

# Décompositions parcimonieuses et persistantes de signaux multicanaux. Applications aux signaux MEEG.

Matthieu Kowalski, Bruno Torrèsani

Laboratoire d'Analyse, Topologie et Probabilités, Université de Provence  
Centre de Mathématique et Informatique,  
39 rue Joliot-Curie, 13453, Marseille Cedex 13, France  
{kowalski, torresani}@cmi.univ-mrs.fr

**Résumé** : On considère le problème de la régression parcimonieuse de signaux multicanaux sur des repères temps-fréquences. Les signaux multicanaux sont décomposés sur un unique repère en utilisant des coefficients vectoriels (i.e. multicanal). La parcimonie apparaît habituellement grâce à des approches type "basis pursuit denoising" : la régression s'effectue en minimisant une fonctionnel qui fait intervenir un terme d'attache aux données  $\ell_2$ , et une pénalité  $\ell_1$  sur les coefficients.

Dans cette contribution, cette dernière est remplacée par une norme mixte, qui favorise la parcimonie à l'intérieur d'un canal et la persistance au travers les différents canaux (i.e. un coefficient donné est considéré "actif" dans un groupe de canaux assez large). On montre que l'optimisation des fonctionnels correspondants se résout à l'aide d'algorithmes de seuillage généralisé itératif.

On présente en plus des adaptations pour l'analyse et le débruitage de signaux EEG et MEG, où l'information topographique des canaux est prise en compte.

## 1 Introduction

Les approches variationnelles du problème de régression parcimonieuse d'un signal dans un dictionnaire de forme d'ondes bénéficient actuellement d'un certain engouement. Une des raisons de cette popularité vient de la simplicité des algorithmes correspondants. Des résultats récents montrent que les algorithmes de seuillages et de seuillage itératif permettent souvent d'obtenir une approximation parcimonieuse d'un signal, sans demander une puissance de calcul démesurée, et dont la convergence est prouvée.

L'hypothèse de parcimonie pour la décomposition d'un signal dans un repère ou une base (noté par la suite  $\mathcal{F} = \{\varphi_\lambda, \lambda \in \Lambda\}$ ) d'un espace  $\mathcal{H}$ , est souvent traduite par la minimisation d'une norme  $\ell_p$  (avec  $p < 2$ ) sur la suite des coefficients, et un terme  $\ell_2$  d'attache aux données. Autrement dit, une décomposition parcimonieuse d'un signal  $x \in \mathcal{H}$  dans un repère  $\mathcal{F}$  de  $\mathcal{H}$  peut être obtenue en résolvant

$$\min \left[ \left\| x - \sum_{\lambda \in \Lambda} \alpha_{\lambda} \varphi_{\lambda} \right\|_{\mathcal{H}}^2 + \mu \|\alpha\|_p^p \right] \quad (1)$$

où  $\mu$  est un paramètre de Lagrange. Sous forme matricielle, ce problème s'écrit

$$\min \left[ \|x - \Phi\alpha\|_{\mathcal{H}}^2 + \mu \|\alpha\|_p^p \right] \quad (2)$$

où  $\Phi$  est la matrice dont les colonnes sont les vecteurs  $\varphi_{\lambda}$  du repère  $\mathcal{F}$ .

Bien que ce papier se place d'un point de vue "signal", les formulations variationnelles de régression telles que (2) sont identiques à celles formulées dans les problèmes d'apprentissage. Pour l'analogie, le *signal observé*  $x$  correspond à la *variable à expliquer* et la matrice  $\Phi$  du *dictionnaire aux variables explicatives*.

Ce problème peut être résolu de plusieurs manières. Par exemple, lorsque  $\mathcal{F}$  est une base orthonormée, la solution est donnée par un seuillage généralisé judicieux des coefficients  $\langle x, \varphi_{\lambda} \rangle$ . Pour le choix standard  $p = 1$ , ce seuillage généralisé est en fait le seuillage doux popularisé par Donoho et ses collaborateurs. En apprentissage, on reconnaît le problème du LASSO (Tibshirani, 1996) aussi connu dans la communauté "signal" sous le nom de *basis pursuit denoising* (Chen *et al.*, 1998). Pour  $p \neq 1$ , on obtient une formule de seuillage différente.

Dans le cas plus général d'un repère, les algorithmes itératifs tels que BCR (Block Coordinate Relaxation (Bruce *et al.*, 1998)), ou les itérations de Landweber (voir Daubechies *et al.*, 2004), sont des solutions dont la convergence est prouvées. Ils sont très utilisés dans la communauté du traitement du signal en raison de leur simplicité de mise en oeuvre et de leur flexibilité. En apprentissage les algorithmes de type homotopie (Osborne *et al.*, 2000) ou le LARS (Efron *et al.*, 2004) sont en général préférés. Quelques considérations sur la vitesse de convergence comparée du LARS et du seuillage itératif de Landweber sont disponibles dans (Daubechies *et al.*, 2007).

On s'intéresse ici aux problèmes pour lesquels l'indice  $\lambda$  des éléments du repère est en fait un double indice, où les deux indices jouent un rôle complètement différent. Plus précisément, la parcimonie est vue comme une fonction d'un seul des deux indices, mais pas de l'autre. De tels problèmes ont déjà été étudiés par plusieurs auteurs, en particulier dans le contexte de la régression à l'aide de variables groupées (Yuan & Lin, 2006). Cette dernière approche est un cas particulier de la situation plus générale étudiée ici, dans laquelle un mélange de parcimonie et de persistance se traduit par un choix réfléchi d'une norme mixte. Selon le choix de cette norme mixte, la parcimonie est favorisée à l'intérieur des groupes, ou bien à travers l'ensemble des groupes. On présentera quelques exemples dans lesquels une approche est mieux adaptée qu'une autre.

Les signaux multicanaux sont de bons exemples pour lesquels des ensembles de doubles indices entrent en jeu : un indice est associé au repère dans lequel le signal multicanal est décomposé, et un autre indice est associé au canal. Les signaux multicanaux sont donc de bons candidats à l'application de techniques de régression parcimonieuse et structurée.

