

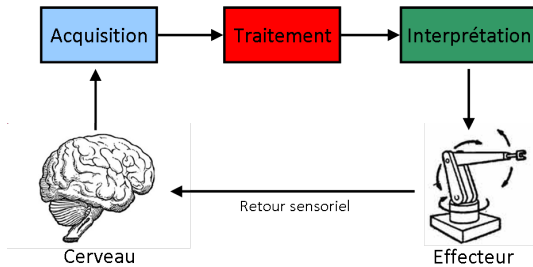
# Classification de potentiels évoqués P300 par géométrie riemannienne pour les interfaces cerveau-machine EEG.

A. Barachant, M. Congedo, G. Van Veen et C. Jutten

Team ViBS (Vision and Brain Signal Processing),  
GIPSA-lab, CNRS, Grenoble University, France

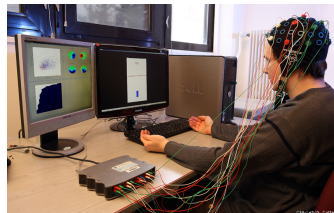
GRETSI 2013

# Interfaces Cerveau Machine (ICM / BCI)

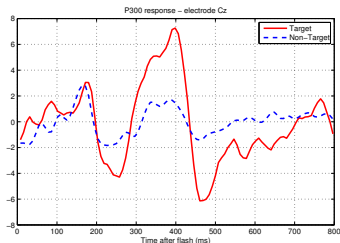


## Applications :

- Médicales
- Grand public

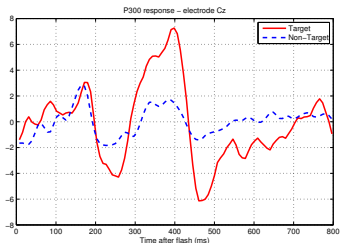


Interface cerveau-machine (ICM) P300 :



**Un problème de classification binaire - présence / absence du P300.**

## Interface cerveau-machine (ICM) P300 :



**Un problème de classification binaire - présence / absence du P300.**

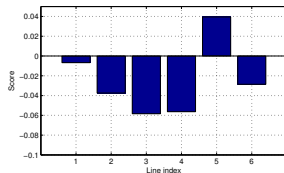
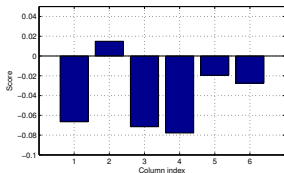
## En pratique :

- Découpage des signaux EEG pour chaque stimulus :

$$\mathbf{X}_i \in \mathbb{R}^{C \times N}$$

- Score de classification :

$$y_i = f(\mathbf{X}_i)$$



- Intersection de la ligne / colonne de score maximal.

## Une grande variabilité entre sujets et sessions !

⇒ Le classifieur optimal est spécifique au sujet et à la session.



### Pour :

- Adapté au sujet.
- Adapté au système.
- Performances presque optimales.

### Contre :

- Avant chaque session
- Ennuyeux
- Gaspillage de données

**Jeux Vidéo  $\implies$  calibration rédhibitoire !**

ICM sans Calibration

- 1 Initialisation avec un classifieur générique.
- 2 Adaptation en ligne.

**Prérequis :**

- **Bonne généralisation entre sessions / sujets**
- **Calibration à partir de peu de données.**
- Pas d'hyperparamètres à régler.

## Stepwise LDA :

- Sous échantillonnage
  - Vectorisation
  - Sélection de Modèle
  - Classification LDA
- 😊 Bonne généralisation entre sessions / sujets
- ☹ Nécessite beaucoup de données

## XDAWN + LDA Régularisée :

- Filtrage spatial
  - Sous échantillonnage
  - Vectorisation
  - LDA Régularisée
- 😊 Nécessite peu de données
- ☹ Mauvaise généralisation entre sessions / sujets

**Ces méthodes ne conviennent pas !**



### **Extraction de caractéristiques :**

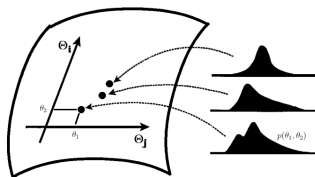
- Construire un espace où les classes sont séparables suivant une métrique usuelle.
- Une bonne manière de procéder !

