

# Segmentation temporelle de signaux à l'aide du Lasso

Zaïd HARCHAOUI & Céline LÉVY-LEDUC

LTCI, Télécom Paris & CNRS  
46 rue Barrault, 75634 Paris cedex 13 France  
zharchao@enst.fr, levyledu@enst.fr

**Résumé** – Nous proposons de ramener la détection rétrospective de ruptures multiples à un problème de sélection de variables, que nous résolvons à l'aide de l'algorithme LAR/LASSO suivi d'une version réduite de l'algorithme de programmation dynamique. Les résultats obtenus sont satisfaisants à la fois sur des données simulées et réelles.

**Abstract** – We propose to cast multiple change-point estimation as a variable selection problem. Our method is implemented in practice by combining the LAR algorithm and a reduced version of the dynamic programming algorithm. We obtain competitive results on artificial and real datasets.

## 1 Introduction

Estimer les instants de ruptures au sein d'un signal constitue une problématique majeure dans de nombreux domaines d'applications, tels que la segmentation de signaux audio [9] et l'analyse des signaux EEG [4]. Le problème consiste à "découper" un signal temporel en segments de durées *a priori* différentes mais homogènes, au sens où la quantité d'intérêt demeure à peu près constante au cours du temps dans chaque segment. En effet, la segmentation temporelle de signaux peut, la plupart du temps, être ramenée au cas de l'estimation des instants de ruptures dans un signal constant par morceaux [6][4].

Une abondante littérature a été consacrée à ce sujet, tant du point de vue séquentiel [1] que du point de vue rétrospectif [14], qui seul nous intéresse ici. La plupart des approches rétrospectives s'appuient sur un algorithme dit de programmation dynamique (DP), qui permet de retrouver  $K$  ruptures parmi  $n$  observations avec une complexité en temps de  $\mathcal{O}(Kn^2)$ . Cet aspect constitue un sérieux obstacle au traitement de signaux de grande taille.

Nous nous proposons ici de traiter le problème de l'estimation rétrospective d'instants de ruptures en adoptant une perspective différente. En effet, estimer les instants de ruptures peut se ramener à l'aide d'une reparamétrisation à un problème de sélection de variables pertinentes. Ceci nous permet alors de mettre à profit les avancées pratiques [7] et théoriques [16][13] les plus récentes sur la régression avec une pénalité de type Lasso [15]. Ainsi, le Lasso nous permet de développer un algorithme efficace de première estimation d'instants de ruptures. Cette première estimation est ensuite raffinée à l'aide d'une version réduite rDP de DP.

Dans un premier temps, nous mettons en lumière le lien entre l'estimation rétrospective d'instants de ruptures et la sélection de variables pour une matrice d'expérience que l'on précisera. Puis, nous décrivons étape par étape la méthode Cachalot que nous proposons. Enfin, nous évaluons

les performances de notre méthode sur des données simulées et réelles.

## 2 Modèle

Nous nous intéressons ici à l'estimation des instants de ruptures  $\tau_k^*$  d'observations  $(Y_t)$  suivant le modèle :

$$\begin{aligned} Y_t &= \mu_k^* + \varepsilon_t, \quad t = 1, \dots, n \\ \text{avec} \quad & \tau_{k-1}^* + 1 \leq t \leq \tau_k^*, \quad 1 \leq k \leq K^*, \quad \tau_0^* = 0 \end{aligned} \quad (1)$$

où  $\varepsilon_t$  représente un bruit additif.

Nous proposons d'ajouter au critère des moindres carrés une pénalisation  $\ell_1$  de l'amplitude des sauts comme suit :

$$\begin{cases} \text{Minimiser} & \sum_{t=1}^n (Y_t - \mu_t)^2 \\ \mu_1, \dots, \mu_n & \\ \text{sous la contrainte} & \sum_{t=1}^{n-1} |\mu_{t+1} - \mu_t| \leq s. \end{cases} \quad (2)$$

En effet, la pénalisation  $\ell_1$  favorisant naturellement la parcimonie, la formulation permet de contrôler le nombre et l'amplitude des sauts *via* le paramètre  $s$ .

