49.2	60.7	52.7
S9	97.3	98.0	97.1	98.0	62.8	83.9
P1	72.5	65.4	67.2	65.4	81.8	56.1
Mean	**84.9**	**79.2**	**83.6**	**77.2**	**72.3**	**67.6**

Table 3
Balanced classification accuracy for cross-session data using session 1 as training data and session 2 as test data (TrS1-TeS2), and using session 2 as training data and session 1 as test data (TrS2-TeS1), using dataset I.

	$RFCB_{Ref}$		$RFCB_{Aut}$		FCB		TSSF	
	TrS1-TeS2	**TrS2-TeS1**	**TrS1-TeS2**	**TrS2-TeS1**	**TrS1-TeS2**	**TrS2-TeS1**	**TrS1-TeS2**	**TrS2-TeS1**
S1	93.4	81.1	92.0	81.1	93.4	85.4	79.6	82.5
S2	93.2	83.4	85.5	77.5	89.9	57.7	69.9	62.3
S3	96.6	83.8	96.3	83.8	95.8	87.6	78.2	75.8
S4	79.3	84.4	78.5	84.2	80.1	77.3	65.0	62.9
S5	72.2	85.2	70.6	81.4	66.3	70.3	57.1	51.4
S6	71.2	86.2	71.2	82.7	56.1	58.2	45.9	48.9
S9	94.9	85.6	93.1	79.6	93.1	82.7	85.9	61.7
P1	76.8	86.6	71.1	58.1	69.7	62.6	58.8	53.9
Mean	**84.7**	**84.5**	**82.3**	**78.6**	**80.6**	**72.7**	**67.6**	**62.4**

Table 4
Balanced classification accuracy for cross-subject data using subjects S9 and S3 as training data, using dataset I

	Subjects	$RFCB_{Ref}$	$RFCB_{Aut}$	FCB	TSSF
Train S9	S1	88.4	88.4	84.7	66.0
	S2	81.6	74.6	68.3	53.3
	S3	93.2	92.1	89.8	60.2
	S4	82.4	82.4	78.8	58.3
	S5	63.6	63.6	58.1	55.2
	S6	66.9	63.4	66.3	56.9
	P1	60.5	60.5	60.1	50.0
	Mean	**76.7**	**75.0**	**72.3**	**57.1**
Train S3	S1	87.8	87.8	70.0	60.1
	S2	72.8	72.8	65.9	55.2
	S4	78.5	78.5	73.0	44.9
	S5	70.6	70.6	64.1	54.6
	S6	76.2	76.2	58.6	45.0
	S9	82.9	76.9	79.8	38.9
	P1	60.1	60.1	60.1	50.9
	Mean	**75.6**	**74.7**	**67.4**	**49.9**

= 0.001, and p = 0.0004). When compared to the user-dependent model (cross-session scenario), $RFCB_{Aut}$ showed a decrease in classification accuracy of about 6%.

5.1.5. Robustness of the spatial filter coefficients

Fig. 1 shows the spatial filters of two subjects for the two sessions, to analyze the robustness of the spatial filter coefficients across sessions and across subjects. Subjects S9 and S6 were selected as they represent the best and worse cross-session scenarios, S9 had the best cross-session generalization and S6 had the worse cross-session generalization. We compared the filter coefficients obtained using RFCB and FCB methods, estimated from the data gathered in session 1 and session 2 from these two subjects. It was expected that: 1) for cross-subject generalization, the coefficient indexes of spatial filters of both session 1 and session 2 were similar for the subject with good generalization and that they were different for the subject with poor generalization, and 2) the correlation between the spatial filters of the two sessions obtained with RFCB was greater than for FCB, since the RFCB method presented better classification. The results corroborated the first expectation, that is, for subject S9, the spatial filter coefficients obtained from the two sessions are very similar for both RFCB and FCB methods (correlation coefficient, r = 0.935, and r = 0.927, respectively). For subject S6, the correlation between the coefficient indexes decreased for both methods (correlation coefficient, r = 0.621, and r = 0.616, for RFCB, and FCB respectively). The results did not support the second expectation, once RFCB only presented a marginal correlation improvement.

