

# Deux méthodes de déconvolution et séparation simultanées ; application à la reconstruction des amas de galaxies

Two approaches for the simultaneous separation and deblurring ; application to astrophysical data

**S. Anthoine<sup>1</sup>, E. Pierpaoli<sup>2</sup>, I. Daubechies<sup>3</sup>**

<sup>1</sup>LATP – CMI – Université de Provence – Aix-Marseille  
39 rue Joliot-Curie, 13453 Marseille Cedex 13, France  
anthoine@cmi.univ-mrs.fr

<sup>2</sup>California Institute of Technology  
Mail Code 130-33, Pasadena, CA, 91125, USA  
pierpa@caltech.edu

<sup>3</sup>Program in Applied and Computational Mathematics  
Princeton University, Princeton, NJ, 08544, USA  
ingrid@math.princeton.edu

Manuscrit reçu le 11 octobre 2005

Résumé et mots clés

Nous présentons deux approches pour résoudre le problème de séparation et de déconvolution simultanées de mélanges de composantes. La première est basée sur une description statistique des coefficients d'ondelettes des signaux. La seconde consiste à minimiser une fonctionnelle variationnelle. Nous appliquons ces deux méthodes à la reconstruction des amas de galaxies par l'effet Sunyaev-Zel'dovich dans le cadre de la mission d'observation du fond diffus cosmique par *ACT*. Nous trouvons que pour des paramètres appropriés, les deux méthodes donnent des résultats comparables et que prendre en compte le caractère non-gaussien des amas très intenses améliore nettement leur reconstruction.

Estimation/séparation/déconvolution de signaux, approche statistique, approche variationnelle, ondelettes, astrophysique.

Abstract and key words

Two approaches are presented to solve the problem of simultaneously deconvolving and separating mixtures of components. The first one uses a statistical description of the wavelet coefficients of the signals. The second one consists in minimizing a variational functional. Both methods are applied to the reconstruction of Sunyaev-Zel'dovich galaxy clusters from Cosmic Microwave Background experiments such as *ACT*. We find that both methods, when tuned, yield similar results and that the reconstruction of intense clusters is substantially improved when their non-gaussianity is taken into account.

Signal estimation/separation/deconvolution, statistical approach, variational approach, wavelets, astrophysics.

Remerciements

Ce travail a été en grande partie fait au Program in Applied and Computational Mathematics à Princeton University, États-Unis, dans le cadre de la thèse de S.A, financée en partie par Burroughs Welcome Fund.



# 1. Introduction

Le problème que nous nous proposons de résoudre est le suivant : nous souhaitons estimer plusieurs composantes, notées  $x^1, x^2, \dots, x^M$ , à partir de plusieurs observations, notées  $y^1, y^2, \dots, y^L$ . Chaque observation est une version floue et bruitée d'un mélange (combinaison linéaire) des composantes originales. Autrement dit,  $y^l$ , l'observation numéro  $l$ , s'écrit :

$$y^l = \sum_{m=1}^M f^{m,l} x^m * b^l + n^l \quad (1)$$

où  $*$  désigne l'opérateur de convolution, les scalaires  $f^{m,l}$  ainsi que les noyaux de convolution  $b^l$  sont supposés connus, et  $n^l$  est un bruit de variance connue :  $\sigma_l^2$ .

Les observations du fond diffus cosmique (rayonnement fossile) en astrophysique sont un exemple de données correspondant à ce modèle. Typiquement, on acquiert plusieurs images de la même portion de ciel, notées  $y_{l,l=1\dots L}$ , à des longueurs d'onde différentes, notées  $\nu_{l,l=1\dots L}$ . On a donc  $y^l = y(\nu_l)$ . Chacune de ces images  $y_l$  est une superposition de différents phénomènes physiques tels que le fond diffus cosmique, les amas de galaxies, les points sources infrarouges... La contribution de chacun de ces phénomènes dans l'observation  $y_l$  dépend non seulement du point de l'espace observé mais aussi de la longueur d'onde d'observation  $\nu_l$ . Les dépendances en espace et en longueur d'onde sont en fait découplées, impliquant que la contribution du  $m$ -ième phénomène peut être vue comme le produit  $f^{m,l} x^m$  d'un scalaire  $f^{m,l}$  appelé la dépendance fréquentielle du phénomène  $m$  par une carte spatiale  $x^m$  indépendante de la longueur d'onde. Le processus d'acquisition de l'image sur le télescope se traduit par une convolution par un noyau  $b^l$ . Enfin les imperfections d'observation ainsi que les phénomènes physiques non considérés sont modélisés par les bruits  $n^l$ . Connaissant les dépendances fréquentielles  $f^{m,l}$ , les noyaux de convolution  $b^l$  et une estimation de la taille des bruits  $n^l$ , nous souhaitons donc estimer les cartes spatiales  $x_m, m=1\dots M$  à partir des images observées  $y_{l,l=1\dots L}$  vérifiant l'équation (1).

Le filtre de Wiener, habituellement utilisé pour estimer une composante à partir d'une observation, peut être adapté au cas où l'on cherche à estimer plusieurs composantes à partir de plusieurs observations. Optimal lorsque les composantes peuvent être décrites comme des processus gaussiens, il donne de bons résultats lorsque cette hypothèse est une bonne approximation. Si le modèle gaussien est approprié pour le fond diffus cosmique, ce n'est pas le cas pour les amas de galaxies, qui sont des structures spatialement localisées avec un pic d'intensité central et des structures filamenteuses vers l'extérieur. En conséquence, utiliser une méthode de type filtre de Wiener donne des estimations peu précises des amas de galaxies. De plus, ce manque de précision pollue l'estimation du fond diffus cosmique. Ainsi, la mauvaise modélisation des structures spatiales des amas de galaxies cause des problèmes. Il semble donc utile de dévelop-

per des méthodes générales qui prennent en compte les caractéristiques spatiales de chaque signal, plutôt que d'avoir un modèle commun. La structure spatiale d'un signal peut être décrite de différentes manières. Pour le filtre de Wiener, cette description est statistique, mais on peut aussi utiliser d'autres cadres, comme par exemple le cadre fonctionnel pour décrire les propriétés des signaux. Dans ce cadre, les signaux sont modélisés par des fonctions et leur appartenance à des espaces fonctionnels permet de préciser leurs propriétés. Dans cet article, nous présentons deux méthodes pour traiter l'équation (1) en tenant compte des propriétés spatiales de chaque composante  $x^m$ , la première méthode utilise une description statistique des signaux et la seconde une description fonctionnelle. Nous procédons ensuite à une étude de leurs performances pour le problème particulier de l'extraction des amas de galaxies dans le cadre de la mission *ACT* de mesure des anisotropies du fond diffus cosmique.