### **Mais :**

- " Consomme" trop de données (SWLDA)
- Est trop spécifique (XDAWN)

**Une autre approche : Trouver la métrique naturelle pour le problème en question.**

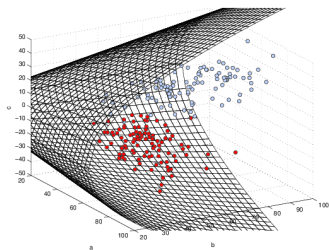
- Distributions d'une même famille :
- ⇒ Points d'une variété Riemannienne



- Une métrique possible : l'**information de Fisher**.
- Métrique  $\Rightarrow$  Produit scalaire  $\Rightarrow$  Distance

- Modèle :

$$\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \Sigma_i)$$



- Dimension :  $C(C + 1)/2$
- Métrique :  $\frac{1}{2} \text{Tr} \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_l} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_k} \right)$

- Distance :

$$\delta_R(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \|\log(\boldsymbol{\Sigma}_1^{-1/2} \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1/2})\|_F = \left[ \sum_{c=1}^C \log^2 \lambda_c \right]^{1/2}$$

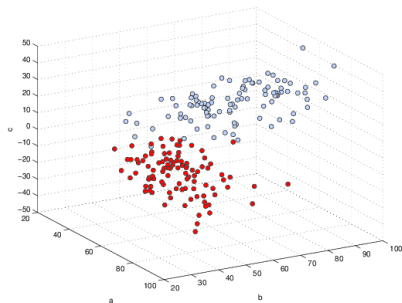
- Moyenne :

$$\mathfrak{G}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_I) = \operatorname{argmin}_{\boldsymbol{\Sigma} \in P(n)} \sum_{i=1}^I \delta_R^2(\boldsymbol{\Sigma}, \boldsymbol{\Sigma}_i).$$

## Distance minimale aux centres (MDM)

Apprentissage

Classification

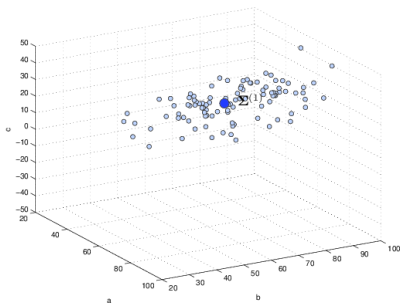


## Distance minimale aux centres (MDM)

### Apprentissage

- Estimation des centres par moyenne géométrique :
  - ①  $\Sigma^{(1)}$  pour la classe (1)

### Classification

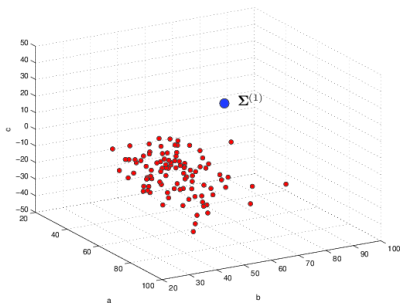


## Distance minimale aux centres (MDM)

### Apprentissage

- Estimation des centres par moyenne géométrique :
  - ①  $\Sigma^{(1)}$  pour la classe (1)

### Classification

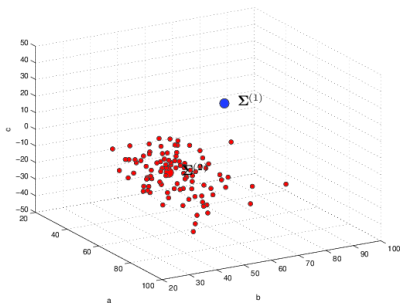


## Distance minimale aux centres (MDM)

### Apprentissage

- Estimation des centres par moyenne géométrique :
  - 1  $\Sigma^{(1)}$  pour la classe (1)
  - 2  $\Sigma^{(2)}$  pour la classe (2)

