Estimation de signaux par noyaux d'ondelettes

Estimating signals using multiple wavelet kernels

Vincent Guigue , Alain Rakotomamonjy et Stéphane Canu
Laboratoire Perception, Systèmes, Information, avenue de l'Université, 76801 St Étienne du Rouvray
Vincent.Guigue@insa-rouen.fr

Manuscrit reçu le 14 octobre 2005

Résumé et mots clés

Cet article présente une méthode de régression pour les signaux non uniformément échantillonnés basée sur les ondelettes. Nous utilisons une formulation issue de l'apprentissage supervisé et des méthodes à noyaux qui combine une fonction coût $\mathcal{L}_2$ et une régularisation $\mathcal{L}_1$ multi-échelles. L'utilisation de l'algorithme *Least Angle Regression* pour la résolution du problème est à la fois efficace et intéressante, elle permet de calculer le chemin complet de régularisation et d'introduire de nouvelles solutions pour régler le compromis biais-variance.

Régularisation $\mathcal{L}_1$, Noyaux multiples, Ondelettes, Régression.

Abstract and key words

This paper addresses the problem of regression in the case of non-uniform sampled signals. Our method is based on supervised learning theory, we propose to use $\mathcal{L}_2$ estimation with wavelet kernels combined with $\mathcal{L}_1$ multiscale regularization. The use of Least Angle Regression as solver enable us to propose new solutions to set the regularization parameter.

Regularization $\mathcal{L}_1$, Multiple Kernels, Wavelets, Regression.

Remerciements

Ce travail est financé en partie par le programme IST de la communauté européenne, avec le réseau d'excellence PASCAL, IST-2002-506778. Cette publication reflète uniquement le point de vue des auteurs.

1. Introduction

Le problème de l'approximation de signaux à partir de données non uniformément échantillonnées apparaît dans plusieurs contextes tels que les systèmes présentant une fréquence d'échantillonnage fluctuante voire aléatoire ou la reconstruction de signaux qui comportent des données manquantes. En traitement d'image, il est également fréquent d'avoir à reconstruire des images dont les échantillons ne sont pas uniformément distribués (*i.e.* données géophysiques, tomographie, etc ...)
Dans cette étude, nous proposons une méthode d'approximation de signaux bruités en se plaçant dans le contexte des espaces de Hilbert à noyau reproduisant et en utilisant comme *a priori* sur la régularité du signal la parcimonie de l'estimateur. Ici, les espaces d'approximation considérés sont des sous-espaces de $\mathcal{L}_2(\mathbb{R})$ générés par un ensemble fini d'ondelettes [RC05].
Plusieurs solutions existent pour faire face à ce type de problème mais elles comportent chacune des inconvénients. Les approches basées sur le *Matching Pursuit* [Mal97] mènent à des résultats sous-optimaux, tandis que le *Basis Pursuit* [CDS98] fait appel à des méthodes *backward* très coûteuses en temps de calcul. Le *wavelet shrinkage* [DJ94] était initialement limité aux signaux uniformément échantillonnés et l'extension aux signaux quelconques nécessite en réalité de ré-échantillonner les signaux [KS00]. Les méthodes existantes utilisant les noyaux d'ondelettes [AAP04] combinent une résolution de type *Expectation-Maximization* (EM) coûteuse avec l'utilisation de plusieurs paramètres de régularisation. Notre but est triple : nous voulons utiliser une formulation multi-noyaux pour faire face efficacement aux signaux présentant des aspects multi-résolutions, nous souhaitons ensuite diminuer la complexité temporelle de la résolution de ce problème et enfin, nous cherchons à éliminer les différents paramètres afin de permettre le traitement automatique de grandes masses de signaux.
Nous nous plaçons dans le contexte d'une estimation de la fonction de régression classique, *i.e* nous disposons d'un ensemble de données bruitées $\{x_j, y_j\}_{j=1\ldots n}$ avec : $y_j = f^\star(x_j) + b_j$ où les b_j sont des variables aléatoires et $x_j \in \mathbb{R}$. Notre objectif est d'obtenir une estimation de $f^\star$ à partir des échantillons bruités y_j. Les x_j ne sont pas équidistants. Nous cherchons la fonction f de l'espace des hypothèses $\mathcal{H}$ qui minimise le risque empirique régularisé :

$$R_{reg}[f] = \frac{1}{n}\sum_{j=1}^{n}(y_j - f(x_j))^2 + \lambda\Omega(\|f\|) \tag{1}$$

où $\Omega(f)$ mesure la régularité de la solution et où λ est le compromis biais-variance. Si $\mathcal{H}$ est un espace de Hilbert à noyau reproduisant alors ce problème d'approximation régularisée d'une fonction non uniformément échantillonnée rentre dans le cadre du théorème du représentant de Kimerdolf et Wahba [KW71] et la forme de la solution est donc :

$$f(x) = \sum_{j=1}^{n}\beta_j K(x_j, x) \tag{2}$$

où K est le noyau reproduisant de l'espace $\mathcal{H}$ et où les β_i sont les coefficients de régression. Nous avons ensuite étendu ce schéma aux noyaux multiples en utilisant la stratégie de Vincent et Bengio [VB02]. La solution est alors de la forme :

$$f(x) = \sum_{i=1}^{N}\sum_{j=1}^{n}\beta_{ij} K_i(x_j, x) \tag{3}$$

où N désigne le nombre de noyaux considérés.
L'utilisation de noyaux d'ondelettes pour la régression, combinée à une régularisation $\Omega(\|f\|)$ a déjà fait l'objet d'études [AAP04, RC05]. Nous introduisons dans cet article une formulation originale, qui pénalise la somme des valeurs absolues des coefficients β_i :

$$\Omega(\|f\|) = \sum_i |\beta_i| \tag{4}$$

Cette formulation est appelée LASSO (*Least Absolute Shrinkage and Selection Operator*) [Tib96]. Nous utilisons l'algorithme du LARS (*Least Angle Regression Stepwise*) [EHJT04] pour la résolution. La régularisation $\mathcal{L}_1$ est un gage de parcimonie tandis que le calcul du chemin complet de régularisation, *via* le LARS, permet de définir de nouveaux critères pour trouver le compromis biais-variance optimal. Cette démarche permet d'améliorer les performances et le temps de calcul pour le LASSO tout en proposant des critères auto adaptatifs pour le réglage du compromis biais-variance, afin de rendre la méthode complètement non paramétrique.
L'utilisation combinée de noyaux multiples, d'une formulation LASSO et d'une résolution itérative LARS aboutit au cadre du *Kernel Basis Pursuit* [GRC05]. Le problème à résoudre se pose comme suit :

$$\min_{\beta}\sum_{j=1}^{n}\left(y_j - \sum_{i=1}^{N}\sum_{k=1}^{n}\beta_{ik}K_i(x_k, x_j)\right)^2 \quad \sum_{i,k}|\beta_{ik}| \le t \tag{5}$$

Nous détaillons la méthode utilisée en section 2 en étudiant successivement la construction des noyaux d'ondelettes, le passage aux noyaux multiples et la régularisation multi-échelles *via* l'algorithme du LARS. La section 3 décrit le réglage du compromis biais-variance t de l'équation (5) en utilisant la propriété de calcul du chemin complet de régularisation. Les résultats sont présentés en section 4, ils montrent l'intérêt d'une telle démarche pour l'estimation de signaux présentant des aspects multi-résolutions et des ruptures mais aussi pour des signaux plus classiques. Par rapport à des SVM [SS02] optimisés par validation croisée, les performances de reconstruction sont meilleures dans la plupart des cas et la parcimonie des solutions est systématiquement améliorée tout en réduisant le nombre de paramètre pour l'utilisateur.

