

K-produits, un nouveau critère pour le traitement du signal et des images

MICHEL TERRE¹, LUC FETY¹, NICOLAS PAUL²

¹ Conservatoire National des Arts et Métiers
292 rue Saint Martin, 75003 Paris, France

² EDF R&D

6 quai Watier, 78401 Chatou, France

¹michel.terre@cnam.fr, luc.fety@cnam.fr,

²nicolas.paul@edf.fr

Résumé – Cet article passe en revue plusieurs domaines d'applications du critère de classification "*K-produits*" introduit dans [11]. Ce critère est appliqué à l'estimation de mélange de lois, à l'extraction de lignes droites dans des images binaires et à l'estimation aveugle de canal en radiocommunications. Plusieurs algorithmes d'optimisation du critère sont présentés et des comparaisons avec d'autres approches sont conduites.

Abstract – This article presents several applications of the clustering criterion "*K-products*" introduced in [11]. This criterion is applied to the unsupervised mixture estimation, the straight lines extraction in binary images and the blind channel estimation in radiocommunications. Some optimization algorithms are proposed and comparisons with other known methods are presented.

1 Introduction

Devoir classer un ensemble de N observations $\{x_i\}_{i \in [1, N]}$ dans K classes est un problème classique [1] (*clustering*) qui se rencontre dans de nombreux domaines d'applications du traitement du signal ou du traitement d'images.

Ce problème de classification est aussi souvent présenté comme l'estimation de mélanges de lois : Les mélanges de lois [2][3] constituent un outil général utilisé dans les problèmes de modélisation et d'apprentissage statistique.

Les méthodes classiques d'estimation de mélanges de lois, telles que l'algorithme EM [4] ou l'algorithme *K-means* [5][6], proposent des rechercher un ensemble de paramètres (espérances et variances des classes par exemple) qui optimisent des critères tels que la log-vraisemblance, éventuellement contrainte [7], ou la somme des carrés des distances par classe (*K-means*).

Les optima globaux de ces critères sont en général non biaisés, ils tendent vers l'ensemble des paramètres du mélange lorsque le nombre d'observations tend vers l'infini, mais ils sont difficiles à atteindre. En effet ces critères possèdent de nombreux points stationnaires différents de leur optimum global. En pratique, même dans le cas le plus simple où les supports des densités de probabilités des composantes du mélange sont disjoints, l'algorithme EM ou l'algorithme *K-means* doivent être relancés [8][9] plusieurs fois pour obtenir une estimation correcte des paramètres du mélange.

Pour résoudre ce problème, nous avons proposé d'introduire un nouveau critère, appelé *K-produits* [11]. Pour des observations de dimension 1 ou 2, il a été montré que *K-produits* n'admet aucun minimum local non global et que le minimum global de *K-produits* est accessible analytiquement.

Dans cet article nous commencerons au paragraphe 2 par rappeler la définition et les propriétés de ce critère

K-produits, puis nous exposerons trois domaines d'applications : l'estimation des mélanges de lois, l'extraction de ligne droites en traitement d'image et l'estimation aveugle de canal en radiocommunications.

2 *K-produits* : critère et minimisation

2.1 Observations réelles ou complexes

Partant d'un ensemble d'observations réelles ou complexes $\{x_i\}_{i \in [1, N]}$ d'un mélange de K composantes, l'algorithme proposé estime les espérances $\mathbf{m} = (m_1, \dots, m_K)$ des K lois. L'algorithme est basé sur la minimisation du critère polynomial $J_{KP}(\mathbf{u})$ défini par :

$$J_{KP}(\mathbf{u}) = \sum_{i=1}^N \prod_{k=1}^K |x_i - u_k|^2. \quad (1)$$

Une justification intuitive de ce critère provient du fait que, dans le cas limite où les variances des composantes sont nulles, alors toutes les observations sont égales à un des m_k . Dès lors $J_{KP}(\mathbf{m}) = 0$ et $J_{KP}(\mathbf{u})$, qui est un critère positif, est ainsi minimal pour $\mathbf{u} = \mathbf{m}$.