Cet article est constitué de deux parties principales outre la présente introduction et la conclusion. La première partie (section 2) est consacrée à la description des deux méthodes que nous proposons. La méthode présentée au paragraphe 2.2 s'inscrit dans le cadre statistique. Elle est inspirée de [9] et a déjà présentée succinctement dans [7]. Elle exploite le fait que les coefficients d'ondelettes varient de manière cohérente à travers les échelles et à des positions proches les unes des autres. Chaque composante est caractérisée par les statistiques de « voisinages » de coefficients en ondelettes et une estimation bayésienne aux moindres carrés est calculée. Nous décrivons au paragraphe 2.3 une seconde méthode qui utilise le cadre fonctionnel. Elle consiste à trouver les minimiseurs d'une fonctionnelle variationnelle composée d'un terme d'erreur par observation et d'un terme de pénalisation par composante. Chaque terme de pénalisation est adapté à la composante correspondante : c'est la norme dans un espace fonctionnel approprié. Nous décrivons un algorithme itératif qui permet de minimiser cette fonctionnelle variationnelle. La seconde partie de cet article (section 3) est dédiée à l'application des deux méthodes à un problème spécifique : l'estimation des cartes d'amas de galaxies à partir de mesures multifréquences des anisotropies du fond diffus cosmique. Nous présentons des résultats pour chacune des méthodes sur des données simulées avec les paramètres du projet *ACT*. Nous spécifions au paragraphe 3.1 la provenance et les caractéristiques des données sur lesquelles l'étude est faite. Les paramètres utilisés pour chacune des méthodes sont décrits aux paragraphes 3.2.1 et 3.2.2. Enfin les résultats obtenus pour chaque méthode sont comparés qualitativement et quantitativement au paragraphe 3.3. Pour ce qui est de l'étude quantitative, nous définissons et utilisons des mesures non standard adaptées aux cartes d'amas de galaxies.

## 2. Deux méthodes de déconvolution de mélanges de composantes

### 2.1. Notations

Notons d'abord qu'étant donné que l'application présentée met en jeu des données de type image, nous plaçons dans le cadre de données 2-dimensionnelles. En réalité les méthodes présentées s'appliquent en d'autres dimensions.

Nous considérons des décompositions de type ondelettes pour le signal  $x$  que nous notons de la façon suivante :

$$x = \sum_{q \in \mathbb{Z}} \langle x, \phi_q \rangle \phi_q + \sum_{j=0}^J \sum_{q \in 2^{-j} \mathbb{Z}^2} \langle x, \psi_{j,q} \rangle \psi_{j,q}$$

où  $\phi$  est la fonction d'échelle, les  $\phi_q(\cdot) = \phi(\cdot - q)$  ses translatées en espace ;  $\psi$  est l'ondelette mère et les  $\psi_{j,q} = \psi(2^j \cdot - q)$  sont ses translatées et dilatées ; et  $\langle \cdot, \cdot \rangle$  désigne le produit scalaire. Ici,  $j$  est un scalaire désignant l'échelle et  $q$  un vecteur indiquant une position dans le plan de l'image. Nous notons  $x_{j,q} = \langle x, \psi_{j,q} \rangle$  le coefficient d'ondelette du signal  $x$  à l'échelle  $j$  et la position  $q$ .

Enfin nous utiliserons des caractères gras pour les vecteurs (autres que les positions) et les matrices.

### 2.2. Estimation bayésienne de voisinages de coefficients d'ondelettes

Cette méthode a été succinctement décrite dans [7] et nous la reprenons ici plus en détails. Les ondelettes sont des fonctions bien localisées en espace ainsi qu'en fréquence. Un coefficient d'ondelette  $x_{j,q}$  donne de l'information sur la structure du signal  $x$  à l'échelle  $j$  et au point  $q$ . Ainsi une description du comportement des coefficients en ondelette doit permettre de modéliser les propriétés spatiales d'un signal  $x$ . Nous décrivons d'abord le modèle statistique que nous utilisons puis nous dérivons l'estimateur bayésien correspondant à notre problème. Pour plus de clarté, cette dérivation est faite en plusieurs étapes : d'abord dans le cadre du débruitage d'une unique composante ( $M = L = 1, T = \text{Id}$ ), puis dans le cas général.

#### 2.2.1 Modèle statistique

L'une des propriétés de la transformée en ondelette ayant participé à son succès est le fait que les coefficients obtenus sont décorrélés. En effet, ceci permet de procéder à des opérations

telles que le seuillage, une opération peu coûteuse et efficace pour les problèmes de débruitage ou de compression car chaque coefficient peut être traité individuellement du fait de la décorrélation. Cependant nombreux travaux [14, 4, 11, 9, 10] ont souligné que cette décorrélation n'était pas synonyme d'indépendance mais qu'il existe au contraire des dépendances spatiales entre les coefficients d'ondelettes. Intuitivement, cela s'explique par le fait que si le signal  $x$  a une structure particulière au point  $q_0$  et à l'échelle  $j_0$ , alors les coefficients d'ondelettes à des échelles ou des positions proches sont en général aussi affectés. Par exemple, si  $x$  a une discontinuité au point  $q_0$ , tous les coefficients d'ondelettes  $\{x_{j,q_0}\}_j$  sont grands. On peut donc espérer faire une meilleure estimation du coefficient  $x_{j,q}$  en considérant un ensemble de coefficients à des échelles proches de  $j$  et des points proches de  $q$ , autrement dit un « voisinage » de coefficients.

