

Approximation des signaux : approches variationnelles et modèles aléatoires.

Matthieu Kowalski

LATP

15 octobre 2008



sous la direction de Bruno Torrèsani

Introduction : approximation parcimonieuse

Les entités ne doivent pas être multipliées par delà ce qui est nécessaire

Guillaume d'Occam

But

Analyser, expliquer, représenter... les signaux.

Exemples

Transcription automatique, séparation de sources, codage...

Problème : Comment représenter le signal et sélectionner "l'information" ?

Principe de parcimonie : expliquer un signal avec un minimum d'éléments.



Exemples de représentation d'un signal audio

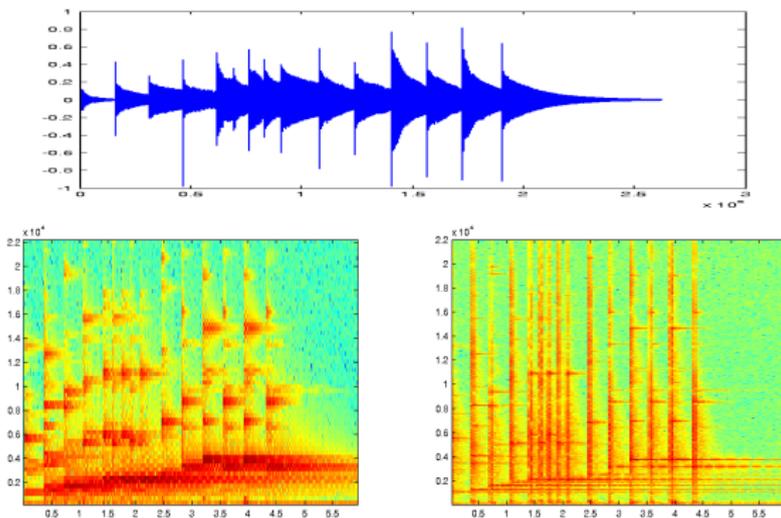


Figure : Images temps-fréquence. Haut : signal, En bas à gauche : représentation adaptée aux transitoires. En bas à droite, représentation adaptée aux tonals

Les caractéristiques d'intérêt sont rarement observables directement.

- 1 Introduction : approximation parcimonieuse
- 2 Approches variationnelles
 - Normes mixtes
 - Modèle hybride ou multicouches
 - Algorithmes de seuillage itératif
 - Application à la séparation tonale/transitoire
- 3 Modèles probabilistes
 - Retour sur le modèle hybride
 - Modèles pour les cartes de significances et les coefficients
 - Estimation du modèle
 - Application au débruitage
- 4 Conclusions et perspectives

Notations et définitions

Quelques notations

- Soit $s \in \mathbb{C}^M$ un signal
- Soit $\Phi \in \mathbb{C}^{M \times N}$, $M \leq N$ la matrice d'un dictionnaire $\{\varphi_k\}$ (ie un ensemble sur-complet), construit comme un ensemble d'atomes temps-fréquence.
- Soit $y = s + b$ une mesure bruitée du signal s .

Définition : coefficients de synthèse

Soit $\alpha \in \mathbb{C}^N$ tel que $s = \Phi\alpha = \sum_k \alpha_k \varphi_k$.

Les α_k sont appelés les *coefficients de synthèse*.

si $N > M$, il existe une infinité d'écriture possible

Définition : coefficients d'analyse

On appelle *coefficients d'analyse* les $\{\langle y, \varphi_k \rangle\} = \Phi^T y$

Parcimonie : approche synthèse

But : Trouver une “bonne représentation” \hat{s} de s tel que $\hat{s} = \Phi \hat{\alpha}$

Hypothèse : s admet une représentation parcimonieuse dans le dictionnaire choisi.

Solution “idéale” :

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \|\alpha\|_0 \quad \text{sc} \quad s = \Phi \alpha$$

Observation bruitée :

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \|y - \Phi \alpha\|_2^2 + \lambda \|\alpha\|_0$$

Problème : très difficile à résoudre en temps fini \Rightarrow on relâche la contrainte ℓ_0 en une contrainte ℓ_1

Problème du **LASSO** (Tibshirani 96) ou **Basis Pursuit Denoising** (Chen et al. 98) :

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \|y - \Phi \alpha\|_2^2 + \lambda \|\alpha\|_1$$

Liens analyse/synthèse et avec le max de vraisemblance

Point de vue Bayésien

$$\min_{\alpha} \{ \|y - \Phi\alpha\|_2^2 + \lambda \|\alpha\|_1 \}$$

$$\Leftrightarrow \max_{\alpha} \left\{ \underbrace{e^{-\|y - \Phi\alpha\|_2^2}}_{\text{"bruit gaussien"}} \underbrace{\prod_k e^{-\lambda|\alpha_k|}}_{\text{a priori Laplacien}} \right\}$$

Hypothèse : Φ correspond à une base orthonormée $\Phi^T = \Phi^{-1}$.

Solution

Soit $\tilde{y} = \Phi^T y$. La solution est obtenue par seuillage doux *coefficient par coefficient* :

$$\hat{\alpha}_k = \arg(\tilde{y}_k) (|\tilde{y}_k| - \lambda)^+$$

et $\hat{s} = \Phi\hat{\alpha}$

Exemples de représentation d'un signal audio

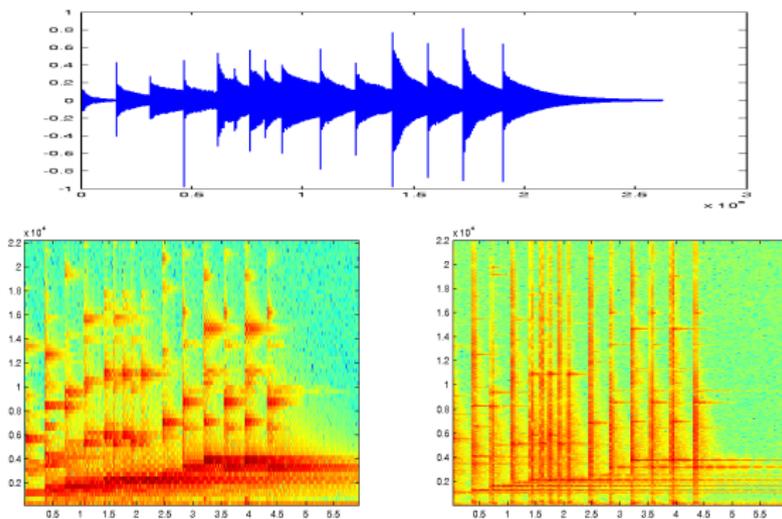


Figure : *Images temps-fréquence. Haut : signal, En bas à gauche : représentation adaptée aux transitoires. En bas à droite, représentation adaptée aux tonals*