En développant l'équation (1), on obtient [11] :

$$J_{KP}(\mathbf{u}) = \sum_{i=1}^N \left| x_i^K + c_1 x_i^{K-1} + c_2 x_i^{K-2} + \dots + c_K \right|^2 \quad (2)$$

$$\text{Avec : } c_p = (-1)^p \sum_{i_1=1}^K \sum_{i_2=i_1+1}^K \dots \sum_{i_p=i_{p-1}+1}^K u_{i_1} u_{i_2} \dots u_{i_p}$$

La minimisation de $J_{KP}(\mathbf{u})$ est donc équivalente à la prédiction linéaire de x_i^K à partir de la base de prédiction : $\{x_i^{K-1}, x_i^{K-2}, \dots, x_i, 1\}$.

On introduit alors les moments estimés des observations :

$$m_{p,q} = \frac{1}{N} \sum_{i=1}^N x_i^p \bar{x}_i^q \quad (3)$$

La solution de la prédiction linéaire conduit à :

$$\begin{pmatrix} m_{0,0} & m_{0,1} & m_{0,2} & \dots & m_{0,K-1} \\ m_{1,0} & m_{1,1} & & & \\ m_{2,0} & & m_{2,2} & & \\ \vdots & & & \ddots & \\ m_{K-1,0} & & & & m_{K-1,K-1} \end{pmatrix} \begin{pmatrix} c_K \\ c_{K-1} \\ c_{K-2} \\ \vdots \\ c_1 \end{pmatrix} = \begin{pmatrix} m_{0,K} \\ m_{1,K} \\ m_{2,K} \\ \vdots \\ m_{K-1,K} \end{pmatrix} \quad (4.)$$

La résolution de ce système linéaire donne alors directement les termes $\{c_i\}_{i \in [1,K]}$. Le minimum de $J_{KP}(\mathbf{u})$, représenté par (\hat{m}_k) , est le vecteur des racines d'un polynôme de coefficients $\{c_i\}_{i \in [1,K]}$:

$$(\hat{m}_k) \equiv \text{Racines}(z^K + c_1 z^{K-1} + c_2 z^{K-2} + \dots + c_K) \quad (5.)$$

Application au mélange de lois

Les figures 1 et 2 ci-dessous montrent les minima du critère K -produits dans le cas d'observation d'un mélange de 5 lois Gaussiennes bidimensionnelles avec différentes matrices de covariances.

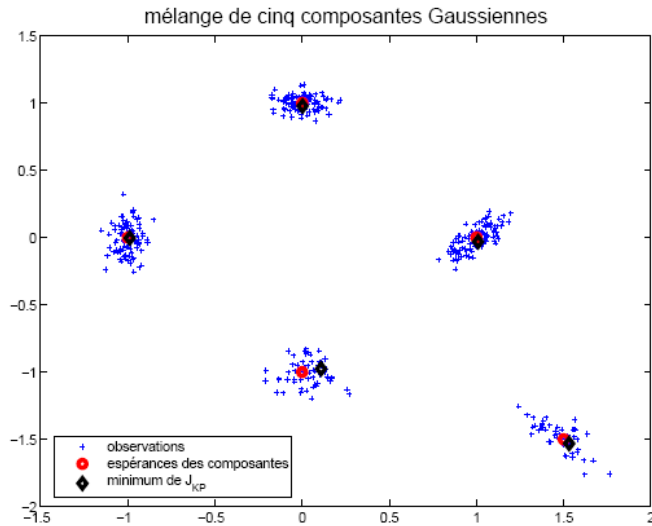


Figure 1 : Observations issues de 5 lois gaussiennes complexes (axes réels et imaginaires) de variance 0.1.

