# Learning from different sources

## Machine Learning for NeuroImaging Workshop

Marie Szafranski

Joint work with Yves Grandvalet and Alain Rakotomamonjy

école nationale supérieure d'informatique pour l'industrie et l'entreprise

# Illustration in BCI

## P300 Speller

| | | | | | |
|---|---|---|---|---|---|
| A | B | C | D | E | F |
| G | H | I | J | K | L |
| M | N | O | P | Q | R |
| S | T | U | V | W | X |
| Y | Z | 1 | 2 | 3 | 4 |
| 5 | 6 | 7 | 8 | 9 | – |

- A person chooses a character in the grid (say V)
- His cerebral activity is measured with EEG
- Lines and columns are randmonly intensified (12 times)
- When the character is lighted, the person has to count

## Outcome

**A potentiel P300 occurs in the EEG signals**

# Illustration in BCI

## P300 Speller

| | | | | | |
|---|---|---|---|---|---|
| A | B | C | D | E | F |
| G | H | I | J | K | L |
| M | N | O | P | Q | R |
| S | T | U | V | W | X |
| Y | Z | 1 | 2 | 3 | 4 |
| 5 | 6 | 7 | 8 | 9 | _ |

- A person chooses a character in the grid (say V)
- His cerebral activity is measured with EEG
- Lines and columns are randmonly intensified (12 times)
- When the character is lighted, the person has to count

## Outcome

**A potentiel P300 occurs in the EEG signals**

# Illustration in BCI

## P300 Speller

| | | | | | |
|---|---|---|---|---|---|
| A | B | C | D | E | F |
| G | H | I | J | K | L |
| M | N | O | P | Q | R |
| **S** | **T** | **U** | **V** | **W** | **X** |
| Y | Z | 1 | 2 | 3 | 4 |
| 5 | 6 | 7 | 8 | 9 | – |

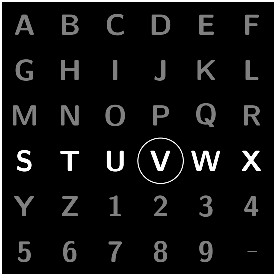
- A person chooses a character in the grid (say V)
- His cerebral activity is measured with EEG
- Lines and columns are randmonly intensified (12 times)
- When the character is lighted, the person has to count

## Outcome

**A potentiel P300 occurs in the EEG signals**

# Illustration in BCI

## P300 Speller




- A person chooses a character in the grid (say V)
- His cerebral activity is measured with EEG
- Lines and columns are randmonly intensified (12 times)
- When the character is lighted, the person has to count

## Outcome

**A potentiel P300 occurs in the EEG signals**

# Illustration in BCI

## P300 Speller

| | | | | | |
|---|---|---|---|---|---|
| A | B | C | D | E | F |
| G | H | I | J | K | L |
| M | N | O | P | Q | R |
| S | T | U | V | W | X |
| Y | Z | 1 | 2 | 3 | 4 |
| 5 | 6 | 7 | 8 | 9 | _ |

- A person chooses a character in the grid (say V)
- His cerebral activity is measured with EEG
- Lines and columns are randmonly intensified (12 times)
- When the character is lighted, the person has to count

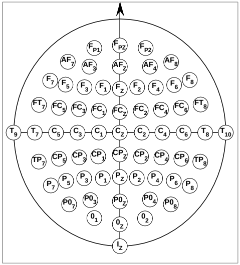
## Outcome

**A potentiel P300 occurs in the EEG signals**

# Illustration in BCI

## Signals acquisition

**And the sources appear...**




- EGG signals are collected over
  64 electrodes → 64 sources
- Each signal is sampled, leading to
  14 temporal frames → 64 sources × 14 measures
- Additional information : P300 occurs
  $\simeq$ 300 ms → frame 7

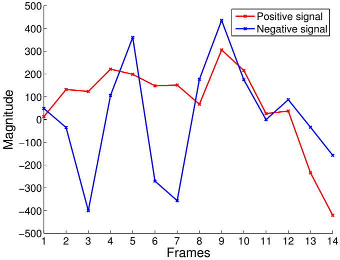
## Question

**Are all the sources and frames significant to recognize a P300 ?**

# Illustration in BCI

## Signals acquisition

And the sources appear...




- EGG signals are collected over
  64 electrodes — 64 sources
- Each signal is sampled, leading to
  14 temporal frames — 64 sources $\times$ 14 measures
- Additional information : P300 occurs
  $\simeq$ 300 ms — frame 7

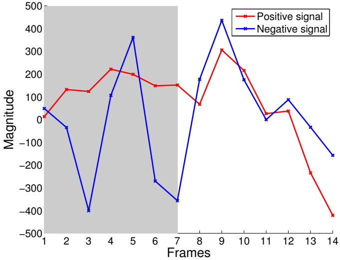
## Question

Are all the sources and frames significant to recognize a P300 ?

# Illustration in BCI

## Signals acquisition

**And the sources appear. . .**




- EGG signals are collected over
  64 electrodes — 64 sources
- Each signal is sampled, leading to
  14 temporal frames — 64 sources $\times$ 14 measures
- Additional information : P300 occurs
  $\simeq$ 300 ms — frame 7

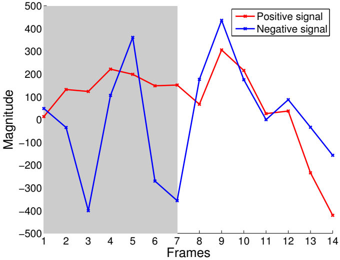
## Question

**Are all the sources and frames significant to recognize a P300 ?**

# Illustration in BCI

## Signals acquisition

**And the sources appear...**



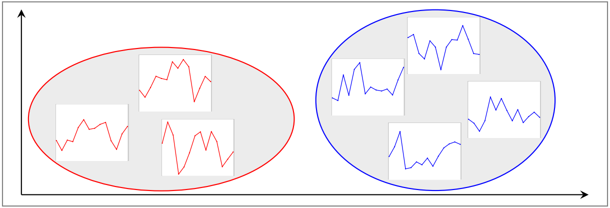

- EGG signals are collected over
  64 electrodes — 64 sources
- Each signal is sampled, leading to
  14 temporal frames — 64 sources $\times$ 14 measures
- Additional information : P300 occurs
  $\simeq$ 300 ms — frame 7

