Formulating Divergence Framework for Multiclass Motor Imagery EEG Brain Computer Interface

Satyam Kumar [1]
Tharun Kumar Reddy [1]
Vipul Arora [1]
Laxmidhar Behera [1]

[1]Indian Institute Of Technology Kanpur

ICASSP 2020
Barcelona, Spain

Overview

1. Brain Computer Interface
2. Motor Imagery BCI
   - Spatial Filtering
   - Dataset
3. Stationarity in BCI

Outline

1. Brain Computer Interface
2. Motor Imagery BCI
   - Spatial Filtering
   - Dataset
3. Stationarity in BCI

Motivation: A realtime BCI

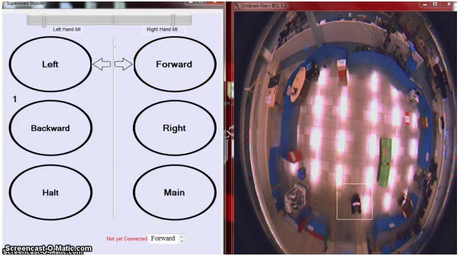


Figure: Source: https://youtu.be/FiQPgRe1NR8

- BCI based on Binary Classification of Motor Imagery
- Subject is controlling the mobile robot using his brain signals corresponding to imagined movements.

Brain-computer interface (BCI)

*"Brain computer interfaces (BCIs) are systems enabling communication between brain and a device that enables signals from brain to direct some external act (**Courtesy**: Jacques Vidal ,1973[1])"*

- Prof. Vidal outlined seminal theoretical and technical suggestions for direct brain-computer communication.
- Prof. Vidal outlined all the elements necessary to build a working BCI.

Figure: Source:shorturl.at/gjEKW

[1] Vidal, J.J. 1973. Toward direct brain-computer communication. Annual Review of Biophysics and Bioengineering 2:157–180

BCI (continued...)

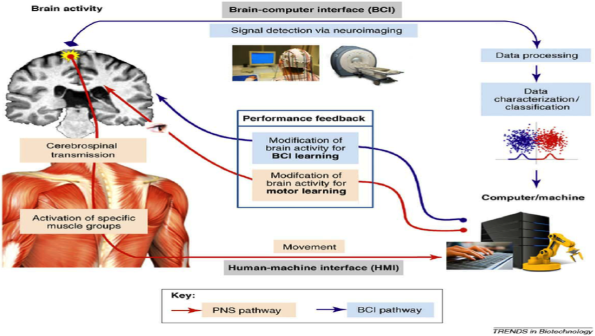


Figure: Source:[2]

- Limit to EEG (non-invasive, cheap and good temporal resolution!!

[2]Min, B.K., Marzelli, M.J. and Yoo, S.S., 2010. Neuroimaging-based approaches in the brain–computer interface. Trends in biotechnology, 28(11), pp.552-560.

- P300 Event Related Potentials [3]
- SSVEP (Evoked potentials)(shorturl.at/koY49)
- **Sensorimotor rhythms (SMR)**
  - SMR refers to signals recorded over SensoriMotor Cortex (SMC) in response to motor activities.
  - SMR signal can be modulated through Motor Imagination tasks.
  - SMRs are generally measured in the alpha (8-12 Hz) and beta (13-30 Hz) bands.
  - Low SNR and higher number of EEG channels, classification using rule based thresholding with increase/decrease in alpha, beta power (ERD/ERS) is difficult !!
  - Solution: Multivariate signal analysis along with advanced machine learning techniques

[3] Kübler, A., 2017, January. Quo vadis P300 BCI?. In 2017 5th International Winter Conference on Brain-Computer Interface (BCI) (pp. 36-39). IEEE.

Motor Imagery (MI)

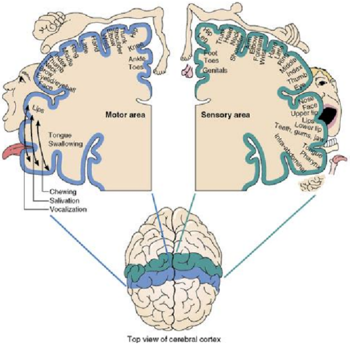


Figure: Somatotopic mapping (Source: shorturl.at/eouzH)

- Imagining the movement of the right or left hands/legs.
- Distinct body parts (leg, tongue, and hand are mapped distant from each other) compared to similar body parts (hand and fingers, which are close to each other in the mapping.)