- 1 Introduction : approximation parcimonieuse
- 2 Approches variationnelles
 - Normes mixtes
 - Modèle hybride ou multicouches
 - Algorithmes de seuillage itératif
 - Application à la séparation tonale/transitoire
- 3 Modèles probabilistes
 - Retour sur le modèle hybride
 - Modèles pour les cartes de significances et les coefficients
 - Estimation du modèle
 - Application au débruitage
- 4 Conclusions et perspectives

Normes mixtes : Définition

Définition

Étant donné une suite doublement indicée $\{\alpha_{g,m}$, on appelle norme mixte $\ell_{g,m}$ de α la norme

$$\|\alpha\|_{p,q} = \left(\sum_g \left(\sum_m |\alpha_{g,m}|^p \right)^{q/p} \right)^{1/q}$$

Remarques

- $\ell_{p,q}$ est une vraie norme pour $p, q \geq 1$.
- Les cas $p = +\infty$ ou $q = \infty$ sont obtenus en remplaçant la norme correspondante par le supremum.
- On peut définir les quasi-normes correspondantes pour $p, q < 1$.

Régression parcimonieuse et persistante

On s'intéresse au problème d'optimisation suivant

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \|y - \Phi\alpha\|_2^2 + \lambda \|\alpha\|_{p_2, q_2}^{q_2}$$

Remarque

Ce problème est convexe pour $p, q \geq 1$ et strictement convexe pour $p, q > 1$.

Le point de vue de Bayes

$$\begin{aligned}\hat{\alpha} &= \underset{\alpha}{\operatorname{argmax}} e^{-\|y - \Phi\alpha\|_2^2} e^{-\lambda \|\alpha\|_{p, q}^q} \\ &= \underset{\alpha}{\operatorname{argmax}} e^{-\|y - \Phi\alpha\|_2^2} \prod_g \exp\{-\lambda \|\alpha_g\|_p^q\}\end{aligned}$$

⇒ Indépendance entre les groupes de coefficients $\alpha_g = (\alpha_{g,1} \dots \alpha_{g,m} \dots)$
Le découplage est sur les groupes et non sur les coefficients

De nouveaux opérateurs de seuillage

On se place dans le cas où Φ est *orthogonale*. On note $\tilde{y} = \Phi^T y$

Solution du LASSO $\min_{\alpha} \|y - \Phi\alpha\|_2^2 + \lambda\|\alpha\|_1$

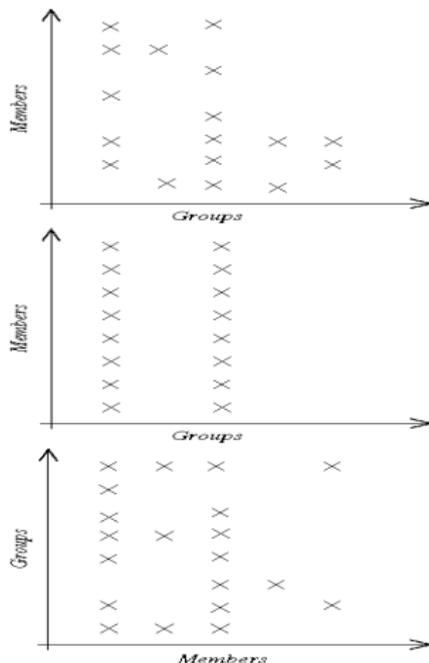
$$\hat{\alpha}_{g,m} = \arg(\tilde{y}_{g,m}) (|\tilde{y}_{g,m}| - \lambda)^+$$

Solution du G-LASSO $\min_{\alpha} \|y - \Phi\alpha\|_2^2 + \lambda\|\alpha\|_{2,1}$

$$\hat{\alpha}_{g,m} = \tilde{y}_{g,m} \left(1 - \frac{\lambda}{\|\tilde{y}_g\|_2}\right)^+$$

Solution du E-LASSO $\min_{\alpha} \|y - \Phi\alpha\|_2^2 + \lambda\|\alpha\|_{1,2}$

$$\hat{\alpha}_{g,m} = \arg(\tilde{y}_{g,m}) \left(|\tilde{y}_{g,m}| - \frac{\lambda}{1 + \lambda L_g} \|\tilde{y}_g\| \right)^+$$



Exemples de représentation d'un signal audio

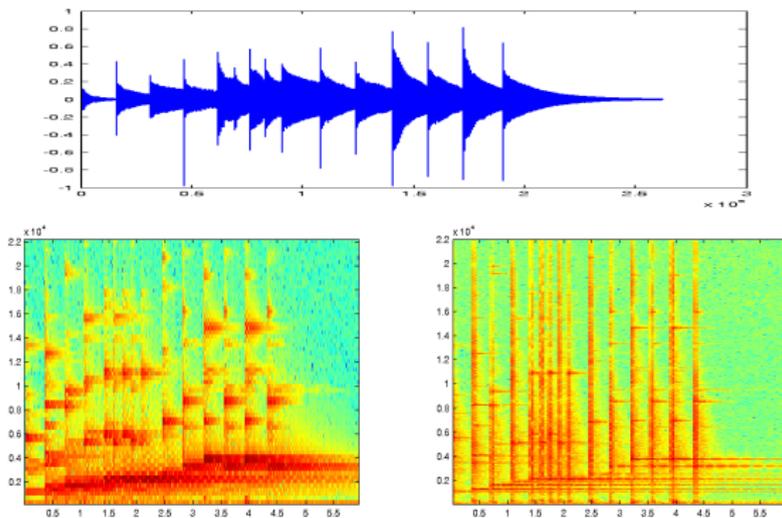


Figure : *Images temps-fréquence. Haut : signal, En bas à gauche : représentation adaptée aux transitoires. En bas à droite, représentation adaptée aux tonals*

- 1 Introduction : approximation parcimonieuse
- 2 Approches variationnelles
 - Normes mixtes
 - **Modèle hybride ou multicouches**
 - Algorithmes de seuillage itératif
 - Application à la séparation tonale/transitoire
- 3 Modèles probabilistes
 - Retour sur le modèle hybride
 - Modèles pour les cartes de significances et les coefficients
 - Estimation du modèle
 - Application au débruitage
- 4 Conclusions et perspectives

Décomposition hybride

On suppose que le signal s que l'on cherche s'écrit sous la forme

$$s = s_1 + s_2$$

On suppose de plus que chacune des deux couches admet une décomposition parcimonieuse dans un dictionnaire adaptée.

On forme un nouveau dictionnaire par l'union des deux dictionnaires

$$\mathcal{D} = \mathcal{U} \cup \mathcal{V}$$

ainsi, le modèle de synthèse s'écrit :

$$s = U\alpha + V\beta$$

Il existe toujours une infinité de décomposition possible

Problème : identifier α et β .

