

Séparation et déconvolution aveugle de signaux bruités : modélisation par mélange de gaussiennes.

Olivier Bermond, Éric Moulines, Jean-François Cardoso.

ENST, département Signal / CNRS URA 820, 46 rue Barrault, 75634 Paris cedex 13, France

(bermond, moulines, cardoso@sig.enst.fr)

RÉSUMÉ

Cet article propose un algorithme d'identification aveugle de systèmes linéaires bruités. Le principe est celui de l'EM, dont le but est de maximiser la vraisemblance de l'observation en maximisant une fonction auxiliaire, l'étape 'M' de maximisation est simple, la difficulté restant l'étape 'E' de calcul de cette fonction auxiliaire. Nous montrons dans l'article que la modélisation de la distribution des entrées par un mélange de Gaussiennes permet une implantation de l'étape 'E' de manière exacte et (raisonnablement) rapide. Nous fournissons un procédé d'estimation *conjointe* des paramètres du système et de l'entrée, qui se généralise au problème de déconvolution aveugle, et qui offre une voie à l'estimation semi-paramétrique.

ABSTRACT

This submission proposes an algorithm for blind identification of noisy linear systems. It follows the principle of the EM algorithm, whose aim is to maximize the likelihood of the observations, by iterative maximization of an auxiliary function. While the M-step is quite simple, the E-step remains a difficult task in many situations. We will show that modeling the input distribution as a mixture of Gaussian allows an exact implementation of the E-step, with a reasonably small numerical cost. Our algorithm is able to *jointly* estimate the parameters of the system as well as the parameters of the input; it is extended to the case of blind deconvolution, and offers a route to semi-parametric estimation.

1 Introduction

Nous nous intéresserons au modèle général de déconvolution aveugle, en précisant l'algorithme pour la séparation de sources, puis en indiquant l'extension au cas du modèle de déconvolution. On observe $\{x_t\}$ (vecteur $m \times 1$) supposé suivre le modèle :

$$x_t = \sum_{k=0}^L A(k)s_{t-k} + b_t$$

Où les matrices $A(k)$ sont inconnues, $\{s_t\}$ est l'entrée (non observée), processus i.i.d. de vecteurs $n \times 1$ dont les composantes sont indépendantes, de densité $p(s, \eta) = \prod_{i=1}^n p_i(s_i, \eta_i)$, et $\{b_t\}$ est le bruit (non observé) lié à l'observation, processus i.i.d. supposé gaussien ; $b_t \sim \mathcal{N}(0, J)$, J étant définie positive. On suppose en outre que $\{s_t\}$ et $\{b_t\}$ sont indépendants.

Le problème est donc l'estimation du paramètre $\theta = (A(0), \dots, A(L), J, \eta)$. Plusieurs méthodes ont déjà été proposées pour le résoudre (cf [1, 2, 3, 4]). Notre approche est celle du maximum de vraisemblance, pour laquelle l'algorithme EM est une méthode itérative particulièrement bien adaptée (cf [5]). Cet algorithme se base sur les données complètes du modèle, en l'occurrence $\{x_t, s_t\}$, qui, si elles étaient observées permettraient une simple identification entrée-sortie. La première partie de cette contribution explicite l'utilisation de l'algorithme EM pour la séparation de sources, dans le cas où la loi d'entrée est modélisée par un mélange de lois Gaussiennes. Nous indiquerons dans un deuxième temps l'extension au cas de la déconvolution aveugle par l'introduction d'un processus de contraste où l'observation est séparée par paquets. Nous présenterons enfin quelques simulations.

2 Séparation de sources

Nous considérons le modèle de séparation de sources, c'est à dire le cas où $L = 0$. On observe alors :

$$x_t = A s_t + b_t$$

On se propose d'estimer *conjointement* les paramètres du mélange (A, J) (paramètres d'intérêt) et le paramètre de l'entrée η (paramètre de nuisance), à partir de l'observation $\mathbf{x} = (x_1, \dots, x_T)'$.

