

Algorithmes rapides de restauration de signaux avec prise en compte des discontinuités

Alain DE CESARE * et Rachid ZEBOUJ **

* INSERM U66, CHU Pitié - 91 Bd de l'Hôpital - 75013 PARIS - decesare@imed.jussieu.fr

** ENSEA-ETIS - 6 Avenue du Ponceau - 95014 CERGY-PONTOISE - zeboudj@ensea.fr

RÉSUMÉ

Cette communication présente des algorithmes pour la restauration des signaux stationnaires par morceaux. Une approche algorithmique basée sur la programmation dynamique conduit à des résultats rapides et optimaux. Deux algorithmes sont présentés. Le premier est d'une mise en œuvre simple, le second en est une amélioration permettant d'éviter l'énumération de la plupart des états qui ne contribuent pas à l'élaboration de la solution optimale. Ces deux algorithmes fournissent la solution exacte d'un problème d'optimisation non convexe en variable mixte avec des complexités algorithmiques respectives en $O(n^2)$ et $O(n)$.

ABSTRACT

This paper presents new algorithms for restoration of piecewise stationary signals restoration. A dynamic programming approach is used to obtain fast computations and optimal results. Two algorithms are presented : the first one is characterized by its easy implementation. The second algorithm derives from the previous one and is based on suppression of most state enumeration which are not used for calculating the optimal solution. These algorithms deliver the exact solution to a non-convex optimization problem. The time-complexities are of $O(n^2)$ and $O(n)$ respectively.

1 - Introduction

On présente de nouveaux algorithmes rapides et exacts pour restaurer des signaux comportant des discontinuités dont le nombre et la position sont a priori inconnus. De nombreux auteurs se sont intéressés à ce problème de restauration et ont proposé différents modèles et algorithmes aussi bien pour des signaux monodimensionnels que pour des images [1][2][3][4]. Nous proposons ici de nouveaux algorithmes d'estimation du maximum a posteriori (MAP) utilisant la programmation dynamique pour le modèle de la « chaîne faible » (« weak string ») introduit par Blake et Zisserman dans [1].

2 - Modèle de la chaîne faible

Ce modèle est une chaîne de Markov cachée du premier ordre pour laquelle la dépendance entre deux sites voisins est soit Gaussienne soit décorrélée, conditionnellement à un processus de ligne. Ce modèle s'applique à des signaux stationnaires par morceaux. La détermination du MAP permet de lisser les segments stationnaires (filtrage) tout en déterminant les points de rupture (segmentation).

On note $Y = \{y_0, \dots, y_{n-1}\}$ le signal observé, $X = \{x_0, \dots, x_{n-1}\}$ le signal restauré, et $L = \{l_1, \dots, l_{n-1}\}$ un vecteur le variables binaires (processus de ligne) indiquant la continuité ($l = 0$) ou la discontinuité ($l = 1$) entre deux données adjacentes (on appelle ces sites des transitions).

L'énergie du modèle est donnée par :

$$E(X, L) = \sum_{k=0}^{n-1} (x_k - y_k)^2 + \varepsilon \sum_{k=1}^{n-1} \left[(1 - l_k)(x_k - x_{k-1})^2 + l_k \cdot \chi \right] \quad (1)$$

Le premier terme, local à chaque donnée est un écart quadratique entre le signal et l'observation. Le second terme est une énergie interne constituée d'une somme d'énergies locales à chaque transition : soit il y a continuité, et la contribution à l'énergie du site est une fonction quadratique de l'écart entre les données adjacentes au site, soit il y a discontinuité et l'énergie locale prend une valeur fixée χ . Le terme ε est un paramètre de régularisation qui permet de fixer la part de chaque type d'énergie. La restauration consiste à trouver les vecteurs X et L qui minimisent $E(X, L)$. On trouvera dans [1] une description plus détaillée du modèle et une discussion sur le choix des paramètres. Un modèle plus général ainsi qu'une estimation des hyperparamètres ε et χ par un algorithme EM (Expectation-Maximisation) sont décrits dans [5].