5.1.6. Analysis of divergences methods

In this study, the AIRM is used to define the distance between two SPD matrices because of its invariance property. However, there are other divergence methods that have invariant properties similar to the Riemannian metric, that could have been used. We

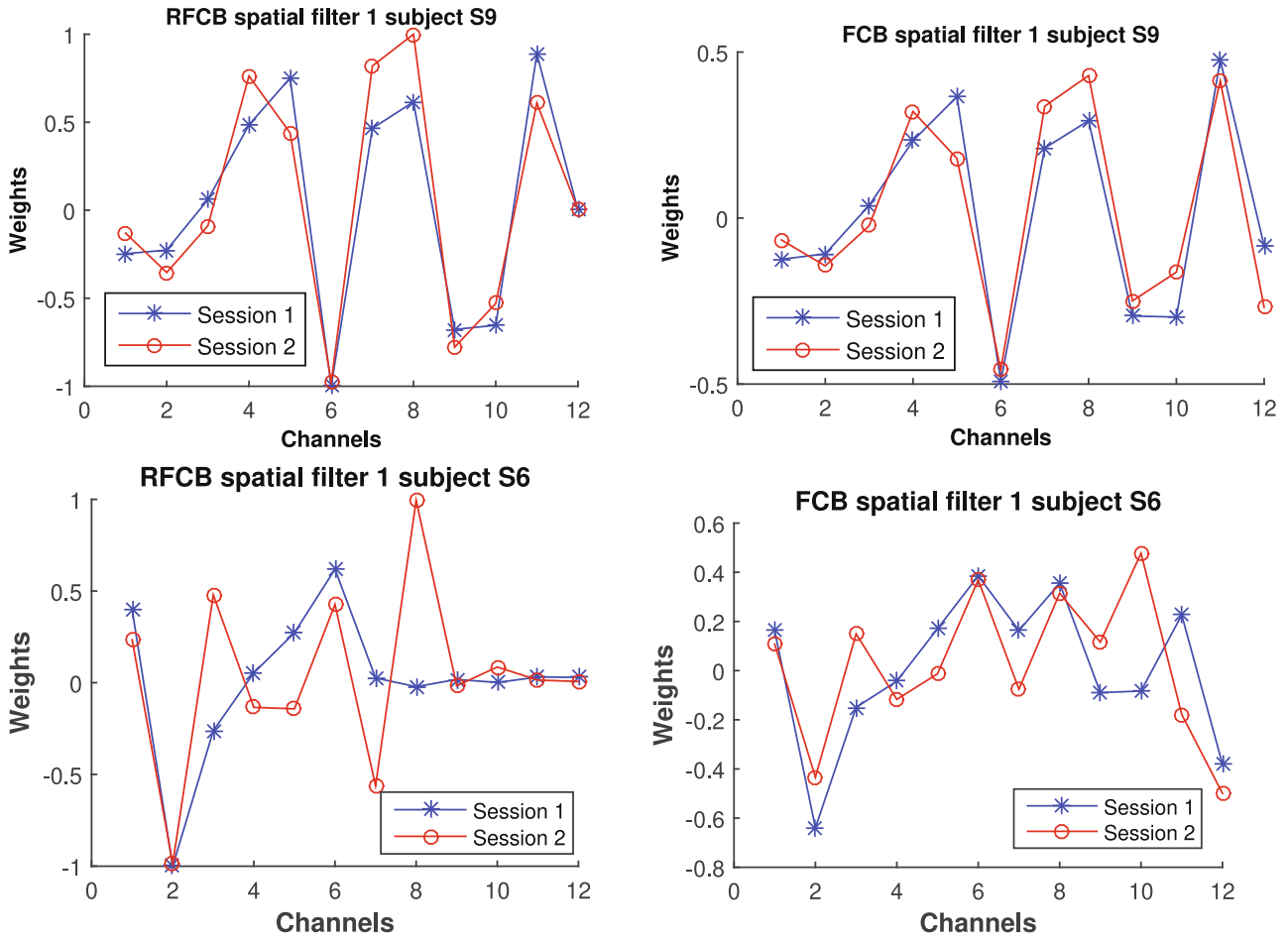

Fig. 1. The coefficients of the first spatial filters estimated using RFCB method (left), and FCB method (right) for subject S9 (with the best generalization), and subject S6 (with the worse generalization) using data from session 1 and session 2.