## Question

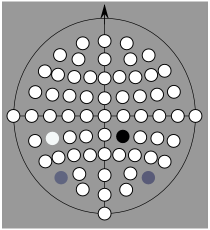
**Are all the sources and frames significant to recognize a P300 ?**

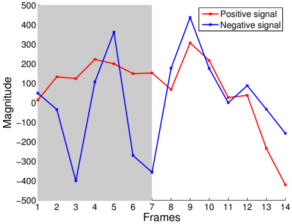
# Goal

## 1. Discriminate the signals that contains a P300



## 2. Identify the significant electrodes and the significant frames





# Outline

**Some machine learning tools**

**Composite Kernel Learning**

**BCI experiments**

# Dataset and classification

## Dataset

- A training set of $n$ samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$
- Observation

$$\boldsymbol{x}_i = (x_i^1, \ldots, x_i^M) \ , \ \boldsymbol{x}_i \in \mathcal{X}$$

- Label

$$y_i \in \{-1, +1\}$$

## Classify any $x \rightarrow$ learn $f$ such that

$$x \xrightarrow{f(x)} \begin{cases} \text{if } f(x) > 0 : \\ \text{if } f(x) < 0 : \end{cases}$$

# Dataset and classification

## Dataset

- A training set of $n$ samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$
- Observation

$$\boldsymbol{x}_i = (x_i^1, \ldots, x_i^M) \;\; , \; \boldsymbol{x}_i \in \mathcal{X}$$

- Label

$$y_i \in \{-1, +1\}$$

## Classify any $\boldsymbol{x} \rightarrow$ learn $f$ such that

$\boldsymbol{x}$  $\xrightarrow{f(\boldsymbol{x})}$

if $f(\boldsymbol{x}) > 0$ : 

if $f(\boldsymbol{x}) < 0$ : 

# Support Vector Machines (SVM)

## Find $f$ with linear SVM

## Hinge loss function


### Maximal margin hyperplane

$$\min_{f,b} \quad \frac{1}{2}\|f\|^2 + C\sum_{i=1}^{n} h(y_i, f(\boldsymbol{x}_i), b)$$

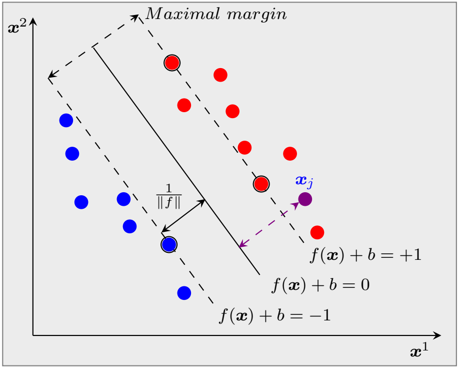
- $C$ : regularization parameter

  $\searrow C$ allows to $\nearrow$ number of errors $Ch(y_i, f(\boldsymbol{x}_i))$ becomes insignificant

- $h(y_i, f(\boldsymbol{x}_i)) = [1 - y_i.(f(\boldsymbol{x}_i) + b)]_+$

# Support Vector Machines (SVM)

## Find $f$ with linear SVM




### Maximal margin hyperplane

$$\min_{f,b} \quad \frac{1}{2}\|f\|^2 + C\sum_{i=1}^{n} h(y_i, f(\boldsymbol{x}_i), b)$$

- $C$ : regularization parameter

  $\searrow C$ allows to $\nearrow$ number of errors
  $Ch(y_i, f(\boldsymbol{x}_i))$ becomes insignificant

- $h(y_i, f(\boldsymbol{x}_i)) = [1 - y_i.(f(\boldsymbol{x}_i) + b)]_+$

# Reproducing Kernel Hilbert Spaces

## From input space $\mathcal{X}$ to feature space $\mathcal{H}$

Let $\phi : \mathcal{X} \to \mathcal{H}, \quad \boldsymbol{x} \mapsto \phi(\boldsymbol{x})$

- $\mathcal{H}$ is a RKHS with

  a dot product and a norm

  $$\|.\|^2_{\mathcal{H}} = \langle ., . \rangle_{\mathcal{H}}$$

  a symmetric and p. d. reproducing kernel

  $$K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$$

  with a reproducing property

  $$\forall f \in \mathcal{H}, \quad f(\boldsymbol{x}) = \langle f, K(\boldsymbol{x}, .) \rangle_{\mathcal{H}}$$

- Representer theorem

  Each minimizer $f^\star$ of $\min_f \; \frac{1}{2}\|f\|^2 + C\sum_{i=1}^n h(y_i, f(\boldsymbol{x}_i))$ can be written

  $$f^\star(\boldsymbol{x}) = \sum_{i=1}^n \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i), \qquad \text{with } \alpha_i \in \mathbb{R}$$

# Kernels

## Specify $K$ rather than $\phi$

**The kernel trick**

$$f^\star(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i)\,, f^\star \in \mathcal{H} \qquad \rightarrow \qquad K(\boldsymbol{x}, \boldsymbol{x}') = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle_{\mathcal{H}}$$

- $K$ as an implicit mapping of the feature space $\mathcal{H}$
- $K$ as a similarity measure of two observations in $\mathcal{H}$

## Some popular kernels

- Linear kernel $K(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{x}^T \boldsymbol{x}'$
- Polynomial kernel $K_d(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x}^T \boldsymbol{x}' + c)^d$
- Gaussian kernel $K_\sigma(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{\sigma^2})$
- ...

# Outline

Some machine learning tools

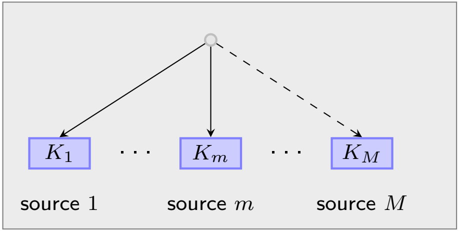
Composite Kernel Learning

BCI experiments

# Multiple Kernel Learning

[Lanckriet et al., 2004 ; Bach et al., 2004 ; Sonnenburg et al., 2006 ; ...]