3 - Algorithmique : état de l'art

L'estimation du MAP d'une chaîne faible consiste à résoudre un problème non trivial d'optimisation non convexe en variables mixtes. On peut utiliser différents types d'algorithmes pour trouver le couple (X,L) optimal :

- Des algorithmes généraux d'optimisation de fonctions non convexes : algorithmes stochastiques (recuit simulé) ou déterministes (GNC de complexité $O(n\sqrt{\varepsilon})$ [1]), de nature itérative, qui fournissent une solution sous-optimale. On trouvera dans [6] une comparaison entre ces algorithmes appliqués à la chaîne faible.
- L'algorithme de Viterbi qui nécessite une quantification des valeurs du signal X en M niveaux. On obtient alors la solution optimale pour la quantification choisie en $O(nM^2)$.
- L'« algorithme naïf », consistant à énumérer les 2^{n-1} vecteurs L possibles, et à calculer pour chacun d'eux le vecteur X optimal. On retiendra alors le couple (X_L, L) d'énergie minimale. Le temps de calcul de cet algorithme en $O(2^n)$ le rend impraticable pour des valeurs de n importantes.
- L'algorithme de Blake [6] basé sur la programmation dynamique donne la solution optimale avec un temps de calcul en $O(Kn^2)$, où K est un majorant du nombre de discontinuité. C'est dans la littérature, mis à part l'« algorithme naïf », le seul algorithme permettant d'obtenir la solution exacte. Du fait de sa relative lenteur (par rapport au GNC par exemple), Blake l'utilise comme référence pour tester la qualité de la solution fournie par d'autres algorithmes.

En résumé, il n'existe dans la littérature aucun algorithme à la fois rapide (linéaire) et exact pour résoudre le problème de minimisation de E(X, L).

4 - Un algorithme de complexité quadratique

Nous présentons dans un premier temps un algorithme en $O(n^2)$ qui utilise les résultats de [6]. L'énergie totale connaissant L peut s'écrire comme la somme des énergies sur chaque segment défini par L et de l'énergie due aux discontinuités. En posant conventionnellement $l_0 = l_n = 1$, les segments concernés sont les intervalles [i, j] tels que $l_i = l_{j+1} = 1$ et $\forall k \in]i, j]$ $l_k = 0$. L'énergie du modèle sur le segment [i, j] s'écrit :

$$E_{i,j} = \underset{X}{\text{Min}} \left[\begin{array}{l} \sum_{k=i}^j (x_k - y_k)^2 \\ + \varepsilon \sum_{k=i+1}^j (x_k - x_{k-1})^2 \end{array} \right] \quad (2)$$

Le point important est que l'on a une relation de récurrence permettant pour tout i, de calculer en $O(1)$ $E_{i,j}$ et x_j à partir de $E_{i,j-1}$ et de x_{j-1} , (avec $E_{i,i}(L) = 0$ et $x_i = y_i$), sans avoir besoin de calculer les autres valeurs du signal (il s'agit en fait d'un filtre de Kalman rapide [7]).

$$\begin{array}{l} E_0 = -\varepsilon\chi \\ \text{Pour } k = 1 \text{ à } n \\ \{ \\ \quad m[k] = \underset{j=0, \dots, k-1}{\text{Argmin}}(E_j + E_{j,k+1}) \\ \quad E_k = E_{m[k]} + E_{m[k],k-1} + \varepsilon\chi \\ \} \end{array}$$

Algorithme 1

On obtient l'algorithme 1 dans lequel, $m[k]$ est l'indice de la dernière discontinuité de la modélisation optimale du signal limité aux k premiers points et E_k son énergie. A l'itération k, on calcule l'énergie optimale du signal limité aux k premiers points. Pour cela, on détermine le dernier segment apparaissant dans le modèle optimal, en testant toutes les possibilités de dernière discontinuité, et en choisissant celle qui minimise l'énergie : le segment ($m[k], k-1$). On obtient alors les valeurs à 1 du vecteur L en remontant la liste chaînée des discontinuités par l'algorithme 2 :

$$\begin{array}{l} \text{Pour } k = 1 \text{ à } n-1 \\ \{ \\ \quad l[k] = 0 \\ \} \\ t = m[n] \\ \text{Tant que } t > 0 \\ \{ \\ \quad l[t] = 1 \\ \quad t = m[t] \\ \} \end{array}$$

Algorithme 2

Le calcul de X connaissant L se fait alors segment par segment. Le temps de calcul pour obtenir X est alors en $O(n)$, la solution du problème sans discontinuité se ramenant à un résoudre un système linéaire tridiagonal [8].