On s'attarde particulièrement sur des applications aux signaux neurophysiologiques tels que l'électroencéphalographie (EEG) et la magnétoencéphalographie (MEG). L'EEG

(resp. MEG) est la mesure de l'activité électrique (resp. magnétique) du cerveau enregistrée par un ensemble de capteur disposés sur le scalp. Les signaux EEG et MEG sont donc naturellement multicanaux, ce qui en fait de bons candidats pour la régression parcimonieuse et persistante. De plus, puisque les repères temps-fréquence sont connus pour leurs bonnes performances d'analyse de ces signaux, les coefficients considérés sont au final indicés par trois indices au lieu de deux, ce qui offre de nouvelles possibilités de modélisation. On discutera de quelques unes de ces possibilités dans ce papier.

## 2 Persistance et parcimonie sous forme variationnelle

On s'intéresse ici aux situations dans lesquelles les coefficients de la décomposition d'un signal dans  $\mathcal{F}$  sont repérés par un double indice. Concrètement, ceci n'est pas une restriction vue que chacun peut choisir de numéroter les coefficients comme bon lui semble. Dans les cas considérés, il arrive que la parcimonie ne soit souhaitée que pour un seul des deux indices, tandis qu'on attend des propriétés de *persistance* pour le second. De tels mélanges de parcimonie et de persistance peuvent être traduits dans des approches variationnelles en minimisant une norme mixte bien choisie pour les coefficients. Dans un souci de simplicité, on se limitera aux mélanges de normes  $\ell_1$  et  $\ell_2$ .

### 2.1 Normes mixtes

À partir de maintenant, on se limite au cas de la dimension finie :  $\mathcal{H} = \mathbb{R}^N$ , avec  $N = KL$ . On note  $\|\cdot\|$  la norme euclidienne. Étant donné une suite doublement indicée  $\{\alpha_{k\ell}, k = 1, \dots, K, \ell = 1 \dots L\}$ , la norme  $\ell_{p,q}$  correspondante est définie telle que

$$\|\alpha\|_{p,q} = \left( \sum_{k=1}^K \left( \sum_{\ell=1}^L |\alpha_{k\ell}|^p \right)^{q/p} \right)^{1/q} \quad (3)$$

Le problème de régression *parcimonieuse et persistante* correspondant

$$\min \left[ \left\| x - \sum_{\lambda \in \Lambda} \alpha_\lambda \varphi_\lambda \right\|_{\mathcal{H}}^2 + \mu \|\alpha\|_{p,q}^q \right] \quad (4)$$

est étudié dans (Kowalski, 2007), où des résultats comparables aux résultats classiques de régression parcimonieuse sont obtenus. En particulier, il est montré que dans le cas où  $\mathcal{F}$  est une base orthonormée, le minimum de la fonctionnelle pénalisée par une norme  $\ell_{2,1}$  sur les coefficients est donnée par l'opérateur de seuillage généralisé

$$\alpha_{k\ell} = \gamma_{k\ell} M_k, \quad M_k = \left( 1 - \frac{\mu}{\|\gamma_k\|_2} \right)^+, \quad (5)$$

où  $\gamma_{k\ell} = \langle x, \varphi_{k\ell} \rangle$ , et  $\gamma_k = (\gamma_{k1}, \dots, \gamma_{kL})$  est la suite (simplement indicée)  $\gamma_{k\ell}$  pour un  $k$  donné. L'opérateur  $^+$  est défini tel que  $x^+ = x$  pour  $x > 0$  et 0 pour  $x \leq 0$ .

Ce problème (et sa solution) a été introduit auparavant sous le nom de group-LASSO (voir Yuan & Lin, 2006, et ses références). On peut voir aisément que cela correspond à un seuillage généralisé, qui sélectionne ou rejette globalement une “ligne” entière de coefficients, pour un indice  $k$  fixé, lorsqu’on les multiplie point par point avec une fonction *masque*  $M$ .

De manière similaire, le minimum de la fonctionnelle pénalisée par une norme  $\ell_{1,2}$  sur les coefficients, est donnée par un autre opérateur de seuillage généralisé

$$\alpha_{k\ell} = \text{sign}(\gamma_{k\ell}) \left( |\gamma_{k\ell}| - \frac{\mu}{1 + \mu L_k} \|\gamma_k\| \right)^+, \quad (6)$$

avec  $\|\gamma_k\| = \sum_{l=1}^{L_k} \check{\gamma}_{k,l}$ , où  $\check{\gamma}_{k,l}$  sont les coefficients  $\gamma$  classés en ordre décroissant pour un  $k$  donné. La quantité  $L_k$  est choisie telle que  $\check{\gamma}_{k,L_k+1} \leq \lambda \left( \sum_{\ell=1}^{L_k+1} (\check{\gamma}_{k,\ell} - \check{\gamma}_{k,L_k+1}) \right)$  et  $\check{\gamma}_{k,L_k} > \lambda \left( \sum_{\ell=1}^{L_k} (\check{\gamma}_{k,\ell} - \check{\gamma}_{k,L_k}) \right)$ . Cet opérateur de seuillage joue, en quelque sorte, un rôle un peu plus subtil. Il coïncide avec un opérateur de seuillage doux où le seuil dépend des coefficients : il est adapté à une ligne donnée de coefficients. Plus précisément, moins la suite  $\{\gamma_{k\ell}, \ell = 1 \dots L\}$  est parcimonieuse, plus la norme  $\ell_1$  correspondante, ainsi que son seuil, sont élevés. Par conséquent, ce seuillage a tendance à favoriser des lignes de coefficients parcimonieuses  $\{\alpha_{k\ell}, \ell = 1 \dots L\}$ .

Ces résultats peuvent être adaptés aux situations où la famille  $\mathcal{F}$  n’est pas une base orthonormée, mais plus généralement un repère, grâce aux algorithmes itératifs adéquats. De tels algorithmes, comme FOCUSS et les seuillages itératifs de Landweber, ont été étudiés en détail dans (Kowalski, 2007).

