

Une méthode bayésienne pour la localisation et la séparation de sources de formes connues

Ali Mohammad-Djafari

Laboratoire des Signaux et Systèmes (CNRS-SUPELEC-UPS)

École Supérieure d'Électricité

Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France

Tél : 01-69 85 17 41 Fax : 01-69 41 30 60 E-mail : djafari@lss.supelec.fr

RÉSUMÉ

Nous considérons ici un problème de déconvolution où le signal d'entrée est supposé être une superposition d'un nombre fini de signaux de forme connue avec des paramètres de positions, d'amplitudes et d'échelles inconnus. Le problème devient alors celui de l'estimation de ces paramètres. La méthode proposée est basée sur l'estimation au sens du MAP avec des choix spécifiques pour des lois *a priori*. L'originalité du travail est plutôt dans ces choix. Il s'agit d'une méthode originale pour la résolution d'un problème peu classique qui comporte à la fois une étape de déconvolution et une étape de séparation de sources. L'optimisation du critère MAP est faite à l'aide d'un algorithme de descente du type Newton-Raphson après une initialisation appropriée. Quelques résultats de simulation illustrent les performances de la méthode.

ABSTRACT

We consider here a deconvolution problem where the input signal is assumed to be the superposition of a finite number of signals with known parametric expressions with unknown parameters of position, amplitude and scale. The problem then become the estimation of these parameters. The proposed method is based on the MAP estimation approach with specific choices for the prior laws. The originality of this work is more in this choice. The main objective of this work is to propose an original method for giving a solution to an uncommon problem which contains both a deconvolution step and a source separation step. The proposed algorithm is a descent algorithm based on Newton-Raphson scheme after an appropriate initialisation. Some simulation results show the performances of the proposed method.

1 Introduction

Dans grand nombre d'applications du traitement du signal, le signal mesuré $y(t)$ peut être modélisé comme :

$$y(t) = h(t) * x(t) + b(t) = [h * x](t) + b(t), \quad (1)$$

où $x(t)$ est le signal d'entrée, $h(t)$ est la réponse impulsionnelle du système de mesure et $b(t)$ représente le bruit. La déconvolution consiste à estimer $x(t)$ à partir d'un nombre limité de mesures $y(t_i)$. Les solutions proposées par la plupart des méthodes classiques peuvent être considérées comme des cas particuliers des solutions que l'on obtient en minimisant un critère régularisé [1] de la forme :

$$J(x) = Q(x) + \lambda \Omega(x), \quad (2)$$

avec

$$Q(x) = \frac{1}{2\sigma_b^2} \int |y(t) - [h * x](t)|^2 dt \quad (3)$$

et

$$\Omega(x) = \int \phi \left(\sum_k p_k(t) x^{(k)}(t) \right) dt, \quad (4)$$

où $x^{(k)}(t)$ est la k -ième dérivée de $x(t)$, $p_k(t)$ sont des fonctions positives connues et $\phi(\cdot)$ une fonction positive, symétrique et non-décroissante, comme par exemple $\phi(\cdot) = |\cdot|^2$.

Dans certaines applications comme en géophysique [2, 3], en radar [4] ou en mesures physico-chimiques [5], nous savons que le signal $x(t)$ est constitué de la superposition d'un nombre fini de signaux de formes connues comme par exemple :

$$x(t) = \sum_{k=1}^K a_k g_k(t - t_k), \quad (5)$$

avec

$$g_k(t) = \frac{1}{\sqrt{2\pi s_k}} \exp \frac{-t^2}{2s_k} \quad \text{ou} \quad g_k(t) = \frac{s_k^2}{s_k^2 + t^2}. \quad (6)$$

Le problème de la déconvolution devient ainsi celui de l'estimation des paramètres $\{a_k, t_k, s_k, k = 1, \dots, K\}$ et leur nombre K . Nous proposons de traiter ce problème en utilisant l'approche bayésienne et plus spécifiquement l'estimation au sens du maximum *a posteriori* (MAP).

2 Méthode proposée

Considérons le modèle d'observation :

$$y(t) = h(t) * x(t; \mathbf{a}, \mathbf{t}, \mathbf{s}) + b(t), \quad (7)$$

où $x(t)$ est de la forme :

$$x(t; \mathbf{a}, \mathbf{t}, \mathbf{s}) = \sum_{k=1}^K a_k g_k(t - t_k), \quad (8)$$

avec

$$g_k(t) = \exp \frac{-t^2}{2s_k}. \quad (9)$$

et le problème de l'estimation des paramètres $\mathbf{a}, \mathbf{t}, \mathbf{s}, K$.