Outline

EEG sources Model

Signal $\mathbf{x}(t)$ recorded at the scalp is modeled as a (noisy) linear mixture.

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

$\mathbf{s}(t) \in \mathbb{R}^D$ are neural sources and $\mathbf{A} \in \mathbb{R}^{D \times D}$

Motor imagery based BCIs aim to focus on the sources $\mathbf{s}(t)$ of sensorimotor rhythm modulation

An estimate of these neural sources is obtained by applying spatial filters $\mathbf{W} = [\mathbf{w}_1, .., \mathbf{w}_d] \in \mathbb{R}^{D \times d}$ to the data.

$$\mathbf{s}(t) = \mathbf{W}^\intercal \mathbf{x}(t) \tag{2}$$

This zero correlation assumption can be expressed mathematically by restricting $\mathbf{W}$ to be decomposable into whitening matrix and an orthogonal projection matrix.

- Solution:Spatial filtering to find un-correlated sources

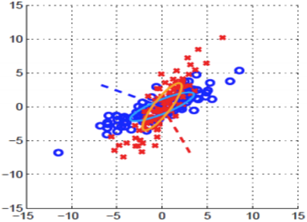
Outline

1. Brain Computer Interface
2. Motor Imagery BCI
   - Spatial Filtering
   - Dataset
3. Stationarity in BCI

Common Spatial Pattern (Blankertz et al. 2008[4])

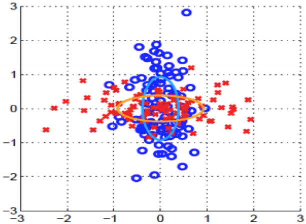
“ CSP method calculates the spatial filters which maximizes variance of band pass Multivariate signals across one class while minimizing the variance across other class ”

$$J(w) = \frac{W^T C_1 W}{W^T C_2 W}$$



(a) before CSP filtering

$S=W^TX$



(b) after CSP filtering

[4] Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M. and Muller, K.R., 2007. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal processing magazine, 25(1), pp.41-56.

CSP theory

$\mathbf{X}_{n,c} \in \mathbb{R}^{C \times T}$ $n \in \{1, 2, .., N_c\}$ $c \in \{1, 2\}$ (n^{th} EEG trial, n^{th} class), C channels and T time samples per trial and N_c trials in each class 'c'.

$$\mathbf{\Sigma}_c = \frac{1}{N_c} \sum_{n=1}^{N_c} \mathbf{X}_{n,c} \mathbf{X}_{n,c}^T \qquad c \in \{1, 2\} \tag{3}$$

The CSP algorithm aims to learn a linear transform W such that the ratio of variances of transformed data across the classes is extremized.

$$\underset{\mathbf{W}}{\operatorname{argmax}} J(\mathbf{W}) = \frac{\mathbf{W}^T \mathbf{\Sigma}_1 \mathbf{W}}{\mathbf{W}^T \mathbf{\Sigma}_2 \mathbf{W}} \quad s.t \quad \mathbf{W}^T \mathbf{\Sigma}_2 \mathbf{W} = \mathbf{I} \tag{4}$$

CSP: Theory (continued...)

Above optimization problem in (4) can be solved using the following generalized eigenvalue problem:

$$\boldsymbol{\Sigma}_1 \mathbf{W} = \lambda \boldsymbol{\Sigma}_2 \mathbf{W} \tag{5}$$

The eigenvectors obtained from (5) constitute the linear transform that maximizes (4). These eigenvectors are generally known as spatial filters. Thus each EEG trial $\mathbf{X}_i$ is mapped to $\mathbf{Z}_i$ such that

$$\mathbf{Z}_i = \mathbf{W}^T \mathbf{X}_i \tag{6}$$

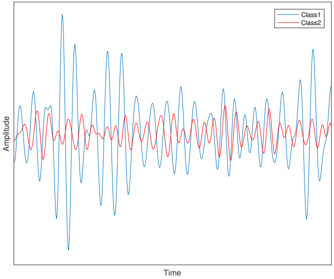
CSP: Theory (Continued...) Feature Extraction

