

Estimation de signaux par noyaux d'ondelettes

Vincent GUIGUE Alain RAKOTOMAMONJY Stéphane CANU

Lab. Perception, Systèmes, Information - CNRS - FRE 2645

Avenue de l'Université, 76801 St Étienne du Rouvray

Tél : 02 32 95 97 03 Fax : 02 32 95 97 08

Vincent.Guigue@insa-rouen.fr

Résumé – Cet article présente une méthode de régression pour les signaux non uniformément échantillonnés. Notre méthode propose une approche de type apprentissage supervisé, qui combine une approximation \mathcal{L}_2 par noyaux d'ondelettes et une régularisation \mathcal{L}_1 multi-échelles. L'utilisation de l'algorithme *Least Angle Regression* pour la résolution du problème permet d'introduire de nouvelles manières de régler le compromis biais-variance.

Abstract – 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 kernel 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.

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 noyaux reproduisants 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. 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. L'extension aux signaux quelconques nécessite en réalité de ré-échantillonner les signaux [KS00]. Les méthodes existantes utilisant les noyaux d'ondelettes [AAP04] combine une résolution de type EM (coûteuse) avec l'utilisation de plusieurs paramètres de régularisation. Notre but est double : d'une part, nous souhaitons diminuer la complexité temporelle de la résolution de ce problème, d'autre part nous proposons des solutions pour régler automatiquement les paramètres de la méthode d'approximation.

Nous avons utilisé la formulation du LASSO (*Least Absolute Shrinkage and Selection Operator*) [Tib96], basée sur une régularisation \mathcal{L}_1 , et l'algorithme du LARS (*Step-wise Least Angle Regression*) [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.

2 Méthode

Nous nous plaçons dans le contexte d'une estimation de la fonction de régression classique *i.e*, nous avons un ensemble de données $\{x_i, y_i\}_{i=1..n}$ avec : $y_i = f(x_i) + b_i$ où les b_i sont des variables aléatoire et $x_i \in \mathbb{R}$. Notre objectif est d'obtenir une estimation de f à partir des échantillons bruités y_i . Les x_i ne sont pas équidistants. Nous cherchons la fonction f de l'espace des hypothèses \mathcal{H} qui minimise le risque empirique régularisé :

$$R_{reg}[f] = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \Omega(f) \quad (1)$$

où $\Omega(f)$ mesure la régularité de la solution et où λ est le compromis biais-variance. L'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 :

$$f(x) = \sum_{i=1}^n \beta_i K(x_i, x) \quad (2)$$

où K est le noyau reproduisant de l'espace de Hilbert \mathcal{H} et où les β_i sont les coefficients de régression.

L'utilisation de noyaux d'ondelettes pour la régression, combinée à une régularisation $\Omega(f)$ a déjà fait l'objet d'études [AAP04, RC05]. La méthode présentée dans cet

article apporte plus de rapidité, ainsi que la possibilité de régler automatiquement tous les paramètres.

2.1 Noyaux d'ondelettes

L'application du théorème du représentant suppose que \mathcal{H} soit un espace de Hilbert à noyaux reproduisants. Les conditions suffisantes d'auto-reproduction [RC05] sont données ci-dessous :

- Tout ensemble fini $\{\psi_n\}_{i=1\dots n}$ de vecteurs d'un espace de Hilbert \mathcal{H} est une structure oblique (ou frame) du sous-espace qu'elle génère.
- si par ailleurs, l'ensemble des ψ_n est borné et de norme finie dans \mathcal{H} alors, l'espace engendré est à noyau reproduisant.

Dans ce cas, le noyau de cette espace s'écrit :

$$k(x, y) = \sum_{i=1}^n \bar{\psi}_i(x) \psi_i(y)$$

où $\bar{\psi}$ est la frame duale de ψ . Ainsi, on peut en conclure qu'une famille finie (orthogonale ou non) d'ondelettes génère un espace à noyau reproduisant. Cette espace peut donc être utilisé dans le cadre du théorème de Kimerdolf-Wahba pour résoudre un problème d'approximation.

Si l'espace d'hypothèses \mathcal{H} est décrit par un ensemble d'ondelettes orthogonales ψ , alors cet espace se décompose en N sous espaces : $\mathcal{H} = \mathcal{H}_1 \oplus \dots \oplus \mathcal{H}_N$ où chaque espace de Hilbert à noyau reproduisant \mathcal{H}_i est l'espace engendré par une famille d'ondelettes de même échelle i .