Notons tout de suite qu'il y a un lien entre ce problème et ceux de la séparation de source d'un côté et celui de la déconvolution impulsionnelle de l'autre côté. En effet, dans le cas où $h(t) = \delta(t)$ on retrouve le problème de la séparation de sources et, si de plus $s_1 = s_2 = \dots = s_K = s$, une constante connue, on retrouve un problème de la déconvolution

impulsionnelle où l'entrée est caractérisée par les impulsions $\{(t_k, a_k), k = 1, \dots, K\}$ et le filtre par la réponse impulsionnelle $g(t) = \exp\frac{-t^2}{2\sigma}$.

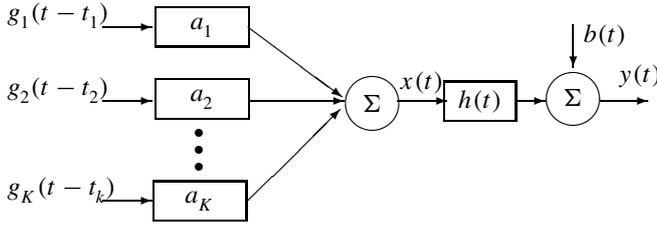


Fig. 1. Modèle de génération de données

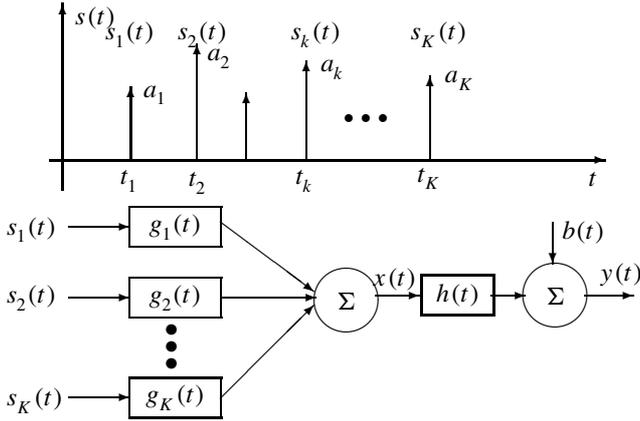


Fig. 2. Modèle de séparation de sources

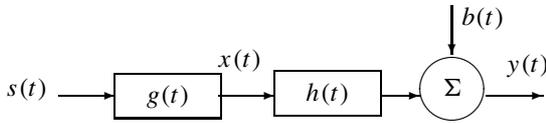


Fig. 3. Modèle de déconvolution impulsionnelle

Revenons au problème dans son cadre général. Dans un premier temps supposons que K soit connu (mais nous reviendrons sur ce point plus tard) et que nous cherchons à estimer $(\mathbf{a}, \mathbf{t}, \mathbf{s})$. Nous proposons alors d'effectuer cette tâche à l'aide de l'approche bayésienne et plus précisément à l'aide de l'estimation au sens du MAP :

$$(\hat{\mathbf{a}}, \hat{\mathbf{t}}, \hat{\mathbf{s}}) = \arg \max_{(\mathbf{a}, \mathbf{t}, \mathbf{s})} p(\mathbf{a}, \mathbf{t}, \mathbf{s} | \mathbf{y}) \quad (10)$$

où

$$\ln p(\mathbf{a}, \mathbf{t}, \mathbf{s} | \mathbf{y}) = cte + \ln p(\mathbf{y} | \mathbf{a}, \mathbf{t}, \mathbf{s}) + \ln p(\mathbf{a}) + \ln p(\mathbf{t}) + \ln p(\mathbf{s}),$$

si on suppose \mathbf{a} , \mathbf{t} et \mathbf{s} indépendants.

Pour attribuer ces différentes loi de probabilités *a priori*, nous faisons les hypothèses suivantes :

- Le bruit $b(t)$ est supposé centré, blanc, gaussien et de variance connue σ_b^2 . Cette hypothèse nous donne le terme de vraisemblance :

$$p(\mathbf{y} | \mathbf{a}, \mathbf{t}, \mathbf{s}) \propto \exp -Q(\mathbf{a}, \mathbf{t}, \mathbf{s}) ;$$

avec

$$Q(\mathbf{a}, \mathbf{t}, \mathbf{s}) = \frac{1}{2} \int |y(t) - h(t) * x(t) ; \mathbf{a}, \mathbf{t}, \mathbf{s}|^2 dt. \quad (11)$$