## 2.2 Signaux multicanaux

Les signaux multicanaux fournissent des exemples simples auxquels les techniques de régression parcimonieuse décrite auparavant peuvent être appliquées. Dans de tels situations, on cherche souvent des décompositions pour lesquelles tous les canaux utilisent le même repère : le modèle suppose des atomes scalaires et des coefficients multicanaux

$$\mathbf{x} = \sum_{\lambda,c} \alpha_{\lambda,c} \varphi_{\lambda}, \quad c = 1, \dots, N_c \quad (7)$$

où  $N_c$  est le nombre de canaux. Il est souvent assez naturel de supposer que tous (ou presque tous) les canaux partagent les mêmes atomes actifs  $\varphi_{\lambda}$  (i.e les atomes avec un coefficient non nul).

Supposer la parcimonie à l’intérieur des canaux et une persistance à travers les canaux devient alors une idée naturelle, qui conduit à considérer des seuillages généralisés tels que (5), avec une norme  $\ell_1$  sur l’indice des atomes, et une norme  $\ell_2$  sur l’indice des canaux. Cette approche a été utilisée dans (Kowalski & Torrèsani, 2007) pour le débruitage de signaux audio multicanaux, et par plusieurs autres groupes (voire Teschke & Ramlau, 2007, Fornasier & Rauhut, 2007).

### 2.3 Localiser l'estimateur $\ell_{2,1}$ : le group-LASSO fenêtré

Dans certaines situations, tous les canaux n'ont pas besoin d'être actifs en même temps. Par exemple, dans les signaux EEG ou MEG, il arrive régulièrement que seulement quelques capteurs localisés dans une région donnée sur le scalp fournissent un signal significatif, alors que les autres sont supposés inactifs. Lorsque les sous-groupes de canaux qui sont simultanément actifs ou inactifs sont connus par avance, les coefficients peuvent être réindexés de manière à ce que l'estimateur classique du group-LASSO puisse être utilisé. Un exemple trivial d'une telle réindexation est montré sur la figure 1. Ici, la réindexation est simplement une séparation des groupes en sous-groupes, mais des réindexations plus complexes peuvent aussi être considérées.

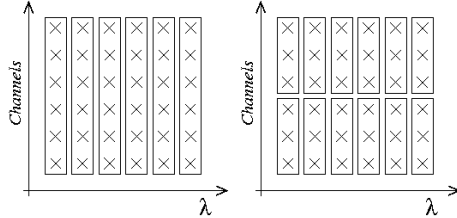


FIG. 1 – Un exemple de réindexation des coefficients.

Cependant, ce n'est pas la situation la plus générale, et les groupes de canaux actifs ne sont en général pas connus à l'avance, ni même fixés. C'est pour cette raison qu'une alternative à "fenêtre glissante" de l'approche précédente peut être appréciée.

Pour cela, on considère tout d'abord le cas où le repère  $\mathcal{F}$  est une base orthonormée. On suppose qu'une information supplémentaire sur les canaux est disponible, disant pour chaque canal  $c$ , quels sont les autres canaux qui sont "actifs" ou "inactifs" en même temps que  $c$ <sup>1</sup>. Ceci génère un système de voisinage, qui associe à chaque canal  $c = 1, \dots, N_c$  un groupe  $\mathcal{N}(c)$  de canaux "proches". Pour un canal  $c$  donné, il semble alors raisonnable d'utiliser seulement son voisinage dans  $\mathcal{N}(c)$  pour estimer sa décomposition parcimonieuse sur la base  $\mathcal{F}$ , tout en exploitant la persistance à l'intérieur de  $\mathcal{N}(c)$ . En adaptant l'estimateur du group-LASSO (5), cela suggère de calculer

$$\alpha_{\lambda c} = \gamma_{\lambda c} \left( 1 - \frac{\mu}{\|\gamma_{\lambda;c}\|_{\ell_2(\mathcal{N}(c))}} \right)^+, \quad (8)$$

où l'on note  $\gamma_{\lambda;c}$  la sous-suite  $\{\gamma_{\lambda c'}, c' \in \mathcal{N}(c)\}$ . On souligne le fait que contrairement à l'approche de réindexation ci-dessus, chaque canal utilise son propre voisinage. Si l'on compare (8) à l'équation (5), la fonction de masque dépend maintenant du canal. Malheureusement, ce nouvel opérateur généralisé n'est plus associé à un système variationnel simple.

Lorsque le repère considéré n'est plus une base orthonormée, les algorithmes itératifs tels que BCR ou le seuillage itératif de Landweber peuvent être adaptés à cette nouvelle situation, qui donnent des estimateurs par seuillage généralisé itératif.

<sup>1</sup>C'est typiquement une situation rencontrée avec les signaux EEG ou MEG, où les capteurs proches les uns des autres sur le scalp sont à même de produire des signaux quasi-identiques.

## 2.4 Introduire de la persistance dans l'estimateur $\ell_{1,2}$

Comme on l'a vu précédemment, la pénalité  $\ell_{1,2}$  sur les coefficients est complètement différente de la pénalité  $\ell_{2,1}$  qui conduit à la régression group-LASSO : elle favorise la parcimonie à l'intérieur des groupes de coefficients plutôt que la parcimonie à travers les groupes. Par exemple, la formule de seuillage (6) va sélectionner un petit nombre de coefficients à l'intérieur de chaque groupe. Pour fixer les idées, on suppose qu'un seul coefficient est sélectionné à l'intérieur de chaque groupe. Ce coefficient, peut varier d'un groupe à l'autre, puisque dans la norme choisie, rien ne l'empêche de le faire. Si l'on veut favoriser la persistance des coefficients retenus, une approche similaire à la précédente peut être développée, en tenant compte des voisins des coefficients considérés. Ici aussi, on commence par associer à n'importe quel index  $k$  une famille  $\mathcal{N}(k)$  de voisins. Alors, pour un  $k$  fixé, on résout le problème de minimisation suivant, avec une pénalité  $\ell_{1,2}$  sur les coefficients du vecteur  $\gamma_{\mathcal{N}(k)} = \{\gamma_{k'\ell}, \ell = 1, \dots, K, k' \in \mathcal{N}(k)\}$ .