### Classification



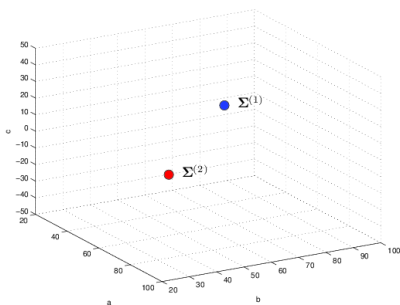


## Distance minimale aux centres (MDM)

### Apprentissage

- Estimation des centres par moyenne géométrique :
  - 1  $\Sigma^{(1)}$  pour la classe (1)
  - 2  $\Sigma^{(2)}$  pour la classe (2)

### Classification



## Distance minimale aux centres (MDM)

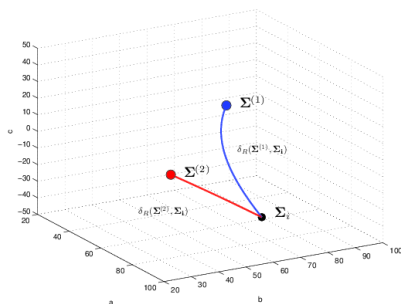
### Apprentissage

- Estimation des centres par moyenne géométrique :
  - 1  $\Sigma^{(1)}$  pour la classe (1)
  - 2  $\Sigma^{(2)}$  pour la classe (2)

### Classification

- Une matrice  $\Sigma_i$  de classe  $y_i$  inconnue.

$$y_i = \delta_R(\Sigma^{(2)}, \Sigma_i) - \delta_R(\Sigma^{(1)}, \Sigma_i)$$



- Invariance par projection congruente, i.e,  $\forall \mathbf{V} \in \mathcal{G}l(C)$

$$\delta_R(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \delta_R(\mathbf{V}^T \boldsymbol{\Sigma}_1 \mathbf{V}, \mathbf{V}^T \boldsymbol{\Sigma}_2 \mathbf{V}).$$

- Séparation de sources, mélange instantané :

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \implies \boldsymbol{\Sigma} = \mathbf{A}\boldsymbol{\Sigma}_s\mathbf{A}^T$$

## Conséquence

Les distances de Riemann dans l'espace des capteurs et dans l'espace des sources sont égales !

$$\delta_R(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \delta_R(\boldsymbol{\Sigma}_{s_1}, \boldsymbol{\Sigma}_{s_2}).$$

- Sources indépendantes : matrices de covariance diagonales.

$$\delta_R(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = \|\log(\sigma_{s1}) - \log(\sigma_{s2})\|_2$$

- La distance Riemannienne mesure la différence des log-variance des sources :
  - Sans estimer la matrice de mélange !
  - Optimale pour  $C$  sources

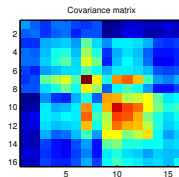
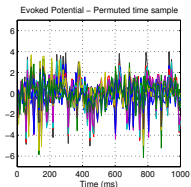
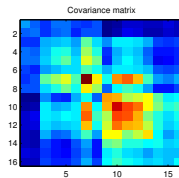
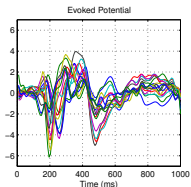
**Parfait lorsque les caractéristiques sont spatio-fréquentielles !**

- 1 A. Barachant, S. Bonnet, M. Congedo and C. Jutten, "Multiclass Brain-Computer Interface Classification by Riemannian Geometry," in IEEE Transactions on Biomedical Engineering, vol. 59, no. 4, p. 920-928, 2012.
- 2 A. Barachant, S. Bonnet, M. Congedo and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications", in NeuroComputing, vol. 112, p. 172-178, 2013.

