



HAL
open science

Sparse Wars

Matthieu Kowalski

► **To cite this version:**

Matthieu Kowalski. Sparse Wars. Traitement du signal et de l'image [eess.SP]. Université paris sud, 2018. tel-01868307

HAL Id: tel-01868307

<https://theses.hal.science/tel-01868307>

Submitted on 20 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris-Sud
Laboratoire des Signaux et Systèmes

Habilitation à Diriger des Recherches

Sparse Wars

Matthieu Kowalski

11 Juin 2018

Jury:

Christine De Mol	<i>Professeure de l'Université Libre de Bruxelles</i>	Rapporteur
Jérôme Idier	<i>Directeur de Recherche CNRS</i>	Rapporteur
Gabriel Peyré	<i>Directeur de Recherche CNRS</i>	Rapporteur
Francis Bach	<i>Directeur de Recherche INRIA, Professeur associé de l'ENS</i>	Président du Jury
Rémi Gribonval	<i>Directeur de recherche INRIA</i>	Examinateur
Michel Kieffer	<i>Professeur de l'Université Paris-Sud</i>	Examinateur



Table des matières

I Sparse Wars	7
1 A long long time ago...	9
1.1 Cadre général : les problèmes inverses	9
1.1.1 Définition <i>a contrario</i>	9
1.1.2 Approche par optimisation et utilisation de dictionnaire	10
1.1.2.1 Formulation générale	10
1.1.2.2 Choix de l'attache aux données	10
1.1.2.3 Choix de la pénalité	10
1.1.2.4 Utilisation d'un dictionnaire	10
1.1.3 Approche Bayésienne	12
1.2 Résumé de mes contributions	13
1.2.1 Organisation du document	13
1.2.2 Autres contributions	14
2 L-One : a sparse war story	17
2.1 Un résumé historique de la parcimonie	17
2.1.1 Du seuillage au ℓ_1	17
2.1.2 Résolution du Basis Pursuit Denoising par seuillage itératif	18
2.1.3 Du ℓ_1 au ℓ_0	19
2.2 Structures et problèmes inverses	19
2.2.1 Parcimonie pour les problèmes inverses	19
2.2.2 Pénalités structurantes	20
2.2.3 Modèles en couches	21
2.3 Frise chronologique	22
3 Revenge of the IST	23
3.1 Descente proximale	24
3.1.1 Algorithme	24
3.1.2 Opérateurs proximaux et fonctions de seuillage	25
3.2 Variantes par blocs	26
4 Bayes strikes back	29
4.1 Un modèle génératif parcimonieux : le Bernoulli-Gaussien	30
4.1.1 Modèle et débruitage	30
4.1.2 Choix optimaux du seuil	31
4.1.3 Choix pratique du seuil en débruitage	37
4.1.4 Un algorithme de type IST pour le modèle BG	37

Table des matières

4.1.5	Problèmes inverses : du Bernoulli-Gaussien au Sparse Bayesian Learning	39
4.2	Synthèse temps-fréquence de rang faible	39
4.2.1	De la NMF au LRTFS	40
4.2.2	Compressive LRTFS	44
5	Return of Shrinkage	47
5.1	Fonctions de seuillage et algorithme de seuillage itératif	48
5.1.1	Exemples d'opérateurs de seuillage	49
5.1.2	IST avec un opérateur de seuillage quelconque	50
5.2	Parcimonie Sociale	51
5.3	Différents opérateurs de seuillage	53
6	Applications awaken	55
6.1	Signaux audio	56
6.1.1	Débruitage	56
6.1.2	Dé-saturation	58
6.2	Localisation de sources M/EEG	58
6.2.1	Modèle direct et problème inverse	58
6.2.2	Résultats	60
6.2.2.1	Simulations	60
6.2.2.2	Signaux réels	60
7	To be continued . . .	65
7.1	Optimisation, problèmes inverses et approches Bayésiennes	65
7.2	Apprentissage de dictionnaire par LRTFS	66
7.3	Applications aux signaux audios	66
7.4	Applications aux signaux M/EEG et neuroimagerie	67
7.5	Déconvolution aveugle des données PFS	67
II	Notice individuelle	79
8	Résumé	81
8.1	Synthèse des activités	81
8.2	Titres universitaires	82
8.3	Parcours	83
9	Activités d'enseignement	85
9.1	Description des interventions principales	85
9.1.1	Traitement du Signal	85
9.1.2	Algorithmique/C	85
9.1.3	Probabilité/Statistiques	85
9.2	Tableau récapitulatif	85

Table des matières

10 Activités scientifiques	87
10.1 Comités	87
10.1.1 Comités de pilotage et d'organisation	87
10.1.2 Comités de sélection	87
10.1.3 Jurys de thèse	87
10.1.4 Projets financés	88
10.1.5 Rayonnement	89
10.1.5.1 Chercheur invité	89
10.1.5.2 Orateur invité	89
10.2 Évaluations	90
10.2.1 Relectures	90
10.2.2 Expertises de projets	90
11 Encadrements	91
11.1 Post-doctoral	91
11.2 Doctoral	91
11.2.1 Thèses soutenues	91
11.2.2 Thèses en cours	92
11.3 Master	92
11.3.1 Master 2	92
11.3.2 Master 1	92
11.4 Licence 3	93
12 Liste des publications	95

Résumé

Il y a bien longtemps, dans une galaxie très, très lointaine. L'ordre des GPi développe un nouvel art de résolution des problèmes inverses : la synthèse parcimonieuse. Nous présenterons ici une forme primordiale de cet art : l'algorithme de seuillage itératif combiné à n'importe quelle règle de seuillage. Allié à la parcimonie sociale, permettant une prise en compte plus perméable des structures, nous verrons comment il éclipse la parcimonie simple de type Basis Pursuit (Denoising) dans la résolution de problèmes inverses. Puis nous méditerons sur ses liens avec l'art Bayésien afin d'ériger une nouvelle forme de structure fondée sur la synthèse parcimonieuse de rang faible, permettant d'interpréter la décomposition en matrice non négative comme un nouvel a priori. Nous mesurerons également la force de ces approches sur un problème particulier : la désaturation audio. Nous illustrerons aussi les succès de ces approches sur d'autres problèmes inverses tels que la séparation de sources ou la localisation de sources M/EEG.

Première partie .

Sparse Wars

1 ■ A long long time ago...

Les méthodes étudiées et développées au cours de ma thèse ont trait à l'approximation parcimonieuse et structurée de signaux. Depuis la fin de ma thèse et durant mon année d'ATER, j'ai élargi mes activités de recherches aux problèmes inverses, ce qui m'a permis d'intégrer le GPI (Groupe Problèmes Inverses) du L2S. Afin d'aborder de tels problèmes, j'exploite les caractéristiques des représentations parcimonieuses à la synthèse des signaux dans des dictionnaires redondants, pouvant être des dictionnaires d'atomes temps-fréquence, d'ondelettes ou autres. Cette hypothèse de parcimonie est prise en compte dans des modèles d'optimisation, convexe ou non, qui permettent d'obtenir l'estimation d'un ou plusieurs signaux comme minimum d'une fonction de coût.

1.1. Cadre général : les problèmes inverses

1.1.1. Définition *a contrario*

D'un point de vue de traicteur de signal, un problème inverse consiste à estimer un signal originel à partir d'observations souvent incomplètes et bruitées d'un ou plusieurs signaux liés physiquement. Un modèle mathématique souvent utilisé est alors le suivant. Soient $\mathbf{y} \in \mathbb{R}^M$, $\mathbf{x} \in \mathbb{R}^N$ et soit $\mathcal{A} : \mathbb{R}^N \rightarrow \mathbb{R}^M$, un opérateur, tels que :

$$\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{e}, \quad (1.1)$$

où $\mathbf{e} \in \mathbb{R}^M$ représente un bruit additif quelconque. Dans ce modèle, les observations sont représentées par \mathbf{y} . Dans beaucoup de problèmes, l'opérateur \mathcal{A} peut être modélisé : c'est le *problème direct*. Le problème inverse se définit donc *a contrario* [DEM+01], le but étant d'estimer \mathbf{x} . Cependant, \mathcal{A} n'est pas toujours connu, et peut même devenir l'objet d'intérêt, connaissant \mathbf{x} et \mathbf{y} . Lorsque ni \mathcal{A} ni \mathbf{x} ne sont connus, on parle de problème *aveugle*. Dans la majorité des cas, on supposera l'opérateur \mathcal{A} linéaire, que l'on notera simplement à l'aide de sa matrice $\mathbf{A} \in \mathbb{R}^{MN}$. Le modèle (1.1) s'écrit alors

$$\mathbf{y} = \mathbf{Ax} + \mathbf{e}. \quad (1.2)$$

Enfin, on distinguera aussi deux cas importants : le cas sur-déterminé, lorsque $M > N$ et le cas sous-déterminé lorsque $M < N$.

De façon générale, beaucoup d'applications peuvent se formuler comme un problème inverse : le but est d'estimer des données à partir d'observations. Cependant, les problèmes

1. A long long time ago...

inverses auxquels on s'intéresse sont mal posés. Si l'on se réfère à Jacques Hadamard, un problème est bien posé s'il vérifie les trois conditions :

1. l'existence de la solution,
2. l'unicité de la solution,
3. la stabilité de la solution.

Le but d'un problème inverse est alors de trouver une méthode assurant ces trois conditions.

1.1.2. Approche par optimisation et utilisation de dictionnaire

1.1.2.1. Formulation générale

Un cadre pratique que j'ai largement utilisé lors de mes travaux pour aborder les problèmes inverses consiste à rechercher la solution d'un problème d'optimisation :

$$\hat{\mathbf{x}} = \operatorname{argmin} \mathcal{L}(\mathbf{y}, \mathbf{A}, \mathbf{x}) + \lambda \Omega(\mathbf{x}), \quad (1.3)$$

où

- \mathcal{L} est un terme d'attache aux données modélisant le lien entre les observations \mathbf{y} et les signaux \mathbf{x} à travers l'opérateur \mathbf{A} .
- Ω est un terme de régularisation, qui reflète un *a priori* sur les données.
- $\lambda > 0$ est un hyperparamètre permettant de balancer l'importance entre l'attache aux données et la régularisation.

1.1.2.2. Choix de l'attache aux données

Parmi les choix classiques pour l'attache aux données \mathcal{L} , on trouve la norme ℓ_2 , permettant de prendre en compte les bruits blancs gaussiens ; la norme ℓ_1 , plus robuste, mieux adaptée aux bruits impulsionnels ; La divergence de Kullback-Leibler, adaptée aux bruits de type Poisson. Dans ce manuscrit, on utilisera essentiellement l'attache aux données de type norme ℓ_2 .

1.1.2.3. Choix de la pénalité

La pénalité Ω modélise les *a priori*, physiques ou mathématiques, sur le signal à estimer. La pénalité la plus classique est la régularisation de Tikhonov [TAJ77] qui force l'estimation d'un signal d'énergie minimale. En image, une pénalité classique est la variation totale, qui consiste à mesurer la norme ℓ_1 de son gradient. Ce choix favorise les images constantes par morceaux.

1.1.2.4. Utilisation d'un dictionnaire

Le choix d'une régularisation sur le signal \mathbf{x} n'est pas toujours aisé. Il est alors utile de choisir un *dictionnaire* permettant de mettre en évidence les caractéristiques d'intérêt d'un signal. L'exemple le plus simple et le plus connu de dictionnaire est la représentation dans

1. A long long time ago...

une base de Fourier, qui permet d'avoir une représentation fréquentielle du signal. La notion de dictionnaire peut être confondue avec la notion de *frames*, en dimension infinie, afin d'assurer la stabilité d'analyse et de reconstruction d'un signal. La dimension finie nous permet de confondre un dictionnaire et sa matrice formée par ses éléments, que l'on nommera atomes [MZ93]. Si l'on note $\mathcal{D} = \{\boldsymbol{\varphi}_k\}_{k=1}^K$ le dictionnaire formé par les atomes $\boldsymbol{\varphi}_k \in \mathbb{C}^K$, on peut associer deux opérateurs à ce dictionnaire :

- L'opérateur d'*analyse* Φ^* tel que

$$\Phi^* \mathbf{x} = \{\langle \mathbf{x}, \boldsymbol{\varphi}_k \rangle\}_{k=1}^K.$$

- L'opérateur de *synthèse* Φ tel que

$$\Phi \boldsymbol{\alpha} = \sum_{k=1}^K \alpha_k \boldsymbol{\varphi}_k.$$

Si $\Phi^* \Phi = \Phi \Phi^* = \mathbf{I}_N$, alors le dictionnaire se réduit à une base orthonormée.

Au-delà de Fourier, de nombreux dictionnaires ont été construits. On pourra se référer à MALLAT [Mal08] pour les exemples suivants :

- Les transformées en cosinus discrètes modifiées (MDCT, pour Modified Discrete Cosine Transform).
- Les dictionnaires de Gabor, ou transformée de Fourier à court terme, encore appelés transformée de Fourier à fenêtre glissante. Ces dictionnaires permettent d'avoir une représentation temps-fréquence du signal.
- Les dictionnaires de Gabor non stationnaires [Bal+11] et en particulier les dictionnaires de Gabor à Q-constant [Hol+13].
- Les bases d'ondelettes et les dictionnaires d'ondelettes non décimés, qui permettent d'avoir une base orthonormée sensible aux singularités présentes dans un signal.
- Les variations autour des ondelettes, développées essentiellement en image, qui permettent d'être sensible aux contours ainsi qu'à leurs orientations. On pourra citer de façon non exhaustive les curvlets [Can+06], les bandelets [MP07], etc.¹

Un cadre théorique pratique de construction des dictionnaires est celui des *frames* [Dau92]. L'article [Bal+13] offre un point de vue général sur les possibilités offertes avec les dictionnaires temps-fréquence.

Le cadre par optimisation (1.3) s'adapte parfaitement à l'approche *synthèse*. Il est en effet pratique de modéliser le signal d'intérêt \mathbf{x} comme une somme *d'atomes élémentaires* choisis dans un dictionnaire, tel que $\mathbf{x} = \Phi \boldsymbol{\alpha}$, où Φ est l'opérateur de synthèse associé au dictionnaire et $\boldsymbol{\alpha}$ les coefficients de synthèse. Le problème d'optimisation devient alors

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \mathcal{L}(\mathbf{y}, \mathbf{A}, \Phi \boldsymbol{\alpha}) + \Omega(\boldsymbol{\alpha}), \quad (1.4)$$

et une estimation de \mathbf{x} est $\hat{\mathbf{x}} = \Phi \hat{\boldsymbol{\alpha}}$. Le terme de régularisation porte alors directement sur les coefficients de synthèse du signal. Lorsque le dictionnaire est bien choisi, le choix d'un *a priori* de parcimonie de type ℓ_1 permet de bonnes estimations. Le chapitre 2 donne un bref aperçu des possibilités pour le choix de la régularisation Ω .

1. Voir <http://www.laurent-duval.eu/siva-wits-where-is-the-starlet.html> pour trouver l'*-let idéale.

1.1.3. Approche Bayesienne

Dans le cadre Bayesien, l'estimation par "probabilité inverse" [Rob06] permet de modéliser les incertitudes de chaque paramètre dans un cadre probabiliste. Ainsi, tout peut être vu comme un variable aléatoire modélisée par une loi de probabilité, ce qui correspond au choix du *prior*. Ainsi, en plus de l'équation du modèle direct (1.2), le modèle se complète par les choix

- D'une distribution *a priori*, appelée prior par la suite, sur α : $p(\alpha)$.
- D'un prior sur \mathbf{e} : $p(\mathbf{e})$, qui permet d'obtenir la densité de \mathbf{y} conditionnellement à α : $p(\mathbf{y}|\alpha)$.

L'inférence bayesienne consiste alors à estimer la densité *a posteriori* de notre variable d'intérêt, en prenant en compte les incertitudes modélisées par les priors. Cette densité *a posteriori* est donnée par le théorème de Bayes :

$$p(\alpha|\mathbf{y}) = \frac{p(\mathbf{y}|\alpha)p(\alpha)}{p(\mathbf{y})}.$$

De cette densité *a posteriori*, modélisant elle-même l'incertitude sur l'estimation, on peut construire deux estimateurs ponctuels :

1. L'espérance *a posteriori* : $\hat{\alpha} = E[p(\alpha|\mathbf{y})]$, qui a l'avantage de minimiser l'erreur quadratique moyenne. Cet estimateur est souvent appelé *estimateur MMSE* (Minimum Mean Squared Error).
2. Le Maximum A Posteriori (MAP) $\hat{\alpha} = \operatorname{argmax}_{\alpha} p(\alpha|\mathbf{y})$, souvent plus simple à calculer que l'estimateur MMSE.

On montre facilement que l'approche par optimisation (1.4) se confond avec l'estimateur du MAP dans le cadre Bayesien. Si l'on choisit les prior suivants

1. $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_0 \mathbf{I}_M)$: $p(\mathbf{y}|\alpha) = \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{\|\mathbf{y} - \mathbf{A}\Phi\alpha\|^2}{2\sigma_0^2}}$,
2. $p(\alpha) = e^{-\gamma\Omega(\alpha)}$,

alors l'estimateur du MAP peut s'écrire comme

$$\begin{aligned} \hat{\alpha} &= \operatorname{argmax}_{\alpha} p(\alpha|\mathbf{y}) = \operatorname{argmax}_{\alpha} p(\mathbf{y}|\alpha)p(\alpha) = \operatorname{argmin}_{\alpha} -\log p(\mathbf{y}|\alpha) - \log p(\alpha) \\ &= \operatorname{argmin}_{\alpha} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\Phi\alpha\|^2 + \lambda\Omega(\alpha), \end{aligned} \quad (1.5)$$

avec $\lambda = \gamma/\sigma_0^2$.

Si le lien entre le choix du prior, l'estimateur MAP et le choix de la régularisation dans l'approche par optimisation est évidente et permet souvent de confondre les deux approches, il est montré dans [GM13] qu'il existe aussi un choix de régularisation Ω permettant de confondre l'estimateur MMSE et l'estimateur par optimisation. De fait, il faut se méfier des interprétations Bayésiennes *a posteriori*. Dans le formalisme Bayesien, le prior modélise l'incertitude que l'on a sur les données. Il y a donc une certaine liberté dans le choix de ce prior mais il est dangereux d'interpréter ce prior comme une vérité sur la distribution des données. Ainsi, l'interprétation de la pénalité ℓ_1 comme un prior Laplacien doit

1. A long long time ago...

être nuancée. S'il est peut-être juste de considérer que l'incertitude sur la connaissance des coefficients peut se modéliser par une Laplacienne (tout comme elle pourrait se modéliser par une autre loi telle qu'une loi de Student-t, un mélange de Gaussiennes, etc.), il ne faut pas en conclure que les coefficients de synthèse sont des réalisations issues d'une Laplacienne comme souligné dans [Sta+13] et analysé dans [Nik07]. Dans un formalisme *purement Bayésien*, si le but est d'obtenir un estimateur efficace au sens du MMSE, l'estimateur du MAP est inadapté pour les problèmes inverses. En effet, dans un contexte où le nombre de paramètres croît en fonction du nombre d'observations, l'équivalence entre l'estimateur du MAP et l'estimateur du maximum de vraisemblance ne s'applique pas. Cette considération sur la définition d'un estimateur de Bayes confirme l'analyse faite dans [GM13] sur l'interprétation d'un prior choisi dans un cadre d'optimisation.

1.2. Résumé de mes contributions

J'ai choisi de ne présenter dans ce document que mes contributions sur la résolution des problèmes inverses à l'aide de modèles parcimonieux à la synthèse, ainsi que deux applications que je trouve significatives par les résultats obtenus. Ces contributions sont résumées dans la section suivante qui présente également la structure du document. Mes autres contributions, en particulier celles qui ont fait l'objet d'un encadrement, sont résumées dans la section 1.2.2.

1.2.1. Organisation du document

Dans le chapitre 2, on retrace un historique du L1 et de la parcimonie dans le contexte du traitement du signal et des problèmes inverses. On y voit les évolutions des pénalités qui permettent de prendre en compte les structures apparaissant dans les signaux. On se focalise en particulier sur l'approche *synthèse*.

Le chapitre 3 expose l'approche par seuillage itératif, qui sera beaucoup utilisé par la suite, afin de résoudre les problèmes inverses avec prise en compte de la parcimonie.

Le chapitre 4 présente dans un cadre Bayésien le modèle "idéal" de parcimonie Bernoulli Gaussien, généralisé par le Sparse Bayesian Learning (SBL). Dans ce cadre, on analyse, dans un premier temps, le seuillage dur pour le débruitage. Le modèle du SBL, combiné au modèle composite Gaussien utilisé comme modèle génératif dans les approches NMF (Non Negative Matrix Factorization), permet de définir un cadre probabiliste permettant d'interpréter la NMF comme un prior Low Rank.

Le chapitre 5 revient sur l'algorithme de seuillage itératif. On montre comment utiliser directement une fonction de seuillage plutôt que de modéliser la pénalité sur les coefficients de synthèse. On étudie ensuite des fonctions de seuillage prenant en compte une structure de voisinage, qui sera appelé *social sparsity*.

Le chapitre 6 donne un aperçu des applications aux problèmes inverses auxquelles je me suis intéressé avec d'un côté les applications sur les signaux audio en débruitage et dé-saturation, et de l'autre la localisation de sources magnéto-electro-encephalographie (M/EEG), dans le cadre d'une collaboration sur le long terme avec Alexandre Gramfort.

1.2.2. Autres contributions

Séparation de sources audio. Une grande part de mon activité de recherche récente s'est concentrée sur la séparation de sources audio. Ma première contribution a concerné la séparation des mélanges convolutifs sous-déterminés (non aveugle), avec une approche par synthèse parcimonieuse temps-fréquence. Ce travail a été effectué lors d'une collaboration avec Emmanuel Vincent et Rémi Gribonval [KVG10], puis approfondie de manière expérimentale lors du stage de M1 de Fangchen Feng [FK14a]. Les résultats obtenus montrent qu'utiliser le modèle de mélange convolutif dans le domaine temporel permet d'améliorer grandement la séparation par rapport à l'approche en bande étroite usuelle dans le domaine temps-fréquence. Ces travaux ont été le point de départ de la thèse de Fangchen Feng soutenue en septembre 2017. Fangchen Feng a obtenu plusieurs résultats significatifs en séparation de source. Les trois principaux résultats sont :

1. La démonstration de l'équivalence entre les approches de type ICA (Independent Component Analysis) et les approches parcimonieuses pour la séparation aveugle de mélanges instantanés. L'algorithme ainsi obtenu fonctionne aussi bien sur les mélanges sur-déterminés que sur les mélanges sous-déterminés [FK15a].
2. Le développement d'un algorithme pour les mélanges convolutifs aveugles avec un modèle convolutif en bande étroite. Tout en étant plus robuste au bruit, l'algorithme permet un gain de 1dB à 2dB en terme de SDR/SIR/ISR par rapport à l'état de l'art.
3. L'algorithme STRAUSS (Sparsity and low-Rank Amplitude based Source Separation) qui exploite les hypothèses de parcimonie et de rang faible pour la séparation aveugle des mélanges convolutifs en bande étroite. STRAUSS permet d'améliorer significativement le SDR et le SIR sur les signaux de musique [FK17].

La thèse de Fangchen Feng a été soutenue en mai 2017.

Publications associées : [KVG10; FK14a; FK15a; FK17]

Séparation harmonique/percussif. La thèse de Clément Laroche étudie le problème de séparation des sources percussives des sources harmoniques (où harmonique est employé par abus de langage, par opposition aux sources percussives). Plusieurs méthodes de décomposition du spectrogramme des signaux de musique ont été proposées, exploitant les différences de structures dans le domaine temps-fréquence. Ces approches reposent sur la NMF et ses variantes "Projective-NMF" et "Orthogonal-NMF" (voir [Cho08] et les références). Une première décomposition a été proposée [Lar+15a], où les sources harmoniques sont bien identifiées par des composantes "presque" orthogonales dans le spectrogramme tandis que les sources percussives sont capturées par une simple NMF. Plus précisément, si on note V le spectrogramme des sources observées, V_H le spectrogramme des sources harmoniques et V_P le spectrogramme des sources percussives, on cherche une décomposition

$$V \simeq V_H + V_P,$$

telle que V_H est approchée par une décomposition de type Projective-NMF, tandis que V_P est approchée par une simple NMF. Ainsi, on cherche une décomposition telle que

$$V \simeq W_H W_H^T V + W_P A_P.$$

1. A long long time ago...

Cette approche a été améliorée en proposant un apprentissage de la matrice W_P qui peut être vue comme un dictionnaire contenant des spectrogrammes élémentaires de sources percussives. W_P a été construit à partir de dictionnaires spécifiques de batteries en appliquant une simple NMF sur une partie de la base "ENST-Drum" [Lar+16]. Cette approche a ensuite été améliorée par une décomposition NMF convolutive [Lar+17], permettant d'apprendre des spectres élémentaires beaucoup plus riches. Les résultats obtenus permettent d'améliorer nettement les approches de l'état de l'art sur les morceaux contenant effectivement de la batterie.

La thèse de Clément Laroche, soutenue en mai 2017, a été financée par le projet digiteo SESAM. Clément faisait partie du projet PICS ImPulse, grâce auquel il a pu effectuer un séjour à Columbia au LabRosa.

Publications associées : [Lar+17; Lar+16; Lar+15a]

Représentation temps-fréquence : une approche variationnelle. Lors de mon séjour à Vienne au Erwin Schrödinger Institute en 2012, j'ai débuté une collaboration avec Hau-Tieng Wu sur les représentations temps-fréquence dites idéales, qui permettent de suivre aux mieux les fréquences instantanées d'un signal. Nous avons développé pour cela une approche variationnelle qui s'affranchit de la transformée de Fourier. En effet, nous avons défini la représentation temps-fréquence d'un signal comme le minimum de :

$$\begin{aligned} \mathcal{H}(F, G) := & \int \left| \Re \int F(t, \omega) d\omega - f(t) \right|^2 dt \\ & + \mu \iint_I |\partial_t F(t, \omega) - i2\pi\omega F(t, \omega) + G(t, \omega)\partial_\omega F(t, \omega)|^2 dt d\omega \\ & + \lambda \|F\|_{L^1(I)} + \gamma \|G\|_{L^2(I)}. \end{aligned}$$

où f est le signal à décomposer, F sa représentation fréquentielle et G un terme permettant de capturer les variations temps-fréquence rapides.

Ce travail a pu aboutir lors du stage d'Adrien Meynard en M1, qui poursuit aujourd'hui sur une thèse que je co-encadre avec Bruno Torrèsani.

Publications associées : [KMW16]

2 ■ L-One : a sparse war story

Ce chapitre permet de contextualiser l'ensemble de mes travaux de recherche en présentant une histoire de la parcimonie en traitement du signal et son utilisation pour les problèmes inverses.

Contributions principales :

1. Un historique des méthodes parcimonieuses.
2. Une revue des approches existantes en *synthèse* [Bal+13].

Collaborations et projets associés : L'article de revue générale publié dans IEEE Signal Processing Magazine [Bal+13] a été construit principalement à l'Erwin Schrodinger Institute à Vienne lors du workshop *Modern Methods of Time-Frequency Analysis II*.

Références personnelles citées dans ce chapitre :

- [KT09a; Kow09], rédigés pendant ma thèse, qui étudient les normes mixtes sur deux niveaux comme critère de parcimonie.
- [GKH12], qui introduit une norme mixte sur trois niveaux pour un problème inverse M/EEG.
- L'article [Bal+13] dresse un état de l'art sur les méthodes temps-fréquence d'un point de vue synthèse.

2.1. Un résumé historique de la parcimonie

2.1.1. Du seuillage au ℓ_1

Avant la généralisation de l'utilisation des dictionnaires sur-complets, les travaux se sont focalisés sur l'estimation d'un signal bruité dans une base orthonormée. Soit l'observation bruitée

$$\mathbf{y} = \mathbf{x} + \mathbf{e}, \tag{2.1}$$

avec \mathbf{e} un bruit que l'on supposera blanc et gaussien.

2. L-One : a sparse war story

L'utilisation de bases orthogonales d'ondelettes ou MDCT (Modified Discrete Cosine Transform) ont permis l'analyse d'opérateurs de type *seuillage doux* ou *seuillage dur* et d'obtenir des résultats précis en terme de performance de débruitage, selon la classe de signal considérée. On pourra se reporter aux travaux fondateurs de l'analyse par seuillage dans des bases d'ondelettes de DONOHO [Don95], DONOHO et JOHNSTONE [D+94] et DONOHO et JOHNSTONE [DJ95].

Cette estimation par seuillage pose les fondations des méthodes parcimonieuses : le seuillage force la mise à zéro de certains coefficients considérés comme peu informatifs. Cependant, l'utilisation de bases orthonormées de type ondelettes limite le choix des atomes. L'algorithme de Best Basis, apparu avec les paquets d'ondelettes (cf. [Mal08, Chapitre 8]), permet de construire de façon adaptative au signal la meilleure base orthonormée au sens de la minimisation d'un critère d'entropie [CW92]. Cependant, in fine, seules des composantes orthogonales sont estimées ce qui peut être trop restrictif dans le choix des atomes.