- Les amplitudes $a_k > 0$ sont supposées *a priori* indépendantes et suivent une loi Gamma :

$$p(\mathbf{a}) = \prod_{k=1}^K p_k(a_k),$$

avec

$$p_k(z) = \frac{1}{\Gamma(\alpha)} z^{\alpha-1} \exp -z, \quad \alpha > 0,$$

Cette hypothèse assure que les amplitudes soient positives et le paramètre α contrôle leur moyenne et leur variance. En effet, la moyenne et la variance d'une variable aléatoire Z suivant cette loi sont :

$$EZ = \text{Var } Z = \alpha.$$

- Les positions $t_k > 0$ sont supposées former une chaîne de Markov :

$$p(\mathbf{t}) = p(t_1) \prod_{k=2}^K p(t_k | t_{k-1}),$$

avec

$$p(t_1) = \frac{1}{\Gamma(\beta)} t_1^{\beta-1} \exp -t_1, \quad \beta > 0,$$

et

$$p(t_k | t_{k-1}) = \frac{1}{\Gamma(\beta)} (t_k - t_{k-1})^{\beta-1} \exp -(t_k - t_{k-1}),$$

Cette hypothèse assure que les positions $t_k > 0$ ainsi que $t_k - t_{k-1} > 0$ et le paramètre β contrôle leur moyenne et leur variance.

- Les variances $s_k > 0$ sont aussi supposées suivre une loi Gamma :

$$p(\mathbf{s}) = \prod_{k=1}^K p_k(s_k)$$

avec

$$p_k(z) = \frac{1}{\Gamma(\gamma)} z^{\gamma-1} \exp -z, \quad \gamma > 0,$$

Là aussi, cette hypothèse assure la positivité des $s_k > 0$ et le paramètre γ contrôle leur moyenne et leur variances.

Avec ces hypothèses il n'est pas difficile de montrer que l'estimation au sens du MAP de ces paramètres devient :

$$(\hat{\mathbf{a}}, \hat{\mathbf{t}}, \hat{\mathbf{s}}) = \arg \min_{(\mathbf{a}, \mathbf{t}, \mathbf{s})} J(\mathbf{a}, \mathbf{t}, \mathbf{s}) \quad (12)$$

où

$$J(\mathbf{a}, \mathbf{t}, \mathbf{s}) = \frac{1}{\sigma_b^2} Q(\mathbf{a}, \mathbf{t}, \mathbf{s}) + H_1(\mathbf{a}) + H_2(\mathbf{t}) + H_3(\mathbf{s}) \quad (13)$$

et

$$H_1(\mathbf{a}) = (1 - \alpha) \sum_{k=1}^K \log(a_k) + \sum_{k=1}^K a_k,$$

$$H_2(\mathbf{t}) = (1 - \beta) \sum_{k=1}^K \log(t_k - t_{k-1}) + \sum_{k=1}^K (t_k - t_{k-1}), \text{ avec } t_0 = 0,$$

$$H_3(\mathbf{s}) = (1 - \gamma) \sum_{k=1}^K \log(s_k) + \sum_{k=1}^K s_k.$$

Les différents composantes Q , H_1 , H_2 et H_3 sont dues, respectivement, aux choix de la loi du bruit, de $p(\mathbf{a})$, de $p(\mathbf{t})$ et de $p(\mathbf{s})$.

Le principal problème est maintenant l'optimisation effective de ce critère qui n'est pas forcément unimodal. Nous pouvons cependant calculer analytiquement son gradient et sa matrice

Hessienne. On peut alors envisager une méthode d'optimisation locale avec une initialisation appropriée en espérant que celle-ci se trouve dans le domaine d'attraction du minimum global. On peut aussi faire recours aux techniques de recuit simulé ou de relaxation déterministes (algorithme du type 'Iterative Conditional Descent' (ICD)) où on ne modifie qu'un seul paramètre à la fois. Dans ce travail, nous avons examiné les deux types d'algorithmes.

Dans le cas d'un algorithme du type gradient l'initialisation devient critique. Pour ceci, nous proposons d'utiliser une méthode de déconvolution classique (par exemple la régularisation au sens du Tikhonov éventuellement avec la contrainte de positivité) pour obtenir une solution initiale pour le signal d'entrée $x(t)$. Ensuite, partant de cette solution, nous pouvons extraire le nombre de sources K et une estimation initiale pour les paramètres $(\mathbf{a}, \mathbf{t}, \mathbf{s})$. Bien entendu, il n'y a aucune garantie pour que cette solution initiale soit dans le domaine d'attraction du minimum global du critère. Cependant, les résultats de simulations ont montré que la méthode proposée fournit des résultats suffisamment fiables et robustes vis-à-vis du bruit.