Si l'on applique la même approche que précédemment, la formule de seuillage généralisée (6) devient alors

$$\alpha_{k\ell} = \text{sign}(\gamma_{k\ell}) \left( |\gamma_{k\ell}| - \frac{\mu}{1 + \mu L |\mathcal{N}(k)|} \|\gamma_{\mathcal{N}(k)}\|_1 \right)^+ . \quad (9)$$

où  $|\mathcal{N}(k)|$  est le cardinal de l'ensemble  $\mathcal{N}(k)$ . Ici aussi, ce seuillage généralisé n'est pas associé à une approche variationnelle simple.

## 2.5 Décomposition multicouches

Les différents estimateurs introduits précédemment permettent d'introduire certaines structures dans la régression des signaux. Cependant, dans une formulation telle que (4) et les diverses formules de seuillage présentées, seul un type de structure peut être introduit à la fois, dans un dictionnaire choisi.

Or, suivant le dictionnaire utilisé pour représenter le signal, celui-ci peut présenter plusieurs «couches» avec des structures différentes. Par exemple, une image peut se décomposer comme la somme de ses «bords» et ses «textures». Il peut alors être judicieux de représenter un signal comme la somme de plusieurs couches.

Les décompositions multicouches parcimonieuses de signaux ont été introduites tout d'abord par Berger et ses collaborateurs (Berger *et al.*, 1994). Elles ont été régulièrement utilisées depuis quelques années dans plusieurs domaines, comme la décomposition de signaux audio (Daudet & Torrèsani, 2002, Févotte *et al.*, 2008) ou encore en traitement d'image (on pourra se reporter par exemple à l'analyse en composante morphologique décrite dans (Elad *et al.*, 2005)). La principale idée derrière la décomposition multicouches est de chercher une approximation parcimonieuse d'un signal dans un dictionnaire dont les éléments sont indexés, i.e. une union de sous-dictionnaire. Une fois la régression effectuée, les synthèses partielles dans les sous-dictionnaires donnent les estimations des différentes couches du signal.

On se limitera ici à la décomposition d'un signal selon deux couches et un résidu (ce dernier n'admettant pas de décomposition parcimonieuse) :

$$x = x_1 + x_2 + r$$

Les décompositions multicouches sont habituellement effectuées en utilisant des dictionnaires construits comme l'union de bases (ou de repères). On suppose alors que la couche  $x_1$  (resp.  $x_2$ ) admet une décomposition dans un repère  $\{\varphi\}$  (resp.  $\{\psi\}$ ). On cherche alors à estimer le signal dans le dictionnaire construit comme l'union des deux bases  $\{\varphi\}$  et  $\{\psi\}$ .

Dans ce cas, la fonction objective à optimiser est sous la forme

$$\Gamma(\alpha, \beta) = \left\| x - \sum_{\lambda} \alpha_{\lambda} \varphi_{\lambda} - \sum_{\lambda'} \beta_{\lambda'} \psi_{\lambda'} \right\|_2^2 + \mu_1 \Phi_1[\alpha] + \mu_2 \Phi_2[\beta], \quad (10)$$

où les pénalités  $\Phi_1$  et  $\Phi_2$  sont construites à partir de normes. Le choix de ces normes est dictées par l'information *a priori* sur les deux couches. Les couches sont alors estimées par

$$x_1 = \sum_{\lambda} \alpha_{\lambda} \varphi_{\lambda}, \quad x_2 = \sum_{\lambda'} \beta_{\lambda'} \psi_{\lambda'}.$$

Afin de minimiser  $\Gamma$ , on se limite aux cas où  $\{\varphi\}$  et  $\{\psi\}$  sont deux bases. Dans ce cas, on peut appliquer l'algorithme BCR (Block Coordinated Relaxation de Bruce *et al.*, 1998) décrit ci-dessous. On note  $\mathcal{S}_1$  et  $\mathcal{S}_2$  les opérateurs de seuillage généralisés associés respectivement aux pénalités  $\Phi_1$  et  $\Phi_2$ .

**Algorithme 1**

**Soit**  $\alpha^{(0)}$  et  $\beta^{(0)}$  donnés.

**Faire**

- Estimation de la couche 1 :  $r_1^{(m)} = x - \sum_{\lambda'} \beta_{\lambda'}^{(m)} \psi_{\lambda'}$   
 Pour tout  $\lambda$ , poser :  $\alpha_{\lambda}^{(m+1)} = \mathcal{S}_1^{\mu}(\langle r_1^{(m)}, \varphi_{\lambda} \rangle)$
- Estimation de la couche 2 :  $r_2^{(m)} = x - \sum_{\lambda} \alpha_{\lambda}^{(m)} \varphi_{\lambda}$   
 Pour tout  $\lambda'$ , poser :  $\beta_{\lambda'}^{(m+1)} = \mathcal{S}_2^{\mu'}(\langle r_2^{(m)}, \psi_{\lambda'} \rangle)$

**Jusqu'à convergence**

Les opérateurs de seuillages généralisés  $\mathcal{S}_1$  et  $\mathcal{S}_2$  sont donnés par (5) et (6), correspondant à la norme choisie. On peut montrer (comme résumé dans Kowalski & Torrèsani, 2007) que cet algorithme converge vers le minimum de (10). Si  $\{\varphi\}$  et  $\{\psi\}$  ne sont plus des bases, l'algorithme ci-dessus n'est pas applicable en l'état. On peut alors appliquer un seuillage de Landweber itératif tel que décrit dans (Kowalski, 2007) pour les normes mixtes, dont la convergence vers le minimum recherché est prouvée.