La notion d'estimation parcimonieuse s'est précisée avec la question de la représentation, puis de l'estimation, d'un signal dans un dictionnaire temps-fréquence de type Gabor ou STFT (Short Time Fourier Transform). Le principe d'incertitude d'Heisenberg implique des interférences temps-fréquence qui peuvent nuire à l'interprétation d'un signal lors de son analyse temps-fréquence. Si des méthodes comme la réallocation temps-fréquence ont été développées, avec un certain succès pour l'analyse, l'algorithme de Matching Pursuit [MZ93] a introduit la notion de *l'analyse par la synthèse*. La question posée n'est plus "*comment analyser au mieux un signal ?*" mais "*comment le synthétiser au mieux ?*". Cet algorithme sélectionne de manière gloutonne les atomes les plus corrélés (et donc les plus informatifs) avec le signal. Son principal inconvénient est de procéder à une analyse locale et gloutonne du signal, plutôt qu'une analyse globale. C'est dans ce contexte qu'a été introduit le Basis Pursuit [Che95 ; CDS98 ; CDS01]

$$\min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad \mathbf{x} = \Phi\alpha. \quad (2.2)$$

Le critère ℓ_1 a été introduit comme un critère d'entropie (et donc de parcimonie), convexe, facilement optimisable par programmation linéaire. Une propriété du Basis Pursuit est de sélectionner la meilleure base *non orthogonale*, au sens du critère ℓ_1 , pour synthétiser le signal. Ces travaux insistent sur le caractère de l'estimation à la *synthèse* et non à l'*analyse* du signal. Le Basis Pursuit a été étendu à l'estimation dans du bruit avec le Basis Pursuit Denoising (BPDN)

$$\min_{\alpha} \frac{1}{2} \|\mathbf{y} - \Phi\alpha\|_2^2 + \lambda \|\alpha\|_1. \quad (2.3)$$

Ce critère a été introduit dans le même temps en apprentissage statistique avec le Lasso [Tib96].

2.1.2. Résolution du Basis Pursuit Denoising par seuillage itératif

Dans l'article original [CDS98] du Basis Pursuit, les auteurs montrent l'équivalence du problème à un programme linéaire et proposent de le résoudre à l'aide du simplexe ou de la méthode des points intérieurs. Cependant, ces algorithmes passent difficilement à

2. L-One : a sparse war story

l'échelle. La difficulté du problème (2.3) vient de la non-différentiabilité de la fonctionnelle. Pour contourner cette difficulté, des auteurs ont proposé de remplacer la pénalité ℓ_1 par un critère différentiable approchant, comme la fonction de Hubert. Il est montré dans [Nes05] que cette approche n'est pas optimale.

Entre 2003 et 2005, trois approches différentes [FN03 ; DDD04 ; CW05] ont été proposées et aboutissent au même algorithme de seuillage itératif présenté plus en détail dans le chapitre 3. Dans [FN03], le problème (2.3) est reformulé à l'aide d'une variable (bien) cachée, permettant d'appliquer un algorithme d'Espérance-Minimisation (EM, pour Expectation-Minimization). Une approche de type Majoration-Minimisation, dont la convergence est montrée par l'application du théorème de point fixe d'Opial, est proposée dans [DDD04]. Enfin, dans [CW05], les approches d'analyse convexe sont utilisées avec la notion d'opérateur de proximité (ou opérateur proximal). Notons que [CW05] s'intéresse à une classe de problème bien plus générale que le problème du BPDN (2.3). On reviendra dans le chapitre 3 sur une version plus générale de l'algorithme de seuillage itératif.

2.1.3. Du ℓ_1 au ℓ_0

Le problème du Basis Pursuit s'est rapidement généralisé à la question "*quelle est la synthèse parcimonieuse optimale d'un signal ?*". Cette question a été formulée comme la minimisation de la pseudo-norme ℓ_0 [Don01] :

$$\min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad \mathbf{x} = \Phi\alpha, \quad (2.4)$$

ainsi que dans sa version *débruitage*

$$\min_{\alpha} \frac{1}{2} \|\mathbf{y} - \Phi\alpha\|_2^2 + \lambda \|\alpha\|_0. \quad (2.5)$$

Ce problème purement combinatoire, et donc très difficilement solvable, a le mérite de poser clairement la définition de l'estimation parcimonieuse idéale, à savoir celle qui fait intervenir le moins de coefficients possible. Beaucoup de travaux de recherche ont alors cherché *quand* cette solution pouvait être atteinte et sous quelles conditions. Plusieurs critères ont alors été proposés, sous lesquels il est démontré que le Matching Pursuit et le Basis Pursuit obtiennent effectivement la solution idéale au sens ℓ_0 – on se reportera par exemple à [Tro04 ; Tro06]. Enfin, la notion de parcimonie a donné naissance au concept d'*échantillonnage compressé*¹, et donc de l'identifiabilité d'un signal à partir d'un nombre restreint de mesures [Don06 ; CW08].

2.2. Structures et problèmes inverses

2.2.1. Parcimonie pour les problèmes inverses

Le BPDN a pour but l'estimation des coefficients de synthèse dans un dictionnaire. Un dictionnaire pouvant se confondre avec ses opérateurs linéaires de synthèse ou d'analyse,

1. Différentes appellations existent en anglais : *compressive sensing*, *compressed sensing*, *compressed sampling*, etc.

2. L-One : a sparse war story

la généralisation du problème du BPDN (2.3) aux problèmes inverses linéaires (1.2) est immédiate. Supposons que l'on dispose d'une observation sous la forme du problème direct (1.2). Si l'on suppose que le signal à estimer \mathbf{x} admet une décomposition parcimonieuse dans un dictionnaire Φ , on a vu au chapitre 1 que le problème direct s'écrit

$$\mathbf{y} = \mathbf{Ax} = \mathbf{A}\Phi\boldsymbol{\alpha} . \quad (2.6)$$

Une estimation des coefficients de synthèse $\boldsymbol{\alpha}$ de \mathbf{x} peut être alors obtenue par

$$\min_{\boldsymbol{\alpha}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\Phi\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 . \quad (2.7)$$

L'algorithme de seuillage itératif (cf. chapitre 3) peut ensuite être utilisé pour estimer les coefficients de synthèse $\boldsymbol{\alpha}$.

Si cette approche permet d'obtenir souvent de bons résultats en problème inverse (cf. les résultats obtenus en *compressive sensing* [CW08] par exemple), la question de prendre en compte une certaine structure sur les coefficients de synthèse se pose rapidement. En effet, une interprétation statistique prudente de la norme ℓ_1 suppose que tous les coefficients de synthèse sont indépendants. Or, si l'on regarde, par exemple, le spectrogramme d'un signal de musique, une telle hypothèse semble très restrictive : l'envie de regrouper les coefficients se fait rapidement sentir.

2.2.2. Pénalités structurantes

Les pénalités ℓ_2 et ℓ_1 donnent deux types de sélection des coefficients de synthèse

- La norme ℓ_2 répartit l'énergie sur l'ensemble des coefficients.
- la norme ℓ_1 sélectionne quelques coefficients significatifs, tels que les atomes ainsi sélectionnés soient peu corrélés (sur le principe de la meilleure base non orthogonale).

Coefficients corrélés. Il peut être souhaitable d'avoir un comportement intermédiaire, c'est-à-dire sélectionner quelques atomes fortement corrélés tout en conservant une certaine parcimonie. C'est dans ce but qu'a été introduit l'*Elastic-Net* [EN, ZH05], qui construit la pénalité Ω comme une combinaison des pénalités ℓ_2 et ℓ_1 :

$$\Omega_{EN}(\boldsymbol{\alpha}) = \lambda \|\boldsymbol{\alpha}\|_2^2 + \mu \|\boldsymbol{\alpha}\|_1 .$$

Cette pénalité a ensuite donné naissance au k -support norm [AFS12], qui est défini comme la meilleure relaxation convexe permettant de contrôler la taille du support et la norme ℓ_2 des coefficients. D'autres pénalités favorisant la sélection d'atomes fortement corrélés ont été étudiées, telles que la pénalité OSCAR [BR08], plus tard généralisée par la pénalité Ordered Weighted ℓ_1 [FN16].

Coefficients explicitement structurés / structure a priori. Une autre approche importante est la possibilité de regrouper des coefficients. La notion de parcimonie apparaît alors entre groupes de coefficients plutôt que sur les coefficients eux-mêmes. Ce type de

2. L-One : a sparse war story

structure suppose que les coefficients ont un double indice $k = (i, j)$, ce qui apparaît naturellement avec des dictionnaires temps-fréquence. La pénalité la plus connue est certainement le *Group-Lasso* [GL, YL06] :

$$\Omega_{GL}(\boldsymbol{\alpha}) = \lambda \|\boldsymbol{\alpha}\|_{2,1} = \lambda \sum_i \sqrt{\sum_j |\alpha_{i,j}|^2}.$$

Cependant, le même type de pénalité a été étudiée simultanément dans [Cot+05], où la norme ℓ_1 externe est remplacée par une norme ℓ_q , pour $q \leq 1$. Ce type de pénalité $\ell_{2,q}$ tombe dans une classe plus large étudiée pendant ma thèse sous le nom de norme mixte $\ell_{p,q}$ [KT09a; Kow09]. Le choix particulier $p = 1$ et $q = 2$ amène une sélection des coefficients par groupe, sous le nom d'Elitist-Lasso, renommé Exclusive-Lasso par d'autres auteurs dans un contexte d'apprentissage [ZJH10]. Cette pénalité par groupe a essentiellement été étudiée et / ou utilisée dans le cadre de groupes sans intersection. Le cas de groupes recouvrants est plus complexe, au moins d'un point de vue optimisation, et a été étudié plus spécifiquement par Jacob *et al.* [JOV09a]. Dans le contexte des problèmes inverses Magneto-Electro-Encéphalographie (M/EEG), avec A. Gramfort, nous avons utilisé une extension des normes mixtes sur trois niveaux [GKH12]. Un autre cas intéressant est la possibilité d'obtenir de la parcimonie à l'intérieur des groupes tout en conservant une parcimonie sur les groupes eux-mêmes. Ce type de pénalités hiérarchiques proposées dans [JAB11] a été étudiée plus spécifiquement sous le nom de *Hi-Lasso* [Hi-L, Spr+11] lorsquel seulement deux niveaux de hiérarchies sont considérés :

$$\Omega_{Hi-L}(\boldsymbol{\alpha}) = \lambda \|\boldsymbol{\alpha}\|_{2,1} + \mu \|\boldsymbol{\alpha}\|_1.$$

À cette liste non exhaustive de pénalités, on peut ajouter l'étude des fonctions sous-modulaires pour la prise en compte de structures complexes [Bac10]. Enfin, on pourra se reporter aux articles de review [Kyr+15] et [Bac+12, Section 1.3] pour une présentation détaillée des principales pénalités existantes.

2.2.3. Modèles en couches

Un autre type de structure est donné par les modèles additifs en couches :

$$\begin{aligned} \mathbf{x} &= \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{e} \\ &= \boldsymbol{\Phi}_1 \boldsymbol{\alpha}_1 + \boldsymbol{\Phi}_2 \boldsymbol{\alpha}_2 + \mathbf{e}, \end{aligned}$$

où les deux signaux \mathbf{x}_1 et \mathbf{x}_2 sont supposés admettre une décomposition parcimonieuse $\boldsymbol{\alpha}_1$ et $\boldsymbol{\alpha}_2$ dans des dictionnaires $\boldsymbol{\Phi}_1$ et $\boldsymbol{\Phi}_2$ différents. Typiquement, cela permet l'usage de deux dictionnaires temps-fréquence avec des fenêtres différentes, permettant ainsi de prendre en compte les composantes bien localisées temporellement (avec une fenêtre courte) et les composantes bien localisées en fréquence (avec une fenêtre longue). Ce type de modèle a été introduit pour les signaux audio sous le nom de *modèle hybride* par Daudet et Torrésani [DT02]. C'est la même idée qu'on retrouve en image sous le nom de *Morphological Component Analysis* [MCA, Ela+05].

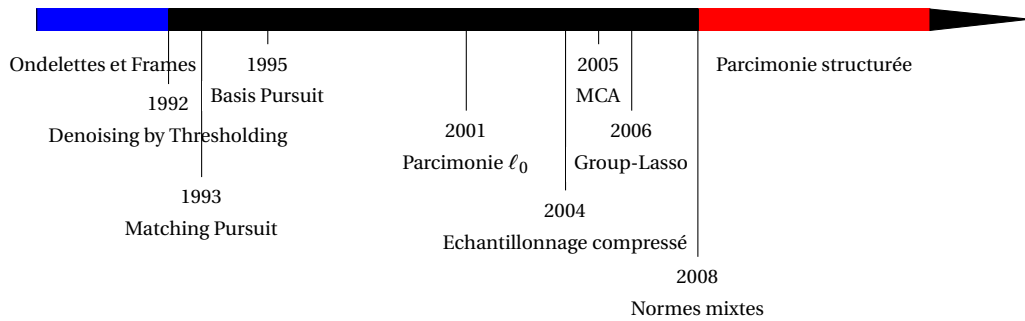
2. L-One : a sparse war story

Ainsi, on peut imaginer des pénalités différentes pour chacune des couches, afin de faire ressortir des structures différentes. Par exemple, on pourra choisir de regrouper tous les coefficients en fréquence pour un temps donné afin de trouver des composantes *transitoire* bien localisée en temps, et se contenter de prendre en compte la parcimonie pour les composantes *tonales* bien localisés en fréquence. On aboutit alors au problème général

$$\alpha_1, \alpha_2 = \operatorname{argmin}_{\alpha_1, \alpha_2} \frac{1}{2} \|\mathbf{y} - \mathbf{A}(\Phi_1 \alpha_1 + \Phi_2 \alpha_2)\|^2 + \lambda_1 \Omega_1(\alpha_1) + \lambda_2 \Omega_2(\alpha_2). \quad (2.8)$$

2.3. Frise chronologique

Les dates choisies sont celles, en général, des articles ou thèse de référence, elles sont donc indicatives. Cette frise se concentre sur la parcimonie, en laissant volontairement de côté le développement des dictionnaires tels que les curvelets, ridgelets et autres \star -lets qui nécessiteraient une autre frise.



3 ■ Revenge of the IST

L'algorithme de seuillage itératif, IST pour Itérative Shrinkage/Thresholding, est un algorithme simple et efficace (dans sa version accélérée) pour résoudre les problèmes de type (1.4), lorsque l'opérateur proximal associé à la pénalité est calculable efficacement.

Dans la mesure où cet algorithme tient une place centrale dans mes travaux, ce chapitre commence par en faire une description synthétique dans le cadre des fonctions convexes en dimension finie. On pourra se reporter à [Bac+12] pour une présentation générale des algorithmes d'optimisation utilisée avec les pénalités favorisant la parcimonie. Ce chapitre est aussi l'occasion de faire le lien entre les algorithmes classiques utilisés pour les décompositions parcimonieuses tels que le Matching Pursuit et les variantes de IST.

D'autre part, j'ai étudié dans [Kow16] des conditions essentiellement théoriques pour conjuguer les descentes proximales à la manière de l'algorithme de gradient conjugué employé pour les fonctions non quadratiques, que je ne présenterai pas ici.

Contributions principales :

1. Une présentation des approches de type "Matching Pursuit" comme des variantes par blocs de l'algorithme IST.
2. Un algorithme de seuillage itératif par contraintes actives (AS, pour Active-Set [Kow+11]).

Collaborations et projets associés :

Le projet MAORI a été le cadre principal pour les discussions et mes travaux autour de l'optimisation. C'est dans ce cadre que j'ai collaboré avec Sandrine Anthoine, Alexandre Gramfort et Pierre Weiss sur l'algorithme AS-ISTA [Kow+11].

Références personnelles citées dans ce chapitre :

- [GKH12], qui présente l'opérateur de proximité et la norme duale d'une norme mixte sur trois niveaux.
- [Kow+11], qui présente l'algorithme AS-ISTA.

3.1. Descente proximale

On présente d'algorithmme sous sa forme générale, avant de montrer que les opérateurs de proximité correspondent effectivement à des fonctions de seuillage.

3.1.1. Algorithme

L'algorithme de seuillage itératif a d'abord été développé spécifiquement pour le problème de type BPDN (2.3). Avec une approche de type Maximisation-Minimisation, Daubechies *et al.* montrent que le problème BPDN peut être résolu avec un algorithme de type "itérations de Landeweber" avec seuillage doux [DDD04]. Il est montré dans [CW05] que ce type d'algorithmme se généralise à une classe de fonctions F convexes beaucoup plus large, faisant intervenir un terme convexe et Lipschitz différentiable f et une pénalité convexe Ω :

$$F(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha}) + \lambda\Omega(\boldsymbol{\alpha}).$$

Plus particulièrement, l'opérateur de seuillage doux est l'opérateur de proximité (ou opérateur proximal) de la pénalité ℓ_1 . On rappelle ci-après la définition de l'opérateur proximal.

Définition 3.1

(Opérateur de proximité)

Soit $\Omega : \mathbb{R}^K \rightarrow \mathbb{R}$ une fonction convexe. On appelle opérateur de proximité (ou opérateur proximal) de $\lambda\Omega$, noté $\text{prox}_{\lambda\Omega}$, la fonction

$$\begin{aligned} \text{prox}_{\lambda\Omega} : \mathbb{R}^K &\rightarrow \mathbb{R}^K \\ \mathbf{y} &\mapsto \text{prox}_{\lambda\Omega}(\mathbf{y}) = \underset{\mathbf{z}}{\text{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{z}\|^2 + \lambda\Omega(\mathbf{z}). \end{aligned}$$

Cette définition peut se généraliser directement aux pénalités non convexe. Ainsi, l'opérateur de proximité peut s'interpréter comme la résolution du problème inverse (2.6) avec les choix particuliers

- $\mathbf{A} = \mathbf{I}$;
- Φ est une base orthonormée .

L'algorithme de descente proximale s'écrit alors simplement :

$$\boldsymbol{\alpha}^{(t+1)} = \text{prox}_{\frac{\lambda}{L}\Omega} \left(\boldsymbol{\alpha}^{(t)} + \frac{1}{L} \nabla f(\boldsymbol{\alpha}^{(t)}) \right). \quad (3.1)$$

avec $L = \|\nabla f\|$.

L'article de Tseng [Tse10] donne une preuve très simple de convergence de l'algorithme permettant d'obtenir aussi sa vitesse de convergence en $\mathcal{O}(1/t)$. Cette analyse permet d'étendre la convergence de l'algorithme de descente proximale (3.1) lorsque la fonction f est L -Lipschitz différentiable mais pas convexe. Une autre propriété remarquable, pouvant être obtenue directement avec cette analyse, est la convergence de l'algorithme relaxé :

$$\begin{aligned} \boldsymbol{\gamma}^{(t)} &= \boldsymbol{\alpha}^{(t)} + \mu^{(t)} (\boldsymbol{\alpha}^{(t)} - \boldsymbol{\alpha}^{(t-1)}) \\ \boldsymbol{\alpha}^{(t+1)} &= \text{prox}_{\frac{\lambda}{L}\Omega} (\boldsymbol{\gamma}^{(t)} + \nabla f(\boldsymbol{\gamma}^{(t)})/L) \end{aligned} \quad (3.2)$$

3. Revenge of the IST

avec la suite des paramètres de relaxation $\mu^{(t)} < 1$. L'emploi d'un tel paramètre de relaxation (par exemple $\mu^{(t)} = 0.9$, quelque soit t) permet d'accélérer significativement la convergence en pratique. Une étude plus précise de l'algorithme relaxé permet d'aboutir à une vitesse de convergence en $\mathcal{O}(1/t^2)$ avec l'algorithme FISTA (Fast Iterative Shrinkage/Thresholding Algorithm) [BT09]. Dans une analyse plus récente [CD15], les auteurs proposent un choix pour la suite plus simple $\mu^{(t)} = \frac{t}{t+4}$, permettant d'obtenir la convergence des itérés (et non simplement de l'image des itérés).

3.1.2. Opérateurs proximaux et fonctions de seuillage

On montre ici que les schémas itératifs donnés par les équations (3.1) et (3.2) sont effectivement des algorithmes de seuillage itératifs. Pour montrer cela, on définit d'abord les fonctions de seuillage, au sens d'Antoniadis [Ant+07].

Définition 3.2 (Opérateur de seuillage)

$\mathbb{S}(\cdot; \lambda)$ est un opérateur de seuillage si et seulement si :

1. $\mathbb{S}(\cdot; \lambda)$ est une fonction impaire. On note $\mathbb{S}_+(\cdot; \lambda)$ sa restriction à \mathbb{R}_+ .
2. $0 \leq \mathbb{S}_+(x; \lambda) \leq x, \forall x \in \mathbb{R}_+$.
3. \mathbb{S}_+ est croissante sur \mathbb{R}_+ , et $\lim_{x \rightarrow +\infty} \mathbb{S}_+(x; \lambda) = +\infty$.

Il est montré dans [Ant+07] qu'on peut associer à n'importe quel opérateur de seuillage \mathbb{S} une pénalité $\Omega_{\mathbb{S}}$ (non nécessairement convexe), telle que \mathbb{S} soit l'opérateur proximal de $\Omega_{\mathbb{S}}$. Les propriétés de non-expansivité de l'opérateur proximal et la convexité de Ω nous permettent d'énoncer directement la proposition suivante, sorte de réciproque à la remarque précédente.

Proposition 3.1 (Équivalence entre opérateur proximal et fonction de seuillage)

Soit $\Omega : \mathbb{R}^N \rightarrow \mathbb{R}$ une fonction convexe paire. Alors $\text{prox}_{\lambda\Omega}$ est un opérateur de seuillage au sens d'Antoniadis.

On donne alors l'algorithme de seuillage itératif relaxé (algorithme 1) pour résoudre les problèmes de type 1.3 avec une attache aux données ℓ_2 :

$$\min_{\alpha} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\Phi\alpha\|_2^2 + \lambda\Omega(\alpha).$$

On désigne par

$$\begin{aligned} \mathbb{S} : \mathbb{R}^K &\rightarrow \mathbb{R}^K \\ \alpha &\mapsto \mathbb{S}(\alpha, \lambda) = \text{prox}_{\lambda\Omega}(\alpha) \end{aligned}$$

l'opérateur de seuillage associé à la pénalité Ω (i.e. son opérateur proximal),

Un des avantages des fonctions convexes est de fournir un moyen simple pour vérifier si le minimum est atteint en calculant le *saut de dualité* qui repose sur le théorème de dualité de Fenchel-Rockafellar [Roc70]. Dans [GKH12], nous avons calculé le conjugué de Fenchel pour n'importe quelle norme mixte, afin d'avoir un critère d'optimalité pour les algorithmes utilisés en M/EEG.

3. Revenge of the IST

Algorithm 1: Algorithme de seuillage itératif (IST)

```

1 Set  $L = \|\mathbf{A}\|_2^2 \|\Phi\|_2^2$  (or a larger value)
2 Set  $\boldsymbol{\gamma}^{(0)} \in \mathbb{R}^N$  and  $t = 0$ 
3 repeat
4    $\boldsymbol{\alpha}^{(t+1)} = \mathbb{S}(\boldsymbol{\gamma}^{(t)} + \frac{1}{L} \Phi^* \mathbf{A}^* (\mathbf{y} - \mathbf{A} \Phi \boldsymbol{\gamma}^{(t)}), \lambda)$ 
5    $\boldsymbol{\gamma}^{(t+1)} = \boldsymbol{\alpha}^{(t+1)} + \mu^{(t)} (\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{\alpha}^{(t)})$ 
6    $t = t + 1$ 
7 until convergence;
```

3.2. Variantes par blocs

L'algorithme 1 procède à une descente de gradient puis un seuillage sur l'ensemble des coefficients de synthèse $\boldsymbol{\alpha}$. Dans le même temps que les algorithmes de seuillage itératifs tels que 1, des algorithmes de descente par coordonnées ou par blocs ont été proposés. On pourra se référer à [TY09] pour une revue générale des descentes par coordonnées, et en particulier à la discussion sur le choix du parcours des coordonnées dans l'algorithme concernant les règles de Gauss-Seidel (simple cycle sur les coordonnées) ou celles de Gauss-Southwell. On rappelle la règle de Gauss-Southwell-r.

Définition 3.3 (Règle de Gauss-Southwell-r)

Soit la descente restreinte aux coordonnées de l'ensemble \mathcal{J}

$$d(\boldsymbol{\alpha}, \mathcal{J}) = \operatorname{argmin}_d \left\{ f(\boldsymbol{\alpha}) + \frac{1}{2} \|d\|^2 + \Omega(\boldsymbol{\alpha}) \mid d_j = 0 \forall j \notin \mathcal{J} \right\}.$$

\mathcal{J} suit la règle de Gauss-Southwell-r s'il existe $0 < \nu \leq 1$ tel que

$$\|d(\boldsymbol{\alpha}, \mathcal{J})\|_\infty \geq \nu \|d(\boldsymbol{\alpha})\|_\infty,$$

où $d(\boldsymbol{\alpha})$ est la descente sur toutes les coordonnées.

En particulier, le Matching Pursuit peut se voir comme une descente proximale par coordonnées avec un choix de type Gauss-Southwell. On rappelle dans l'algorithme 2 le principe du Matching Pursuit pour sélectionner les atomes $\boldsymbol{\varphi}$ d'un dictionnaire Φ et trouver une décomposition parcimonieuse de

$$\mathbf{y} = \Phi \boldsymbol{\alpha}.$$

On note $\boldsymbol{\delta}_k$ le vecteur de Dirac à la position k .

L'étape de mise à jour des coefficients $\boldsymbol{\alpha}$ peut se voir comme une descente de gradient sur la fonction $\boldsymbol{\alpha} \rightarrow \frac{1}{2} \|\mathbf{y} - \Phi \boldsymbol{\alpha}\|^2$, restreinte à la coordonnée k_t . Le choix de la coordonnée k_t , choisie comme celle maximisant la corrélation avec le dictionnaire, correspond exactement à la mise à jour de Gauss-Southwell-r avec $\nu = 1$ et $\#\mathcal{J} = 1$. Ainsi, le Matching Pursuit correspond exactement à la minimisation des moindres carrés, avec une descente par co-

3. Revenge of the IST

Algorithm 2: Matching Pursuit

```

1 Initialisation :  $\mathbf{r}^{(0)} = \mathbf{y}$ ,  $\boldsymbol{\alpha}^{(0)} = \mathbf{0}_N$ ,  $\mathbf{x}^{(0)} = \mathbf{0}_T$ 
2 repeat
3    $k_t = \operatorname{argmax}_\lambda |\langle \mathbf{r}^{(t)}, \boldsymbol{\varphi}_k \rangle|$ 
4    $\boldsymbol{\alpha}^{(t)} = \boldsymbol{\alpha}^{(t)} + \langle \mathbf{r}^{(t)}, \boldsymbol{\varphi}_{k_t} \rangle \boldsymbol{\delta}_{k_t}$ 
5    $\mathbf{x}^{(t)} = \mathbf{x}^{(t)} + \langle \mathbf{r}^{(t)}, \boldsymbol{\varphi}_{k_t} \rangle \boldsymbol{\varphi}_{k_t}$ 
6    $\mathbf{r}^{(t+1)} = \mathbf{y} - \mathbf{x}^{(t)} = \mathbf{r}^{(t)} - \langle \mathbf{r}^{(t)}, \boldsymbol{\varphi}_{k_t} \rangle \boldsymbol{\varphi}_{k_t}$ 
7    $t = t + 1$ 
8 until convergence;

```

ordonnée, suivie en général d'une règle *d'early stopping* [Pre98; Mat+17]. Le même raisonnement peut s'appliquer au Matching Pursuit Orthogonal [PRK93].

Dans [Kow+11], nous avons proposé une méthode de descente proximale par blocs avec une règle de Gauss-Southwell-r adaptée au Lasso et qui s'étend directement au Group-Lasso. Cette adaptation de l'algorithme IST, nommé AS-ISTA pour Active-Set Iterative Shrinkage/Thresholding Algorithm, a été inspirée par une étude d'un algorithme d'ensemble actif donné par Roth et Fisher [RF08]. Cet algorithme permet de travailler sur un nombre de coefficients réduit à chaque itération, ce qui améliore considérablement le temps de calcul sur les problèmes rencontrés en M/EEG. La sélection des coordonnées se fait en regardant quelles sont celles qui violent le plus les contraintes d'optimalité du problème. Un tel algorithme est adapté aux problèmes où la matrice \mathbf{A} est fortement sous-déterminé et lorsque la matrice dictionnaire Φ peut être stockée entièrement. Les algorithmes d'ensemble actif [Bac+11] permettent ainsi de travailler à chaque itération sur des problèmes de taille réduite. On donne les étapes de AS-ISTA dans l'algorithme 3, avec $\Phi = \mathbf{I}$. On note pour un ensemble d'indices \mathcal{A} , $\mathbf{A}_{\mathcal{A}}$ la sous matrice \mathbf{A} formée par les colonnes d'indices contenus dans \mathcal{A} , et $\boldsymbol{\alpha}_{\mathcal{A}}$ le sous vecteur de $\boldsymbol{\alpha}$ formé par les coordonnées d'indices de \mathcal{A} .

Un autre avantage des algorithmes de descente par blocs est la possibilité de calculer un pas de descente sur le sous-problème, généralement mieux conditionné. Le pas calculé est alors plus grand et permet d'accélérer l'algorithme en pratique. Le choix d'une descente par coordonnée (ou par bloc) dépend aussi énormément de l'environnement de programmation. Si avec MatLab une implémentation classique sera souvent plus efficace qu'une descente par coordonnées, les conclusions peuvent s'inverser avec une implémentation dans un autre langage.