Le critère ci-dessus peut être réécrit sous la forme classique d'une régression avec une pénalité de type Lasso à partir des observations suivantes :

$$Y_n = X_n \beta^n + \varepsilon_n, \quad (3)$$

$Y_n$  étant un vecteur de taille  $n \times 1$ ,  $X_n$  une matrice triangulaire inférieure de taille  $n \times n$  de termes non nuls égaux à 1 et  $\beta^n$  étant un vecteur de taille  $n \times 1$  ayant toutes ses composantes nulles à l'exception de celles ayant pour indices les instants de ruptures  $\tau_k^*$ . Le problème se réécrit alors :

$$\begin{cases} \text{Minimiser} & \|Y_n - X_n \beta\|_2^2 \\ \beta & \\ \text{sous la contrainte} & \|\beta\|_1 \leq s. \end{cases} \quad (4)$$

Ce problème peut être aisément résolu pour toutes les valeurs de  $s$  à l'aide de la famille d'algorithmes LAR.

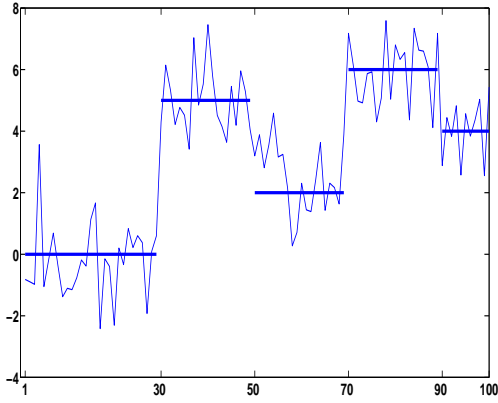


FIG. 1 – Exemple d’observations ( $Y_n$ ) et signal déterministe sous-jacent

### 3 Description de la méthode

Considérons des observations ( $Y_n$ ) satisfaisant le modèle (3) représentées dans la Figure 1. La méthode que l’on propose peut se décomposer en trois étapes.

#### Étape 1 : Algorithme LAR/LASSO

Les ruptures potentielles apparaissent au fur et à mesure que l’on parcourt le chemin de régularisation du problème :

$$\begin{cases} \text{Minimiser}_{\beta} & \|Y_n - X_n\beta\|_2^2 \\ \text{sous la contrainte} & \|\beta\|_1 \leq s \end{cases} \quad (5)$$

Les 9 premiers instants de ruptures ainsi obtenus à l’aide de la modification du LAR indiquée dans [7] sont  $S = \{21, 23, 28, 29, 30, 50, 69, 70, 90\}$ . L’évolution des valeurs des coefficients au fur et à mesure que l’on parcourt le chemin de régularisation sont représentées dans les Figures 2 et 3.

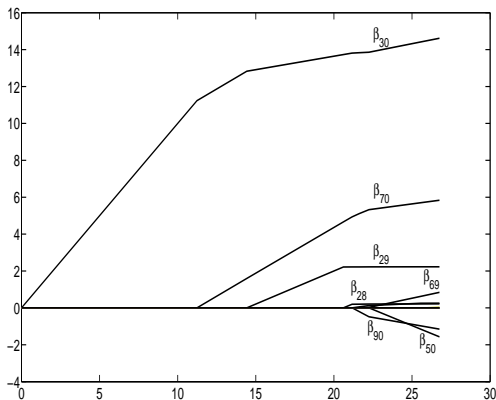


FIG. 2 – Profils des coefficients obtenus par le Lasso le long du chemin de régularisation

En général, les vrais instants de ruptures sont inclus dans  $S$ . Cependant, plusieurs instants de ruptures non pertinents peuvent également y figurer. L’étape suivante permet d’identifier les vraies ruptures dans  $S$ .

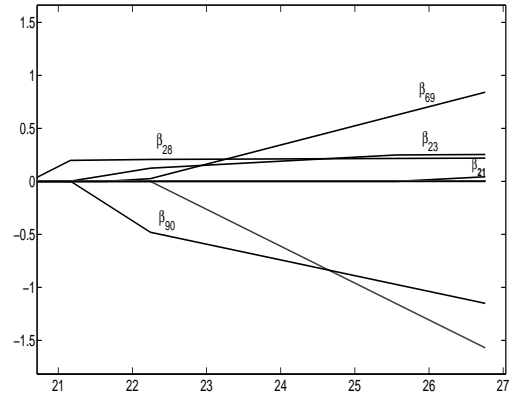


FIG. 3 – Zoom de la figure 2

$K$	$J(K)$	$(\hat{\tau}_1, \dots, \hat{\tau}_K)$
0	696.28	$\emptyset$
1	249.24	30
2	209.94	(30,70)
3	146.29	(30,50,69)
4	120.21	(30,50,70,90)
5	118.22	(30,50,69,70,90)
6	116.97	(21,30,50,69,70,90)
7	116.66	(21,29,30,50,69,70,90)
8	116.65	(21,23,29,30,50,69,70,90)
9	116.64	(21,23,28,29,30,50,69,70,90)