Si l'entrée $\mathbf{s} = (s_1, \dots, s_T)'$ était observée, le problème serait celui de l'identification entrée-sortie, et dans ce cas l'estimateur au maximum de vraisemblance s'obtient à l'aide des statistiques :

$$R_{xs} = \frac{1}{T} \sum_t x_t s_t' \quad \text{et} \quad R_{ss} = \frac{1}{T} \sum_t s_t s_t'$$

Seule l'observation de la sortie étant disponible, l'estimateur au maximum de vraisemblance est calculé récursivement en remplaçant les signaux d'entrée par leurs espérances conditionnelles ; il s'agit de l'**algorithme EM** dont les étapes sont les suivantes :

- **Étape E** : calcul de la fonctionnelle

$$Q(\theta, \theta') = E[\log p(\mathbf{x}, \mathbf{s}, \theta) | \mathbf{x}, \theta']$$

C'est l'étape la plus délicate ; puisque la distribution à posteriori $p(s|x, \theta')$ est en général difficile à estimer. Une possibilité est la technique d'intégration Monte-Carlo (cf [6]). Nous allons voir que la modélisation de $q(s, \eta)ds$ par un mélange de Gaussiennes permet un calcul exact.

– **Etape M** : Maximisation de la fonctionnelle ;

$$\theta_{k+1} = \text{Arg max}_{\theta} Q(\theta|\theta_k)$$

Remarquons que la maximisation de la fonctionnelle Q à chaque étape entraîne un accroissement de la vraisemblance, ce qui constitue un point essentiel de l'algorithme EM.

$\{x_t\}$ est un processus i.i.d. et la loi jointe de l'entrée et de l'observation est de forme simple :

$$\begin{aligned} \log p(\mathbf{x}, \mathbf{s}, \theta) &= \sum_{t=1}^T \log p(x_t|s_t, \theta)q(s_t, \eta) \\ &= \sum_{t=1}^T \log \phi(x_t, As_t, J) + \sum_{t=1}^T \log q(s_t, \eta) \end{aligned}$$

où l'on a noté (avec $|M| = \det M$) :

$$\log \phi(x, m, R) = -\frac{1}{2}(x - m)'R^{-1}(x - m) - \frac{1}{2} \log |2\pi R|$$

La réestimation des paramètres (A, J) à chaque étape fait intervenir la loi de l'entrée à postérieure, et plus précisément les deux quantités :

$$\begin{aligned} \hat{R}_{xs} &= \frac{1}{T} \sum_t x_t E[s_t|x_t, \theta]' & (1) \\ \hat{R}_{ss} &= \frac{1}{T} \sum_t E[s_t s_t'|x_t, \theta]. \end{aligned}$$

qui sont les espérances conditionnelles des statistiques nécessaires à l'identification entrée-sortie. Le calcul de la fonctionnelle Q est donc ramené à celui de l'espérance et de la covariance conditionnelle de \mathbf{s} sachant \mathbf{x} et l'estimé courant.

En général, le calcul de cette loi à postérieure est très difficile à mettre en œuvre, sauf dans le cas d'une loi d'entrée discrète, pour lequel les intégrales dans l'équation 1 deviennent des sommes finies (cf [7]). On peut approcher les intégrales nécessaires par une méthode numérique où Monte Carlo stochastique, mais les sources ne sont plus distribuées indépendamment sous la loi à postérieure, et cela rend ces méthodes extrêmement délicates.

Lorsque la loi d'entrée est un mélange de Gaussiennes, en revanche, nous allons voir qu'il est relativement aisé de calculer les moments d'intérêt de manière exacte.

Mélanges de Gaussiennes : on suppose que chacune des sources s_i (on omettra l'indice temporel par souci de lisibilité) est distribuée suivant un mélange fini de lois gaussiennes, c'est à dire :

$$\begin{aligned} p_i(s_i; \eta_i) &= \sum_{j=1}^{q_i} \pi_{ij} \phi(s_i; \mu_{ij}, \sigma_{ij}^2) \quad \eta_i = (\pi_i, \xi_i), \\ \pi_i &= [\pi_{i1}, \dots, \pi_{iq_i}], \quad \xi_i = (\mu_{i1}, \sigma_{i1}^2, \dots, \mu_{iq_i}, \sigma_{iq_i}^2) \end{aligned}$$

Quand on utilise des mélanges, il est pratique de considérer une variable cachée discrète z_i , à valeurs dans $\mathcal{X}_i = [1, \dots, q_i]$ avec la distribution $P(z_i = j) = \pi_{ij}$, $1 \leq j \leq q_i$, et telle que la densité conditionnelle de s_i sachant $z_i = j$ soit précisément $p(s_i|z_i = j) = \phi(s_i; \mu_{ij}, \sigma_{ij}^2)$. La distribution

conjointe des variables (x, s, z) se factorise alors de manière intéressante :

$$\begin{aligned} p(x, s, z; \theta) &= p(x|s; A, J)p(s|z; \xi)p(z; \pi), & (2) \\ &= \phi(x; As, J)\phi(s; \mu_z(\xi), \Gamma_z(\xi))\pi_z, \\ \text{avec } \mu_z(\xi) &= [\mu_{1z_1}, \mu_{2z_2}, \dots, \mu_{nz_n}], \\ \Gamma_z(\xi) &= \text{diag}[\sigma_{1z_1}^2, \sigma_{2z_2}^2, \dots, \sigma_{nz_n}^2]. \end{aligned}$$