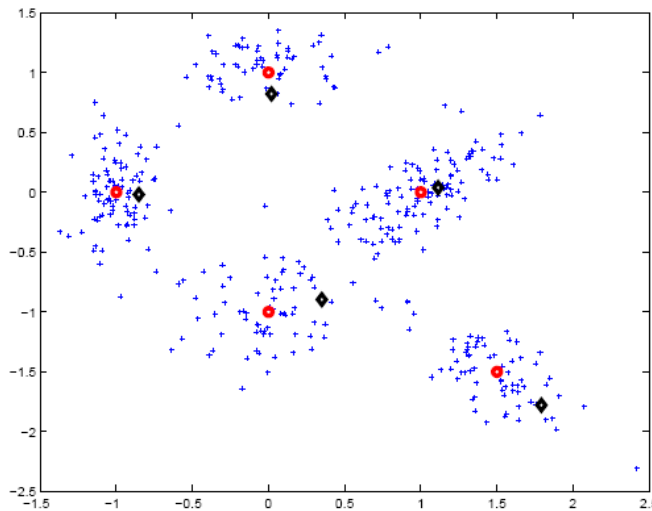


Figure 2 : Observations issues de 5 lois gaussiennes complexes (axes réels et imaginaires) de variance 0.2.

On observe un biais relativement faible qui dépend de la covariance des composantes du mélange. Pour minimiser ce biais, on peut recourir à une dernière étape

d'allocation des observations aux modes, au sens des plus proches voisins. Il suffit alors de recalculer les moyennes des classes pour obtenir une estimation nettement moins biaisée des espérances du mélange.

Concernant le biais il a aussi été montré [3] que dans le cas d'observations bidimensionnelles, pour $K=2$ et en considérant un mélange de deux lois gaussiennes complexes équiprobables, de moyennes et variances respectives $(a, \sigma_1^2 I_{2 \times 2})$ et $(-a, \sigma_2^2 I_{2 \times 2})$, alors la solution analytique de K -produits conduisait à $\hat{a} = a \left(\sqrt{1 + \rho^2} + \rho \right)$ avec $\rho = \frac{\sigma_2^2 - \sigma_1^2}{\|a\|^2 + \sigma_1^2 + \sigma_2^2}$. Dans ce cas particulier, on constate même que l'estimateur sera sans biais si $\sigma_1^2 = \sigma_2^2$.

2.2 Observations de dimensions supérieures à 2

La solution exposée précédemment fonctionne pour des observations unidimensionnelles (réelles) ou bidimensionnelles (complexes). Par contre elle ne se généralise pas directement aux dimensions supérieures. Dans de tels cas, on peut utiliser un algorithme de relaxation. Pour cela on écrit le critère de la manière suivante :

$$J_{KP}(\mathbf{u}) = \sum_{i=1}^N C_{k,i} \|x_i - u_k\|^2 \quad (6.)$$

$$\text{avec } C_{k,i} = \prod_{\substack{k'=1 \\ k' \neq k}}^K \|x_i - u_{k'}\|^2$$

On adapte alors itérativement les différentes estimations des modes. En notant "ite" l'indice de l'itération on a :

$$u_k^{ite} = \sum_{i=1}^N \left(\frac{C_{k,i}^{ite-1}}{\sum_{m=1}^K C_{k,m}^{ite-1}} \right) x_i \quad (7.)$$

L'unicité du minimum du critère a été prouvée pour les dimensions 1 et 2 [3], mais la preuve de cette unicité n'est pas encore démontrée pour les dimensions supérieures.

Jusqu'à maintenant les simulations, effectuées pour ces dimensions supérieures, n'ont jamais fait apparaître de minima secondaires.

2.3 Modèle paramétrique

Le critère K -product peut être étendu dans le cas de fonctions paramétriques.

$$J_{KP}(\mathbf{u}) = \sum_{i=1}^N \prod_{k=1}^K f_k(z_i) \quad (8.)$$

La fonction $f_k(\cdot)$ fait intervenir un vecteur de paramètres u_k . L'observation, notée ici z_i , est éventuellement multidimensionnelle. On peut illustrer ce cas avec le problème d'extraction de lignes droites à partir d'une image binaire.