La plupart des méthodes prenant en compte la dépendance des coefficients n'utilisent en fait que la relation entre un coefficient  $x_{j,q}$  et son parent  $x_{j-1,q}$  [14, 4, 11], ce qui permet de bien modéliser des structures de type discontinuité qui sont importantes dans les images naturelles. Dans notre problème, d'une part les objets que nous décrivons (fond cosmique, amas de galaxies) ont des formes géométriques d'étendue bien déterminée, et d'autre part la présence de l'opérateur de convolution vient corréler les coefficients d'ondelette dans les observations. Il est donc préférable de considérer des voisinages d'ondelettes plus étendus comme proposé dans [9, 10]. Un exemple de voisinage de  $x_{j,q}$  est :  $\mathbf{x} \equiv \mathbf{x}_{j,q} = (x_{j,q}, x_{j,q-1_x}, x_{j,q+1_x}, x_{j,q-1_y}, x_{j,q+1_y}, x_{j-1,q})$ , c'est-à-dire, le vecteur contenant le coefficient  $x_{j,q}$ , son parent  $x_{j-1,q}$  et ses quatre plus proches voisins à la même échelle.

Plusieurs modèles pour la distribution jointe de ces vecteurs ont été proposés. En particulier, [4, 13] utilisent des mélanges de deux gaussiennes, l'une à forte variance permettant de capturer la queue lourde de la distribution marginale des coefficients d'ondelettes, et l'autre à faible variance modélisant la parcimonie de la transformée (nombreux coefficients proches de zéro). Là encore, nous souhaitons un modèle plus flexible car nous aurons à faire à des mélanges de signaux. Nous suivons la proposition de [9] et modélisons le voisinage  $\mathbf{x}_{j,q}$  par un mélange de gaussiennes :

$$\mathbf{x}_{j,q} \equiv \sqrt{z} \mathbf{u} \text{ en distribution} \quad (2)$$

où  $z$  est une variable aléatoire scalaire, indépendante de  $\mathbf{u}$ , de moyenne 1 et de distribution  $p_z$  et  $\mathbf{u}$  est un vecteur gaussien centré, de covariance  $\mathbf{C}_u$ .

Les paramètres d'un tel mélange de gaussiennes sont la probabilité  $p_z$  et la matrice de covariance  $\mathbf{C}_u$  (car l'espérance de  $z$  est fixée à 1). S'il est raisonnable de penser que nos signaux ont les mêmes propriétés en tout point de l'espace – les amas de galaxies sont distribués plus ou moins uniformément dans l'espace –, ce n'est pas vrai pour les échelles. Par exemple, la taille typique d'un amas de galaxies est de l'ordre de quelques minutes d'arc. Les coefficients d'ondelettes aux échelles  $j$  cor-

respondant à cette taille se comporteront donc de manière bien différente des autres. C'est pourquoi nous supposons que le modèle de l'éq.(2) dépend uniquement de l'échelle  $j$  et du signal  $x$  mais pas de la localisation en espace  $q$ . Fixer le modèle complet revient donc à se donner la probabilité  $p_z$  et la matrice de covariance  $\mathbf{C}_u$  pour chaque l'échelle  $j$  et pour chaque composante  $x^m$ . Il convient de noter ici qu'un paramètre est caché : la taille exacte du voisinage  $x_{j,q}$ . Nous reviendrons sur la détermination de ces trois paramètres au paragraphe 2.2.3. Ce modèle étant fixé, nous expliquons d'abord comment calculer l'estimateur bayésien aux moindres carrés.

### 2.2.2. Dérivation de l'estimateur

Nous travaillons sous l'hypothèse que les bruits  $n^l$  sont des processus gaussien (bruits blancs ou colorés) indépendants deux à deux et indépendants des processus  $x^m$ . Le voisinage  $\mathbf{x}_{j,q}^m$  est estimé pour chaque échelle  $j$ , localisation  $q$  et composante  $x^m$  et l'on en garde que l'estimation du premier coefficient  $x_{j,q}^m$ . L'estimation de  $x^m$  est alors synthétisée à partir des coefficients d'ondelettes estimés.

#### Débruitage d'une composante

Plaçons-nous tout d'abord dans le cas où l'on estime une unique composante  $x$  à partir d'une observation bruitée  $y = x + n$  ( $M = L = 1, T = \text{Id}$ ). Les « voisinages » vérifient alors

$$\mathbf{y}_{j,q} = \mathbf{x}_{j,q} + \mathbf{n}_{j,q}. \quad (3)$$

Si de plus, on suppose  $p_z = \delta_{\{z=z_0\}}$ , alors les voisinages sont tous des vecteurs gaussiens et l'on a  $\mathbf{y}_{j,q} \stackrel{\text{dist.}}{\equiv} \sqrt{z_0} \mathbf{u}_{j,q} + \mathbf{n}_{j,q}$ . L'estimateur bayésien de  $\mathbf{x}$  n'est autre que le filtre de Wiener appliqué aux voisinages d'ondelettes observés :

$$E\{\mathbf{x}_{j,q} | \mathbf{y}_{j,q}, z = z_0\} = z_0 \mathbf{C}_{x,j} (z_0 \mathbf{C}_{x,j} + \mathbf{C}_{n,j})^{-1} \mathbf{y}_{j,q}. \quad (4)$$

où  $\mathbf{C}_{x,j}$  (resp.  $\mathbf{C}_{n,j}$ ) désigne la matrice de covariance des voisinages d'ondelettes du signal  $x$  (resp. du bruit  $n$ ) à l'échelle  $j$ . (Notons que  $\mathbf{C}_{x,j} = \mathbf{C}_{u,j}$  car l'espérance de  $z$  est 1.)

En levant le conditionnement par  $z$ , on obtient donc pour  $p_z$  quelconque une combinaison linéaire de filtres de Wiener, les poids étant donnés par la règle de Bayes :

$$E\{\mathbf{x}_{j,q} | \mathbf{y}_{j,q}\} = \int_0^\infty p(z = z_0 | \mathbf{y}_{j,q}) E\{\mathbf{x}_{j,q} | \mathbf{y}_{j,q}, z = z_0\} dz_0 \quad (5)$$

$$\text{où } p(z = z_0 | \mathbf{y}_{j,q}) = \frac{p(\mathbf{y}_{j,q} | z = z_0) p_z(z_0)}{\int p(\mathbf{y}_{j,q} | z = z') p_z(z') dz'} \quad (6)$$

et  $p(\mathbf{y}_{j,q} | z = z')$  est une gaussienne centrée de covariance  $z' \mathbf{C}_{x,j} + \mathbf{C}_{n,j}$ . Ce cas a été présenté dans [9].

