

# On the interpretation of linear Riemannian tangent space model parameters in M/EEG

Reinmar J. Kobler<sup>1,2</sup>, Jun-Ichiro Hirayama<sup>1</sup>, Lea Hehenberger<sup>2,3</sup>, Catarina Lopes-Dias<sup>2,3</sup>,  
 Gernot R. Müller-Putz<sup>2,3</sup> and Motoaki Kawanabe<sup>1</sup>

**Abstract**—Riemannian tangent space methods offer state-of-the-art performance in magnetoencephalography (MEG) and electroencephalography (EEG) based applications such as brain-computer interfaces and biomarker development. One limitation, particularly relevant for biomarker development, is limited model interpretability compared to established component-based methods. Here, we propose a method to transform the parameters of linear tangent space models into interpretable patterns. Using typical assumptions, we show that this approach identifies the true patterns of latent sources, encoding a target signal. In simulations and two real MEG and EEG datasets, we demonstrate the validity of the proposed approach and investigate its behavior when the model assumptions are violated. Our results confirm that Riemannian tangent space methods are robust to differences in the source patterns across observations. We found that this robustness property also transfers to the associated patterns.

## I. INTRODUCTION

Magnetoencephalography (MEG) and electroencephalography (EEG) capture a linear mixture of brain and noise signals [1]. In a supervised setting, where the goal is to infer a target signal from the power of latent oscillatory sources, component-based methods like common spatial patterns (CSP) [2,3] for discrete targets or source power co-modulation (SPOC) [4] for continuous targets are widely used [5]. Recently, they have been outperformed by Riemannian tangent space methods in several datasets [6–9].

Key factors for the success of Riemannian tangent space methods are that the features, namely covariance matrices, lie on a Riemannian manifold and the commonly used geometric metric is invariant to affine transformations [6,10]. The tangent space is a vector space with a Euclidean metric that locally approximates the Riemannian manifold around a reference matrix, typically the geometric mean of a dataset. Consequently, standard linear machine learning techniques for Euclidean vector spaces can be used in the tangent space.

One limitation of the Riemannian approaches is their lack of interpretability in terms of contributing brain sources [9,11]. For component-based methods, there are established techniques to interpret the model parameters in terms of spatial patterns [12]. In this work, we show that the parameters of linear regression and classification methods in the Riemannian tangent space can be transformed to interpretable spatial patterns in the M/EEG channel space in a similar fashion as for component-based methods. Thereby, the tradeoff between performance and interpretability can be overcome.

<sup>1</sup>RIKEN AIP, Kyoto, Japan. <sup>2</sup>Graz BCI Racing Team, <sup>3</sup>Institute of Neural Engineering, Graz University of Technology, Graz, Austria. E-mail: {reinmar.kobler, motoaki.kawanabe}@riken.jp

In the next section, we introduce the underlying generative model for a regression problem, briefly outline a recently proposed tangent space regression algorithm [8], followed by the proposed method to convert the parameters to interpretable patterns. The section ends with a description of conducted simulations, analyzed datasets, and baseline methods.

## II. MATERIALS AND METHODS

### A. Generative model

The M/EEG signals  $\mathbf{x}_i(t)$  are typically modelled as a linear mixture of sources plus additive noise [1]

$$\mathbf{x}_i(t) = \mathbf{A}_s \mathbf{s}_i(t) + \mathbf{n}_i(t) \quad (1)$$

where  $\mathbf{s}_i(t) \in \mathbb{R}^Q$  denotes the source signal time activity of observation  $i$  (epoch, session, subject, etc.) and  $\mathbf{n}_i(t) \in \mathbb{R}^P$  the additive noise. The matrix  $\mathbf{A}_s \in \mathbb{R}^{P \times Q}$  contains the  $Q$  source patterns [12].

As in [8], we assume that the brain signals arise from activity of uncorrelated sources. The noise  $\mathbf{n}_i(t) = \mathbf{A}_n \boldsymbol{\nu}_i(t)$  is stationary, uncorrelated with the sources, and spans a subspace ( $\mathbf{A}_n \in \mathbb{R}^{P \times P-Q}$ ) that is shared across observations. The generative model can then be written as:

$$\mathbf{x}_i(t) = \mathbf{A}_s \mathbf{s}_i(t) + \mathbf{A}_n \boldsymbol{\nu}_i(t) = \mathbf{A} \boldsymbol{\eta}_i(t) \quad (2)$$

$\mathbf{A} \in \mathbb{R}^{P \times P}$  includes the source and noise patterns and is assumed to be invertible. The vector  $\boldsymbol{\eta}_i(t) \in \mathbb{R}^P$  is the concatenation of the latent source and noise signals. A scalar target signal  $y_i$  at observation  $i$  is then modeled as a function of the latent sources' powers:

$$y_i = \mathbf{b}^T f(\mathbf{p}_i) + b_0 + \varepsilon_i \quad (3)$$

where  $\mathbf{p}_i = \mathbb{E}\{\mathbf{s}_i^2(t)\}_t \in \mathbb{R}^Q$  contains the sources' powers at observation  $i$ ,  $f(\cdot)$  is a known function,  $\mathbf{b} \in \mathbb{R}^Q$  a weight vector,  $b_0$  a bias term, and  $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$  noise. Here, we consider  $f(\cdot) = \log(\cdot)$  as log-linear relationships are often encountered in oscillatory brain activity [13].