Formulation variationnelle

On cherche à estimer s par l'optimisation de la fonctionnelle

$$\hat{\alpha}, \hat{\beta} = \underset{\alpha, \beta}{\operatorname{argmin}} \|y - \sum_k \alpha_k u_k - \sum_\ell \beta_\ell v_\ell\|_2^2 + \lambda_1 \|\alpha\|_{p_1, q_1}^{q_1} + \lambda_2 \|\beta\|_{p_2, q_2}^{q_2}$$

$$\hat{\alpha}, \hat{\beta} = \underset{\alpha, \beta}{\operatorname{argmin}} \|y - U\alpha - V\beta\|_2^2 + \lambda_1 \|\alpha\|_{p_1, q_1}^{q_1} + \lambda_2 \|\beta\|_{p_2, q_2}^{q_2}$$

Ainsi, on obtient une séparation en deux couches :

$$\hat{s}_1 = U\hat{\alpha} \quad \hat{s}_2 = V\hat{\beta}$$

- 1 Introduction : approximation parcimonieuse
- 2 Approches variationnelles
 - Normes mixtes
 - Modèle hybride ou multicouches
 - **Algorithmes de seuillage itératif**
 - Application à la séparation tonale/transitoire
- 3 Modèles probabilistes
 - Retour sur le modèle hybride
 - Modèles pour les cartes de significances et les coefficients
 - Estimation du modèle
 - Application au débruitage
- 4 Conclusions et perspectives

Algorithme BCR

Hypothèse : U et V orthogonales.

La solution au problème précédent est obtenue par l'algorithme itératif suivant

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

Faire

- Estimation de la couche 1 :
$$r_1^{(m)} = x - \sum_{\ell} \beta_{\ell}^{(m)} v_{\ell}$$

Pour tout k , poser :
$$\alpha_k^{(m+1)} = \operatorname{argmin}_{\alpha} \|y - U\alpha\|_2^2 + \lambda_1 \|\alpha\|_{p_1, q_1}^{q_1}$$
- Estimation de la couche 2 :
$$r_2^{(m+1)} = x - \sum_k \alpha_k^{(m+1)} u_k$$

Pour tout ℓ , poser :
$$\beta_{\ell}^{(m+1)} = \operatorname{argmin}_{\beta} \|y - V\beta\|_2^2 + \lambda_1 \|\beta\|_{p_2, q_2}^{q_2}$$

Jusqu'à convergence

Algorithme BCR

Théorème (MK,BT 08), utilise les résultats de (Tseng 93) et (Bruce et al. 98)

Soit $1 \leq p_1, q_1$ et $1 \leq p_2, q_2$,

$$\Psi(\alpha, \beta) = \|y - U\alpha - V\beta\| + \lambda_1 \|\alpha\|_{p_1, q_1}^{q_1} + \lambda_2 \|\beta\|_{p_2, q_2}^{q_2}$$

Soit \mathcal{U} et \mathcal{V} deux bases orthogonales de \mathbb{R}^N . Soit $\alpha^{(0)} \in \mathbb{R}^N$ et $\beta^{(0)} \in \mathbb{R}^N$ quelconques.

L'algorithme BCR converge vers un minimum global de Ψ .

Itérations de Landweber - 1

Cas général d'une fonctionnelle du type :

$$\min_{\alpha} \left[\Psi(\alpha) := \|y - \mathcal{F}(\alpha)\|_2^2 + \lambda \|\alpha\|_{p,q}^q \right] .$$

où \mathcal{F} est un opérateur linéaire quelconque (ex. convolution).

Pour **découpler** les équations variationnelles, on introduit, pour un γ donné :

$$\Psi^{\text{sur}}(\alpha, \gamma) = \|y - \mathcal{F}(\alpha)\|_2^2 + C \|\alpha - \gamma\|_2^2 - \|\mathcal{F}(\alpha) - \mathcal{F}(\gamma)\|_2^2 + \lambda \|\alpha\|_{p,q}^q$$

avec $C > \|\mathcal{F}^* \mathcal{F}\|$

Ψ^{sur} est **strictement convexe** en α et en γ . Le minimum en α pour γ fixé est donné par :

$$\alpha = \mathcal{S}^{\lambda/C} [(\mathcal{F}^*(y) + C\gamma - \mathcal{F}^*(\mathcal{F}(\gamma))) / C]$$

où $\mathcal{S}^{\lambda/C}$ est l'opérateur de seuillage associé à la norme mixte.

Itérations de Landweber - 2

Pour $\alpha^{(0)}$ quelconque, l'algorithme de seuillage itératif de Landweber est :

$$\alpha^{(m+1)} = \underset{\alpha}{\operatorname{argmin}} \Psi^{\text{sur}}(\alpha, \alpha^{(m)}) \quad (1)$$

$$= \mathcal{S}^{\lambda/C} \left[\left(\mathcal{F}^*(y) + C\alpha^{(m)} - \mathcal{F}^*(\mathcal{F}(\alpha^{(m)})) \right) / C \right] \quad (2)$$

(Daubechies et al. 04) ont montré la convergence pour une norme ℓ_1 .

Pour les normes mixtes ?

Itérations de Landweber - 3

Théorème, convergence faible (MK 08)

Soit $\mathcal{F} : \ell_2(\mathbb{C}) \rightarrow \mathcal{H}$, un opérateur linéaire tq $\|\mathcal{F}^* \mathcal{F}\| < C$. Soit $\mathbf{y} \in \mathcal{H}$. Alors, la suite des $(\alpha^{(n)})$, $\alpha^{(0)}$ quelconque, converge faiblement vers un point fixe, qui est un minimiseur de la fonctionnelle :

$$\Psi(x) = \frac{1}{2} \|y - \mathcal{F}(\alpha)\|_2^2 + \frac{\lambda}{q} \|\alpha\|_{p,q}^q \text{ pour } 1 \leq p, q \leq 2$$

Théorème (MK,08)

Il existe une suite $\xi = (\xi_{g,m})$ de \mathbb{R}_+ telle que

$$\Psi(x) = \|y - \mathcal{F}(\alpha)\|_2^2 + \lambda \|\alpha\|_{p,q}^q \text{ et } \tilde{\Psi}(x) = \|y - \mathcal{F}(\alpha)\|_2^2 + \lambda \sum_{g,m} \xi_{g,m} |\alpha|$$

ont le même minimum atteint au même point α^* . De plus

$$\xi_{g,m} = |\alpha^*|^{p-1} \|\alpha^*\|_{p,q}^q \text{ si } \alpha_{g,m} \neq 0 .$$

Itérations de Landweber - 4

Les théorèmes précédents permettent d'énoncer

Théorème, convergence forte (MK 08)

La suite $\alpha^{(n)}$ des itérées générée par l'algorithme de seuillage itératif de Landweber avec $\alpha^{(0)}$ choisi arbitrairement dans $\ell_2(\mathbb{C})$, converge fortement vers un point fixe qui est un minimiseur de la fonctionnelle Φ .