Le travail effectué sur AS-ISTA a permis une amélioration pratique appréciable. Cependant, nous avons butté sur des résultats théoriques que nous souhaitons montrer, comme savoir si une coordonnée ajoutée à une itération puis retirée plusieurs itérations plus tard, peut-être ajoutée à nouveau ou bien considérée comme définitivement retirée. Les expériences suggèrent ce dernier comportement.

Enfin, les algorithmes de descente par coordonnées se prêtent très facilement à une adaptation stochastique, où le choix des coordonnées peut se faire aléatoirement [RVV14]. Malheureusement, ces algorithmes ne sont pas adaptés aux problèmes faisant intervenir

3. Revenge of the IST

Algorithm 3: AS-ISTA

1 Initialization : $\mathcal{A}^{(0)} \subset \{k \text{ s.t. } |\langle \boldsymbol{\varphi}_k, \mathbf{y} \rangle| > \lambda\}$

2 **repeat**

1. Let $\tilde{\boldsymbol{\alpha}}_{\mathcal{A}^{(t)}}^{(0)} = \boldsymbol{\alpha}_{\mathcal{A}^{(t)}}^{(t)}$, and $\mathbf{A}^{(t)} = \mathbf{A}_{\mathcal{A}^{(t)}}$.

2. Iterates $J^{(t)}$ times

$$\tilde{\boldsymbol{\alpha}}_{\mathcal{A}^{(t)}}^{(j+1)} = \mathbb{S} \left(\tilde{\boldsymbol{\alpha}}_{\mathcal{A}^{(t)}}^{(j)} + \mathbf{A}^{(t)*} \left(\mathbf{y} - \mathbf{A}^{(t)} \tilde{\boldsymbol{\alpha}}_{\mathcal{A}^{(t)}}^{(j)} \right), \lambda \right)$$

3. $\boldsymbol{\alpha}_{\mathcal{A}^{(t)}}^{(t+1)} = \tilde{\boldsymbol{\alpha}}_{\mathcal{A}^{(t)}}^{(J)}$

4. Compute the dual variable $\mathbf{z}^{(t+1)} = \mathbf{y} - \mathbf{A} \boldsymbol{\alpha}^{(t+1)}$

5. Let $\mathcal{a}^{(t+1)} \subset \{k \text{ such that } |\langle \boldsymbol{\varphi}_k, \mathbf{y} \rangle| > \lambda\}$

6. Update the active set $\mathcal{A}^{(t+1)} = \text{supp}(\boldsymbol{\alpha}^{(t+1)}) \cup \left\{ \underset{k}{\text{argmax}} |\langle \boldsymbol{\varphi}_k, \mathbf{y} \rangle| \right\} \cup \mathcal{a}^{(t+1)}$

3 **until** convergence;

un dictionnaire comme les dictionnaires temps-fréquence, où les calculs d'analyse et de synthèse se font à l'aide d'un algorithme rapide, sans avoir besoin de stocker la matrice du dictionnaire.

4 ■ Bayes strikes back

S'il est réducteur de circonscrire le cadre Bayésien à la définition de *priors*, le choix d'un modèle probabiliste fournit un cadre pratique pour générer des données selon un modèle précis mais aussi pour évaluer d'un point de vue statistique les performances d'un estimateur (non nécessairement Bayésien).

Dans l'introduction, nous avons vu qu'interpréter la norme ℓ_1 comme un prior Laplacien dans les modèles parcimonieux est généralement inadapté. Un modèle probabiliste simple permettant d'obtenir des données effectivement parcimonieuses est donné par le modèle Bernoulli-Gaussien [Sou+11; DHD12]. Un autre modèle proposé par le *Sparse Bayesian Learning* [SBL, WR04] suppose que les coefficients de synthèse sont générés de manière indépendante selon des lois normales dont la variance peut être différente pour chacun des coefficients. Ce modèle permet effectivement d'estimer des données parcimonieuses.

Dans la première section de ce chapitre, on analyse les performances du seuillage dur à la lumière du modèle Bernoulli-Gaussien. La seconde section construit sur le SBL et sur le modèle composite gaussien pour proposer le modèle Low Rank Time-Frequency Synthesis (LRTFS) permettant d'introduire une structure de rang faible sur les coefficients de synthèse dans un dictionnaire temps-fréquence.

Contributions principales :

1. Une étude du seuillage dur à l'aide du modèle Bernoulli-Gaussien.
2. Un algorithme de type IST pour le Bernoulli-Gaussien.
3. Une présentation du modèle Low-Rank Time Frequency Synthesis (LRTFS).

Collaboration : L'étude du modèle LRTFS a débuté lors d'une invitation par Cédric Févotte à l'université de Nice en décembre 2013. Les premières discussions ont eu lieu à l'Erwin Schrodinger Institute à Vienne lors du workshop *Modern Methods of Time-Frequency Analysis II*.

Références personnelles citées dans ce chapitre :

- [KR11a], qui présente un algorithme de type IST construit sur le modèle Bernoulli-Gaussien.

4. Bayes strikes back

- [FK14b; FK15b; KF17], qui présente le modèle LRTFS et ses variantes ainsi que leur utilisation pour les problèmes inverses.

4.1. Un modèle génératif parcimonieux : le Bernoulli-Gaussien

Après la description du modèle Bernoulli-Gaussien, on analyse le débruitage par seuillage dur dans ce cadre. Un algorithme de type IST avec détermination automatique des paramètres est présenté dans le cas d'une union de bases orthonormée. Enfin, le modèle Bernoulli-Gaussien est généralisé par les approches de type *Sparse Bayesian Learning*.

4.1.1. Modèle et débruitage

Soit une observation $\mathbf{y} \in \mathbb{R}^N$ d'un signal $\mathbf{x} \in \mathbb{R}^N$ dégradé par un bruit $\mathbf{e} \in \mathbb{R}^N$

$$\mathbf{y} = \mathbf{x} + \mathbf{e}. \quad (4.1)$$

On suppose que le signal \mathbf{x} et le bruit sont des réalisations qui suivent

- Un modèle Bernoulli-Gaussien pour \mathbf{x} : $\forall n, x_n \sim \mathcal{B}\mathcal{N}(p, 0, \sigma_x^2)$,
- Un bruit blanc gaussien pour \mathbf{e} : $\forall n, e_n \sim \mathcal{N}(0, \sigma_0^2)$.

Autrement dit, on a pour toutes les coordonnées $n \in \{1, \dots, N\}$

$$y_n = X_n x_n + e_n,$$

avec

- $x_n \sim (1 - X_n)\delta_0 + X_n \mathcal{N}(0, \sigma_x^2)$, $X_n \sim \mathcal{B}(p)$,
- $e_n \sim \mathcal{N}(0, \sigma_0^2)$.

On a alors comme rapport signal à bruit en entrée

$$\text{iSNR}(\mathbf{x}, \mathbf{y}) = 10 \log \left(\frac{\mathbb{E}[\|\mathbf{x}\|^2]}{\mathbb{E}[\|\mathbf{e}\|^2]} \right) = 10 \log \left(p \frac{\sigma_x^2}{\sigma_0^2} \right).$$

Supposons qu'on soit capable de retrouver parfaitement le support de \mathbf{x} en mettant simplement à zéro les y_n tels que $X_n = 0$. Cet estimation *oracle* a pour SNR :

$$\begin{aligned} \text{SNR}(\mathbf{x}, \hat{\mathbf{x}}) &= 10 \log \left(\frac{\mathbb{E}[\|\mathbf{x}\|^2]}{\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]} \right) \\ &= 10 \log \left(\frac{|S| \sigma_x^2}{|S| \sigma_0^2} \right) \\ &= \text{iSNR} - 10 \log(p). \end{aligned}$$

Ce simple calcul montre qu'un signal effectivement parcimonieux peut être débruité efficacement dès lors qu'on est capable de retrouver son support. Une estimation simple du support peut être obtenue par seuillage. On s'intéresse dans la section suivante à l'estimation du signal \mathbf{x} par seuillage dur sur le signal observé \mathbf{y} .

4. Bayes strikes back

4.1.2. Choix optimaux du seuil

L'estimateur du MAP pour la variable indicatrice X_n est donné par¹

$$\hat{X}_n = 1 \quad \text{si} \quad P(X_n = 1|y_n) > P(X_n = 0|y_n),$$

ou autrement dit, $\hat{X}_n = 1$ si

$$\frac{p}{\sqrt{2\pi(\sigma_0^2 + \sigma_x^2)}} e^{-\frac{|y_n|^2}{2(\sigma_0^2 + \sigma_x^2)}} > \frac{1-p}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{|y_n|^2}{2\sigma_0^2}}.$$

C'est-à-dire

$$|y_n|^2 > 2\sigma_0^2 \left(\frac{\sigma_0^2}{\sigma_x^2} + 1 \right) \log \left(\frac{1-p}{p} \sqrt{1 + \frac{\sigma_x^2}{\sigma_0^2}} \right). \quad (4.2)$$

Le théorème suivant permet de (re)montrer que cet estimateur permet effectivement de minimiser la somme des deux types d'erreurs possibles sur le support.

Théorème 4.1 (Optimalité 0-1 du MAP)

Soient $\mathbf{x} \sim \mathcal{B}\mathcal{N}(p, 0, \sigma_x^2)$ et $\mathbf{e} \sim \mathcal{N}(0, \sigma_0^2)$ avec $\mathbf{y} = \mathbf{x} + \mathbf{e}$. L'estimateur du MAP minimise $(|\Delta \cap \tilde{S}| + |\tilde{\Delta} \cap S|)$, c'est-à-dire $P(\hat{X} = 1, X = 0) + P(\hat{X} = 0, X = 1)$.

Démonstration. On a d'une part

$$\begin{aligned} P(\hat{X} = 1, X = 0) &= (1-p)P(\hat{X} = 1|X = 0) \\ &= 2(1-p) \int_{\lambda}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(y)^2}{2\sigma_0^2}} dy, \end{aligned}$$

et d'autre part

$$\begin{aligned} P(\hat{X} = 0, X = 1) &= pP(\hat{X} = 0|X = 1) \\ &= 2p \int_0^{\lambda} \frac{1}{\sqrt{2\pi(\sigma_0^2 + \sigma_x^2)}} e^{-\frac{(y)^2}{2(\sigma_0^2 + \sigma_x^2)}} dy. \end{aligned}$$

On dérive par rapport à λ

$$\frac{dP(\hat{X} = 1, X = 0) + P(\hat{X} = 0, X = 1)}{d\lambda} = -(1-p)2 \frac{1}{\sqrt{2\pi(\sigma_0^2)}} e^{-\frac{\lambda^2}{2\sigma_0^2}} + 2p \frac{1}{\sqrt{2\pi(\sigma_0^2 + \sigma_x^2)}} e^{-\frac{\lambda^2}{2(\sigma_0^2 + \sigma_x^2)}}$$

et l'annulation de la dérivée donne

$$\lambda = 2\sigma_0^2 \left(\frac{\sigma_0^2}{\sigma_x^2} + 1 \right) \log \left(\frac{1-p}{p} \sqrt{1 + \frac{\sigma_x^2}{\sigma_0^2}} \right).$$

□

1. Pour les détails, on pourra se référer par exemple à [KR11a].

4. Bayes strikes back

Ce résultat est une démonstration ad-hoc pour le modèle Bernoulli-Gaussien dans la mesure où l'estimateur MAP est l'*estimateur de Bayes* associé à la fonction de coût 0-1 [Rob06]. On montre que l'emploi du seuillage par MAP n'est pas le choix du seuil optimal au sens du MMSE pour le seuillage dur, qu'on note

$$\mathbb{S}_{\mathcal{H}}(\mathbf{y}, \lambda) = \mathbf{y} \cdot \mathbb{1}_{|y| > \lambda}.$$

Ce résultat de non-optimalité n'a rien de surprenant si l'on considère l'estimateur de Bayes pour le coût ℓ_2 qui est donné par la moyenne *a posteriori* et par la marginalisation du support [TYE11]. Or cet estimateur de Bayes, en raison de la marginalisation sur le support, ne donne pas une estimation strictement parcimonieuse de \mathbf{x} . Le théorème suivant donne la valeur théorique du seuil optimal (au sens MMSE) pour l'opérateur de seuillage dur.

Théorème 4.2 (Seuil optimal au sens MMSE pour le seuillage dur)

Soit

$$\lambda_{Risk} = \operatorname{argmin}_{\lambda} \mathbb{E} [\|\mathbf{x} - \mathbb{S}_{\mathcal{H}}(\mathbf{y}, \lambda)\|^2],$$

alors

$$\lambda_{Risk}^2 = \lambda_{MAP}^2 + \frac{2\sigma_0^2(\sigma_x^2 + \sigma_0^2)}{\sigma_x^2} \log \left(\frac{(\sigma_0^2 + \sigma_x^2)^2}{\sigma_x^4 - \sigma_0^4} \right).$$

Démonstration. On a

$$\|\mathbf{x} - \hat{\mathbf{x}}\|^2 = \sum_{n=1}^N |X_n x_n - \hat{X}_n \hat{x}_n|^2 = \sum_{n: X_n=1, \hat{X}_n=1} |e_n|^2 + \sum_{n: X_n=0, \hat{X}_n=1} |e_n|^2 + \sum_{n: X_n=1, \hat{X}_n=0} |x_n|^2,$$

c'est-à-dire

$$\begin{aligned} \frac{1}{N} \mathbb{E} [\|\mathbf{x} - \hat{\mathbf{x}}\|^2] &= \frac{1}{N} \sum_{n=1}^N \mathbb{E} [|x_n - \hat{x}_n|^2] \\ &= \mathbb{E} [e^2, X=1, \hat{X}=1] + \mathbb{E} [e^2, X=0, \hat{X}=1] + \mathbb{E} [x^2, X=1, \hat{X}=0]. \end{aligned}$$

Calcul de $\mathbb{E} [e^2, X=1, \hat{X}=1]$. On a

$$\begin{aligned} p_{e, X=1, \hat{X}=1}(u) &= \int_{\mathbb{R}} p_{e, X=1, \hat{X}=1, y}(u) dy \\ &= \int_{\mathbb{R}} P(\hat{X}=1|y) p_{y|e, X=1}(y) p_e(u) P(X=1) dy \\ &= p \int_{\mathbb{R}} \mathbb{1}_{|y| > \lambda} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(y-u)^2}{2\sigma_x^2}} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{u^2}{2\sigma_0^2}} dy \\ &= p \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{u^2}{2\sigma_0^2}} \int_{|y| > \lambda} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(y-u)^2}{2\sigma_x^2}} dy, \end{aligned}$$

4. Bayes strikes back

ce qui donne

$$\begin{aligned}
 E[e^2, X=1, \hat{X}=1] &= \int_{\mathbb{R}} u^2 p \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{u^2}{2\sigma_0^2}} \int_{|y|>\lambda} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(y-u)^2}{2\sigma_x^2}} dy du \\
 &= p \int_{|y|>\lambda} \int_{\mathbb{R}} u^2 \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{u^2}{2\sigma_0^2}} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(y-u)^2}{2\sigma_x^2}} du dy \\
 &= p \int_{|y|>\lambda} \int_{\mathbb{R}} u^2 \frac{A(y)}{\sqrt{2\pi\Sigma^2}} e^{-\frac{(u-\mu(y))^2}{2\Sigma^2}} du dy,
 \end{aligned}$$

avec

$$\mu(y) = \frac{y\sigma_0^2}{\sigma_0^2 + \sigma_x^2}, \quad \Sigma^2 = \frac{\sigma_0^2\sigma_x^2}{\sigma_x^2 + \sigma_0^2}, \quad A(y) = \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_0^2)}} e^{-\frac{y^2}{\sigma_x^2 + \sigma_0^2}}.$$

On obtient au final

$$\begin{aligned}
 E[e^2, X=1, \hat{X}=1] &= p \int_{|y|>\lambda} \int_{\mathbb{R}} u^2 \frac{A}{\sqrt{2\pi\Sigma^2}} e^{-\frac{(u-\mu(y))^2}{2\Sigma^2}} du dy \\
 &= p \int_{|y|>\lambda} A(y) (\Sigma^2 + \mu(y)^2) dy \\
 &= p \int_{|y|>\lambda} A(y) \left(\Sigma^2 + \left(\frac{y\sigma_0^2}{\sigma_0^2 + \sigma_x^2} \right)^2 \right) dy.
 \end{aligned}$$

Calcul de $E[e^2, X=0, \hat{X}=1]$. On a

$$\begin{aligned}
 p_{e, X=0, \hat{X}=1}(u) &= \int_{\mathbb{R}} p_{e, X=0, \hat{X}=1, y}(u) dy \\
 &= \int_{\mathbb{R}} p_{e|X=0, \hat{X}=1, y}(u) P(\hat{X}=1|X=0, y) P(X=0|y) p_y(y) dy \\
 &= \int_{\mathbb{R}} \delta_y(u) \mathbb{1}_{|y|>\lambda} \frac{1-p}{p_y(y)} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{y^2}{2\sigma_0^2}} p_y(y) dy \\
 &= (1-p) \int_{|y|>\lambda} \delta_y(u) \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{y^2}{2\sigma_0^2}} dy,
 \end{aligned}$$

4. Bayes strikes back

ce qui donne

$$\begin{aligned}
 E[e^2, X=0, \hat{X}=1] &= \int_{\mathbb{R}} u^2 (1-p) \int_{|y|>\lambda} \delta_y(u) \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{y^2}{2(\sigma_0^2)}} dy du \\
 &= (1-p) \int_{|y|>\lambda} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{y^2}{2\sigma_0^2}} \int_{\mathbb{R}} u^2 \delta_y(u) du dy \\
 &= (1-p) \int_{|y|>\lambda} y^2 \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{y^2}{2(\sigma_0^2)}} dy.
 \end{aligned}$$

Calcul de $E[x^2, X=1, \hat{X}=0]$. On a

$$\begin{aligned}
 p_{x, \hat{X}=0, X=1}(x) &= \int_{\mathbb{R}} p_{x, \hat{X}=0, X=1, y}(x) dy \\
 &= \int_{\mathbb{R}} P(\hat{X}=0|y) p_{y|x, X=1}(y) p(X=1) p_{x|X=1}(x) P(X=1) dy \\
 &= p \int_{\mathbb{R}} \mathbb{1}_{|y|<\lambda} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(y-x)^2}{2\sigma_0^2}} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{x^2}{2\sigma_x^2}} dy \\
 &= \frac{p}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{x^2}{2\sigma_x^2}} \int_{|y|<\lambda} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(y-x)^2}{2\sigma_0^2}} dy,
 \end{aligned}$$

ce qui donne

$$\begin{aligned}
 E[x^2, X=1, \hat{X}=0] &= \int_{\mathbb{R}} u^2 p_{x, \hat{X}=0, X=1}(u) du \\
 &= \int_{\mathbb{R}} u^2 \frac{p}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{u^2}{2\sigma_x^2}} \int_{|y|<\lambda} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(y-x)^2}{2\sigma_0^2}} dy du \\
 &= p \int_{|y|<\lambda} \int_{\mathbb{R}} u^2 \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{u^2}{2\sigma_x^2}} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(y-u)^2}{2\sigma_0^2}} du dy \\
 &= p \int_{|y|<\lambda} \int_{\mathbb{R}} u^2 \frac{A'(y)}{\sqrt{2\pi\Sigma'^2}} e^{-\frac{(u-\mu'(y))^2}{2\Sigma'^2}} du dy,
 \end{aligned}$$

avec

$$\mu'(y) = \frac{y\sigma_x^2}{\sigma_0^2 + \sigma_x^2}, \quad \Sigma'^2 = \frac{\sigma_0^2\sigma_x^2}{\sigma_x^2 + \sigma_0^2} = \Sigma^2, \quad A'(y) = \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_0^2)}} e^{-\frac{y^2}{\sigma_x^2 + \sigma_0^2}} = A(y).$$

4. Bayes strikes back

Au final, on obtient

$$\begin{aligned}
 E[x^2, X=1, \hat{X}=0] &= p \int_{|y|<\lambda} \int_{\mathbb{R}} u^2 \frac{A'(y)}{\sqrt{2\pi\Sigma'^2}} e^{-\frac{(u-\mu'(y))^2}{2\Sigma'^2}} du dy \\
 &= p \int_{|y|<\lambda} A(y) \left(\frac{\sigma_0^2 \sigma_x^2}{\sigma_x^2 + \sigma_0^2} + \mu'(y)^2 \right) dy \\
 &= p \int_{|y|<\lambda} \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_0^2)}} e^{-\frac{y^2}{\sigma_x^2 + \sigma_0^2}} \left(\frac{\sigma_0^2 \sigma_x^2}{\sigma_x^2 + \sigma_0^2} + \left(\frac{y\sigma_x^2}{\sigma_0^2 + \sigma_x^2} \right)^2 \right) dy.
 \end{aligned}$$

Calcul de $E[|x - \hat{x}|^2]$ et conclusion. On a

$$\begin{aligned}
 E[|x - \hat{x}|^2] &= 2p \int_{\lambda}^{+\infty} \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_0^2)}} e^{-\frac{y^2}{2(\sigma_x^2 + \sigma_0^2)}} \left(\frac{\sigma_0^2 \sigma_x^2}{\sigma_x^2 + \sigma_0^2} + \left(\frac{y\sigma_0^2}{\sigma_0^2 + \sigma_x^2} \right)^2 \right) dy \\
 &\quad + 2(1-p) \int_{\lambda}^{+\infty} y^2 \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{y^2}{2\sigma_0^2}} dy \\
 &\quad + 2p \int_0^{\lambda} \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_0^2)}} e^{-\frac{y^2}{\sigma_x^2 + \sigma_0^2}} \left(\frac{\sigma_0^2 \sigma_x^2}{\sigma_x^2 + \sigma_0^2} + \left(\frac{y\sigma_x^2}{2(\sigma_0^2 + \sigma_x^2)} \right)^2 \right) dy \\
 &= 2p \frac{\sigma_0^2 \sigma_x^2}{\sigma_x^2 + \sigma_0^2} + 2p \int_{\lambda}^{+\infty} \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_0^2)}} e^{-\frac{y^2}{2(\sigma_x^2 + \sigma_0^2)}} y^2 \left(\frac{\sigma_0^2}{\sigma_0^2 + \sigma_x^2} \right)^2 dy \\
 &\quad + 2p \int_0^{\lambda} \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_0^2)}} e^{-\frac{y^2}{2(\sigma_x^2 + \sigma_0^2)}} y^2 \left(\frac{\sigma_x^2}{\sigma_0^2 + \sigma_x^2} \right)^2 dy \\
 &\quad + 2(1-p) \int_{\lambda}^{+\infty} y^2 \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{y^2}{2\sigma_0^2}} dy.
 \end{aligned}$$

4. Bayes strikes back

En dérivant

$$\begin{aligned}
\frac{dE[|x - \hat{x}|^2]}{d\lambda} &= -2p \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_0^2)}} e^{-\frac{\lambda^2}{\sigma_x^2 + \sigma_0^2}} \left(\frac{\sigma_0^2 \sigma_x^2}{\sigma_x^2 + \sigma_0^2} + \left(\frac{\lambda \sigma_0^2}{\sigma_0^2 + \sigma_x^2} \right)^2 \right) \\
&\quad - 2(1-p)\lambda^2 \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{\lambda^2}{2\sigma_0^2}} \\
&\quad + 2p \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_0^2)}} e^{-\frac{\lambda^2}{\sigma_x^2 + \sigma_0^2}} \left(\frac{\sigma_0^2 \sigma_x^2}{\sigma_x^2 + \sigma_0^2} + \left(\frac{\lambda \sigma_x^2}{\sigma_0^2 + \sigma_x^2} \right)^2 \right) \\
&= 2p \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_0^2)}} e^{-\frac{\lambda^2}{\sigma_x^2 + \sigma_0^2}} \left(\left(\frac{\lambda \sigma_x^2}{\sigma_0^2 + \sigma_x^2} \right)^2 - \left(\frac{\lambda \sigma_0^2}{\sigma_0^2 + \sigma_x^2} \right)^2 \right) \\
&\quad - 2(1-p)\lambda^2 \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{\lambda^2}{2\sigma_0^2}},
\end{aligned}$$

puis en annulant la dérivée

$$\begin{aligned}
0 &= \frac{dE[|x - \hat{x}|^2]}{d\lambda} \\
&= p \frac{1}{\sqrt{\sigma_x^2 + \sigma_0^2}} e^{-\frac{\lambda^2}{\sigma_x^2 + \sigma_0^2}} \left(\frac{\sigma_x^4 - \sigma_0^4}{(\sigma_0^2 + \sigma_x^2)^2} \right) - (1-p) \frac{1}{\sqrt{\sigma_0^2}} e^{-\frac{\lambda^2}{2\sigma_0^2}},
\end{aligned}$$

on obtient

$$\begin{aligned}
p \frac{1}{\sqrt{\sigma_x^2 + \sigma_0^2}} e^{-\frac{\lambda^2}{2(\sigma_x^2 + \sigma_0^2)}} \left(\frac{\sigma_x^4 - \sigma_0^4}{(\sigma_0^2 + \sigma_x^2)^2} \right) &= (1-p) \frac{1}{\sqrt{\sigma_0^2}} e^{-\frac{\lambda^2}{2\sigma_0^2}} \\
\Leftrightarrow e^{\frac{\lambda^2 \sigma_x^2}{2\sigma_0^2(\sigma_x^2 + \sigma_0^2)}} &= \frac{1-p}{p} \sqrt{\frac{\sigma_x^2 + \sigma_0^2}{\sigma_0^2}} \frac{(\sigma_0^2 + \sigma_x^2)^2}{\sigma_x^4 - \sigma_0^4} \\
\Leftrightarrow \frac{\lambda^2 \sigma_x^2}{2\sigma_0^2(\sigma_x^2 + \sigma_0^2)} &= \log \left(\frac{1-p}{p} \sqrt{\frac{\sigma_x^2 + \sigma_0^2}{\sigma_0^2}} \frac{(\sigma_0^2 + \sigma_x^2)^2}{\sigma_x^4 - \sigma_0^4} \right) \\
\Leftrightarrow \lambda^2 &= \frac{2\sigma_0^2(\sigma_x^2 + \sigma_0^2)}{\sigma_x^2} \log \left(\frac{1-p}{p} \sqrt{\frac{\sigma_x^2 + \sigma_0^2}{\sigma_0^2}} \frac{(\sigma_0^2 + \sigma_x^2)^2}{\sigma_x^4 - \sigma_0^4} \right) \\
\lambda_{Risk}^2 &= \lambda_{MAP}^2 + \frac{2\sigma_0^2(\sigma_x^2 + \sigma_0^2)}{\sigma_x^2} \log \left(\frac{(\sigma_0^2 + \sigma_x^2)^2}{\sigma_x^4 - \sigma_0^4} \right).
\end{aligned}$$

□

4. Bayes strikes back

La différence entre λ_{MAP} et λ_{Risk} est donc importante dès lors que $\sigma_x \approx \sigma_0$. En pratique, on peut souvent considérer σ_x grand devant σ_0 , la différence entre les deux estimateurs devenant alors négligeable pour les valeurs habituellement rencontrées pour σ_0 .

En conclusion, l'estimateur du MAP est donc approprié pour le modèle Bernoulli-Gaussien. En effet, l'estimation du support par cet estimateur est tout à fait justifiée par son optimalité pour le coût 0-1. Une fois ce support supposé connu, l'estimateur du MAP est confondu avec l'estimateur de la moyenne a posteriori pour le modèle Gaussien.

4.1.3. Choix pratique du seuil en débruitage

Les seuils λ_{MAP} et λ_{Risk} précédents nécessitent de connaître les paramètres p, σ_x, σ_0 du modèle (4.1). Pour le seuillage dur, une généralisation de l'approche SURE (Stein Unbiased Risk Estimator), nommée SCORE (Stein CONSistent Risk Estimator), a été effectuée dans [DPF13]. Cependant, une simple estimation des paramètres du modèle Bernoulli par un algorithme d'Espérance-Maximisation (EM, pour Expectation-Maximization) donne des résultats très satisfaisants, en permettant un débruitage automatique d'un signal parcimonieux dans une base. Cette approche par algorithme EM a de plus l'avantage de fournir un calcul simple de l'estimateur de la moyenne pour le modèle (4.1). En effet, les variables indicatrices X_n jouant le rôle des variables cachées, l'algorithme évalue à l'itération t les quantités $E[X_k | y_k, \sigma_x^t, \sigma_0^t]$. En notant $\hat{\sigma}_x, \hat{\sigma}_0$ les paramètres estimés, on obtient alors l'estimateur de la moyenne :

$$x_n^{MMSE} = E[X_k | y_k, \hat{\sigma}_x, \hat{\sigma}_0] \frac{\hat{\sigma}_x^2}{\hat{\sigma}_x^2 + \hat{\sigma}_0^2} y_n. \quad (4.3)$$

Cependant, les différences en terme de SNR entre l'estimateur de la moyenne et l'estimateur MAP apparaissent négligeables en pratique (de l'ordre de quelques dixièmes de dB).

4.1.4. Un algorithme de type IST pour le modèle BG

Lors d'une collaboration avec Thomas Rodet [KR11a], nous avons proposé d'estimer les coefficients de synthèse dans une union de dictionnaires orthonormés avec un algorithme de type EM qui s'apparente à un algorithme IST. Le modèle considéré est le suivant :

$$\mathbf{y} = \Phi \boldsymbol{\alpha} + \mathbf{e},$$

avec pour tout k $\alpha_k \sim \mathcal{B}\mathcal{N}(p, 0, \sigma_\alpha^2)$ et $\mathbf{e} \sim \mathcal{N}(0, \sigma_0^2 \mathbf{I}_M)$. L'algorithme contient toujours une étape de type descente de gradient suivie d'un seuillage effectué lui-même en deux étapes : une étape de sélection du support avec une règle de type MAP (4.2), puis une pondération "à la Wiener" connaissant ce support, le modèle se réduisant alors à un simple modèle Gaussien. Les étapes sont résumées dans l'algorithme 4.