3 Algorithmes d'optimisation proposés

3.1. Optimisation locale L'algorithme proposé est une version modifiée de l'algorithme de Newton-Raphson appliqué d'une manière successive à chaque groupe de paramètres \mathbf{a} , \mathbf{t} et \mathbf{s} . Cet algorithme peut être résumé ainsi :

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} + \Delta\boldsymbol{\theta}^{(n)} \text{ avec } \boldsymbol{\theta} = \mathbf{a}, \mathbf{t} \text{ ou } \mathbf{s} \quad (14)$$

et où :

$$\Delta\boldsymbol{\theta}^{(n)} = -\lambda [\mathbf{D}^{(n)}]^{-1} \nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}^{(n)}), \quad (15)$$

et

$$\mathbf{D}^{(n)} = \text{diag} \left[\frac{\partial^2 J}{\partial \theta_1^2}(\boldsymbol{\theta}^{(n)}), \dots, \frac{\partial^2 J}{\partial \theta_m^2}(\boldsymbol{\theta}^{(n)}) \right] \quad (16)$$

Pour donner quelques détails de calcul des différents termes et de la mise en œuvre de l'algorithme, notons :

$$\delta y(t) = y(t) - h(t) * x(t) = y(t) - \sum_{k=1}^K [h * g_k](t - t_k). \quad (17)$$

On a alors les relations suivantes :

$$Q(\mathbf{a}, \mathbf{t}, \mathbf{s}) = \frac{1}{2} \int |\delta y(t)|^2 dt \quad (18)$$

$$\frac{\partial Q}{\partial \theta_k} = - \int \left[h * \frac{\partial g_k}{\partial \theta_k} \right] (t - t_k) \delta y(t) dt$$

$$\frac{\partial^2 Q}{\partial \theta_k^2} (=) - \int \left[h * \frac{\partial^2 g_k}{\partial \theta_k^2} (-) \left(h * \frac{\partial g_k}{\partial \theta_k} \right)^2 \right] (t - t_k) \delta y(t) dt$$

$$\text{avec } \theta_k = a_k, t_k \text{ ou } s_k.$$

(19)

3.2. Optimisation globale La principale idée ici est de modifier un paramètre à la fois et vérifier si le critère diminue. Si ceci est le cas on accepte la modification, sinon on l'accepte avec une probabilité décroissante. Le principal avantage est qu'ici il n'est pas nécessaire de calculer le gradient du critère, mais son inconvénient est qu'elle nécessite, en général, un nombre beaucoup plus important d'itérations.

4 Sélection du modèle

Bien qu'il existe un très grand nombre de méthodes théoriques [6] basées sur la théorie de l'information (entropie, entropie croisée) ou de codage (minimum length description code), nous proposons ici de choisir l'ordre K du modèle par :

$$\hat{K} = \arg \min_k J_k(\hat{\mathbf{a}}, \hat{\mathbf{t}}, \hat{\mathbf{s}}) \quad (20)$$

où

$$(\hat{\mathbf{a}}, \hat{\mathbf{t}}, \hat{\mathbf{s}}) = \arg \min_{(\mathbf{a}, \mathbf{s}, \mathbf{t})} J_k(\mathbf{a}, \mathbf{s}, \mathbf{t}) \quad (21)$$

et où $J_k(\mathbf{a}, \mathbf{s}, \mathbf{t})$ est donnée par

$$J_k(\mathbf{a}, \mathbf{t}, \mathbf{s}) = \frac{1}{\sigma_b^2} Q(\mathbf{a}, \mathbf{t}, \mathbf{s}) + H_1(\mathbf{a}) + H_2(\mathbf{t}) + H_3(\mathbf{s}) + k[\Gamma(\alpha) + \Gamma(\beta)\Gamma(\gamma)].$$

Ceci signifie que nous devons répéter l'algorithme décrit plus haut pour K allant de K_{\min} à K_{\max} . Ceci peut être interprété comme l'estimation au sens du MAP de K avec une loi *a priori* uniforme pour K . Il est cependant possible de choisir une autre loi *a priori* (par exemple une loi binomiale) pour K , ce qui revient à pondérer J_k dans le schéma précédent.