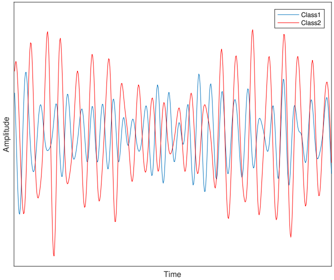
For feature extraction, a small number of spatial filters are selected, typically the first m and last m columns of the transformation matrix $\mathbf{W}$ (The spatial filters are sorted in decreasing order corresponding to their eigenvalues). Thus for a given trial $\mathbf{X}_i$ the mapped matrix $\mathbf{Z}_i$ has $2m$ rows. A feature vector $\mathbf{f}_i$ corresponding to trial i can thus be formed as

$$\mathbf{f}_i^n = \log\left(\frac{var(\mathbf{Z}_i^n)}{\sum_{n=1}^{2m} var(\mathbf{Z}_i^n)}\right) \tag{7}$$

where $\mathbf{Z}_i^n$ is the nth row of the projected matrix and n varies from 1 to $2m$. These are also called as 'Log Variance Features'.

CSP in Action

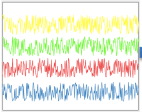




- Variance after spatial filtering is high for class2 and low for class1 enhancing discriminativity.
- Limitations: CSP looks at average variances, ignoring nonstationarities(within-class variability and within session variability).

CSP Framework: Complete Pipeline

- CSP parameters
  - Time window extraction
  - Frequency Band
  - Number of spatial filters



Band-Pass Filtering → CSP Feature Extraction → Classifier SVM/LDA → Output

- Epoch Extraction
- Alpha, Beta, Customized
- Spatial Filters
- Log Variance feature
- Binary Class

- Generally six spatial filters are selected using prespecified criteria

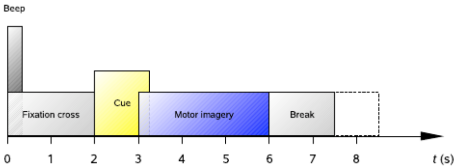
Outline

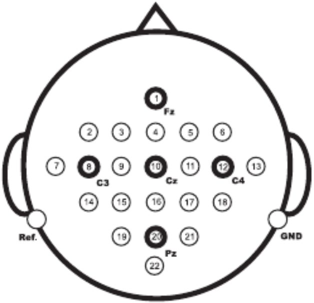
1. Brain Computer Interface
2. Motor Imagery BCI
   - Spatial Filtering
   - Dataset
3. Stationarity in BCI

Offline analysis: Database

	BCI Competition IV Dataset IIa
Task	Left hand, Right hand, Foot and Tongue
Subjects	9
Channels	22
Training set	288 (72 for each class)
Test set	288 (72 for each class)

Offline analysis: Experimental Protocol (Source:BCI Competition IV Dataset IIa)





Observations!!

- Classification of Motor Imagery (Multi-class)
- For motor imagery BCI, mainly features are extracted through spatial characteristics (enhanced SNR through spatial filtering).
- Non-stationarities in EEG signal prone to generate errors in spatial filters.[5]

[5] Samek, W., 2016. On robust spatial filtering of EEG in nonstationary environments. it-Information Technology, 58(3), pp.150-154.

Outline

Stationarity Optimization

Divergence based approach to optimize stationarity

Non-stationarity ([Von Bünau et al. 2009[7]])

- EEG signals are highly non-stationary (Joint probability distribution changes with time).
- Difficult to classify using ML algorithms
  - Feature distribution continuously changes
    - Changes occurring within each experimental sessions (due to eye blink, head movement etc).
    - Between different subjects.
    - Across experimental sessions.

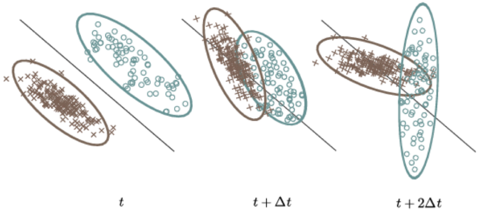


Figure: Source: [6]

  - Different training and test distributions
- Solution: Extract stationary features

[6] Samek, W., 2016. On robust spatial filtering of EEG in non-stationary environments. it-Information Technology, 58(3), pp.150-154.