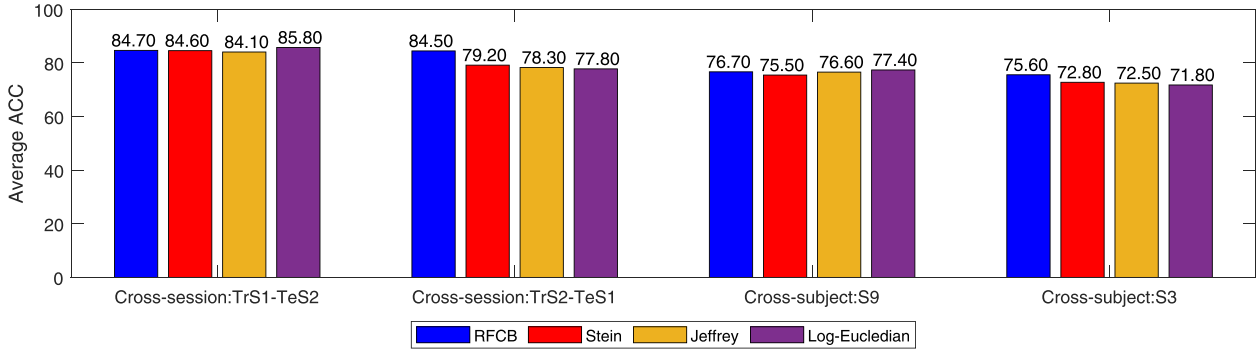

Fig. 2. Balanced classification accuracy of RFCB algorithm using Riemannian metric, Stein divergence, Jeffrey divergence, and log-euclidean distance, using dataset I.

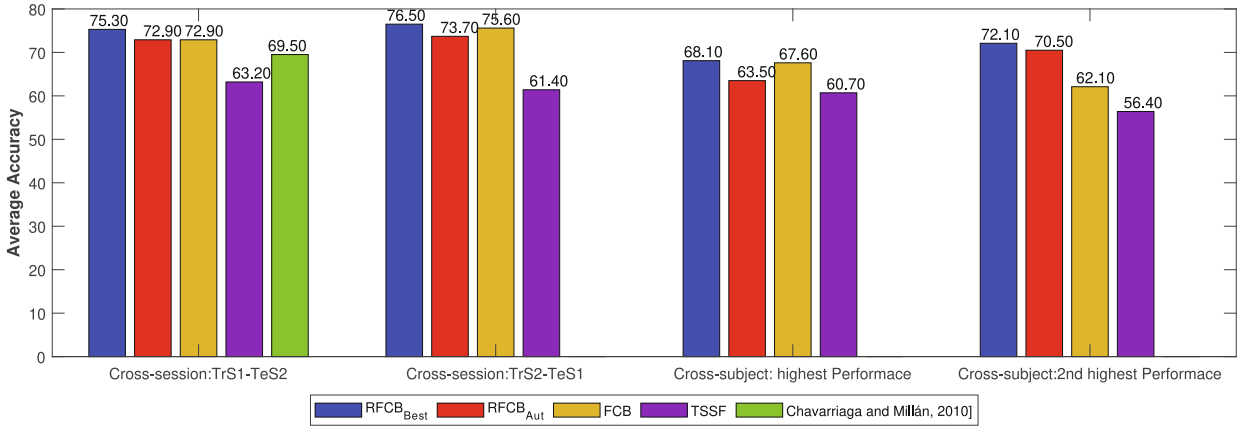

Fig. 3. Balanced classification accuracy using dataset II.

compare in Fig. 2 the use of other divergence methods in order to explore their impact on the performance of the spatial filter RFCB and to sustain our selection. The methods compared were Stein divergence [51], Jeffrey divergence [52], and log-euclidean distance [53] using dataset I.

The mean classification accuracy of the RFCB algorithm with AIRM is around 85% for both sessions. There is a decrease of 5.4%, 5.9%, and 8.0%, for Stein divergence, Jeffrey divergence, and log-Euclidean distance respectively. This means that RFCB using the Riemannian distance is more robust regarding the variability between sessions and regarding the number of error samples in the training data. In cross-subject analysis, all methods had similar performance when data from subject S9 is used for training, and AIRM had higher performance when data from subject S3 is used for training. These results showed that that RFCB with Riemannian distance is the most robust to training data variability. The classification difference between AIRM and the log-euclidean distance is statistically significant (one-tailed paired t-test, $p = 0.039$, and $p = 0.027$), and for Stein and Jeffrey divergence the differences only approach statistical significance (one-tailed paired t-test, $p = 0.083$, $p = 0.068$, $p = 0.077$, and $p = 0.083$).