Bilan

À la parcimonie s'ajoutent

- Des structures à l'intérieur de groupes
- Un modèle en couches additives

De plus

- On peut construire une variante de FOCUSS (Rao et al. 99) pour les normes mixtes et les modèles en couches (MK 08)
- Généraliser les opérateurs de seuillage ou "shrinkage" dans le cas des bases orthogonales (MK,BT 08)

Mais

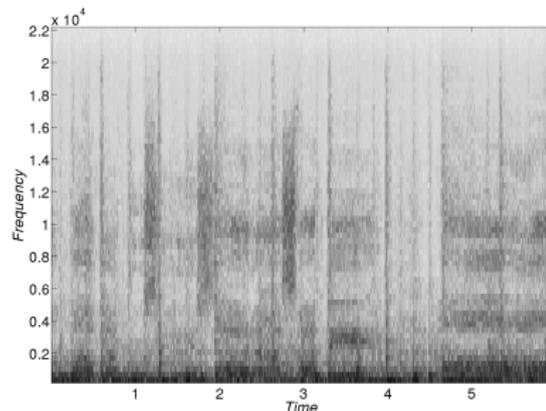
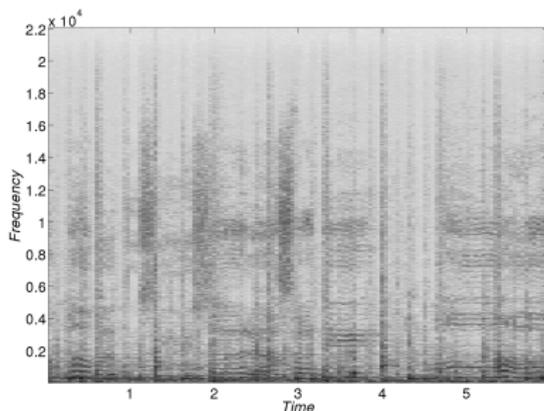
On contrôle mal la parcimonie : le degré de parcimonie intervient à travers les paramètres λ qui sont parfois difficiles à ajuster.

- 1 Introduction : approximation parcimonieuse
- 2 Approches variationnelles
 - Normes mixtes
 - Modèle hybride ou multicouches
 - Algorithmes de seuillage itératif
 - Application à la séparation tonale/transitoire
- 3 Modèles probabilistes
 - Retour sur le modèle hybride
 - Modèles pour les cartes de significances et les coefficients
 - Estimation du modèle
 - Application au débruitage
- 4 Conclusions et perspectives

Séparation tonale/transitoire - 1

Extrait de *Mamavatu* de Susheela Raman. Longueur des fenêtre des Bases MDCT choisies :

- Pour la couche tonale : 4096 échantillons (93 ms) (Gauche)
- Pour la couche transitoire : 128 échantillons (3 ms) (Droite)



Séparation tonale/transitoire - 2

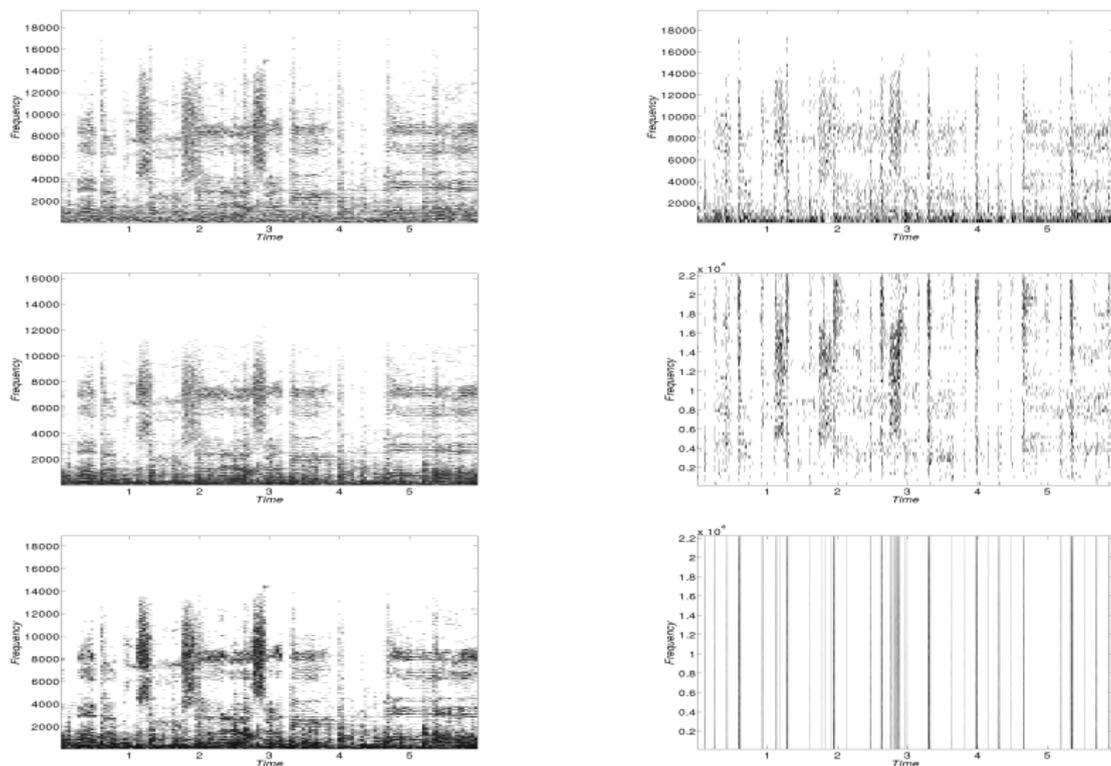


Figure : Gauche : couche tonale. Droite : couche transitoire. De haut en bas : LASSO/LASSO, LASSO/ELASSO, LASSO/GLASSO.

- 1 Introduction : approximation parcimonieuse
- 2 Approches variationnelles
 - Normes mixtes
 - Modèle hybride ou multicouches
 - Algorithmes de seuillage itératif
 - Application à la séparation tonale/transitoire
- 3 Modèles probabilistes
 - Retour sur le modèle hybride
 - Modèles pour les cartes de significances et les coefficients
 - Estimation du modèle
 - Application au débruitage
- 4 Conclusions et perspectives

Formulation du problème

On travaille sur un signal $x \in \mathbb{R}^N$ pouvant se décomposer sous la forme

$$x = \textit{transitoire} + \textit{tonal} + \textit{residu}$$

On cherche alors x sous la forme :

$$x = \sum_n \alpha_n v_n + \sum_m \beta_m u_m$$

où

- $\mathcal{U} = \{u_m\}_{m=1}^N$ est une base “adaptée” à la couche tonale ;
- $\mathcal{V} = \{v_n\}_{n=1}^N$ est une base “adaptée” à la couche transitoire ;

Comment trouver une “bonne” représentation à partir des coefficients d’analyse ?