L'estimation des variables indicatrices et des paramètres peut être vue comme une étape de type *Classification-EM* [CG92] pour estimer les paramètres d'un mélange de deux gaussiennes. En remplaçant cette étape par un algorithme EM, on obtient la réécriture donnée dans l'algorithme 5.

4. Bayes strikes back

Algorithm 4: IST-BG

1 Initialisation : $L = \|\Phi\|^2$, $\alpha = \mathbf{0}$

2 **repeat**

- E-step : $\mathbf{z} = \alpha + \frac{1}{L}\Phi^*(\mathbf{y} - \Phi\alpha)$
- M-step :
 1. Estimation du support :

$$\forall \lambda, X_\lambda = \begin{cases} 1 & \text{if } |z_\lambda| > \sqrt{\frac{2(\sigma_0^2 + L\sigma_\alpha^2)}{L\sigma_\alpha^2} \ln \left[\frac{1-p}{p} \sqrt{\frac{\sigma_0^2 + L\sigma_\alpha^2}{\sigma_0^2}} \right]} \\ 0 & \text{if not} \end{cases}$$

2. Estimation des coefficients : $\forall n, \alpha_n = \frac{L\sigma_\alpha^2}{L\sigma_\alpha^2 + \sigma_0^2} X_n z_n$

3. Estimation des paramètres : $\sigma^2 = \frac{\sum X_\lambda |\alpha_\lambda|^2}{|\Lambda|} \sigma_0^2 = \frac{\|\mathbf{y} - \Phi\alpha\|^2}{N} \quad p = \frac{|\Lambda|}{N_1}$

3 **until** convergence;

Algorithm 5: IST-BG version EM

1 Initialisation : $L = \|\Phi\|^2$, $\alpha = \mathbf{0}$

2 **repeat**

1. $\mathbf{z} = \alpha + \frac{1}{L}\Phi^*(\mathbf{y} - \Phi\alpha)$
2. Estimation de $P(X_n = 1), \sigma_0^2, \sigma_\alpha^2$ par un algorithme EM sur les variables cachées \mathbf{z}
3. Estimation du support : $X_n = 1$ si $P(X_n = 1) > 0.5$
4. Estimation des coefficients : $\forall n, \alpha_n = \frac{L\sigma_\alpha^2}{L\sigma_\alpha^2 + \sigma_0^2} X_n z_n$

3 **until** convergence;

Lorsque Φ est une base orthonormée, l'algorithme se réduit à l'estimation des paramètres du modèle et des probabilités $P(X_n = 1)$. On retrouve alors l'estimation par la moyenne *a posteriori* (4.3).

L'utilisation du modèle Bernoulli-Gaussien à été étudiée dans [Sou+11] avec un algorithme de type single best replacement, dans [DHD12] avec une approche en champ moyen ou encore dans [ZBJ09] avec une approximation continue du modèle Bernoulli-Gaussien. La section suivante présente le modèle SBL [Tip01; WR04], qui peut être vu comme une généralisation du modèle Bernoulli-Gaussien, mais qui en pratique s'étend plus facilement aux problèmes inverses.

4.1.5. Problèmes inverses : du Bernoulli-Gaussien au Sparse Bayesian Learning

En considérant la distribution discrète de Bernoulli comme un Dirac, et donc une Gaussienne dont la variance tend vers 0, on peut voir dans le modèle proposé par le SBL [Tip01 ; WR04] une généralisation du Bernoulli-Gaussien. L'idée du SBL est de supposer que tous les coefficients suivent une loi normale, tout en laissant libre la variance pour chaque coefficient. Reprenons le modèle de synthèse dans un dictionnaire Φ

$$\mathbf{y} = \Phi \boldsymbol{\alpha} + \mathbf{e}. \quad (4.4)$$

Le modèle SBL s'écrit simplement

$$\alpha_k \sim \mathcal{N}(0, \sigma_k^2). \quad (4.5)$$

Ainsi, si $\sigma_k \rightarrow 0$, on obtient effectivement un signal parcimonieux. Ce modèle, proposé par [Tip01] puis étudié plus spécifiquement dans un cadre d'estimation en signal par [WR04] a cet avantage de rester dans un cadre Gaussien², et génère effectivement de la parcimonie. On peut remarquer qu'en ajoutant un simple modèle sur la distribution des variances tel que

$$\forall k \quad P(\sigma_k = \sigma) = p \quad \text{et} \quad P(\sigma_k = 0) = 1 - p,$$

on retombe sur le modèle Bernoulli-Gaussien.

L'approche SBL nécessite d'estimer à la fois les coefficients de synthèse $\boldsymbol{\alpha}$ et les paramètres de variances σ_k pour tout k . En notant le vecteur de variance $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)$, une approche MAP classique avec estimation jointe donne l'estimation

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\sigma}} \frac{1}{2} \|\mathbf{y} - \Phi \boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\sigma} \circ \boldsymbol{\alpha}\|_2^2, \quad (4.6)$$

où \circ dénote la multiplication de deux vecteurs éléments par éléments. On reconnaît ici l'approche par moindres-carrés itérée (IRLS [Dau+10]).

Une autre procédure d'estimation possible pour le SBL est l'estimation par maximum de vraisemblance marginalisée (MMLE, pour Marginalized Maximum Likelihood Estimation), aussi appelée estimation par maximum de vraisemblance de type II [Tip01 ; WR04]. Elle repose sur la minimisation de $-\log p(\mathbf{y}|\boldsymbol{\sigma})$ et implique la marginalisation en $\boldsymbol{\alpha}$ de la vraisemblance jointe. Cet approche à l'inconvénient de recourir à une inversion explicite d'une matrice de taille K^2 et passe donc difficilement à l'échelle en pratique.

4.2. Synthèse temps-fréquence de rang faible

Si l'approche SBL a l'avantage d'une estimation simplifiée par rapport au modèle Bernoulli-Gaussien, les deux approches n'impliquent aucune structure sur les coefficients. Une généralisation du modèle Bernoulli-Gaussien permettant de modéliser des structures

2. On voit ici qu'on peut construire des priors tenant compte de la parcimonie et du caractère gaussien des données comme exigé en cosmologie (voir la discussion [Sta+13]). Les expériences confirmeront les remarques faites dans [Sta+13].

4. Bayes strikes back

(à l'aide de chaîne de Markov par exemple) sont les machines de Blotzman [DHD12]. On expose dans la sous-section suivante le modèle de synthèse temps-fréquence à rang faible (LRTFS, pour Low Rank Time-Frequency Synthesis) qui repose sur le modèle SBL, et permet de prendre en compte des structures de rang faible telles que rencontrées dans les approches de factorisation en matrices non négatives³ (NMF, pour Non Negative Matrix Factorization).

4.2.1. De la NMF au LRTFS

La NMF a rencontré un vif succès dans le traitement du signal. En effet, en décomposant un signal par une combinaison linéaire de motifs de base fréquentiels, cette approche permet d'obtenir un modèle structuré simple. Plus précisément, on note $\mathbf{V} \in \mathbb{R}^{FT}$, avec $FT = K$, la matrice du spectrogramme du signal \mathbf{y} . Cette matrice contient donc le module des coefficients d'analyse $\Phi^* \mathbf{y}$, en conservant la structure temps-fréquence inhérente au dictionnaire de Gabor. L'approche NMF consiste à approcher \mathbf{V} comme le produit de deux matrices non négatives $\mathbf{W} \in \mathbb{R}^{FR}$ et $\mathbf{H} \in \mathbb{R}^{RT}$ de rang faible ($R \ll \min(F, T)$) :

$$\mathbf{V} \simeq \mathbf{W}\mathbf{H}, \quad \mathbf{W}, \mathbf{H} \geq 0. \quad (4.7)$$

L'avantage de la positivité est d'interdire les interférences entre les différents motifs fréquentiels définis par les colonnes \mathbf{w}_f de \mathbf{W} . Ainsi, chaque motif fréquentiel du spectrogramme \mathbf{V} se voit comme l'addition de motifs fréquentiels élémentaires. L'inconvénient principal de l'approche NMF est la perte de toute information relative à la phase. Pour résoudre en partie ce problème, le modèle *Gaussian Composite Model* [GCM, FBD09a] propose un modèle probabiliste génératif du spectrogramme. L'estimation de ce modèle aboutit à la décomposition NMF par minimisation de la divergence d'Itakura-Saito, et permet la reconstruction des composantes latentes de rang 1 par simple filtrage de Wiener.

Nous avons proposé dans [FK14b; FK15b; KF17] un modèle génératif des échantillons temporels, qui permet :

- de résoudre le problème d'estimation de la phase, en partant du modèle GCM,
- d'étendre l'utilisation de la NMF aux problèmes inverses.

Reprenons le problème (4.4) avec un dictionnaire temps-fréquence de type Gabor

$$\mathbf{y} = \Phi \boldsymbol{\alpha} + \mathbf{e} = \sum_{t,f} \alpha_{t,f} \boldsymbol{\varphi} + \mathbf{e}.$$

En s'inspirant des modèles SBL et GCM, on définit le modèle LRTFS comme

$$\mathbf{x} = \Phi \boldsymbol{\alpha} + \mathbf{e}, \quad (4.8)$$

$$\alpha_{ft} \sim \mathcal{N}_c(\mathbf{0}, [\mathbf{W}\mathbf{H}]_{ft}) \quad (4.9)$$

$$\mathbf{e} \sim \mathcal{N}_c(\mathbf{0}, \lambda \mathbf{I}_T), \quad (4.10)$$

où \mathbf{W} et \mathbf{H} sont deux matrices non négatives de taille FR et RT . Les coefficients de synthèse α_{ft} sont donc supposés indépendants, connaissant \mathbf{W} et \mathbf{H} . On note de plus $\mathbf{v} \in \mathbb{R}^{FT}$ le vecteur de dimension K d'éléments $v_k = v_{ft} = [\mathbf{W}\mathbf{H}]_{ft}$.

3. La traduction appropriée est *matrices positives*.

4. Bayes strikes back

Si l'on ignore la structure de rang faible de \mathbf{v} , les équations (4.8)-(4.10) reviennent exactement au SBL. La différence fondamentale entre le modèle LRTFS proposé et le SBL est la prise en compte d'une structure de rang faible. En effet, dans le SBL, chaque coefficient α_n a un paramètre libre de variance ν_n . Cette propriété est fondamentale pour obtenir un effet de parcimonie [Tip01]. Dans le modèle LRTFS (4.8)-(4.10) les variances sont dorénavant liées de sorte que $\nu_m = \nu_{f_t} = [\mathbf{WH}]_{f_t}$.

On propose d'estimer \mathbf{W} , \mathbf{H} et $\boldsymbol{\alpha}$ par maximum de vraisemblance joint (MJLE, pour Maximum Joint Likelihood Estimation), ce qui revient à l'estimation de maximum de vraisemblance de type I [WR04]. MJLE se base sur la minimisation de la fonction objectif

$$\begin{aligned} C_{JL}(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H}, \lambda) &\stackrel{\text{def}}{=} -\log p(\mathbf{y}, \boldsymbol{\alpha} | \mathbf{W}, \mathbf{H}, \lambda) \\ &= -\log p(\mathbf{x} | \boldsymbol{\alpha}, \lambda) - \log p(\boldsymbol{\alpha} | \mathbf{W}, \mathbf{H}) \\ &= \frac{1}{\lambda} \|\mathbf{x} - \Phi \boldsymbol{\alpha}\|_2^2 + \sum_{f_n} \left[\frac{|\alpha_{f_n}|^2}{[\mathbf{WH}]_{f_n}} + \log [\mathbf{WH}]_{f_n} \right] + cst \\ &= \frac{1}{\lambda} \|\mathbf{x} - \Phi \boldsymbol{\alpha}\|_2^2 + D_{IS}(|\boldsymbol{\alpha}|^2 | \mathbf{v}) + \log(|\boldsymbol{\alpha}|^2) + cst. \end{aligned} \quad (4.11)$$

Le premier terme de l'équation (4.11) mesure classiquement la disparité entre le signal mesuré et son approximation. Le second terme assure que les coefficients de synthèse sont de rang faible, étant donnée leur structure matricielle temps-fréquence.

On décrit alors l'algorithme de minimisation alterné permettant d'obtenir un point stationnaire de $C_{JL}(\boldsymbol{\theta})$, où $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H}\}$. Ce type d'approche se justifie d'autant plus dans le cadre Bayésien, où les paramètres sont estimés par maximum de vraisemblance et les coefficients par moyenne *a posteriori*, confondus avec le maximum *a posteriori*, grâce au modèle Gaussien.

L'optimisation en $\boldsymbol{\alpha}$ étant donnés les autres paramètres se réduit à

$$\min_{\boldsymbol{\alpha}} \frac{1}{\lambda} \|\mathbf{x} - \Phi \boldsymbol{\alpha}\|_2^2 + \sum_{f_n} \frac{|\alpha_{f_n}|^2}{[\mathbf{WH}]_{f_n}}, \quad (4.12)$$

dont la solution analytique

$$\hat{\boldsymbol{\alpha}} = [\Phi^H \Phi + \lambda \text{diag}(\mathbf{v})^{-1}]^{-1} \Phi^H \mathbf{x},$$

requiert simplement la résolution d'un système linéaire. L'inversion directe de la matrice étant difficile dans les applications pratiques de traitement du signal compte-tenu de la taille des signaux, on utilise une procédure numérique. Nous avons choisi une implémentation de type seuillage itératif (cf. Chapitre 3) qui a l'avantage d'être simple et aussi rapide qu'une approche de type gradient-conjugué.

L'optimisation en \mathbf{W} et \mathbf{H} étant donné $\boldsymbol{\alpha}$ revient à évaluer

$$\min_{\mathbf{W}, \mathbf{H} \geq 0} \sum_{f_n} D_{IS}(|\alpha_{f_n}|^2 | [\mathbf{WH}]_{f_n}), \quad (4.13)$$

qui est un problème de type IS-NMF (décomposition NMF avec divergence d'Itakura-Saito). Ce problème d'optimisation non convexe est généralement résolu par mise à jour

4. Bayes strikes back

Algorithm 6: Alternate minimization for LRTFS

```

1 Set  $L = \|\Phi\|_2^2$  (or a larger value)
2 Compute the synthesis coefficients  $\mathbf{y} = \Phi^H \mathbf{x}$  (with matrix form  $\mathbf{Y}$ )
3 Set  $\alpha^{(0)} = \mathbf{y}$ 
4 Initialize  $\mathbf{W}^{(0)}$  and  $\mathbf{H}^{(0)}$  with the absolute values of the complex SVD of  $\mathbf{Y}$ 
5 Set  $i = 0$ 
6 repeat
7   %% Update  $\mathbf{W}$  and  $\mathbf{H}$  with MM
8   Compute spectrogram  $\mathbf{S}^{(i)} = [|\alpha_{fn}^{(i)}|^2]_{fn}$ 
9   Initialize inner loop :  $\mathbf{W} = \mathbf{W}^{(i)}, \mathbf{H} = \mathbf{H}^{(i)}$ 
10  repeat
11     $\mathbf{W} \leftarrow \mathbf{W} \circ \frac{[\mathbf{S}^{(i)} \circ (\mathbf{W}\mathbf{H})^{\circ-2}] \mathbf{H}^T}{[(\mathbf{W}\mathbf{H})^{\circ-1}] \mathbf{H}^T}$ 
12     $\mathbf{H} \leftarrow \mathbf{H} \circ \frac{\mathbf{W}^T [\mathbf{S}^{(i)} \circ (\mathbf{W}\mathbf{H})^{\circ-2}]}{\mathbf{W}^T [(\mathbf{W}\mathbf{H})^{\circ-1}]}$ 
13  until convergence;
14  Leave inner loop :  $\mathbf{W}^{(i+1)} = \mathbf{W}, \mathbf{H}^{(i+1)} = \mathbf{H}$ 
15  Set  $\mathbf{v}^{(i+1)} = \text{vect}[\mathbf{W}^{(i+1)} \mathbf{H}^{(i+1)}]$ 
16  %% Update  $\alpha$  with accelerated ISA
17  Initialize inner loop :  $\mathbf{a}^{(0)} = \mathbf{z}^{(0)} = \alpha^{(i)}$ 
18  Set  $j = 0$ 
19  repeat
20    % Descend
21     $\mathbf{z}^{(j+1/2)} = \mathbf{a}^{(j)} + \frac{1}{L} \Phi^H (\mathbf{x} - \Phi \mathbf{a}^{(j)})$ 
22    % Shrink
23     $\mathbf{z}^{(j+1)} = \frac{\mathbf{v}^{(i+1)}}{\mathbf{v}^{(i+1)} + \lambda/L} \circ \mathbf{z}^{(j+1/2)}$ 
24    % Accelerate
25     $\mathbf{a}^{(j+1)} = \mathbf{z}^{(j+1)} + \frac{j+1}{j+5} (\mathbf{z}^{(j+1)} - \mathbf{z}^{(j)})$ 
26     $j \leftarrow j + 1$ 
27  until convergence;
28  Leave inner loop :  $\alpha^{(i+1)} = \mathbf{z}^{(j+1)}$ 
29 until convergence;

```

alternée de \mathbf{W} et \mathbf{H} à l'aide d'algorithme de majoration-minimisation [CI11]. Cela revient en pratique à des mises à jour multiplicatives données dans l'algorithme 6.

LRTFS suppose que les coefficients de synthèse α_{ft} suivent le GCM (voir l'équation (4.9)). Ainsi, α_{ft} peut être écrit comme la somme de composantes latentes gaussiennes de sorte que $\alpha_{fn} = \sum_r \alpha_{rft}$, avec $\alpha_{rft} \sim N_c(0, w_{fr} h_{rt})$. En notant α_r le vecteur colonne de dimension F composé des coefficients $\{\alpha_{rft}\}_{ft}$, l'équation. (4.8) peut s'écrire comme

$$\mathbf{y} = \sum_r \Phi \alpha_r + \mathbf{e} = \sum_r \mathbf{c}_r + \mathbf{e}, \quad (4.14)$$

4. Bayes strikes back

où $\mathbf{c}_r = \Phi \boldsymbol{\alpha}_r$. Les composantes \mathbf{c}_r représentent l'expression temporelle du motif spectral \mathbf{w}_r , la r^{eme} colonne de \mathbf{W} . Étant donnée l'estimation de $\boldsymbol{\alpha}$, \mathbf{W} et \mathbf{H} , les composantes peuvent être reconstruites comme $\hat{\mathbf{c}}_r^{\text{MMSE}} = \Phi \hat{\boldsymbol{\alpha}}_r^{\text{MMSE}}$ avec

$$\hat{\boldsymbol{\alpha}}_r^{\text{MMSE}} \stackrel{\text{def}}{=} \mathbb{E}[\boldsymbol{\alpha}_r | \mathbf{y}, \hat{\boldsymbol{\theta}}] = \mathbb{E}[\boldsymbol{\alpha}_r | \hat{\boldsymbol{\alpha}}, \hat{\mathbf{W}}, \hat{\mathbf{H}}] \quad (4.15)$$

dont les coefficients sont donnés par

$$\hat{\alpha}_{rft}^{\text{MMSE}} = \frac{\hat{w}_{fr} \hat{h}_{rt}}{[\hat{\mathbf{W}}\hat{\mathbf{H}}]_{ft}} \hat{\alpha}_{ft}. \quad (4.16)$$

Avec ces estimations, les composantes latentes sont reconstruites à l'aide d'un masque de Wiener temps-fréquence sur les coefficients de synthèse.

On illustre les performances de l'approche LRTFS par rapport à une décomposition IS-NMF standard sur l'exemple de piano utilisé dans [FBD09b]. Cette séquence a une structure simple : quatre notes sont jouées en même temps lors de la première mesure, puis toutes les paires possibles composées par ces notes sont jouées lors des mesures suivantes. Les résultats sont résumés sur la figure 4.1. LRTFS est capable de retrouver les quatre notes dans les quatre premières composantes. La cinquième composante retrouve la partie transitoire produite par le marteau et la pédale. Les cinq composantes restantes ne sont pas audibles en raison de la procédure de débruitage de l'estimation. IS-NMF n'arrive pas à retrouver cette partie transitoire, et sépare la première note sur deux composantes. Le bruit d'entrée en lui étalé sur les cinq dernières composantes.

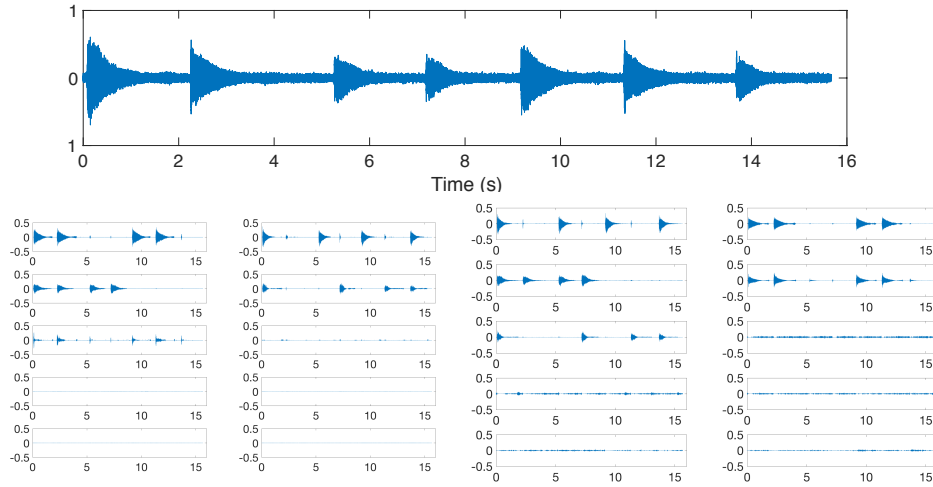


FIGURE 4.1. – Décomposition d'une séquence de piano construite sur quatre notes. En haut, le signal d'entrée. À gauche, la décomposition LRTFS. À droite, la décomposition IS-NMF. Les composantes temporelles sont montrées par ordre décroissant d'énergie.

4. Bayes strikes back

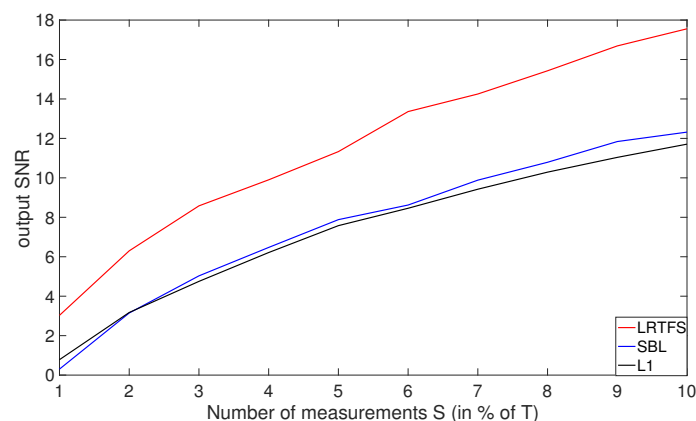


FIGURE 4.2. – Estimation de la séquence de piano après échantillonnage compressé par LRTFS, SBL/IRLS et regularisation ℓ_1 .

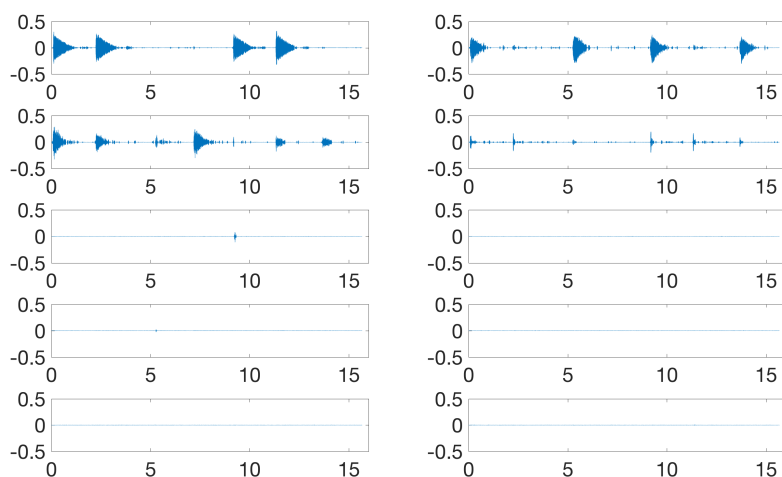


FIGURE 4.3. – Composantes latentes (de "rang 1") de la séquence de piano estimée par LRTFS après échantillonnage compressé. Les composantes temporelles sont affichées par ordre décroissant d'énergie.

4.2.2. Compressive LRTFS

Un avantage majeur de l'approche LRTFS sur la NMF, est de pouvoir s'utiliser comme prior dans un problème inverse. Un exemple simple de problème inverse pour lequel les approches parcimonieuses sont idéales est le compressed sensing. On montre ici que la parcimonie peut être remplacée efficacement par le prior de rang faible exposé précédemment.

4. Bayes strikes back

Soit $\mathbf{x} \in \mathbb{C}^T$ le signal source, qu'on suppose mesuré par un opérateur linéaire $\mathbf{A} \in \mathbb{C}^{S \times T}$, avec $S < T$, donnant le signal $\mathbf{y} \in \mathbb{C}^S$. On suppose le modèle d'observation suivant

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \quad (4.17)$$

$$= \mathbf{A}\Phi\boldsymbol{\alpha} + \mathbf{e}, \quad (4.18)$$

où $\Phi \in \mathbb{C}^{T \times M}$ est un dictionnaire temps-fréquence, $\boldsymbol{\alpha}$ les coefficients de synthèse de \mathbf{x} et \mathbf{e} un terme de résidu supposé blanc gaussien. \mathbf{A} est choisi comme une matrice aléatoire gaussienne (connue). Avec un prior du type LRTFS, l'approche MJLE revient à minimiser la fonction objectif suivante :

$$\begin{aligned} C_{CS}(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H}) &= -\log p(\mathbf{b}, \boldsymbol{\alpha} | \mathbf{W}, \mathbf{H}, \lambda) \\ &= \frac{1}{\lambda} \|\mathbf{b} - \mathbf{A}\Phi\boldsymbol{\alpha}\|_2^2 + D_{IS}(|\boldsymbol{\alpha}|^2 | \mathbf{W}\mathbf{H}) + \log(|\boldsymbol{\alpha}|^2) + cst. \end{aligned} \quad (4.19)$$

Minimiser $C_{CS}(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H})$ revient à minimiser la fonction $C_{JL}(\boldsymbol{\alpha}, \mathbf{W}, \mathbf{H})$ donné par (4.11) en remplaçant Φ par $\Gamma = \mathbf{A}\Phi$.

On évalue les qualités de reconstruction de la séquence de piano, précédemment utilisée, de longueur 15,6 s et échantillonnée à 11025 Hz, selon un nombre de mesures S allant de $T/100$ à $T/10$. On compare les méthodes d'estimation par LRTFS (pour $K \in \{5, 8, 10, 15, 20, 30\}$), SBL type-I et régularisation ℓ_1 . Les résultats sont résumés sur les figures 4.2 et 4.3 où l'on voit que l'approche LRTFS permet un gain significatif par rapport aux approches simplement parcimonieuses. Toutes les valeurs de K donnent le même SNR à 0.5 dB près.

5 ■ Return of Shrinkage

La définition d'une pénalité appropriée permettant de prendre en compte les structures voulues dans les approches par optimisation n'est pas toujours évidente. Nous avons vu dans le chapitre 3 comment estimer un signal dans un problème inverse à l'aide d'un algorithme IST, la fonction de seuillage étant obtenue comme l'opérateur de proximité de la pénalité. On se pose ici la question de savoir dans quelle mesure il est possible de choisir directement la fonction de seuillage dans l'algorithme IST plutôt que de définir la pénalité. Dans la première section, on présente l'étude de l'algorithme IST avec n'importe quelle règle de seuillage au sens d'Antoniadis, règle qui considère chaque coefficient individuellement. Notons que ces résultats peuvent se généraliser directement au seuillage par groupe, sur des groupes de coefficients indépendants. La section suivante introduit la notion de *parcimonie sociale* avec des opérateurs de seuillage structurés prenant en compte le voisinage d'un coefficient.

Contributions principales :

1. La définition d'un algorithme de seuillage itératif et de fonctions de seuillage [Kow14].
2. La définition de la *Parcimonie Sociale* [KSD13].

Collaborations et projets associés : L'étude des algorithmes de seuillage itératif a bénéficié du cadre donné par le projet MAORI que j'ai coordonné. L'étude des opérateurs de parcimonie sociale a été faite lors d'une collaboration avec Monika Dörfler et Kai Siedenburg, principalement lors de mes visites à Vienne.

Références personnelles citées dans ce chapitre :

- [Kow14], qui étudie l'algorithme IST avec une fonction de seuillage quelconque.
- [KT09b], [KT09a], [KSD13], dans lesquels les opérateurs de seuillage structurés sont introduits puis étudiés sous le terme de *parcimonie sociale*. Les articles [KT09b] et [KT09a] ont été écrits pendant ma thèse.

