

# Approximation structurée appliquée à l'analyse des signaux M/EEG

Matthieu Kowalski

L2S - Univ. Paris-Sud

En collaboration avec Alexandre Gramfort

10 septembre 2010

- 1 Introduction : M/EEG et problèmes inverses
- 2 Approche variationnelle
- 3 Optimisation convexe
- 4 Résultats
- 5 Conclusion

- 1 Introduction : M/EEG et problèmes inverses
- 2 Approche variationnelle
- 3 Optimisation convexe
- 4 Résultats
- 5 Conclusion

# Problèmes inverses

## Exemples

séparations de sources audio, tomographie, localisation de sources M/EEG

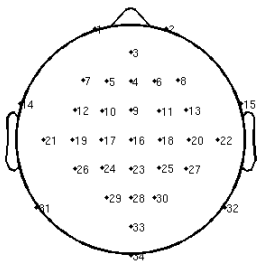
## Problèmes

on cherche à reconstruire des signaux à partir d'observations incomplètes et/ou bruitées

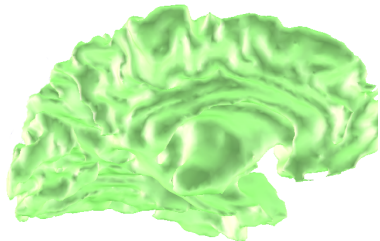
## modélisation en deux étapes

- 1 Trouver un bon modèle pour le problème direct, qui permet de lier les signaux d'entrées aux sorties mesurées ;
- 2 Inverser le modèle précédent, c'est à dire estimer les signaux d'entrées à partir des seules mesures à disposition.

# Problème inverse en MEEG



? ↓?



À partir d'un enregistrement M/EEG sur le scalp comment localiser les sources dans le cerveau ?

enregistrements :  $M \in \mathbb{R}^{N \times T}$

Matrice de "diffusion" connue :  $G \in \mathbb{R}^{N \times I}$

sources :  $X \in \mathbb{R}^{I \times T}$

$$M = GX + b$$

avec  $N \approx 150$  et  $I \approx 18000$

Problème très sous-déterminé : besoin d'*a priori* très forts sur  $X$

- 1 Introduction : M/EEG et problèmes inverses
- 2 Approche variationnelle**
- 3 Optimisation convexe
- 4 Résultats
- 5 Conclusion

# Généralités

On cherche une estimation  $\hat{X}$  comme optimum d'une fonction coût :

$$\hat{X} = \operatorname{argmin} f_1(X) + \lambda f_2(X),$$

avec

- Un terme d'attache aux données,  $f_1$ .
- Un terme de régularisation,  $f_2$ , ou *a priori*.
- Un paramètre  $\lambda > 0$ .

## Exemple d'attache aux données

$$f_1(X) = \frac{1}{2} \|M - GX\|_2^2$$

## Exemple de régularisation

$$f_2(X) = \|X\|_2^2 = \sum_{i,t} |X_{i,t}|^2$$

# Localisation de sources M/EEG

Modèle de mélange :

$$M = GX + B$$

avec

- $K$  conditions expérimentales
- $N$  Mesures sur le scalp de longueur  $T$  :  $M \in \mathbb{R}^{N \times KT}$
- Système de diffusion  $G \in \mathbb{R}^{N \times I}$  connu
- $I$  Sources  $X \in \mathbb{R}^{I \times KT}$
- Bruit  $B$

## Difficulté

Le système est *très* sous-déterminé : il faut retrouver plusieurs milliers de sources à partir d'une centaine de mesures sur le scalp.

Solution d'énergie minimale :

$$X^* = \underset{X}{\operatorname{argmin}} \|M - GX\|_F^2 + \lambda \|X\|_{\mathbf{w};F}^2, \lambda \in \mathbb{R}_+,$$



# Autres *a priori*

## Normes $\ell_p$

$$\|X\|_p^p = \sum_{i,t} |X_{i,t}|^p$$

En particulier :

- $p = 2$  vue précédemment.
- $p = 1$  pour introduire de la parcimonie. **Non adapté ici !**

## Les normes $\ell_{p,q}$

$$\|X\|_{p,q}^q = \sum_g \left( \sum_m |X_{g,m}|^p \right)^{q/p}$$

## Autres *a priori* : normes mixtes

Deux normes mixtes particulières :