On a choisi cet algorithme, car la construction d'un dictionnaire par union de bases est simple, et l'algorithme s'interprète facilement : une couche est estimée à partir du signal auquel on a retirée l'estimation de l'autre couche.

La simplicité de cet algorithme permet de remplacer les opérateurs  $\mathcal{S}_1$  et  $\mathcal{S}_2$  par leur version fenêtrée (8) ou (9), afin d'introduire de la localité ou de la persistance à l'intérieur d'une couche. Si, dans ce cas, l'algorithme converge en pratique avec les expériences faites, la preuve de convergence est toujours à l'étude.

### 3 Application aux signaux EEG et MEG

#### 3.1 Introduction

Les signaux provenant des expériences de neurosciences tels que les EEG ou les MEG sont de bons exemples pour tester les idées développées dans ce papier : ce sont des signaux multicanaux<sup>2</sup>, auxquels des techniques de régression parcimonieuse ont déjà été appliquées (voir par exemple Studer *et al.*, 2006, pour une adaptation du matching pursuit multicanal, et ses références pour d'autres contributions). De plus, ils sont souvent bien représentés dans des bases ou des repères temps-fréquence ; cela produit des décompositions dans des systèmes atomiques avec trois indices : le temps, la fréquence, et le canal.

$$\mathbf{x} = \sum_{t,f,c} \alpha_{t,f,c} \varphi_{t,f} . \quad (11)$$

Un des principaux problèmes dans l'analyse de signaux *EEG*, est de séparer l'information utile de diverses sources de bruits et de caractéristiques moins importantes. C'est en raison de cela que les techniques de régression multicanales appropriée aux signaux issus de plusieurs capteurs, et qui font intervenir une modélisation adéquate, présentent un intérêt certain. On décrit par la suite les premières expériences faites en utilisant des techniques de modélisation décrites dans ce papier.

Dans les signaux EEG dont on parle ci-dessous, 34 canaux sont à notre disposition, avec une topographie<sup>3</sup> donnée sur la figure 2. Le taux d'échantillonnage est de 1000 Hz. Les signaux multicanaux originaux sont représentés sur la figure 3.

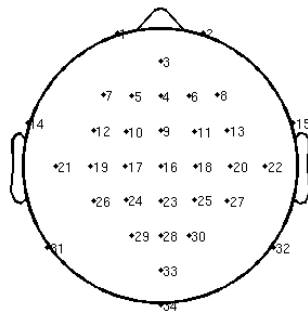


FIG. 2 – Localisation des électrodes pour les signaux EEG considérés.

Les signaux sont constitués d'environ 130 essais d'une durée de 1.2 s. Les essais résultent d'une expérience réalisée dans les mêmes conditions. Ils ont été concaténés afin d'être traités globalement. Sur cet ensemble de données particulier, les portions de signaux qui présentent un intérêt particulier sont les potentiels évoqués, qui sont des signaux de durée courte, et qui apparaissent suite à un stimulus (à raison d'un stimulus

<sup>2</sup>Typiquement, les signaux EEG sont donnés par un nombre de capteurs situé entre 20 et 50, tandis que les signaux MEG en demande encore plus

<sup>3</sup>La topographie exploitée dans des schémas de group-LASSO fenêtré comme vu plus haut.

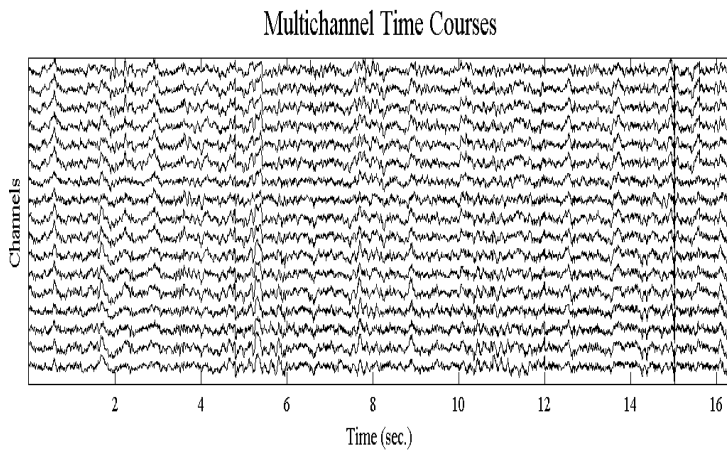


FIG. 3 – Signaux EEG originaux.

par essai). Ces signaux ont une très faible amplitude comparé au fond généré par les divers signaux biologiques et le bruit de mesure. Une technique souvent utilisée est alors de traiter la moyenne des essais. Cette moyenne a pour inconvénient une perte de précision due à la variabilité des essais. L'analyse d'un seul essai est donc un challenge majeur dans ce contexte.

Dans la suite, ces signaux sont décomposés en utilisant diverses méthodes de décomposition multicanale décrite précédemment, en utilisant des bases MDCT (voir Vetterli & Kovacevic, 1995, pour plus de détails sur les bases MDCT).

### 3.2 Régression LASSO et group-LASSO

Afin de comparer les estimations données par le LASSO et le group-LASSO, on a utilisé des bases MDCT avec une fenêtre de longueur 256 ms. Le paramètre de régularisation a été choisi de manière à ce que l'énergie du résidu soit, à peu de chose près, égale dans les deux cas. Comme on peut le voir, les deux techniques de débruitage ont tendance à filtrer les hautes fréquences. Cependant, la qualité de l'estimateur du group-LASSO est meilleur aussi bien numériquement que visuellement, puisque les propriétés de persistance à travers les canaux ressortent. Afin de rendre les figures le plus clair possible, seuls les signaux des capteurs 3,4,7,8,9,12,13,17,18,21,22,23,26,27,29,30,33 sont représentés ici et dans les figures suivantes.