Modèle hybride

Définition : modèle hybride

Soit $\mathcal{U} = \{u_\delta\}$ et $\mathcal{V} = \{v_\lambda\}$ deux bases de \mathbb{R}^N . On appelle modèle hybride une décomposition de $x \in \mathbb{R}^N$ telle que

$$x = \sum_{\lambda \in \Lambda} \alpha_\lambda v_\lambda + \sum_{\delta \in \Delta} \beta_\delta u_\delta + r$$

où

- Λ et Δ sont appelées **carte de significances**, supposées de “petite taille” (ie de faible cardinalité).
- $r \in \mathbb{R}^N$ est un résidu dont la décomposition dans $\mathcal{D} = \mathcal{U} \cup \mathcal{V}$ n'est pas parcimonieuse.

Modèle hybride aléatoire

Définition : modèle hybride aléatoire

Un modèle hybride aléatoire est un modèle hybride, pour lequel on se donne

- 1 Un modèle probabiliste discret pour Λ et Δ ;
- 2 Un modèle probabiliste pour les coefficients α_λ , $\lambda \in \Lambda$ et β_δ , $\delta \in \Delta$, conditionnellement aux cartes.

- Quelle modèle choisir pour les cartes de signifiante ?
- Quelle modèle choisir pour les coefficients ?

- 1 Introduction : approximation parcimonieuse
- 2 Approches variationnelles
 - Normes mixtes
 - Modèle hybride ou multicouches
 - Algorithmes de seuillage itératif
 - Application à la séparation tonale/transitoire
- 3 Modèles probabilistes
 - Retour sur le modèle hybride
 - Modèles pour les cartes de significances et les coefficients
 - Estimation du modèle
 - Application au débruitage
- 4 Conclusions et perspectives

Modèle de carte : Bernoulli

Le modèle le plus simple est un modèle de Bernoulli :

$$\left\{ \begin{array}{l} n \in \Lambda \text{ avec proba. } p \\ n \notin \Lambda \text{ avec proba. } 1 - p \end{array} \right. \quad \left\{ \begin{array}{l} m \in \Delta \text{ avec proba. } \tilde{p} \\ m \notin \Delta \text{ avec proba. } 1 - \tilde{p} \end{array} \right.$$

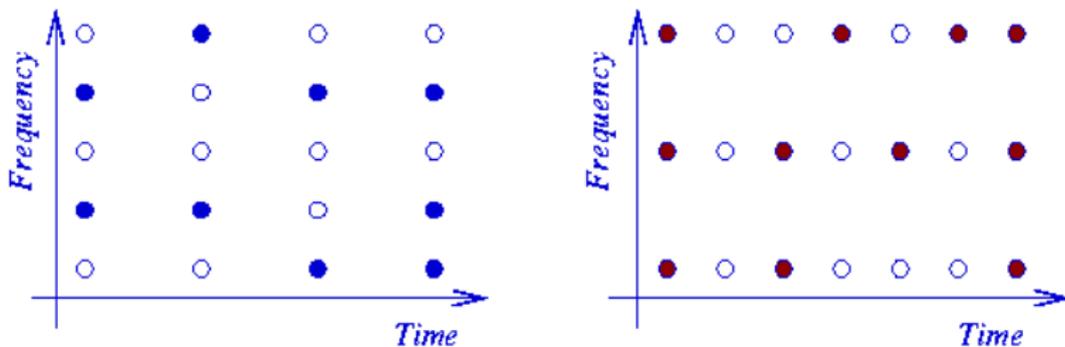


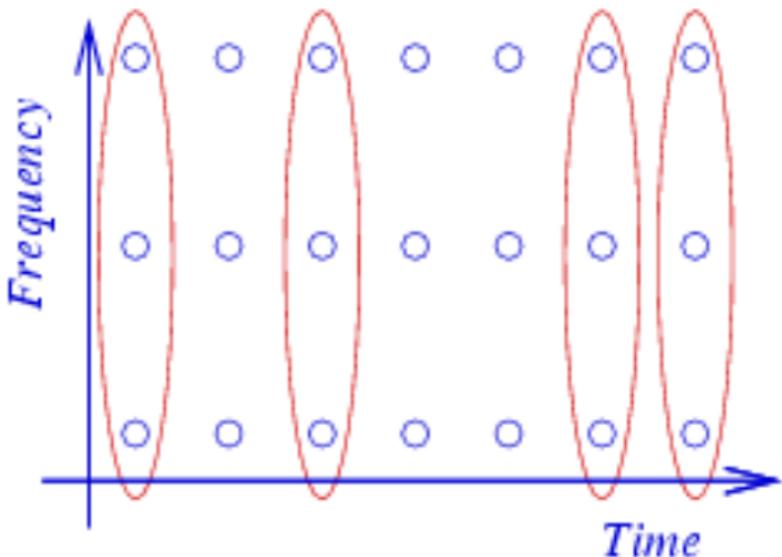
Figure : Exemple de deux distributions temps-fréquence. Gauche : carte avec une bonne résolution fréquentielle. Droite : carte avec une bonne résolution temporelle.

Modèle de carte : Bernoulli-Hierarchique

Une observation expérimentale

Les transitoires sont bien localisés en temps.

Dans un 1er temps, tous les atomes aux temps k sont supposés appartenir à la carte avec la proba. p_1 .



Un modèle de Bernoulli-Hierarchique pour les transitoires

Dans un 2eme temps, afin de préserver la parcimonie, les atomes sélectionnés sont distribués selon une loi de Bernoulli :

Pour $k \in \Lambda$, les atomes à la fréquence ν appartiennent à la carte avec une proba. p_2 .