2. Méthode

Nous détaillons dans cette section la constuction des noyaux d'ondelettes, le choix de notre stratégie de régularisation et l'algorithme de résolution utilisé dans cet article. Les noyaux d'ondelettes permettent de faire le lien entre les méthodes à noyaux qui sont réputées pour leur performance et leur souplesse d'utilisation[1] et l'analyse multi-résolutions qui tire partie des propriétés des ondelettes pour débruiter, approximer et représenter les signaux non stationnaires. Nous avons ensuite opté pour une régularisation $\mathcal{L}_1$ pour des raisons de parcimonie, de qualité et d'efficacité. Au niveau de la méthode de résolution, nous faisons appel à une version modifiée de l'algorithme *Least Angle Regession* pour sa rapidité et le calcul du chemin complet de régularisation.

2.1. Noyaux d'ondelettes

Pour pouvoir mettre en œuvre le théorème de la représentation dans le cadre de l'estimation par ondelettes, nous nous sommes interéssés à la question des conditions nécessaires et suffisantes pour construire un espace de Hilbert à noyau reproduisant à partir d'une famille quelconque de fonctions. Pour répondre à ce point, nous nous sommes placés dans le cadre théorique des familles génératrices et redondantes d'un espace de Hilbert.

Définition 2.1.1. *Un ensemble de vecteurs* $\{\phi_n\}_{n\in\Gamma}$ *est une structure oblique ou* frame *d'un espace de Hilbert* $\mathcal{H}$ *si il existe deux constantes* $A > 0$ *et* $\infty > B \geq A > 0$ *telles que :*

$$\forall f \in \mathcal{H}, \qquad A||f||^2_{\mathcal{H}} \leq \sum_{n\in\Gamma} |\langle f,\phi_n\rangle_{\mathcal{H}}|^2 \leq B||f||^2_{\mathcal{H}}$$

Cette frame *est dit «* frame tendue *» si A et B sont égaux.*

Une *frame* d'un espace de Hilbert est associée à une *frame* duale $\{\bar{\phi}_n\}_{n\in\Gamma}$ qui est telle que :

$$f = \sum_{n\in\Gamma} \langle f,\bar{\phi}_n\rangle_{\mathcal{H}}\phi_n = \sum_{n\in\Gamma} \langle f,\phi_n\rangle_{\mathcal{H}}\bar{\phi}_n \tag{6}$$

Dans le cas d'une *frame* tendue, la relation suivante est également vérifiée :

$$\bar{\phi}_n = \frac{1}{A}\phi_n$$

Comme nous le constatons à travers ces définitions et ces propriétés, une structure oblique permet la représentation des éléments d'un espace de Hilbert $\mathcal{H}$ à travers une famille de vecteurs qui peut être redondante. En ce sens, l'utilisation d'une *frame* permet une plus grande flexibilité pour décrire des espaces de Hilbert.

1. Choix de la fonction coût, du type de régularisation, traitement des signaux non uniformément...

Ainsi, nous avons cherché à savoir dans quel cas, un espace de Hilbert décrit par une *frame* est en fait un espace à noyau reproduisant. Nous avons également montré [RC05] que :

Théorème 2.1.1. *Soit un espace de Hilbert* $\mathcal{H}$ *de fonctions de* $\mathbb{R}^\Omega$ *muni du produit scalaire* $\langle\cdot,\cdot\rangle$ *et* $\{\phi_n\}_{n\in\Gamma}$ *une* frame *de cette espace alors* $\mathcal{H}$ *est un espace de Hilbert à noyau reproduisant si et seulement si :*

$$\forall x \in \Omega, \qquad \left\| \sum_{n\in\Gamma} \bar{\phi}_n(\cdot)\phi_n(x) \right\|_{\mathcal{H}} < \infty \tag{7}$$

Ce premier théorème n'est toujours pas constructif mais permet néanmoins de donner quelques conditions sur la structure oblique pour obtenir un espace de Hilbert à noyau reproduisant. À partir de ce théorème, il devient simple de montrer qu'un ensemble fini de fonctions bornées $\{\phi_n\}_{n\in\Gamma}$ définies sur Ω appartenant à un espace de Hilbert engendre un espace de Hilbert à noyau reproduisant dont le noyau s'écrit :

$$\forall x,x' \in \Omega, \quad K(x,x') = \sum_{n\in\Gamma} \bar{\phi}_n(x)\phi_n(x') \tag{8}$$

À partir de cette propriété, il est donc facile de créer un noyau défini positif associé à l'espace d'hypothèses engendré par les fonctions $\{\phi_n\}_{n\in\Gamma}$. L'avantage d'un tel noyau vient de l'introduction de connaissances *a priori* dans le noyau à travers les fonctions génératrices de l'espace d'hypothèses. Par exemple, la figure 1 montre deux noyaux des espaces engendrés respectivement par les familles :

$$\left\{\phi_n(x) = \frac{\sin(\pi(x-n))}{\pi(x-n)}\right\}_{n=1}^{N} \quad \text{et} \quad \left\{\phi_n(x) = \frac{1}{\sqrt{2^j}}\psi(2^j x - n)\right\}_{n=1}^{N} \tag{9}$$

où ψ est une ondelette de type « Symmlet » [Mal97]. Comme on peut le constater la représentation du noyau et donc la mesure de similarité dépendent des fonctions génératrices.
Bien que dans le cas pratique on se limite toujours à un ensemble fini de fonctions génératrices pour construire le noyau, nous avons également examiné les conditions pour qu'un ensemble infini de fonctions génère un espace de Hilbert à noyau reproduisant. Nous avons montré que, sous réserve de certaines conditions faibles, une famille infinie de fonctions $\{\phi_n\}_{n\in\Gamma}$, munie d'un produit scalaire approprié définit un EHNR dont le noyau s'écrit [RC05] :

$$K(x,x') = \sum_n \bar{\phi}_n(x)\phi_n(x') = \sum_n \lambda_n\phi_n(x)\phi_n(x') \tag{10}$$

À partir de ce cadre, il est possible de montrer, par exemple, que l'espace engendré par une famille de dimension infinie d'ondelettes orthogonales à support compact, muni d'un produit scalaire approprié est un espace de Hilbert à noyau reproduisant. Dans ce cas, la famille $\phi_n(t)$ est en fait constituée d'une famille

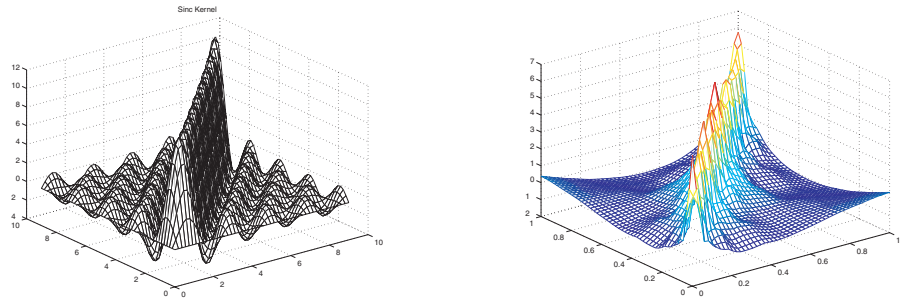


Figure 1. Noyaux sinc et noyaux d'ondelettes.