Il est facile de voir (cf [8]) que la distribution à postérieure de s sachant x, z est Gaussienne, de moyenne $\alpha_{xz}(\theta)$ et de covariance $\Delta_z(\theta)$ respectivement données par :

$$\begin{aligned} \alpha_{xz}(\theta) &= \mu_z(\xi) + \Gamma_z(\xi)A'R_z(\theta)(x - A\mu_z(\xi)), & (3) \\ \Delta_z(\theta) &= \Gamma_z(\xi) - \Gamma_z(\xi)A'R_z(\theta)A\Gamma_z(\xi) \\ R_z(\theta) &= (A\Gamma_z(\xi)A' + J)^{-1}. \end{aligned}$$

D'autre part, grâce à la formule de Bayes, la distribution à postérieure de z sachant x , $\pi_{zx}(\theta) = p(z|x; \theta)$ est :

$$\begin{aligned} \pi_{zx}(\theta) &= \frac{p(x|z; \theta)\pi_z}{p(x; \theta)} & (4) \\ &\propto \pi_z \phi(x; A\mu_z(\xi), A\Gamma_z(\xi)A' + J). \end{aligned}$$

Ainsi, on obtient pour chaque valeur de l'étiquette z et de manière exacte $\pi_{zx}(\theta)$ et les moments conditionnels $\alpha_{xz}(\theta)$ et $\Delta_z(\theta)$. Il suffit alors de sommer sur z (somme finie de $q = \prod q_i$ termes) pour retrouver les quantités \hat{R}_{xs} et \hat{R}_{ss} , grâce à :

$$p(s|x; \theta) = \sum_z \pi_{zx}(\theta)\phi(s, \alpha_{xz}(\theta), \Delta_z(\theta))$$

Finalement, la log-vraisemblance complète utilisée pour la fonctionnelle de l'EM se décompose en :

$$\begin{aligned} \log p(\mathbf{x}, \mathbf{s}, \mathbf{z}; \theta) &= \sum_{t=1}^T [\log \phi(x_t; As_t, J) \\ &+ \sum_{i=1}^n \sum_{j=1}^{q_i} \log \phi(s_i(t); \mu_{ij}, \sigma_{ij}^2) \delta\{z_i(t) = j\} \\ &+ \sum_{i,j} \log \pi_{ij} \delta\{z_i(t) = j\}], \end{aligned}$$

La fonction de réestimation $Q(\theta|\theta')$ se sépare en trois termes qui se calculent explicitement grâce aux équations 3, 4, e.

$$\begin{aligned} Q_1(\theta|\theta') &= \sum_z \sum_{t=1}^T \pi_{zx}(\theta') \times \\ &\int \log \phi(x_t; As, J)\phi(s; \alpha_{xz}(\theta), \Delta_{xz}(\theta))ds, \\ Q_2(\theta|\theta') &= \sum_{t=1}^T \sum_{i=1}^n \sum_{j=1}^{q_i} \int \log \phi(s_i; \mu_{ij}, \sigma_{ij}^2) \times \\ &p(s_i, z_i = j|x_t; \theta')ds, \\ Q_3(\theta|\theta') &= \sum_{t=1}^T \sum_{i=1}^n \sum_{j=1}^{q_i} \log \pi_{ij} p(z_i = j|x_t; \theta'). \end{aligned}$$

Maximiser chacun de ces termes se fait de manière explicite, nous n'indiquons aucune formule de réestimation par souci de concision.

3 Déconvolution aveugle

Nous considérons à présent le cas de déconvolution aveugle, pour lequel on peut étendre les résultats du paragraphe précédent. L'estimation au maximum de vraisemblance pour le problème de déconvolution est plus difficile que pour celui de séparation de sources, car on a plus d'indépendance pour les sources, et par conséquent la factorisation de la vraisemblance complète n'est plus possible.

