# Overcomplete source separation using Laplacian Mixture Models

Nikolaos Mitianoudis and Tania Stathaki *Members, IEEE*

**Abstract**

In this letter, the authors explore the use of *Laplacian Mixture Models* (LMMs) to address the overcomplete Blind Source Separation problem in the case that the source signals are very sparse. A two-sensor setup was used to separate an instantaneous mixture of sources. A *hard* and a *soft* decision scheme were introduced to perform separation. The algorithm exhibits good performance as far as separation quality and convergence speed are concerned.

**Index Terms**

Overcomplete source separation, mixture models, Expectation-Maximisation (EM) algorithm.

**EDICS Category: 1.ICAB, 1.STAT**

## I. Introduction

Assume a set of sensors $\underline{x}(n) = [x_1(n), \ldots, x_M(n)]^T$, observing a number of source signals $\underline{s}(n) = [s_1(n), \ldots, s_N(n)]^T$. In this letter, we will assume noiseless instantaneous mixing, i.e.

$$\underline{x}(n) = A\underline{s}(n) \tag{1}$$

where $A$ denotes the *mixing matrix*. The source separation problem consists of estimating the original sources $\underline{s}(n)$, given the observed signals $\underline{x}(n)$. In the case of equal number of sources and sensors ($N = M$), a number of robust approaches using Independent Component Analysis (ICA) have been proposed in literature [7]. In the *overcomplete* source separation case ($M < N$),

Manuscript received June 26, 2004; revised xxxx xx, 200x.

The authors are with the Electrical and Electronic Engineering Department, Imperial College London, Exhibition Road, London SW7 2AZ, UK (e-mail: n.mitianoudis@imperial.ac.uk).

the source separation problem consists of two sub-problems: i) estimate the mixing matrix $A$, ii) estimate the source signals $\underline{s}(n)$.

In general, the linear blind source separation problem can have two theoretical issues: the *identifiability* and the *separability* of the problem. *Identifiability* describes the capability of estimating the structure of the linear model up to a scale and permutation. *Separability* refers to the capability of retrieving the sources using the estimate of the mixing model. According to Eriksson and Koivunen [4], in the case of overcomplete ICA, it is still possible to identify the mixing matrix from the knowledge of $\underline{x}$ alone, although it is not possible to uniquely recover the sources $\underline{s}$. However, if we assume a specific probability distribution for $\underline{s}$, one can obtain estimates of the sources, by maximising the likelihood of $p(\underline{x}|A, \underline{s})$. Eriksson and Koivunen [4] proved that the general linear ICA model is *unique* up to the following assertions: a) The model is separable, b) all source variables are nonGaussian and $rank(A) = M$ and c) none of the source variables have characteristic function featuring a component in the form $\exp(Q(u))$, where $Q(u)$ is a polynomial of degree at least $2$ .

Several approaches were proposed to address the overcomplete source separation problem in the past. Lewicki [6] provided a complete Bayesian approach, assuming Laplacian source priors to estimate both the mixing matrix and the sources in the time domain. Clustering solutions were introduced by Hyvärinen [5] and Bofill-Zibulevsky [2]. Davies and Mitianoudis [3] employed the MDCT (Modified Discrete Cosine Transform) to obtain a sparse representation of the data (see figure 1 for a two sensors - three sources scenario). They proposed a two-state Gaussian Mixture Model (GMM) to represent the source densities and the possible additive noise and used an *Expectation Maximization*(EM)-type algorithm to perform separation with reasonable performance.

In this paper, the authors explore the case of a two-sensor setup with no additive noise, where the source separation problem becomes a one-dimensional (1D) optimal detection problem. The phase difference between the two-sensor data is employed. A *Laplacian Mixture Model*(LMM) is fitted to the phase difference between the two sensors, using an EM-type algorithm. The LMM model is then used to perform separation using either a *soft* or a *hard threshold*.

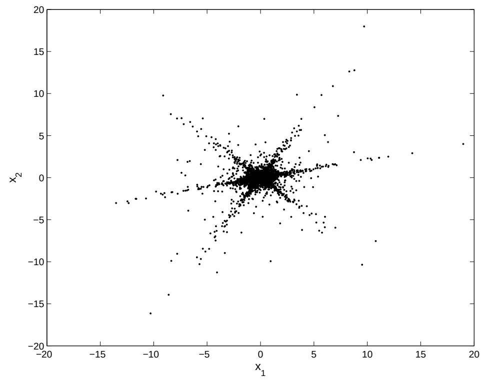


Fig. 1. Scatter plot of the two sensor signals in the sparse MDCT domain.