Given a set of paired observations  $\{(\mathbf{x}_i(t), y_i)_{i=1, \dots, N}\}$  our goal is to predict the target signal for unseen data, and identify the patterns  $\mathbf{A}_s$  of the encoding sources.

As in previous works [4,8], we will use the between-sensor covariance matrices as features. Assuming zero-mean signals, the covariance matrices can be computed as:

$$\mathbf{C}_i = \frac{1}{T} \mathbf{X}_i \mathbf{X}_i^T \in \mathbb{R}^{P \times P} \quad (4)$$

where the columns of the matrix  $\mathbf{X}_i \in \mathbb{R}^{P \times T}$  contain  $T$  temporal samples. The covariance matrices  $\mathbf{C}_i$  are in the manifold of positive definite matrices  $\mathcal{S}_P^{++}$ . If we assume that the source signals are zero-mean and uncorrelated, their covariance matrix is diagonal  $\mathbb{E}\{\mathbf{s}_i(t)\mathbf{s}_i^T(t)\}_t = \text{diag}(\mathbf{p}_i)$ . If they are also uncorrelated with the noise, i.e.  $\mathbb{E}\{\mathbf{s}_i(t)\boldsymbol{\nu}_i^T(t)\}_t = 0$ , and w.l.o.g. the noise sources are uncorrelated, the sensor covariance matrices can be expressed as:

$$\mathbf{C}_i = \mathbf{A}\mathbf{E}_i\mathbf{A}^T \quad (5)$$

where  $\mathbf{E}_i = \mathbb{E}\{\boldsymbol{\eta}_i(t)\boldsymbol{\eta}_i^T(t)\}_t$  is a diagonal matrix, whose diagonal elements are  $\mathbf{p}_i$ .

### B. Riemannian tangent space regression model

Equipping the manifold  $\mathcal{S}_P^{++}$  with the geometric metric gives a Riemannian manifold structure to  $\mathcal{S}_P^{++}$ . For the generative model, defined in (1), the Riemannian tangent space embedding of the covariance matrices  $\mathbf{C}_i$  gives a consistent estimator for  $y_i$ , if the function  $f(\cdot)$  is the logarithm [8]. The embedding  $\mathbf{v}_i$  is computed as:

$$\mathbf{v}_i = \text{proj}_{\bar{\mathbf{C}}}(\mathbf{C}_i) = \text{upper} \left( \log \left( \bar{\mathbf{C}}^{-1/2} \mathbf{C}_i \bar{\mathbf{C}}^{-1/2} \right) \right) \quad (6)$$

where  $\bar{\mathbf{C}}$  is the geometric mean [6] of the covariance matrices  $\{\mathbf{C}_1, \dots, \mathbf{C}_N\}$ , the logarithm for  $\mathbf{C} \in \mathcal{S}_P^{++}$  is  $\log(\mathbf{C}) = \mathbf{U}\text{diag}(\log(\lambda_1), \dots, \log(\lambda_P))\mathbf{U}^T$  with  $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ , and the invertible mapping  $\text{upper}(\mathbf{C}) \in \mathbb{R}^{P(P+1)/2}$  extracts the upper triangular elements of a symmetric matrix, with the off-diagonal elements weighted by the factor  $\sqrt{2}$ . The weighting ensures that  $\|\text{upper}(\mathbf{C})\|_2 = \|\mathbf{C}\|_F$ . The Euclidean distance  $\|\mathbf{v}_k - \mathbf{v}_l\|_2$  in the tangent space approximates the geometric distance between  $\mathbf{C}_k$  and  $\mathbf{C}_l$  in the vicinity of  $\bar{\mathbf{C}}$  [6].

Since the relation between  $\mathbf{v}_i$  and  $y_i$  is linear [8], any linear method can be used to fit  $\mathbf{b}_c$  so that a cost function between  $\hat{y}_i = \mathbf{b}_c^T \mathbf{v}_i + b_0$  and  $y_i$  is minimized. For M/EEG datasets the observation ( $N$ ) to feature ( $P(P+1)/2$ ) ratio is typically small, requiring regularization.

### C. M/EEG channel space model patterns

Given a fitted model with parameters  $(\mathbf{b}_c, \bar{\mathbf{C}})$ , the patterns  $\mathbf{A}_s$  of the encoding sources can be identified in a three step procedure (algorithm 1). First, the tangent space pattern  $\mathbf{d}_c$  associated to  $\mathbf{b}_c$  is computed according to [12], using  $\mathbf{C}_v = \mathbb{E}\{\mathbf{v}_i\mathbf{v}_i^T\}_i$ . Next,  $\mathbf{d}_c$  is projected back to the covariance matrix space to obtain  $\mathbf{C}_d \in \mathcal{S}_P^{++}$ . Finally, the general eigenvalue problem for  $\mathbf{C}_d$  and  $\bar{\mathbf{C}}$  is solved.