5.1. Fonctions de seuillage et algorithme de seuillage itératif

Le chapitre 3 nous a permis d'introduire le cadre de l'optimisation pour la résolution de problèmes inverses à l'aide de la synthèse parcimonieuse dans un dictionnaire. On a vu qu'en pratique, l'estimation se fait par un algorithme de seuillage itératif, possiblement relaxé, simple à implémenter. Cependant, l'utilisation de pénalité convexe comme le ℓ_1 a le défaut d'introduire un biais sur l'estimation des coefficients qui est souvent corrigé en pratique par une approche en deux étapes [WNF09] :

1. Estimation du support par minimisation ℓ_1 :

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1$$

$$\hat{\Delta} = \operatorname{supp} \{\boldsymbol{\alpha}\} .$$

2. Estimation de la valeur des coefficients de synthèse non nuls, sachant le support :

$$\hat{\boldsymbol{\alpha}}_{\Delta} = \underset{\boldsymbol{\alpha}_{\Delta}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\boldsymbol{\Phi}_{\Delta}\boldsymbol{\alpha}_{\Delta}\|_2^2 . \quad (5.1)$$

Une autre approche est d'essayer de minimiser directement le problème ℓ_0 (2.4), en remplaçant le seuillage doux par un seuillage dur. Dans le même temps, tout un panel d'opérateurs de seuillage ont été étudiés [Ant+07]. La question de pouvoir utiliser n'importe quel opérateur de seuillage dans un algorithme de seuillage itératif a donc naturellement été étudiée dans [Kow14]. En voici les résultats principaux qui abordent la question sous l'angle des fonctions semi-convexes. Cela permet d'étudier la convergence des algorithmes itératifs quelque soit l'attache aux données (à condition qu'elle soit Lipschitz différentiable). Cela montre de plus qu'il est possible de relaxer l'algorithme ce qui permet d'accélérer grandement l'algorithme en pratique.

L'outil principal pour donner les résultats de convergence est donc la notion de fonction semi-convexe (voir par exemple [CH00] et les références), qu'on rappelle ici

Définition 5.1 (Fonction c -semi-convexe)

Une fonction $f : \mathbb{R}^N \rightarrow \mathbb{R}$ est dite c -semi-convexe s'il existe une constante finie $c \geq 0$ telle que

$$\tilde{f}(x) := f(x) + \frac{c}{2} \|x\|^2$$

est une fonction convexe.

L'autre notion essentielle, qui sera reliée à la notion de pénalité semi-convexe, est la définition d'un opérateur de seuillage [Ant+07] donné au chapitre 3 et qu'on rappelle ci-après.

Définition 5.2 (Opérateur de seuillage)

$\mathbb{S}(\cdot; \lambda)$ est un opérateur de seuillage ssi

1. $\mathbb{S}(\cdot; \lambda)$ est une fonction impaire. On note $\mathbb{S}_+(\cdot; \lambda)$ sa restriction à \mathbb{R}_+ ,

5. Return of Shrinkage

2. $0 \leq \mathbb{S}_+(x; \lambda) \leq x, \forall x \in \mathbb{R}_+$,
3. \mathbb{S}_+ est croissante sur \mathbb{R}_+ , et $\lim_{x \rightarrow +\infty} \mathbb{S}_+(x; \lambda) = +\infty$.

Il est montré dans [Ant+07] que l'on peut associer une pénalité P à n'importe quel opérateur de seuillage, de sorte que $\mathbb{S} = \text{prox}_P$ avec

$$P(x; \lambda) = \int_0^{|x|} \mathbb{S}_+^{-1}(t; \lambda) - t \, dt,$$

où $\mathbb{S}^{-1}(t; \lambda) = \sup\{x; \mathbb{S}(x; \lambda) \leq t\}$ et $\mathbb{S}^{-1}(-t; \lambda) = -\mathbb{S}^{-1}(t; \lambda)$. On montre dans le théorème suivant qu'une telle pénalité est semi-convexe.

Théorème 5.1 (Semi-convexité des pénalités associées aux opérateurs de seuillage)

La pénalité $P(\cdot; \lambda)$ associée à un opérateur de seuillage $\mathbb{S}(\cdot; \lambda)$ est au moins 1-semi-convexe. De plus, si \mathbb{S} est continue et son quotient aux différences est borné par a , alors P est au moins $1 - \frac{1}{a}$ semi-convexe. De plus, P est ℓ -fortement convexe ssi $a \leq \frac{1}{1+\ell}$.

La preuve est disponible dans [Kow14].

5.1.1. Exemples d'opérateurs de seuillage

On donne ici trois grands exemples de règles de seuillage : le seuillage dur, SCAD et le seuillage de la "garotte non négative" (NNG, pour Non Negative Garotte) connu aussi comme seuillage de Wiener Empirique (EW, pour Empirical Wiener), ainsi que leur pénalité associée. Ces règles de seuillage sont représentées sur la figure 5.1.1, où l'on peut voir que les seuillages SCAD et NNG sont des sortes de compromis entre les seuillages doux et dur.

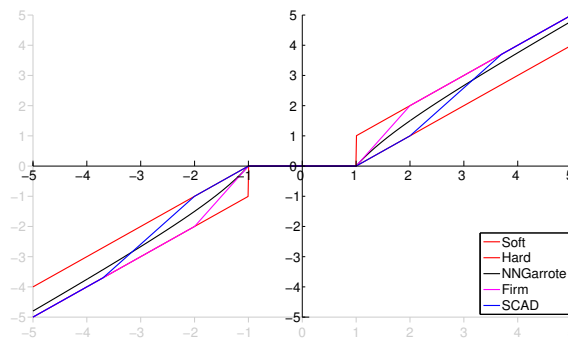


FIGURE 5.1. – Quelques règles de seuillage.

5.1.2. IST avec un opérateur de seuillage quelconque

On peut montrer la convergence de l'algorithme IST lorsque \mathbb{S} n'est plus nécessairement un opérateur de proximité d'une fonction convexe mais un opérateur de seuillage tel que défini à la définition 5.2. Du point de vue l'optimisation, on peut considérer la minimisation de la fonction

$$F(\boldsymbol{\alpha}) = \mathcal{L}(\boldsymbol{\alpha}) + P(\boldsymbol{\alpha}; \lambda), \quad (5.2)$$

où \mathcal{L} est une fonction L -Lipschitz différentiable et P une fonction c -semi-convexe d'opérateur de seuillage associé $\mathbb{S}(\cdot; \lambda)$. On rappelle l'algorithme IST associé au problème de minimisation de F dans l'algorithme 7 pour sa version simple, et dans l'algorithme 8 pour sa version relaxée.

Algorithm 7: Algorithme de seuillage itératif (IST).

```

1 Set  $L = \|\nabla f\|_2^2$  (or a larger value)
2 Set  $\boldsymbol{\alpha}^{(0)} \in \mathbb{R}^N$  and  $t = 0$ 
3 repeat
4    $\boldsymbol{\alpha}^{(t+1)} = \mathbb{S}(\boldsymbol{\alpha}^{(t)} + \frac{1}{L}\nabla\mathcal{L}(\boldsymbol{\alpha}^{(t)}), \lambda)$ 
5    $t = t + 1$ 
6 until convergence;
```

Algorithm 8: Algorithme de seuillage itératif relaxé.

```

1 Set  $L = \|\mathbf{A}\|_2^2 \|\Phi\|_2^2$  (or a larger value)
2 Set  $\boldsymbol{\gamma}^{(0)} \in \mathbb{R}^N$  and  $t = 0$ 
3 repeat
4    $\boldsymbol{\alpha}^{(t+1)} = \mathbb{S}(\boldsymbol{\gamma}^{(t)} + \frac{1}{L}\nabla\mathcal{L}(\boldsymbol{\gamma}^{(t)}), \lambda)$ 
5    $\boldsymbol{\gamma}^{(t+1)} = \boldsymbol{\alpha}^{(t+1)} + \mu^{(t)}(\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{\alpha}^{(t)})$ 
6    $t = t + 1$ 
7 until convergence;
```

Alors, on peut énoncer les théorèmes suivants, dont les preuves sont disponibles dans l'article [Kow14]).

Théorème 5.2 (Convergence de IST)

Toute sous-suite convergente de $\{\boldsymbol{\alpha}^{(t)}\}$ générée par l'algorithme IST 7 converge vers un point critique de F , où P est la pénalité c -semi-convexe associée à \mathbb{S} .

Théorème 5.3 (Convergence de IST relaxé)

Si \mathcal{L} est convexe et $\mu^{(t)} < \sqrt{1 - c/L}$, alors toute sous-suite convergente de $\{\boldsymbol{\alpha}^{(t)}\}$ générée par l'algorithme IST relaxé 8 converge vers un point critique de F .

5. Return of Shrinkage

Cependant, comme il est souvent plus simple de choisir la fonction de seuillage plutôt que la pénalité semi-convexe associée, on donne le corollaire suivant.

Théorème 5.4 (Corollaire)

Soit \mathbb{S} un opérateur de seuillage et \mathcal{L} une fonction convexe L -Lipschitz différentiable. Alors l'algorithme IST 7 converge. De plus, si $\mathbb{S}' \leq a$, alors l'algorithme IST relaxé 8, avec $\mu^{(t)} < \frac{1}{a}$, converge.

On peut remarquer que, avec l'opérateur de seuillage dur, on doit choisir un pas inférieur à $\frac{1}{L}$. Le résultat précédent sur la convergence de l'algorithme relaxé ne peut donc pas s'appliquer.

5.2. Parcimonie Sociale

Afin d'exploiter certaines structures sur les coefficients de synthèse, on construit sur les opérateurs de seuillage classiques, tels que définis dans la section précédente, de nouveaux opérateurs de seuillage prenant en compte le voisinage d'un coefficient. Pour un index k d'un coefficient de synthèse, on associe un voisinage pondéré $\mathcal{N}(k) = \{k' \in \mathcal{I} : w_{k'}^{(k)} \neq 0\}$ avec les poids $w_{k'}^{(k)}$ tels que $w_{k'}^{(k)} \geq 0$ et $\sum_{k' \in \mathcal{N}(k)} w_{k'}^{(k)} = 1$. On illustre cette notion de voisinage sur la figure 5.2.

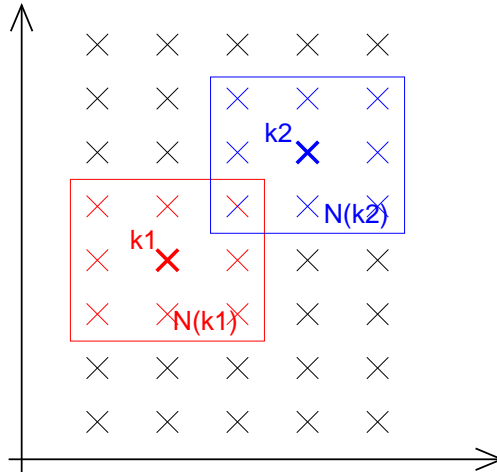


FIGURE 5.2. – Le voisinage du coefficient k_1 est donné par la fenêtre rouge, et le voisinage du coefficient k_2 par la fenêtre bleue. Ces deux voisinages partagent un même coefficient.

Une fois ces voisinages définis, on peut définir l'opérateur du group-lasso fenêtré (WG-

5. Return of Shrinkage

Lasso) comme

$$\mathbb{S}_\lambda^{wgl}(\boldsymbol{\alpha}) : \mathbb{C}^N \rightarrow \mathbb{C}^N$$

$$\boldsymbol{\alpha} \mapsto \underline{\boldsymbol{\alpha}},$$

tel que pour tout k ,

$$\underline{\alpha}_k = \alpha_k \left(1 - \frac{\lambda}{\sqrt{\sum_{k' \in \mathcal{N}(k)} w_{k'}^{(k)} |\alpha_{k'}|^2}} \right)^+. \quad (5.3)$$

L'idée d'un tel opérateur de seuillage est de sélectionner un coefficient si l'énergie de son voisinage est assez grande, et inversement, de pouvoir seuiller un coefficient isolé. Une telle notion de voisinage se retrouve dans [CS01], où une règle similaire est introduite dans le contexte de l'estimation SURE en ondelettes. L'opérateur du WG-Lasso a été introduit lors de ma thèse, dans [KT09b] et [KT09a].

Dans [KSD13], on montre comment relier ces opérateurs aux opérateurs de proximité usuels. Ils sont donc d'une certaine manière liés à la minimisation d'une fonctionnelle. Pour cela, on montre que les groupes implicitement créés par les voisinages peuvent être définis formellement à l'aide d'un opérateur d'expansion. Un tel opérateur plonge les coefficients dans un espace de grande dimension en les recopiant, de sorte que les groupes définis par les voisinages deviennent des groupes indépendants comme pour le group-lasso.

Définition 5.3 (Opérateur d'expansion)

Soit $\boldsymbol{\alpha} \in \mathbb{C}^N$. Soit $\mathcal{E} : \mathbb{C}^N \rightarrow \mathbb{C}^{N \times N}$ un opérateur d'expansion tel que

$$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_N) \mapsto$$

$$(\sqrt{w_1^{(1)}} \alpha_1, \dots, \sqrt{w_N^{(1)}} \alpha_N, \dots, \sqrt{w_1^{(N)}} \alpha_1, \dots, \sqrt{w_N^{(N)}} \alpha_N),$$

avec $w_i^{(j)} \geq 0$ et $\sum_j w_i^{(j)} = 1$ pour tout i .

Les conditions choisies sur les poids nous permettent d'assurer le fait que \mathcal{E} est une isométrie.

On note par la suite \mathbf{E} la matrice $N^2 \times N$ associée à \mathcal{E} . Si l'on peut revenir dans l'espace original à l'aide de l'opérateur \mathbf{E}^T , on utilisera par la suite l'inverse à gauche suivant :

$$\mathbf{D} : \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^N$$

$$\mathbf{z} = (z_1^1, \dots, z_N^1, \dots, z_1^N, \dots, z_N^N) \mapsto \mathbf{x}$$

$$\text{tel que } \forall k, x_k = \frac{1}{\sqrt{\sum_j w_k^{(j)}}} z_k^k, \quad (5.4)$$

$\mathbf{DE} = I$ et \mathbf{ED} est une projection oblique bi-orthogonale. On peut alors re-définir le WG-Lasso comme

5. Return of Shrinkage

$$\mathbb{S}_\lambda = \mathbf{D} \circ \text{prox}_{\lambda \|\cdot\|_{2,1}} \circ \mathbf{E},$$

où l'on retrouve l'opérateur de proximité du group-lasso. L'itération principale de l'algorithme de seuillage itératif se récrit alors

$$\begin{aligned} \mathbf{z}^{(k)} &= \mathbf{E} \mathbf{D} \text{prox}_{\frac{\lambda}{\gamma} \|\cdot\|_{2,1}} \left(\tilde{\mathbf{z}}^{(k-1)} \right) \\ \boldsymbol{\alpha}^k &= \mathbf{D} \mathbf{z}^k \\ \text{avec } \tilde{\mathbf{z}}^{(k-1)} &= \mathbf{z}^{(k-1)} + \frac{\mathbf{E}}{\gamma} \boldsymbol{\Phi}^* (\mathbf{y} - \boldsymbol{\Phi} \mathbf{E}^T \mathbf{z}^{(k-1)}) \end{aligned}$$

qui peut être vu comme un pas de descente proximale suivi d'une projection oblique.

Cette remarque permet de définir un nouvel opérateur utilisant l'inverse à gauche \mathbf{E}^T au lieu de \mathbf{D} . On obtient alors un nouvel opérateur **orth-WG-Lasso** :

$$\begin{aligned} \mathbb{S}_\lambda(\boldsymbol{\alpha}) : \mathbb{C}^N &\rightarrow \mathbb{C}^N \\ \boldsymbol{\alpha} &\mapsto \underline{\boldsymbol{\alpha}}, \end{aligned}$$

où pour tout k ,

$$\underline{\alpha}_k = \alpha_k \sum_j w_k^{(j)} \left(1 - \frac{\lambda}{\sqrt{\sum_{j' \in \mathcal{N}(j)} w_{j'}^{(j)} |\alpha_{j'}|^2}} \right)^+. \quad (5.5)$$

L'algorithme 7 peut alors être vu comme la composition d'une descente proximale classique suivie d'une projection orthogonale

$$\begin{aligned} \mathbf{z}^{(k)} &= \mathbf{E} \mathbf{E}^T \text{prox}_{\frac{\lambda}{\gamma} \|\cdot\|_*} \left(\tilde{\mathbf{z}}^{(k-1)} \right) \\ \boldsymbol{\alpha}^k &= \mathbf{E}^T \mathbf{z}^k \\ \text{où } \tilde{\mathbf{z}}^{(k-1)} &= \mathbf{z}^{(k-1)} + \frac{\mathbf{E}}{\gamma} \boldsymbol{\Phi}^* (\mathbf{y} - \boldsymbol{\Phi} \mathbf{E}^T \mathbf{z}^{(k-1)}). \end{aligned} \quad (5.6)$$

L'avantage d'un tel opérateur est donc d'assurer la convergence de l'algorithme vers un point fixe en appliquant [Com04, théorème 6.3 et corollaire 6.5].

5.3. Différents opérateurs de seuillage

Dans cette section, on résume les différents opérateurs de seuillage (avec ou sans voisinage) les plus fréquemment utilisés, en particulier dans le chapitre 6 suivant dédié aux applications. On présente les opérateurs coordonnée par coordonnée k , où la double indexation $k = (g, m)$ est utilisée pour les pénalités de type Group-Lasso, avec g représentant l'indice des groupes et m l'indice des coefficients dans un groupe.

5. Return of Shrinkage

Lasso/Seuillage doux

$$\mathbb{S}^L(\alpha_k; \lambda) = \alpha_k \left(1 - \frac{\lambda}{|\alpha_k|}\right)^+ \quad (5.7)$$

Empirical Wiener/Seuillage NonNegative Garotte

$$\mathbb{S}^{EW}(\alpha_k) = \alpha_k \left(1 - \frac{\lambda^2}{|\alpha_k|^2}\right)^+ \quad (5.8)$$

Group-Lasso

$$\mathbb{S}^{GL}(\alpha_{gm}) = \alpha_{gm} \left(1 - \frac{\lambda}{\sqrt{\sum_m |\alpha_{g,m}|^2}}\right)^+ \quad (5.9)$$

Group Empirical Wiener/ NonNegative Garotte

$$\mathbb{S}^{GEW}(\alpha_{gm}) = \alpha_{gm} \left(1 - \frac{\lambda^2}{\sum_m |\alpha_{g,m}|^2}\right)^+ \quad (5.10)$$

Windowed Group-Lasso

$$\mathbb{S}^{WGL}(\alpha_k) = \alpha_k \left(1 - \frac{\lambda}{\sqrt{\sum_{k' \in \mathcal{N}(k)} |\alpha_{k'}|^2}}\right)^+ \quad (5.11)$$

Persistent Empirical Wiener

$$\mathbb{S}^{PEW}(\alpha_k) = \alpha_k \left(1 - \frac{\lambda^2}{\sum_{k' \in \mathcal{N}(k)} |\alpha_{k'}|^2}\right)^+ \quad (5.12)$$

orthogonal Windowed Group-Lasso

$$\mathbb{S}^{OWGL}(\alpha_k) = \alpha_k \sum_{k' \in \mathcal{N}(k)} \left(1 - \frac{\lambda}{\sqrt{\sum_{k'' \in \mathcal{N}(k')} |\alpha_{k''}|^2}}\right)^+ \quad (5.13)$$

orthogonal Persistent Empirical Wiener

$$\mathbb{S}^{OPEW}(\alpha_k) = \alpha_k \sum_{k' \in \mathcal{N}(k)} \left(1 - \frac{\lambda^2}{\sum_{k'' \in \mathcal{N}(k')} |\alpha_{k''}|^2}\right)^+ \quad (5.14)$$

6 ■ Applications awaken

J'ai choisi de me concentrer sur deux classes d'applications auxquelles j'ai contribué pendant mes années de recherche au L2S : le problème de restauration des signaux audio (le débruitage et la désaturation) que je présente dans la première section, et la localisation de source magnéto-électroencéphalographique (M/EEG) présentée dans la section suivante. Mes contributions en séparation de sources, qui mériteraient un chapitre entier, sont volontairement omises dans la mesure où elles sont, pour la plupart, présentées de façon précise dans le manuscrit de thèse de Fangchen Feng [Fen17].

Ce chapitre présente deux classes d'application auxquelles j'ai contribué pendant mes années de recherche au L2S : les applications liées aux signaux audio et la localisation de source magnéto-électroencéphalographique (M/EEG). Pour les signaux audio, j'ai choisi de présenter les résultats obtenus en débruitage et en dé-saturation, principalement lors de mes collaborations avec Monika Dörfler et Kai Siedenburg [KSD13; SKD14]. Les résultats obtenus sur le problème de localisation de sources M/EEG ont été obtenus lors mes collaborations avec Alexandre Gramfort [GK09; Gra+11a; Str+11; GKH12; Gra+13]. Lors de ma délégation INRIA dans l'équipe Parietal, j'ai regardé avec Alexandre Gramfort la prise en compte de bruit coloré pour la localisation de sources M/EEG. Des premiers résultats non montrés ici ont été publiés dans [KG15a].

Contributions principales :

1. Résultats sur le débruitage et la dé-saturation en restauration audio [KSD13], [SKD14].
2. Résultats sur la localisation de sources M/EEG [GKH12; Gra+13].

Collaborations et projets associés : Les travaux sur la restauration audio font suite à la collaboration avec Monika Dörfler et Kai Siedenburg sur la parcimonie sociale. Les contributions en localisation de sources sont issues d'une longue collaboration avec Alexandre Gramfort débuté lorsqu'on était chacun en thèse. Elles ont toutes deux profité du projet MAORI dont j'étais coordinateur.

Références personnelles citées dans ce chapitre :

- [SKD14] illustre les opérateurs de type WG-Lasso sur le problème de débruitage.
- [SKD14] est consacré au problème de dé-saturation audio.
- [GKH12] étudie l'emploi des normes mixtes pour la localisation de source M/EEG, tandis que [Gra+13] introduit la synthèse temps-fréquence pour ce même problème.

6. Applications awaken

- [GK09; Gra+11a; Str+11] sont des articles de conférences reprenant les deux idées principales introduits précédemment.

6.1. Signaux audio

Les résultats sur le débruitage permettent d'évaluer la pertinence de différents opérateurs de seuillage. Le lien avec les approches historiques de type soustraction spectrale est souligné. Je me limite ici au bruit blanc Gaussien. L'extension aux bruits colorés dans les approches par seuillage simple est directe, et fait l'objet d'un sujet de travaux pratiques dans un cours sur la restauration des signaux audio dans le master ATSI (TP "De-Hissing").

L'application de désaturation présentée ensuite peut se formuler comme un problème inverse linéaire, pour lequel les opérateurs de seuillage qui ne sont pas associés à une pénalité convexe obtiennent d'excellents résultats, les seconds échouant sur cette tâche. Les résultats obtenus ont permis d'améliorer l'état de l'art de plusieurs dB en terme de SNR.

6.1.1. Débruitage

Le débruitage de signal peut-être considéré comme le problème inverse le plus simple. Il s'écrit :

$$\mathbf{y} = \mathbf{x} + \mathbf{e} = \Phi \boldsymbol{\alpha} + \mathbf{e},$$

où \mathbf{e} est supposé Gaussien de densité spectrale $E(f)$. Si l'on considère le problème dans le domaine transformé de Gabor à l'analyse, on a :

$$Y(t, f) = X(t, f) + E(f).$$

En supposant E et X indépendants, on a

$$E[|Y(t, f)|^2] = E[|X(t, f)|^2] + E[|N(f)|^2].$$

La densité spectrale du bruit, qu'on notera $E[|N(f)|^2] = \sigma_f^2$, peut être estimée par la méthode de Welsh [Wel67]. On obtient l'estimateur par soustraction spectrale :

$$X(t, f) = Y(t, f) \left(1 - \frac{\lambda^2 \sigma_f^2}{|Y(t, f)|^2} \right)^{1/2},$$

estimateur qui se généralise par [GRC02] ,

$$X(t, f) = Y(t, f) \left(1 - \frac{(\lambda \sigma_f)^\gamma}{|Y(t, f)|^\gamma} \right)^\beta, \quad (6.1)$$

avec comme choix courant $\gamma = 2$ et $\beta = 1$, ce qui revient au seuillage de type Empirical Wiener présenté dans les sections 5.1.1 et 5.3. Les choix $\gamma = 1$ et $\beta = 1$ reviennent au seuillage doux.

Nous avons comparé les différents opérateurs de seuillage ainsi que l'approche par seuillage itératif pour le débruitage d'un bruit blanc. Le signal choisi est un extrait de quintet de Jazz comportant batterie, basse, piano, saxophone et trompette. Le WG-Lasso est

6. Applications awaken

utilisé avec un voisinage de 4 coefficients avant et après le coefficient central d'intérêt. On compare les résultats à l'algorithme de Block-Thresholding [YMB08]. La figure 6.1 donne les résultats de l'algorithme IST (cf. chapitre 3 algorithme 1 et les divers opérateurs). La figure 6.2 compare les résultats pour les seuillages Lasso, WG-Lasso, Empirical Wiener (EW) et Persistent Empirical Wiener (PEW) avec et sans itération pour un SNR d'entrée de 10 dB. On remarque que pour ce problème en particulier, il vaut mieux faire du simple seuillage, sans itération. L'avantage du seuillage simple dans un dictionnaire temps-fréquence, est sa généralisation directe au débruitage d'un bruit Gaussien coloré : il suffit de jouer sur une pondération du seuil en fonction de la densité spectrale comme le montre l'équation (6.1).

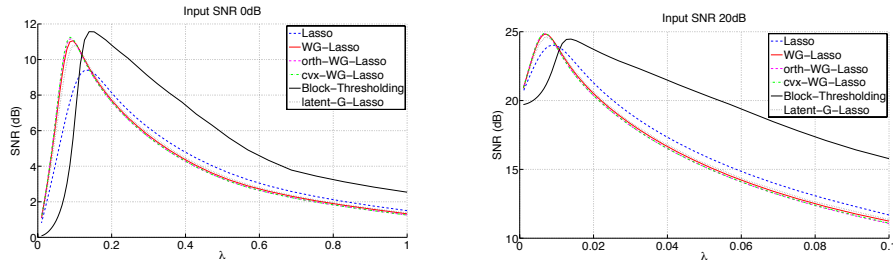


FIGURE 6.1. – SNR vs hyperparameter λ pour IST et six opérateurs de seuillage : Lasso, WG-Lasso, orth-WG-Lasso, cvx-WG-Lasso, latent-G-Lasso and Block-Thresholding sur un extrait de 6 sec d'un signal audio contenant de la batterie, de la basse, du piano, du saxophone et de la trompette.

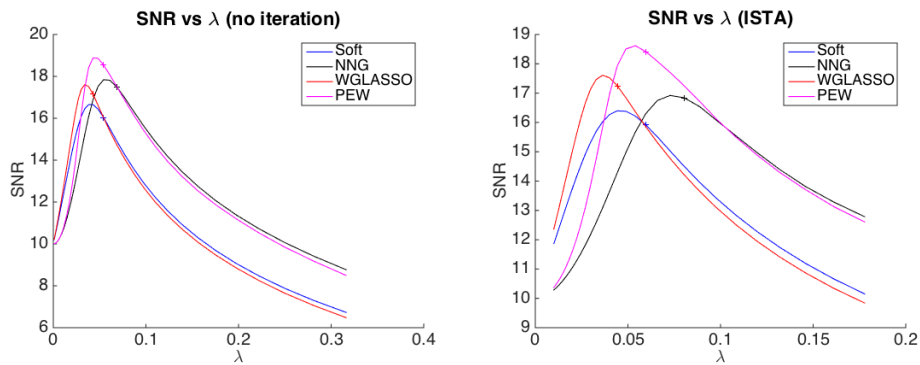


FIGURE 6.2. – Seuillage simple (gauche) dans un dictionnaire de Gabor vs Approche itérative (droite). SNR vs hyperparamètres λ pour quatre opérateurs (Lasso, WG-Lasso, EW and PEW).

6. Applications awaken

6.1.2. Dé-saturation

Le problème de dé-saturation audio a été présenté comme un cas particulier d'inpainting audio dans [Adl+12], et peut se formuler comme le problème suivant :

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\boldsymbol{\alpha}\|_0 \quad \text{s.t.} \quad \|\mathbf{y}^r - \mathbf{M}^r \boldsymbol{\Phi} \boldsymbol{\alpha}\|_2^2 \leq \epsilon. \quad (6.2)$$

Ici, \mathbf{y}^r sont les échantillons non saturés du signal observé. La matrice \mathbf{M}^r est une sous-matrice de l'identité qui permet de sélectionner les échantillons non clippés. Dans [Adl+12], le problème de minimisation (6.2) est résolu en appliquant un Matching Pursuit Orthogonal qui est ensuite projeté sur les contraintes.