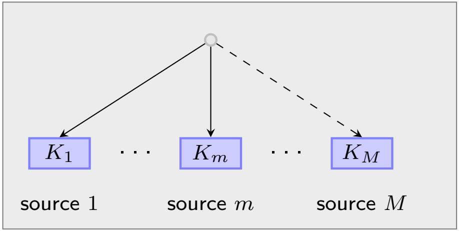
## Deal with several kernels



# Multiple Kernel Learning

[Lanckriet et al., 2004 ; Bach et al., 2004 ; Sonnenburg et al., 2006 ; ...]

## Deal with several kernels




## Use a convex combination

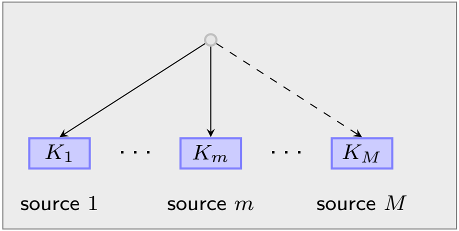
$$\bar{K}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{m=1}^{M} \sigma_m K_m(\boldsymbol{x}, \boldsymbol{x}')\,,$$

- $\sigma_m$, the coef. of the combination
- $K_m$, the r. k. of $\mathcal{H}_m$

# Multiple Kernel Learning

[Lanckriet et al., 2004 ; Bach et al., 2004 ; Sonnenburg et al., 2006 ; ...]

## Deal with several kernels




## Use a convex combination

$$\bar{K}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{m=1}^{M} \sigma_m K_m(\boldsymbol{x}, \boldsymbol{x}'),$$

- $\sigma_m$, the coef. of the combination
- $K_m$, the r. k. of $\mathcal{H}_m$

## Select the most relevant kernels

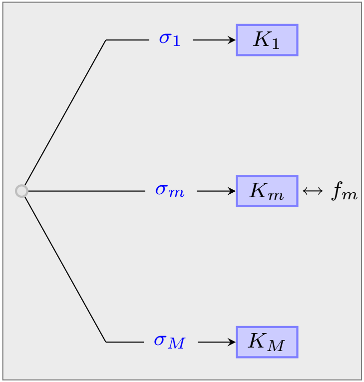
## Add a $\ell_1$ constraint

$$\sum_{m=1}^{M} \sigma_m = 1 \;, \sigma_m \geq 0 \;, \forall m$$

# Multiple Kernel Learning

## From MKL...

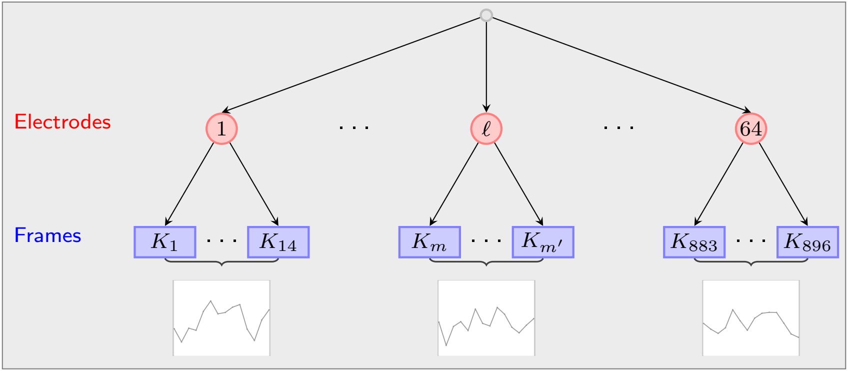
[Rakotomamonjy et al., 2007]




$$\begin{cases} \displaystyle\min_{\{f_m\},b,\boldsymbol{\sigma}} \quad \frac{1}{2}\sum_m \frac{\|f_m\|^2_{\mathcal{H}_m}}{\sigma_m} + C\sum_{i,m} h(\boldsymbol{x}_i, f_m(\boldsymbol{x}_i), b) \\ \text{s. t.} \quad \displaystyle\sum_m \sigma_m \leq 1 \ , \ \ \sigma_m \geq 0 \end{cases}$$

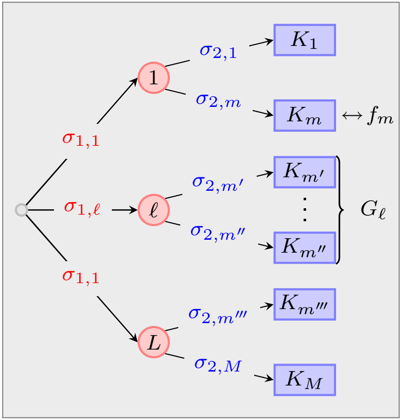
# From Multiple Kernel Learning to Composite Kernel Learning

## Work with an organization among kernels



# From Multiple Kernel Learning to Composite Kernel Learning

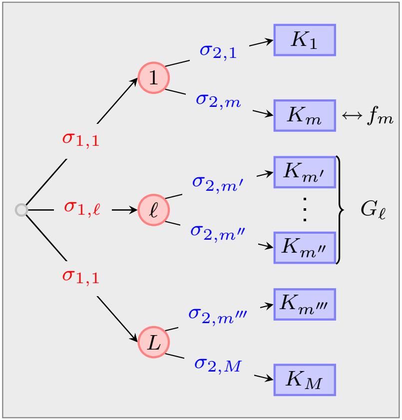
## . . . to CKL




$$\begin{cases} \min\limits_{\{f_m\},b,\boldsymbol{\sigma}} \quad \dfrac{1}{2}\sum\limits_m \dfrac{\|f_m\|^2_{\mathcal{H}_m}}{\sigma_m} \\ \qquad\qquad + C \sum\limits_{i,m} h(y_i, f_m(\boldsymbol{x}_i), b) \\ \text{s. t.} \end{cases}$$