## II. A TWO-SENSOR APPROACH

In figure 1, one could see the scatter plot of the two sensor signals, in the case of two sensors and three sources. To get a sparser representation of the data, we applied the *Modified Discrete Cosine Transform* (MDCT) on the observed signals. Other transforms with similar properties, such as the Wavelet Transform, can also be employed [8]. The need for sparser representations in overcomplete source separation is discussed more rigorously in [3]. Observing the scatter plot, we can see that the two-dimensional (2D) problem can be mapped to a 1D problem, as the only important parameter is the angle $\theta_n$ of each point, i.e. the phase difference between the two sensors.

$$\theta_n = \text{atan}\frac{x_2(n)}{x_1(n)} \tag{2}$$

Using only the phase difference is equivalent to mapping all the observed data points on the unit-circle. The whole idea resembles the processing of all the observed data points mapped to the half-unit $N$-dimensional sphere, as proposed by Bofill and Zibulevsky [2]. In figure 2, we plot the histogram of the observed data angle $\theta_n$. We can see that the strong superGaussian characteristics of the individual components in the MDCT domain are preserved in $\theta_n$. Observing figure 2, we can model the observed density $p(\theta_n)$, by fitting a *Laplacian Mixture Model* (LMM). Subsequently, each of the Laplacians in the mixture will represent each individual source. Using the estimated Laplacians, we can perform source separation by optimal detection schemes.

## III. Laplacian Mixture Modelling

The Laplacian density is usually represented by

$$\mathcal{L}(\theta, c, \theta_0) = ce^{-2c|\theta-\theta_0|} \tag{3}$$

where $\theta$ represents the center (mean) and $c > 0$ controls the "width" of the density respectively. A *Laplacian Mixture Model* (LMM) is defined as follows:

$$p(\theta) = \sum_{i=1}^{N} \alpha_i \mathcal{L}(\theta, c_i, \theta_i) = \sum_{i=1}^{N} \alpha_i c_i e^{-2c_i|\theta-\theta_i|} \tag{4}$$

where $\alpha_i$, $\theta_i$, $c_i$ are the weights, centers and widths of each Laplacian. Effectively, $\sum_{i=1}^{N} \alpha_i = 1$. A common method used to train a mixture model is the *Expectation-Maximization* (EM) algorithm.

## IV. Training using the EM algorithm

In this section, we derive the EM algorithm to train a LMM, based on Bilmes's analysis [1]. In [1], Bilmes presents a procedure to find Maximum Likelihood Mixture density parameters using the EM. Assuming $T$ samples for $\theta_n$ and Laplacian Mixture densities (4), we can extend Bilmes's analysis to formulate EM's cost function for Laplacian Mixtures, as follows:

$$J(c_i, \theta_i) = \sum_{n=1}^{T} \sum_{i=1}^{N} (\log c_i - 2c_i|\theta_n - \theta_i|) p(i|\theta_n) \tag{5}$$

where $p(i|\theta_n)$ represents the probability of $\theta_n$ belonging to the $i^{th}$ Laplacian. The updates for $p(i|\theta_n)$ and $\alpha_i$ are given by the following formulas:

$$p(i|\theta_n) = \frac{\alpha_i c_i e^{-2c_i|\theta_n-\theta_i|}}{\sum_{i=1}^{N} \alpha_i c_i e^{-2c_i|\theta_n-\theta_i|}} \tag{6}$$

$$\alpha_i^+ \leftarrow \frac{1}{T} \sum_{n=1}^{T} p(i|\theta_n) \tag{7}$$

To find the updates for $\theta_i^+$ and $c_i^+$, we have to solve the equations $\partial J/\partial \theta_i = 0$ and $\partial J/\partial c_i = 0$ (see Appendix I). The updates are given by:

$$\theta_i^+ \leftarrow \frac{\sum_{n=1}^{T} \frac{\theta_n}{|\theta_n-\theta_i|} p(i|\theta_n)}{\sum_{n=1}^{T} \frac{1}{|\theta_n-\theta_i|} p(i|\theta_n)} \tag{8}$$

$$c_i^+ \leftarrow \frac{\sum_{n=1}^{T} p(i|\theta_n)}{2\sum_{n=1}^{T} |\theta_n - \theta_i| p(i|\theta_n)} \tag{9}$$