d'ondelettes $\{\psi_{j,k}\}$ et le noyau devient donc :

$$K(x,x') = \sum_{j,k} \alpha_{j,k} \psi_{j,k}(x) \psi_{j,k}(x') \tag{11}$$

Par ailleurs, il est possible de décomposer ce noyau K comme étant la somme de plusieurs noyaux K_j d'espaces de Hilbert associée à une échelle j donnée de la famille d'ondelettes :

$$K_j(x,x') = \sum_{k} \alpha_{j,k} \psi_{j,k}(x) \psi_{j,k}(x') \tag{12}$$

Cette approche permet donc de générer des noyaux multi-échelles voire des noyaux multiéchelles localisés dans le temps, en ne considérant qu'une famille d'ondelettes d'une échelle j et de translation $k \in K \subset \mathbb{Z}$.

2.2. Utilisation de noyaux multiples

Indépendamment des algorithmes (régression à vecteur support [SS02] ou réseau de régularisation [EPP00] qui se différencient par leurs fonctions coût[2]), le choix du noyau pose des difficultés lorsque la fonction à estimer présente une structure multi-échelles. En effet, dans ce cas, l'utilisation d'un noyau unique aura tendance, à la fois, à sur-apprendre et sous-apprendre les données, car il ne pourra s'adapter à toutes les échelles du signal. La solution consiste à estimer la fonction sur un ensemble d'espace de Hilbert $\{\mathcal{H}_i\}_{i=1,\ldots,N}$ de noyaux reproduisants respectifs $\{K_i\}_{i=1,\ldots,N}$ afin de modéliser toutes les échelles.

En reprenant la non linéarisation de Vincent et Bengio [VB02], nous considérons chaque $K_i(x,.)$ comme une source d'information. Nous construisons donc un noyau multi-échelles comme la concaténation de différents noyaux élémentaires :

$$K = [\, K_1 \ldots K_i \ldots K_N \,] \tag{13}$$

Chaque source d'information $K_i(x_j,\cdot)$ est caractérisée par un point x_j (référence temporelle) et un facteur d'échelle i. La fonction de régression s'écrit alors :

$$f(x) = \sum_{i=1}^{N} \sum_{j=1}^{n} \beta_{ij} K_i(x_j,x), \quad f \in \mathcal{H}_1 + \ldots + \mathcal{H}_N \tag{14}$$

Comme nous verrons en section 2.4, les sources d'informations sont sélectionnées sur le critère de corrélation avec le résidu à chaque itération. Afin de conserver un processus de sélection équitable entre les sources issues des différentes échelles, nous les normalisons de sorte que leur écart-type soit unitaire.

2.3. Régularisation multi-échelles et LARS

La régularisation est souvent associée à la parcimonie de la solution. La meilleure manière de l'améliorer est de pénaliser la pseudo norme $\Omega(f) = ||\beta||_{\mathcal{L}_0}$, c'est-à-dire de compter le nombre de coefficients β non nuls dans l'équation (5). Cependant, ce problème est NP-difficile et les solutions classiques consistent à utiliser soit la norme $\mathcal{L}_2$, soit la norme $\mathcal{L}_1$ en approximant parcimonieusement la solution.

2.3.1. Régularisation $\mathcal{L}_2$ et $\mathcal{L}_1$

Plusieurs approches appliquent une régularisation $\mathcal{L}_2$ sur des noyaux multiples. Les solutions dérivées du *Matching Pursuit*, comme [VB02], sont très efficaces mais sous-optimales car ce sont des stratégies gloutonnes. L'algorithme du *Back-Fitting* [HT90] permet de trouver la solution exacte du problème régularisé $\mathcal{L}_2$ mais le coût de calcul est important et la formulation utilisée multiplie les paramètres de compromis (un par noyau) :

$$\min_{\beta} \sum_{i=1}^{n} \|y_i - \sum_{j=1}^{N} f_j(x_i)\|^2 + \lambda_1 \|f_1\|^2_{\mathcal{L}_2} + \ldots + \lambda_N \|f_N\|^2_{\mathcal{L}_2} \qquad f_j(x) = \sum_{k=1}^{n} \beta_{jk} K_j(x_k,x) \tag{15}$$

Parmi les stratégies parcimonieuses, basées sur une régularisation $\mathcal{L}_1$, nous nous focaliserons sur la formulation du LASSO

2. Les machines à vecteur support utilisent la fonction coût de hinge alors que les réseaux de régularisation utilisent les moindres carrés.

[Tib96] qui combine une fonction coût des moindres carrés et une pénalisation de la norme[3] $\mathcal{L}_1$ de l'estimateur. L'intérêt de cette formulation est notamment discutée dans [DDM04] et [Ng04]. Chen a proposé l'algorithme du *Basis Pursuit* [CDS98] sur ce schéma et Grandvalet a montré que l'*Adaptive Ridge Regression* (ARR, $\Omega(f) = \sum_{ij} \lambda_i \beta_{ij}^2$) est équivalente au LASSO [Gra98].

La plupart des problèmes d'apprentissage sont régularisés en $\mathcal{L}_2$ et il existe beaucoup d'outils pour les traiter. À l'inverse, la formulation $\mathcal{L}_1$ demandait jusqu'à maintenant des résolutions coûteuses, basées sur la programmation linéaire [CDS98] ou sur des versions améliorées de l'algorithme EM (*Expectation-Maximization*) [Gra98]. L'algorithme *Least Angle Regression Stepwise* (LARS) [EHJT04] offre de nouvelles perspectives pour ce genre de problème. La rapidité du LARS vient de la combinaison entre parcimonie et méthode pas à pas *forward*. Les premières itérations sont peu coûteuses, elles nécessitent la résolution de système linéaires de faible dimension. Le coût augmente au fil des itérations, mais la régularisation $\mathcal{L}_1$, gage de parcimonie, limite le nombre d'itérations nécessaires. À l'inverse, les méthodes *backward* [Gra98, CDS98] nécessitent la résolution d'un système linéaire de grande dimension pour réduire les coefficients inutiles à zéro. Étant donné qu'une seule source d'information est ajoutée (ou retirée) à chaque itération, il est possible de mettre à jour la solution au lieu de la recalculer entièrement à la manière de la formulation simple-SVM [LCV+04]. Finalement, la formulation du LARS est la suivante :

$$\min_{\beta} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{N} \sum_{k=1}^{n} \beta_{jk} K_j(x_k, x_i) \right)^2 \quad \sum_{j,k} |\beta_{jk}| \leq t \tag{16}$$

Cette formulation est équivalente à (1) avec $\Omega(\|f\|) = \|\beta\|_{\mathcal{L}_1}$.

2.4. Fonctionnement de *Least Angle Regression Stepwise*

L'intérêt de l'algorithme du LARS est de proposer un ensemble de solutions pour ce problème, correspondant à toutes les valeurs possibles de t dans (16). La solution peut ainsi être choisie *a posteriori*, parmi un ensemble de solutions optimales. Nous expliquons ici le fonctionnement de l'algorithme de [EHJT04], plus de détails sont disponibles dans [Gui05].