#### Déconvolution de mélanges de composantes

Nous avons étendu le résultat précédent au cas des mélanges de composantes décrits par l'éq. (1) (cf. [7], [1]). À échelle  $j$  et

position  $q$  fixées, les  $M$  « voisinages »  $\{\mathbf{x}_{j,q}^m\}_{m=1,\dots,M}$  sont estimés simultanément pour toutes les composantes  $x^m$  à partir des  $L$  voisinages observés  $\{\mathbf{y}_{j,q}^l\}_{l=1,\dots,L}$ . Pour chaque  $m$ , le voisinage  $\mathbf{x}_{j,q}^m$  est modélisé par un mélange de gaussiennes :

$$\mathbf{x}_{j,q}^m \stackrel{\text{dist.}}{\equiv} \sqrt{z^{m,j}} \mathbf{u}^{m,j}, \quad (7)$$

qui est déterminé par la probabilité  $p_{z^{m,j}}$  et la matrice de covariance  $\mathbf{C}_u^{m,j}$ . Cette fois-ci, la présence des convolutions empêche d'écrire de manière exacte une équation du type

$$\mathbf{y}_{j,q}^l = f^l g^l (\{\mathbf{x}_{j,q}^m\}_{m=1,\dots,M}) + \mathbf{n}_{j,q}^l \quad (8)$$

avec  $f^l g^l$  linéaire car les noyaux de convolution  $b^l$  ne sont pas nécessairement à support finis. (Potentiellement tous les coefficients  $x_{j,q}^m$  contribuent au coefficient  $y_{j,q}^l$ ). Néanmoins, cette approximation est valable dès que la taille des voisinages est assez grande (voir [1] pour plus de détails). Le choix de la taille des voisinages dépend donc des noyaux de convolution  $b^l$  ainsi que de l'échelle  $j$ .

Comme précédemment, si les  $p_{z^{m,j}}$  sont des fonctions  $\delta$  pour toutes les composantes, on retrouve l'hypothèse gaussienne. Dans ce cas, l'estimation des  $\mathbf{x}_{j,q}^m$  est donnée par le filtre de Wiener multi-objets/multi-observations sur les voisinages de coefficients d'ondelettes. Notons que ces filtres prennent en compte le mélange par les coefficients  $f^{m,l}$  et les convolutions par les  $b^l$ . L'estimation du modèle complet est une combinaison linéaire de tels filtres, les poids étant donnés par les probabilités à postériori :

$$p(\{z^m\}_{m=1,\dots,M} | \{\mathbf{y}_{j,q}^l\}_{l=1,\dots,L}). \quad (9)$$

Comme pour le cas monodimensionnel, les probabilités à postériori sont calculées en utilisant la règle de Bayes, et en conséquence sont des combinaisons de gaussiennes multidimensionnelles  $p(\{\mathbf{y}_{j,q}^l\}_{l=1,\dots,L} | \{z^m\}_{m=1,\dots,M})$  avec les probabilités à priori  $p_{z^m}, m = 1, \dots, M$ . Nous notons ici encore que les coefficients de mélange  $f^{m,l}$  et les convolutions par les  $b^l$  interviennent dans les probabilités  $p(\{\mathbf{y}_{j,q}^l\}_{l=1,\dots,L} | \{z^m\}_{m=1,\dots,M})$ . Une description détaillée des formules obtenues se trouve dans [7, 1].

### 2.2.3. Paramètres

Comme nous l'avons vu précédemment, la taille des voisinages considérés doit dépendre de l'échelle  $j$  (car la transformée a un pas de translation dépendant de  $j$ ), ainsi que des noyaux de convolution  $b_l$  (puisque la convolution corréle les coefficients). D'autre part, le calcul de l'estimateur bayésien requiert les matrices de covariances des voisinages à chaque échelle pour chaque composante et pour le bruit ainsi que des probabilités  $p_{z^m}$ . Le bruit étant gaussien, les matrices de covariance pour les voisinages du bruit dépendent de l'ondelette choisie ainsi que de la covariance initiale du bruit qui est supposée connue. Si ces données sont connues sous forme analytique, on peut calculer

théoriquement les matrices de covariance pour les termes de bruit, sinon, on les calcule numériquement au préalable. En théorie, il est aussi possible d'estimer les matrices de covariances des voisinages des composantes  $x^m$  à partir des observations. Néanmoins cette méthode ne marche pas nécessairement en pratique, car certains signaux ont des intensités trop faibles. Il est alors nécessaire de déterminer ces matrices au préalable sur des simulations de chaque composante  $x^m$ . C'est la méthode que nous emploierons pour notre application à l'extraction des amas de galaxies. Nous expliquons dans le paragraphe 3.2.1 comment déterminer les densités de probabilités  $p_z$  dans le cadre de cette application.

### 2.3. Minimisation d'une fonctionnelle variationnelle par un algorithme itératif

Dans cette section, nous proposons une autre méthode pour estimer les signaux  $x^m$  dans laquelle la description des propriétés spatiales de ces signaux passe par l'utilisation d'espace fonctionnels. Nombreux travaux en traitement d'images ont prouvé le bien-fondé de l'utilisation de tels espaces, l'un des pionniers étant [12]. On peut citer comme autre exemple [3], qui a mis en évidence le fait que chercher une solution dans l'espace des fonctions à variations bornées permet d'obtenir des estimations d'images naturelles où les contours sont bien préservés. La méthode que nous proposons ici consiste à minimiser une fonctionnelle variationnelle. Les solutions ne sont pas recherchées dans un espace particulier mais simplement dans  $L^2$ . Les espaces fonctionnels interviennent dans la fonctionnelle elle-même par l'intermédiaire de termes dit de régularisation. Le but étant que les estimations trouvées ont une petite norme vis-à-vis de ces termes de régularisation. Ainsi les espaces fonctionnels sont définis de manière à ce que des signaux plausibles sont des éléments de petite norme dans ces espaces.

Nous considérons en particulier des espaces fonctionnels définis par le comportement de coefficients sur un système de décomposition. Par exemple, l'espace de Besov  $B_1^{1,1}$  peut-être défini comme l'espace des fonctions vérifiant que leurs décomposition en ondelettes est une suite sommable.

Dans la suite, nous nous plaçons dans le cas où chaque signal d'intérêt ( $x^m$ ,  $y^l$ ,  $b^l$  et  $n^l$ ) est une fonction de carré intégrable.