Dans les expériences numériques considérées ici, les paramètres  $\mu$  ont été choisis tels que les signaux estimés aient le même pourcentage d'énergie par rapport au signal original (la valeur exacte, ici de 50%, a été choisie afin d'illustrer au mieux les différences entre les deux approches). Les différences entre les deux signaux estimés apparaissent clairement sur les figures 4 et 5. Comme on pouvait s'y attendre, le group-LASSO multicanal fait ressortir les composantes qui sont présentes sur tous les capteurs simultanément, tandis que le LASSO multicanal a une structure moins lisse, bien que les

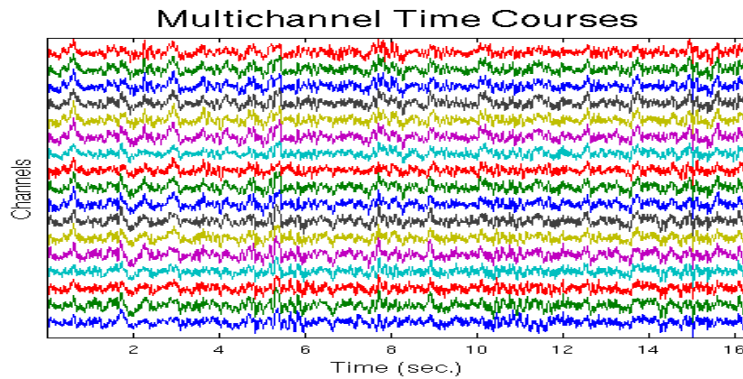


FIG. 4 – Signaux EEG débruités par le LASSO.

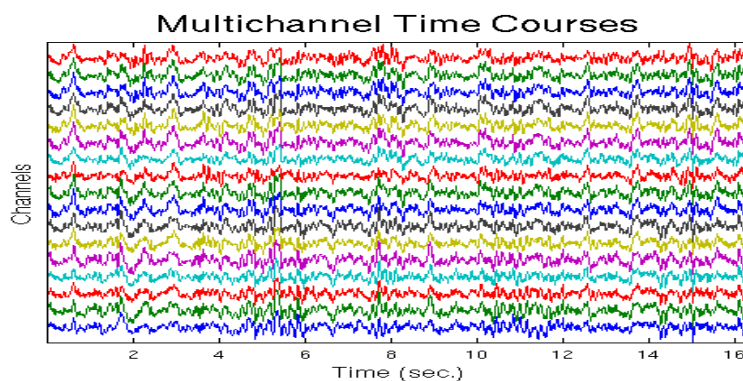


FIG. 5 – Signaux EEG débruités par le group-LASSO.

deux signaux aient tous les deux la même énergie.

Si l'on se rapporte à l'équation (5), l'indice  $k$  est ici un indice temps-fréquence ; par conséquent, l'estimation du group-LASSO équivaut à multiplier point par point les coefficients MDCT avec un masque temps-fréquence  $M_{t,f}$  fixé. Parallèlement, l'estimateur du LASSO donne un masque temps-fréquence pour chaque canal. Bien qu'ils ne soient pas montrés ici, le masque du group-LASSO est significativement différent de la moyenne des masques donnée par le LASSO. La moyenne des masques du LASSO revient essentiellement à un filtre passe-bas, tandis que le masque du group-LASSO fait apparaître plus de structures.

Les résultats obtenus avec le group-LASSO fenêtré décrits à la section 2.3 sont reportés dans un autre travail.

### 3.3 Décomposition multicouches

Afin d'exploiter les approches décrites précédemment, on peut définir plusieurs façons de regrouper les coefficients (afin de favoriser la parcimonie dans les groupes, ou à travers les groupes) que l'on illustre sur la figure 6. Dans de telles situations, les choix naturels pour les groupes sont des lignes ou des plans.

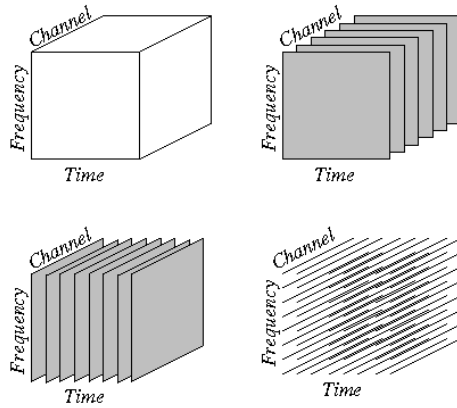


FIG. 6 – L'ensemble des indices temps-fréquence-canal (en haut à gauche) et divers regroupements (plans ou lignes).

On se limite par la suite aux décompositions multicanales. Le dictionnaire utilisé est une union de deux bases orthonormées MDCT, avec une différence significative pour la résolution temps-fréquence des fenêtres.

Pour les composantes localisée en fréquence, l'hypothèse est qu'à l'intérieur de chaque plan  $temps \times canal$ , seul quelques fréquences remarquables vont apparaître, qui ne sont pas forcément les mêmes pour chaque plans. On utilise alors le troisième regroupement représenté sur la figure 6 (dans le coin inférieur gauche) et on utilise la norme mixte

$$\Phi_1[\alpha] = \sum_{t,c} \left( \sum_f |\alpha_{tfc}| \right)^2. \quad (12)$$

Pour cette composante, on choisit une base MDCT avec une fenêtre de longueur 256 ms et correspond à la base  $\{\varphi\}$  de la section 2.5.

Pour estimer les composantes localisées en temps, l'estimateur du group-LASSO semble le mieux adapté, pour sélectionner des groupes complets à des dates fixes, et ainsi favoriser la persistance en fréquence et à travers les canaux. Cela correspond donc à la norme mixte

$$\Phi_2[\beta] = \sum_t \sqrt{\sum_{f,c} |\beta_{tfc}|^2}. \quad (13)$$

La base MDCT choisie pour cette composante à une taille de fenêtre de 32 ms et correspond à la base  $\{\psi\}$  de la section 2.5.