On propose de re-formuler le problème sous forme non contrainte afin d'utiliser l'algorithme IST. De façon analogue à la matrice \mathbf{M}^r , on note \mathbf{M}^c la matrice qui sélectionne les coefficients clippés. On note ainsi $\boldsymbol{\theta}^{clip}$ le vecteur des échantillons clippés qui prennent tous la même valeur $\pm\theta^{clip}$ suivant le signe de l'échantillon original dans \mathbf{y} . On souhaite alors résoudre le problème d'optimisation suivant

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y}^r - \mathbf{M}^r \boldsymbol{\Phi} \boldsymbol{\alpha}\|_2^2 + \frac{1}{2} [\boldsymbol{\theta}^{clip} - \mathbf{M}^c \boldsymbol{\Phi} \boldsymbol{\alpha}]_+^2 + \lambda P(\boldsymbol{\alpha}), \quad (6.3)$$

avec

$$[\boldsymbol{\theta}^{clip} - \mathbf{x}]_+^2 = \sum_{k:\theta_k^{clip} > 0} (\theta_k^{clip} - x_k)_+^2 + \sum_{k:\theta_k^{clip} < 0} (-\theta_k^{clip} + x_k)_+^2.$$

Ainsi, la pénalité $\frac{1}{2} [\boldsymbol{\theta}^{clip} - \mathbf{M} \boldsymbol{\Phi} \boldsymbol{\alpha}]_+^2$ force les échantillons restaurés à prendre une valeur au-dessus du seuil de saturation. Cette fonction fait bien apparaître une partie Lipschitz différentiable et une pénalité (possiblement non convexe), nous permettant d'appliquer l'algorithme IST.

La figure 6.3 présente les résultats obtenus en dé-saturation. On représente l'amélioration obtenue sur SNR par rapport au SNR du signal saturé en fonction du niveau de saturation. Les résultats ont été moyennés sur dix signaux différents. Pour cette application, les opérateurs de type EW ou PEW permettent une très bonne estimation du signal original, tandis que les opérateurs de seuillage classique échouent totalement.

6.2. Localisation de sources M/EEG

Après une brève description du problème direct, pour introduire les notations habituellement utilisées en M/EEG, et de l'approche synthèse dans un dictionnaire temps-fréquence, je présente la fonctionnelle utilisée basée sur la pénalité Hi-Lasso [JAB11; Spr+11]. Les performances sont illustrées sur des simulations avant d'être validées sur des signaux réels lors d'une stimulation du système auditif.

6.2.1. Modèle direct et problème inverse

Le problème de localisation de sources M/EEG peut se formuler comme

$$\mathbf{M} = \mathbf{G}\mathbf{X} + \mathbf{E}, \quad (6.4)$$

6. Applications awaken

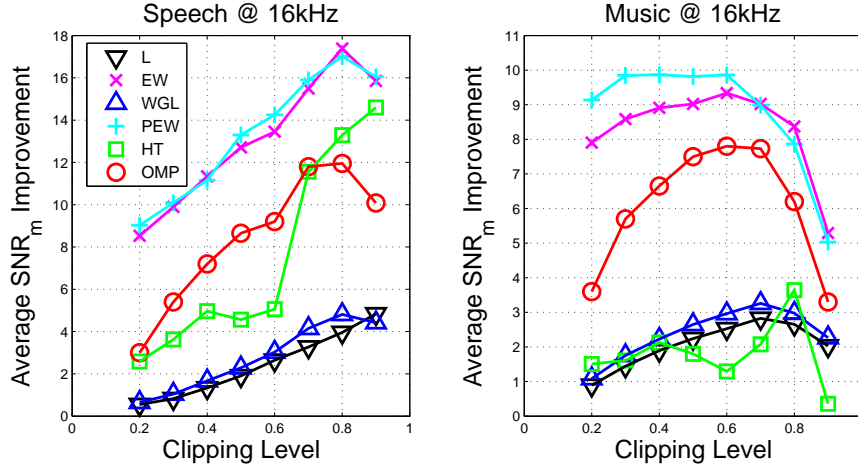


FIGURE 6.3. – SNR_{miss} moyen pour 10 signaux de parole (gauche) et de musique (droite) pour différents niveaux de saturation et divers opérateurs de seuillage. Les voisinages sont de 3 (pour la parole) et 7 coefficients (pour la musique) selon l'axe temporel.

avec

- $\mathbf{M} \in \mathbb{R}^{MT}$ les mesures des signaux M /EEG sur le scalp.
- $\mathbf{X} \in \mathbb{R}^{NT}$ les sources neuronales.
- $\mathbf{G} \in \mathbb{R}^{MN}$ la matrice de gain, qui modélise la propagation du champ électromagnétique du cortex cérébrale jusqu'au scalp. Cette matrice est connue.
- $\mathbf{E} \in \mathbb{R}^{MT}$ est un bruit additif supposé blanc et gaussien.

On considère ici l'approche synthèse dans un dictionnaire temps-fréquence

$$\mathbf{M} = \mathbf{G}\mathbf{X} + \mathbf{E} = \mathbf{G}\mathbf{Z}\mathbf{\Phi}^* + \mathbf{E} , \quad (6.5)$$

où $\mathbf{\Phi}^* \in \mathbb{C}^{K \times T}$ est un dictionnaire de K atomes de Gabor, $\mathbf{Z} \in \mathbb{C}^{P \times K}$ sont les coefficients de synthèse. Le problème (6.5) peut s'optimiser de la façon suivante :

$$\mathbf{Z}^* = \underset{\mathbf{Z}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{M} - \mathbf{G}\mathbf{Z}\mathbf{\Phi}^*\|_{\text{Fro}}^2 + \lambda \Omega(\mathbf{Z}) , \quad \lambda > 0 . \quad (6.6)$$

Lorsque $\Omega(\mathbf{Z}) = \|\mathbf{Z}\|_1$, le problème (6.6) est équivalent au problème du BPDN. Ce choix correspond au modèle MCE (Minimum Current Estimate) et souffre d'une inconsistance des sources actives au cours du temps. Les approches de type Group-Lasso avec une pénalité ℓ_{21} ont été proposées dans [OHG09] puis [GKH12]. Cependant, les séries temporelles obtenues sont bruitées, la parcimonie s'exprimant seulement à travers les sources sélectionnées. Afin de retrouver une structure parcimonieuse aussi sur les lignes de \mathbf{Z} , et non seulement sur les colonnes, nous avons utilisé une pénalité de type Hi-Lasso :

$$\lambda \Omega(\mathbf{Z}) = \lambda_{\text{space}} \|\mathbf{Z}\|_{21} + \lambda_{\text{time}} \|\mathbf{Z}\|_1 , \quad \lambda_{\text{space}} > 0, \lambda_{\text{time}} > 0 . \quad (6.7)$$

La figure 6.4 illustre les différents types de parcimonie obtenus selon le choix de Ω .

6. Applications awaken

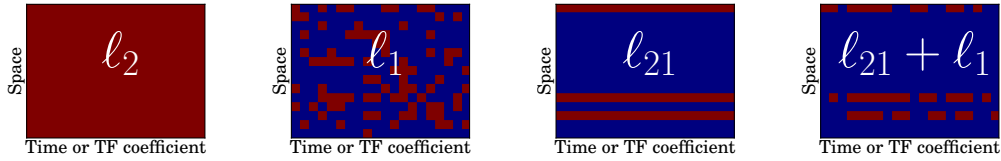


FIGURE 6.4. – Motifs de parcimonie favorisée par les différents priors. Le rouge indique les coefficients non nuls. Pénalité ℓ_2 : aucun coefficient à zéro. Pénalité ℓ_1 : zéros diffus et non structurés. Pénalité ℓ_{21} : structure de bloc en ligne. Pénalité $\ell_{21} + \ell_1$: structure par blocs avec une parcimonie intra-ligne.

6.2.2. Résultats

6.2.2.1. Simulations

On compare sur la figure 6.5 les différentes approches sur une configuration réaliste de type EEG avec un modèle de tête sphérique et 60 électrodes placées selon le système international 10-5.

6.2.2.2. Signaux réels

La figure 6.6 montre le modèle de tête et la reconstruction obtenue avec la norme mixte ℓ_{21} et l'approche temps-fréquence avec le Hi-Laso sur un enregistrement EEG. L'approche par synthèse temps-fréquence permet de séparer les composantes à haute fréquence.

La figure 6.7 montre quand à elle les résultats sur des signaux combinés MEG et EEG lors d'une stimulation du système auditif gauche. L'approche temps-fréquence proposée permet de trouver deux régions du cerveau active dans le cortex auditif.

6. Applications awaken

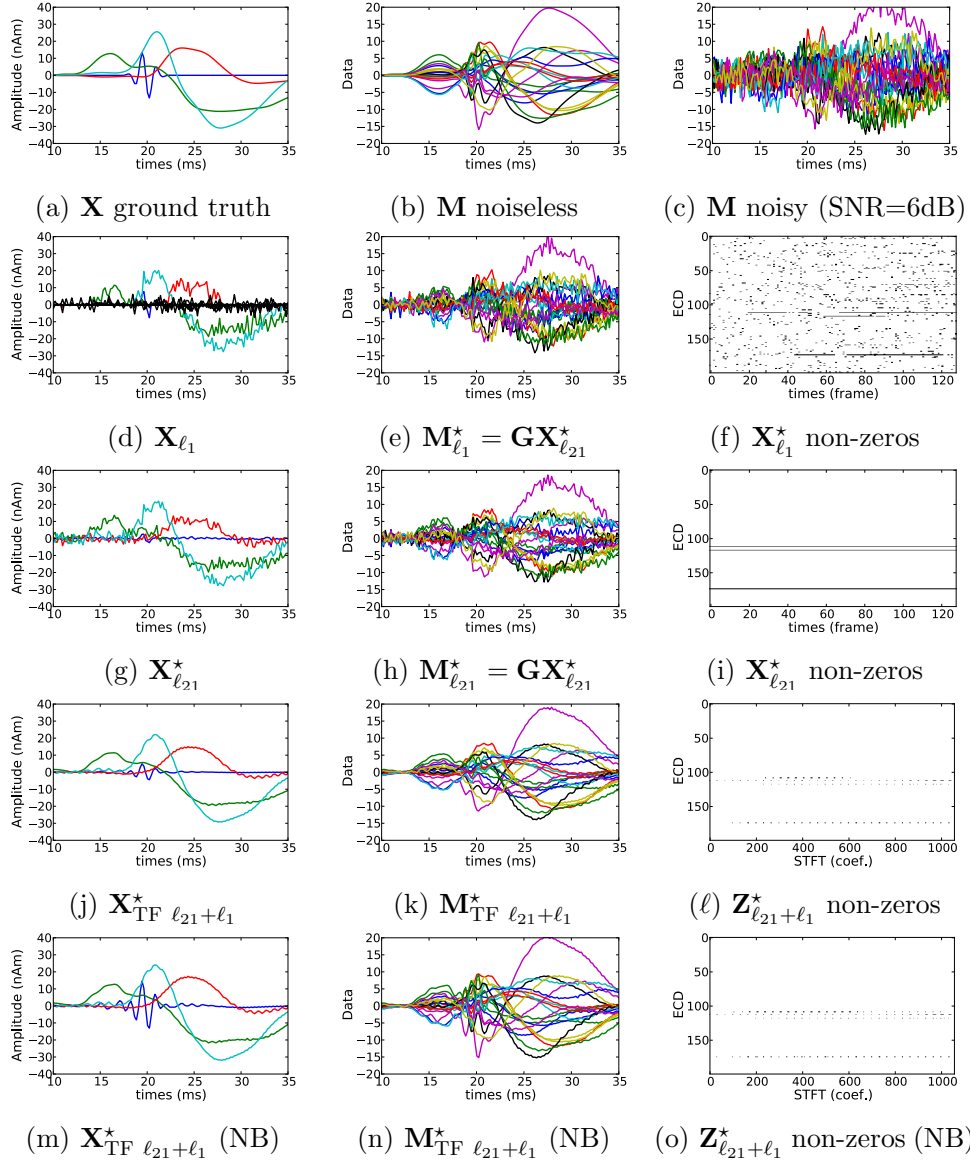


FIGURE 6.5. – Résultats de simulations avec un $SNR = 6$ dB. (a) Activation des sources. (b) Mesures non bruitées. (c) Mesures bruitées. (d-e-f) Estimation avec ℓ_1 . (g-h-i) Estimation avec ℓ_{21} [OHG09]. (j-k-l) Estimation avec un prior temps-fréquence composite. (f-i-l) Cartes d'activation obtenues pour les trois priors différents, comme expliqué sur la figure 6.4. Les résultats en (j) montrent comment le prior temps-fréquence composite améliore (d) et (g); (l) présente aussi un plus grand niveau de parcimonie par rapport à (f) et (i).

6. Applications awaken

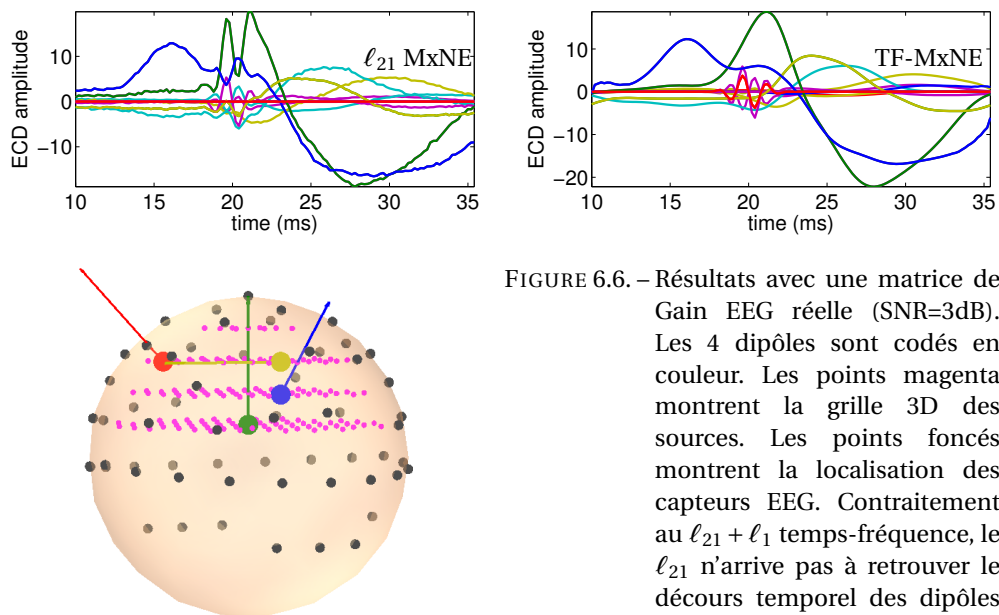
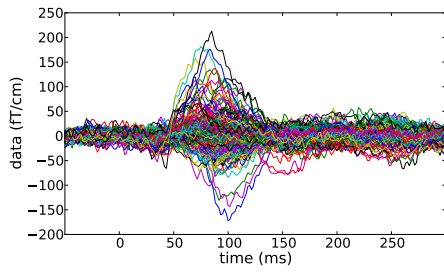
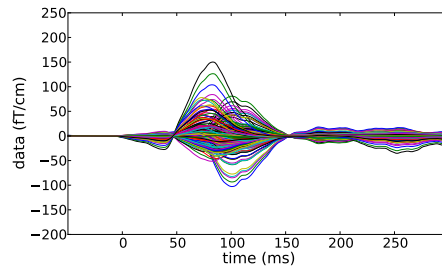


FIGURE 6.6. – Résultats avec une matrice de Gain EEG réelle (SNR=3dB). Les 4 dipôles sont codés en couleur. Les points magenta montrent la grille 3D des sources. Les points foncés montrent la localisation des capteurs EEG. Contrairement au $\ell_{21} + \ell_1$ temps-fréquence, le ℓ_{21} n'arrive pas à retrouver le décours temporel des dipôles verts profonds.

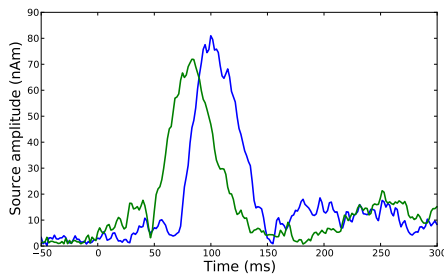
6. Applications awaken



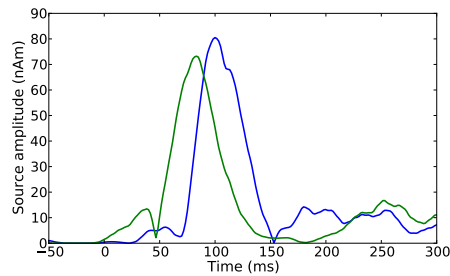
(a) MEG data (Gradiometers only)



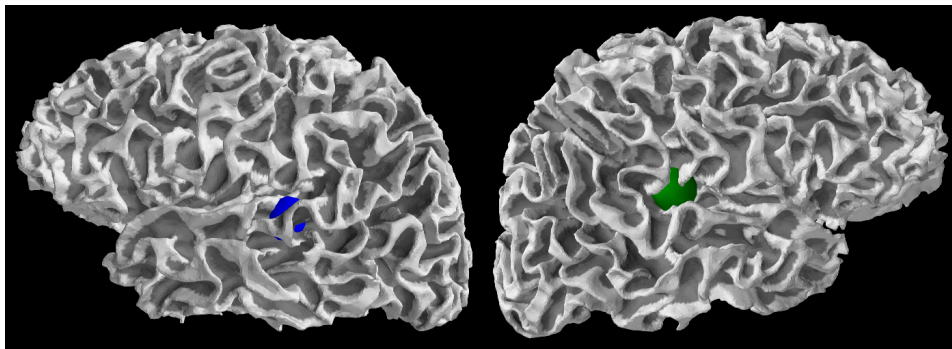
(b) $\mathbf{GX}_{\text{TF-MxNE}}^*$ (explained data)



(c) $\mathbf{X}_{\text{MxNE}}^*$



(d) $\mathbf{X}_{\text{TF-MxNE}}^*$



(e) Source locations in bilateral auditory cortices

FIGURE 6.7. – Résultats obtenus avec l'approche temps-fréquence et Hi-Lasso, et l'approche norme mixte simple pour une stimulation du système auditif de l'oreille gauche avec des données MEG/EEF non filtrés. $\mathbf{G} \in \mathbb{R}^{364 \times 22494}$. On retrouve deux régions actives dans le cerveau sur le cortex auditif. L'approche temps-fréquence trouve des décours temporels doux et des zéros durant la ligne de base.

7

■ To be continued...

Les travaux présentés dans ce manuscrit sont concentrés sur mes apports méthodologiques pour les méthodes parcimonieuses du point de vue de l'approche synthèse', en considérant d'une part les aspects probabilistes sous l'angle Bayésien et d'autre part l'optimisation d'une fonction de coût. Les algorithmes proposés reposent sur un algorithme général de seuillage itératif simple à mettre en oeuvre. J'ai voulu replacer ces travaux dans le contexte très large des approches parcimonieuses pour les problèmes inverses. Malgré la confusion parfois existante dans la littérature, je me suis attaché à différencier la résolution d'un problème inverse à l'aide de la parcimonie de l'application d'échantillonnage compressée. Cette dernière approche est, de mon point de vue, avant tout un problème d'échantillonnage pour lequel la parcimonie permet de s'affranchir du critère de Shannon.

Pour les années à venir, mon projet de recherche s'appuie sur l'approche que j'ai adoptée pour mener mes travaux jusqu'à présent, à savoir l'articulation autour des trois axes *modélisation, études mathématiques et mise en oeuvre sur des problèmes concrets*.

7.1. Optimisation, problèmes inverses et approches Bayésiennes

Le choix des approches variationnelles demande une étude théorique des algorithmes permettant de minimiser efficacement une fonctionnelle, qu'elle soit convexe ou non, différentiable ou non. L'étude des algorithmes de seuillage itératif à l'aide d'une fonction de seuillage quelconque offre un cadre simple et général pour traiter les problèmes inverses avec a priori de parcimonie. Plusieurs questions restent cependant en suspens : le choix des opérateurs est souvent dicté par l'intuition et/ou un compromis efficacité algorithmique et pratique. Cependant, si le choix du niveau de seuillage est souvent discuté (approches SURE, mini-max, etc.), la justification théorique du choix de l'opérateur de seuillage est rarement étudiée.

Afin de discuter du choix des paramètres au sens large (niveau de seuillage, opérateur, etc.), ou de l'analyse d'un modèle, on peut distinguer :

- les approches de type analyse harmonique, où l'estimation se place dans le cadre d'hypothèses sur les régularités des fonctions (type espace de Besov), qui permettent de justifier les approches de type mini-max ;
- les résultats d'identification d'un signal parcimonieux selon la structure du dictionnaire ;

7. To be continued...

- les approches de type Stein Unbiased Risk Estimator (SURE) et leur généralisation aux problèmes inverses (GSURE) ;
- les modèles Bayésiens, où le choix d'un a priori adéquat permet l'obtention d'une estimation totalement non supervisée, souvent au prix d'un certain coût de calcul.

Le lien entre les approches de type optimisation et Bayésienne a été discuté dans le chapitre d'introduction 1. Cependant, le cadre probabiliste utilisé dans les approches Bayésiennes apporte aussi un éclairage sur les algorithmes de seuillage, comme montré au chapitre 4 avec le modèle Bernoulli-Gaussien. L'extension de l'algorithme 5 de type IST présenté au chapitre 4 aux problèmes inverses est une piste privilégiée pour mes travaux futurs. Dans le même temps, il me semble intéressant d'étudier dans quelle mesure un modèle Bayésien permet d'analyser les performances d'une approche par optimisation, dans le même esprit que l'étude du seuillage dur dans le cadre Bernoulli-Gaussien. Enfin, l'existence d'une pénalité associée à un estimateur MMSE étant montrée dans [GM13], je souhaiterais voir dans quelle mesure une telle pénalité peut être effectivement construite, toujours dans le cadre d'un problème inverse général.

7.2. Apprentissage de dictionnaire par LRTFS

La seconde piste que je souhaite explorer rapidement est l'apprentissage de dictionnaire dans le contexte de synthèse temps-fréquence à faible rang développée avec Cédric Févotte et présentée au chapitre 4. En effet, la matrice \mathbf{W} peut-être vue comme une matrice de dictionnaire dans le cadre NMF. L'idée est donc d'apprendre une telle matrice sur une base de signaux (par exemple, une base de signaux de piano), afin de s'adapter au mieux aux motifs temps-fréquence associés. L'apprentissage de dictionnaire a largement été étudié pour les problèmes inverses, essentiellement en image (avec, par exemple, l'approche k-svd [AEB06] ou bien encore les approches en ligne [Mai+09]). Cependant, une telle approche pour les signaux audio a toujours butté sur la taille des signaux rencontrés (44100 échantillons pour une seconde de signal en haute fidélité). La matrice \mathbf{W} dans le modèle LRTFS peut-être de taille fixe, dépendant de la précision du dictionnaire envisagée. Si l'on reprend l'exemple du piano, on souhaiterait idéalement apprendre un \mathbf{W} contenant k motifs de bases (k motifs par touche de piano).

Clément Laroche a commencé l'exploration d'une telle approche pour les signaux de batterie pendant de sa thèse. Alors que ces signaux "large bande" semblent plus difficiles à représenter dans les dictionnaires de Gabor que des signaux "harmoniques", les résultats obtenus sur l'application de séparation de sources harmoniques/percussives à l'aide d'une variante de la NMF sont très encourageants.

7.3. Applications aux signaux audios

Les signaux audio sont au coeur de beaucoup d'applications, en particulier depuis l'émergence de la communauté MIR (Music Information Retrieval) ces dernières années. Dans ce contexte, les problèmes inverses les plus évidents sont la restauration des signaux audio et la séparation de sources.

7. To be continued...

La séparation de sources audio, dans le cas le plus général, est un problème convolutif et sous-déterminé. Cela implique de réussir à déconvoluer les signaux en plus de les séparer, sans connaître le système de mélange (cas aveugle). Ces travaux se placent dans la continuité de la thèse de Fangchen Feng, qui a permis d'explorer très largement différentes approches, essentiellement en bande étroite. L'extension aux modèles en bande large reste totalement ouverte.

Un autre aspect des applications aux signaux audio que je souhaite explorer est l'utilisation dans un cadre "légiste" (*forensic*). J'ai eu l'occasion de discuter avec le groupe audio de la police scientifique à Écully. Ils exploitent aujourd'hui les données avec les logiciels classiques de traitement audio. La discussion a permis de mettre en lumière des problèmes concrets, comme la mise en place de protocoles d'écoutes afin de minimiser les effets de réverbérations et de bruit. De même, les problèmes de débruitage rencontrés lors d'enregistrement en environnement non contrôlé sortent largement du cadre du bruit Gaussien (même coloré), voire du modèle de bruit stationnaire.

7.4. Applications aux signaux M/EEG et neuroimagerie

Mon intérêt pour cette deuxième application s'inscrit dans la poursuite de ma collaboration avec Alexandre Gramfort, spécialiste de l'analyse de données M/EEG, et permet une complémentarité idéale pour attaquer ces problèmes inverses. En effet, l'analyse exploratoire de réseaux fonctionnels implique un changement radical dans l'analyse des données M/EEG. Le travail se fait sur données brutes et non-moyennées, avec des techniques d'analyse spectrale, ce qui pose des problèmes méthodologiques de quantification et de contrôle statistique des interactions, mais aussi des problèmes pratiques liés au volume de données à traiter.

Au-delà des signaux M/EEG, je souhaiterais étendre mon champ d'application à la neuro-imagerie de manière générale, et développer la collaboration avec l'équipe PARIE-TAL de Neurospin débutée lors de ma demi-délégation INRIA.

7.5. Déconvolution aveugle des données PFS

Ma collaboration avec Frédéric Schmidt du laboratoire GEOPS a pour but de développer des outils pour la calibration et l'analyse des données PFS (Planetary Fourier Spectrometer) à bord de Mars Express. Les spectres enregistrés sont en effet corrompus par les vibrations mécaniques de la sonde. Plus précisément, les interférogrammes dépendent d'une modulation qui correspond aux déviations de la configuration réelle de l'instrument par rapport au système optique idéal. Les données PFS ont été peu exploitées en raison du problème des micro-vibrations. Les travaux les plus récents basés sur une simulation numérique de l'instrument montre que les effets sont réduits par un moyennage des spectres. Cependant, cette opération réduit considérablement la résolution spatio-temporelle des données et limite la compréhension des phénomènes dus au méthane.

Outre les approches par déconvolution aveugle existante, l'exploitation de toutes les données ouvre un certain nombre de pistes originales. Par exemple, il existe des mesures

7. *To be continued...*

de calibrage : l'instrument pointe alors vers l'espace. Si ces mesures peuvent être considérées comme stationnaire, elles pourraient être utilisées pour aider l'estimation du noyau de convolution. De plus, les mesures sont faites de manière quasi-continue autour de la planète. L'évolution douce de toutes ces mesures devrait permettre l'amélioration de l'estimation du noyau de convolution.

Bibliographie

- [DEM+01] Guy DEMOMENT, Jérôme IDIER, Jean-Francois GIOVANNELLI et Ali MOHAMMAD-DJAFARI. “Problèmes inverses en traitement du signal et de l’image”. In : *Techniques de l’ingénieur. Télécoms* TE5235 (2001), TE5235–1.
- [TAJ77] A. Nikolaevich TIKHONOV, V. Iakovlevich ARSENIN et F. JOHN. *Solutions of ill-posed problems*. T. 14. Winston Washington, DC, 1977.
- [MZ93] S. G. MALLAT et Z. ZHANG. “Matching pursuits with time-frequency dictionaries”. In : *IEEE Transactions on signal processing* 41.12 (1993), p. 3397–3415.
- [Mal08] S. MALLAT. *A wavelet tour of signal processing : the sparse way*. Academic press, 2008.
- [Bal+11] Peter BALAZS, Monika DÖRFLER, Florent JAILLET, Nicki HOLIGHAUS et G VELASCO. “Theory, implementation and applications of nonstationary Gabor frames”. In : *Journal of computational and applied mathematics* 236.6 (2011), p. 1481–1496.
- [Hol+13] Nicki HOLIGHAUS, Monika DÖRFLER, Gino Angelo VELASCO et Thomas GRILL. “A framework for invertible, real-time constant-Q transforms”. In : *IEEE Transactions on Audio, Speech, and Language Processing* 21.4 (2013), p. 775–785.
- [Can+06] E. CANDÈS, L. DEMANET, D. DONOHO et L. YING. “Fast discrete curvelet transforms”. In : *Multiscale Modeling & Simulation* 5.3 (2006), p. 861–899.
- [MP07] S. MALLAT et G. PEYRÉ. “A review of bandlet methods for geometrical image representation”. In : *Numerical Algorithms* 44.3 (2007), p. 205–234.
- [Dau92] I. DAUBECHIES. *Ten lectures on wavelets*. SIAM, 1992.
- [Bal+13] Peter BALAZS, Monika DOERFLER, Matthieu KOWALSKI et Bruno TORRESANI. “Adapted and adaptive linear time-frequency representations : a synthesis point of view”. In : *IEEE Signal Processing Magazine* 30.6 (2013), p. 20–31.
- [Rob06] Christian ROBERT. *Le choix bayésien : Principes et pratique*. Springer Science & Business Media, 2006.
- [GM13] R. GRIBONVAL et P. MACHART. “Reconciling” priors” &” priors” without prejudice?” In : *Advances in Neural Information Processing Systems*. 2013, p. 2193–2201.
- [Sta+13] J-L. STARCK, D. DONOHO, J. FADILI et A. RASSAT. “Sparsity and the Bayesian perspective”. In : *Astronomy & Astrophysics* 552 (2013), A133.
- [Nik07] M. NIKOLOVA. “Model distortions in Bayesian MAP reconstruction”. In : *Inverse Problems and Imaging* 1.2 (2007), p. 399.