Soit la base d'apprentissage matricielle X constituée des $n \times N$ sources d'information vectorielles $X_i \in \mathbb{R}^n$ décrites en section 2.2 :

$$X = [\, X_1 \quad \cdots \quad X_i \quad \cdots \quad X_d \,] \in \mathbb{R}^{n \times nN} \tag{17}$$

La fonction coût $\mathcal{L}$, sous sa forme vectorielle s'écrit :

$$\mathcal{L} = \|y - X\beta\|^2, \quad y \in \mathcal{Y}^n, X \in \mathbb{R}^{n \times nN}, \beta \in \mathbb{R}^{nN} \tag{18}$$

La méthode de résolution la plus classique pour ce genre de problème convexe consiste à chercher la solution de :

$$\frac{\partial \mathcal{L}}{\partial \beta} = -2X^\top (y - X\beta) = 0 \tag{19}$$

$$\Longleftrightarrow X^\top (y - X\beta) = 0 \tag{20}$$

qui aboutit à la solution optimale au sens des moindres carrés. Afin de chercher tous les compromis au sens du paramètre t, nous allons résoudre ce problème de manière itérative, en partant d'un ensemble vide de sources puis en ajoutant successivement les sources permettant d'améliorer le plus la solution.

2.4.1. Notations et hypothèses

Notons $R = y - X\beta = y - f(x)$ le résidu engendré par la solution f. Dans ce cas, le terme $X^\top (y - X\beta)$ de l'équation (20) désigne simplement la corrélation entre les sources et le résidu. Afin que les corrélations des différentes sources au résidu puissent être comparées, il est nécessaire que les sources soient normalisées, c'est-à-dire de moyennes nulles et de variances unitaires. Nous utiliserons les exposants pour désigner l'itération où nous nous trouvons. Par exemple, à l'itération i, nous avons : $f^{(i)}, R^{(i)}, \beta^{(i)}$.

Le LARS est un algorithme itératif procédant par sélection de sources d'information : nous noterons $\mathcal{A}$ l'ensemble des sources sélectionnées, dit ensemble actif. $\bar{\mathcal{A}}$ est le complémentaire de $\mathcal{A}$.

2.4.2. Résolution itérative

1. Au départ, nous avons un ensemble de sources actives $\mathcal{A}$ qui est vide :

$$\mathcal{A} = \emptyset, \quad \bar{\mathcal{A}} = \{1, \ldots, d\}, \quad \forall i, \beta_i^{(0)} = 0, R^{(0)} = y \tag{21}$$

2. Nous sélectionnons la source qui viole le plus la contrainte (20), c'est-à-dire celle qui permet de faire baisser le plus le résidu, à valeur de β_i constante[4]. Il s'agit également de la source la plus corrélée avec le résidu.

$$\mathcal{A} \leftarrow \{\mathcal{A}, i\}, \text{ avec : } i \text{ tel que : } \max_{i \in \bar{\mathcal{A}}} |X_i (y - X\beta)| \tag{22}$$

$$\bar{\mathcal{A}} \leftarrow \bar{\mathcal{A}} - i \tag{23}$$

Cette opération est illustrée en figure 2(b).

3. À l'intérieur des sources actives, nous cherchons le vecteur $u_{\mathcal{A}} \in \mathbb{R}^n$ suivant lequel va évoluer la solution f :

$$f^{(k+1)} = f^{(k)} + \gamma u_{\mathcal{A}} \in \mathbb{R}^n \tag{24}$$

3. Le noyau ne forme pas une famille de fonctions orthogonales, la somme des valeurs absolues des coefficients β_i n'est donc pas une norme mais une approximation de la norme $\mathcal{L}_1$. Des études empiriques [DDM04, Ng04] ont néanmoins montré l'intérêt de cette formulation.

4. C'est ce point précis qui montre que nous sommes en train de minimiser la norme $\mathcal{L}_1$ de f.

Ce vecteur représente la projection du résidu R dans l'espace engendré par $X_{\mathcal{A}}$.
Listons maintenant les propriétés liées à $u_{\mathcal{A}}$:

- $u_{\mathcal{A}}$ s'écrit comme une combinaison linéaire des sources de l'ensemble actif puisqu'il s'agit d'une projection du résidu sur ces sources :

$$u_{\mathcal{A}} = \sum_{i=1}^{|\mathcal{A}|} \omega_i X_i = X_{\mathcal{A}}\omega \in \mathbb{R}^n, \quad \omega \in \mathbb{R}^{|\mathcal{A}|} \tag{25}$$

- $u_{\mathcal{A}}$ est equi-corrélé à toutes les sources de l'ensemble actif :

$$\forall i \in \mathcal{A}, X_i^\top u_{\mathcal{A}} = \alpha, \quad \alpha \in \mathbb{R} \tag{26}$$

où α est une constante.

En écriture vectorielle, nous avons finalement :

$$X_{\mathcal{A}}^\top u_{\mathcal{A}} = \alpha \mathbb{1}^{|\mathcal{A}|} \in \mathbb{R}^{|\mathcal{A}|} \tag{27}$$

En remplaçant maintenant $u_{\mathcal{A}}$ par l'expression de (25) :

$$X_{\mathcal{A}}^\top X_{\mathcal{A}}\omega = \alpha \mathbb{1}^{|\mathcal{A}|} \tag{28}$$

$$\omega = \left(X_{\mathcal{A}}^\top X_{\mathcal{A}}\right)^{-1} \alpha \mathbb{1}^{|\mathcal{A}|} \tag{29}$$

D'où :

$$u_{\mathcal{A}} = X_{\mathcal{A}} \left(X_{\mathcal{A}}^\top X_{\mathcal{A}}\right)^{-1} \alpha \mathbb{1}^{|\mathcal{A}|} \tag{30}$$

Or $u_{\mathcal{A}}$ est un vecteur unitaire, donc :

$$\|u_{\mathcal{A}}\| = 1 \Longleftrightarrow \alpha = \frac{1}{\sqrt{\mathbb{1}^{|\mathcal{A}|}(X_{\mathcal{A}}^\top X_{\mathcal{A}})^{-1}\mathbb{1}^{|\mathcal{A}|}}} \tag{31}$$

4. Une fois établie la direction dans laquelle va évoluer la solution, reste le calcul du pas γ pour définir $f^{(k+1)}$. Nous déterminons le plus petit pas, de sorte qu'une (et une seule) source de $\bar{\mathcal{A}}$ soit corrélée au résidu autant que les sources de $\mathcal{A}$:

$$\gamma = \min_{i \in \bar{\mathcal{A}}, \gamma > 0} \left\{ \frac{X_j^\top R^{(k)} - X_i^\top R^{(k)}}{X_i^\top u_{\mathcal{A}} - X_j^\top u_{\mathcal{A}}}, \frac{X_i^\top R^{(k)} + X_j^\top R^{(k)}}{X_j^\top u_{\mathcal{A}} + X_i^\top u_{\mathcal{A}}} \right\}, \quad j \in \mathcal{A} \tag{32}$$

