Common Spatial Patterns for BCI (Spatial Filters)

Alberto Llera, Bert Kappen

Radboud University of Nijmegen

October 15, 2009

Introduction

- Assume a two class classification problem.
- The idea of CSP is to proyect the data in a space in which the difference between the variance of the two classes is maximal.
- CSP is based in the simultaneous diagonalization of two symmetric matrices, originally described by K. Fukunaga in "Introduction to statistical pattern recognition"

Simultaneous diagonalization of two covariance matrices

- Let C_1, C_2 be respectivelly the covariance matrix of two different class trials.
- Define $C_t := C_1 + C_2$.
- Find matrixes W and diagonal λ such that

$$W'C_tW = \lambda \tag{1}$$

- Define the "whitening transformation" P as

$$P := 1/\sqrt{(}\lambda^{-1}). \tag{2}$$

- Then

$$P'W'C_tWP = I \tag{3}$$

Simultaneous diagonalization of two covariance matrices

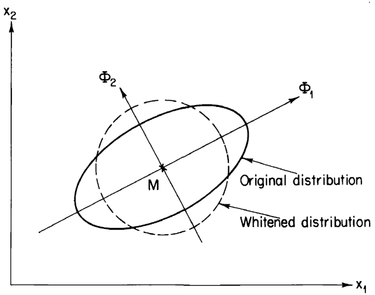


Fig. 2-2 Whitening process.

- After the whitening transformation, the covariance matrix is invariant under any orthonormal transformation O since

$$O'IO = I \tag{4}$$

Simultaneous diagonalization of two covariance matrices

- Since $C_t := C_1 + C_2$

$$P'W'(C_1 + C_2)WP = I \tag{5}$$

$$(P'W'C_1WP) + (P'W'C_2WP) = I \tag{6}$$

- Let

$$P'W'C_1WP = K_1 \tag{7}$$

$$P'W'C_2WP = K_2 \tag{8}$$

where K_1 and K_2 are not necessarily diagonal.

Simultaneous diagonalization of two covariance matrices

- Find matrices S_2 and diagonal D_2 such that

$$S_2' K_2 S_2 = D_2 \tag{9}$$

- Then

$$S_2' P' W' C_t W P S_2 = S_2' I S_2 = I \tag{10}$$

- And then

$$S_2' K_1 S_2 = I - D_2 \tag{11}$$

CSP

- Define $M := WPS_2$. Then

$$M'C_2M = D_2 \tag{12}$$

and

$$M'C_1M = I - D2 \tag{13}$$

- The eigenvector asociated with the biggest eigenvalue for C_2 is also asociated with the smallest eigevalue for C_1, etc...
- Select C_1 or C_2. Choose the k biggest and the k smallest eigenvalues (for some $k \leq n$) and their respective eigevectors (which are our filters).
- Build a matrix $B \epsilon M_{n \times 2k}$ with the selected eigenvectors as columns.
- Proyect the data into this matrix.

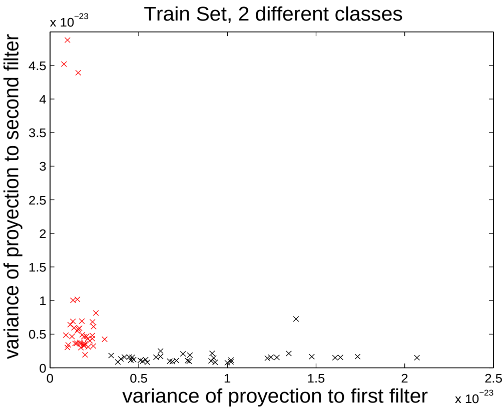
Processing of the data

- MEG data.
- Automatic artifacts removal.
- High pass filter 30 Hz.
- Reduce trials length. New trials go from 0.1 to 0.7 sec. after device feedback.
- Normalize data (substracting mean).
- Apply CSP.
- Calculate variance of the proyected data as features for classification.