TAB. 1 –

#### Étape 2 : Version réduite de l’algorithme de programmation dynamique

Il s’agit à présent de déterminer les positions de  $K$  ruptures choisies parmi  $K_{\max}$  ruptures potentielles. Ceci est fait en minimisant le critère des moindres carrés suivant :

$$J(K) = \min_{\substack{\tau_1 < \dots < \tau_K \\ \text{t.q. } \tau_1, \dots, \tau_K \in S}} \sum_{k=1}^K \sum_{i=\tau_{k-1}+1}^{\tau_k} (Y_i - \overline{Y}_{[\tau_{k-1}+1:\tau_k]})^2$$

où  $\overline{Y}_{[\tau_{k-1}+1:\tau_k]}$  est la moyenne empirique sur le segment  $[\tau_{k-1} + 1, \tau_k]$  avec la convention  $\tau_0 = 1$ . Pour cela, on adapte l’algorithme de programmation dynamique [8] [2] décrit dans [10], Chapitre 12, p. 450 pour que les instants de ruptures soient recherchés uniquement dans  $S$ . On obtient ainsi pour chaque  $K$ ,  $J(K)$  et les instants de ruptures correspondants  $(\tau_1, \dots, \tau_K)$  notés  $(\hat{\tau}_1, \dots, \hat{\tau}_K)$ . Les résultats sont résumés dans le Tableau 1.

#### Étape 3 : Sélection de modèle

Enfin, le problème est de sélectionner le nombre d’instants de ruptures. Les valeurs de  $J(K)$  obtenues dans le Tableau 1 montrent qu’à partir de  $K = 4$  (qui est le vrai nombre de ruptures dans cet exemple),  $J$  ne décroît quasiment plus. On propose donc de choisir  $\hat{K}$ , l’estimateur du nombre de ruptures comme suit :

$$\hat{K} = \min_{k \geq 1} \left\{ \frac{J(k+1)}{J(k)} \geq 1 - \nu \right\},$$

où  $\nu > 0$ . Les valeurs du rapport  $J(k+1)/J(k)$  pour  $k = 0, \dots, 8$  obtenues pour  $\nu = 0.05$  sont représentées

dans la Figure 4. On en conclut que  $\hat{K} = 4$  et que les ins-

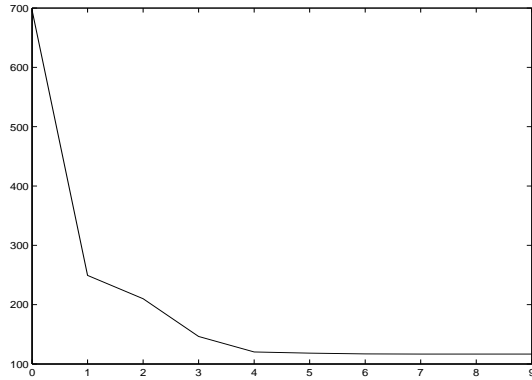


FIG. 4 – Evolution de  $J$  en fonction de  $K$

tants de ruptures sont (30, 50, 70, 90) grâce aux résultats du Tableau 1.

En résumé, la méthode que nous baptisons **Cachalot** (pour *CATching CHAnge-points with LassO*), suit les étapes suivantes.

### Cachalot

#### Entrée

- Observations  $Y \in \mathbb{R}^n$
- Borne supérieure  $K_{\max}$  sur le nombre de ruptures
- Seuil de sélection de modèle  $\nu$

#### Traitement

1. Calculer les  $K_{\max}$  premiers coefficients non-nuls ( $\beta_{\tau_1}, \dots, \beta_{\tau_{K_{\max}}}$ ) apparaissant le long du chemin de régularisation du LAR/LASSO.
2. Lancer rDP sur l'ensemble des ruptures potentielles proposées ( $\tau_1, \dots, \tau_{K_{\max}}$ ).
3. Sélectionner le plus grand sous-ensemble de ruptures potentielles ( $\tau_1, \dots, \tau_{K_{\max}}$ ) ordonnées par rDP pour lequel  $\rho_k < 1 - \nu$ .