On dispose alors de couples d'observations, qui sont les coordonnées : $z_i = (x_i, y_i)$, de points appartenant à des droites. On cherche un ensemble de K équations de droites qui "expliquent" au mieux cet ensemble de coordonnées. Le problème revient donc à identifier les K couples $u_k = \{a_k, b_k\}_{k \in [1, K]}$ tels que $y_i = a_k x_i + b_k$. Issu de K -produits, nous proposons donc le critère suivant :

$$J_{LKP}(a, b) = \sum_{i=1}^N \prod_{k=1}^K \frac{(y_i - a_k x_i - b_k)^2}{(1 + a_k^2)} \quad (9.)$$

Ce nouveau critère représente les produits des distances des observations aux K équations de droites. L'optimisation peut être effectuée au moyen d'un algorithme de relaxation qui cherche à déterminer successivement les paramètres des K droites.

Il a été montré [13] que cette stratégie offre une alternative intéressante à la Transformée de Hough [14] étendue au cas de plusieurs droites. Un exemple d'application est donné sur la figure 3.

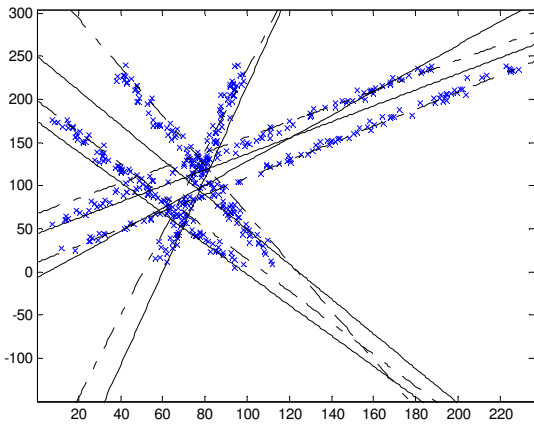


Figure 3 : 5 lignes droites, variance du bruit ajouté sur les coordonnées des pixels $\sigma^2=7$, Transformée de Hough en trait plein, LK-Produits en traits pointillé, observations représentées par des croix.

On constate sur cet exemple que la Transformée de Hough converge parfois vers des minima secondaires du critère qu'elle cherche à optimiser. Ceci se traduit, en définitive, par des équations de lignes droites qui se croisent, afin de correspondre au mieux aux observations.

3 Application à l'estimation aveugle de canal

On considère ici la réception de symboles de communications appartenant à une modulation d'amplitude et de phase *MDAQ* (*QAM*). Il s'agit d'identifier, sans séquence d'apprentissage ("en aveugle"), la réponse complexe H du canal de propagation. On reçoit ainsi des symboles x_i :

$$x_i = H s_i + w_i \quad (10.)$$

où w_i représente un bruit complexe additif et H la réponse du canal. Ce modèle simple correspond idéalement à l'observation faite sur une porteuse, à l'issue de la FFT de réception, dans le cas typique d'une transmission en OFDMA. On propose alors d'identifier la transformation qu'il faut appliquer à la constellation

de départ pour rendre compte au mieux des observations.

Si on note $(s_k)_{k \in [1, K]}$ les K symboles de cette constellation, alors on peut proposer une modification [15] du critère K -produits :

$$J_{HKP}(\hat{H}) = \sum_{i=1}^N \prod_{k=1}^K |x_i - \hat{H} s_k|^2 \quad (11.)$$

K représente alors le nombre de symboles de la modulation utilisée (4, 16, 64, ...).

Une approche relativement similaire a été proposée dans [16], pour le cas d'une réception sur plusieurs antennes. Dans cette référence les auteurs cherchent à agir sur les observations et non sur la constellation comme nous le faisons ici. Ils restreignent de plus leurs travaux au cas des modulations dont tous les symboles sont de même puissance : constellations circulaires de type modulation de phase *MDP* (*PSK*). Des équivalences avec le critère du maximum de vraisemblance sont proposées.