5.2. Dataset II

In this section, we present the results obtained with dataset II [50]. We evaluated the cross-session and cross-subject generalization. For cross-subject analysis, we select two training subjects that have great cross-validation classification, and good generalization between sessions following the same approach explained in Section 5.1. Contrary to the dataset I in which the best subjects are equal for the three methods, here the two best subjects are different for each method. We selected subjects S1 and S2 for RFCB method, subjects S5 and S6 for FCB method, and subjects S4 and S3 for the TSSF method. Fig. 3 shows the mean accuracy of all subjects. Regarding cross-session data, the mean classification accuracy of both sessions were 75.9%, 73.3%, 74.3%, and 62.3% for $RFCB_{Ref}$, $RFCB_{Aut}$, FCB, and TSSF respectively. There was low performance variability across-session, average values of 1.3%, 0.6%, 2.7%, and 1.8% respectively. These results showed that the $RFCB_{Ref}$ method was less sensitive to the variation between sessions and it had a slightly higher classification accuracy.

For cross-subject generalization, results showed high classification variability when the data of subjects with the highest or second-highest performance (combining accuracy and generalization) was used as the training set, and RFCB approaches achieved higher classification accuracy when a subject with second-highest performance was used as the training set. In order to assess whether the defined criteria for selecting the subject to be used as a training set produced the best results, we used the data of other subjects as the training set. The best results were obtained using subjects with the highest performance, second-highest performance, and fourth performance for TSSF, RFCB, and FCB methods. Therefore, in the future other criteria to select the training set should be explored.

6. Discussion and conclusion

In this study, we proposed a new spatial filter method based on Riemannian geometry that uses the invariance properties of Riemannian distance to deal with cross-session and cross-subject

variability. The robustness of the RFCB method was assessed in cross-session and cross-subject classification scenarios using an in–house dataset and a benchmark dataset. $RFCB_{Aut}$ approach achieved a stable classification accuracy across-session for both datasets with values about 80.4% and 73.3% respectively.

6.1. Classification analysis

Regarding the in–house dataset results, when session 1 is used as a training set the cross-session average classification accuracy is similar to the one obtained from in intra-session, and when session 2 is used as a training set the cross-session accuracy is 5.4% higher than the one obtained in intra-session. These results show the robustness of the $RFCB_{Ref}$ algorithm to the changes occurring between sessions and to the variability of the number of error training samples. For the $RFCB_{Aut}$ approach (automatic selection of m) we verified similar results with a small decrease in performance. Comparing cross-subject vs intra-session, for $RFCB_{Ref}$ and $RFCB_{Aut}$ approaches we verified that the average classification accuracy is about 6% and 7% less, using either data from subjects S3 or S9 as training sets.

The RFCB method also produced better results in the benchmark dataset. The classification accuracy is higher (5.8%) than the one achieved in the original work [50], which strengthens the effectiveness of the proposed method. In order to analyze the effect of parameter m in classification, we compare the results obtained with $RFCB_{Ref}$ and $RFCB_{Aut}$ approaches. $RFCB_{Ref}$ had higher classification accuracy than $RFCB_{Aut}$ in both datasets. For cross-session data, the performance was statistically better except when using session 2 as training set and session 1 as test set (one-tailed paired test, $p = 0.058$) using the in–house dataset. For cross-subject data, the difference is not statistically significant. There is great variability in the classification performance using different values of the parameter m, which highlights its importance. The best values of m, i.e., the ones providing the best classification accuracies for each participant and methods are shown in Table 5. These results and the difference obtained between $RFCB_{Ref}$ and $RFCB_{Aut}$ emphasize the need to further explore others ways to automatically define the parameter m.

6.2. Comparison to closely related works

Table 6 presents the works closely related to our study, i.e., those that use transfer learning either for cross-session or cross-subject generalization, in the context of ErrP classification. Table 6 includes for each work the number of tested subjects, feature extraction methods, classification algorithms, evaluation metric and the best classification accuracies achieved. For feature extrac-

Table 5
The best value of the selected parameter m for each participant (using dataset I).

	Intra-session		**Cross-session**			
	Session 1	**Session 2**	**TrS1-TeS2**		**TrS2-TeS1**	
Subjects	**RFCB**	**RFCB**	**RFCB_Ref**	**RFCB_Aut**	**RFCB_Ref**	**RFCB_Aut**
S1	9	2	10	9	2	2
S2	2	10	10	2	2	10
S3	9	3	11	9	4	3
S4	11	10	7	11	10	10
S5	9	10	11	9	4	10
S6	3	8	5	5	2	8
S9	5	9	8	11	6	9
P1	11	9	5	3	6	9
Mean	7.4	7.6	8.4	7.4	4.5	7.6

Table 6
Summary of related works that use transfer learning either for cross-session generalization or cross-subject generalization of ErrP classifier. For each work, only the best accuracy reported is presented here.