# From Multiple Kernel Learning to Composite Kernel Learning

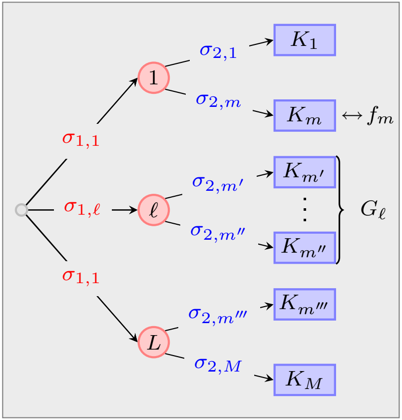
## . . . to CKL




$$
\left\{
\begin{array}{ll}
\displaystyle \min_{\{f_m\},b,\boldsymbol{\sigma}} & \displaystyle \frac{1}{2}\sum_m \frac{||f_m||^2_{\mathcal{H}_m}}{\sigma_m} \\
 & \displaystyle \qquad\qquad + C\sum_{i,m} h(y_i, f_m(\boldsymbol{x}_i), b) \\
\text{s. t.} & \displaystyle \sum_m \sigma_m \leq 1 \;,\; \sigma_m \geq 0
\end{array}
\right.
$$

# From Multiple Kernel Learning to Composite Kernel Learning

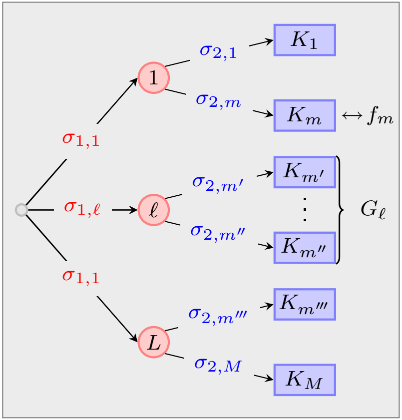
## . . . to CKL




$$
\begin{cases}
\displaystyle\min_{\{f_m\},b,\boldsymbol{\sigma}} & \displaystyle\frac{1}{2}\sum_m \frac{||f_m||^2_{\mathcal{H}_m}}{\sigma_m} + C\sum_{i,m} h(y_i, f_m(\boldsymbol{x}_i), b) \\
\text{s. t.} & \sigma_m = \sigma_{1,\ell}^p\,\sigma_{2,m}^q \\
 & \displaystyle\sum_\ell \sigma_{1,\ell} \le 1 \;,\; \sigma_{1,\ell} \ge 0 \\
 & \displaystyle\sum_m \sigma_{2,m} \le 1 \;,\; \sigma_{2,m} \ge 0
\end{cases}
$$

# From Multiple Kernel Learning to Composite Kernel Learning

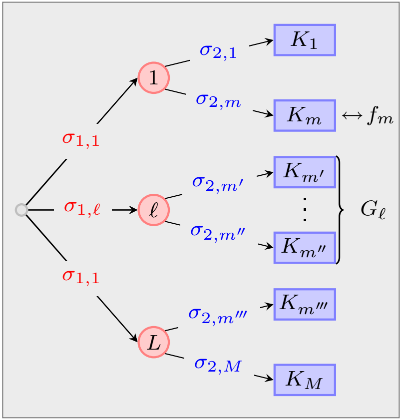
## ... to CKL




$$\left\{\begin{array}{ll} \min\limits_{\{f_m\},b,\boldsymbol{\sigma}} & \dfrac{1}{2}\sum\limits_m \dfrac{||f_m||^2_{\mathcal{H}_m}}{\sigma_m} \\ & \quad + C\sum\limits_{i,m} h(y_i, f_m(\boldsymbol{x}_i), b) \\ \text{s. t.} & \sigma_m = \sigma_{1,\ell}^p\,\sigma_{2,m}^q \\ & \sum\limits_\ell \sigma_{1,\ell} \leq 1 \;,\; \sigma_{1,\ell} \geq 0 \\ & \sum\limits_m \sigma_{2,m} \leq 1 \;,\; \sigma_{2,m} \geq 0 \end{array}\right.$$

# From Multiple Kernel Learning to Composite Kernel Learning

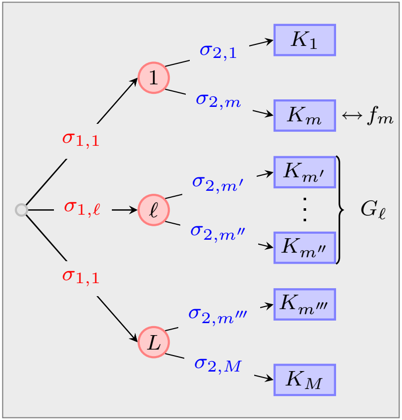
## . . . to CKL




$$\left\{ \begin{array}{ll} \min\limits_{\{f_m\},b,\boldsymbol{\sigma}} & \displaystyle \frac{1}{2} \sum_m \frac{||f_m||^2_{\mathcal{H}_m}}{\sigma_m} \\ & \displaystyle \qquad + C \sum_{i,m} h(y_i, f_m(\boldsymbol{x}_i), b) \\ \text{s. t.} & \sigma_m = \sigma_{1,\ell}^p \, \sigma_{2,m}^q \\ & \displaystyle \sum_\ell \sigma_{1,\ell} \le 1 \ , \quad \sigma_{1,\ell} \ge 0 \\ & \displaystyle \sum_m \sigma_{2,m} \le 1 \ , \quad \sigma_{2,m} \ge 0 \end{array} \right.$$

# From Multiple Kernel Learning to Composite Kernel Learning

## . . . to CKL




$$
\begin{cases}
\displaystyle\min_{\{f_m\},b,\boldsymbol{\sigma}} & \displaystyle\frac{1}{2}\sum_m \frac{\|f_m\|^2_{\mathcal{H}_m}}{\sigma_m} + C\sum_{i,m} h(y_i, f_m(\boldsymbol{x}_i), b) \\
\text{s. t.} & \sigma_m = \sigma_{1,\ell}^{p}\,\sigma_{2,m}^{q} \\
& \displaystyle\sum_\ell \left(\left(\sum_{m\in G_\ell} \sigma_m^{\frac{1}{q}}\right)^{q}\right)^{\frac{1}{p+q}} \leq 1 \\
& \sigma_m \geq 0
\end{cases}
$$