- Le Group-Lasso  $\|X\|_{2,1} = \sum_i \sqrt{\sum_t |X_{i,t}|^2}$
- L'Elitist-Lasso  $\|X\|_{1,2}^2 = \sum_i (\sum_t |X_{i,t}|)^2$

[Ou *et al.*]

$$\hat{X} = \operatorname{argmin} \|M - GX\|_2^2 + \lambda \sum_i \sqrt{\sum_t |X_{i,t}|^2}$$

# Vers une autre sélection des groupes

“deux expériences ne peuvent avoir lieu en même temps”

## Définition du prior

$X$  : a des éléments indexés par un index *triple* :  $x_{i,k,t}$   
 $i$  indice l'espace,  $k$  la condition et  $t$  le temps.

$$\|X\|_{\underline{w};212}^2 = \sum_i \left( \sum_k \left( \sum_t w_{i,k,t} |x_{i,k,t}|^2 \right)^{1/2} \right)^2 .$$

où  $w_{i,k,t} > 0$  sont des poids.

## Application M/EEG : Fonctionnelle à optimiser

$$\hat{X} = \underset{X}{\operatorname{argmin}} \|M - GX\|_F^2 + \lambda \|X\|_{\underline{w};212}^2, \lambda \in \mathbb{R}_+,$$

- 1 Introduction : M/EEG et problèmes inverses
- 2 Approche variationnelle
- 3 Optimisation convexe**
- 4 Résultats
- 5 Conclusion

# Opérateur de proximité

## Définition

Soit  $\phi : \mathbb{R}^P \rightarrow \mathbb{R}$  une fonction convexe semie-continue inférieurement. L'opérateur de proximité associé à  $\phi$  et  $\lambda \in \mathbb{R}_+$ , noté  $\text{prox}_{\lambda\phi} : \mathbb{R}^P \rightarrow \mathbb{R}^P$  est donné par

$$\text{prox}_{\lambda\phi}(\mathbf{y}) = \underset{\mathbf{x} \in \mathbb{R}^P}{\text{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\phi(\mathbf{x}) .$$

## FISTA

## Algorithme

Initialisation :  $X^{(0)} \in \mathbb{R}^{I \times KT}$ ,  $Z^{(1)} = X^{(0)}$ ,  $\tau^{(1)} = 1$ ,  $k = 1$ ,  $\mu = 1/\|G^*G\|$   
 Répéter

$$X^{(k)} = \text{prox}_{\mu\lambda f_2} \left( Z^{(k)} + \mu G^*(M - GZ^{(k)}) \right)$$

$$\tau^{(k+1)} = \frac{1 + \sqrt{1 + 4\tau^{(k)^2}}}{2}$$

$$Z^{(k+1)} = X^{(k)} + \frac{\tau^{(k)} - 1}{\tau^{(k)}} (X^{(k)} - X^{(k-1)})$$

## Prox de la norme $\ell_{212}$

Soit  $\mathbf{y} \in \mathbb{R}^{IKT}$  indicé par un triple indice  $(i, k, t)$ . Soit  $\mathbf{w}$  une séquence de poids strictement positifs tels que  $\forall t, w_{i,k,t} = w_{i,k}$ . Pour chaque  $i$ , soit  $w_{i,k'_i}, [y_{i,k'_i}] = \sqrt{w_{i,k'_i} \sum_t |y_{i,k'_i,t}|^2}$  et  $r_{i,k'_i} = [y_{i,k'_i}]/w_{i,k'_i}$  ordonnés tels que, pour un  $i$  fixé,  $\forall k'_i, r_{i,k'_i+1} \leq r_{i,k'_i}$ . Alors,  $\mathbf{x} = \text{prox}_{\lambda \|\cdot\|_{\mathbf{w};212}}(\mathbf{y})$  est donné coordonnée par coordonnée  $(i, k, t)$  par

$$x_{i,k,t} = y_{i,k,t} \left( 1 - \frac{\lambda \sqrt{w_{i,k}} \sum_{k'_i=1}^{K_i} [y_{i,k'_i}]}{1 + K_{\mathbf{w}_i} \lambda \|\mathbf{y}_{i,k}\|_{\mathbf{w}_i;2}} \right)^+,$$

avec  $K_{\mathbf{w}_i} = \sum_{k'_i=1}^{K_i} w_{i,k'_i}^2$  et l'indice  $K_i$  est le nombre tel que

$$\lambda \sum_{k'_i=1}^{K_i} w_{i,k'_i}^2 (r_{i,k'_i} - r_{i,K_i}) < r_{i,K_i} \leq \lambda \sum_{k'_i=1}^{K_i+1} w_{i,k'_i}^2 (r_{i,k'_i} - r_{i,K_i}),$$

et  $\mathbf{y}_{i,k}$  la sous-suite issue de  $\mathbf{y}$  pour  $(i, k)$  fixé.