Results in subject one

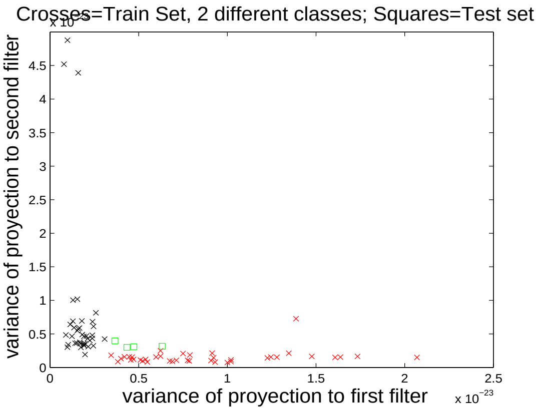
- 50-60 % of classification using 6 filters and FLD as classifier.
- Similar results using a perceptron as classifier.
- Another researchers have shown that the use of a SVM can improve this results but not very significantly.
- What is wrong with CSP?

Train set proyected to two filters



- Proyecting the train data in 2 dim. we see that it already is linearly separable.
- How does it generalize?

Train and Test set proyected to two filters



- The generalization to test data is not as good as we would like, however it increses as we increase dimension.

Overfitting

- CSP performs very good in the train set, but does not generalize good to new data, i.e. it suffers from overfitting.
- Can we add a regularizer to control overfitting?

Rayleigh quotient

- The solution for the eigevector with the biggest eigevalue can be found by maximizing the Rayleigh quotient:

$$argmax_w \quad \frac{w' C_2 w}{w' C_t w} \tag{14}$$

- Its optima are the solutions of the generalized eigenvalue problem (Fukunaga).
- So we can rewrite the CSP in terms of the Rayleigh quotient and add a regularizer to w.

RCSP-Regularized CSP (J.Farquhar)

- Find first filter w:

$$argmax_w \quad \frac{w'C_2w}{w'C_tw} - \lambda\frac{|w|_1}{|w|_2} \tag{15}$$

- Once a filter is found, we can find subsequent filters by deflating C_2

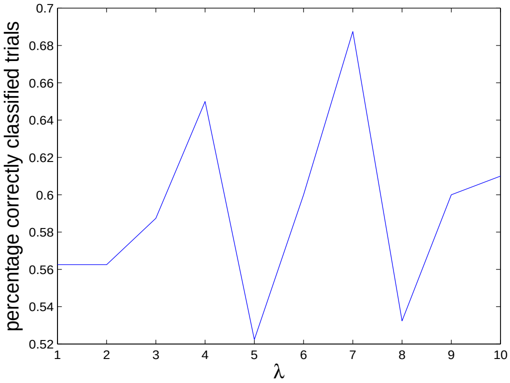
$$C_2 \leftarrow C_2(I - \frac{w'wC_t}{w'C_tw}) \tag{16}$$

- Iterate till find all wished filters.

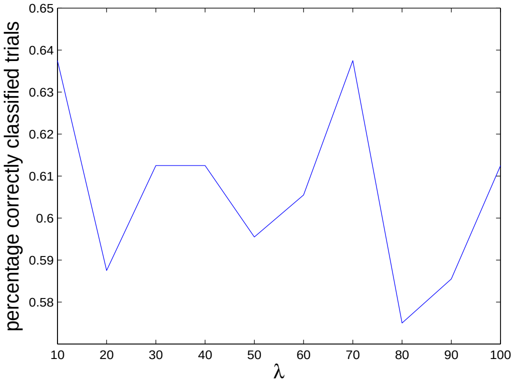
Results RCSP subject one

- We applied RCSP to our data set.
- Balance data.
- 20 fold cross-validation.