En pratique, nous avons utilisé l'algorithme suivant :

Pour $k = 1 : K_{\max}$

à – Initialiser $\boldsymbol{\theta}_k$ à partir des valeurs obtenues à l'étape $k - 1$ et une nouvelle valeur obtenue à partir de la position du maximum du résidu δy

– Estimer $\boldsymbol{\theta}_k$ en minimisant J_k par un algorithme de descente du gradient et noter la valeur optimale de J_k

Fin de boucle.

à À la fin des itérations choisir la valeur de k correspondant au minimum des J_k et la solution $\boldsymbol{\theta}_k$ qui en résulte.

5 Simulations

Pour montrer les performances de la méthode proposée nous avons simulé un signal d'entrée $x(t)$ avec les paramètres : $\mathbf{a} = [1; 2; 1.4]$, $\mathbf{t} = [200; 250; 300]$, $\mathbf{s} = [10; 15; 20]$ une réponse impulsionnelle $h(t)$ et calculé $[h * x](t)$ auquel nous avons ajouté un bruit blanc gaussien pour obtenir $y(t)$ avec SNR=20dB (Fig. 4).

Fig. 5, 6 et 7 montrent les différentes étapes de la méthode proposée : le résultats de la déconvolution, initialization, estimation finale et l'évolution du critère au cours des itérations.

Pour s'assurer que la méthode marche convenablement, nous avons construit un deuxième signal $x(t)$ avec les paramètres $\mathbf{a} = [1; 2; 1.4; .7]$, $\mathbf{t} = [200; 250; 300; 350]$, $\mathbf{s} = [10; 15; 20; 10]$

et de nouveau appliqué la méthode proposée sans aucune modification. On peut constater que la méthode a bien trouvé le nombre de sources et que le résultat de l'estimation des paramètres semble bien correcte.

Références

- [1] G. Demoment, "Image reconstruction and restoration : Overview of common estimation structure and problems," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-37, pp. 2024–2036, December 1989.
- [2] J. Fuchs and H. Chuberre, "A Deconvolution Approach to Source Localization," *IEEE Transactions on Signal Processing*, vol. SP-42, pp. 1462–1470, June 1994. FD.

- [3] H. El-Sherief, "Adaptive least squares for parametric spectral estimation and its application to pulse estimation and deconvolution of seismic data," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-16, pp. 299–303, 1986.
- [4] Holmes, "Blind deconvolution of quantum-limited incoherent imagery : Maximum-likelihood approach," *Journal of the Optical Society of America*, vol. 9, July 1993.
- [5] A. Kumar, C. Sotak, C. du Moulin, and G. Levy, "Software for deconvolution of overlapping spectral peaks and quantitative analysis by 13c fourier transform NMR spectroscopy," *Comput. Enhanced Spectrosc. (GB)*, vol. 1, pp. 107–114, Apr. 1983.
- [6] Q. T. Zhang, K. M. Wong, P. C. Yip, and J. L. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 37, pp. 1557–1567, octobre 1989.

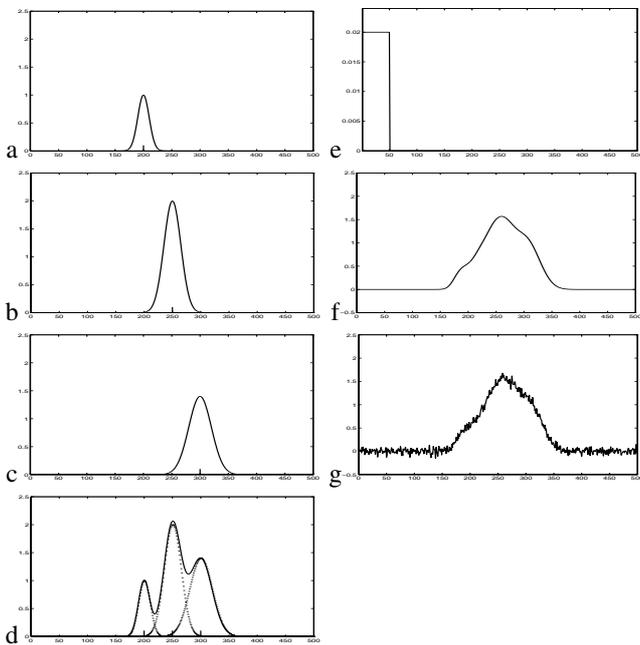


Fig. 4. Différentes étapes de construction des signaux :
 a,b,c,d) $g_1(t), g_2(t), g_3(t)$ et $x(t)$;
 e,f,g) $h(t), y(t)$ sans et avec bruit

