

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

Thèmes : 3.3 Ondelettes, multirésolution; 4.1 Détection et estimation statistiques.

Problème : approximation de signaux à partir de données non uniformément échantillonnées.

Originalité : nous avons combiné une approximation \mathcal{L}_2 par noyaux d'ondelettes et une régularisation \mathcal{L}_1 multi-échelles pour la formulation du problème. Nous avons introduit une heuristique pour déterminer automatiquement le compromis biais-variance.

Résultats : amélioration du temps de calcul et des performances sur des signaux artificiels classiques, utilisés en particulier par Donoho et Johnstone [4].

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 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. Nous avons utilisé la formulation du LASSO (*Least Absolute Shrinkage and Selection Operator*) [11], basée sur une régularisation \mathcal{L}_1 , et l'algorithme du LARS (*Stepwise Least Angle Regression*) [5] 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. Notre but est d'améliorer les performances et le temps de calcul par 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 sont non 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 [8] :

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

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

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 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 noyaux 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 re-prenant la non linéarisation de Vincent et Bengio [12], nous considérons chaque $K_i(x, \cdot)$ comme une source d'information. Nous construisons donc un noyau multi-échelles comme l'assemblage 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 :

$$f(x) = \sum_i \sum_j \beta_{ij} K_i(x_j, x) \tag{3}$$

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 [11], soit la norme ℓ_2 en approximant parcimonieusement la solution [12].

Grandvalet a montré que l'*Adaptive Ridge Regression* (ARR) ($\Omega(f) = \sum_{ij} \lambda_i \beta_{ij}^2$) est équivalente au LASSO ($\Omega(f) = \sum_{ij} |\beta_{ij}|$) [6]. 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 y faire face. A l'inverse, la formulation ℓ_1 demandait jusqu'à maintenant des résolutions coûteuses, basées sur la programmation linéaire [3] ou sur des versions améliorées de EM [6]. L'algorithme *Stepwise Least Angle Regression* (LARS) [5] 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* [6, 3] 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 [10].

2.3 Optimisation du compromis biais-variance

Le LARS calcule le chemin complet de régularisation, ce qui permet de régler le compromis biais-variance dynamiquement [1]. 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 les alternatives suivantes :

- le ν -LARS, où ν est un seuil sur le pourcentage de points pouvant être sélectionnés comme supports de la solution,
- le critère de Ljung [9], où un seuil est fixé sur l'auto-correlation du résidu qui mesure sa ressemblance à un bruit blanc,
- l'échelle piège. Ce dernier critère repose sur le mécanisme de sélection des sources d'information du LARS, qui est lié à la corrélation entre ces sources et le résidu à chaque itération. L'idée est d'introduire des sources d'information piégées comme des variables aléatoires (LARS-RV) [2], ou de se servir de l'échelle des hautes fréquences (LARS-HF). Lorsque ces sources seront sélectionnées, cela signifie qu'elles sont plus corrélées au résidu que n'importe qu'elle autre source d'information. C'est à dire que le résidu est corrélé respectivement à du bruit où à un point très précis, ce qui est interprété comme du sur-apprentissage. Cette dernière solution est particulièrement intéressante puisque complètement non-paramétrique.

3 Résultats

Nous avons comparé le LARS, avec plusieurs critères d'arrêt, par rapport aux ϵ -SVM et au *back-fitting* [7] pour construire une estimation des signaux artificiels présentés figure 1. il s'agit de signaux classiques, utilisés par Donoho et Johnstone [4]. L'échantillonnage est aléatoire sur l'intervalle $[0, 1]$ et les données sont bruitées. Nous avons utilisé une

base d'apprentissage de 100 points et une base de test 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.

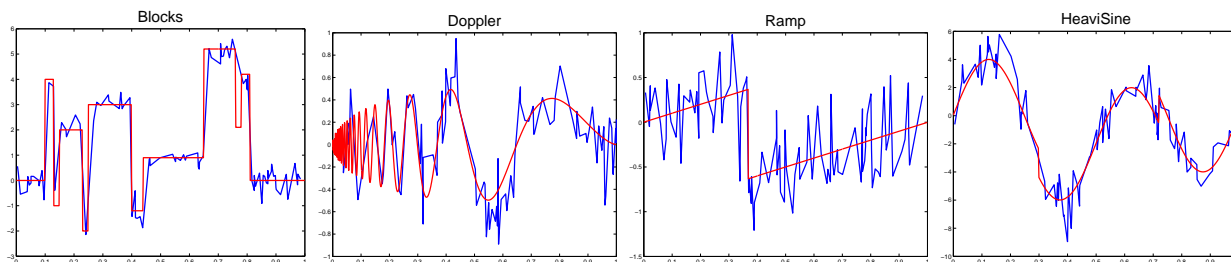


FIG. 1 – Signaux à estimer et données d'apprentissage.

Algorithme	Back-fitting	ϵ -SVM	LARS- $\sum_i \beta_i $	ν -LARS	LARS-RV	LARS-HF
Blocks	0.85 ± 0.19	1.18 ± 0.20	0.83 ± 0.34	0.67 ± 0.15	0.67 ± 0.15	0.71 ± 0.21
	800	35.10	45.02	38	42.35	27.87
	1	0	0	13	11	5
Doppler	0.033 ± 0.0063	0.039 ± 0.0065	0.036 ± 0.0068	0.030 ± 0.0060	0.031 ± 0.0059	0.035 ± 0.0075
	800	55.33	17.10	28	32.80	21.60
	12	0	0	14	11	3
Ramp	0.028 ± 0.0069	0.026 ± 0.0025	0.028 ± 0.0070	0.026 ± 0.0060	0.025 ± 0.0061	0.024 ± 0.0062
	800	37.11	25.63	22	33.93	27.33
	4	8	0	4	7	7
HeavySine	0.46 ± 0.11	0.48 ± 0.12	0.52 ± 0.21	0.40 ± 0.12	0.32 ± 0.11	0.35 ± 0.15
	800	51.33	42.25	35	45.11	19.03
	3	0	0	0	15	12

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 la solution. Nombre de meilleures performances sur les 30 itérations.

4 conclusion

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 multi-résolutions comme les *blocks* et le *doppler*. Il reste néanmoins à valider cette méthode sur des cas réels. Cette approche est équivalente aux méthodes d'estimation par rétrécissement des coefficients d'ondelettes proposé par Donoho et al [4], tout en prenant en compte le cas où le signal est non uniformément échantillonné.

Références

- [1] 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.
- [2] 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.
- [3] 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.
- [4] Donoho D. and Johnstone I. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81 :425–455, 1994.
- [5] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of statistics*, 32(2) :407–499, 2004.
- [6] Y. Grandvalet. Least absolute shrinkage is equivalent to quadratic penalization. In *ICANN*, pages 201–206, 1998.
- [7] T. Hastie and R. Tibshirani. Generalized additive models. *Statistical Science*, 1 :297–318, 1986.
- [8] G. Kimeldorf and G. Wahba. Some results on Tchebycheffian spline functions. *J. Math. Anal. Applic.*, 33 :82–95, 1971.
- [9] L. Ljung. *System Identification - Theory for the User*. 1987.
- [10] 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.
- [11] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal. Statist.*, 58(1) :267–288, 1996.
- [12] P. Vincent and Y. Bengio. Kernel matching pursuit. *Machine Learning Journal*, 48(1) :165–187, 2002.