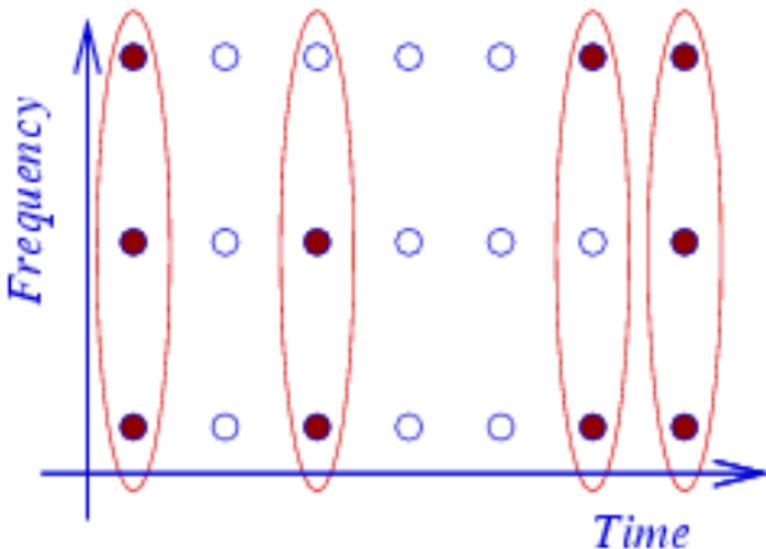


Figure : *Modèle de Bernoulli-Hiérarchique.*

Exemples de représentation d'un signal audio

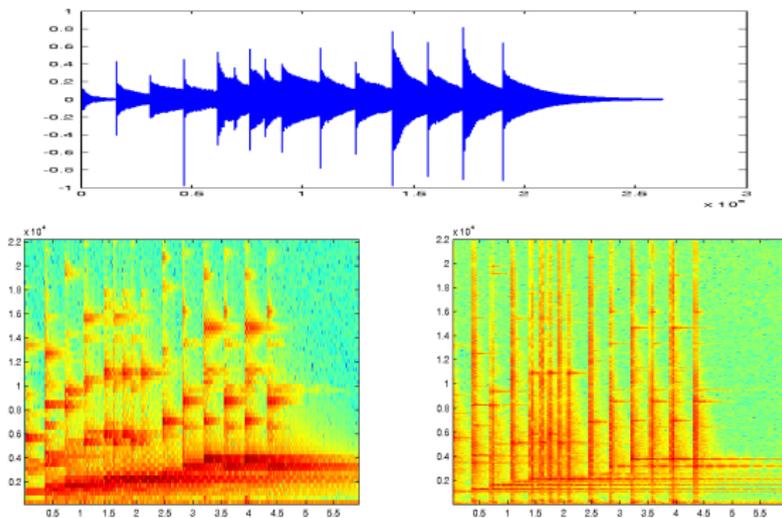


Figure : Images temps-fréquence. Haut : signal, En bas à gauche : représentation adaptée aux transitoires. En bas à droite, représentation adaptée aux tonals

Modèle - Coefficients

Coefficients i.i.d. Gaussiens

Les coefficients de *synthèse* α_λ et β_δ sont des V.A. i.i.d. suivant une loi normale

$$\alpha_\lambda | \lambda \in \Lambda \sim \mathcal{N}(0, \sigma_\lambda^2) \quad \text{et} \quad \beta_\delta | \delta \in \Delta \sim \mathcal{N}(0, \tilde{\sigma}_\delta^2) .$$

Prise en compte de la décroissance en fréquence

n étant un indice temps-fréquence, on note $n = (t, f)$. On suppose

$$\sigma_{tf} = \sigma \nu(f)^\gamma ,$$

où ν est une fonction qui fixe un profil fréquentiel.

Résidu

Le résidu est modélisé comme un bruit blanc gaussien $\mathcal{N}(0, \sigma_0^2)$.

- 1 Introduction : approximation parcimonieuse
- 2 Approches variationnelles
 - Normes mixtes
 - Modèle hybride ou multicouches
 - Algorithmes de seuillage itératif
 - Application à la séparation tonale/transitoire
- 3 Modèles probabilistes
 - Retour sur le modèle hybride
 - Modèles pour les cartes de significances et les coefficients
 - **Estimation du modèle**
 - Application au débruitage
- 4 Conclusions et perspectives

Coefficients d'analyse

Rappel du modèle

$$x = \sum_{\lambda \in \Lambda} \alpha_{\lambda} v_{\lambda} + \sum_{\delta \in \Delta} \beta_{\delta} u_{\delta} + r$$

- $\{v_n\}$ et $\{u_m\}$ sont deux bases ; Λ et Δ sont les cartes de significances
- α_{λ} et β_{Δ} sont les **coefficients de synthèse**.

Coeff. de synthèse VS Coeff. d'analyse

On travaille sur les **coefficients d'analyse** :

$$b_n = \langle x, u_n \rangle$$

$$= \text{contribution de } \left(\sum_{\lambda \in \Lambda} \alpha_{\lambda} v_{\lambda} + r \right) + \begin{cases} \beta_n & \text{si } n \in \Delta \\ 0 & \text{sinon} \end{cases}$$

Comment estimer les paramètres du modèle à partir des coefficients d'analyse ?

Analyse des Coefficients d'analyse

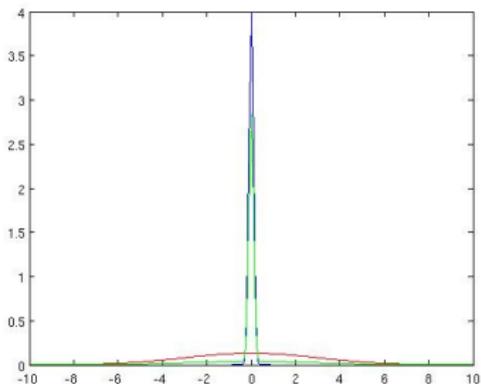
Une approche de la distribution des b_n

Conditionnellement à la carte Λ

$$b_n \sim p\mathcal{N}(0, \tilde{\sigma}_n^2 + \sigma^2\gamma_n(\Lambda) + \sigma_0^2) + (1 - p)\mathcal{N}(0, \sigma^2\gamma_n(\Lambda) + \sigma_0^2)$$

$$\text{où } \gamma_n(\Lambda) = \sum_{\lambda \in \Lambda} \nu_n^2 \langle u_n, v_\lambda \rangle^2.$$

Les $\gamma_n(\Lambda)$ sont des V.A. qui mesurent “l’influence” des “atomes actifs” de la base \mathbf{V} sur u_n .



Les poids γ

Dans un premier temps, on suppose $\sigma_n = \sigma \forall n$.