**Sortie** Estimation des instants de ruptures  $\hat{\tau}_1, \dots, \hat{\tau}_{\hat{K}}$ .

## 4 Discussion

Notons tout d'abord que notre nouvelle formulation de l'estimation d'instant de ruptures a été précédemment évoquée à plusieurs reprises, par exemple dans [12] et [3]. D'autre part, il est possible de montrer dans un cadre asymptotique bien choisi qu'avec une probabilité tendant vers 1 l'ensemble des instants de ruptures potentielles proposées par le LAR/LASSO contient les vrais instants de ruptures. Ceci conforte notre approche dans la mesure où il ne semble pas possible de garantir que l'ensemble des ruptures potentielles ne contiendra pas d'instant de ruptures "parasites". C'est pourquoi la version réduite de DP est indispensable pour affiner l'ensemble des ruptures potentielles.

## 5 Comparaison aux méthodes existantes

### 5.1 Simulations

Nous proposons d'évaluer les performances de notre méthode (**A**) et de la méthode décrite dans [11] (**B**) en termes de précision, rappel, et probabilité de fausse alarme [9] sur des données simulées. Nous mesurons les performances moyennes en simulant 30 configurations d'instant de ruptures, 30 jeux de valeurs de niveaux de segments, et 10 tirages de bruit blanc gaussien d'écart-type  $\sigma$ . Pour une configuration d'instant de ruptures et de niveaux de segments donnée, l'écart-type du bruit a été fixé à

$$\sigma = m \min_k (\mu_{k+1}^* - \mu_k^*)$$

pour différentes valeurs de  $m$ , où  $\mu_k^*$  est le niveau du  $k$ -ième segment. Les Tableaux 2, 3, et 4 résument les performances moyennes pour un nombre de ruptures  $K^* = 5, 15$  et un niveau de bruit  $m = 0.1, 0.5, 1.0, 1.5$  respectivement en termes de précision, rappel et fausse alarme. Les résultats révèlent que notre méthode **A** obtient des performances très compétitives. En effet, les performances en précision et probabilité de fausse alarme sont améliorées, tandis que les performances en rappel sont maintenues à un niveau équivalent à la méthode **B**.

$m =$	$K^* = 5$		$K^* = 15$	
	A	B	A	B
0.1	0.81±0.15	0.71±0.29	0.95±0.05	0.86±0.13
0.5	0.8±0.16	0.73±0.29	0.95±0.05	0.86±0.13
1.0	0.78±0.17	0.71±0.27	0.93±0.06	0.85±0.13
1.5	0.73±0.19	0.66±0.28	0.93±0.06	0.84±0.13

TAB. 2 – Performances en termes de précision

$m =$	$K^* = 5$		$K^* = 15$	
	A	B	A	B
0.1	0.99±0.02	0.99±0.02	0.99±0	0.99±0
0.5	0.98±0.04	0.99±0.03	0.99±0.01	0.99±0.01
1.0	0.95±0.08	0.94±0.08	0.97±0.03	0.97±0.04
1.5	0.85±0.16	0.87±0.15	0.94±0.06	0.94±0.06

TAB. 3 – Performances en termes de rappel

$m =$	$K^* = 5$		$K^* = 15$	
	A	B	A	B
0.1	0.13±0.03	0.23±0.2	0.34±0.02	0.42±0.13
0.5	0.13±0.03	0.22±0.2	0.33±0.02	0.41±0.13
1.0	0.13±0.03	0.21±0.18	0.33±0.02	0.4±0.13
1.5	0.13±0.03	0.21±0.2	0.31±0.03	0.4±0.15

TAB. 4 – Performances en termes de probabilité de fausse alarme

### 5.2 Données d'exploration géophysique

Nous proposons d'analyser les performances de notre méthode sur des données d'exploration géophysique (*well-log data*, cf. Figures 5 et 6) décrites dans [14] et [5]. Ces

données comprennent une proportion non-négligeable de données aberrantes, que nous éliminons à l'aide d'un filtrage médian [5]). Nous mettons ensuite en oeuvre Cachalot sur les données pré-traitées pour  $K_{\max} = 200$  et  $1 - \nu = 0.99$ . Les traits verticaux représentent les estimations des instants de ruptures. Les estimations sont très proches de celles indiquées dans [5](page 206), obtenues à l'aide d'un modèle probabiliste bayésien évolué.