Afin de favoriser la persistance en temps d'une fréquence, on a décidé d'utiliser la version fenêtrée de la norme  $\ell_{1,2}$ . Le système de voisinage est construit comme suit. Soit  $\ell_k$  l'ensemble de toutes les fréquences au temps  $k$ . On considère alors toutes les fréquences contenues dans l'ensemble  $\mathcal{N}(k) = \{\ell_{k-2}, \ell_{k-1}, \ell_k, \ell_{k+1}, \ell_{k+2}\}$ , au lieu de considérer seulement  $\ell_k$  à l'indice  $k$  donné. Finalement, on utilise l'opérateur de seuillage généralisé (9) et (5) pour estimer les deux couches.

L'algorithme 1 décrit en section 2.5 a été utilisé dans cette configuration. Pour la clarté des comparaisons, les résultats obtenus avec group-LASSO multicanal classique sont aussi illustrés. Les paramètres ont été choisis de façon à obtenir des résidus dont l'énergie est comparable (environ 10% de l'énergie du signal original). De plus, les poids relatifs des deux couches, donnés par le rapport  $\mu_2/\mu_1$ , ont été choisis tels que les composantes localisées en temps soient restreintes à de très courts signaux transitoires (et ont donc une énergie faible).

Les composantes localisées en fréquences estimées par les deux méthodes sont difficilement différenciables visuellement. On a donc choisi de ne représenter qu'une seule d'entre elles sur la figure 7. Les seules différences entre les deux peuvent se remarquer sur les hautes fréquences (non montrées ici en raison du peu d'intérêt qu'elles présentent).

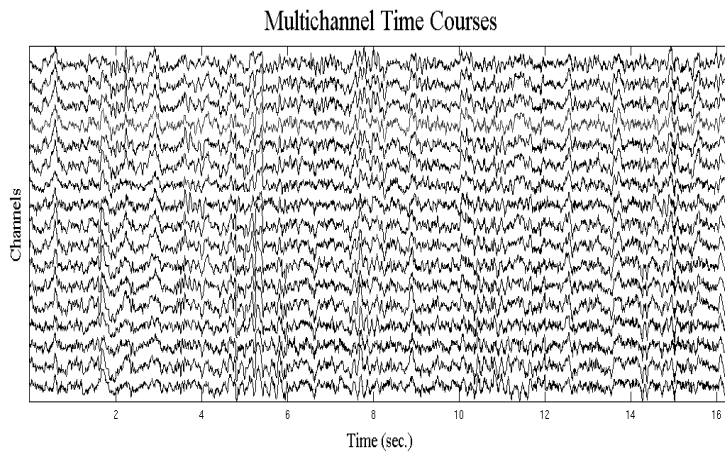


FIG. 7 – Composantes localisées en fréquence pour les signaux EEG représentés sur la figure 3, estimées avec l'algorithme décrit en section 3.3.

Les différences apparaissent plus visiblement sur les figures des composantes localisées en temps. En effet, l'estimateur du group-LASSO multicanal illustré sur la figure 8 met en relief des variations plus nettes, et des composantes très courtes et très oscillantes. Ces dernières disparaissent sur l'estimation donnée par l'algorithme 1. Plus précisément, elles sont capturées par la couche localisée en fréquence.

Ceci apparaît plus clairement si l'on se concentre sur un seul canal. On représente sur la figure 10 les deux estimations de la couche localisée en temps pour le canal 3. On remarque très clairement que l'approche proposée donne une solution plus parcimonieuse

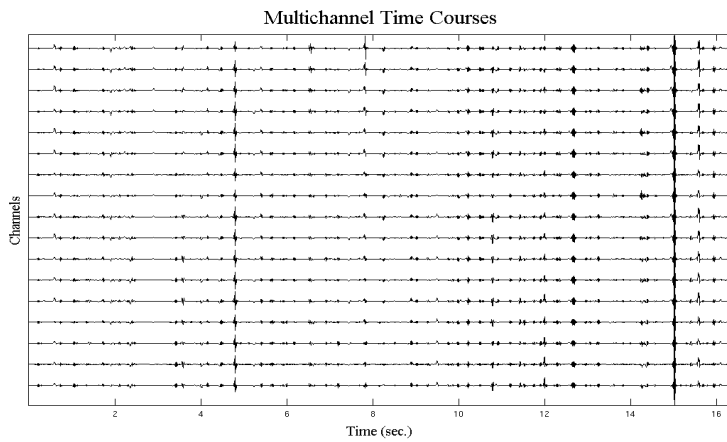


FIG. 8 – Composantes localisées en temps pour les signaux EEG représentés sur la figure 3, en utilisant l'estimateur du group-LASSO multicanal.

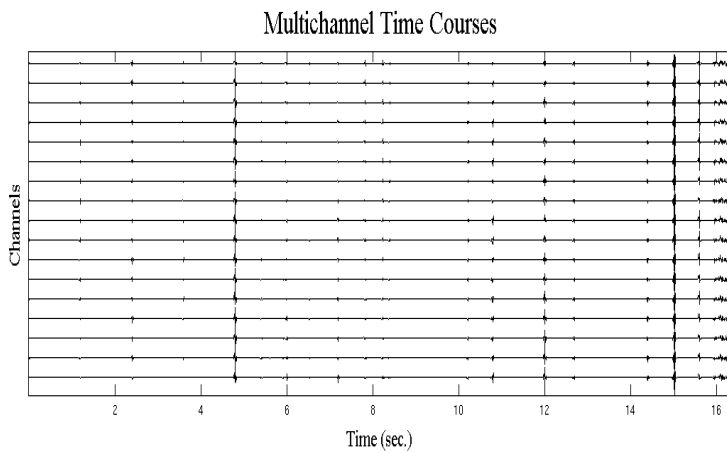


FIG. 9 – Composantes localisées en temps pour les signaux EEG représentés sur la figure 3, en utilisant l'algorithme de la section 3.3.

en temps.