[7] Von Bünau, P., Meinecke, F.C., Király, F.C. and Müller, K.R., 2009. Finding stationary subspaces in multivariate time

Stationary feature extraction [Von Bünau et al. 2009[9]]

- Stationary subspace analysis (SSA)
  - Preserve stationarity in small chunks of data
  - Extracts stationary sources.

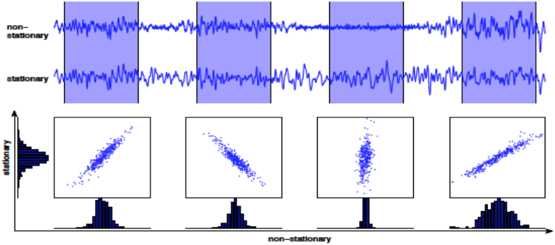


Figure: Source:[8]

[8] Bünau, P.V., 2012. Stationary Subspace Analysis: Towards understanding non-stationary data.

[9] Von Bünau, P., Meinecke, F.C., Király, F.C. and Müller, K.R., 2009. Finding stationary subspaces in multivariate time series. Physical review letters, 103(21), p.214101.

Quantifying Stationarity

- Given D distinct sources. A set of 'd' latent sources are stationary, if joint distribution of latent sources stays the same in time.

$$\underbrace{Y}_{\text{latent sources}} = V^T \underbrace{X}_{\text{raw signal}} \qquad V \in \mathbb{R}^{D\times d}$$

- Time series is stationary, if its first two moments (mean and covariance) are constant over time.
- Natural measure of stationarity is based on the distance of the mean and covariance matrices.
- Pairwise Kullback-Leibler divergence between distributions of the latent sources (projected data) and average of projected data.
  - Kullback-Leibler (KL) divergence measures the distance between two distribution

$$L(V) = \sum_{i=1}^{N} D_{kl}[\mathcal{N}(\mu_i, \Sigma_i) \parallel \mathcal{N}(\bar{\mu}, \bar{\Sigma})]$$

Quantifying Stationarity

- The KL divergence between two D variate gaussians can be computed as:

$$D_{kl}(\mathcal{N}(\mu_{\mathbf{0}}, \mathbf{\Sigma_0}) \parallel \mathcal{N}(\mu_{\mathbf{1}}, \mathbf{\Sigma_1})) = \frac{1}{2}\Bigg(\log\left(\frac{det(\mathbf{\Sigma_1})}{det(\mathbf{\Sigma_0})}\right) + trace({\mathbf{\Sigma_1}}^{-1}\mathbf{\Sigma_0}) + (\mu_{\mathbf{1}} - \mu_{\mathbf{0}})^{\intercal}{\mathbf{\Sigma_1}}^{-1}(\mu_{\mathbf{1}} - \mu_{\mathbf{0}}) - D\Bigg) \tag{8}$$

where log(.) denotes the logarithm operator.

- By SSA, we get the stationary sources as

$$\underbrace{Y}_{\text{latent sources}} = V^T X \quad V \in \mathbb{R}^{D\times d}$$

- Optimizing D_{kl} gives $\mathbf{V}$!

Preserving stationarity

- SSA+CSP [Von Bünau et al. 2010 [10]]
  - Apply stationary subspace analysis
  - CSP for feature extraction
  - Cons: Different class have different characteristics so discriminative direction can be non-stationary
- groupSSA [Samek et al. 2012[11]]
  - Preserving stationarity within the classes of the session
  - Discriminativity and stationarity still not captured jointly!!

What if we simultaneously optimize stationarity and discriminability ?

[10] Von Bünau, P., Meinecke, F.C., Scholler, S. and Müller, K.R., 2010, August. Finding stationary brain sources in EEG data. In 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology (pp. 2810-2813). IEEE.