#### 2.3.1. Modèle

Posons  $X = (x^1, x^2, \dots, x^M)$ ,  $Y = (y^1, y^2, \dots, y^L)$  et  $N = (n^1, n^2, \dots, n^L)$ , alors l'équation (1) s'écrit :  $Y = KX + N$  et  $K$  est un opérateur linéaire continu de  $(L^2)^M$  dans  $(L^2)^L$ . Nous cherchons donc à « inverser »  $K$  tout en contrôlant la taille du terme dû au bruit après inversion ( $K^{-1}N$ ). Comme  $K$  n'a pas nécessairement un inverse borné, le mieux que l'on puisse espérer est un schéma régularisant i.e.

$$\forall X, \lim_{\epsilon \rightarrow 0} \sup_{\|N\| < \epsilon} \|K^{-1}(KX + N) - X\| \rightarrow 0. \quad (10)$$

Un tel schéma assure que si la qualité des observations augmente ( $\|N\| \rightarrow 0$ ), l'estimation devient de plus en plus précise. La minimisation d'une fonctionnelle variationnelle permet de définir un tel schéma. Par exemple, la minimisation de

$$J_{L^2}(X) = \|KX - Y\|_{(L^2)^L}^2 + \gamma \|X\|_{(L^2)^M}^2 \quad (11)$$

$$J_{L^2}(X) = \sum_{l=1}^L \|(KX)^l - y^l\|_{L^2}^2 + \gamma \sum_{m=1}^M \|X^m\|_{L^2}^2, \quad (12)$$

avec  $\gamma$  strictement positif, est une méthode régularisante pour notre problème. Minimiser cette fonctionnelle revient à chercher des composantes  $X^m$  dont l'image par l'opérateur  $K$  est aussi près que possible des observations  $Y^l$  (premier terme de l'éq.(11)) et de norme aussi petite que possible dans  $L^2$  (second terme de l'éq.(11)). Toutefois, elle traite chacune des composantes de la même façon. Qui plus est, le choix de la norme  $L^2$  n'est pas nécessairement adéquat pour toutes les composantes  $X^m$ . Par exemple, ajouter des oscillations change peu la norme quadratique d'un objet. Si  $K$  est un filtre basse fréquence, alors ajouter des oscillations hautes fréquences à une composante  $X^m$  change peu la fonctionnelle. Il est donc très possible d'obtenir une estimation  $X^m$  oscillant à haute fréquence. Ceci n'est pas acceptable par exemple pour la composante représentant les amas de galaxies puisqu'on sait qu'ils sont localisés spatialement.

Nous proposons donc de minimiser une fonctionnelle variationnelle du type suivant :

$$J_S(X) = \|KX - Y\|_{(L^2)^L}^2 + \sum_{m=1}^M \gamma^m \|X^m\|_{\mathcal{G}^m} \quad (13)$$

$$J_S(X) = \sum_{l=1}^L \|(KX)^l - y^l\|_{L^2}^2 + \sum_{m=1}^M \gamma^m \|X^m\|_{\mathcal{G}^m} \quad (14)$$

où les  $\gamma_m$  sont strictement positifs pour tout  $m$  et  $\mathcal{G}^m$  est un espace fonctionnel choisi de telle manière que les éléments ayant les mêmes propriétés que  $x^m$  ont une petite norme dans cet espace. Par exemple, on choisira un espace de Sobolev pour un signal régulier à variations lentes comme le fond diffus cosmique et un espace de Besov pour un signal localisé avec des transitions rapides comme les amas de galaxies.

#### 2.3.2. Estimation

En toute généralité, nous ne savons pas minimiser une telle fonctionnelle. Mais s'il existe  $M$  bases orthonormées  $\{\phi_\lambda^m\}_{\lambda \in \Lambda}$ , pour  $m = 1, \dots, M$  telles que pour chaque espace fonctionnel  $\mathcal{G}^m$ , il existe des poids strictement positifs  $w_\lambda^m = \{w_\lambda^m\}_{\lambda \in \Lambda}$  et  $p^m$  tels que :

$$\|\cdot\|_{\mathcal{G}^m} = \left( \sum_{\lambda \in \Lambda} w_\lambda^m |\langle \cdot, \phi_\lambda^m \rangle|^{p^m} \right)^{\frac{1}{p^m}}, \quad 1 \leq p^m \leq 2 \quad (15)$$

alors il existe un algorithme itératif permettant de trouver des minimiseurs de l'équation (14). Cet algorithme a été introduit par exemple dans [5] dans le cas d'une composante et d'une



observation. Nous l'avons généralisé au cas de plusieurs composantes et observations [1]. La généralisation étant assez directe, les propriétés de convergence et de régularisation prouvées dans [5] s'en suivent. L'algorithme itératif suivant converge fortement vers un minimiseur de l'équation (14) :

- $X_o$  quelconque.
- $\forall n > 0, \forall m = 1, \dots, M :$

$$X_{n+1}^m = \sum_{\lambda \in \Lambda} S_{\theta, p}^{m, p_m} \left( \langle [cX_n + K^*g - K^*KX_n]^m, \phi_\lambda^m \rangle \right) \phi_\lambda^m, \tag{16}$$

où  $\theta_\lambda^m = \gamma^m w_\lambda^m$ ,  $K^*$  est l'adjoint de  $K$  et  $c$  vérifie  $c > \|K^*K\|$  ( $\|\cdot\|$  est la norme opérateur).

Les fonctions  $S_{\theta, p}$  sont définies par :

$$S_{\theta, p} \left( x + \frac{\theta}{2} \text{sign}(x) |x|^{p-1} \right) = x \tag{17}$$

Nous utilisons en particulier  $p^m = 1$  ou  $p^m = 2$ .

Pour  $p = 1$ ,  $S_{\theta, 1}$  est le seuillage doux de paramètre  $\frac{\theta}{2}$  :

$$S_{\theta, 1}(x) = \begin{cases} x \times \text{sign}(x) \frac{\theta}{2} & \text{si } x \geq \frac{\theta}{2} \\ 0 & \text{sinon} \end{cases} \tag{18}$$

Et pour  $p = 2$ ,  $S_{\theta, 2}$  est l'atténuation :

$$S_{\theta, 2}(x) = \frac{x}{1 + \frac{\theta}{2}} \tag{19}$$

Après initialisation arbitraire, l'algorithme consiste donc à chaque itération à calculer  $H = cX_n + K^*g - K^*KX_n$  (qui est dans  $(L^2)^M$ ). Chacune des  $M$  composantes de  $H$  est alors traitée séparément. On décompose  $H^m$  sur son système  $\{\phi_\lambda^m\}_{\lambda \in \Lambda}$ . La fonction  $S_{\theta, p}^{m, p_m}$  est appliquée à chaque coefficient obtenu. Par exemple, on seuille chaque coefficient si  $p^m = 1$ , ou on atténue chaque coefficient si  $p^m = 2$ . Enfin  $X^m$  est recomposée à partir de ces coefficients modifiés. Soulignons le fait que les systèmes de décomposition  $\{\phi_\lambda^m\}_{\lambda \in \Lambda}$  peuvent être différents pour chaque composante, par exemple, on peut traiter la première composante  $x^1$  sur la base de Fourier mais préférer une base d'ondelettes pour la seconde composante  $x^2$ .