5. Calcul de la nouvelle solution suivante :

$$f^{(k+1)} = f^{(k)} + \gamma u_{\mathcal{A}} \in \mathbb{R}^n \tag{33}$$

Identification des β associés à cette solution en utilisant (25) :

$$f^{(k+1)} = X\beta^{(k+1)} = X\beta^{(k)} + \gamma u_{\mathcal{A}} = X\beta^{(k)} + \gamma X_{\mathcal{A}}\omega \tag{34}$$

d'où une mise à jour des $\beta_{\mathcal{A}}$ correspondant aux variables actives :

$$\beta_{\mathcal{A}}^{(k+1)} = \beta_{\mathcal{A}}^{(k)} + \gamma\omega \in \mathbb{R}^{|\mathcal{A}|} \tag{35}$$

6. Lorsqu'un coefficient β_i change de signe entre les $\beta^{(k)}$ et les $\beta^{(k+1)}$, cela signifie que la corrélation entre une des variables et le résidu s'est annulée puis a changé de signe. Afin de trouver la solution du LASSO, il est nécessaire de déterminer le pas γ correspondant à l'annulation du coefficient β concerné pour retirer la variable qui lui est liée de l'ensemble actif.
Nous déterminons alors un pas γ tel que $\beta_i^{(k+1)} = 0$, c'est-à-dire :

$$\gamma = \frac{\beta_i^{(k+1)} - \beta_i^{(k)}}{\omega_i} = \frac{-\beta_i^{(k)}}{\omega_i} \tag{36}$$

La variable i est éliminée de l'ensemble actif :

$$\mathcal{A} \leftarrow \mathcal{A} - i, \quad \bar{\mathcal{A}} \leftarrow \bar{\mathcal{A}} + i \tag{37}$$

Et une nouvelle optimisation est effectuée : détermination de $u_{\mathcal{A}}$, calcul de γ, mise à jour de la solution f.

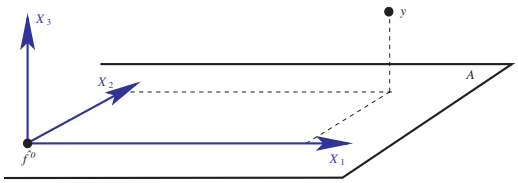
7. Le processus est itéré jusqu'à ce que la borne t de l'équation (16) soit atteinte, toutes les étapes correspondant à des valeur de t inférieure sont décrite en utilisant cet algorithme de résolution.

2.4.3. Modification du LARS

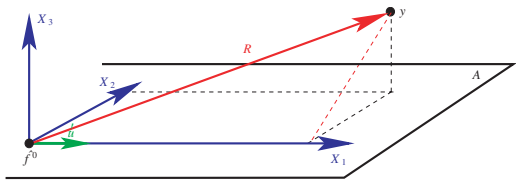
Afin d'améliorer la parcimonie et les résultats en régression de la méthode, nous avons introduit une modification dans l'algorithme du LARS. Lors de la dernière itération de la méthode, le pas est calculé suivant la méthode des moindres carrés classiques (*cf.* figure 3). Cela revient en réalité à utiliser le LARS comme une méthode de sélection de sources d'information puis à faire une projection sur cette famille de fonctions. Dans la pratique, seul le calcul du dernier pas γ change et l'efficacité de la méthode n'est pas altérée.

3. Réglage du compromis biais-variance

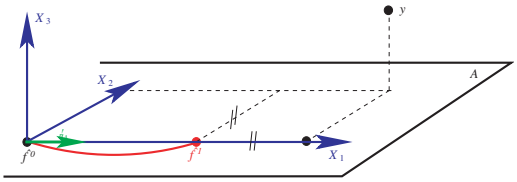
Le compromis biais-variance permet de stopper la phase d'apprentissage avant d'intégrer tous les β dans la construction de la solution. L'intérêt est double : d'une part, il s'agit d'éviter les phénomènes de sur-apprentissage qui surviennent lorsque la solution est trop complexe, d'autre part, cela nous permet de réduire le temps de calcul nécessaire pour obtenir la solution optimale. Le réglage du paramètre de compromis t dans l'équation (16) constitue donc un enjeu essentiel.
Pour faire face à ce problème nous utilisons une propriété du LARS : le calcul du chemin complet de régularisation. Chacune des itérations du LARS aboutit à une solution optimale au sens d'un paramètre t fixé. Il est donc possible de régler le compromis biais-variance dynamiquement [BTJ041], en étudiant l'évolution de la solution à chaque itération.



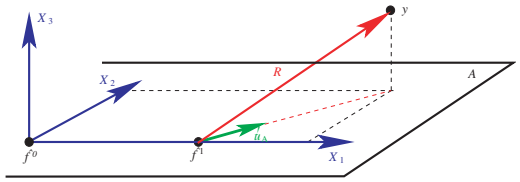
(a) Espace des variables ou sources d'information. Les signaux sont ici décrits dans un espace à 3 dimensions ($d = 3$). Cet exemple illustre le cas très particulier d'un espace orthogonal.



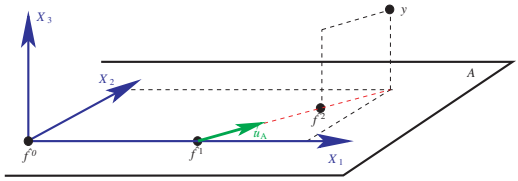
(b) Les projections successives du résidu sur les variables X_1, X_2 etX_3 permettent d'isoler la variable la plus corrélée au résidu. Elle est ajoutée à l'ensemble des variables actives A. Définition de la direction u_A dans l'espace engendré par les variables de A.



(c) Définition d'une nouvelle solution en faisant un pas dans la direction u_A. Le pas est défini de façon à établir une équi-corrélation entre le résidu, les variables de l'ensemble actif et une nouvelle variable.



(d) L'opération est réitérée, le résidu projeté sur les variables de l'ensemble actif, pour définir u_A. La figure met en évidence le fait que tous les angles entre u_A et les variables actives sont égaux. Toutes les variables actives sont donc continuellement équi-corrélées au résidu.



(e) La définition du pas γ va jusqu'à equi-corrélation.

Figure 2. Explication graphique du fonctionnement de l'algorithme Least Angle Regression Stepwise *(LARS). Les sources d'information sont des variables qui forment un espace dans lequel nous évoluons pas à pas pour construire linéairement une approximation des étiquettes y.*

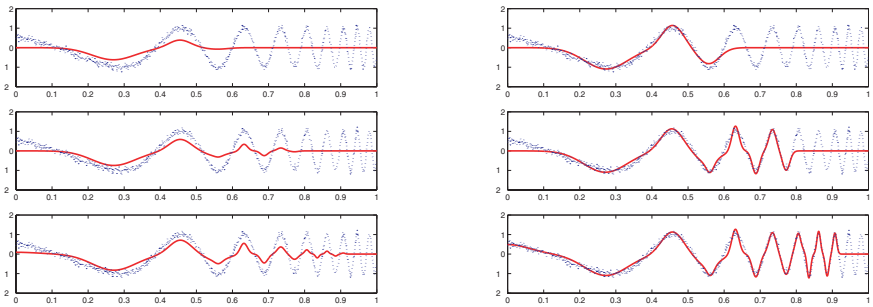


Figure 3. Illustration de la modification de l'algorithme Least Angle Regression Stepwise *à trois instants distincts lors de l'estimation de la fonction* $\cos(\exp(\omega t))$. *Version originale à gauche, version modifée à droite.*

3.1. Autres définitions de seuil

Le critère le plus classique pour régulariser le LARS consiste à fixer une borne sur la somme des $|\beta_i|$. Cependant, ce critère est particulièrement abstrait et difficile à régler. Nous proposons donc d'utiliser d'autres définitions de seuil pour stopper la phase d'apprentissage :

- le ν-LARS, où ν est un seuil sur le pourcentage de points pouvant être sélectionnés comme supports ($\beta_i \neq 0$) de la solution,
- le critère de Ljung [Lju87], où un seuil est fixé sur l'auto-corrélation du résidu qui mesure sa ressemblance à un bruit blanc.