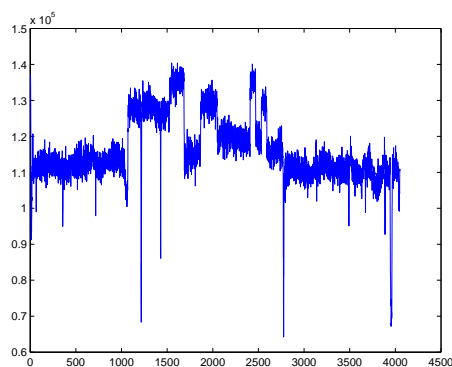


FIG. 5 – Données brutes

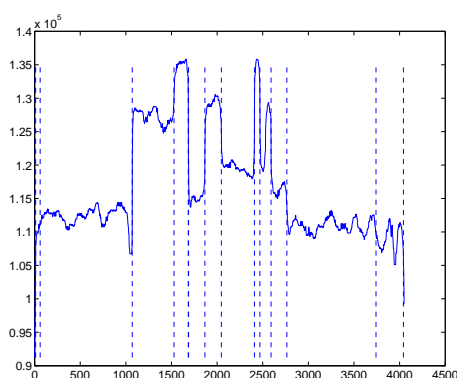


FIG. 6 – Données pré-traitées, et instants de ruptures proposés par Cachalot

## 6 Conclusion et perspectives

On a proposé une nouvelle méthode de détection de ruptures consistant à voir ce problème comme un problème de sélection de variables que l'on résout grâce à l'algorithme LAR/LASSO. Notre méthode s'avère performante par rapport aux méthodes existantes. On peut d'autre part montrer que, dans un cadre asymptotique bien choisi, notre méthode est consistante. Nous souhaitons maintenant obtenir des résultats théoriques plus précis et proposer une extension au cas multivarié.

## Références

[1] M. Basseville and N. Nikiforov. *The detection of abrupt changes*. Information and System sciences series. Prentice-Hall, 1993.

- [2] R. Bellman. On the approximation of curves by line segments using dynamic programming. *Communications of the ACM*, 4(6), 1961.
- [3] L. Boysen, A. Kempe, A. Munk, V. Liebscher, and O. Wittich. Consistencies and rates of convergence of jump penalized least squares estimators. *Annals of Statistics*, In revision.
- [4] B. Brodsky and B. Darkhovsky. *Non-parametric statistical diagnosis : problems and methods*. Kluwer Academic Publishers, 2000.
- [5] O. Cappe, E. Moulines, and T. Ryden. *Inference in Hidden Markov Models (Springer Series in Statistics)*. Springer-Verlag New York, Inc., 2005.
- [6] E. Carlstein, H. Mueller, and e. Siegmund, D. *Change-point Problems*. IMS Monograph 23. Institute of Mathematical Statistics, Hayward, CA, 1994.
- [7] B. Efron, T. Hastie, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32 :407–499, 2004.
- [8] W. D. Fisher. On grouping for maximum homogeneity. *Journal of the American Statistical Society*, 53 :789–798, 1958.
- [9] O. Gillet, S. Essid, and G. Richard. Sicvir : On the correlation of automatic audio and visual segmentation of music videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 2007.
- [10] S. M. Kay. *Fundamentals of statistical signal processing : detection theory*. Prentice-Hall, Inc., 1993.
- [11] M. Lavielle. Using penalized contrasts for the change-points problems. *Signal Processing*, 85(8) :1501–1510, 2005.
- [12] E. Mammen and S. Van De Geer. Locally adaptive regression splines. *Annals of Statistics*, 1997.
- [13] N. Meinshausen and B. Yu. Lasso-type recovery of sparse representations for high-dimensional data. Preprint 2006.
- [14] J. Ruanaidh and W. Fitzgerald. *Numerical Bayesian Methods Applied to Signal Processing*. Statistics and Computing. Springer, 1996.
- [15] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58(1) :267–288, 1996.
- [16] P. Zhao and B. Yu. On model selection consistency of lasso. *Journal Of Machine Learning Research*, 7, 2006.