Under the generative model, the resulting eigenvectors correspond to the patterns  $\mathbf{A}$  and the eigenvalues  $(\lambda_j)_{j=1, \dots, P}$  are a function of the unknown, weight vector  $\mathbf{b}$  (see proof in the appendix). Specifically, the eigenvalues are:

$$\lambda_j(\mathbf{b}) = \begin{cases} \exp(b_j/\|\mathbf{b}\|^2) & j \leq Q \\ 1 & \text{otherwise} \end{cases} \quad (7)$$

where the first  $Q$  eigenvalues correspond to the encoding sources  $\mathbf{s}(t)$ . In practice, the sources with the strongest coupling, i.e., largest  $|b_j|$ , are found via sorting the eigenvalues according to the criterion  $\max(\lambda_j, 1/\lambda_j)$ . The number of

sources with significant coupling ( $Q$ ) can be identified via a shuffling procedure.

---

#### Algorithm 1: Channel space model patterns

---

**Input** : tangent space weight vector  $\mathbf{b}_c$ ,  
tangent space projection matrix  $\bar{\mathbf{C}}$ ,  
tangent space feature covariance matrix  $\mathbf{C}_v$

**Output**:  $\mathbf{A} \in \mathbb{R}^{P \times P}$ ,  $\boldsymbol{\lambda} \in \mathbb{R}^P$

---

- 1  $\mathbf{d}_c = \mathbf{C}_v \mathbf{b}_c \sigma_{\hat{y}}^{-2}$  with  $\sigma_{\hat{y}}^2 = \mathbf{b}_c^T \mathbf{C}_v \mathbf{b}_c$
  - 2  $\mathbf{C}_d = \text{proj}_{\bar{\mathbf{C}}}^{-1}(\mathbf{d}_c) = \bar{\mathbf{C}}^{1/2} \exp(\text{upper}^{-1}(\mathbf{d}_c)) \bar{\mathbf{C}}^{1/2}$
  - 3  $\mathbf{A}, \boldsymbol{\lambda} = \text{eigh}(\mathbf{C}_d, \bar{\mathbf{C}})$
- 

### D. Model fitting and evaluation

In real M/EEG datasets, the covariance matrices can be rank-deficient ( $\mathbf{C}_i \notin \mathcal{S}_P^{++}$ ). In this case, the geometric metric is not defined [6]. As a remedy, we reduced the dimensions from  $P$  to  $K$  via projecting the covariance matrices to the subspace spanned by the first  $K$  principal components of the covariance matrices' arithmetic mean [11]. In this subspace, we computed the geometric mean and the tangent space features, according to (6). Next, the features were z-scored. Depending on the dataset, the linear weight vector  $\mathbf{b}_c$  was estimated either via ridge regression or penalized logistic regression. The associated cost functions were the mean absolute error (MAE) between  $y_i$  and  $\hat{y}_i$  or the balanced classification accuracy. The train/test-splitting scheme depended on the dataset. All parameters (spatial filters, geometric mean, z-scoring, linear weights) were fitted to training data. The optimal regularization parameter of the regression/classification model was determined using an inner generalized cross-validation (CV) scheme. We considered 25 candidate values which were log-spaced in the range  $[10^{-5}, 10^3]$ . In the next sections, we refer to this sequence of operations as the RIEMANN pipeline.

To put the results into context, we compared the performance to SPOC/CSP and a naïve method, denoted DIAG here. For SPOC, we used the SPOC lambda algorithm [4] to estimate  $k$  components. After spatial filtering with SPOC or CSP, the logarithm of the covariance matrices' diagonal elements were the features. The subsequent steps (z-scoring, regression/classification) were identical. The DIAG pipeline was identical to the SPOC/CSP pipeline except for the initial spatial filtering step. Consequently, the method computed log-band power features in channel space. We computed patterns for the RIEMANN and SPOC/CSP pipelines according to section II-C and [12].

### E. Experiments

We conducted three experiments to demonstrate the validity of the approach and analyzed the behavior when the model assumptions are violated.

1) *Simulations*: In two regression problem simulations we investigated the algorithms' properties in identifying a single encoding source. First, we varied the power  $\sigma^2$  of the additive Gaussian noise  $\varepsilon_i$  in (3). Second, we introduced a model violation by making the patterns dependent on the