Le ν-LARS permet à l'utilisateur de régler le paramètre de compromis plus facilement, en utilisant sa connaissance *a priori* sur la parcimonie de la solution recherchée. À l'inverse, nous avons abandonné la formulation basée sur le critère de Ljung rapidement : l'utilisateur n'a pas plus d'*a priori* pour une borne sur ce critère que pour la borne t sur la somme des valeurs absolues des β.

3.2. Critères auto-adaptatifs

L'idée est d'ajouter des sources d'information piégées, que nous ne souhaitons pas utiliser pour construire la solution. Lorsque ces sources sont sélectionnées par le LARS, nous arrêtons la procédure d'apprentissage. Cette formulation permet d'éliminer la notion de borne et donc de paramètre à régler. Nous proposons deux heuristiques pour construire ces sources pièges :

LARS-VA Nous introduisons des vecteurs de variables aléatoires (VA) gaussiennes parmi les sources d'informations à la manière de [BBE+03]. Lorsque ces sources seront sélectionnées, cela signifie qu'elles sont plus corrélées au résidu que n'importe quelle autre source d'information. Le résidu peut alors être assimilé à du bruit et l'apprentissage stoppé.

LARS-HF Comme le montre la figure 4, les régions basses fréquences du signal sont expliquées en premier. Lorsque nous commençons à utiliser le noyau hautes fréquences (HF), les phénomènes de plus basses fréquences ont déjà été expliqués. Dans le noyau HF, les points d'apprentissage n'ont d'influence que dans une très petite sphère, c'est-à-dire essentiellement sur eux-mêmes. La sélection d'une source d'information HF peut donc être interprétée comme une tentative d'expliquer l'amplitude d'un unique point du signal original. Dans la plupart des cas, nous commençons ainsi à expliquer le bruit contenu dans le signal, c'est-à-dire à faire du sur-apprentissage : nous arrêtons donc la phase d'apprentissage.
Cette dernière famille de solution est particulièrement intéressante puisqu'elle permet d'avoir un algorithme non-paramétrique. Pour améliorer la robustesse de la méthode en évitant les cas particuliers défavorables, nous attendons d'avoir sélectionné trois sources piégées avant de stopper la phase d'apprentissage.

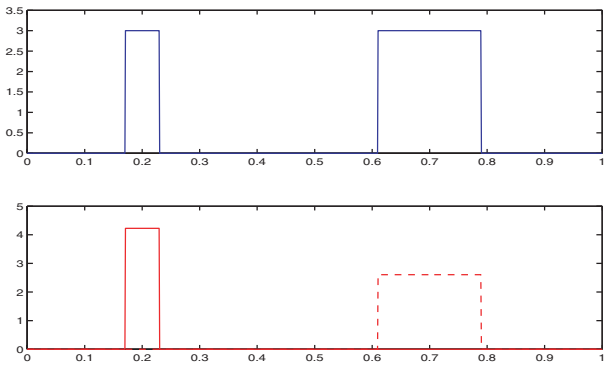


Figure 4. Illustration du mécanisme de sélection des sources d'information. Le signal S à décrire est en haut, de couleur bleu. Le dictionnaire est composé de deux fonctions analysantes g_1 et g_2 (respectivement en trait plein et pointillé sur la figure du bas), parfaitement adapté au signal à décrire et normalisé pour avoir un écart-type unitaire. La corrélation entre la première fonction et le signal est : $g_1^\top S = 773.14$. La corrélation entre la deuxième fonction et le signal est : $g_2^\top S = 1437.6$. La partie de forte énergie du signal sera donc expliquée en premier.

4. Résultats

Nous avons comparé le LARS, avec plusieurs critères d'arrêt, par rapport aux ε-SVM et au *back-fitting* [HT86] pour construire une estimation des signaux artificiels présentés figure 6. il s'agit de signaux classiques, utilisés par Donoho et Johnstone [DJ94]. Nous les avons regoupés suivant leurs caractéristiques dominantes, afin de mettre en évidence les forces et les faiblesses de notre approche sur différents types de signaux. L'échantillonnage est aléatoire sur l'intervalle [0, 1] et les données sont bruitées. Nous avons utilisé un signal de 400 points pour l'apprentissage et une base de test (non bruitée) de 1000 points.
Le critère d'évaluation est l'erreur moyenne au sens des moindres carrés. Les bornes du LARS ($\sum_i |\beta_i|$, ν) sont optimisées par validation croisée. Les paramètres des SVM (ε, C et le choix du noyau) sont optimisés par triple validation croisée sur l'ensemble d'apprentissage, chaque échelle est donc traitée séparément. L'algorithme du *back-fitting* nécessite le réglage d'autant de paramètres qu'il y a de noyaux. Ici encore, nous avons utilisé la validation croisée pour estimer les compromis optimaux. Nous avons utilisé 10 échelles d'ondelettes de type « Symmlet » (4 moments) [Mal97], correspondant aux paramètres de dilatation des ondelettes de 0 à 9. Les résultats sont présentés dans le tableau 1.
Les méthodes LARS, basées sur les noyaux d'ondelettes multi-résolutions donnent les meilleurs résultats sur les bases de données qui présentent des aspects multi-résolutions comme $\cos(\exp(\omega t))$, Doppler ou blocks. L'avantage est particulièrement significatif par rapport au SVM (entre 21 % et 48 %), la formulation multi-noyaux permettant de *respecter* les différentes fréquences lors de la reconstruction du signal (figure 5). La régularisation utilisée participe également à la qualité de la solution comme le montre l'écart de performance entre le *back-fitting* et le LARS. Le phénomène est similaire sur les signaux qui comportent des ruptures, grâce à l'utilisation des échelles haute-fréquence pour marquer ces ruptures. À l'inverse, les SVM se montrent très efficaces sur les bases de données régulières au sens de l'erreur des moindres carrés (signal HeavySine). L'ε-tube autour de la solution est bien adapté à ce