Author Reference	Number of subjects	Feature extraction method	Classifier	Evaluation Metric	Intra-session	Cross-session	Cross-subject
[50]	6	CAR and downsampling	Gaussian	Accuracy	-	69.5 (offline)	
[7]	12	Arithmetic mean of ten partially overlapping time windows	Linear discriminant analysis	Balanced Accuracy	88.3 (offline)	-	72.7 (offline)
[57]	26	Savitzky-Golay filter	Ensemble of linear discriminant analysis	Accuracy	81.4 (offline)	-	74.0 (pseudo-online)
[58]	26	Temporal, spectral, and spatial domains features	Feedforward neural network	Accuracy	-	-	76.7 (offline)
[54]	23	CAR, downsampling, and filtered by FFT	SVM with radial basis function (RBF) kernel	Accuracy	-	81.0 (online)	
[55]	6	CAR	Deep Convolutional neural network	Accuracy	-	80.1 (offline)	79.9 (offline)
[59]	10	CAR, baseline correction, and downsampling	Shrinkage-LDA classifier	Balanced Accuracy	81.0 (offline)		71.0 (offline)
[60]	15	Temporal features	Random forest classifiers	Accuracy			89.0 (offline)
[56]	8	xDAWN	Support vector machine	Balanced Accuracy	82.0 (offline)		75.0 (offline)
[39]	10	FCB	Naïve Bayes	Balanced Accuracy	84.5 (offline)	90.0 (online)	
Our study	**8 and 6**	**RFCB**	**Naïve Bayes**	**Balanced Accuracy**	**82.1 (offline)**	**80.4 (offline)**	**74.8 (offline)**

tion, techniques can be separated into two groups: 1) EEG time samples from all or a selected number of channels, after eventual downsampling and/or spatial filtering using common average reference (CAR) [50,54,55], and 2) optimal statisitical spatial filtering [56,39]. Classification is based on different approaches, mainly on classical machine learning techniques, such as Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM), and only one work applies a Deep Convolutional Neural Network (CNN). The evaluation metric is not always the balanced accuracy, which may bias the performance and makes difficult the direct comparison. Our study is the only one evaluating the methods in three scenarios: intra-session, cross-validation cross-session generalization, and cross-subject generalization. Excluding our previous work (which used a data augmentation approach and a double-ErrP detection) we achieved the highest classification accuracy for cross-session generalization, with 80.4%. For cross-subject analysis, our approach also compared favorably with the studies using the balanced accuracy metric, with 74.8%. Of the studies that evaluated their approaches with unbalanced accuracy, the performance ranged between 74% and 89%, however these values can have been biased by imbalanced classes. Nevertheless, it should be noted that a direct comparison between studies is not straightforward because there is great variability in classification performance between different BCI paradigms/applications (e.g., "interaction ErrPs" vs. "observation ErrPs", "screen feedback" vs. "robot action"). To better understand this performance variability, we are already working on a systematic comparison of methods proposed in the literature that have been applied to ErrPs elicited by different paradigms.

6.3. Conclusion

A new method called RFCB was proposed, lying on the Riemannian SPD manifold, seeking to improve the cross-session and cross-subject classification in the context of error-related potentials. The results are promising outperforming two other methods, FCB and TFSS, used for comparison, and comparing very favorably with the state of the art. Cross-subject classification is still relatively low, which motivates further research.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Tangent space spatial filter (TSSF)

The TSSF algorithm [17] obtains the filter on the tangent space and projects it to the manifold. We first compute the covariance matrices (P_i) using (Eq. 1), then the Riemannian mean (P_Ω) is calculated as

$$P_\Omega = \mathcal{G}(P_1 \dots P_N) = \arg min_{P \in \mathcal{P}(n)} \sum_i^N \delta_R^2(P, P_i) \tag{23}$$

The Riemannian mean is used as the reference point for the tangent space:

$$Log_{P_\Omega}(P_i) = S_i = P_\Omega^{\frac{-1}{2}} log\left(P_\Omega^{\frac{1}{2}} P_i P_\Omega^{\frac{-1}{2}}\right) P_\Omega^{\frac{1}{2}} \tag{24}$$

The tangent vector (S) is computed, and the filters can be obtained using different criterion. Here we used the Fisher Linear Discriminant criterion. Therefore, we compute the within-class and between-class scatter matrix and then we solve the generalized eigenvalue decomposition (GED) problem. The filter is the eigenvector sorted according to the largest eigenvalue and it is mapped back onto the manifold (P_w). After we solve the generalized eigenvalue decomposition problem: GED(P_w, P_Ω). The optimum spatial filter (V) is the eigenvector sorted according to the largest absolute value of the logarithm of the eigenvalues.