5 - Algorithme linéaire

L'algorithme précédent a l'avantage de la simplicité, mais il a l'inconvénient d'être quadratique. Au prix d'une implémentation un peu plus lourde, on va donner une version qui permet de réduire considérablement le nombre des énumérations.

Si on résout le problème (2) en contraignant x_j à prendre une valeur arbitraire x et que l'on fait la minimisation pour les autres variables (x_i, \dots, x_{j-1}) , on constate alors que l'énergie s'écrit comme une fonction quadratique de x : $E_{ij}(x) = f_{j-i+1} \cdot (x - x_j)^2 + E_{ij}$ (voir [6] et [8]). Dans cette expression le facteur f ne dépend que de la longueur du signal : $j-i+1$.

Pour le problème avec discontinuité, l'énergie optimale du signal limité aux k premiers points se met alors sous la forme suivante: $E_k(x) = \text{Min}_{j=0, \dots, k-1} (E_j + E_{j,k-1}(x)) + \epsilon\chi$.

Dans cette expression, l'indice j désigne la position de la dernière discontinuité. En fait seul un nombre très limité de valeurs de l'indice j sont utiles au calcul. La figure 1 illustre ce mécanisme (avec des notations simplifiées) : la fonction correspondant à l'indice $j = 2$, $E_2(x)$ est inutile pour le calcul de la fonction $E(x) = \text{Min}_j E_j(x)$, car pour tout point x elle est minorée soit par $E_1(x)$, soit par $E_3(x)$.

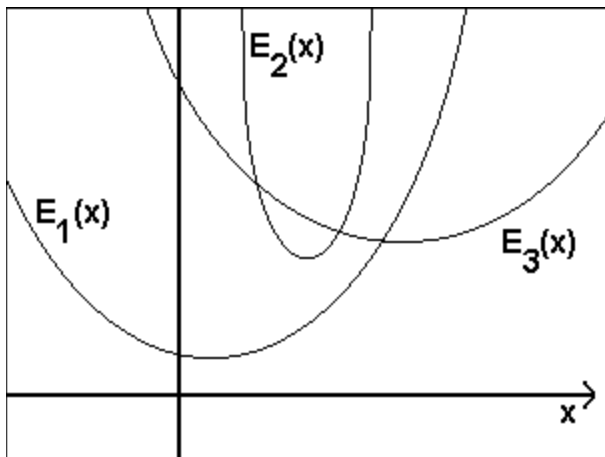


Figure 1 - Exemple d'indice inutile.

Il est nécessaire de conserver tous les indices contribuant à définir $E_k(x)$, et pas seulement celui qui permet d'obtenir le minimum à l'itération courante k . En effet les valeurs du signal n'ont pas encore été toutes utilisées par l'algorithme, et la valeur de x dans la solution optimale est inconnue. Par contre quelque soit les valeurs futures, l'énergie de la solution optimale sur les sites présent et passés ne dépendent que de x (processus Markovien du premier ordre) qui sera déterminé en fin d'algorithme, et cette énergie vaudra $E_k(x)$. Le fait que des indices j soit inutiles montrent que quelles que soient les valeurs futures du signal, la dernière

discontinuité précédent k ne peut pas être celle d'indice j . Il est alors immédiat que pour tous les successeurs de k , la dernière discontinuité ne peut pas non plus être j .

En résumé, à l'itération k et à chaque indice j est associé la fonction d'énergie $E_{jk}(x)$ et les indices inutiles sont ceux pour lesquels :

$$\forall x \in \mathfrak{X}, \exists i < k, E_{ik}(x) < E_{jk}(x).$$

On appelle J_k l'ensemble des indices utiles au calcul de la solution optimale à l'itération k . Dans l'algorithme 1 la dernière discontinuité optimale était recherchée parmi tous les sites de transition précédents. Ici on ne la recherchera que dans l'ensemble J_k (algorithme 3).