## V. SEPARATION

Once the LMM is trained, we can use optimal detection theory and the estimated individual Laplacians to get estimates of the sources. Essentially, the center of each Laplacian $\theta_i$ represents a column of the mixing matrix $A$, in the form of $[\cos\theta_i \ \ \sin\theta_i]^T$. To perform separation, we can use a *hard* or *soft decision threshold*.

### A. Hard threshold - "Winner takes all"

The "Winner takes all" strategy simply attributes each point of the scatter plot to one of the sources. This is performed by setting a hard threshold at the intersections between the trained Laplacians. In figure 2, you can see the fitted Laplacians in a two sensors - three sources example and the hard threshold imposed.

### B. Soft threshold

One can relax the hard threshold strategy, by allowing points belong to more than one source simultaneously. A soft thresholding strategy can attribute only the points that constitute a ratio $q$ (i.e. 0.8-0.9) of each Laplacian to the corresponding source. This is that the $i^{th}$ source will be associated with the points $\theta_n$ for which $p(\theta_n) \geq (1-q)\alpha_i c_i$. This will allow points to belong to more than one sources. However, for this scheme to be effective, the estimated Laplacians need to be fairly concentrated around $\theta_i$ (quite small variance). In the opposite case, there will be classification mistakes. The solutions are either to decrease $q$ or apply a hard threshold.

### C. Edge effects

There is an issue about some edge effects on the Laplacian Mixture Modelling. The Laplacian density, as described in (3), is valid $\forall\ \theta \in (-\infty, +\infty)$. However, the range of $\theta_n$ is bounded to $(-\pi/2, \pi/2)$. Assume that you have a concentration of points close to these boundaries. The EM algorithm will fit a Laplacian around this cluster without any problem, as the algorithm is not imposing any restriction on $\theta$. As a result, this Laplacian will be extended slightly on $(-3\pi/2, -\pi/2)$ or $(\pi/2, 3\pi/2)$. The problem comes when we need to perform separation. Limiting this Laplacian at $-\pi/2$ or $\pi/2$, implies that we are imposing a hard threshold at these points, which might not be very accurate. Therefore, we need to map the parts that exist in

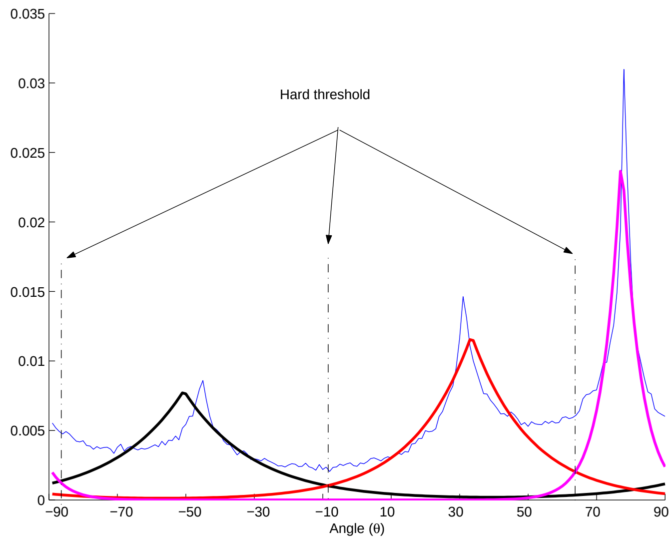


Fig. 2. A plot of the observed $\theta_n$, as in eq. (2). Three Laplacian densities were fitted by the proposed algorithm. A hard threshold employed for separation is depicted.

$(-3\pi/2, -\pi/2)$ or $(\pi/2, 3\pi/2)$ to $(-\pi/2, \pi/2)$. This can be achieved by using a "modified" density $G(\theta)$ (see figures 2, 3):

$$G(\theta) = \mathcal{L}(\theta) + \mathcal{L}(\theta - \pi) + \mathcal{L}(\theta + \pi), \;\; \forall \; \theta \in (-\pi/2, \pi/2) \tag{10}$$

## VI. Experiments - Performance

To test the efficiency of the proposed scheme, we created two artificial instantaneous mixtures of two sensors - three audio sources and two sensors - four audio sources.