# Critère d'arrêt

On voit souvent des critères du type :

$$\frac{\|x^{(k)} - x^{(k-1)}\|}{\|x^{(k)}\|} < \varepsilon .$$

Inconvénients :

- Cette quantité peut vite devenir très petite alors qu'on est encore très loin du minimiseur.
- On trouve de plus qu'en pratique, le choix d'un « bon »  $\varepsilon$  dépend fortement du paramètre  $\lambda$  de la fonctionnelle.

Un “bon” critère est le saut de dualité :

$$\eta^{(k)} = \mathcal{F}_p(X^{(k)}) - \mathcal{F}_d(Y^{(k)}) \geq 0 .$$

où  $\mathcal{F}_d$  est une fonction duale associée à la fonction à optimiser, et  $Y^k$  une variable dual associée à  $X^k$  tel qu'à l'optimum le saut de dualité soit nul.



# Dualité de Fenchel

## Conjugué de Fenchel

Soit  $f$  une fonction convexe. Le conjugué de fenchel de  $f$ , notée  $f^*$  est définie par :

$$f^*(v) = \sup_u \langle v, u \rangle - f(u) .$$

$v$  s'appelle la variable conjuguée.

## Norme duale

Soit  $\|\cdot\|$  une norme. On appelle norme duale la norme notée  $\|\cdot\|^*$  définie par

$$\|u\|^* = \sup_v \{u^T v \mid \|v\| = 1\} .$$

## Norme duale d'une norme mixte

Le duale de la norme  $\|\cdot\|_{p_1, \dots, p_n}$  est la norme  $\|\cdot\|_{p'_1 \dots p'_n}$  telle que , si  $p_i > 1$ ,  $\frac{1}{p_i} + \frac{1}{p'_i} = 1$ , et si  $p_i = 1$ ,  $p'_i = \infty$ .

# Conjugué de Fenchel

## Conjugué de Fenchel d'une norme mixte

- ① Le conjugué de Fenchel d'une norme est l'indicatrice de la norme conjuguée. Ainsi, le dual de la fonction  $u \mapsto \|u\|_{p_1, \dots, p_n}$  est la fonction

$$v \mapsto \chi_{\|v\|_{p'_1, \dots, p'_n}^*} = \begin{cases} 1 & \text{si } \|v\|_{p'_1, \dots, p'_n} \leq 1 \\ 0 & \text{sinon.} \end{cases}$$

où  $\|v\|_{p'_1, \dots, p'_n}$  est la norme mixte conjuguée.

- ② Le conjugué de Fenchel de la fonction  $u \mapsto \frac{1}{2} \|u\|_{p_1, \dots, p_n}^2$  est la fonction

$$v \mapsto \frac{1}{2} \|v\|_{p'_1, \dots, p'_n}^2.$$

# Dualité de Fenchel

## Théorème de dualité

Soit  $f$  une fonction convexe de  $\mathbb{R}^m$  et  $g$  une fonction concave de  $\mathbb{R}^n$ . Soit  $G$  un opérateur linéaire de  $\mathbb{R}^m$  dans  $\mathbb{R}^n$ , alors

$$\inf_{x \in \mathbb{R}^m} \{f(x) - g(Gx)\} = \sup_{u \in \mathbb{R}^n} \{f^*(G^*u) - g^*(u)\},$$

où  $f^*$  (resp.  $g^*$ ) est le conjugué de Fenchel de  $f$  (resp.  $g$ ), et  $G^*$  l'opérateur adjoint de  $G$ .