$J_0 = \emptyset$
 Pour $k = 1$ à n
 {
 $J_k = J_{k-1} \cup \{k-1\}$
 Supprimer dans J_k les indices inutiles
 $m[k] = \underset{j \in J_k}{\text{Argmin}} (E_j + E_{j,k-1})$
 $E_k = E_{m[k]} + E_{m[k],k-1} + \epsilon\chi$
 }

Algorithme 3

6 - Exemple

Un exemple de restauration est donné figures 2a à 2c pour un signal de synthèse obtenu en ajoutant un bruit blanc Gaussien à un signal constant par morceaux.

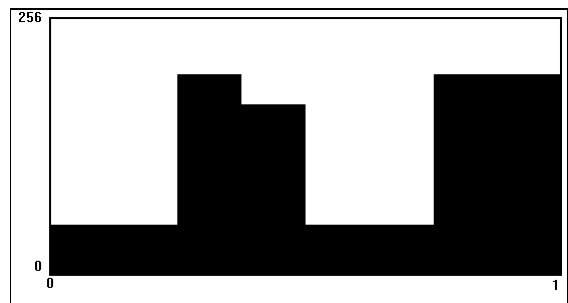


Figure 2a - Signal de synthèse non bruité

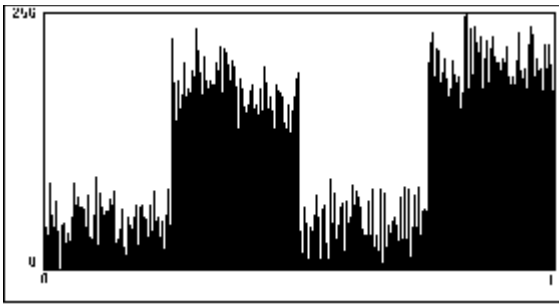


Figure 2b - Signal bruité à traiter

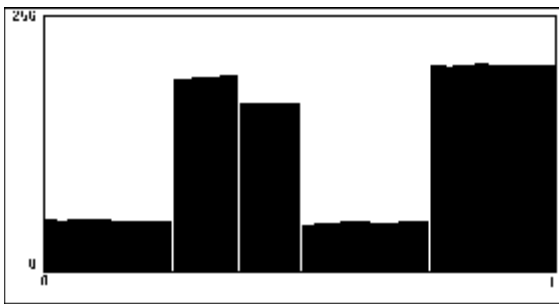


Figure 2c - Signal restauré avec indication de la position des 4 discontinuités détectées

[6] - A. Blake : Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction. IEEE transactions on Pattern Analysis and Machine Intelligence, vol 11, n°1, pp.2-12, 1989.

[7] - G. Demoment : Equations de Chandrasekhar et algorithmes rapides pour le traitement du signal et des images. Traitement du signal, vol 6, n°2, pp. 103-115, 1989.

[8] - A. De Cesare : Algorithmes rapides de restauration des signaux : Application à l'imagerie médicale. Thèse Paris XI-Orsay, 1996.

7 - Conclusion

Quand les valeurs des paramètres sont adaptés au signal, le nombre moyen d'indices conservés est de l'ordre de 3, à chaque itération. On constate expérimentalement que l'algorithme est linéaire : on traite environ 50000 points par seconde (sur un Pentium 133 sous Linux) pour des signaux du type de celui présenté figure 2. La preuve théorique de la linéarité reste un problème ouvert qui nécessiterait d'obtenir un majorant du cardinal de l'ensemble J_k indépendant de k et de la taille du signal.

8 - Références

[1] - A. Blake and A. Zisserman : Visual reconstruction. MIT Press, cambridge MA, 1987.

[2] - S. Geman and D. Geman : Stochastic relaxation, Gibbs distribution and bayesian restoration of images. IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 6, n°6, pp.721-741, 1984.

[3] - R.L. Stevenson, B.E. Schmitz and E.J. Delp : Discontinuity preserving regularization of inverse visual problems. IEEE transactions on systems, man and cybernetics, vol. 24, n°3, pp. 455-469, 1994.

[4] - D. Geman and C. Yang : Nonlinear image recovery with half-quadratic regularization. IEEE transactions on Image Processing, vol. 4, n° 7, 1995.

[5] - M. Fayolle and J. Idier : EM parameters estimation for a piecewise AR. Proc. ICASSP, pp.3545-3548, 1997.