### 2.3.3 Paramètres

Pour cette méthode, les paramètres à définir pour chaque composant  $x^m$  sont :

- La base de décomposition  $\{\phi_\lambda^m\}_{\lambda \in \Lambda}$ .
- L'exposant  $p^m$ , avec  $1 \leq p^m \leq 2$ .
- Les poids  $w^m = \{w_\lambda^m\}_{\lambda \in \Lambda}$ , strictement positifs.
- Le paramètre de régularisation  $\gamma^m$ , strictement positif.

Les trois premiers paramètres définissent l'espace fonctionnel qui est adapté aux propriétés spatiales de  $x^m$  et sont donc à définir en fonction de l'application considérée. Nous les décrivons pour l'application à l'extraction des amas de galaxies dans le

paragraphe 3.2.2. Les coefficients  $\gamma^m$  sont choisis de manière à ce que chacun des termes dans la fonctionnelle (14) aient le même ordre de grandeur.

## 3. Application : reconstruction des amas de galaxies

### 3.1. Données

Les deux méthodes décrites dans la section précédente ont été appliquées à des données simulées qui ont les caractéristiques du projet d'observation du fond diffus cosmique par ACT. Nous considérons le cas de deux composantes :  $x^1$  est le fond diffus cosmique et  $x^2$  les amas de galaxies. Le fond diffus cosmique est simulé par un champ gaussien dont la puissance spectrale est ajustée avec les paramètres de WMAP (c.f. [2]). Les cartes d'amas de galaxies sont obtenues à partir des simulations hydrodynamique de Zhang *et al.* [15]. Ces simulations nous été fournies par Kevin Huffenberger. Nous disposons d'observations à trois fréquences différentes : 145, 217 et 265 GHz. Les noyaux de convolution  $b^l$  sont gaussiens, d'écart-type 1.7, 1.1 et 0.93 minute d'arc. Le bruit est blanc, d'écart type 3.4, 3.63 et  $4.37 \mu K$  par minute d'arc carré. Les simulations sont au nombre de 24 et couvrent chacune une portion de ciel de  $1.19 \times 1.19$  degrés.

Un exemple d'observation à 265 GHz est donné dans la Figure [Fig.1] (haut), ainsi que la contribution du fond diffus cosmique (bas gauche) et des amas de galaxies (bas droite).

### 3.2. Paramètres

#### 3.2.1. Paramètres de la méthode statistique

Le signal venant des amas de galaxies est trop peu puissant par rapport au fond diffus cosmique ou au bruit pour que les matrices de covariance soient estimées précisément à partir des observations. Elles sont calculées à partir de simulations de chaque composante.

Le fond diffus cosmique est bien modélisé par un signal gaussien, donc la probabilité  $p_z$  pour ce signal est fixée à  $\delta_{\{z=1\}}$  pour toute échelle. Pour les amas de galaxies, nous avons comparé quatre possibilités pour  $p_z$  dans [7, 1] et en présentons ici deux : celle dérivée des simulations et  $\delta_{\{z=1\}}$ . Dans le cas où l'on utilise  $p_z = \delta_{\{z=1\}}$ , les voisinages pour les amas de galaxies sont supposés gaussiens. Au contraire, la probabilité  $p_z$  dérivée des simulations induit un modèle non-gaussien.

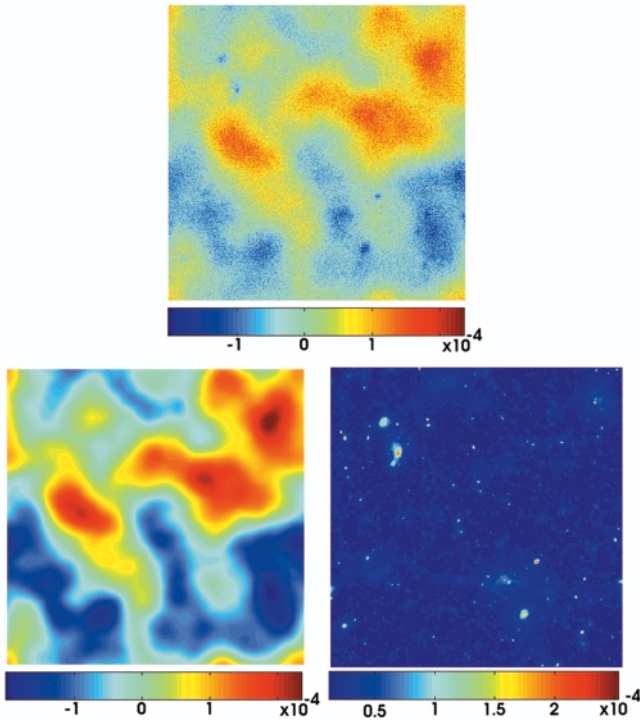


Figure 1. Données : en haut, une observation  $g_i$ , puis la contribution de deux composantes à cette observation : en bas à gauche celle du fond diffus cosmique, en bas à droite, celle des amas de galaxies. (Échelles de couleur en Kelvin.).

### 3.2.2. Paramètres de la méthode variationnelle

Le fond diffus cosmique est pénalisé par une norme de type Sobolev avec des poids ajustés à sa puissance spectrale. Plus précisément, nous utilisons une norme  $l^2$  sur les coefficients d'ondelettes, avec des poids  $w_{j,q}$  ne dépendant que de l'échelle  $j$ , inversement proportionnels à la puissance spectrale à celle échelle. Cela permet de pénaliser les variations rapides et de prendre en compte la puissance spectrale du fond diffus cosmique. Pour les amas de galaxies, nous utilisons une norme  $l^1$  sur les coefficients d'ondelettes (ce qui est équivalent à une norme dans un espace de Besov  $B_1^{1,1}$ ), favorisant la parcimonie de la décomposition du signal en ondelettes et les transitions rapides.