Estimer la fonction sur \mathcal{H} nécessite l'utilisation des noyaux des \mathcal{H}_i , pour modéliser chacune des échelles. En reprenant la non linéarisation de Vincent et Bengio [VB02], nous considérons chaque $K_i(x, \cdot)$ comme une source d'information. Nous construisons donc un noyau multi-échelles comme la concaténation de différents noyaux élémentaires : $K = [K_1 \dots K_i \dots K_N]$. 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 :

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

2.2 Régularisation multi-échelles et LARS

La régularisation est souvent associée à la parcimonie de la solution. La meilleure manière de l'améliorer est de minimiser $\|\beta\|_{\ell_0}$, c'est à dire de prendre $\Omega(f) = \|\beta\|_{\ell_0}$. Cependant, ce problème est NP-difficile et deux solutions classiques consistent à utiliser soit la norme ℓ_1 [Tib96], soit la norme ℓ_2 en approximant parcimonieusement la solution [VB02].

Grandvalet a montré que l'*Adaptive Ridge Regression* (ARR, $\Omega(f) = \sum_{ij} \lambda_i \beta_{ij}^2$) est équivalente au LASSO ($\Omega(f) = \sum_{ij} |\beta_{ij}|$) [Gra98]. Dans le cas présent, l'ARR utilise une régularisation multi-échelles, λ_i étant fixé pour chaque K_i . La formulation du LASSO permet donc d'obtenir simplement une régularisation multi-échelles en s'affranchissant d'un grand nombre d'hyper-paramètres.

La plupart des problèmes d'apprentissage sont régularisés en ℓ_2 et il existe beaucoup d'outils pour les traiter. A l'inverse, la formulation ℓ_1 demandait jusqu'à maintenant des résolutions coûteuses, basées sur la programmation linéaire [CDS98] ou sur des versions améliorées de EM [Gra98]. L'algorithme *Stepwise Least Angle Regression* (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, en partant d'un ensemble vide. Les premières itérations sont peu coûteuses du fait de la petite taille de l'ensemble actif des sources d'information. Le coût augmente au fil des itérations, mais la régularisation ℓ_1 , gage de parcimonie, limite le nombre d'itérations nécessaires. A 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'un seul point est ajouté (ou retiré) de l'ensemble actif à chaque itération, il est possible de mettre à jour la matrice inversée au lieu de la recalculer entièrement à la manière de la formulation simple-SVM [LCV⁺04].

Finalement, le problème du LARS s'écrit :

$$\min_{\beta} \sum_{i=1}^n \|y_i - \sum_{j=1}^N \sum_{k=1}^n \beta_{jk} K_j(x_k, x_i)\|^2 \quad (4)$$

$$\sum_{j,k} |\beta_{jk}| \leq t$$

Cette formulation est équivalente à (1) avec $\Omega(f) = \|\beta\|_{\ell_1}$.

2.3 Modification du LARS

Afin d'améliorer la parcimonie et les résultats en classification 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. 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.

2.4 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 (4) constitue donc un enjeu essentiel.

Pour faire face à ce problème nous utilisons une propriété du LARS : le calcul le 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 [BTJ04].

Le critère le plus classique pour régulariser le LARS consiste à fixer une borne sur la somme des $|\beta_i|$. Cepen-

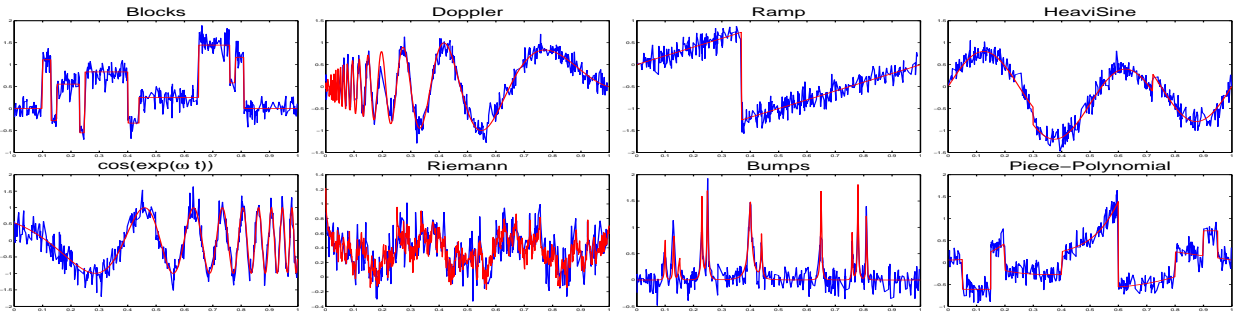


FIG. 1 – Signaux à estimer et données d’apprentissage bruitées.

nant, ce critère est particulièrement abstrait et difficile à régler.

2.4.1 Autres définitions de seuil

Nous proposons 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-correlation du résidu qui mesure sa ressemblance à un bruit blanc.

2.4.2 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. Nous proposons deux heuristiques pour construire ces sources pièges :

- LARS-RV. Nous introduisons des vecteurs de variables aléatoires 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. Nous utilisons alors les échelles hautes fréquences comme des échelles pièges. En effet, dans le noyau correspondant aux hautes fréquences (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 un début de sur-apprentissage.