Dans le cas d'un signal d'entrée discret, on peut utiliser des méthodes propres au modèle général de Chaines de Markov Cachées (on peut avoir recours à des algorithmes de type "forward-backward" notamment). Nous présentons une autre méthode, basée sur le principe de la séparation de l'observation en paquets (*split data likelihood*), utilisé par Ryden dans le cadre de l'estimation pour des CMC en dimension finie (cf [9]). Il s'agit principalement de considérer que l'observation est constituée de paquets de taille fixe, indépendants entre eux. Toujours dans un souci de clarté nous supposons $m = n = 1$, l'extension au cas vectoriel ne présente pas de difficultés (en théorie du moins). On observe donc $\mathbf{x} = (x_1, \dots, x_T)'$ suivant le modèle :

$$x_t = A_0 s_t + \dots + A_L s_{t-L} + b_t = \mathcal{A}s(t-L:t) + b_t$$

Soit m un entier, et $p_m(x(1:m); \theta)$ la loi jointe du premier paquet de taille m :

$$p_m(x(1:m); \theta) \propto \sigma^{-m} \int \prod_{i=1}^m \phi(x_i; \sum_{k=0}^L A_k s_{i-k}, J) \times p(s_{1-L}; \eta) \cdots p(s_m; \eta) ds_{1-L} \cdots ds_m.$$

$J = \sigma^2 Id$ est la covariance du bruit. La fonction de contraste que nous utiliserons est l'information de Kullback-Leibler entre les distributions des m -marginales :

$$K(\theta, \theta') = E_\theta \left[\log \frac{p_m(x(1:m); \theta')}{p_m(x(1:m); \theta)} \right]$$

Remarque : L'information de Kullback est connu comme bonne fonction de contraste dès que l'on a l'identifiabilité du modèle (le minimum sur θ' étant alors réalisé pour le seul $\theta' = \theta$) :

$$p_m(x(1:m); \theta) = p_m(x(1:m); \theta') \text{ ssi } \theta = \theta'$$

Cette condition est vérifiée pour une taille de paquet supérieure à la taille du filtre, $m \geq L$ (cf [4]).

Considérons à présent le processus de contraste :

$$l_n(\mathbf{x}, \theta') = \sum_{t=0}^{\lfloor T/m \rfloor} \log p_m(x(tm : tm + m - 1); \theta')$$

et $\hat{\theta}_n$ un maximum de $l_n(\mathbf{x}, \theta')$ sur Θ (l'ensemble des valeurs admissible pour le paramètre). Sous certaines hypothèses classiques de régularité, on peut montrer que cet estimateur est consistant et asymptotiquement normal (la preuve étant une généralisation de celle dans le cas i.i.d.).

Remarque : On peut aussi bien considérer des processus de contrastes un peu plus généraux en faisant se recouper les

paquets :

$$l_n(\mathbf{x}; \theta') = \sum_{r=0}^{\lfloor T-m/r \rfloor} \log p_m(x(tr : tr + m - 1); \theta') \quad 1 \leq r \leq m$$

Les estimateurs obtenus à partir de ce type de contrastes restent consistant et asymptotiquement normaux, en outre, leur variance asymptotique dépendant de r , le recouplement peut améliorer l'estimation.

Pour l'implantation de l'algorithme, on se base encore une fois sur une approche de type EM, en modélisant la distribution du signal d'entrée par un mélange de Gaussiennes. On est ainsi ramené au problème de séparation de sources, avec m capteurs et $m + L$ sources, et les résultats précédents s'appliquent, avec quelques changements dans les formules de réestimation (dûs à la structure de la matrice de séparation \mathcal{A}).

4 Simulations

On se place dans le cadre de la séparation de sources, avec $n = 3$ sources, $m = 4$ capteurs, les sources sont bimodales ; c'est à dire la distribution d'une composante est $p_i(s_i) = \frac{1}{2}[\phi(s_i, -\mu, \sigma^2) + \phi(s_i, \mu, \sigma^2)]$. on a pris $\mu = 1, \sigma = 0.3$. Dans les figures, on s'intéresse à l'estimation de la matrice de mélange ; $\hat{A}^{(k)}$ doit converger vers une matrice de permutation (la numérotation des sources étant immatérielle, on ne peut estimer A qu'à une permutation des colonnes près), (M désignant la pseudo-inverse de la matrice M), et la norme de $V * \hat{A}^{(k)}$ quantifie l'angle entre les espaces engendrés par A et son estimée (V est une base de l'orthogonal de l'espace engendré par les colonnes de A).