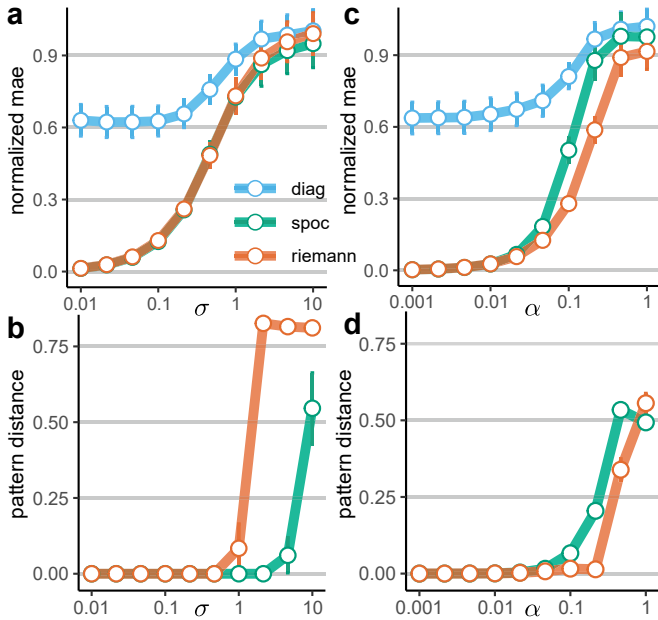


Fig. 1. **Simulation results.** Effect of the target noise power ( $\sigma^2$ ) on the mean absolute error (MAE) (a) and the pattern distance (b). The methods are color-coded. Error bars show the standard error of the mean. a, The MAE was normalized by the MAE of a dummy method that used the expected value of the target signal  $\mathbb{E}\{y_i\}_i$  as predictor. c and d, as in a and b for different pattern noise power ( $\alpha^2$ ) levels.

observation index  $\mathbf{A}_i = \mathbf{A} + \mathbf{N}_i$  with  $(\mathbf{N}_i)_{jk} \sim \mathcal{N}(0, \alpha^2)$ . The patterns were computed as  $\mathbf{A} = \exp(\mathbf{B})$  with  $(\mathbf{B})_{jk} \sim \mathcal{N}(0, 1)$ . The other parameters were identical to [8]. Because the covariance matrices had full rank, we omitted the PCA step for the RIEMANN pipeline.

Ten-fold CV was used to fit and evaluate the models. In addition to the MAE cost function between  $y_i$  and  $\hat{y}_i$ , we computed distances between the true  $\mathbf{a}_s$  and its estimate  $\hat{\mathbf{a}}_s$ . The pattern distance was defined as  $1 - |\hat{\mathbf{a}}_s^T \mathbf{a}_s| / (|\hat{\mathbf{a}}_s| \cdot |\mathbf{a}_s|)$ . A distance of 0 means that the topographies are identical up to a scalar factor  $\gamma \in \mathbb{R}$ .

2) *Cortico-muscular coherence dataset:* We analyzed an MEG dataset that studied cortico-muscular coherence (CMC) [14]. The publicly available dataset contains recordings of a single participant during one session. In a trial-based task, the participant contracted her left hand and exerted a constant force against a lever. The trials were interleaved with short breaks. As in [8], we analyzed the dataset in a continuous setting with the goal to decode the EMG envelope from the MEG beta band activity of 151 gradiometers. The considered data and preprocessing steps were similar to [8]. In a nutshell, we set the single bipolar EMG channel power as target signal  $y_i$ , and extracted oracle approximating shrinkage (OAS) regularized [15] covariance matrices  $\mathbf{C}_i$  for beta band ([15, 30] Hz) activity in overlapping windows ( $T = 1.5$  s, overlap = 1.25 s). We applied 10-fold CV to evaluate the goodness of fit and used the coefficient of determination  $R^2$  as metric.

3) *Multi-session BCI dataset:* This dataset was recorded during a longitudinal (26 sessions, 15 months) BCI study with a tetraplegic user [16]. The analyzed data contains

EEG signals (32 channels), recorded during a trial-based paradigm. In each trial, the user performed 1 of 4 distinct mental tasks and received discrete feedback, provided by an adaptive BCI. Here, we analyzed the two tasks with the strongest patterns (feet motor imagery and mental subtraction). The data preprocessing and trial rejection methods were identical to [16]. The preprocessed and cleaned data comprised activity in 4 frequency bands during a 2-s epoch per trial (1438 trials). For each epoch and band, one OAS regularized covariance matrix  $\mathbf{C}_i$  was computed. The tangent space projection was computed independently for each frequency band. Thereafter, the individual feature vectors were concatenated and used to predict the target class. The models were evaluated using a leave-one-session-out CV scheme.

#### F. Software

The software and analysis scripts are publicly available [https://github.com/rkobler/interpret\\_lin\\_rts\\_mdls](https://github.com/rkobler/interpret_lin_rts_mdls), and are based on the code of [8] and the python packages Scikit-Learn [17], MNE [18] and PyRiemann [19].

### III. RESULTS AND DISCUSSION

The simulation results are summarized in Fig. 1. As expected, the regression scores in Fig. 1a,c are similar to the results reported in [8]. They confirm that the DIAG method is not a consistent estimator for the generative model considered here, and that the RIEMANN method is more robust to pattern noise than SPOC.

The higher robustness to pattern noise of the RIEMANN method generally translated to lower pattern distances compared to SPOC (Fig. 1d). Regarding the target noise (Fig. 1b), SPOC was more robust to higher noise levels, as the distance of the RIEMANN method increased abruptly for  $\sigma \geq 1$ . Note that for  $\sigma = 1$  the noise term in (3) started to dominate the data term, resulting in poor out-of-sample predictions for both methods (Fig. 1a).