Bibliographie

- [KVG10] Matthieu KOWALSKI, Emmanuel VINCENT et Rémi GRIBONVAL. “Beyond the narrowband approximation : Wideband convex methods for under-determined reverberant audio source separation”. In : *IEEE Transactions on Audio, Speech, and Language Processing* 18.7 (2010), p. 1818–1829.
- [FK14a] Fangchen FENG et Matthieu KOWALSKI. “Hybrid model and structured sparsity for under-determined convolutive audio source separation”. In : *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE. 2014, p. 6682–6686.
- [FK15a] Fangchen FENG et Matthieu KOWALSKI. “An unified approach for blind source separation using sparsity and decorrelation”. In : *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE. 2015, p. 1736–1740.
- [FK17] Fangchen FENG et Matthieu KOWALSKI. “Sparsity and low-rank amplitude based blind Source Separation”. In : *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. 2017.
- [Cho08] Seungjin CHOI. “Algorithms for orthogonal nonnegative matrix factorization”. In : *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence)*. *IEEE International Joint Conference on*. IEEE. 2008, p. 1828–1832.
- [Lar+15a] Clément LAROCHE, Matthieu KOWALSKI, Helene PAPADOPOULOS et Gaël RICHARD. “A structured nonnegative matrix factorization for source separation”. In : *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE. 2015, p. 2033–2037.
- [Lar+16] Clément LAROCHE, Hélène PAPADOPOULOS, Matthieu KOWALSKI et Gaël RICHARD. “Genre specific dictionaries for harmonic/percussive source separation”. In : *ISMIR The 17th International Society for Music Information Retrieval Conference*. 2016.
- [Lar+17] Clément LAROCHE, Hélène PAPADOPOULOS, Matthieu KOWALSKI et Gaël RICHARD. “Drum extraction in single channel audio signals using multi-layer non negative matrix factor deconvolution”. In : *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. 2017.
- [KMW16] Matthieu KOWALSKI, Adrien MEYNARD et Hau-Tieng WU. “Convex Optimization approach to signals with fast varying instantaneous frequency”. In : *Applied and Computational Harmonic Analysis* (2016).
- [Kow09] Matthieu KOWALSKI. “Sparse regression using mixed norms”. In : *Applied and Computational Harmonic Analysis* 27.3 (2009), p. 303–324.
- [KT09a] Matthieu KOWALSKI et Bruno TORRÉSANI. “Sparsity and persistence : mixed norms provide simple signal models with dependent coefficients”. In : *Signal, image and video processing* 3.3 (2009), p. 251–264.
- [GKH12] Alexandre GRAMFORT, Matthieu KOWALSKI et Matti HÄMÄLÄINEN. “Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods”. In : *Physics in medicine and biology* 57.7 (2012), p. 1937.

Bibliographie

- [DJ95] D. L. DONOHO et I. M. JOHNSTONE. “Adapting to unknown smoothness via wavelet shrinkage”. In : *Journal of the american statistical association* 90.432 (1995), p. 1200–1224.
- [Don95] D. L. DONOHO. “De-noising by soft-thresholding”. In : *IEEE transactions on information theory* 41.3 (1995), p. 613–627.
- [D+94] D. L. DONOHO, I. M. JOHNSTONE et al. “Ideal denoising in an orthonormal basis chosen from a library of bases”. In : *Comptes rendus de l’Académie des sciences. Série I, Mathématique* 319.12 (1994), p. 1317–1322.
- [CW92] R. R. COIFMAN et M. V. WICKERHAUSER. “Entropy-based algorithms for best basis selection”. In : *IEEE Transactions on information theory* 38.2 (1992), p. 713–718.
- [CDS98] S. S. CHEN, D. L. DONOHO et M. A. SAUNDERS. “Atomic Decomposition by Basis Pursuit”. In : *SIAM Journal on Scientific Computing* 20.1 (1998), p. 33–61.
- [CDS01] S. S. CHEN, D. L. DONOHO et M. A. SAUNDERS. “Atomic decomposition by basis pursuit”. In : *SIAM review* 43.1 (2001), p. 129–159.
- [Che95] S. S. CHEN. “Basis pursuit [Ph. D. thesis]”. In : *Department of Statistics*. Stanford University, 1995.
- [Tib96] R. TIBSHIRANI. “Regression Shrinkage and selection via the lasso”. In : *Journal of the Royal Statistical Society Serie B* 58.1 (1996), p. 267–288.
- [Nes05] Y. NESTEROV. “Smooth minimization of non-smooth functions”. In : *Mathematical programming* 103.1 (2005), p. 127–152.
- [FN03] M. A. T. FIGUEIREDO et R. D. NOWAK. “An EM algorithm for wavelet-based image restoration”. In : *IEEE Transactions on Image Processing* 12.8 (2003), p. 906–916.
- [CW05] P. L. COMBETTES et V. R. WAJS. “Signal recovery by proximal forward-backward splitting”. In : *Multiscale Modeling & Simulation* 4.4 (2005), p. 1168–1200.
- [DDD04] I. DAUBECHIES, M. DEFRISE et C. DE MOL. “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint”. In : *Communications on pure and applied mathematics* 57.11 (2004), p. 1413–1457.
- [Don01] D. L. DONOHO. “Sparse components of images and optimal atomic decompositions”. In : *Constructive Approximation* 17.3 (2001), p. 353–382.
- [Tro04] J. A. TROPP. “Greed is good : Algorithmic results for sparse approximation”. In : *IEEE Transactions on Information theory* 50.10 (2004), p. 2231–2242.
- [Tro06] J. A. TROPP. “Just relax : Convex programming methods for identifying sparse signals in noise”. In : *IEEE transactions on information theory* 52.3 (2006), p. 1030–1051.
- [Don06] D. L. DONOHO. “Compressed sensing”. In : *IEEE Transactions on information theory* 52.4 (2006), p. 1289–1306.
- [CW08] E. J. CANDÈS et M. B. WAKIN. “An introduction to compressive sampling”. In : *IEEE signal processing magazine* 25.2 (2008), p. 21–30.

Bibliographie

- [ZH05] H. ZOU et T. HASTIE. “Regularization and variable selection via the elastic net”. In : *Journal of the Royal Statistical Society : Series B (Statistical Methodology)* 67.2 (2005), p. 301–320.
- [AFS12] A. ARGYRIOU, R. FOYGEL et N. SREBRO. “Sparse Prediction with the k -Support Norm”. In : *Advances in Neural Information Processing Systems*. 2012, p. 1457–1465.
- [BR08] H. D BONDELL et B. J. REICH. “Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR”. In : *Biometrics* 64.1 (2008), p. 115–123.
- [FN16] M. FIGUEIREDO et R. NOWAK. “Ordered weighted l1 regularized regression with strongly correlated covariates : Theoretical aspects”. In : *Artificial Intelligence and Statistics*. 2016, p. 930–938.
- [YL06] M. YUAN et Y. LIN. “Model selection and estimation in regression with grouped variables”. In : *Journal of the Royal Statistical Society : Series B (Statistical Methodology)* 68.1 (2006), p. 49–67.
- [Cot+05] Shane F COTTER, Bhaskar D RAO, Kjersti ENGAN et Kenneth KREUTZ-DELGADO. “Sparse solutions to linear inverse problems with multiple measurement vectors”. In : *IEEE Transactions on Signal Processing* 53.7 (2005), p. 2477–2488.
- [ZJH10] Y. ZHOU, R. JIN et S. HOI. “Exclusive lasso for multi-task feature selection”. In : *International conference on artificial intelligence and statistics*. 2010, p. 988–995.
- [JOV09a] L. JACOB, G. OBOZINSKI et J.-P. VERT. “Group lasso with overlap and graph lasso”. In : *Proceedings of the 26th annual international conference on machine learning*. ACM. 2009, p. 433–440.
- [JAB11] R. JENATTON, J.-Y. AUDIBERT et F. BACH. “Structured variable selection with sparsity-inducing norms”. In : *Journal of Machine Learning Research* 12.Oct (2011), p. 2777–2824.
- [Spr+11] P. SPRECHMANN, I. RAMIREZ, G. SAPIRO et Y. C. EL DAR. “C-HiLasso : A collaborative hierarchical sparse modeling framework”. In : *IEEE Transactions on Signal Processing* 59.9 (2011), p. 4183–4198.
- [Bac10] Francis R BACH. “Structured sparsity-inducing norms through submodular functions”. In : *Advances in Neural Information Processing Systems*. 2010, p. 118–126.
- [Kyr+15] A. KYRILLIDIS, L. BALDASSARRE, M. EL HALABI, Q. TRAN-DINH et V. CEVHER. “Structured sparsity : Discrete and convex approaches”. In : *Compressed Sensing and its Applications*. Springer, 2015, p. 341–387.
- [Bac+12] Francis BACH, Rodolphe JENATTON, Julien MAIRAL, Guillaume OBOZINSKI et al. “Optimization with sparsity-inducing penalties”. In : *Foundations and Trends® in Machine Learning* 4.1 (2012), p. 1–106.

Bibliographie

- [DT02] L. DAUDET et B. TORRÉSANI. “Hybrid Representations for Audiophonic Signal Encoding”. In : *Signal Processing* 82.11 (2002). Special issue on Image and Video Coding Beyond Standards, p. 1595–1617.
- [Ela+05] M. ELAD, J.-L. STARCK, P. QUERRE et D. L. DONOHO. “Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)”. In : *Applied and Computational Harmonic Analysis* 19.3 (2005), p. 340–358.
- [Kow16] Matthieu KOWALSKI. “Proximal algorithm meets a conjugate descent”. In : *Pacific journal of optimization* 12.3 (2016), p. 669–695.
- [Kow+11] Matthieu KOWALSKI, Pierre WEISS, Alexandre GRAMFORT et Sandrine ANTHOINE. “Accelerating ISTA with an active set strategy”. In : *OPT 2011 : 4th International Workshop on Optimization for Machine Learning*, 2011, p. 7.
- [Tse10] P. TSENG. “Approximation accuracy, gradient methods, and error bound for structured convex optimization”. In : *Mathematical Programming* 125.2 (2010), p. 263–295.
- [BT09] A. BECK et M. TEOULLE. “A fast iterative shrinkage-thresholding algorithm for linear inverse problems”. In : *SIAM journal on imaging sciences* 2.1 (2009), p. 183–202.
- [CD15] A. CHAMBOLLE et C. DOSSAL. “On the convergence of the iterates of “FISTA””. In : *Journal of Optimization Theory and Applications* 166.3 (2015), p. 25.
- [Ant+07] Anestis ANTONIADIS et al. “Wavelet methods in statistics : Some recent developments and their applications”. In : *Statistics Surveys* 1 (2007), p. 16–55.
- [Roc70] R.T. ROCKAFELLAR. *Convex Analysis*. Sous la dir. de Princeton University PRESS. 1970.
- [TY09] P. TSENG et S. YUN. “A coordinate gradient descent method for nonsmooth separable minimization”. In : *Mathematical Programming* 117.1 (2009), p. 387–423.
- [Mat+17] S. MATET, L. ROSASCO, S. VILLA et B. L. VU. “Don’t relax : early stopping for convex regularization”. In : *arXiv preprint arXiv :1707.05422* (2017).
- [Pre98] L. PRECHELT. “Early stopping-but when?” In : *Neural Networks : Tricks of the trade* (1998), p. 553–553.
- [PRK93] Y.C. PATI, R. REZAIIFAR et P.S. KRISHNAPRASAD. “Orthogonal matching pursuit : Recursive function approximation with applications to wavelet decomposition”. In : *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE. 1993, p. 40–44.
- [RF08] V. ROTH et B. FISCHER. “The group-lasso for generalized linear models : uniqueness of solutions and efficient algorithms”. In : *Proceedings of the 25th international conference on Machine learning*. ACM. 2008, p. 848–855.
- [Bac+11] F. BACH, R. JENATTON, J. MAIRAL, G. OBOZINSKI et al. “Convex optimization with sparsity-inducing norms”. In : *Optimization for Machine Learning* 5 (2011), p. 19–53.

Bibliographie

- [RVV14] L ROSASCO, S VILLA et B.C. VU. “Convergence of stochastic proximal gradient algorithm”. In : *arXiv preprint arXiv :1403.5074* (2014).
- [Sou+11] C. SOUSSEN, J. IDIER, D. BRIE et J. DUAN. “From Bernoulli–Gaussian deconvolution to sparse signal restoration”. In : *IEEE Transactions on Signal Processing* 59.10 (2011), p. 4572–4584.
- [DHD12] A. DRÉMEAU, C. HERZET et L. DAUDET. “Boltzmann machine and mean-field approximation for structured sparse decompositions”. In : *IEEE Transactions on Signal Processing* 60.7 (2012), p. 3425–3438.
- [WR04] D. P. WIPF et B. D. RAO. “Sparse Bayesian Learning for Basis Selection”. In : *IEEE Transactions on Signal Processing* 52.8 (août 2004), p. 2153–2164.
- [KR11a] Matthieu KOWALSKI et Thomas RODET. “An unsupervised algorithm for hybrid/morphological signal decomposition”. In : *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE. 2011, p. 4112–4115.
- [KF17] Matthieu KOWALSKI et Cédric FÉVOTTE. *Estimation with Low-Rank Time-Frequency Synthesis Models*. Submitted to IEEE Transactions on Signal Processing. 2017.
- [FK15b] Cédric FÉVOTTE et Matthieu KOWALSKI. “Hybrid sparse and low-rank time-frequency signal decomposition”. In : *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE. 2015, p. 464–468.
- [FK14b] Cédric FÉVOTTE et Matthieu KOWALSKI. “Low-rank time-frequency synthesis”. In : *Advances in Neural Information Processing Systems*. 2014, p. 3563–3571.
- [TYE11] Javier S TUREK, Irad YAVNEH et Michael ELAD. “On MMSE and MAP denoising under sparse representation modeling over a unitary dictionary”. In : *IEEE Transactions on Signal Processing* 59.8 (2011), p. 3526–3535.
- [DPF13] Charles-Alban DELEDALLE, Gabriel PEYRÉ et Jalal FADILI. “Stein consistent risk estimator (SCORE) for hard thresholding”. In : *arXiv preprint arXiv :1301.5874* (2013).
- [CG92] Gilles CELEUX et Gérard GOVAERT. “A classification EM algorithm for clustering and two stochastic versions”. In : *Computational statistics & Data analysis* 14.3 (1992), p. 315–332.
- [ZBJ09] Hadi ZAYYANI, Massoud BABAIE-ZADEH et Christian JUTTEN. “An iterative bayesian algorithm for sparse component analysis in presence of noise”. In : *IEEE Transactions on Signal Processing* 57.11 (2009), p. 4378–4390.
- [Tip01] M. E. TIPPING. “Sparse Bayesian Learning and the Relevance Vector Machine”. In : *Journal of Machine Learning Research* 1 (2001), p. 211–244.
- [Dau+10] I. DAUBECHIES, R. DEVORE, M. FORNASIER et C. S. GÜNTÜRK. “Iteratively reweighted least squares minimization for sparse recovery”. In : *Communications on Pure and Applied Mathematics* 63.1 (2010), p. 1–38.

Bibliographie

- [FBD09a] C. FÉVOTTE, N. BERTIN et J.-L. DURRIEU. “Nonnegative matrix factorization with the Itakura-Saito divergence : With application to music analysis”. In : *Neural computation* 21.3 (2009), p. 793–830.
- [CI11] Févotte; C. et J. IDIER. “Algorithms for nonnegative matrix factorization with the beta-divergence”. In : *Neural Computation* 23.9 (sept. 2011), p. 2421–2456.
- [FBD09b] C. FÉVOTTE, N. BERTIN et J.-L. DURRIEU. “Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis”. In : *Neural Computation* 21.3 (mar. 2009), p. 793–830.
- [Kow14] Matthieu KOWALSKI. “Thresholding rules and iterative shrinkage/thresholding algorithm : A convergence study”. In : *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE. 2014, p. 4151–4155.
- [KSD13] Matthieu KOWALSKI, Kai SIEDENBURG et Monika DÖRFLER. “Social sparsity! neighborhood systems enrich structured shrinkage operators”. In : *IEEE transactions on signal processing* 61.10 (2013), p. 2498–2511.
- [KT09b] Matthieu KOWALSKI et Bruno TORRÉSANI. “Sparsity and persistence in time-frequency sound representations”. In : *Wavelets XIII*. T. 7446. SPIE. 2009, 74460F.
- [WNF09] Stephen J WRIGHT, Robert D NOWAK et Mário AT FIGUEIREDO. “Sparse reconstruction by separable approximation”. In : *IEEE Transactions on Signal Processing* 57.7 (2009), p. 2479–2493.
- [CH00] Andrea COLESANTI et Daniel HUG. “Hessian measures of semi-convex functions and applications to support measures of convex bodies”. In : *manuscripta mathematica* 101.2 (2000), p. 209–238.
- [CS01] T. CAI et B. W. SILVERMAN. “Incorporating information on neighbouring coefficients into wavelet estimation”. In : *shankhya : the indian journal of statistic. Special issue on wavelets. Series B*. 63 (2001), p. 127–148.
- [Com04] P.L. COMBETTES. “Solving monotone inclusions via compositions of nonexpansive averaged operators”. In : *Optimization* 53.5-6 (déc. 2004), p. 475–504.
- [Fen17] Fangchen FENG. “Séparation aveugle de source : de l’instantané au convolutif”. Thèse de doct. Université Paris-Saclay, 2017.
- [SKD14] Kai SIEDENBURG, Matthieu KOWALSKI et Monika DORFLER. “Audio declipping with social sparsity”. In : *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE. 2014, p. 1577–1581.
- [Gra+13] Alexandre GRAMFORT, Daniel STROHMEIER, Jens HAUEISEN, Matti S HÄMÄLÄINEN et Matthieu KOWALSKI. “Time-frequency mixed-norm estimates : Sparse M/EEG imaging with non-stationary source activations”. In : *NeuroImage* 70 (2013), p. 410–422.
- [Gra+11a] Alexandre GRAMFORT, Daniel STROHMEIER, Jens HAUEISEN, Matti HAMALAINEN et Matthieu KOWALSKI. “Functional brain imaging with M/EEG using structured sparsity in time-frequency dictionaries”. In : *Information Processing in Medical Imaging*. Springer. 2011, p. 600–611.

Bibliographie

- [Str+11] Daniel STROHMEIER, Alexandre GRAMFORT, Jens HAUEISEN, Matti HÄMÄLÄINEN et Matthieu KOWALSKI. “MEG/EEG source reconstruction based on Gabor thresholding in the source space”. In : *Noninvasive Functional Source Imaging of the Brain and Heart & 2011 8th International Conference on Bioelectromagnetism (NFSI & ICBEM), 2011 8th International Symposium on*. IEEE. 2011, p. 103–108.
- [GK09] Alexandre GRAMFORT et Matthieu KOWALSKI. “Improving m/eeg source localization with an inter-condition sparse prior”. In : *Biomedical Imaging : From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*. IEEE. 2009, p. 141–144.
- [KG15a] Matthieu KOWALSKI et Alexandre GRAMFORT. “Inverse problems with time-frequency dictionaries and non-white Gaussian noise”. In : *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE. 2015, p. 1741–1745.
- [Wel67] Peter WELCH. “The use of fast Fourier transform for the estimation of power spectra : a method based on time averaging over short, modified periodograms”. In : *IEEE Transactions on audio and electroacoustics* 15.2 (1967), p. 70–73.
- [GRC02] Simon GODSILL, Peter RAYNER et Olivier CAPPÉ. “Digital audio restoration”. In : *Applications of digital signal processing to audio and acoustics*. Springer, 2002, p. 133–194.
- [YMB08] Guoshen YU, Stéphane MALLAT et Emmanuel BACRY. “Audio denoising by time-frequency block thresholding”. In : *IEEE Transactions on Signal processing* 56.5 (2008), p. 1830–1839.
- [Adl+12] Amir ADLER, Valentin EMIYA, Maria G JAFARI, Michael ELAD, Rémi GRIBONVAL et Mark D PLUMBLEY. “Audio inpainting”. In : *IEEE Transactions on Audio, Speech, and Language Processing* 20.3 (2012), p. 922–932.
- [OHG09] W. OU, M. HÄMÄLÄINEN et P. GOLLAND. “A Distributed Spatio-Temporal EEG/MEG Inverse Solver”. In : *NeuroImage* 44.3 (fév. 2009), p. 932–946.
- [AEB06] Michal AHARON, Michael ELAD et Alfred BRUCKSTEIN. “*rmk*-SVD : An algorithm for designing overcomplete dictionaries for sparse representation”. In : *IEEE Transactions on signal processing* 54.11 (2006), p. 4311–4322.
- [Mai+09] Julien MAIRAL, Francis BACH, Jean PONCE et Guillermo SAPIRO. “Online dictionary learning for sparse coding”. In : *Proceedings of the 26th annual international conference on machine learning*. ACM. 2009, p. 689–696.
- [PK13a] Hélène PAPADOPOULOS et Matthieu KOWALSKI. “Sparse and structured decomposition of audio signals on hybrid dictionaries using musical priors”. In : *The Journal of the Acoustical Society of America* 134.1 (2013), p. 666–685.
- [PK11] Hélène PAPADOPOULOS et Matthieu KOWALSKI. “Sparse signal decomposition on hybrid dictionaries using musical priors”. In : *ISMIR 2011*. 2011, p–687.
- [FK15c] Fangchen FENG et Matthieu KOWALSKI. “Vers une approche unifiée pour la séparation aveugle de sources en sur et sous-déterminé, basée sur la parcimonie et la décorrélation”. In : *Gretsi 2015*. 2015.

Bibliographie

- [Lar+15b] Clément LAROCHE, Matthieu KOWALSKI, Helene PAPADOPOULOS et Gaël RICHARD. “Méthode Structurée de décomposition en matrices non-négatives appliquée à la séparation de sources audio”. In : *Gretsi 2015*. 2015.
- [Mer+17] Alina-Georgiana MERESCU, Matthieu KOWALSKI, Frédéric SCHMIDT et François LANDAIS. “Estimation du Temps de Résidence Hydrologique : Déconvolution 1D”. In : *Gretsi 2017*. 2017.
- [GPK15] Cyril GAUDEFROY, Hélène PAPADOPOULOS et Matthieu KOWALSKI. “A multi-dimensional meter-adaptive method for automatic segmentation of music”. In : *Content-Based Multimedia Indexing (CBMI), 2015 13th International Workshop on*. IEEE. 2015, p. 1–6.
- [Sch+14] Frédéric SCHMIDT, Irina SHATALINA, Matthieu KOWALSKI, Nicolas GAC, Bortolino SAGGIN et M GIURANNA. “Toward a numerical deshaker for PFS”. In : *Planetary and Space Science* 91 (2014), p. 45–51.
- [KT08] Matthieu KOWALSKI et Bruno TORRÉSANI. “Random models for sparse signals expansion on unions of bases with application to audio signals”. In : *IEEE Transactions on Signal Processing* 56.8 (2008), p. 3468–3481.
- [KG10] Matthieu KOWALSKI et Alexandre GRAMFORT. “A priori par normes mixtes pour les problèmes inverses : Application à la localisation de sources en M/EEG”. In : *Traitement du Signal* 27.1 (2010), p. 51–76.
- [Sha+13] Irina SHATALINA, Frederic SCHMIDT, Bortolino SAGGIN, Nicolas GAC, Matthieu KOWALSKI et Marco GIURANNA. “Analytical model and spectral correction of vibration effects on Fourier transform spectrometer”. In : *SPIE Remote sensing 2013*. T. 8890. 2013, 88900S–88900S.
- [KSR09] Matthieu KOWALSKI, Marie SZAFRANSKI et Liva RALAIVOLA. “Multiple indefinite kernel learning with mixed norm regularization”. In : *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM. 2009, p. 545–552.
- [KT09c] Matthieu KOWALSKI et Bruno TORRÉSANI. “Structured sparsity : from mixed norms to structured shrinkage”. In : *SPARS’09-Signal Processing with Adaptive Sparse Structured Representations*. 2009.
- [KVG08] Matthieu KOWALSKI, Emmanuel VINCENT et Rémi GRIBONVAL. “Under-determined source separation via mixed-norm regularized minimization”. In : *Signal Processing Conference, 2008 16th European*. IEEE. 2008, p. 1–5.
- [KT06] Matthieu KOWALSKI et Bruno TORRÉSANI. “A family of random waveform models for audio coding”. In : *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. T. 3. IEEE. 2006, p. III–III.
- [KT05] Matthieu KOWALSKI et Bruno TORRÉSANI. “A study of Bernoulli and structured random waveform models for audio signals”. In : *SPARS 05*. 2005, TS2–1.
- [KG15b] Matthieu KOWALSKI et Alexandre GRAMFORT. “Problèmes inverses avec dictionnaires temps-fréquences en présence de bruit Gaussien non blanc”. In : *Gretsi 2015*. 2015.

Bibliographie

- [SDK13] Kai SIEDENBURG, Monika DÖRFLER et Matthieu KOWALSKI. "Parcimonie Sociale et application à l'audio inpainting". In : *GRETSI 2013*. 2013.
- [PK13b] Hélène PAPADOPOULOS et Matthieu KOWALSKI. "Décomposition parcimonieuse structurée des signaux audio de musique guidée par l'information experte". In : *GRETSI 2013*. 2013, 4–p.
- [KR11b] Matthieu KOWALSKI et Thomas RODET. "Un algorithme de seuillage itératif non-supervisé pour la décomposition parcimonieuse de signaux dans des unions de dictionnaires". In : *GRETSI 2011*. 2011.
- [Gra+11b] Alexandre GRAMFORT, Daniel STROHMEIER, Jens HAUEISEN, Matti HAMÄLÄINEN et Matthieu KOWALSKI. "Apport des dictionnaires temps-fréquence, de la parcimonie et des structures pour l'imagerie cérébrale fonctionnelle M/EEG". In : *GRETSI 2011*. 2011.
- [KG09] Matthieu KOWALSKI et Alexandre GRAMFORT. "A priori par normes mixtes pour les problèmes inverses Application à la localisation de sources en M/EEG". In : *XXIIe colloque GRETSI (traitement du signal et des images), Dijon (FRA), 8-11 septembre 2009*. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images. 2009.
- [KT09d] Matthieu KOWALSKI et Bruno TORRÉSANI. "Modèles aléatoires hiérarchiques pour la modélisation de signaux audio". In : *XXIIe colloque GRETSI (traitement du signal et des images), Dijon (FRA), 8-11 septembre 2009*. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images. 2009.
- [KT09e] Matthieu KOWALSKI et Bruno TORRÉSANI. "Décompositions parcimonieuses et persistantes de signaux multicanaux. Applications aux signaux MEEG." In : *CAP 2008*. 2009, pages–105.
- [KT07] Matthieu KOWALSKI et Bruno TORRESANI. "Identification de signaux parcimonieux dans un dictionnaire hybride". In : *GRETSI 2007*. 2007, p. 1249.
- [FHT10] J. FRIEDMAN, T. HASTIE et R. TIBSHIRANI. "A note on the group lasso and a sparse group lasso". In : *arXiv preprint arXiv:1001.0736* (2010).
- [Wei08] P. WEISS. "Algorithmes rapides d'optimisation convexe. Applications à la reconstruction d'images et à la détection de changements." Thèse de doct. Université Nice Sophia Antipolis, 2008.
- [EMR07] M. ELAD, P. MILANFAR et R. RUBINSTEIN. "Analysis Versus Synthesis in Signal Priors". In : *Inverse Problems* 23.3 (juin 2007), p. 947–968.
- [JOV09b] L. JACOB, G. OBOZINSKI et J.-P. VERT. "Group Lasso with Overlap and Graph Lasso". In : *ICML*. 2009.
- [OJV12] G. OBOZINSKI, L. JACOB et J.-P. VERT. *Group Lasso with Overlaps : the Latent Group Lasso approach*. Rapp. tech. 2012.