# Two equivalent formulations

## Expression in $\|f_m\|_{\mathcal{H}_m}$

$$\begin{cases} \displaystyle\min_{\{f_m\},\boldsymbol{\sigma}} & \displaystyle\frac{1}{2}\sum_m \frac{1}{\sigma_m}||f_m||^2_{\mathcal{H}_m} \\ \text{s. t.} & \displaystyle\sum_\ell \left(\left(\sum_{m\in G_\ell} \sigma_m^{1/q}\right)^q\right)^{1/(p+q)} \leq 1 \ , \ \sigma_m \geq 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} \displaystyle\min_{\{f_m\}} & \displaystyle\frac{1}{2}\left(\sum_\ell \left(\sum_{m\in G_\ell} ||f_m||^s_{\mathcal{H}_m}\right)^{r/s}\right)^{2/r} \\ \text{with} & \displaystyle s = \frac{2}{q+1} \quad \text{and} \quad r = \frac{2}{p+q+1} \end{cases}$$

# Two equivalent formulations

## Expression in $\|f_m\|_{\mathcal{H}_m}$

$$\Leftrightarrow \begin{cases} \min\limits_{\{f_m\}} & \dfrac{1}{2}\left(\sum\limits_{\ell}\left(\sum\limits_{m\in G_\ell}\|f_m\|^{s}_{\mathcal{H}_m}\right)^{r/s}\right)^{2/r} \\ \text{with} & s = \dfrac{2}{q+1} \quad \text{and} \quad r = \dfrac{2}{p+q+1} \end{cases}$$

## Interpretation

$$g_\ell = \left(\sum_{m\in G_\ell}\|f_m\|^{s}_{\mathcal{H}_m}\right)^{1/s}$$

Sparse in kernels if $s \leq 1$

$$\ell_{(r,s)} = \left(\sum_{\ell=1}^{L} g_\ell^{r}\right)^{1/r}$$

Sparse in sources if $r \leq 1$

# Two equivalent formulations

## Expression in $\|f_m\|_{\mathcal{H}_m}$

$$\Leftrightarrow \begin{cases} \min\limits_{\{f_m\}} & \frac{1}{2}\left(\sum\limits_{\ell}\left(\sum\limits_{m\in G_\ell}\|f_m\|^{s}_{\mathcal{H}_m}\right)^{r/s}\right)^{2/r} \\ \text{with} & s = \frac{2}{q+1} \quad \text{and} \quad r = \frac{2}{p+q+1} \end{cases}$$

## Some particular cases

- Lasso [Tibshirani, 1996] $p = 0$ $q = 1$ $\ell_1$
- Group-lasso [Yuan & Lin, 2006] $p = 1$ $q = 0$ $\ell_{(1,2)}$

# Two equivalent formulations

## Expression in $\|f_m\|_{\mathcal{H}_m}$

$$\Leftrightarrow \begin{cases} \min\limits_{\{f_m\}} & \dfrac{1}{2}\left(\sum\limits_{\ell}\left(\sum\limits_{m\in G_\ell}\|f_m\|^{s}_{\mathcal{H}_m}\right)^{r/s}\right)^{2/r} \\ \text{with} & s = \dfrac{2}{q+1} \quad \text{and} \quad r = \dfrac{2}{p+q+1} \end{cases}$$

## Convexity is not compatible with sparsity on each levels

- Convex if $s \geq$ and $r \geq 1$
- Sparse if $s \leq$ and $r \leq 1$

# Algorithm

## A wrapper approach

- Iterative algorithm — based on [Rakotomamonjy et coll., 2007]
- Alternate optimisation of two problems

1. Optimize the parameters of the SVM, considering fixed coefficients $\{\sigma_m\}$

$$\left\{ \min_{\{f_m, b\}} \quad J(\boldsymbol{\sigma}) = \frac{1}{2} \sum_m \frac{1}{\sigma_m} \|f_m\|^2_{\mathcal{H}_m} + C \sum_{i,m} h(y_i, f(\boldsymbol{x}_i), b) \right.$$

# Algorithm

## A wrapper approach

- Iterative algorithm based on [Rakotomamonjy et coll., 2007]
- Alternate optimisation of two problems

1. Optimize the parameters of the SVM, considering fixed coefficients $\{\sigma_m\}$

$$\left\{ \min_{\{f_m\},b} \quad J(\boldsymbol{\sigma}) = \frac{1}{2} \sum_m \frac{1}{\sigma_m} \|f_m\|^2_{\mathcal{H}_m} + C \sum_{i,m} h(y_i, f(\boldsymbol{x}_i), b) \right.$$

# Algorithm

## A wrapper approach

- Iterative algorithm — based on [Rakotomamonjy et coll., 2007]
- Alternate optimisation of two problems

2. Optimize the coefficients $\{\sigma_m\}$, with parameters of the SVM fixed at step 1

$$\begin{cases} \min\limits_{\{\sigma_m\}} & J(\boldsymbol{\sigma}) \\ \text{s. t.} & \sum\limits_{\ell} \left( \left( \sum\limits_{m} \sigma_m^{1/q} \right)^q \right)^{1/(p+q)} \leq 1 \qquad \sigma_m \geq 0 \qquad \forall m \end{cases}$$

where $J(\boldsymbol{\sigma})$ is the value of the objectif function of the SVM

# Outline

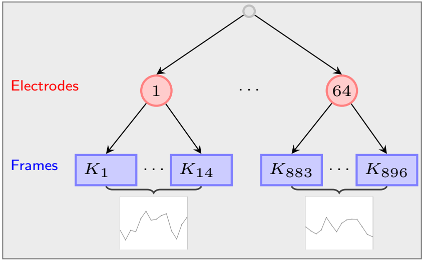
Some machine learning tools

Composite Kernel Learning

**BCI experiments**

# Reminder of the problem

## Signals acquisition




- Signals are collected over 64 electrodes
  64 sources
- Each signal is sampled on 14 frames
  64 sources $\times$ 14 measures
- P300 occurs around $\simeq$ 300 ms
  frame 7