[11] Wojciech Samek, Klaus-Robert Müller, Motoaki Kawanabe, and Carmen Vi- daurre. Brain-computer interfacing in discriminative and stationary subspaces. In Engineering in Medicine and Biology Society (EMBC), 2012 Annual Inter- national Conference of the IEEE, pages 2873–2876. IEEE, 2012.

CSP reframed as an information theoretic problem!![Samek 2013 NIPS[12]]

Theorem

Let $\mathbf{W}_{CSP} \in \mathbb{R}^{D\times d}$ *denote CSP filters and* $\mathbf{\Sigma}_c$ *denote covariance matrix of class* c*. Let* $\mathbf{V}^\intercal = \overline{\mathbf{R}}\mathbf{W} \in \mathbb{R}^{d\times D}$ *be decomposable into a whitening projection* $\mathbf{W} \in \mathbb{R}^{D\times D}$ *(with* $\mathbf{W}(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)\mathbf{W}^\intercal = \mathbf{I}$*) and a truncated orthogonal projection* $\overline{\mathbf{R}} \in \mathbb{R}^{d\times D}$*. Then*

$$\mathbf{V}^* = \arg\max_{\mathbf{V}} sD_{kl}(\mathbf{V}^T\mathbf{\Sigma_1}\mathbf{V} \parallel \mathbf{V}^T\mathbf{\Sigma_2}\mathbf{V}) \tag{9}$$

$$Span(\mathbf{W}_{CSP}) = Span(\mathbf{V}^*) \tag{10}$$

- $sD_{kl}(\mathbf{A} \parallel \mathbf{B})$ *is the symmetric KL divergence between zero mean Gaussians with covariance matrices* $\mathbf{A}$ *and* $\mathbf{B}$*.* $Span(\mathbf{W}_{CSP})$ *stands for subspace spanned by columns of* $\mathbf{W}$*.*

[12] Samek, W., Blythe, D., Müller, K.R. and Kawanabe, M., 2013. Robust spatial filtering with beta divergence. In Advances in Neural Information Processing Systems (pp. 1007-1015).

Information Theoretic CSP

- Information theoretic interpretation of CSP

$$F(\mathbf{V}) = sD_{kl}(\mathbf{V}^T\boldsymbol{\Sigma}_1\mathbf{V} \parallel \mathbf{V}^T\boldsymbol{\Sigma}_2\mathbf{V}) \tag{11}$$

$$\mathbf{V}_{skl} = \arg\max_{\mathbf{V}} F(\mathbf{V}) \tag{12}$$

- sD_{kl} is the symmetric KL divergence. Σ_1 and Σ_2 are the class covariance matrices for class 1 and 2
- Enforcing stationarity in spatial filters
  - **Within Session (WS)** : KL divergence within the classes
- We only discuss to optimize within session stationarity

$$G(\mathbf{V}) = \frac{1}{2N}\sum_{c=1}^{2}\sum_{i=1}^{N} D_{kl}(\mathbf{V}^T\boldsymbol{\Sigma}_{\mathbf{i,c}}\mathbf{V} \parallel \mathbf{V}^T\boldsymbol{\Sigma}_{\mathbf{c}}\mathbf{V}) \tag{13}$$

- A composite objective function to enforce stationarity in spatial filters

$$\Delta(\mathbf{V}) = \underbrace{(1-\lambda)F(\mathbf{V})}_{\text{Information theoretic CSP}} - \underbrace{\lambda G(\mathbf{V})}_{\text{Stationary}} \tag{14}$$

Optimization technique

- Spatial filters can be optimized using following techniques
  - **Subspace Technique** : A group of filters optimized together
  - **Deflation technique** : Sequential optimization of filters
- Filters ($\mathbf{V}$) decomposed as a product of Whitening matrix ($\mathbf{W}$) and Orthogonal matrix ($\mathbf{R}$)

$$\mathbf{V}^\intercal = \tilde{\mathbf{R}}\mathbf{W} \quad \tilde{\mathbf{R}} = \mathbf{I_d R} \quad \mathbf{V} \in \mathbb{R}^{D\times d}, \mathbf{W} \in \mathbb{R}^{D\times D}$$
$$\mathbf{W}^\intercal(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\mathbf{W} = \mathbf{I}$$