Pour minimiser, dans le cas général des modulations *MDAQ*, le critère proposé à l'équation (11), des propriétés de symétrie de ces modulations peuvent être utilisées. Nous savons que $\{s_k, j s_k, -s_k, -j s_k\}$ appartiennent à la constellation. On utilise alors la propriété suivante :

$$(x_i - \hat{H} s_k)(x_i - j \hat{H} s_k)(x_i + \hat{H} s_k)(x_i + j \hat{H} s_k) = x_i^4 - \hat{H}^4 s_k^4 \quad (12.)$$

On introduit ensuite la nouvelle observation $y_i = x_i^4$, le nouveau pseudo symbole $v_k = s_k^4$, et la variable complexe $z = \hat{H}^4$. Dès lors le critère de l'équation (11) devient ($M=K/4$) :

$$J_{HKP}(z) = \sum_{i=1}^N \prod_{k=1}^M |y_i - z v_k|^2 \quad (13.)$$

On introduit alors :

$$c_0 = 1, \quad c_1 = -\sum_{k=1}^M v_k, \quad c_2 = \sum_{i=1}^M \sum_{j=i+1}^M v_i v_j, \dots$$

$$c_p = (-1)^p \sum_{i_1=1}^M \sum_{i_2=i_1+1}^M \dots \sum_{i_p=i_{p-1}+1}^M v_{i_1} v_{i_2} \dots v_{i_p}, \quad c_M = \prod_{k=1}^M v_k,$$

$$m_{p,q} = \frac{1}{N} \sum_{i=1}^N y_i^p \bar{y}_i^q \quad \text{et} \quad a_{n,m} = c_n \bar{c}_m m_{M-n, M-m}.$$

Le critère peut alors être ré-exprimé sous la forme :

$$J_{HKP}(z) = \sum_{n=0}^M \sum_{m=0}^M a_{n,m} z^n \bar{z}^m \quad (14.)$$

A cette étape il est judicieux d'utiliser une représentation polaire $z = \rho e^{j\theta}$. On obtient alors :

$$J_{HKP}(\rho, \theta) = \sum_{n=0}^M \sum_{m=0}^M a_{n,m} \rho^{n+m} e^{j(n-m)\theta} \quad (15.)$$

La minimisation de $J_{HKP}(\rho, \theta)$ suivant les deux paramètres est alors une étape difficile. Une approche par relaxation entre ρ et θ peut alors être proposée. Elle conduit, après quelques changements de variables, aux solutions simples suivantes :

ρ connu

$$J_{HKP}(\theta) = \text{Re} \left\{ \sum_{m=0}^M a_{m,m} \rho^{2m} + 2 \sum_{t=1}^M \sum_{m=0}^{M-t} a_{m+t,m} \rho^{t+2m} e^{jt\theta} \right\} \quad (16.)$$

$$\hat{\theta} = \underset{\theta}{\text{Argmin}} \text{Re} \left\{ \text{IFFT} \left\{ \frac{\alpha_0}{2}, \alpha_1, \alpha_2, \dots, \alpha_M \right\} \right\} \quad (17.)$$

avec :

$$\alpha_t = 2 \sum_{m=0}^{M-t} a_{m+t,m} \rho^{t+2m} \quad (18.)$$

La phase estimée $\hat{\theta}$ est donc simplement donnée par la minimisation de la partie réelle d'une Transformée de Fourier Discrète Inverse (IFFT) :

θ connue

$$J(\rho) = \sum_{t=0}^{2M} \sum_{m=t-M}^t a_{t-m,m} \rho^t e^{j(t-2m)\theta} \quad (19.)$$

$$\hat{\rho} = \underset{\rho \in \{\text{real positive roots of } J(\rho)\}}{\text{Argmin}} \{J(\rho)\} \quad (20.)$$

avec :

$$\beta_t = \sum_{m=t-M}^t a_{t-m,m} e^{j(t-2m)\theta} \quad (21.)$$

On montre alors que, même avec très peu d'observations, cette approche permet d'identifier la réponse du canal avec précision. La figure 4 illustre la variance de l'estimation du canal en fonction de la taille de l'observation (N). Les performances sont comparées à celles d'un algorithme de référence [17] pour l'estimation aveugle de la phase du canal.