## Goal

1. Discriminate the signals that contains a P300
2. Identify the significant electrodes and frames

# Protocol

## Dataset

- 64 electrodes × 14 frames — 896 linear kernels
- Protocol applied on 10 datasets — 10 draws with replacements

## Methods

- SVM $\bar{K} = \sum_{m=1}^{896} 1/m \, K_m$ — **No organization**

  Not sparse
- MKL $\bar{K} = \sum_{m=1}^{896} \sigma_m \, K_m$ — **No organization**

  Sparse in kernels
- CKL $\bar{K} = \sum_{\ell=1}^{64} \sum_{m \in G_\ell} \sigma_{1,\ell} \, \sigma_{2,m} \, K_m$ — **Organization**

  Sparse in electrodes and frames

# Performances and sparsity

## Averaged results

**10 repetitions**

| Algorithms | AUC | # Sources [1] | # Kernels [1] |
|---|---|---|---|
| SVM | $84.6 \pm 0.9$ | $64$ | $896$ |
| MKL | $85.7 \pm 0.9$ | $47.0 \pm 7.9$ | $112.6 \pm 46.2$ |
| CKL | $84.7 \pm 1.1$ | $14.6 \pm 13.1$ | $65.8 \pm 52.2$ |

- Similar performances for the three methods
- Sparsity in sources
  CKL removes about three quarters of the sources
  MKL keeps about three quarters of them
- Sparsity in kernels
  CKL is twice more sparse than MKL

1. Number of sources and kernels involved in the decision function.

# Performances and sparsity

## Averaged results

**10 repetitions**

| Algorithms | AUC | # Sources [1] | # Kernels [1] |
|---|---|---|---|
| SVM | $84.6 \pm 0.9$ | $64$ | $896$ |
| MKL | $85.7 \pm 0.9$ | $47.0 \pm 7.9$ | $112.6 \pm 46.2$ |
| CKL | $84.7 \pm 1.1$ | $14.6 \pm 13.1$ | $65.8 \pm 52.2$ |

- Similar performances for the three methods
- Sparsity in sources
  CKL removes about three quarters of the sources
  MKL keeps about three quarters of them
- Sparsity in kernels
  CKL is twice more sparse than MKL

1. Number of sources and kernels involved in the decision function.

# Performances and sparsity

## Averaged results

10 repetitions

| Algorithms | AUC | # Sources [1] | # Kernels [1] |
|---|---|---|---|
| SVM | $84.6 \pm 0.9$ | $64$ | $896$ |
| MKL | $85.7 \pm 0.9$ | $47.0 \pm 7.9$ | $112.6 \pm 46.2$ |
| CKL | $84.7 \pm 1.1$ | $14.6 \pm 13.1$ | $65.8 \pm 52.2$ |

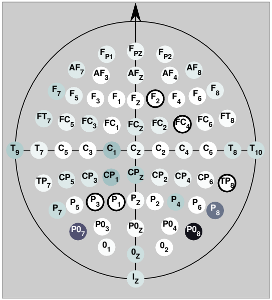
- Similar performances for the three methods
- Sparsity in sources
  CKL removes about three quarters of the sources
  MKL keeps about three quarters of them
- Sparsity in kernels
  CKL is twice more sparse than MKL

1. Number of sources and kernels involved in the decision function.

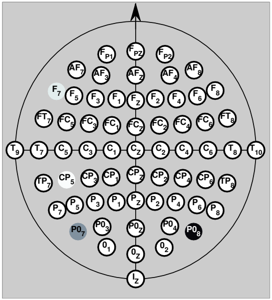
# Electrodes relevance

## Median results

10 repetitions



MKL



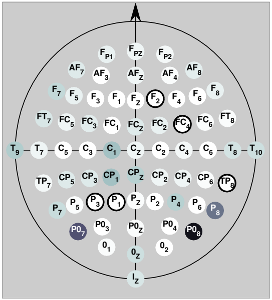
CKL

- The darker the color, the higher the relevance
- Electrodes in white with a black circle are discarded (the relevance is exactly zero)

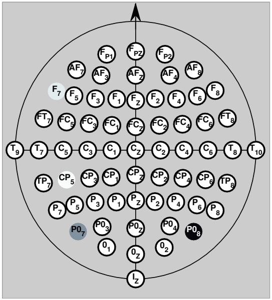
# Electrodes relevance

## Median results

10 repetitions




MKL



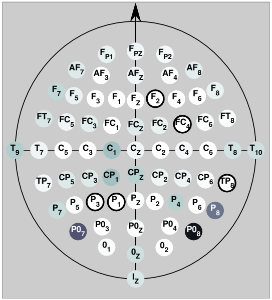

CKL

- CKL – High relevances for the electrodes in the areas of the visual cortex (lateral electrodes $P0_7$ and $P0_8$)

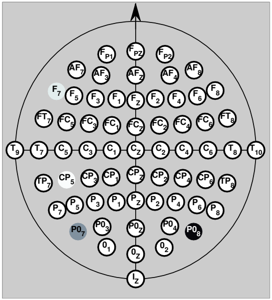
# Electrodes relevance

## Median results

10 repetitions



MKL



CKL

- MKL – Similar behaviour, but also highlights numerous frontal electrodes that are not likely to be relevant for the BCI P300 Speller paradigm

# Electrodes relevance

**Temporal evolution** **One specific repetition**

### MKL

- A global activity over all electrodes
- A more important activity from frame 7

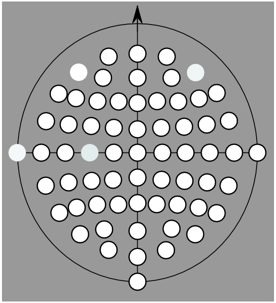
### CKL

- No or few electrodes selected in the first and the last frames
- More electrodes are discarded