Fig. 2 summarizes the CMC dataset results for the SPOC and RIEMANN methods. Both methods achieved similar quantitative (Fig. 2a) and qualitative (Fig. 2b) decoding accuracies. The  $R^2$  score peaked at approx. 0.5 (SPOC: 4 components, RIEMANN: 42). SPOC reached the peak accuracy at a lower number of components because its components are fitted in a supervised fashion.

As the patterns in (Fig. 2c-f) indicate, both methods relied on similar sources. Considered that the sign is ambiguous, pendants of the 4 SPOC patterns (Fig. 2d) can be readily found among the first 8 RIEMANN patterns (Fig. 2f). The first pattern of both methods indicates that they primarily decoded the target from eye artifacts. Knowing that the paradigm had a trial-based structure and there was a strong target signal change in the breaks (Fig. 2b), eye artifacts were likely a confounding source. This result underlines the importance of interpretable models in M/EEG experiments.

The multi-session binary classification dataset results are depicted in Fig. 3. Using 25 sessions to fit the parameters, both methods achieved high accuracies in the test session.

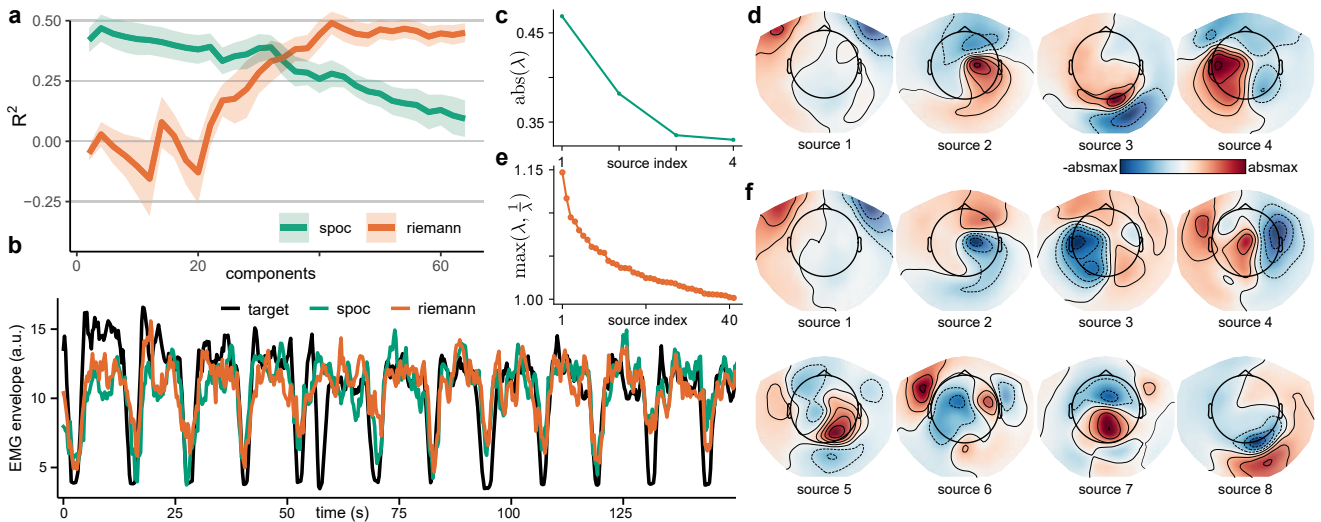


Fig. 2. **Decoding EMG envelope from beta band MEG power.** **a**, Dependence of the CV test set regression score (explained variance,  $R^2$ ) on the number of components. The number of components was increased from 2 to 64 in steps of 2. The methods are color-coded. Shaded areas show the standard error of the mean. **b**, Trajectory of the single-channel EMG envelope (black). The test set trajectories of the SPOC and RIEMANN methods are color-coded. **c**, Eigenvalues ( $\lambda$ ) of the SPOC method with 4 components. The components are sorted in descending order according to the absolute value of the eigenvalue. **d**, Patterns of the 4 SPOC sources. **e**, Eigenvalues ( $\lambda$ ) of the RIEMANN method with 42 components. The sources are ordered in descending order according to  $\max(\lambda, 1/\lambda)$ . **f**, Patterns of the first 8 RIEMANN sources.

The peak accuracies for the RIEMANN and CSP methods were 0.93 (18 components per frequency band) and 0.91 (10). The paired difference across sessions was significant (Fig. 3b).