References

[1] L.F. Nicolas-Alonso, J. Gomez-Gil, Brain computer interfaces, a review, Sensors 12 (2) (2012) 1211–1279..

[2] C.S. Nam, A. Nijholt, F. Lotte, Brain–computer interfaces handbook: technological and theoretical advances, CRC Press, 2018.

[3] P.-J. Kindermans, M. Schreuder, B. Schrauwen, K.-R. Müller, M. Tangermann, True zero-training brain-computer interfacing–an online study, PloS one 9 (7) (2014) e102504.

[4] D.-K. Han, J.-H. Jeong, Domain generalization for session-independent brain-computer interface, arXiv preprint arXiv:2012.03533 (2020)..

[5] M. Congedo, A. Barachant, R. Bhatia, Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review, Brain-Computer Interfaces 4 (3) (2017) 155–174.

[6] M. Krauledat, M. Tangermann, B. Blankertz, K.-R. Müller, Towards zero training for brain-computer interfacing, PloS one 3 (8) (2008) e2967.

[7] F.M. Schönleitner, L. Otter, S.K. Ehrlich, G. Cheng, A comparative study on adaptive subject-independent classification models for zero-calibration error-potential decoding, in: IEEE International Conference on Cyborg and Bionic Systems 2019, 2019.

[8] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowledge Data Eng. 22 (10) (2009) 1345–1359.

[9] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, M. Grosse-Wentrup, Transfer learning in brain-computer interfaces, IEEE Comput. Intell. Magazine 11 (1) (2016) 20–31.

[10] H. Lee, A. Cichocki, S. Choi, Nonnegative matrix factorization for motor imagery eeg classification, in: International Conference on Artificial Neural Networks, Springer, 2006, pp. 250–259.

[11] M. Arvaneh, I. Robertson, T.E. Ward, Subject-to-subject adaptation to reduce calibration time in motor imagery-based brain-computer interface, in: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2014, pp. 6501–6504..

[12] H. Morioka, A. Kanemura, J.-I. Hirayama, M. Shikauchi, T. Ogawa, S. Ikeda, M. Kawanabe, S. Ishii, Learning a common dictionary for subject-transfer decoding with resting calibration, NeuroImage 111 (2015) 167–178.

[13] H. Raza, H. Cecotti, Y. Li, G. Prasad, Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface, Soft. Comput. 20 (8) (2016) 3085–3096.

[14] M. Alamgir, M. Grosse-Wentrup, Y. Altun, Multitask learning for brain-computer interfaces, in, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2010, pp. 17–24.

[15] H. Kang, S. Choi, Bayesian common spatial patterns for multi-subject eeg classification, Neural Networks 57 (2014) 39–50.

[16] P. Zanini, M. Congedo, C. Jutten, S. Said, Y. Berthoumieu, Transfer learning: a riemannian geometry framework with applications to brain–computer interfaces, IEEE Trans. Biomedical Eng. 65 (5) (2017) 1107–1116.

[17] J. Xu, M. Grosse-Wentrup, V. Jayaram, Tangent space spatial filters for interpretable and efficient riemannian classification, J. Neural Eng. 17 (2) (2020) 026043.

[18] M. Harandi, M. Salzmann, R. Hartley, Dimensionality reduction on spd manifolds: The emergence of geometry-aware methods, IEEE Trans. Pattern Anal. Mach. Intell. 40 (1) (2017) 48–62.

[19] M. Yamamoto, K. Sadatnejad, T. Tanaka, M.R. Islam, Y. Tanaka, F. Lotte, Detecting eeg outliers for bci on the riemannian manifold using spectral clustering, in: 42nd Annual International Conferences of the IEEE Engineering in Medicine and Biology Society (EMBC'2020), 2020.

[20] F. Lotte, Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces, Proc. IEEE 103 (6) (2015) 871–890.

[21] S. Kumar, F. Yger, F. Lotte, Towards adaptive classification using riemannian geometry approaches in brain-computer interfaces, in: 2019 7th International Winter Conference on Brain-Computer Interface (BCI), IEEE, 2019, pp. 1–6.

[22] F. Yger, M. Berar, F. Lotte, Riemannian approaches in brain-computer interfaces: a review, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (10) (2016) 1753–1762.

[23] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, F. Yger, A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update, J. Neural Eng. 15 (3) (2018) 031005.