In figures 2,3, we can see the actual distribution of observed $\theta_n$ and the fitted Laplacians. The proposed EM algorithm fitted the Laplacians with relative accuracy. More accuracy is observed in the case of $3$ sources than in the case of $4$ sources, as in the latter case Gaussianity has increased with the number of sources. We managed to separate the sources in either case using the proposed soft thresholding scheme, as the sources were quite concentrated around the estimated $\theta_i$.

In table I, we compare the algorithm's performance (soft and hard thresholding) with Hyvärinen's [5] approach, in terms of average SNR of the separated sources in the two aforementioned experiments. We observe that the algorithm features similar performance to well-established approaches in the field.

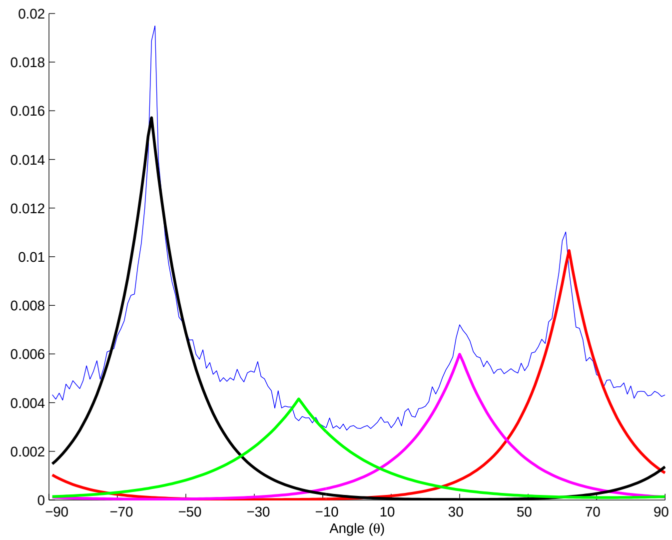


Fig. 3. A plot of the observed $\theta_n$ (2) for the case of four sources and two sensors. Four Laplacian densities were fitted by the proposed algorithm.

The algorithm requires $30-80$ iterations to converge (see figure 4). We speculate that the convergence speed is slightly decreased because we are not able to calculate directly the update for $\theta_i$ (see Appendix I).

TABLE I

AVERAGE SNR ($dB$) COMPARISON BETWEEN HYVÄRINEN'S AND LMM-EM APPROACH USING SOFT AND HARD THRESHOLDS.

| | $2\times3$ | $2\times4$ |
|---|---|---|
| LMM-EM (soft thres.) | 10.3 | 7.17 |
| LMM-EM (hard thres.) | 10.3 | 7.45 |
| Hyvärinen's [5] | 11.2 | 8.1 |

## VII. CONCLUSIONS

In this letter, we present a Laplacian Mixture Modelling approach for overcomplete separation of signals. The signals are processed in the MDCT domain, however, any alternative sparse transform can be applied. We reduce the 2 sensors - many sources problem to an 1D optimal

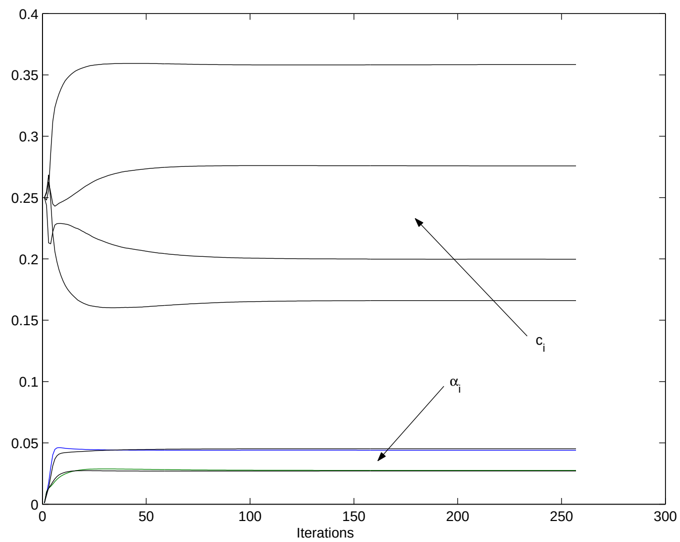


Fig. 4. Convergence of the LMM-EM algorithm for $\alpha_i$ and $c_i$ for the four sources experiment.

detection problem, by fitting a Laplacian Mixture Model using the EM algorithm. The algorithm provided reasonable separation, compared to previous approaches. For future work, it would be interesting to generalise the Laplacian Mixture Modelling and perform separation in the case of more than $2$ sensors.