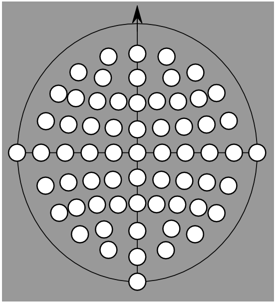
# Electrodes relevance

## Temporal evolution

Let's go



MKL



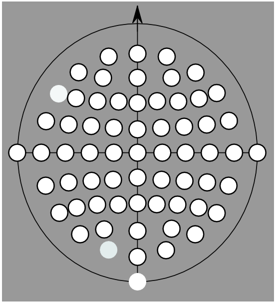
CKL

frame 1

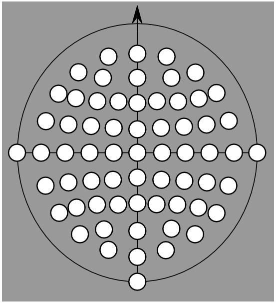
# Electrodes relevance

## Temporal evolution

Let's go



MKL



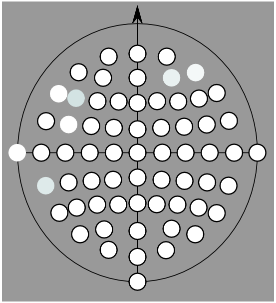
CKL

frame 2

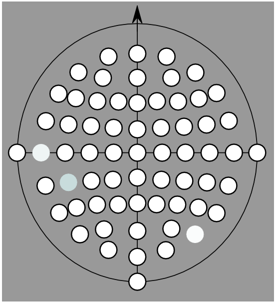
# Electrodes relevance

## Temporal evolution

Let's go



MKL



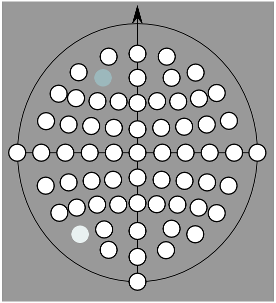
CKL

frame 3

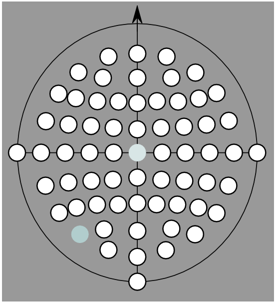
# Electrodes relevance

## Temporal evolution

Let's go



MKL



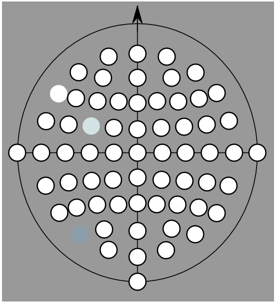
CKL

frame 4

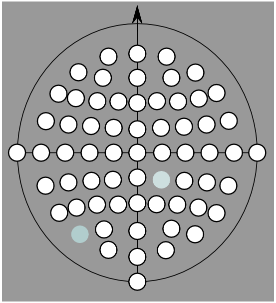
# Electrodes relevance

## Temporal evolution

Let's go



MKL



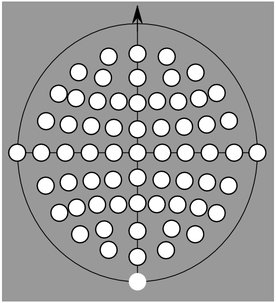
CKL

frame 5

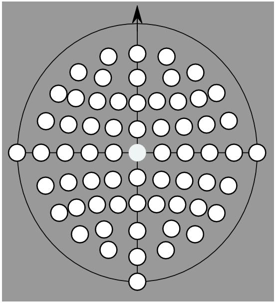
# Electrodes relevance

## Temporal evolution

Let's go



MKL



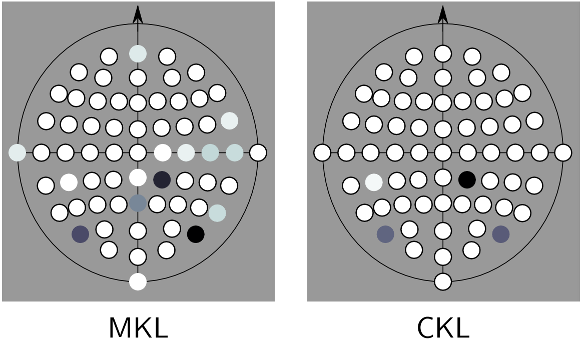
CKL

frame 6

# Electrodes relevance

## Temporal evolution

Let's go

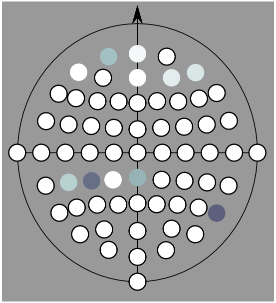



frame 7

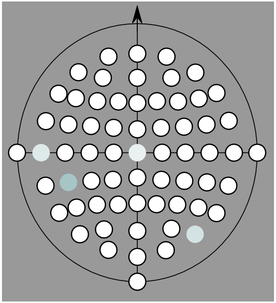
# Electrodes relevance

## Temporal evolution

Let's go



MKL



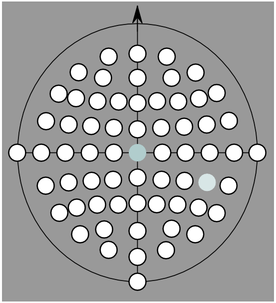
CKL

frame 8

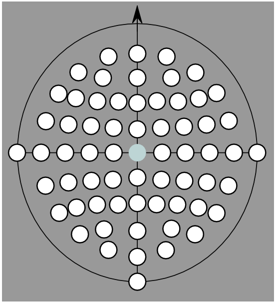
# Electrodes relevance

## Temporal evolution

Let's go



MKL



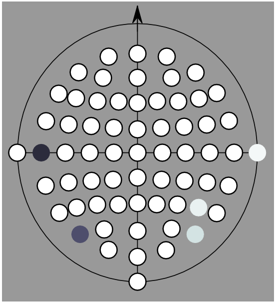
CKL

frame 9

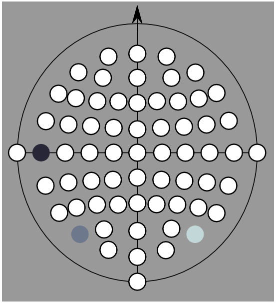
# Electrodes relevance

## Temporal evolution

Let's go



MKL



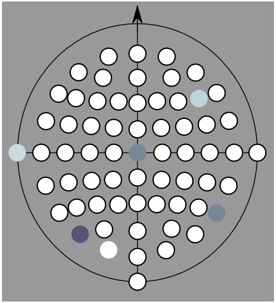
CKL

frame 10

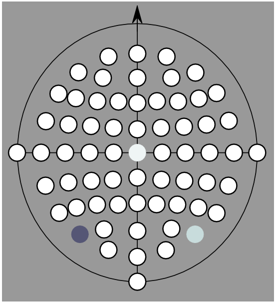
# Electrodes relevance

## Temporal evolution

Let's go



MKL



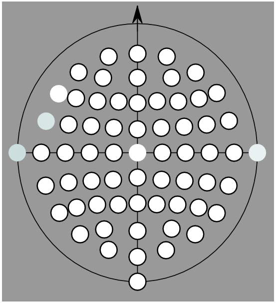
CKL

frame 11

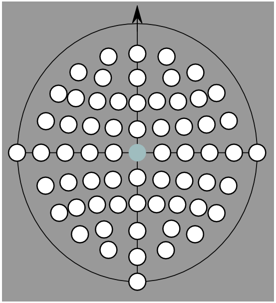
# Electrodes relevance

## Temporal evolution

Let's go



MKL



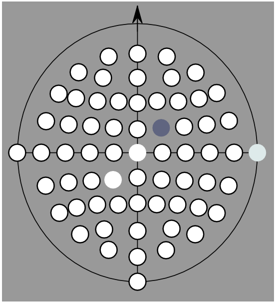
CKL

frame 12