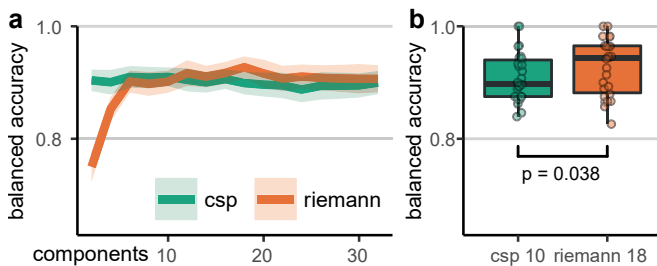


Fig. 3. **Classifying two mental tasks across sessions.** **a**, Dependence of the classification score (balanced accuracy) on the number of components for a leave one session out CV scheme. The methods are color-coded. Shaded areas show the 95% confidence interval of the mean. **b**, Boxplots summarizing the results for the models at the peak in **a**. The differences between the methods were significant (two-sided, paired t-test,  $df = 25$ ,  $\alpha = 0.05$ ).

Fig. 4 shows the associated patterns. The sources with highest eigenvalues (source 1 in the alpha bands) were similar, indicating that both methods agreed in the most discriminative source. The patterns also match with the class-specific, grand-average power modulations, reported in [16]. We observed two differences. First, the RIEMANN patterns were spatially smoother and easier to attribute to single dipolar sources. Generally, a higher fraction of dipolar patterns indicates a better source de-mixing quality [20]. Second, evaluating the eigenvalues in the lower and higher beta band, there was a drastic drop between the first and second eigenvalue for the RIEMANN method. This suggests that the first source contained considerably more discriminative information than the second one. CSP lacked such a drop,

suggesting that it did not identify this beta band source. In this longitudinal dataset, the assumption of stationary patterns is certainly violated as the manually mounted electrode cap location varied across sessions. Because the electrode locations varied across sessions, both differences could be attributed to the fact that the RIEMANN method is more robust to pattern noise (Fig. 1c,d).

#### IV. CONCLUSION

We proposed a method to interpret the model parameters of linear regression and classification methods, operating in Riemannian tangent space. In simulations, we found that the estimated patterns were robust to noise in the patterns across observations. These findings were confirmed in a multi-session EEG dataset. The Riemannian tangent space method not only significantly improved the classification accuracy upon CSP but also extracted sources whose patterns were smoother and more dipolar, which are typical characteristics of sources originating in the brain. In summary, the proposed approach to compute patterns enables an intuitive interpretation of state-of-the-art linear Riemannian tangent space models.

#### REFERENCES

- [1] P. L. Nunez and R. Srinivasan, *Electric Fields of the Brain*. Oxford University Press, 2006, ISBN: 978-0-19-505038-7.
- [2] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 4, pp. 441–446, 2000.
- [3] B. Blankertz *et al.*, "Optimizing Spatial Filters for Robust EEG Single-Trial Analysis," *IEEE Signal Proc. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.
- [4] S. Dähne *et al.*, "SPoC: A novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters," *NeuroImage*, vol. 86, pp. 111–122, 2014.
- [5] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *J. Neural Eng.*, vol. 15, no. 3, 2018.

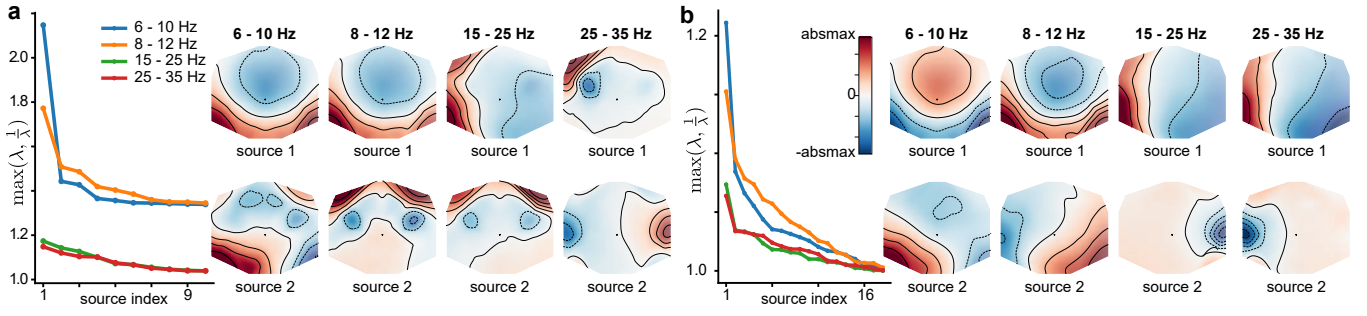


Fig. 4. **Classifying two mental tasks across sessions. Patterns.** **a**, Eigenvalues (left) and patterns (right) of the CSP method with 10 components per frequency band. The 4 frequency bands in the alpha and beta range are color-coded. For each frequency band, the top 2 patterns are plotted. **b**, As in **a** for the RIEMANN method with 18 components.