La première figure présente l'estimation de A conjointement aux paramètres d'entrée (non représentés) pour 50 échantillons et un rapport signal/bruit de l'ordre de 4.5dB, l'initialisation étant aléatoire ; l'algorithme converge en général en une trentaine d'itérations. La deuxième figure présente le cas où les deux modes sont moins séparés ($\sigma = 0.7$), l'initialisation se faisant par l'algorithme JADE basé sur des techniques de cumulants d'ordre quatre ; l'algorithme EM est alors plus performant.

Notre méthode peut s'étendre au cas de l'estimation semi-paramétrique, c'est à dire pour une loi d'entrée inconnue, non nécessairement mélange de Gaussiennes. En fait, on peut approximer la plupart des distributions par un mélange de Gaussiennes, et pour l'estimation des paramètres d'intérêt, il peut suffire que le mélange représente bien les parties les plus importantes de la distribution. Des résultats dans cette direction seront présentés à la conférence.

Références

- [1] P. Comon Independent component analysis, a new concept? *Signal Processing*, volume 36, pp. 287-314, 1994
- [2] O. Shalvi and E. Weinstein System identification based on higher-order statistics submitted to *IEEE Trans. on Signal Proc.*, 1995

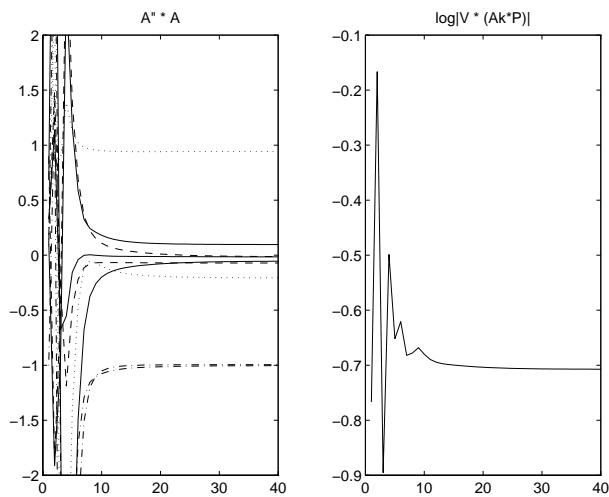


FIG. 1 — Initialisation aléatoire

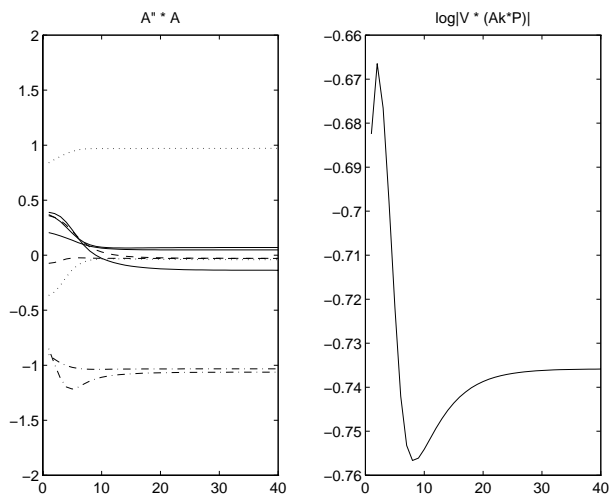


FIG. 2 — Initialisation par Jade (cumulants)

- [3] J. Tugnait and U. Gummadavelli, Blind channel estimation and deconvolution in colored noise using higher-order cumulants In *Proc. SPIE*, volume 2296, 107-125
- [4] G. Giannakis and A. Swami, On estimating noncausal nonminimum phase ARMA models of non-gaussian processes *IEEE Trans. on Acoust. Speech and Sig. Proc.*, volume 38, 478-495, 1990
- [5] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B*, 39 :1-38, 1977.
- [6] M. Tanner *Tools for statistical inference Springer Series in Statistics, Springer-Verlag*, 1993
- [7] A. Belouchrani and J-F Cardoso. Maximum likelihood source separation for discrete sources. In *Proc. EU-SIPCO*, pages 768-771, Edinburgh, September 1994.
- [8] Steven M. Kay(1993). Fundamentals of Statistical Signal Processing. *Prentice Hall. Signal proc. series.*
- [9] Ryden, T. (1994). Consistent and asymptotically normal parameter estimates for hidden Markov models. *Annals of Stat.* **22** 1884-1895.