[24] A. Cherian, S. Sra, A. Banerjee, N. Papanikolopoulos, Efficient similarity search for covariance matrices via the jensen-bregman logdet divergence, in: 2011 International Conference on Computer Vision, IEEE, 2011, pp. 2399–2406.

[25] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Common spatial pattern revisited by riemannian geometry, in: 2010 IEEE International Workshop on Multimedia Signal Processing, IEEE, 2010, pp. 472–476..

[26] S. Kumar, K. Mamun, A. Sharma, Csp-tsm: Optimizing the performance of riemannian tangent space mapping using common spatial pattern for mi-bci, Computers Biology Med. 91 (2017) 231–242.

[27] B. Sun, J. Feng, K. Saenko, Return of frustratingly easy domain adaptation, in: Thirtieth AAAI Conference on Artificial Intelligence, 2016..
[28] H. He, D. Wu, Transfer learning for brain-computer interfaces: A euclidean space data alignment approach, IEEE Trans. Biomed. Eng. (2019).
[29] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Multiclass brain–computer interface classification by riemannian geometry, IEEE Trans. Biomed. Eng. 59 (4) (2011) 920–928.
[30] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Classification of covariance matrices using a riemannian-based kernel for bci applications, Neurocomputing 112 (2013) 172–178.
[31] C.H. Nguyen, P. Artemiadis, Eeg feature descriptors and discriminant analysis under riemannian manifold perspective, Neurocomputing 275 (2018) 1871–1883.
[32] L. Korczowski, M. Congedo, C. Jutten, Single-trial classification of multi-user p300-based brain-computer interface using riemannian geometry, in: 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC), IEEE, 2015, pp. 1769–1772..
[33] F. Li, Y. Xia, F. Wang, D. Zhang, X. Li, F. He, Transfer learning algorithm of p300-eeg signal based on xdawn spatial filter and riemannian geometry classifier, Appl. Sci. 10 (5) (2020) 1804.
[34] B. Rivet, A. Souloumiac, V. Attina, G. Gibert, xdawn algorithm to enhance evoked potentials: application to brain–computer interface, IEEE Trans. Biomed. Eng. 56 (8) (2009) 2035–2043.
[35] G. Pires, U. Nunes, M. Castelo-Branco, Statistical spatial filtering for a P300-based BCI: tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis, J. Neurosci. Methods 195 (2) (2011) 270–281.
[36] R. Chavarriaga, A. Sobolewski, J. d. R. Millán, Errare machinale est: The use of error-related potentials in brain-machine interfaces, Frontiers in Neuroscience 8 (8 JUL) (2014) 1–13. doi:10.3389/fnins.2014.00208..
[37] M. Falkenstein, J. Hohnsbein, J. Hoormann, L. Blanke, Effects of crossmodal divided attention on late erp components. ii. error processing in choice reaction tasks, Electroencephalography Clinical Neurophysiology 78 (6) (1991) 447–455.
[38] J. Teeuw, Comparison of error-related eeg potentials, in: 13th Twente Student Conference on IT, 2010.
[39] A. Cruz, G. Pires, U.J. Nunes, Double errp detection for automatic error correction in an erp-based bci speller, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (1) (2018) 26–36.
[40] F. Iwane, R. Chavarriaga, I. Iturrate, J. d. R. Millán, Spatial filters yield stable features for error-related potentials across conditions, in: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2016, pp. 000661–000666..
[41] A. Cruz, G. Pires, U.J. Nunes, Generalization of errp-calibration for different error-rates in p300-based bcis, in: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2018, pp. 644–649..
[42] A. Cruz, G. Pires, U.J. Nunes, BCI DOUBLE ERRP DATASET, url:https://ieee-dataport.org/open-access/error-related-potentials-primary-and-secondary-errp-and-p300-event-related-potentials-, [Online; accessed July-2020] (2020)..
[43] P.-A. Absil, R. Mahony, R. Sepulchre, Optimization algorithms on matrix manifolds, Princeton University Press, 2009.
[44] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Riemannian geometry applied to bci classification, in: International Conference on Latent Variable Analysis and Signal Separation, Springer, 2010, pp. 629–636.
[45] M. Moakher, A differential geometric approach to the geometric mean of symmetric positive-definite matrices, SIAM J. Matrix Anal. Appl. 26 (3) (2005) 735–747.
[46] M. Sugiyama, Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis, J. Mach. Learning Res. 8 (May) (2007) 1027–1061.
[47] N. Boumal, B. Mishra, P.-A. Absil, R. Sepulchre, Manopt, a matlab toolbox for optimization on manifolds, J. Mach. Learning Res. 15 (1) (2014) 1455–1459.
[48] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, Machine learning 29 (2) (1997) 131–163.
[49] G.P. Pires, Biosignal classification for human interface with devices and surrounding environment, Ph.D. thesis (2011)..
[50] R. Chavarriaga, J. d. R. Millán, Learning from EEG error-related potentials in noninvasive brain-computer interfaces, IEEE Transactions on Neural Systems and Rehabilitation Engineering 18 (4) (2010) 381–388..
[51] S. Sra, Positive definite matrices and the symmetric stein divergence., arXiv preprint arXiv:1110.1773 (2011)..
[52] A. Cherian, S. Sra, A. Banerjee, N. Papanikolopoulos, Jensen-bregman logdet divergence with application to efficient similarity search for covariance matrices, IEEE Trans. Pattern Anal. Mach. Intell. 35 (9) (2012) 2161–2174.
[53] V. Arsigny, P. Fillard, X. Pennec, N. Ayache, Log-euclidean metrics for fast and simple calculus on diffusion tensors, Magnetic Resonance in Medicine: An Official Journal of the International Society for, Magn. Reson. Med. 56 (2) (2006) 411–421.
[54] M. Spüler, M. Bensch, S. Kleih, W. Rosenstiel, M. Bogdan, A. Kübler, Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a p300-bci, Clin. Neurophysiol. 123 (7) (2012) 1328–1337.
[55] S.A.S. Bellary, J.M. Conrad, Classification of error related potentials using convolutional neural networks, in: 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, 2019, pp. 245–249.
[56] S.K. Kim, E.A. Kirchner, Handling few training data: classifier transfer between different types of error-related potentials, IEEE Trans. Neural Syst. Rehabil. Eng. 24 (3) (2015) 320–332.
[57] S. Bhattacharyya, A. Konar, D.N. Tibarewala, M. Hayashibe, A generic transferable eeg decoder for online detection of error potential in target selection, Front. Neurosci. 11 (2017) 226.
[58] J. Tong, Q. Lin, R. Xiao, L. Ding, Combining multiple features for error detection and its application in brain–computer interface, Biomed. Eng. 15 (1) (2016) 1–15.
[59] M. Abu-Alqumsan, C. Kapeller, C. Hintermüller, C. Guger, A. Peer, Invariance and variability in interaction error-related potentials and their consequences for classification, J. Neural Eng. 14 (6) (2017) 066015.
[60] N. Usama, K.K. Leerskov, I.K. Niazi, K. Dremstrup, M. Jochumsen, Classification of error-related potentials from single-trial eeg in association with executed and imagined movements: A feature and classifier investigation, Medical Biolog. Eng. Computing 58 (11) (2020) 2699–2710.