- [6] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review," *Brain-Computer Interfaces*, vol. 4, no. 3, pp. 155–174, 2017.
- [7] L. A. Gemein *et al.*, "Machine-learning-based diagnostics of EEG pathology," *NeuroImage*, vol. 220, 2020.
- [8] D. Sabbagh *et al.*, "Predictive regression modeling with MEG/EEG: from source power to signals and cognitive states," *NeuroImage*, vol. 222, 2020.
- [9] J. Xu, M. Grosse-Wentrup, and V. Jayaram, "Tangent space spatial filters for interpretable and efficient riemannian classification," *J. Neural Eng.*, vol. 17, no. 2, 2020.
- [10] F. Yger, M. Berar, and F. Lotte, "Riemannian Approaches in Brain-Computer Interfaces: A Review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1753–1762, 2017.
- [11] D. Sabbagh *et al.*, "Manifold-regression to predict from MEG/EEG brain signals without source modeling," in *NeurIPS*, 2019.
- [12] S. Haufe *et al.*, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *NeuroImage*, vol. 87, pp. 96–110, 2014.
- [13] G. Buzsáki and K. Mizuseki, "The log-dynamic brain: how skewed distributions affect network operations," *Nat. Rev. Neurosci.*, vol. 15, no. 4, pp. 264–278, 2014.
- [14] J.-M. Schoffelen *et al.*, "Selective Movement Preparation Is Subverted by Selective Increases in Corticomuscular Gamma-Band Coherence," *J. Neurosci.*, vol. 31, no. 18, pp. 6750–6758, 2011.
- [15] Y. Chen *et al.*, "Shrinkage Algorithms for MMSE Covariance Estimation," *IEEE Trans. Signal Proc.*, vol. 58, no. 10, pp. 5016–5029, 2010.
- [16] L. Hehenberger *et al.*, "Long-term mutual training for the CY-BATHLON BCI Race with a tetraplegic pilot: a case study on inter-session transfer and intra-session adaptation," *Front. Hum. Neurosci.*, 2021.
- [17] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *JMLR*, vol. 12, pp. 2825–2830, 2011.
- [18] A. Gramfort *et al.*, "MNE software for processing MEG and EEG data," *NeuroImage*, vol. 86, pp. 446–460, 2014.
- [19] A. Barachant *et al.*, "Multiclass Brain-Computer Interface Classification by Riemannian Geometry," *IEEE. Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 920–928, 2012.
- [20] A. Delorme *et al.*, "Independent EEG Sources Are Dipolar," *PLoS ONE*, vol. 7, no. 2, 2012.

## APPENDIX

*Proof* that the encoding source patterns  $\mathbf{A}_s$  and unknown regression coefficients  $\mathbf{b}$  can be recovered from the tangent space weight vector  $\mathbf{b}_c$ .

We start the proof with expressing the regression model in (3) in terms of the tangent space features at the geometric mean source covariance matrix  $\bar{\mathbf{E}}$ . Since all  $\mathbf{E}_i$  matrices are diagonal, we have that their geometric mean  $\bar{\mathbf{E}} = \text{diag}(\bar{\mathbf{p}}) = \text{diag}(\left(\prod_{i=1}^N p_{ij}\right)^{\frac{1}{N}})$ . Projecting  $\mathbf{E}_i$  to the tangent space at  $\bar{\mathbf{E}}$  yields:

$$\begin{aligned} \tilde{\mathbf{v}}_i &= \text{proj}_{\bar{\mathbf{E}}}(\mathbf{E}_i) = \text{upper}(\log(\bar{\mathbf{E}}^{-\frac{1}{2}} \mathbf{E}_i \bar{\mathbf{E}}^{-\frac{1}{2}})) \\ &= \text{upper}(\text{diag}((\log(p_{ij}) - \log(\bar{p}_j))_{j=1, \dots, P})) \end{aligned} \quad (\text{A.1})$$

Starting from (3), assuming w.l.o.g. that  $y$  is zero-mean, and setting  $f(p_{ij}) = \log(p_{ij})$  and  $b_0 = -\sum_{j=1}^Q b_j \log(\bar{p}_j)$ , it follows that

$$y_i = \mathbf{b}^T f(\mathbf{p}_i) + b_0 + \varepsilon_i = \underbrace{[\mathbf{b}^T, 0]}_{\tilde{\mathbf{b}}^T} \tilde{\mathbf{v}}_i + \varepsilon_i \quad (\text{A.2})$$

Next, we relate the observed tangent space features  $\tilde{\mathbf{v}}_i$  at  $\bar{\mathbf{C}}$  with  $\tilde{\mathbf{v}}_i$  at  $\bar{\mathbf{E}}$ . Due to the invariance property of the geometric mean we have  $\bar{\mathbf{C}} = \mathbf{A} \bar{\mathbf{E}} \mathbf{A}^T$ . We additionally introduce the matrix  $\mathbf{U} = \bar{\mathbf{C}}^{\frac{1}{2}} \mathbf{A}^{-T} \bar{\mathbf{E}}^{-\frac{1}{2}}$  so that

$$\mathbf{U}^T \bar{\mathbf{C}}^{-\frac{1}{2}} \mathbf{C}_i \bar{\mathbf{C}}^{-\frac{1}{2}} \mathbf{U} = \bar{\mathbf{E}}^{-\frac{1}{2}} \mathbf{E}_i \bar{\mathbf{E}}^{-\frac{1}{2}} \quad (\text{A.3})$$

holds for all observations  $i$ . It is straightforward to show that  $\mathbf{U}$  is orthogonal. Consequently, we have:

$$\begin{aligned} \mathbf{U}^T \log(\bar{\mathbf{C}}^{-\frac{1}{2}} \mathbf{C}_i \bar{\mathbf{C}}^{-\frac{1}{2}}) \mathbf{U} &= \log(\bar{\mathbf{E}}^{-\frac{1}{2}} \mathbf{E}_i \bar{\mathbf{E}}^{-\frac{1}{2}}) \\ \mathbf{U}^T \text{upper}^{-1}(\mathbf{v}_i) \mathbf{U} &= \text{upper}^{-1}(\tilde{\mathbf{v}}_i) \end{aligned} \quad (\text{A.4})$$