Proposition : moments des poids γ (MK, BT 08)

$$\mathbb{E}_\Lambda\{\gamma_n(\Lambda)\} = p$$

$$\text{Var}_\Lambda\{\gamma(\Lambda)\} = p(1-p) \sum_m \langle u_n, v_m \rangle^4$$

La variance est contrôlée par la fonction de Babel-4 :

$$B_4 = \sup_n \sum_m \langle u_n, v_m \rangle^4.$$

Les poids γ

Soit \mathbf{M} l'opérateur diagonal tel que $\mathbf{M}v_n = \nu_n v_n$. On peut calculer les premiers moments des poids γ (MK, BT 08)

$$\mathbb{E}_\Lambda\{\gamma_n(\Lambda)\} = p\|\mathbf{M}u_n\|^2$$

$$\text{Var}_\Lambda\{\gamma(\Lambda)\} = p(1-p) \sum_m \langle u_n, \mathbf{M}v_m \rangle^4$$

La variance est contrôlée par la fonction de Babel-4 :

$$B_4 = \sup_n \sum_m \langle u_n, \mathbf{M}v_m \rangle^4.$$

Estimations pratiques

- ① **Estimation des Cartes** L'estimation des deux cartes se fait *indépendamment*.
 - ① **Estimation des paramètres.** La distribution des coefficients est un mélange de Gaussienne, qu'on peut alors estimer par un algorithme EM
 - ② **Classification des coefficients d'analyse.** Les cartes sont alors simplement estimées par classification des coefficients en utilisant le Maximum A Posteriori (\Leftrightarrow **seuillage adaptatif**)
- ② **Estimation des coefficients de synthèse.**

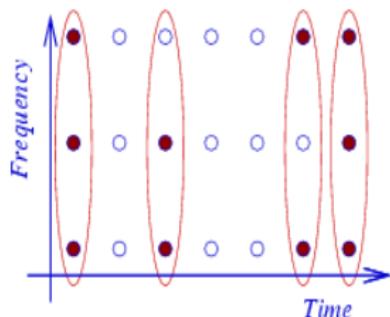
Comment améliorer l'estimation grâce au modèle Bernoulli-Hiérarchique pour la carte transitoire ?

Estimation de la carte transitoire

On distingue le temps et la fréquence : $\Lambda = \Lambda_t \times \Lambda_f$, et on travaille sur les coefficients “temporels” suivant :

$$c_k = \sum_{\nu=1}^{\Lambda_f} a_{k,\nu}^2$$

$$= \begin{cases} \sum_{\nu=1}^{|\Lambda_f|} \alpha_{k,\nu} X_{k,\nu} + \tilde{\pi}_{k,\nu}(\Delta) & \text{si } k \in \Lambda_t \\ \sum_{\nu=1}^{|\Delta_f|} \tilde{\pi}_{k,\nu}(\Delta) & \text{si } k \notin \Lambda_t \end{cases}$$



où $\tilde{\pi}_{k,\nu}(\Delta) \sim \mathcal{N}(0, \tilde{\gamma}_{k,\nu}(\Delta)\tilde{\sigma}^2 + \sigma_0^2)$,

\Rightarrow on approche la distribution des c_k , $k \notin \Lambda$ par une distribution du χ^2 .

Une détection en deux étapes

- 1 Une détection d'outliers permet d'estimer les lignes temporelles de la carte de signifiante (1ere étape du modèle Bernoulli-Hiérarchique).
- 2 Les coefficients restant sont discriminés par les coefficients d'analyse a_n .

Estimations pratiques

- ① **Estimation des Cartes** L'estimation des deux cartes se fait *indépendamment*.

- ① *Estimation de la carte Tonale*

- ① **Estimation des paramètres.** La distribution des coefficients est un mélange de Gaussienne, qu'on peut alors estimer par un algorithme EM
- ② **Classification des coefficients d'analyse.** Les cartes sont alors simplement estimées par classification des coefficients en utilisant le Maximum A posteriori

- ② *Estimation de la carte Transitoire*

- ① Détection d'outliers pour trouver les coefficients temporels ;
- ② On finit l'estimation comme pour la carte Tonale

- ② **Estimation des coefficients de synthèse.**

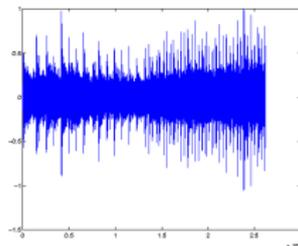
Estimations pratiques

- ① **Estimation des Cartes** L'estimation des deux cartes se fait *indépendamment*.
 - ① *Estimation de la carte Tonale*
 - ① **Estimation des paramètres.** La distribution des coefficients est un mélange de Gaussienne, qu'on peut alors estimer par un algorithme EM
 - ② **Classification des coefficients d'analyse.** Les cartes sont alors simplement estimées par classification des coefficients en utilisant le Maximum A posteriori
 - ② *Estimation de la carte Transitoire*
 - ① Détection d'outliers pour trouver les coefficients temporels ;
 - ② On finit l'estimation comme pour la carte Tonale
- ② **Estimation des coefficients de synthèse.** Les coefficients $\hat{\alpha}$ et $\hat{\beta}$ sont obtenus via une *projection orthogonale* de x sur l'espace engendré par $\hat{\mathcal{D}} = \{v_\lambda, \lambda \in \hat{\Lambda}\} \cup \{u_\delta, \delta \in \hat{\Delta}\}$.

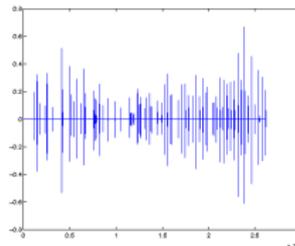
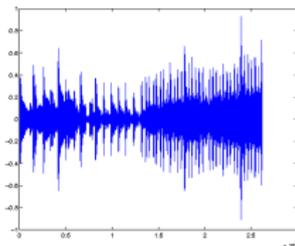
- 1 Introduction : approximation parcimonieuse
- 2 Approches variationnelles
 - Normes mixtes
 - Modèle hybride ou multicouches
 - Algorithmes de seuillage itératif
 - Application à la séparation tonale/transitoire
- 3 Modèles probabilistes
 - Retour sur le modèle hybride
 - Modèles pour les cartes de significances et les coefficients
 - Estimation du modèle
 - Application au débruitage
- 4 Conclusions et perspectives

Xylophone + bruit gaussien

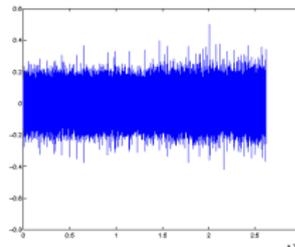
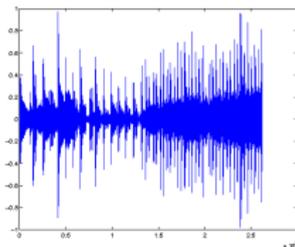
Signal :



Couches tonale et transitoire :

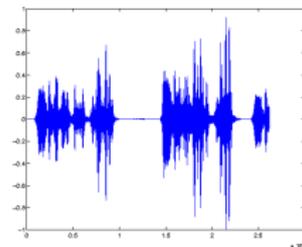


Reconstruction et résidu :