### 3.3. Résultats

Dans cette section, nous comparons les résultats obtenus sur une série de 24 simulations avec les paramètres de ACT. Les trois types de reconstructions considérés sont celles de la méthode variationnelle, celles de la méthode statistique gaussienne avec  $p_z = \delta_{\{z=1\}}$  pour les amas de galaxies, et celles de la méthode statistique non-gaussienne avec  $p_z$  dérivée des simulations pour les amas de galaxies.

Nous avons implémenté chacune des deux méthodes avec deux transformées en ondelettes différentes : la *steerable pyramid* décrite dans [9] et la transformée en ondelettes complexes de [6]. Toutes deux sont des transformées redondantes, qui permettent de décomposer les images en sous-bandes orientées. L'avantage sur une base d'ondelette ordinaire est que l'on obtient une meilleure selectivité directionnelle (le nombre d'orientations peut être choisi), et que l'on se rapproche de l'invariance par translation. Les deux transformées donnent des résultats de qualité égale. Nous donnons ici les résultats avec la *steerable pyramid* et 4 orientations.

#### 3.3.1. Étude qualitative

Nous trouvons que les différentes reconstructions du fond diffus cosmique sont de qualité égale. En revanche, pour les amas de galaxies, les différences sont notables. La figure [FIG. 2] montre que la méthode variationnelle et la méthode statistique non-gaussienne donnent des résultats comparables et que ceux-ci sont nettement meilleurs que les résultats de la méthode statistique gaussienne en terme de reconstruction de la valeur centrale d'amas très intenses. Les différences entre la méthode variationnelle et la méthode statistique non-gaussienne sont plus fines. Nous trouvons que la méthode variationnelle résout les amas plus compacts et permet de reconstruire les amas plus intenses plus précisément que la méthode statistique non-gaussienne. En revanche, les structures filamenteuses sont mieux préservées par cette dernière.

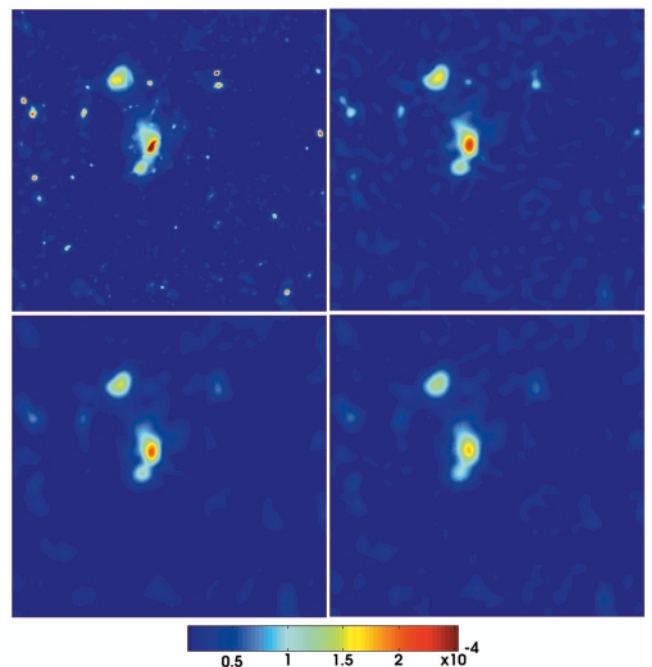


Figure 2. Zoom sur un amas étendu. À haut à gauche l'original, puis 3 reconstructions : En haut à droite : par la méthode variationnelle ; en bas à gauche par la méthode statistique avec  $p_z$  dérivée des simulations ; en bas à droite par la méthode statistique avec  $p_z = \delta_{\{z=1\}}$ . Les cartes ont la même échelle de couleur (en Kelvin).

### 3.3.2. Étude quantitative

Le signal du fond diffus cosmique étant bien modélisé par un processus gaussien isotrope, mesurer l'erreur de reconstruction en terme d'erreur aux moindres carrés entre la reconstruction et l'image simulée est donc approprié et confirme les conclusions qualitatives cités plus haut.

En revanche une carte d'amas de galaxies est un signal avec des objets d'intérêt compacts et brillants (les amas) sur de grandes régions vides. L'erreur de reconstruction définie précédemment n'est plus appropriée car elle donne la même importance à chaque pixel. Pour étudier quantitativement la qualité de nos reconstructions, nous avons donc défini et utilisé des mesures propres à ce signal. Nous procédons à la détection d'objets de type amas (par une méthode de maxima locaux) dans chaque carte d'amas (l'originale et la reconstruite). Après mise en correspondance pour les amas se retrouvant dans les deux cartes, nous pouvons donc établir diverses quantités. Nous choisissons d'abord une quantité observable de référence : l'intensité moyenne de l'amas sur un disque de quelques minutes d'arc, que nous notons  $I$ . Puis nous choisissons un échantillon d'amas originaux : l'ensemble des amas avec  $I > t$  où  $t$  est une valeur seuil. En utilisant les correspondances, nous pouvons donc mesurer :



- la complétude: la proportion d'amas de cet échantillon qui a été reconstruit dans l'estimation (idéalement 1).
- le gain et la déviation au gain : la valeur de gain prédit la véritable valeur de  $I$  d'après celle de l'estimation ( $I_{estimation} = g I_{original}$ ). Idéalement, on souhaite un gain de 1 et une déviation nulle.
- la pureté: la proportion d'amas détectés dans la carte reconstruite qui existent dans l'original (idéalement 1).

Pour les 50 amas les plus intenses dans notre expérience, nous trouvons une pureté et complétude proche 1 pour chaque méthode. Le tableau [TAB. 1] donne les valeurs obtenues pour le gain et sa déviation. Nous retrouvons les conclusions de l'étude qualitative à savoir qu'en terme d'intensité centrale des amas, les méthodes variationnelle ou statistique non-gaussienne sont nettement plus efficaces que l'hypothèse gaussienne, avec un avantage pour la méthode variationnelle.

Cet échantillon de 50 amas dans notre étude correspond à environ 150 amas dans l'expérience ACT, ce qui est un nombre d'amas satisfaisant pour déduire de ces résultats des propriétés astrophysiques.

Tableau 1. Reconstruction des pics d'intensité dans les amas : gain et déviation au gain pour les trois estimations considérées.