## 4 Conclusion

Ce papier montre comment les normes mixtes peuvent être utilisées comme un outil de modélisation, dans le contexte de la régression des signaux multicanaux. Les normes mixtes faisant intervenir un mélange de normes  $\ell_1$  et  $\ell_2$  conduisent à des opérateurs de

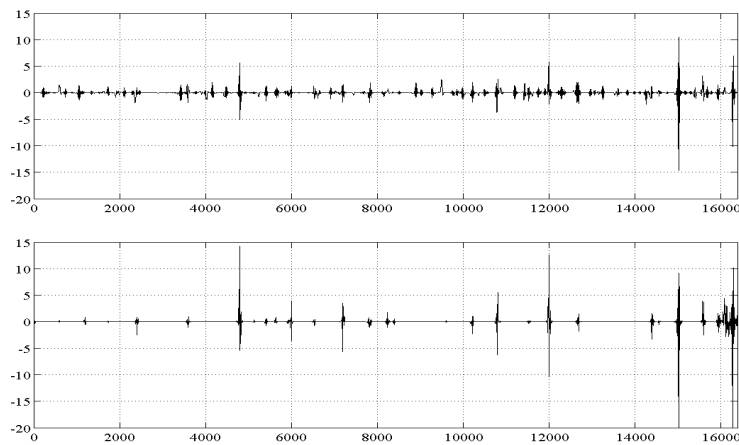


FIG. 10 – Composantes localisées en temps pour un canal particulier des signaux EEG représentés sur la figure 3, en utilisant les deux algorithmes. Haut : group-LASSO, bas : algorithme de la section 3.3.

seuillage généralisés, qui peuvent être interprétés à la lumière des structures que l'on veut introduire dans le modèle.

Les normes mixtes  $\ell_{2,1}$  et  $\ell_{1,2}$  ont des comportements sensiblement différents. La première est parfaitement adaptée aux signaux multicanaux, en raison de sa capacité à regrouper les coefficients. La seconde, plus subtile, permet de garder les coefficients les plus gros (comparés aux autres) à l'intérieur des groupes.

On a aussi introduit de nouveaux opérateurs de seuillage dans le but de localiser ou d'introduire de la persistance dans les estimateurs donnés par les normes mixtes. Même si ces opérateurs ne correspondent pas à un problème variationnel simple, leur utilité semble naturelle.

Finalement, on s'est attardé sur ces opérateurs de seuillage généralisés dans le cadre d'une décomposition multicouche. L'algorithme BCR donne une solution simple, facile à implémenter, et flexible.

## Remerciements

On tient à remercier B. Burle pour toutes les discussions et ses remarques précieuses, ainsi que pour nous avoir fourni les données utilisées pour illustrer ce papier. La figure 2 a été générée par EEGLab (Delorme, 2004) ; les transformées temps-fréquences ont été produites grâce à la Linear Time-Frequency Analysis Toolbox (LTFAT) développée et maintenue par P. Søndergaard (Søndergaard, 2008).

## Références

- BERGER J., COIFMAN R. & GOLDBERG M. (1994). Removing noise from music using local trigonometric bases and wavelet packets. *J. Audio Eng. Soc.*, **42**(10), 808–818.
- BRUCE A. G., SARDY S. & TSENG P. (1998). Block coordinate relaxation methods for nonparametric signal denoising. In *Proceedings of the SPIE - The International Society for Optical Engineering*, number 3391, p. 75–86.
- CHEN S. S., DONOHO D. L. & SAUNDERS M. (1998). Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, **20**(1), 33–61.
- DAUBECHIES I., DEFRISE M. & DE MOL C. (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, **57**(11), 1413 – 1457.
- DAUBECHIES I., FORNASIER M. & LORIS L. (2007). Accelerated projected gradient methods for linear inverse problems with sparsity constraints. *Journal of Fourier Analysis and Application*. to appear.
- DAUDET L. & TORRÉSANI B. (2002). Hybrid representations for audiophonic signal encoding. *Signal Processing*, **82**(11), 1595–1617. Special issue on Image and Video Coding Beyond Standards.
- DELORME A. (2004). Eeglab. <http://www.sccn.ucsd.edu/eeglab/>.
- EFRON B., HASTIE T., JOHNSTONE I. M. & TIBSHIRANI R. (2004). Least angle regression. *The Annals of Statistics*, **32**(2), 407–499.
- ELAD M., STARCK J.-L., DONOHO D. L. & QUERRE P. (2005). Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). *Journal on Applied and Computational Harmonic Analysis*, **19**, 340–358.
- FÉVOTTE C., TORRÉSANI B., DAUDET L. & GODSILL S. J. (2008). Sparse linear regression with structured priors and application to denoising of musical audio. *IEEE Trans. Audio, Speech and Language Processing*, **16**(1), 174–185.
- FORNASIER M. & RAUHUT H. (2007). Recovery algorithm for vector-valued data with joint sparsity constraints. *SIAM (to appear)*.
- KOWALSKI M. (2007). Sparse regression using mixed norms. submitted.
- KOWALSKI M. & TORRÉSANI B. (2007). Sparsity and persistence : mixed norms provide simple signal models with dependent coefficients. submitted.
- OSBORNE M. R., PRESNELL B. & TURLACH B. A. (2000). A new approach to variable selection in least squares problems. *IMA J. Numerical Analysis*, **20**, 389–403.
- SØNDERGAARD P. (2008). LTFAT, the linear time-frequency analysis toolbox. <http://sourceforge.net/projects/ltfat>.
- STUDER D., HOFFMAN H. & KOENIG T. (2006). From eeg dependency multichannel matching pursuit to sparse topographic eeg decomposition. *Journal of Neuroscience Methods*, **153**(2), 261–275.
- TESCHKE G. & RAMLAU R. (2007). An iterative algorithm for nonlinear inverse problems with joint sparsity constraints in vector valued regimes and an application to color image inpainting. Preprint.
- TIBSHIRANI R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Serie B*, **58**(1), 267–288.

- VETTERLI M. & KOVACEVIC J. (1995). *Wavelets and subband coding*. Englewood Cliffs, NJ, USA : Prentice Hall.
- YUAN M. & LIN Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Serie B*, **68**(1), 49–67.