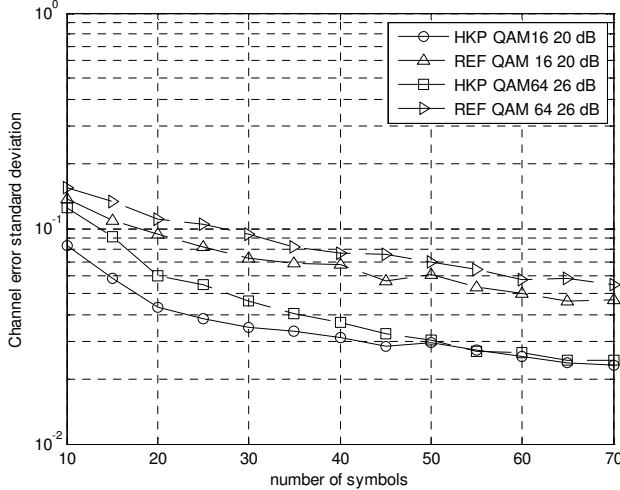


Figure 4 : Comparaisons des performances obtenues avec l'approche K-produits pour l'estimation du canal par rapport à une méthode de référence. Les courbes sont tracées en fonction du nombre de symboles observés.

4 Conclusion

Nous avons passé en revue un ensemble d'applications du critère K -produits pour la classification de données. Des résultats prometteurs ont été présentés pour l'application de ce critère dans des domaines aussi différents que l'estimation non supervisée des mélanges de lois, l'extraction de plusieurs droites dans des images ou l'estimation aveugle d'un canal en radiocommunications.

5 Références

- [1] Xu .R, Wunsch D., "Survey of clustering algorithms," *IEEE trans. on Neural Networks*, vol 16, pp 645-678, 2005.
- [2] O. Cappé, "Modèles de mélange et modèles de Markov cachés pour le traitement automatique de la parole," *notes de cours*, ENST, Paris, 2000.
- [3] N. Paul, "Une nouvelle approche pour l'estimation des mélanges de lois, application à la radio cognitive," *thèse de doctorat*, CNAM, Paris, soutenue le 14 mai 2008.
- [4] G. McLachlan, T. Krishnan, "The EM algorithm and extensions," John Wiley and Sons, New York, 1997.
- [5] J. Hartigan, M. Wong, "A K-means algorithm," *Journal of Applied Statistics*, vol 28, pp 100-108, 1979.
- [6] H. H. Bock, "Clustering methods: A history of the K-means algorithm," in *Selected contributions in data Analysis and Classification*, pp 161-172, Springer Verlag, Berlin 2007.
- [7] D. Ormoneit, V. Tresp, "Averaging maximum penalized likelihood and Bayesian estimation for improving Gaussian mixture probability density estimate," *IEEE trans. on Neural Network*, vol 9, pp 639-649, 1998.
- [8] C. Biernacki, G. Celeux, G. Govaert, "Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models," *Computational Statistics and Data Analysis*, vol 41, pp 561-575, 2003.
- [9] P.S. Bradley, U.M. Fayyad, "Refining initial points for K-means clustering," in *proc. of the 15th Int. Conf. on Machine Learning*, pp 91-99, San Francisco, 1998.
- [11] N. Paul, M. Terré, L. Féty, "A global algorithm to estimate the expectations of the components of an observed univariate mixture," *Advances in Data Analysis and Classification*, Volume 1, pp 201-219, 2007.
- [13] C. Bidot, N. Paul, L. Féty, M. Terré, "A new optimisation criterion for extracting several straight lines in a binary image," in *proc EUSIPCO*, Lausanne, August 2008.
- [14] Henri Maitre, "Contribution to the prediction of performances of the Hough transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 8, Issue 5, pp. 669-674, September 1986.
- [15] M. Terré, L. Féty, N. Paul, "Optimisation of a clustering-based criterion for blind channel estimation," in *proc. SPAWC 2009*, Perugia, June 2009.
- [16] O. Grellier, P. Comon, "Blind separation of Discrete Sources," *IEEE signal processing letters*, vol 5, n°8, pp 212-214 August 1998.
- [17] E. Serpedin, P. Ciblat, G. B. Giannakis, and P. Loubaton, "Performance Analysis of Blind Carrier Phase Estimators for General QAM Constellations," *IEEE trans. on Signal Processing*, vol. 49, no. 8, August 2001.