## Problème avec les Potentiels évoqués :

- L'estimation de la covariance indépendante de la forme d'onde

$$\Sigma_i = \frac{1}{N-1} \mathbf{X}_i \mathbf{X}_i^T$$



## Problème avec les Potentiels évoqués :

- L'estimation de la covariance indépendante de la forme d'onde

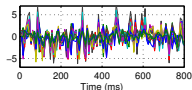
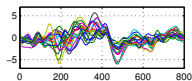
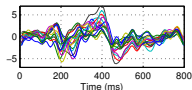
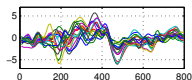
Potentiel évoqué prototypique

$$\mathbf{P}_1 = \frac{1}{M} \sum_m \mathbf{x}_m$$

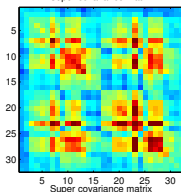
Concaténation :

$$\tilde{\mathbf{x}}_i = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{x}_i \end{bmatrix}$$

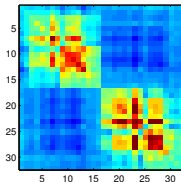
$$\tilde{\Sigma}_i = \begin{bmatrix} \Sigma_{P_1} & \mathbf{C}_{P_1, X_i}^T \\ \mathbf{C}_{P_1, X_i} & \Sigma_i \end{bmatrix}$$



Super covariance matrix



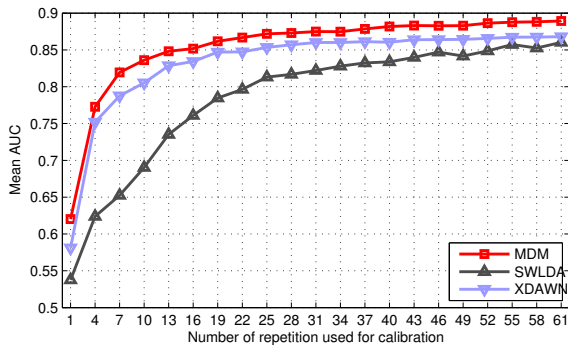
Super covariance matrix



- 23 sujets, 10 Ag/AgCl électrodes.
- Résultat Hors ligne : AUC / Convergence

# Résultat - Jeu de données 1 - Convergence

- 23 sujets, 10 Ag/AgCl électrodes.
- Résultat Hors ligne : AUC / Convergence



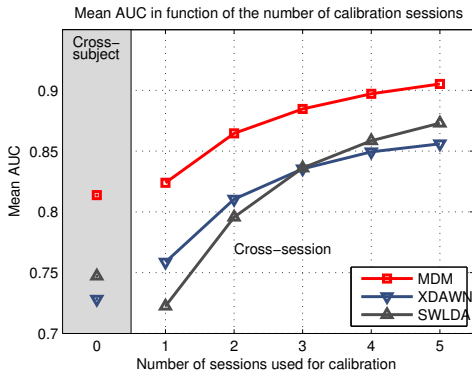


## Résultat - Jeu de données 2

- Pas de calibration / plusieurs sessions
- 4 sujets - 6 sessions - 16 électrodes sèches
- Résultat Hors ligne : AUC / Inter sessions et sujets

# Résultat - Jeu de données 2

- Pas de calibration / plusieurs sessions
- 4 sujets - 6 sessions - 16 électrodes sèches
- Résultat Hors ligne : AUC / Inter sessions et sujets



## Une Nouvelle methode de classification des P300 :

- Amélioration des performances.
  - Convergence rapide.
  - Bonne généralisation entre sujets/sessions.
  - Méthode Simple, pas d'hyperparamètres.
  - Unification des paradigmes (SMR, SSVEP, P300)
- 
- Problèmes en grande dimension ( $C > 64$ )
  - Interprétation des caractéristiques.

Code : <http://github.com/alexandrebarachant/covariancetoolbox>