Aniana Cruz obtained her M.Sc and Ph.D. degrees in Biomedical Engineering from the University of Coimbra, Portugal, in 2011 and 2021. She has been involved in different research projects in human-machine interfaces and the sleep stage. Her current research interests include EEG-based brain-computer interfaces (BCIs), biomedical signal processing, human-machine interaction, error detection, self-paced control, and mental state monitoring.

Gabriel Pires (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Coimbra, Portugal, in 2012. He is currently an Adjunct Professor with the Engineering Department, Polytechnic Institute of Tomar. He is also an Integrated Researcher with the Institute of Systems and Robotics, University of Coimbra. He has been responsible and a member of several funded projects in the areas of human-computer interfaces and assistive mobile robotics. His main research interests include EEG-based brain-computer interfaces (BCIs) and biosignal processing, seeking to improve the reliability and usability of BCIs. He is also the Coordinator of the research Laboratory VITA.IPT-Life Assisted by Intelligent Environments.

Urbano J. Nunes (Senior Member, IEEE) received the Ph. D. degree in electrical engineering from the University of Coimbra, Portugal, in 1995. He is currently a Full Professor with the Department of Electrical and Computer Engineering, University of Coimbra (UC), and a Senior Researcher with the Institute for Systems and Robotics. He has been involved with/responsible for several funded projects at both national and international levels in the areas of human-centered mobile robotics and intelligent vehicles. He has been active in the organization of conferences, such as the General Chair of the 2010 IEEE Intelligent Transportation Systems Conference, the General Chair of the 2012 IEEE/RSJ Intelligent Robots and Systems, and the General Chair of the 2017 IEEE International Symposium on Robot and Human Interactive Communication. He is also a Senior Editor of the IEEE TRANSACTIONS ON INTELLIGENT VEHICLES.