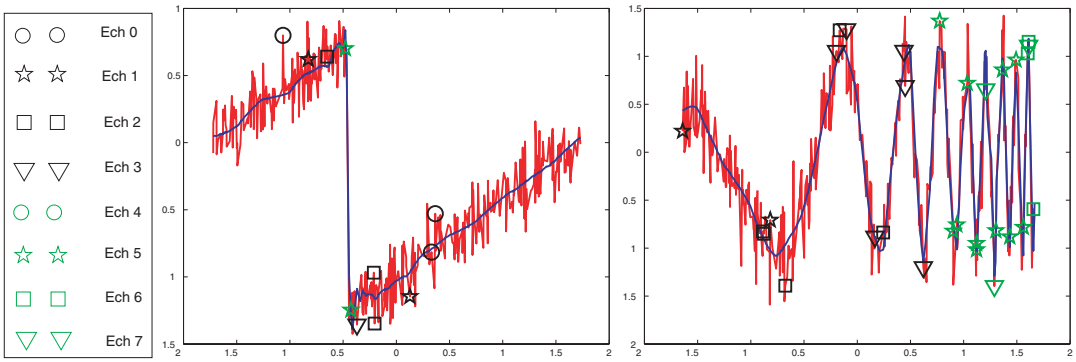
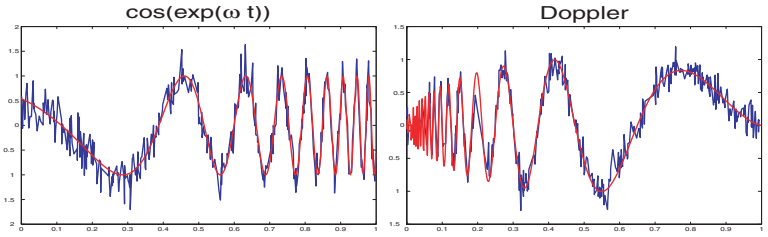
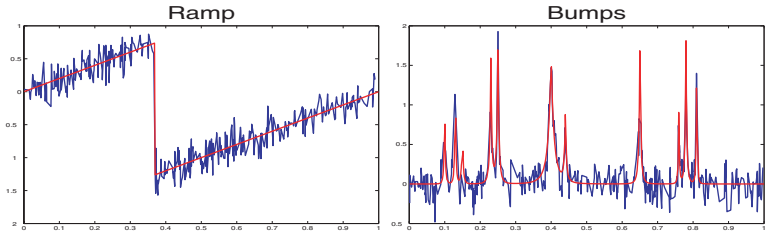


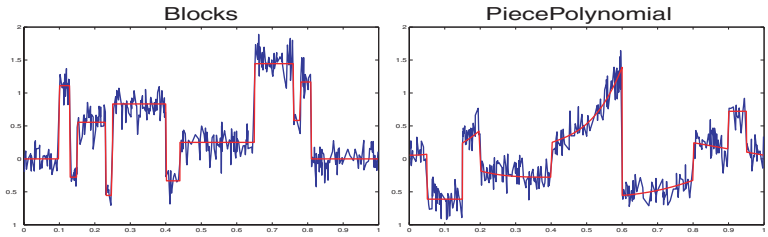
Figure 5. Illustration des solutions obtenues par la méthode du LARS sur les fonctions Ramp et $\cos(\exp(\omega t))$. *L'image de gauche montre comment les échelles de hautes fréquences sont utilisées pour marquer la rupture dans le signal Ramp. L'image de droite illustre le fait que le LARS utilise différentes échelles pour reconstruire différentes portions d'un signal présentant des aspects multi-résolutions.*



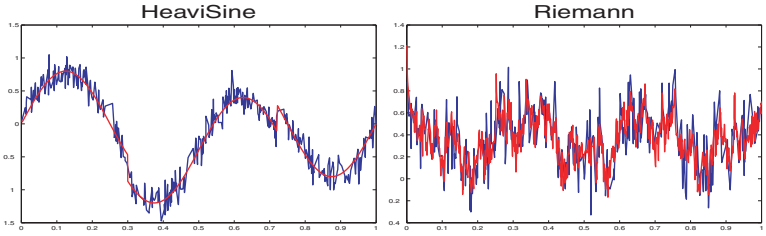
(a) Signaux présentant des aspects multirésolutions.



(b) Signaux présentant de fortes ruptures.



(c) Signaux présentant à la fois des ruptures et des aspects multirésolutions.



(d) Autres signaux plus régulier (au sens de l'erreur des moindres carrés) ou plus chaotique.

Figure 6. Signaux à estimer et données d'apprentissage bruitées.

Tableau 1. Tableau de résultats. Moyenne et écart-type de l'erreur au sens des moindres carrés sur 30 itérations. Moyenne des vecteurs supports nécessaires pour construire la solution. Nombre de meilleures performances sur les 30 itérations (le total dépasse parfois 30, plusieurs solutions LARS étant identiques). Pour plus de compacité, toutes les erreurs ont été multipliées par 1000.

Algorithme	Back-fitting	ε-SVM	LARS- $\sum_i \lvert\beta_i\rvert$	ν-LARS	LARS-VA	LARS-HF
$\cos(\exp(\omega t))$	25 ± 6.4 4000 0	29 ± 5.6 111.1 0	$\mathbf{20 \pm 4.7}$ 42.3 24	20 ± 4.8 42 19	21 ± 4.9 48.7 16	$\mathbf{20 \pm 4.7}$ 43.4 20
Doppler	19 ± 4.1 4000 0	25 ± 5.6 110.7 0	14 ± 4.2 44.1 11	$\mathbf{13 \pm 3.7}$ 42 24	14 ± 4.5 47.1 9	13 ± 3.8 42.5 21
Ramp	12 ± 3.6 4000 0	14 ± 3.7 57.6 2	$\mathbf{8 \pm 3.5}$ 19.2 19	$\mathbf{8 \pm 3.5}$ 17 20	9 ± 3.7 21.4 13	10 ± 3.5 15.8 8
Bumps	24 ± 7.1 4000 0	25 ± 7.3 52.7 0	$\mathbf{20 \pm 5.0}$ 42.2 23	20 ± 5.2 46 19	20 ± 5.2 45.3 19	20 ± 5.1 41.7 18
Blocks	24 ± 4.1 4000 0	26 ± 4.5 107.7 0	21 ± 3.3 50.3 17	$\mathbf{20 \pm 3.4}$ 54 21	23 ± 3.9 60.8 5	22 ± 3.6 49.2 8
Piece-Polynomial	18 ± 5.0 4000 0	19 ± 5.1 68.8 0	$\mathbf{15 \pm 4.9}$ 53.1 18	15 ± 5.3 52 16	15 ± 5.2 55.4 14	15 ± 5.3 59.1 15
HeavySine	3.3 ± 0.8 4000 0	$\mathbf{2.8 \pm 0.5}$ 29 28	3.4 ± 0.8 17.4 0	3.3 ± 0.8 18 2	3.3 ± 0.7 19.6 0	3.4 ± 0.8 20.7 2
Riemann	19 ± 1.6 4000 5	$\mathbf{18 \pm 1.7}$ 57.9 15	19 ± 2.0 51.4 7	19 ± 1.9 50 8	19 ± 1.9 52.1 6	19 ± 1.9 57.5 6

genre de problèmes. La parcimonie de la solution est systématiquement à l'avantage du LARS (entre 10 % et 62 % d'avantage), du fait de la régularisation ℓ_1.

Les résultats des méthodes LARS-VA et LARS-HF sont très intéressants : ils ne font intervenir aucun paramètre de compromis ou borne sur les $|\beta_i|$ et sont toujours proches des meilleurs résultats, à la fois au niveau de l'erreur et de la parcimonie.

5. Conclusions

Les résultats obtenus sont très encourageants : ils montrent la performance de la méthode pour différents types de signaux artificiels, tout en améliorant la rapidité de traitement et en diminuant le nombre de paramètres nécessaires par rapport aux SVM. L'écart de performance est le plus important sur les problèmes nécessitant la prise en compte des ruptures ou des aspects multi-résolutions. L'intérêt des approches LARS-VA et LARS-HF est particulièrement remarquable : ces méthodes permettent d'éliminer le paramètre de régularisation tout en proposant des résultats toujours proches des meilleurs.