## ACKNOWLEDGMENTS

The authors would like to thank Dr. Mike Davies for useful discussions and suggestions.

## APPENDIX I
## CALCULATING UPDATES

*Update for* $\theta_i$

$$\frac{\partial J}{\partial \theta_i} = \sum_{n=1}^{T} (-2c_i \frac{\partial}{\partial \theta_i} |\theta_n - \theta_i|) p(i|\theta_n) = \tag{11}$$

$$= \sum_{n=1}^{T} -2c_i \text{sgn}(\theta_n - \theta_i) p(i|\theta_n) = 0 \Rightarrow \tag{12}$$

$$\sum_{n=1}^{T} \text{sgn}(\theta_n - \theta_i) p(i|\theta_n) = \sum_{n=1}^{T} \frac{\theta_n - \theta_i}{|\theta_n - \theta_i|} p(i|\theta_n) = 0 \tag{13}$$

In the previous equation, it is not possible to find an exact solution for $\theta_n$, because of the $\mathrm{sgn}(\cdot)$ function. However, we can estimate a new update for $\theta_i$, given a previous estimate for $|\theta_n - \theta_i|$. This approximation may decrease the convergence of the EM algorithm.

$$\sum_{n=1}^{T} \frac{\theta_n}{|\theta_n - \theta_i|} p(i|\theta_n) = \sum_{n=1}^{T} \frac{\theta_i}{|\theta_n - \theta_i|} p(i|\theta_n) \tag{14}$$

$$\theta_i^{+} \leftarrow \frac{\sum_{n=1}^{T} \frac{\theta_n}{|\theta_n - \theta_i|} p(i|\theta_n)}{\sum_{n=1}^{T} \frac{1}{|\theta_n - \theta_i|} p(i|\theta_n)} \tag{15}$$

*Update for* $c_i$

$$\frac{\partial J}{\partial c_i} = \sum_{n=1}^{T} (\frac{1}{c_i} - 2|\theta_n - \theta_i|) p(i|\theta_n) = 0 \tag{16}$$

$$\frac{1}{c_i} \sum_{n=1}^{T} p(i|\theta_n) = \sum_{n=1}^{T} 2|\theta_n - \theta_i| p(i|\theta_n) \tag{17}$$

$$c_i^{+} \leftarrow \frac{\sum_{n=1}^{T} p(i|\theta_n)}{2\sum_{n=1}^{T} |\theta_n - \theta_i| p(i|\theta_n)} \tag{18}$$

## References


[1] J.A. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian Mixture and Hidden Mixture Models. Technical report, Department of Electrical Engineering and Computer Science, U.C. Berkeley, California, 1998.

[2] P. Bofill and M. Zibulevsky. Underdetermined blind source separation using sparse representations. *Signal Processing*, 81(11):2353–2362, 2001.

[3] M. Davies and N. Mitianoudis. A simple mixture model for sparse overcomplete ica. *IEE proceedings in Vision, Image and Signal Processing*, 151(1):35–43, 2004.

[4] J. Eriksson and V. Koivunen. Identifiability and seperability of linear ica models revisited. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 23–27, Nara, Japan, 2003.

[5] A. Hyvärinen. Independent component analysis in the presence of gaussian noise by maximizing joint likelihood. *Neurocomputing*, 22:49–67, 1998.

[6] M. Lewicki and T.J. Sejnowski. Learning overcomplete representations. *Neural Computation*, 12:337–365, 2000.

[7] N. Mitianoudis. *Audio Source Separation using Independent Component Analysis*. PhD thesis, Queen Mary London, 2004.

[8] M. Zibulevsky, P. Kisilev, Y.Y. Zeevi, and B.A. Pearlmutter. Blind source separation via multinode sparse representation. *Advances in Neural Information Processing Systems*, 14:1049–1056, 2002.