- On an orthogonal manifold $\mathbf{R}\mathbf{R}^\intercal = \mathbf{I}$
- Objective function now depends on orthogonal matrix $\mathbf{R}$

$$\Delta(\mathbf{R}) = \underbrace{(1-\lambda)F(\mathbf{I}_d\mathbf{R})}_{\text{Information theoretic CSP}} + \underbrace{\lambda G(\mathbf{I}_d\mathbf{R})}_{\text{Stationary}}$$

- $1 < d < D$: Subspace approach, and $d = 1$: sequential optimization or Deflation approach

Optimization on orthogonal manifold [Von Bünau 2009]

- Optimization is performed using gradient descent on orthogonal manifold
- To stay on orthogonal manifold: Multiplicative update is performed

$$\mathbf{R}_{k+1} = \mathbf{B}_k \mathbf{R}_k$$

- At each step of optimization, rotation is absorbed into the data and gradient is calculated at identity
- On orthogonal manifold, gradient at **Identity** is skew symmetric
- Update matrix $\mathbf{B}_k$ is parametrised using exponential of skew symmetric gradient
- After convergence criteria is satisfied, Composite filter matrix is generated by multiplying optimum $\mathbf{R}$ with whitening matrix

Divergence CSP: Summary

- Divergence framework of CSP
  - Step-1: First the subspace is optimized to balance out discriminability and stationarity
  - Step-2: CSP is applied on data transformed through subspace in Step-1
- Generally, initialization of orthogonal matrix is the CSP solution

Divergence CSP framework not suitable for Multi-class BCIs !

Multi-class BCI: Quick Detour

- Wu et al. [13] proposed **One Versus Rest (OVR)** approach for Multiclass BCI (K number of classes)
- Dornhege et al. [14] proposed **Joint Approximate Diagonalisation (JAD)** technique to estimate the spatial filters
- JAD (simultaneous diagonalisation optimized using Cardoso method) is motivated by the idea of joint diagonalisation in binary class (Singular Value Decomposition)

$$\mathbf{W}^\intercal \mathbf{\Sigma}_{c_i} \mathbf{W} = \mathbf{D}_{c_i}, \quad i = 1, \cdots, K \tag{15}$$

$$\sum_{i=1}^{K} \mathbf{D}_{c_i} \approx \mathbf{I} \tag{16}$$

- Generally, JAD outperforms OVR-CSP [15]

[13] Wu, W., Gao, X. and Gao, S., 2006, January. One-versus-the-rest (OVR) algorithm: An extension of common spatial patterns (CSP) algorithm to multi-class case. In 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference (pp. 2387-2390). IEEE.

[14] Dornhege, G., Blankertz, B., Curio, G. and Müller, K.R., 2004. Increase information transfer rates in BCI by CSP extension to multi-class. In Advances in neural information processing systems (pp. 733-740).

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

Information theoretic OVR-CSP (IT-OVR-CSP)(Contribution!!)

- Modelled similar to Information theoretic CSP
- Mathematically, for i^{th} OVR ,

$$F_i^{ovr}(\mathbf{V}) = sD_{kl}(\mathbf{V}^\intercal \mathbf{\Sigma}_i \mathbf{V} \parallel \mathbf{V}^\intercal \mathbf{\Sigma}_{ovr_i} \mathbf{V})$$

$$\mathbf{V}_i^* = \underset{\mathbf{V}}{\arg\max}\ F_i^{ovr}(\mathbf{V})$$

$$\mathbf{V}_{final} = [\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_{2K}]$$

- IT-OVR-CSP can thus be included in divergence framework to enforce stationarity

Information theoretic JAD (IT-JAD)

- Gouy-Pailler et al. [16] proposed information theoretic interpretation of JAD

$$F(\mathbf{V}) = \sum_{c=1}^{K} p_k D_{kl}(\mathbf{V}^\top \mathbf{\Sigma}_c \mathbf{V} \parallel diag(\mathbf{V}^\top \mathbf{\Sigma}_c \mathbf{V}))$$

$$\mathbf{V}^* = \underset{\mathbf{V}}{\arg\min}\ F(\mathbf{V})$$

- $\mathbf{\Sigma}_c$ is the class covariance-matrix of the c^{th} class
- diag($\mathbf{X}$) refers to square diagonal matrix with its elements having value same as that of square matrix $\mathbf{X}$
- IT-JAD can be optimized on orthogonal manifold

[16] Gouy-Pailler, C., Congedo, M., Brunner, C., Jutten, C. and Pfurtscheller, G., 2009. Nonstationary brain source separation for multiclass motor imagery. IEEE transactions on Biomedical Engineering, 57(2), pp.469-478.