Étant données les performances obtenues, la parcimonie des solutions et le faible nombre de paramètres à régler, la stratégie proposée dans cet article est tout à fait appropriée pour la description et le débruitage de grandes masses de signaux non-stationnaires en vue d'application de classification. Contrairement à [DJ94] ou [KS00], la description ainsi obtenue est covariante en translation. De plus, le cadre proposé est beaucoup plus souple que les méthodes classiques de débruitage de type *Wavelet Shrinkage* [DJ94] : nous pouvons aisément changer de fonction coût, de type de régularisation et nous traitons les signaux non uniformément échantillonnés de manière transparente.

Les principales perspectives de ce travail concernent l'optimisation de la forme de l'ondelette mère utilisée, à la manière de

[MDL05] et l'intégration de critères discriminants dans la sélection des sources d'information.

Le but serait de renforcer les atouts de la méthode pour la représentation de signaux non-stationnaires dans les problèmes de classification.

Une autre perspective serait d'étudier un algorithme efficace basé sur le LARS permettant de résoudre un problème de minimisation du type :

$$R_{reg}[f] = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{N} f_j(x_i)\right)^2 + \lambda \sum_{j=1}^{N} \theta_j \|f_j\|^2_{\mathcal{H}_j} + \sum_{j=1}^{N} |\theta_j|,$$

$$f_j(x) = \sum_{i=1}^{n} \beta_{ij} K_j(x, x_i) \qquad (38)$$

Cette formulation permet à la fois d'obtenir une solution parcimonieuse et de préserver un sens au terme régularisant, pour étudier de manière plus théorique les propriétés d'un tel estimateur [AAP04].

Références

[AAP04] U. AMATO, A. ANTONIADIS, and M. PENSKY, Wavelet kernel penalized estimation for non-equispaced design regression. Technical report, Istituto per le Applicazioni del Calcolo Mauro Picone, 2004.

[BBE+03] J. BI, K. BENNETT, M. EMBRECHTS, C. BRENEMAN, and M. SONG, Dimensionality reduction via sparse support vector machines. *Journal of Machine Learning Research*, 3:1229-1243, 2003.

[BTJ04] F.R. BACH, R. THIBAUX, and M.I. JORDAN, Computing regularization paths for learning multiple kernels. In *Advances in Neural Information Processing Systems*, volume 17, 2004.

[CDS98] S.S. CHEN, D.L. DONOHO, and M.A. SAUNDERS, Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33-61, 1998.

[DDM04] I. DAUBECHIES, M. DEFRISE, and C. DE MOL, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Pure and Applied Mathematics*, 2004.

[DJ94] D. DONOHO and I. JOHNSTONE, Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425-455, 1994.

[EHJT04] B. EFRON, T. HASTIE, I. JOHNSTONE, and R. TIBSHIRANI, Least angle regression. *Annals of statistics*, 32(2):407-499, 2004.

[EPP00] T. EVGENIOU, M. PONTIL, and T. POGGIO, Regularization Networks and Support Vector Machines. *Advances in Computational Mathematics*, 13(1):1-50, 2000.

[Gra98] Y. GRANDVALET, Least absolute shrinkage is equivalent to quadratic penalization. In *ICANN*, pages 201-206, 1998.

[GRC05] V. GUIGUE, A. RAKOTOMAMONJY, and S. CANU, Kernel basis pursuit. In *16th European Conference on Machine Learning*, Porto, 2005.

[Gui05] V. GUIGUE, *Méthodes à noyaux pour la représentation et la discrimination de signaux non-stationnaires*. PhD thesis, INSA de Rouen, 2005.

[HT86] T. HASTIE and R. TIBSHIRANI, Generalized additive models. *Statistical Science*, 1:297-318, 1986.

[HT90] T. HASTIE and R. TIBSHIRANI, *Generalized Additive Models*. Chapman and Hall, London, 1990.

[KS00] A. KOVAC and B.W. SILVERMAN, Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *Journal of the American Statistical Association*, 95:172-183, 2000.

[KW71] G. KIMELDORF and G. WAHBA, Some results on Tchebycheffian spline functions. *J. Math. Anal. Applic.*, 33:82-95, 1971.

[LCV+04] G. LOOSLI, S. CANU, S.V.N. VISHWANATHAN, A. J. SMOLA, and M. CHATTOPADHYAY, Une boîte à outils rapide et simple pour les svm. In Michel Liquière and Marc Sebban, editors, *CAp*, pages 113-128. Presses Universitaires de Grenoble, 2004.

[Lju87] L. LJUNG, *System Identification – Theory for the User*. 1987.

[Mal97] S. MALLAT. *A Wavelet Tour Of Signal Processing*. Academic Press, 1997.

[MLD05] A. MAITROT, M.F. LUCAS, and C. DONCARLI, Design of wavelets adapted to signals and application. In *IEEE Internantional Conference an Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphie, USA, March 2005.

[Ng04] A.Y. NG, Feature selection, l1 vs. l2 regularization, and rotational invariance. In *International Conference on Machine Learning*, 2004.

[RC05] A. RAKOTOMAMONJY and S. CANU, Frame, reproducing kernel, regularization and learning. *JMLR*, 6:1485-1515, 2005.

[SS02] B. SCHÖLKOPF and A.J. SMOLA, *Learning with kernels*. MIT Press, 2002.

[Tib96] R. TIBSHIRANI, Regression shrinkage and selection via the lasso. *J. Royal. Statist.*, 58(1):267-288, 1996.

[VB02] P. VINCENT and Y. BENGIO, Kernel matching pursuit. *Machine Learning Journal*, 48(1):165-187, 2002.

Vincent **Guigue**

Ingénieur en mécanique de l'INSA de Rouen, il a ensuite passé un DEA en optimisation et apprentissage statistique. Sa thèse, effectuée au sein du laboratoire PSI de Rouen a porté sur la représentation et la discrimination de signaux non-stationnaires, en particulier sur la reconnaissance de signaux EEG dans le cadre des interfaces cerveaux-machines. Ses recherches s'articulent autour des méthodes à noyaux et de la régularisation $\mathcal{L}_1$.

Alain **Rakotomamonjy**

Ingénieur en électronique et informatique de l'Ecole Supérieure d'Electronique de l'Ouest en 1993, il soutient sa thèse à l'Université d'Orléans en 1997. Depuis 1999, il occupe un poste de maître de conférences à l'INSA de Rouen. Ses sujets de recherche tournent autour des techniques d'apprentissages statistiques, des méthodes à noyaux, de la représentation des signaux en ondelettes et de la classification de signaux non-stationnaires.

Stéphane **Canu**

Docteur de l'Université de Compiègne (1986), il travaille ensuite dans le département d'informatique de l'Université de Technologie. Il rejoint l'INSA de Rouen en 1997 en tant que professeur. Ses travaux s'articulent autour des techniques d'apprentissages statistiques et des méthodes à noyaux pour les problèmes de classification et de régression. Il est impliqué dans trois programmes européens (Neufodi, EMS et EM2S) et dans le réseau d'excellence PASCAL.