Méthode	Statistique		Fonctionnelle
	Gaussienne	Non gaussienne	
Gain	0.69	0.83	0.89
Déviation au gain	0.16	0.11	0.09

## 4. Conclusions

Nous avons présenté deux approches, l'une statistique, l'autre fonctionnelle qui exploitent les propriétés spatiales des composantes  $x^m$  pour résoudre l'équation (1). La méthode statistique repose sur la description par des mélanges de gaussiennes de voisinages de coefficients d'ondelettes des signaux et est adaptée spécialement aux convolutions de mélanges de composantes. La méthode fonctionnelle utilise des espaces  $l^p$  de coefficients sur des systèmes de décompositions qui peuvent être choisis pour chaque composante. Elle est applicable à des cadres plus généraux que l'équation (1).

Nous avons étudié à la fois quantitativement et qualitativement les résultats obtenus par ces deux méthodes pour l'extraction des amas de galaxies par l'effet Sunyaev-Zel'dovich dans le cadre de ACT. Ces études ont montré que les deux méthodes donnent des résultats de qualité suffisante pour des études astrophysiques postérieures. Il est important de noter que les résultats des deux méthodes ont des différences qualitatives. Celles-ci se trouvent accentuées lorsque les conditions d'observations sont moins bonnes ou que d'autres phénomènes astrophysiques sont pris en compte (voir [1] pour plus de détails). Il serait intéressant de développer une approche combinant les qualités des deux méthodes, à savoir l'exploitation de la parcimonie par la méthode variationnelle et la cohérence des voisinages de coefficients d'ondelette de la méthode statistique.

## Références

- [1] ANTHOINE S., «Wavelet-based approaches for the deconvolution of blurred mixtures of components ; Application to astrophysical data», *Ph.D thesis*, Princeton University, 2005.
- [2] BENNETT C. L., HALPERN M., HINSHAW G. *et al.*, «First Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations : Preliminary Maps and Basic Results», *Astrophys.J.Suppl.* Vol.148, #1, 2003.
- [3] CHAMBOLLE A., LIONS P. L., «Image recovery via total variation minimization and related problems», *Numerische Mathematik*, Vol.76, #3, 1997, p.167-188.
- [4] CROUSE M. S., NOWAK R. D., BARANIUK R. G., «Wavelet-based statistical signal processing using hidden Markov models», *IEEE Trans. Signal Processing*, Vol. 46, 1998, p. 886-902.
- [5] DAUBECHIES I., DEFRISE M., DE MOL C., «An iterative thresholding algorithm for linear inverse problems with a sparsity constraint», *Comm. Pure Appl. Math.*, Vol.57, #11, 2004, p. 1413-1457.
- [6] KINGSBURY N.G., «Complex wavelets for shift invariant analysis and filtering of signals», *Journal of Applied and Computational Harmonic Analysis*, vol 10, #3, 2001, p. 234-253.
- [7] PIERPAOLI E., ANTHOINE S., HUFFENBERGER K., DAUBECHIES I., «Reconstructing Sunyaev-Zeldovich clusters in future CMB experiments». *Mon. Not. Roy. Astron. Soc.*, vol.359, #1, 2005, p.261-271.
- [8] PIERPAOLI E., ANTHOINE S., «Finding SZ clusters in the ACBAR maps», *Advances in Space Research*, Vol.36, p.757-761, 2005.



- [9] PORTILLA J., STRELA V., WAINWRIGHT, M., SIMONCELLI E., « Image denoising using a scale mixture of Gaussians in the wavelet domain », *IEEE Trans. Image Proc.*, vol. 12, #11, 2003, p. 1338-1351.
- [10] PORTILLA J., SIMONCELLI E., « Image Restoration using Gaussian Scale Mixtures in the Wavelet Domain », *9<sup>th</sup> IEEE Int'l Conf on Image Processing*, vol. II, 2003, Barcelona, Spain. p. 965-968.
- [11] ROMBERG J., CHOI H., BARANIUK R., « Bayesian Tree-Structured Image Modeling using Wavelet Domain Hidden Markov Models », *IEEE Transactions on Image Processing*, Vol. 10, #7, p. 1056-68, 2001.
- [12] RUDIN L., OSHER S., FATEMI E., « Nonlinear total variation based noise removal algorithms », *Physica D*, Vol.60, 1992, p. 259-268.
- [13] SENDUR L., SELESNICK I. W., « Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency », *IEEE Trans. on Signal Processing*. Vol.50, #11, 2002, p. 2744-2756.
- [14] SHAPIRO I. W., « Embedded image coding using zerotrees of wavelet coefficients », *IEEE Trans. Signal Processing*, vol. 41, 1993, p. 3445-3462.
- [15] ZHANG P, PEN U., WANG B., « The Sunyaev-Zeldovich Effect: Simulations and Observations », *ApJ*, Vol. 577, 2002, p. 555-568.



Sandrine **Anthoine**

Sandrine Anthoine a obtenu un Ph.D en mathématiques appliquées à Princeton University en 2005. Elle est actuellement ATER à l'Université de Provence – Aix-Marseille I. Son domaine de recherche est axé sur les mathématiques appliquées au traitement de l'image.



Elena **Pierpaoli**

Elena Pierpaoli est chercheur en tant que « assistant professor » en cosmologie théorique à l'University of Southern California (États-Unis). Elle a obtenu un master de physique à l'Université de Milan (1994) et un doctorat en astrophysique à SISSA/ISAS (Trieste, Italie, 1998). Elle a été chercheur en tant que « CITA National Fellow » à Vancouver (Canada, 1998-2000) puis « research staff » à Princeton University (2001-2004) et « research faculty » à Caltech (2004-2006).



Ingrid **Daubechies**

Ingrid Daubechies a obtenu une maîtrise (1975) et un doctorat (1980) à l'Université Libre de Bruxelles puis y travaillé comme chercheur jusqu'en 1987. De 1987 à 1994, elle était chercheur (« technical staff ») à AT & T Bell Laboratories. Elle a passé six mois (1990) à l'Université de Michigan et deux ans (1991-1993) à Rutgers University. Elle est actuellement professeur – William R. Keenan Jr. Professor – dans le département de mathématiques et dans le Program in Applied and Computational Mathematics à Princeton University. Son domaine de recherche est axé sur les aspects mathématiques de l'analyse temps-fréquence et ses applications, en particulier les ondelettes.