Enforcing Stationarity in Multiclass

- Within session (WS) stationarity in multiclass can be preserved by minimizing the objective function

$$G(\mathbf{V}) = \frac{1}{KN}\sum_{c=1}^{K}\sum_{i=1}^{N} D_{kl}(\mathbf{V}^{\intercal}\mathbf{\Sigma}_{i,c}\mathbf{V} \parallel \mathbf{V}^{\intercal}\mathbf{\Sigma}_{c}\mathbf{V})$$

- $\mathbf{\Sigma}_c$ is class co-variance matrix of class c, N is total number of trials in each class and K is number of distinct classes

Enforcing Stationarity in Multiclass

- Divergence framework to induce stationarity
  - DivOVR-CSP-WS

$$\mathbf{V}_i^* = \arg\max_{\mathbf{V}} \; (1-\lambda)F_i^{ovr}(\mathbf{V}) - \lambda G(\mathbf{V}) \tag{17}$$

  - DivJAD-WS

$$\begin{aligned}\Delta(\mathbf{V}) = \Bigg[&(1-\lambda)\left(\sum_{k=1}^{K} p_k D_{kl}(\mathbf{V}^\intercal \mathbf{\Sigma}_k \mathbf{V} \parallel diag(\mathbf{V}^\intercal \mathbf{\Sigma}_k \mathbf{V}))\right) \\ &+ \lambda\left(\frac{1}{KN}\sum_{c=1}^{K}\sum_{i=1}^{N} D_{kl}(\mathbf{V}^\intercal \mathbf{\Sigma}_{i,c}\mathbf{V} \parallel \mathbf{V}^\intercal \mathbf{\Sigma}_c \mathbf{V})\right)\Bigg]\end{aligned}$$

$$\mathbf{V}^* = \arg\min_{\mathbf{V}} \; \Delta(\mathbf{V})$$

Optimization Technique: DivJAD-WS

- Similar to DivCSP-WS
  - Subspace approach
  - Deflation Approach
- Sub-optimal solution than CSP approach when the initial orthogonal matrix is not the solution of CSP
- Not a suitable technique for optimizing DivJAD-WS
  - **Modified Subspace Approach**: Simultaneous optimization of Joint Diagonalisation and Stationarity in first 'd' Directions

$$\mathbf{R}^* = \arg\min_{\mathbf{R}} \, (1-\lambda)J(\mathbf{R}) + \lambda J_s(\mathbf{I}_d\mathbf{R}) \tag{18}$$

- Optimal $\mathbf{R}$ is used to construct spatial filters
- ITFE technique used for sorting the filters

Results: DivJAD-WS and DivOVR-WS v/s IT-JAD, OVR

	IT-JAD	IT-OVRCSP	DivJAD-WS$_2$	DivOVR-CSP-WS
S1	78.51	71.01	80.65	75.65
S2	56.13	55.54	58.21	56.85
S3	84.58	82.68	85.60	85.65
S4	54.82	55.24	56.07	55.60
S5	39.05	36.79	39.29	37.08
S6	47.74	47.26	50.00	49.82
S7	78.63	67.08	78.39	68.87
S8	80.83	84.05	84.58	84.76
S9	73.10	73.87	75.77	75.00
Mean CV	65.93	63.72	**67.62**	**65.48**
Mean test	63.46	62.85	**65.16**	**64.08**