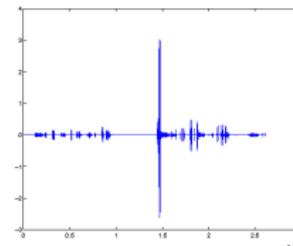
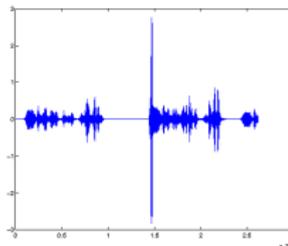


Flûte de Pan

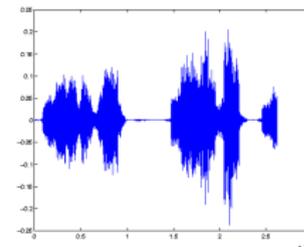
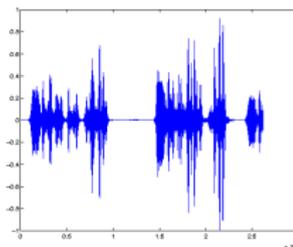
Signal :



Couches tonale et transitoire :

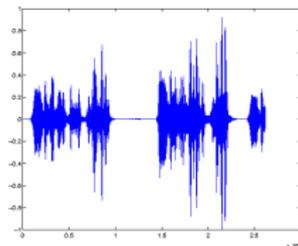


Reconstruction et résidu :

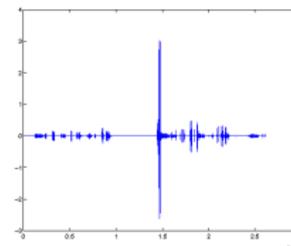
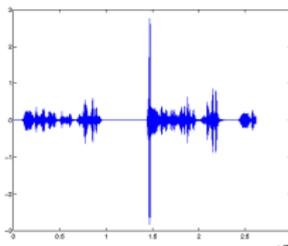


INA 1

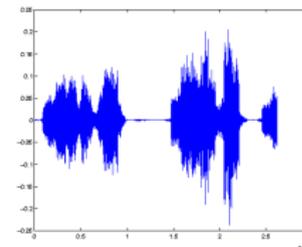
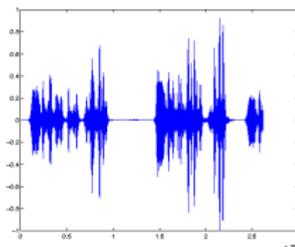
Signal :



Couches tonale et transitoire :

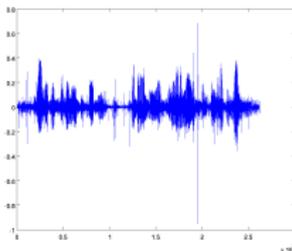


Reconstruction et résidu :

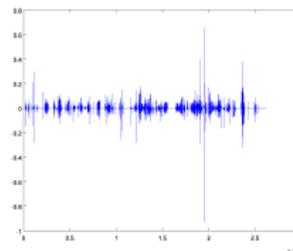
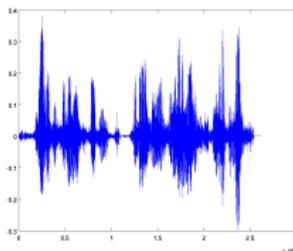


Signal de l'INA (Merci à Vincent Fromont)

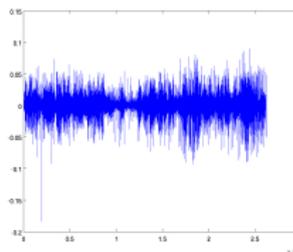
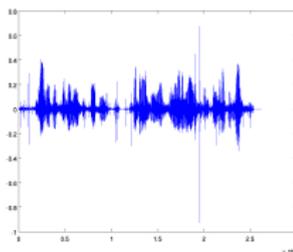
Signal :



Couches tonales et transitoires :



Reconstruction et résidu



- 1 Introduction : approximation parcimonieuse
- 2 Approches variationnelles
 - Normes mixtes
 - Modèle hybride ou multicouches
 - Algorithmes de seuillage itératif
 - Application à la séparation tonale/transitoire
- 3 Modèles probabilistes
 - Retour sur le modèle hybride
 - Modèles pour les cartes de significances et les coefficients
 - Estimation du modèle
 - Application au débruitage
- 4 Conclusions et perspectives

Conclusions

Approche variationnelle

- Introduction de structures via les normes mixtes
- Opérateurs de seuillage “généralisé”
- Plusieurs algorithmes itératifs (avec preuve de convergence)

Modèles aléatoires

- Construction d'un modèle *a priori* prenant en compte :
 - La parcimonie ;
 - Les structures en couches ;
 - Un modèle structuré pour la partie transitoire.
- Estimation des cartes à partir des coefficients d'analyse
 - Algorithme EM
 - Test d'adéquation

Quelques perspectives

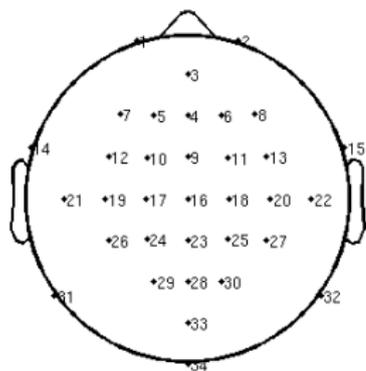
Perspectives théoriques

- Modèle(s) de Bernoulli multicapteurs
- Liens modèles proba/normes mixtes
- Autres modèles aléatoires (Markov, Ising . . .)

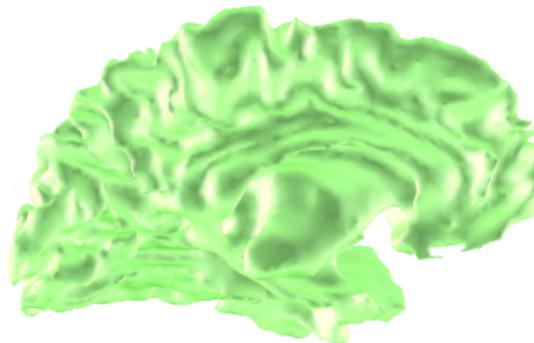
Perspectives d'applications pratiques

- Application à la séparation de sources
- Application aux problèmes inverses en MEEG

Problème inverse en MEEG



? ↓?



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

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

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

sources : $\mathbf{s} \in \mathbb{R}^{N \times T}$

$$\mathbf{y} = A\mathbf{s}$$

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

Problème très sous-déterminé : besoin d'*a priori* très forts sur \mathbf{s}

Piste étudiée avec A. Gramfort :

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \|\mathbf{y} - A\alpha\Phi\|_2^2 + \lambda P(\alpha)$$

Avec P norme mixte "3 niveaux" mélangeant G-LASSO et E-LASSO.