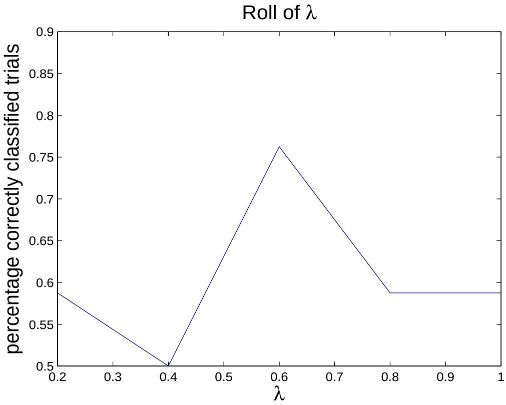
Results RCSP subject one



Results RCSP subject one



More lambdas



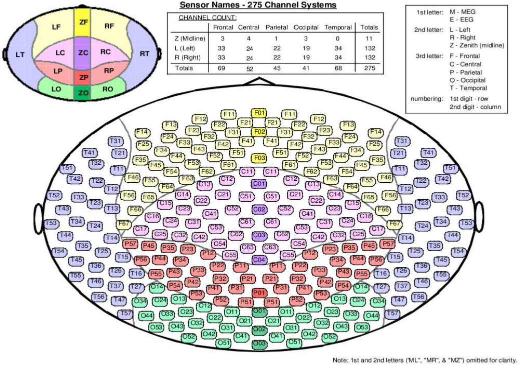
RCSP vs CSP

- RCSP outperforms CSP in our experiments;
- Is still an open question how to fix λ.
- The variance is too big. (10-17 %).

Optimization of the frequency band

- We transform the data to the frequency domain and we average over trials of the same class.
- We plot the class-dependent frequency-power representation.

Sens-layout

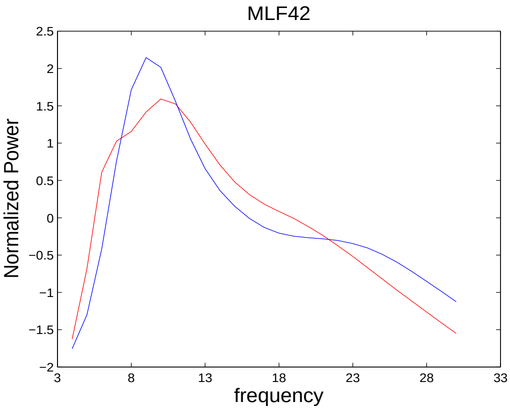


CHANNEL COUNT:

	Frontal	Central	Parietal	Occipital	Temporal	Totals
Z (Midline)	3	4	1	3	0	11
L (Left)	33	24	22	19	34	132
R (Right)	33	24	22	19	34	132
Totals	69	52	45	41	68	275

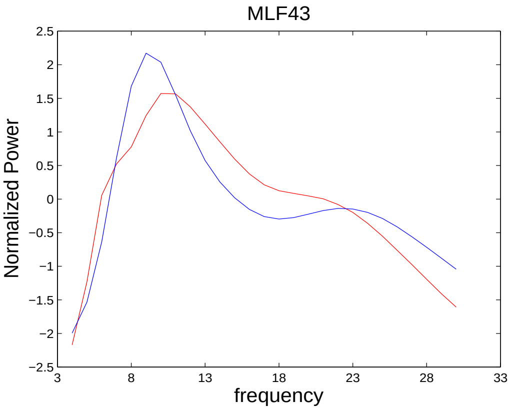
Nice sensors

- blue=correct trials; red= error trials;



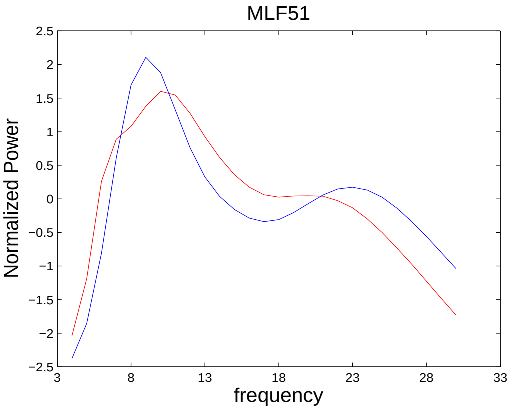
Nice sensors

- blue=correct trials; red= error trials;