Now, we can rewrite the dot product between  $\tilde{\mathbf{b}}$  and  $\tilde{\mathbf{v}}$  in (A.2) as:

$$\begin{aligned} \tilde{\mathbf{b}}^T \tilde{\mathbf{v}}_i &= \text{tr}(\text{upper}^{-1}(\tilde{\mathbf{b}}) \text{upper}^{-1}(\tilde{\mathbf{v}}_i)) \\ &= \text{tr}(\text{upper}^{-1}(\tilde{\mathbf{b}}) \mathbf{U}^T \text{upper}^{-1}(\mathbf{v}_i) \mathbf{U}) \\ &= \text{upper}(\mathbf{U} \text{upper}^{-1}(\tilde{\mathbf{b}}) \mathbf{U}^T)^T \mathbf{v}_i = \mathbf{b}_c^T \mathbf{v}_i \end{aligned} \quad (\text{A.5})$$

where we defined  $\mathbf{b}_c$  as  $\text{upper}(\mathbf{U} \text{upper}^{-1}(\tilde{\mathbf{b}}) \mathbf{U}^T)$  and  $\text{tr}(\cdot)$  computes the trace of a matrix. The weights  $\mathbf{b}_c$  linearly relate the tangent space features  $\mathbf{v}_i$  at  $\bar{\mathbf{C}}$  with the target signal  $y_i$ . Consequently, with  $N \rightarrow \infty$  the estimates of a linear estimation method converge to the true weights  $\mathbf{b}_c$ .

Since the in-product between the pattern and the weight vector in both tangent spaces is 1 [12], it follows that  $\mathbf{d}_c = \text{upper}(\mathbf{U} \text{upper}^{-1}(\tilde{\mathbf{d}}) \mathbf{U}^T)$  where  $\tilde{\mathbf{d}} = \tilde{\mathbf{b}} / \|\tilde{\mathbf{b}}\|^2$  is the pattern associated to  $\tilde{\mathbf{b}}$ . If we project  $\mathbf{d}_c$  from the tangent space to the covariance matrix space, we get:

$$\begin{aligned} \mathbf{C}_d &= \text{proj}_{\bar{\mathbf{C}}}^{-1}(\mathbf{d}_c) = \bar{\mathbf{C}}^{\frac{1}{2}} \exp(\text{upper}^{-1}(\mathbf{d}_c)) \bar{\mathbf{C}}^{\frac{1}{2}} \\ &= \bar{\mathbf{C}}^{\frac{1}{2}} \mathbf{U} \exp(\text{upper}^{-1}(\tilde{\mathbf{d}})) \mathbf{U}^T \bar{\mathbf{C}}^{\frac{1}{2}} \\ &= \bar{\mathbf{C}} \mathbf{A}^{-T} \bar{\mathbf{E}}^{-\frac{1}{2}} \exp(\text{upper}^{-1}(\tilde{\mathbf{d}})) \bar{\mathbf{E}}^{-\frac{1}{2}} \mathbf{A}^{-1} \bar{\mathbf{C}} \\ &= \mathbf{A} \bar{\mathbf{E}}^{\frac{1}{2}} \exp(\text{upper}^{-1}(\tilde{\mathbf{d}})) \bar{\mathbf{E}}^{\frac{1}{2}} \mathbf{A}^T \\ &= \mathbf{A} \text{diag} \left( \left[ \left( e^{d_j} \bar{p}_j \right)_{j=1, \dots, Q}, \left( \bar{p}_j \right)_{j=Q+1, \dots, P} \right] \right) \mathbf{A}^T \end{aligned} \quad (\text{A.6})$$

where we used in the last line the fact that  $\text{upper}^{-1}(\tilde{\mathbf{d}})$  is a diagonal matrix. Computing  $\mathbf{C}_d \bar{\mathbf{C}}^{-1}$ , we get:

$$\mathbf{A} \text{diag} \left( \left[ \left( e^{\frac{d_j}{\|\tilde{\mathbf{b}}\|^2}} \right)_{j=1, \dots, Q}, \left( 1 \right)_{j=Q+1, \dots, P} \right] \right) \mathbf{A}^{-1} \quad (\text{A.7})$$

Hence, via eigen decomposition of  $\mathbf{C}_d \bar{\mathbf{C}}^{-1}$  we can recover the unknown mixing matrix  $\mathbf{A}$  and latent weights  $\mathbf{b}$ . This concludes the proof.