Deuxième partie .
Notice individuelle

8

■ Résumé

Matthieu Kowalski

Maître de conférences U-PSUD

Né le 11 mai 1981
Nationalité française
PACSÉ, 1 enfant

Laboratoire des Signaux et Systèmes (L2S)
UMR 8506 SUPELEC-CNRS-Univ Paris-Sud
3, rue Joliot-Curie
91192 Gif-sur-Yvette Cedex, France
Tél : +33 (0)1 69 85 17 47
Fax : +33 (0)1 69 85 17 65
Email : matthieu.kowalski@u-psud.fr
www.l2s.supelec.fr/perso/kowalski

8.1. Synthèse des activités

Recherche	Problèmes Inverses
Thèmes principaux	Décompositions parcimonieuses, Optimisation
Production	12 articles de revues internationales 1 article de revue nationale invité 25 articles de conférences internationales 12 articles de conférences nationales
Indices bibliométriques	h-index : 14, i10-index : 17 (source : google scholar)
Enseignement	Traitement du signal et Informatique
Niveau master	Représentations parcimonieuses et applications, Multimédia, Traitement du signal
Niveau licence	Traitement du signal, Algorithmique et Langage C, Probabilités et Statistiques

8. Résumé

Établissements	Université Paris-Sud, Université de Provence, École Centrale Marseille
Resp. scientifiques	Principales responsabilités liées aux activités scientifiques
Laboratoire	membre votant du conseil de laboratoire
Projet IMPULSE	Responsable du Projet <i>Inverse problems in Music Source sePa-ration Using expert knowLEdge</i> , financé par le CNRS (projet PICS)
Projet SESAM	Responsable du Projet <i>Savoir Expert pour les Signaux Audio de Musique</i> , financé par <i>Digiteo</i>
Projet MAORI	Responsable du projet <i>MATHématiques de l'Optimisation pouR l'Imagerie (MAORI)</i> , financé par le <i>Programme Gaspard Monge pour l'Optimisation</i> .
Comités de pilotage	de SPARS (membre élu)
Comité organisation	1er et 2nd workshop sur l'optimisation pour le traitement des images et du signal (IHP 2015; Polytechnique 2013); conférence SAMPTA 2009 (Marseille – comité local)

8.2. Titres universitaires

2009	Qualification aux fonctions de maître de conférences Sections 26 et 61 du Conseil National des Universités
2008	Doctorat en mathématiques appliquées Université de Provence – mention très honorable
	Titre Approximation des signaux : approches variationnelles et modèles aléatoires
	<i>Président :</i> Stéphane Canu (PdU, INSA Rouen)
	<i>Rapporteurs :</i> Emmanuel Bacry (CR CNRS, Polytechnique)
	Pierre Vandergheynst (Prof., EPFL)
Jury	<i>Examineurs :</i> Rémi Gribonval (CR INRIA, IRISA)
	Liva Ralaivola (MdC, Université de Provence)
	<i>Directeur de thèse :</i> Bruno Torrèsani (PdU, Université de Provence)
Mots-clés	Approximation parcimonieuse – Seuillage structuré – Normes mixtes – modèles hybrides aléatoires
2005	Diplôme d'ingénieur en génie informatique Université de Technologie de Compiègne Master MVA (Mathématiques Vision Apprentissage) (<i>Mention très bien</i>) ENS Cachan

8. Résumé

8.3. Parcours

2014 – 2016	Délégation INRIA (mi-temps) dans l'équipe Parietal.
Depuis 2009	Maître de conférences (61 ^e section) Université Paris-Sud– L2S (UMR 8506)
2008–2009	Attaché temporaire d'enseignement et de recherche (26 ^e section) Université de provence – LATP (UMR CNRS 7353) (temps-plein)
2005–2008	Doctorant à L'Université de Provence (maintenant Aix-Marseille Université) – Laboratoire d'Analyse Topologie Probabilité (LATP – UMR CNRS 7353)

9

■ Activités d'enseignement

9.1. Description des interventions principales

9.1.1. Traitement du Signal

Cours spécialisé en *Représentations parcimonieuses et applications* du **M2R ATSI** (Automatique, Traitement du Signal et des Images). Cours de traitement du signal pour le **M2 Acoustique**. Responsable de la partie "Signal" du cours *Signaux et Systèmes Linéaires* en **L3-IST** (Information, Systèmes et Technologie). C'est un cours de traitement du signal déterministe classique, qui introduit les représentations de Fourier, le filtrage et le théorème d'échantillonnage.

9.1.2. Algorithmique/C

Ancien responsable du cours d'algorithmique/C en **L3-IST** et en tronc commun de la **3ème année** de la formation **Polytech' Paris-Sud**. Ce cours est un rappel/introduction de l'algorithmique et du langage C à un public très hétérogène (étudiants venant de CPGE, de la formation préparatoire Polytech, divers DUT etc.).

9.1.3. Probabilité/Statistiques

Responsable du cours de Probabilités en **L3-IST** à l'université Paris-Sud. Chargé de TD en **1ère année** d'école d'ingénieur Centrale Marseille (**ECM**) et en **L2-MASS** (Mathématiques Appliquées aux Sciences Sociales). En L2-MASS, j'ai aussi assuré les cours *d'atelier Statistique*. En **L3-MASS** j'ai assuré le cours de *statistique avec le logiciel R*.

9.2. Tableau récapitulatif

Le détail des principales activités est présenté dans le tableau et les descriptifs ci-après, auxquels s'ajoute les charges pédagogiques de jury ou d'encadrement obligatoires. Tous les documents produits (polys de cours, planche de TD, TP etc.) sont disponibles sur ma page web.

9. Activités d'enseignement

	Enseignement	Niveau	Période	Cours/an	TD/an	TP/an	H éq. TD/an
Univ Paris-Sud	Ondelettes et parcimonie	M2	2009/2010 →	12			18
	Traitement du signal	M2	2017/2018 →	15		15	37.5
	Multimédia	M2	2009/2010 → 2011/2012	9		6	19.5
	Traitement du Signal	M2	2009/2010 → 2011/2012	9		6	19.5
	Signaux et Systèmes Linéaire	L3	2010/2011 →	16			24
	Probabilités	L3	2012/2013 →	16	16	8	48
	Algorithmique et Langage C	L3	2009/2010 →	12	10	18	46
	Projets C	L3	2009/2010 →			16	16
	Informatique	L3	2013/2014 →	14		36	57
	Travaux d'étude et de recherche	L3	2012/2013 →		10		10
AMU	Traitement du signal	M1	2008/2009		24		24
	Processus Stochastiques	M1	2008/2009		20		20
	Statistiques sous R	L3	2008/2009	12		36	54
	Statistique	L2	2008 /2009			24	24
	Atelier statistique	L2	2008/2009	6	18		27
	Statistiques descriptives	L2	2008/2009	10	15		30
	Probabilités/Statistiques	L2	2008/2009		36		36
ECM	Probabilités	L3	2005/2006 → 2007/2008		24	36	60

10. ■ Activités scientifiques

10.1. Comités

10.1.1. Comités de pilotage et d'organisation

Depuis 2013	Membre Élu du comité de pilotage de SPARS <i>Signal Processing with Adaptive Sparse Structured Representations</i> . Membres actuels : Bubacarr Bah, Volkan Cevher, Maryam Fazel, Cédric Févotte, Mario Figueiredo (Chair), Laurent Jacques, Nick Kingsbury, Matthieu Kowalski, Gitta Kutyniok, Gabriel Peyré, Mark Plumbley, Yves Wiaux.
2015	Responsable de l'organisation du second <i>Workshop on Optimization for Image and Signal Processing</i> http://www.lss.supelec.fr/MaoriWorkshop/
2013	Responsable de l'organisation du premier <i>Workshop on Optimization for Image and Signal Processing</i> http://www.lss.supelec.fr/MaoriWorkshop/2013
2009	Membre du comité local d'organisation de la conférence SAMPTA à Marseille (http://www.latp.univ-mrs.fr/SAMPTA09/)

10.1.2. Comités de sélection

2015	Membre du comité de sélection pour le poste de maître de conférence à l'univ Paris-Sud (Poste 61MCF1521)
2011	Membre du comité de sélection pour le poste de maître de conférence à l'IUT de Caen (Poste 61MCF0610)

10.1.3. Jurys de thèse

2017	Examineur dans le jury de thèse de <i>Simon Leglaive</i> . Titre de la thèse : <i>Modèles de mélange pour la séparation multicanale de</i>
------	---

10. Activités scientifiques

sources sonores en milieu réverbérant. **Membres du jury** : Cédric Févotte (rapporteur); Laurent Girin (rapporteur); Nancy Berty (examinatrice); Sharon Gannot (examineur); Matthieu Kowalski (examineur); Roland Badeau (directeur de thèse); Gaël Richard (directeur de thèse). **Université** : Toulouse

2017 **Examineur** dans le jury de thèse de *Facundo Costa*. Titre de la thèse : *Bayesian M/EEG Source Localization with possible Joint Skull Conductivity Estimation*. **Membres du jury** : David Brie (rapporteur); Laurent Albera (rapporteur); André Ferrari (président); Matthieu Kowalski (examineur); Thomas Oberlin (examineur); Jean-Yves Tournet (directeur de thèse); Hadj Batatia (co-directeur de thèse). **Université** : Toulouse

2013 **Examineur** dans le jury de thèse de *Thomas Peel*. Titre de la thèse : *Algorithmes de poursuite stochastiques et inégalités de concentration empiriques pour l'apprentissage statistique*. **Membres du jury** : Francois Denis (président); Laurent Daudet (rapporteur); Nicolas Vayatis (rapporteur); Eric Debreuve (examineur); Matthieu Kowalski (examineur); Liva Ralainvola (directeur) Sandrine Anthoine (co-directeur de thèse). **Université** : Aix-Marseille

2013 **Examineur** dans le jury de thèse d'*Alexis Bénichoux*. Titre de la thèse : *Fonctions de coût pour l'estimation des filtres acoustiques dans les mélanges réverbérants*. **Membres du jury** : Laurent Daudet (Président, rapporteur); Mila Nikolova (rapporteur); Emmanuel Bacry (examineur); Matthieu Kowalski (examineur); Rémi Gribonval (directeur de thèse); Emmanuel Vincent (co-directeur de thèse). **Université** : Rennes I

2011 **Examineur** dans le jury de thèse de *Prasad Shudakara Murthy*. Titre de la thèse : *Sparse Models and Convex Optimisation for Convolutional Blind Source Separation*. **Membres du jury** : Jean-Jacques FUCHS (président); Mark PLUMBLEY (rapporteur); Michael ZIBULEVSKY (rapporteur); Abdeldjalil AISSA-EL-BEY (examineur); Matthieu KOWALSKI (examineur); Rémi GRIBONVAL (directeur de thèse). **Université** : Rennes I

10.1.4. Projets financés

Responsable des projets suivants :

Depuis 2015

IMPULSE

Inverse problems in Music Source seParation Using expert knowLEdge

Financement PICS CNRS <http://www.cnrs.fr>

Montant 15 000 €

10. Activités scientifiques

	Durée	3 ans
2014 – 2017	SESAM	<i>Savoir Expert pour les Signaux Audio de Musique</i>
	Financement	Digiteo http://www.digiteo.fr
	Montant	102 000 €
	Durée	3 ans
2013–2016	MAORI	<i>MAthématique de l'Optimisation pour l'Imagerie</i>
	Financement	PGMO Programme Gaspard Monge pour l'Optimisation http://www.fondation-hadamard.fr/PGMO
	Montant	52 000 €
	Durée	3 ans
2010 – 2011		Membre du projets jeune chercheur BISOU <i>BIStochastic Optimisation for mUltiple kernel learning</i>
	Coordination	Sandrine Anthoine (CR CNRS, LATP)
	Financement	Groupe de recherche ISIS

10.1.5. Rayonnement

10.1.5.1. Chercheur invité

2013	Invité par Cédric Févotte à l'Université de Nice (1 semaine)
2012	Invité par Monika Döfler à L'Erwin Schroedinger Institute (Autriche) à l'occasion du workshop Modern Methods of Time-Frequency Analysis II. (3 semaines)
2011	Invité par Monika Döfler à l'Université de Vienne (Autriche) (1 semaine)

10.1.5.2. Orateur invité

2013	Session orale <i>Optimization methods for inverse problems in signal processing</i> du workshop BASP frontier (Chair : Jalal Fadili). Titre : <i>Solving mixed-norm estimate for the MEEG inverse problem using an active-set strategy and proximal algorithm</i> . http://baspfrontiers.epfl.ch
2012	Journée du GDR ISIS <i>Avancées récentes en traitement du signal audio</i> organisée par Valentin Emiya et Cédric Févotte. Titre : <i>'*-lasso therapy : parcimonie et structures, application à la séparation de sources et au débruitage'</i> . http://gdr-isis.fr/index.php?page=reunion&idreunion=177

10. Activités scientifiques

2011	Cours à l'École d'été ERMITES (6ème Ecole Recherche Multimodale d'Information - Techniques & Sciences). http://glotin.univ-tln.fr/ERMITES11/
2010	Journée du GDR ISIS <i>Apprentissage et Parcimonie</i> organisée par Rémi Gribonval et Francis Bach. Titre : <i>Parcimonie et structures pour les décompositions des signaux dans des dictionnaires temps-fréquence</i> http://www.gdr-isis.fr/index.php?page=compte-rendu&idreunion=116

10.2. Évaluations

10.2.1. Relectures

	Évaluation régulière d'articles scientifiques pour les revues et les conférences suivantes :
Revues	IEEE TSP (11 articles), IEEE TALSP (5 articles), Elsevier Signal Processing (13 articles)
Conférences	NIPS, ICASSP, EUSIPCO, ICIP, SAMPTA, DAFX, LVA.

10.2.2. Expertises de projets

	Rapport d'expertise pour l'organisme suivant :
ANR	Programme « Jeunes chercheuses jeunes chercheurs » <i>CE23</i> (2017)
ANR	Programme « Jeunes chercheuses jeunes chercheurs » <i>SIMI 3</i> (2013)

11 ■ Encadrements

11.1. Post-doctoral

2011 H el ene Papadopoulos. Responsable Scientifique lors de son ann ee d'ATER  a l'universit e Paris-Sud et au L2S. A donn e lieu aux publications des articles **[PK13a]** et **[PK11]**.

11.2. Doctoral

11.2.1. Th eses soutenues

2014 – 2017 Fangchen Feng.

Sujet S eparation aveugle de source dans le cas sous d etermin e : de l'instantan e au convolutif.

Encadrement 100%

Soutenance : le 29 septembre 2017 devant le jury compos e de : L. Daudet (pr esident), Laurent Girin (rapporteur), Emmanuel Vincent (rapporteur), Roland Badeau (examineur), Alexandre Gramfort (examineur)

Articles **[FK14a; FK15c; FK15a; FK17]** + 1 article soumis  a Elsevier Signal Processing (r evisions majeures)

Poste actuel Post-Doc  a l'universit e Paris-Diderot au laboratoire AstroParticule et Cosmologie, avec  Eric Chassande-Motin

2014 – 2017 Cl ement Laroche.

Sujet S eparation aveugle de source dans le cas sous d etermin e : de l'instantan e au convolutif.

Encadrement 33%. co-encadrement avec Ga el Richard (HdR) et H el ene Papadopoulos, dans le cadre du projet Digiteo SESAM.

Soutenance : le 16 mai 2017 devant le jury compos e de : L. Daudet (pr esident), S. Marchand (rapporteur), C. F evotte (rapporteur), A. R obel (examineur), O. Gillet (examineur)

11. Encadrements

Articles	[Lar+17; Lar+16; Lar+15a; Lar+15b] + 1 article soumis à IEEE TALSP (révisions mineures)
Poste actuel	Ingénieur de Recherche chez GN Group (Danemark)

11.2.2. Thèses en cours

2016 – 2019	Adrien Meynard
Sujet	synthèse de signaux non-stationnaires large bande.
Encadrement	30%. Co-encadrement avec Bruno Torrèsani (Prof. Aix-Marseille)
2015 – 2018	Alina Meresescu.
Sujet	Déconvolution aveugle des spectromètres issu de l'instrument PFS embarqué à bord de Mars Express.
Encadrement	50%. Co-encadrement avec Frédéric Schmidt (HdR)
Articles	[Mer+17] + 1 article soumis à Elsevier Computer & Geoscience (révisions mineures)

11.3. Master

11.3.1. Master 2

2014	Fangchen Feng	(5 mois)
Titre	Séparation aveugle de source dans le cas sous déterminé : utilisation de la parcimonie et de l'indépendance.	
Formation	M2R ATSI de l'université Paris-Sud	
Encadrement	100%	
2013	Guillaume Mahenc	(5 mois)
Titre	Optimisation convexe pour la reconstruction de signaux à partir de l'amplitude de leur transformée de Fourier à court terme.	
Formation	master ATIAM	
Encadrement	Co-encadrement avec Gaël Richard et Hélène Papadopoulou.	

11.3.2. Master 1

2015	Adrien Meynard	
Titre	Représentations temps-fréquence d'un signal : une approche variationnelle	
Article	[KMW16]	
Encadrement	100%	

11. Encadrements

2015	Cyril Gaudefroy	
	Titre	Music structure estimation ; : Combining musical dimensions, scales and segmentation methods
	Article	[GPK15]
	Encadrement	30%. Co-encadrement avec H�el�ene Papadopoulos.
2013	Fangchen Feng	(3 mois)
	Titre	S�eparation de sources audio : m�elange convolutif sous-d�etermin�e
	Formation	Magist�ere EEA de l'Univ Paris-Sud
	Article	[FK14a] .
	Encadrement	100%
2012	Paul Mondolot	(3 mois)
	Titre	�tude comparatives de d�ebuitage des signaux audios
	Formation	2�e ann�ee de l'�cole Centrale Lyon (niveau master 1)
	Encadrement	100%
2011	Jean Kossaifi	(3 mois)
	Titre	Classification supervis�ee de donn�ees h�et�erog�enes par combinaison de noyaux
	Formation	2�e ann�ee de l'ENSIEE (niveau master 1)
	Encadrement	Co-encadrement � 30% avec Marie Szafranski (MdC, ENSIIE)

11.4. Licence 3

2013	Nicolas Meteye	(2 mois)
	Titre	Reconstruction d'un signal audio : test d'algorithmes sous matlab
	Formation	3�e ann�ee de Licence IST (Paris-Sud)
	Encadrement	100%
2012	Xiang Liu et Zhuocheng LI	(2 mois)
	Titre	�volution de la Toolbox Matlab <i>Underdetermined Reverberant Source Separation</i>
	Formation	3�e ann�ee de Licence IST (Paris-Sud)
	Encadrement	100%

12 ■ Liste des publications

Revue internationale avec comité de lecture

- [Kow16] Matthieu KOWALSKI. “Proximal algorithm meets a conjugate descent”. In : *Pacific journal of optimization* 12.3 (2016), p. 669–695.
- [KMW16] Matthieu KOWALSKI, Adrien MEYNARD et Hau-Tieng WU. “Convex Optimization approach to signals with fast varying instantaneous frequency”. In : *Applied and Computational Harmonic Analysis* (2016).
- [Sch+14] Frédéric SCHMIDT, Irina SHATALINA, Matthieu KOWALSKI, Nicolas GAC, Bortolino SAGGIN et M GIURANNA. “Toward a numerical deshaker for PFS”. In : *Planetary and Space Science* 91 (2014), p. 45–51.
- [Bal+13] Peter BALAZS, Monika DOERFLER, Matthieu KOWALSKI et Bruno TORRESANI. “Adapted and adaptive linear time-frequency representations : a synthesis point of view”. In : *IEEE Signal Processing Magazine* 30.6 (2013), p. 20–31.
- [Gra+13] Alexandre GRAMFORT, Daniel STROHMEIER, Jens HAUEISEN, Matti S HÄMÄLÄINEN et Matthieu KOWALSKI. “Time-frequency mixed-norm estimates : Sparse M/EEG imaging with non-stationary source activations”. In : *NeuroImage* 70 (2013), p. 410–422.
- [KSD13] Matthieu KOWALSKI, Kai SIEDENBURG et Monika DÖRFLER. “Social sparsity! neighborhood systems enrich structured shrinkage operators”. In : *IEEE transactions on signal processing* 61.10 (2013), p. 2498–2511.
- [PK13b] Hélène PAPADOPOULOS et Matthieu KOWALSKI. “Sparse and structured decomposition of audio signals on hybrid dictionaries using musical priors”. In : *The Journal of the Acoustical Society of America* 134.1 (2013), p. 666–685.
- [GKH12] Alexandre GRAMFORT, Matthieu KOWALSKI et Matti HÄMÄLÄINEN. “Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods”. In : *Physics in medicine and biology* 57.7 (2012), p. 1937.

12. Liste des publications

- [KVG10] Matthieu KOWALSKI, Emmanuel VINCENT et Rémi GRIBONVAL. “Beyond the narrowband approximation : Wideband convex methods for under-determined reverberant audio source separation”. In : *IEEE Transactions on Audio, Speech, and Language Processing* 18.7 (2010), p. 1818–1829.
- [Kow09] Matthieu KOWALSKI. “Sparse regression using mixed norms”. In : *Applied and Computational Harmonic Analysis* 27.3 (2009), p. 303–324.
- [KT09d] Matthieu KOWALSKI et Bruno TORRÉSANI. “Sparsity and persistence : mixed norms provide simple signal models with dependent coefficients”. In : *Signal, image and video processing* 3.3 (2009), p. 251–264.
- [KT08] Matthieu KOWALSKI et Bruno TORRÉSANI. “Random models for sparse signals expansion on unions of bases with application to audio signals”. In : *IEEE Transactions on Signal Processing* 56.8 (2008), p. 3468–3481.

Revue nationale avec comité de lecture

- [KG10] Matthieu KOWALSKI et Alexandre GRAMFORT. “A priori par normes mixtes pour les problèmes inverses : Application à la localisation de sources en M/EEG”. In : *Traitement du Signal* 27.1 (2010), p. 51–76.

Conférences internationales avec actes et comité de lecture

- [FK17] Fangchen FENG et Matthieu KOWALSKI. “Sparsity and low-rank amplitude based blind Source Separation”. In : *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. 2017.
- [Lar+17] Clément LAROCHE, Hélène PAPADOPOULOS, Matthieu KOWALSKI et Gaël RICHARD. “Drum extraction in single channel audio signals using multi-layer non negative matrix factor deconvolution”. In : *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. 2017.
- [Lar+16] Clément LAROCHE, Hélène PAPADOPOULOS, Matthieu KOWALSKI et Gaël RICHARD. “Genre specific dictionaries for harmonic/percussive source separation”. In : *ISMIR The 17th International Society for Music Information Retrieval Conference*. 2016.
- [FK15a] Fangchen FENG et Matthieu KOWALSKI. “An unified approach for blind source separation using sparsity and decorrelation”. In : *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE. 2015, p. 1736–1740.
- [FK15c] Cédric FÉVOTTE et Matthieu KOWALSKI. “Hybrid sparse and low-rank time-frequency signal decomposition”. In : *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE. 2015, p. 464–468.

12. Liste des publications

- [GPK15] Cyril GAUDEFROY, Hélène PAPADOPOULOS et Matthieu KOWALSKI. “A multi-dimensional meter-adaptive method for automatic segmentation of music”. In : *Content-Based Multimedia Indexing (CBMI), 2015 13th International Workshop on*. IEEE. 2015, p. 1–6.
- [KG15a] Matthieu KOWALSKI et Alexandre GRAMFORT. “Inverse problems with time-frequency dictionaries and non-white Gaussian noise”. In : *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE. 2015, p. 1741–1745.
- [Lar+15a] Clément LAROCHE, Matthieu KOWALSKI, Helene PAPADOPOULOS et Gaël RICHARD. “A structured nonnegative matrix factorization for source separation”. In : *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE. 2015, p. 2033–2037.
- [FK14a] Fangchen FENG et Matthieu KOWALSKI. “Hybrid model and structured sparsity for under-determined convolutive audio source separation”. In : *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE. 2014, p. 6682–6686.
- [FK14b] Cédric FÉVOTTE et Matthieu KOWALSKI. “Low-rank time-frequency synthesis”. In : *Advances in Neural Information Processing Systems*. 2014, p. 3563–3571.
- [Kow14] Matthieu KOWALSKI. “Thresholding rules and iterative shrinkage/thresholding algorithm : A convergence study”. In : *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE. 2014, p. 4151–4155.
- [SKD14] Kai SIEDENBURG, Matthieu KOWALSKI et Monika DORFLER. “Audio declipping with social sparsity”. In : *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE. 2014, p. 1577–1581.
- [Sha+13] Irina SHATALINA, Frederic SCHMIDT, Bortolino SAGGIN, Nicolas GAC, Matthieu KOWALSKI et Marco GIURANNA. “Analytical model and spectral correction of vibration effects on Fourier transform spectrometer”. In : *SPIE Remote sensing 2013*. T. 8890. 2013, 88900S–88900S.
- [Gra+11a] Alexandre GRAMFORT, Daniel STROHMEIER, Jens HAUEISEN, Matti HAMALAINEN et Matthieu KOWALSKI. “Functional brain imaging with M/EEG using structured sparsity in time-frequency dictionaries”. In : *Information Processing in Medical Imaging*. Springer. 2011, p. 600–611.
- [KR11a] Matthieu KOWALSKI et Thomas RODET. “An unsupervised algorithm for hybrid/morphological signal decomposition”. In : *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE. 2011, p. 4112–4115.
- [Kow+11] Matthieu KOWALSKI, Pierre WEISS, Alexandre GRAMFORT et Sandrine ANTHOINE. “Accelerating ISTA with an active set strategy”. In : *OPT 2011 : 4th International Workshop on Optimization for Machine Learning*. 2011, p. 7.
- [PK11] Hélène PAPADOPOULOS et Matthieu KOWALSKI. “Sparse signal decomposition on hybrid dictionaries using musical priors”. In : *ISMIR 2011*. 2011, p–687.

12. Liste des publications

- [Str+11] Daniel STROHMEIER, Alexandre GRAMFORT, Jens HAUEISEN, Matti HÄMÄLÄINEN et Matthieu KOWALSKI. “MEG/EEG source reconstruction based on Gabor thresholding in the source space”. In : *Noninvasive Functional Source Imaging of the Brain and Heart & 2011 8th International Conference on Bioelectromagnetism (NFSI & ICBEM), 2011 8th International Symposium on*. IEEE. 2011, p. 103–108.
- [GK09] Alexandre GRAMFORT et Matthieu KOWALSKI. “Improving m/eeg source localization with an inter-condition sparse prior”. In : *Biomedical Imaging : From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*. IEEE. 2009, p. 141–144.
- [KSR09] Matthieu KOWALSKI, Marie SZAFRANSKI et Liva RALAIVOLA. “Multiple indefinite kernel learning with mixed norm regularization”. In : *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM. 2009, p. 545–552.
- [KT09c] Matthieu KOWALSKI et Bruno TORRÉSANI. “Sparsity and persistence in time-frequency sound representations”. In : *Wavelets XIII*. T. 7446. SPIE. 2009, 74460F.
- [KT09e] Matthieu KOWALSKI et Bruno TORRÉSANI. “Structured sparsity : from mixed norms to structured shrinkage”. In : *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*. 2009.
- [KVG08] Matthieu KOWALSKI, Emmanuel VINCENT et Rémi GRIBONVAL. “Under-determined source separation via mixed-norm regularized minimization”. In : *Signal Processing Conference, 2008 16th European*. IEEE. 2008, p. 1–5.
- [KT06] Matthieu KOWALSKI et Bruno TORRÉSANI. “A family of random waveform models for audio coding”. In : *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. T. 3. IEEE. 2006, p. III–III.
- [KT05] Matthieu KOWALSKI et Bruno TORRÉSANI. “A study of Bernoulli and structured random waveform models for audio signals”. In : *SPARS 05*. 2005, TS2–1.

Conférences nationales avec actes et comité de lecture

- [Mer+17] Alina-Georgiana MERESESCU, Matthieu KOWALSKI, Frédéric SCHMIDT et François LANDAIS. “Estimation du Temps de Résidence Hydrologique : Déconvolution 1D”. In : *Gretsi 2017*. 2017.
- [FK15b] Fangchen FENG et Matthieu KOWALSKI. “Vers une approche unifiée pour la séparation aveugle de sources en sur et sous-déterminé, basée sur la parcimonie et la décorrélation”. In : *Gretsi 2015*. 2015.

12. Liste des publications

- [KG15b] Matthieu KOWALSKI et Alexandre GRAMFORT. “Problèmes inverses avec dictionnaires temps-fréquences en présence de bruit Gaussien non blanc”. In : *Gretsi 2015*. 2015.
- [Lar+15b] Clément LAROCHE, Matthieu KOWALSKI, Helene PAPADOPOULOS et Gaël RICHARD. “Méthode Structurée de décomposition en matrices non-négatives appliquée à la séparation de sources audio”. In : *Gretsi 2015*. 2015.
- [PK13a] Hélène PAPADOPOULOS et Matthieu KOWALSKI. “Décomposition parcimonieuse structurée des signaux audio de musique guidée par l’information experte”. In : *GRETSI 2013*. 2013, 4–p.
- [SDK13] Kai SIEDENBURG, Monika DÖRFLER et Matthieu KOWALSKI. “Parcimonie Sociale et application à l’audio inpainting”. In : *GRETSI 2013*. 2013.
- [Gra+11b] Alexandre GRAMFORT, Daniel STROHMEIER, Jens HAUEISEN, Matti HÄMÄLÄINEN et Matthieu KOWALSKI. “Apport des dictionnaires temps-fréquence, de la parcimonie et des structures pour l’imagerie cérébrale fonctionnelle M/EEG”. In : *GRETSI 2011*. 2011.
- [KR11b] Matthieu KOWALSKI et Thomas RODET. “Un algorithme de seuillage itératif non-supervisé pour la décomposition parcimonieuse de signaux dans des unions de dictionnaires”. In : *GRETSI 2011*. 2011.
- [KG09] Matthieu KOWALSKI et Alexandre GRAMFORT. “A priori par normes mixtes pour les problèmes inverses Application à la localisation de sources en M/EEG”. In : *XXIIe colloque GRETSI (traitement du signal et des images), Dijon (FRA), 8-11 septembre 2009*. GRETSI, Groupe d’Etudes du Traitement du Signal et des Images. 2009.
- [KT09a] Matthieu KOWALSKI et Bruno TORRÉSANI. “Décompositions parcimonieuses et persistantes de signaux multicanaux. Applications aux signaux MEEG.” In : *CAp 2008*. 2009, pages–105.
- [KT09b] Matthieu KOWALSKI et Bruno TORRÉSANI. “Modèles aléatoires hiérarchiques pour la modélisation de signaux audio”. In : *XXIIe colloque GRETSI (traitement du signal et des images), Dijon (FRA), 8-11 septembre 2009*. GRETSI, Groupe d’Etudes du Traitement du Signal et des Images. 2009.
- [KT07] Matthieu KOWALSKI et Bruno TORRESANI. “Identification de signaux parcimonieux dans un dictionnaire hybride”. In : *GRETSI 2007*. 2007, p. 1249.