Results: DivJAD-WS vs DivOVR-CSP-WS

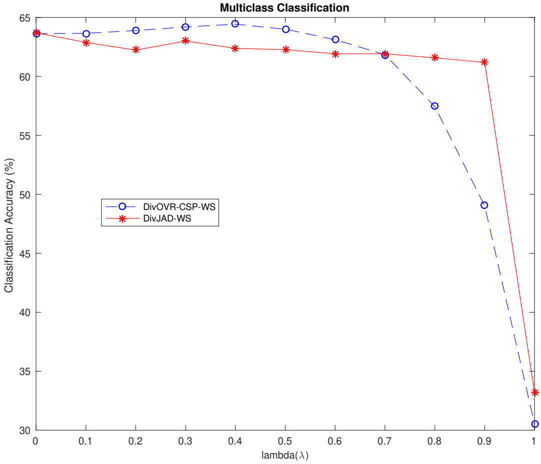


Figure: Effect of varying λ on mean classification accuracies estimated using DivJAD-WS and DivOVR-CSP-WS

Contributions

- A composite framework to incorporate stationarity in multiclass BCI is proposed
- Novel optimization technique for simultaneous optimization of JAD term and stationarity term is proposed

Conclusion and Outlook II

- Efficient selection of optimization parameter λ
- Cross validation strategy to optimize λ
- Integration of Across subject stationarity to estimate robust spatial filters

References

1 Gouy-Pailler, C., Congedo, M., Brunner, C., Jutten, C. and Pfurtscheller, G., 2010. Nonstationary brain source separation for multiclass motor imagery. IEEE transactions on Biomedical Engineering, 57(2), pp.469-478.

2 Grosse-Wentrup, M. and Buss, M., 2008. Multiclass common spatial patterns and information theoretic feature extraction. IEEE transactions on Biomedical Engineering, 55(8), pp.1991-2000.

3 Samek, W., 2016. On robust spatial filtering of EEG in nonstationary environments. it-Information Technology, 58(3), pp.150-154.

4 Von Bünau, P., Meinecke, F.C., Király, F.C. and Müller, K.R., 2009. Finding stationary subspaces in multivariate time series. Physical review letters, 103(21), p.214101.

References

5 Samek, W., Müller, K.R., Kawanabe, M. and Vidaurre, C., 2012, August. Brain-computer interfacing in discriminative and stationary subspaces. In Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE (pp. 2873-2876). IEEE.

6 Wu, W., Gao, X. and Gao, S., 2006, January. One-versus-the-rest (OVR) algorithm: An extension of common spatial patterns (CSP) algorithm to multi-class case. In Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the (pp. 2387-2390). IEEE.

7 Dornhege, G., Blankertz, B., Curio, G. and Müller, K.R., 2004. Increase information transfer rates in BCI by CSP extension to multi-class. In Advances in neural information processing systems (pp. 733-740).

8 Volkmann, J., Joliot, M., Mogilner, A., Ioannides, A.A., Lado, F., Fazzini, E., Ribary, U. and Llinas, R., 1996. Central motor loop oscillations in parkinsonian resting tremor revealed

9 Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M. and Muller, K.R., 2008. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal processing magazine, 25(1), pp.41-56.

10 Lotte, F. and Guan, C., 2011. Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. IEEE Transactions on biomedical Engineering, 58(2), pp.355-362.

11 Ang, K.K., Chin, Z.Y., Zhang, H. and Guan, C., 2008, June. Filter bank common spatial pattern (FBCSP) in brain-computer interface. In Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on (pp. 2390-2397). IEEE.

12 Zhang, Y., Zhou, G., Jin, J., Wang, X. and Cichocki, A., 2015. Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface. Journal of neuroscience methods, 255, pp.85-91.

References

13 Lachaux, J.P., Rodriguez, E., Le Van Quyen, M., Lutz, A., Martinerie, J. and Varela, F.J., 2000. Studying single-trials of phase synchronous activity in the brain. International Journal of Bifurcation and Chaos, 10(10), pp.2429-2439.

14 Lachaux, J.P., Rodriguez, E., Martinerie, J. and Varela, F.J., 1999. Measuring phase synchrony in brain signals. Human brain mapping, 8(4), pp.194-208.

15 Caramia, N., Lotte, F. and Ramat, S., 2014, May. Optimizing spatial filter pairs for EEG classification based on phase synchronization. In International Conference on Audio, Speech and Signal Processing (ICASSP'2014).

Thank You