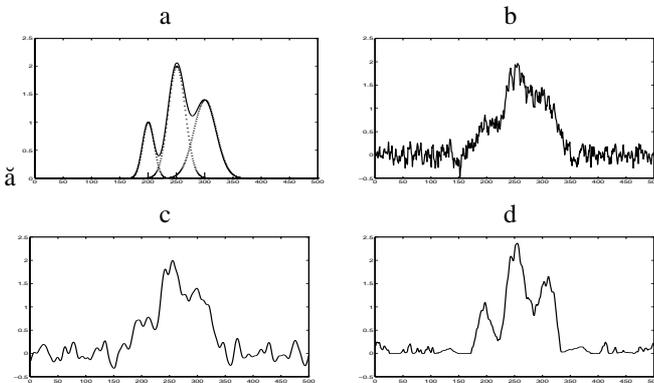


Fig. 5. Résultat de la déconvolution directe par différentes méthodes :
 a) Signal $x(t)$,
 b) Déconvolution par filtre de Wiener,
 c) Déconvolution par régularisation quadratique,
 d) Déconvolution par régularisation entropique

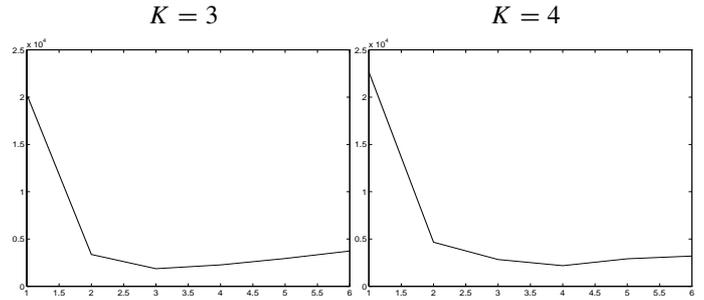


Fig. 7. Critères J_k en fonction du k :
 Les valeurs du critères J_k en fonction du nombre de source K dans deux cas.

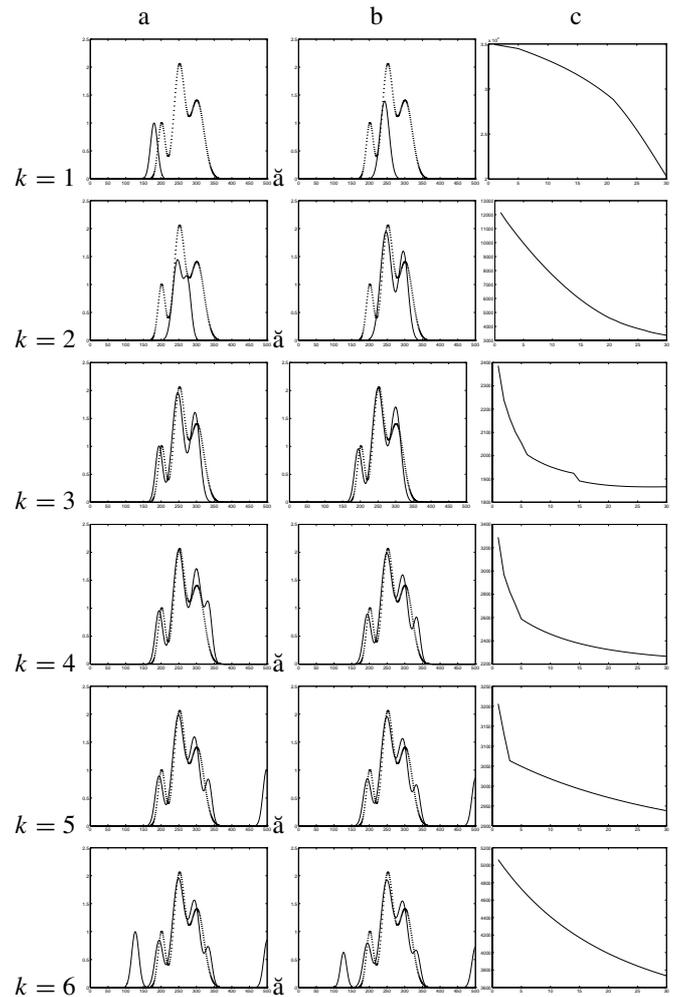


Fig. 7. Résultat de séparation pour $k = 1, \dots, 6$:
 a) Initialisation, b) Résultat final dans chaque étape
 c) Évolution du critère au cours des itérations.