Cette dernière famille de solution est particulièrement intéressante puisqu’elle permet d’avoir un algorithme complètement non-paramétrique.

3 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 1. il s’agit de signaux classiques, utilisés par

Donoho et Johnstone [DJ94]. 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 ($\sum_i |\beta_i|, \nu$) 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*, 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 multirésolution donnent les meilleurs résultats sur les bases de données faisant intervenir un aspect multirésolution ($\cos(\exp(\omega t)), \text{Doppler}, \dots$), mais aussi sur des bases plus classique (Ramp, Bumps...). A l’inverse, les SVM se montrent très efficaces sur les bases de données régulières (HeavySine). L’ ϵ -tube autour de la solution est bien adapté à ce genre de problèmes. La parcimonie de la solution est systématiquement à l’avantage du LARS, du fait de la régularisation ℓ_1 . Les résultats des méthodes LARS-RV et LARS-HF sont très intéressants : ils sont proches des meilleurs résultats dans la plupart des cas, sans faire intervenir de paramètres.

4 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 plus important sur les problèmes nécessitant une approche multirésolution.

Il reste néanmoins à valider cette méthode sur des cas réels. Une autre perspective de ce travail concerne l’optimisation de la forme de l’ondelette mère utilisée. Enfin, notons que l’estimation des signaux par cette méthode peut servir de description pour les problèmes de classification de signaux. Contrairement à [DJ94] ou [KS00], la description ainsi obtenue est covariante en translation.

Algorithme	Back-fitting	ϵ -SVM	LARS- $\sum_i \beta_i $	ν -LARS	LARS-RV	LARS-HF
Blocks	0.024 ± 0.0041	0.026 ± 0.0045	0.021 ± 0.0033	0.020 ± 0.0034	0.023 ± 0.0039	0.022 ± 0.0036
	4000	107.7	50.3	54	60.8	49.2
	0	0	17	21	5	8
Doppler	0.019 ± 0.0041	0.025 ± 0.0056	0.014 ± 0.0042	0.013 ± 0.0037	0.014 ± 0.0045	0.013 ± 0.0038
	4000	110.7	44.1	42	47.1	42.5
	0	0	11	24	9	21
Ramp	0.012 ± 0.0036	0.014 ± 0.0037	0.008 ± 0.0035	0.008 ± 0.0035	0.009 ± 0.0037	0.010 ± 0.0035
	4000	57.6	19.2	17	21.4	15.8
	0	2	19	20	13	8
HeavySine	0.0033 ± 0.0008	0.0028 ± 0.0005	0.0034 ± 0.0008	0.0033 ± 0.0008	0.0033 ± 0.0007	0.0034 ± 0.0008
	4000	29	17.4	18	19.6	20.7
	0	28	0	2	0	2
$\cos(\exp(\omega t))$	0.025 ± 0.0064	0.029 ± 0.0056	0.020 ± 0.0047	0.020 ± 0.0048	0.021 ± 0.0049	0.020 ± 0.0047
	4000	111.1	42.3	42	48.7	43.4
	0	0	24	19	16	20
Riemann	0.019 ± 0.0016	0.018 ± 0.0017	0.019 ± 0.0020	0.019 ± 0.0019	0.019 ± 0.0019	0.019 ± 0.0019
	4000	57.9	51.4	50	52.1	57.5
	5	15	7	8	6	6
Bumps	0.024 ± 0.0071	0.025 ± 0.0073	0.020 ± 0.0050	0.020 ± 0.0052	0.020 ± 0.0052	0.020 ± 0.0051
	4000	52.7	42.2	46	45.3	41.7
	0	0	23	19	19	18
Piece-Polynomial	0.018 ± 0.0050	0.019 ± 0.0051	0.015 ± 0.0049	0.015 ± 0.0053	0.015 ± 0.0052	0.015 ± 0.0053
	4000	68.8	53.1	52	55.4	59.1
	0	0	18	16	14	15

TAB. 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).

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.

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. Brene-man, 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.
- [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.
- [Gra98] Y. Grandvalet. Least absolute shrinkage is equivalent to quadratic penalization. In *ICANN*, pages 201–206, 1998.
- [HT86] T. Hastie and R. Tibshirani. Generalized additive models. *Statistical Science*, 1 :297–318, 1986.
- [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 *CAP*, 2004.
- [Lju87] L. Ljung. *System Identification - Theory for the User*. 1987.
- [Mal97] S. Mallat. *A Wavelet Tour Of Signal Processing*. Academic Press, 1997.
- [RC05] A. Rakotomamonjy and S. Canu. Frame, reproducing kernel, regularization and learning. Technical report, Lab. Perception, Systemes, information, 2005.
- [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.