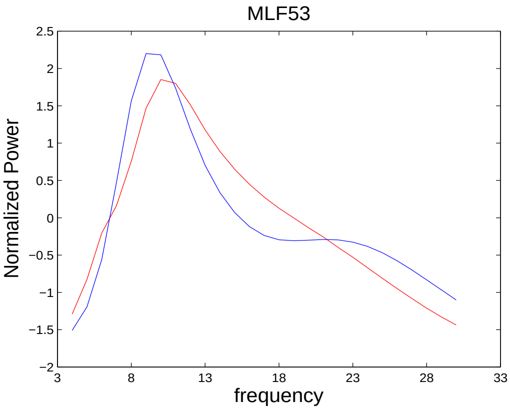
Nice sensors

- blue=correct trials; red= error trials;



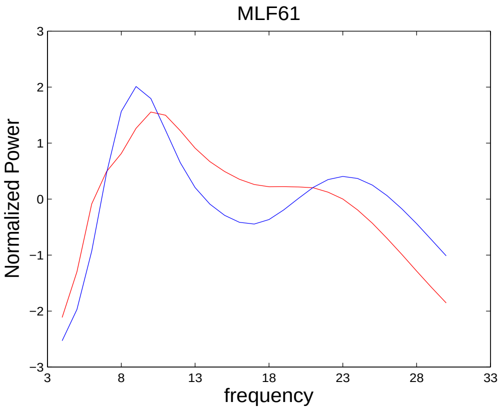
Nice sensors

- blue=correct trials; red= error trials;



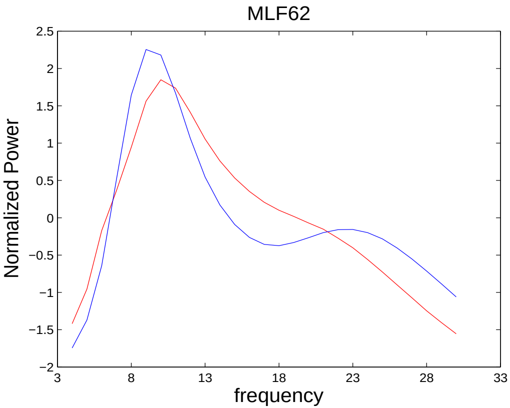
Nice sensors

- blue=correct trials; red= error trials;



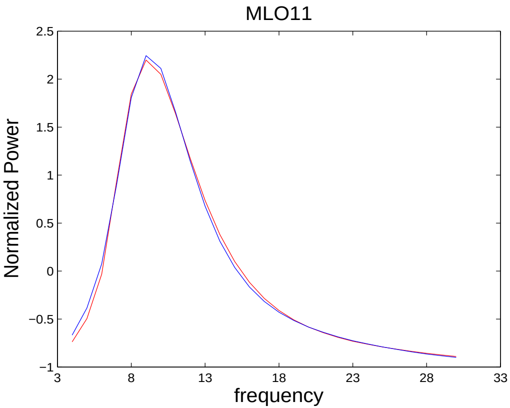
Nice sensors

- blue=correct trials; red= error trials;



Example normal sensors

- blue=correct trials; red= error trials;



Frequency selection

- Our results show a decrease in power in the "error trials" in the alfa band (8-11 Hz), an increase in the low-beta band (13-20 Hz) and again a decrease in the high beta band (25-30 Hz) in left fronto central sensors.
- The decrese in power in the alfa band in front channels can be related with the well known Error Related Negativity in EEG studies.

Coming next

- Apply RCSP in the predefined alfa and low beta bands.
- Implement a SVM.
- AR model before Spatial filters?
- Try ICA.

THANKS

- THANKS

THE END

- FIN