De plus, les conditions de Karush-Khun-Tucker (KKT) sont équivalentes à

$$\begin{aligned} f(x) + f^*(G^*u) &= \langle x, G^*u \rangle, \\ g(Gx) + g^*(u) &= \langle Gx, u \rangle. \end{aligned}$$

# Dualité de Fenchel

Fonction primale :

$$\mathcal{F}_p(X) = \|M - GX\|_F^2 + \lambda f_2(X)$$

Fonction duale :

$$\mathcal{F}_p d(Y) = -\frac{1}{2} \|Y\|_2^2 - \text{Tr}(Y^T M) + f_2^*(G^* Y)$$

avec

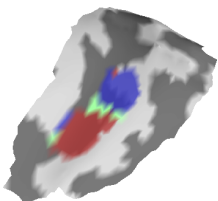
- $Y^{(k)} = GX^{(k)} - M$
- $f_2^*$  est le conjugué de Fenchel de  $\lambda f_2$

- 1 Introduction : M/EEG et problèmes inverses
- 2 Approche variationnelle
- 3 Optimisation convexe
- 4 Résultats**
- 5 Conclusion

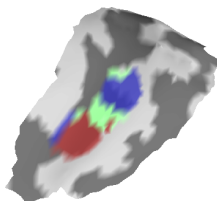
# Simulations



(a) Données simulées



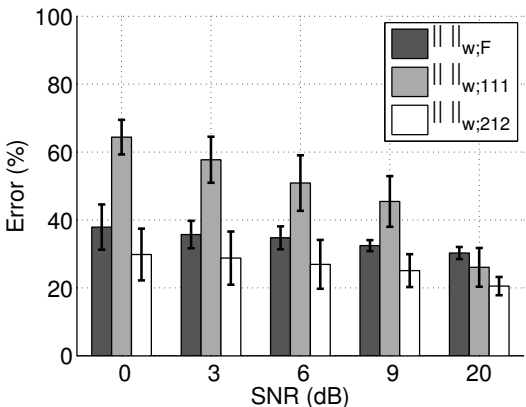
(b) Résultat avec  $\|\cdot\|_{\mathbf{w};F}$  (RI)



(c) Résultat avec  $\|\cdot\|_{\mathbf{w};212}$  (RI)

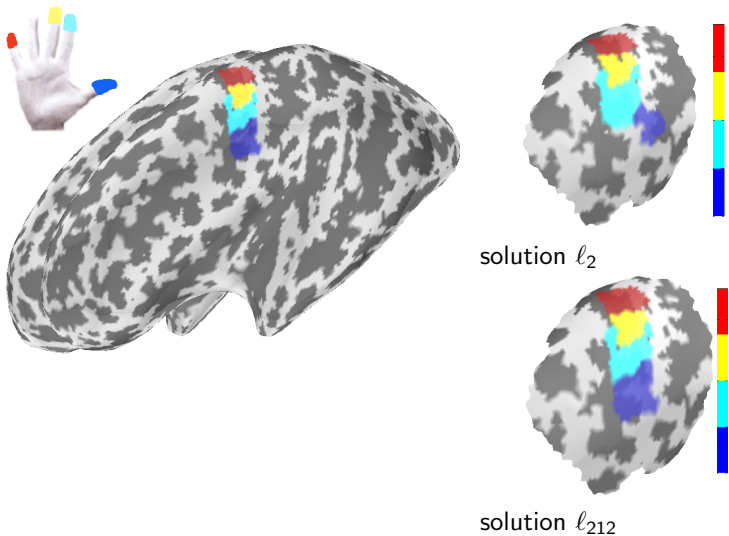
Illustration des résultats sur le cortex sensoriel primaire (S1) (SNR = 20dB). Les régions actives voisines reproduisent l'organisation de S1.

# Simulations



Évaluation des estimées par  $\|\cdot\|_{w;F}$  vs.  $\|\cdot\|_{w;212}$  vs.  $\|\cdot\|_{w;111}$  sur des données somesthésiques synthétiques. L'erreur représente le pourcentage des sources mal identifiés.

# Données MEG





# Conclusion

## A retenir

- *A priori* par une norme mixte sur trois niveaux
- Utilisation des formes variationnelles
- Cadre de l'optimisation convexe

## A venir

- Utilisation de dictionnaires temps-fréquence
- Construction de nouveaux *a priori* adaptés au temps-fréquence
- Accélération des techniques d'optimisation (active set)