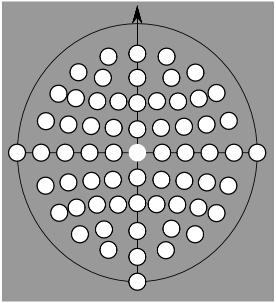
# Electrodes relevance

## Temporal evolution

Let's go



MKL



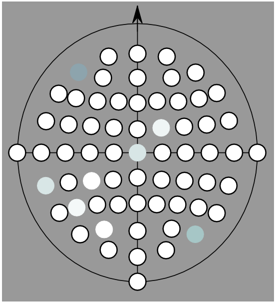
CKL

frame 13

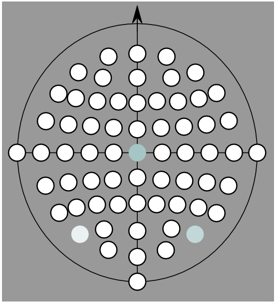
# Electrodes relevance

## Temporal evolution

Let's go



MKL



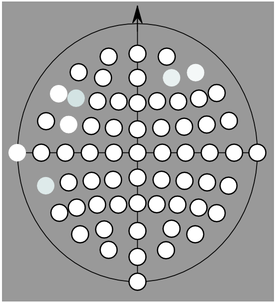
CKL

frame 14

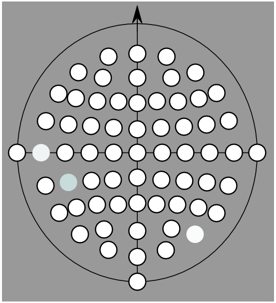
# Electrodes relevance

## Temporal evolution

### Once more, with comments



MKL



CKL

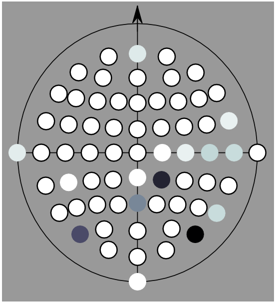
frames 1..6

A low activity at the beginning

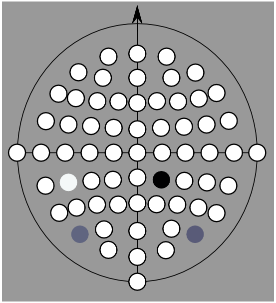
# Electrodes relevance

## Temporal evolution

Once more, with comments



MKL



CKL

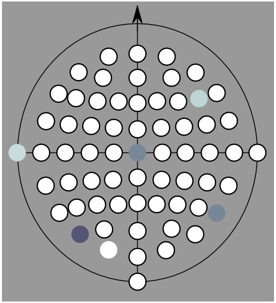
frame 7

Things happen here

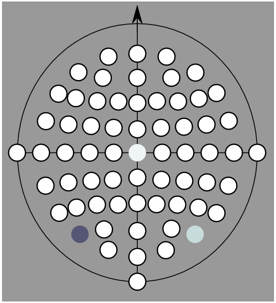
# Electrodes relevance

## Temporal evolution

Once more, with comments



MKL



CKL

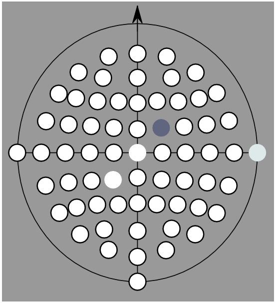
frames 8..12

The activity starts to decrease

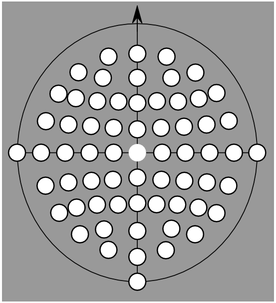
# Electrodes relevance

## Temporal evolution

Once more, with comments



MKL



CKL

frames 13..14

And almost stops

# Conclusion

## Model

- Take into account an organization among kernels
- Identify the significant sources and kernels within the organization

## Going further

- Formalize the model when kernels belong to more than one source

  for instance [Jacob et al., 2009, Jenatton et al., 2010]

- Extend the model for an arbitrary number of levels

  for instance [Yuan and Lin, 2010]

# Conclusion

## Model

- Take into account an organization among kernels
- Identify the significant sources and kernels within the organization

## Going further

- Formalize the model when kernels belong to more than one source

  for instance [Jacob et al., 2009, Jenatton et al., 2010]

- Extend the model for an arbitrary number of levels

  for instance [Yuan and Lin, 2010]

Questions ?