Signal Processing 155 (2019) 281–286

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Short communication

# Nonlinear estimators from ICA mixture models

Gonzalo Safont[a], Addisson Salazar[a], Luis Vergara[a,*], Alberto Rodríguez[b]

[a] Universitat Politècnica de València, Instituto de Telecomunicaciones y Aplicaciones Multimedia, València 46022, Spain
[b] Universidad Miguel Hernández de Elche, Departamento de Ingeniería de Comunicaciones, Spain

ARTICLE INFO

*Article history:*
Received 19 February 2018
Revised 1 October 2018
Accepted 4 October 2018
Available online 5 October 2018

*Keywords:*
ICA
Nonlinear estimators
LMSE
MAP
EEG reconstruction
non-Gaussian mixtures

ABSTRACT

Independent Component Analyzers Mixture Models (ICAMM) are versatile and general models for a large variety of probability density functions. In this paper we assume ICAMM to derive new MAP and LMSE estimators. The first one (MAP-ICAMM) is obtained by an iterative gradient algorithm, while the second (LMSE-ICAMM) admits a closed-form solution. Both estimators can be combined by using LMSE-ICAMM to initialize the iterative computation of MAP-ICAMM .The new estimators are applied to the reconstruction of missed channels in EEG multichannel analysis. The experiments demonstrate the superiority of the new estimators with respect to: Spherical Splines, Hermite, Partial Least Squares, Support Vector Regression, and Random Forest Regression.



## 1. Introduction

Estimation is one of the fundamental problems in statistical signal processing [1]. It is an essential part of many fields like spectral analysis, coding, time series analysis, prediction, interpolation, smoothing. Moreover, it appears in many areas of application. In spite of the huge amount of previous work in statistical estimation methods, new developments are still possible if new statistical models appear, so that new maximum a posteriori (MAP) or least mean square error (LMSE) solutions can be found. In this paper we consider the Independent Component Analyzers Mixture Model (ICAMM) [2–4] of the multivariate probability density function (MPDF) of the observations. ICAMM is a versatile model which encompasses most of the usual MPDF models, including both non-Gaussian and Gaussian Mixture Models. This generality implies that optimal estimators assuming an underlying ICAMM can be very attractive options in a large variety of scenarios. Therefore, we have derived the MAP and LMSE estimators of missing data, considering that the underlying MPDF is properly captured by ICAMM. The MAP estimator will be obtained by and iterative algorithm, and will be called MAP-ICAMM. On the other hand, the LMSE estimator is the expected value of the missing data conditioned to the available data, which we have calculated assuming ICAMM; the corresponding estimator will be called LMSE-ICAMM. The new estimator has been assessed in the reconstruction of missing data of electroencephalographic (EEG) signals measured on subjects while performing a memory and learning task. They have been compared with the method of splines [5], a deterministic method which approximates complex function stepwise by local polynomials. In particular, we have considered the following state-of-the-art methods: Spherical Splines [6], which is the most commonly used method for interpolation in EEG processing in many EEG processing software, such as EEGLAB [7]; Hermite interpolation [8], which is a benchmark smoothing and interpolation method; Partial Least Squares [9], a preferred tool for ill-posed linear estimation problems; Support Vector Regression [10], one of the most popular machine learning tools for regression; and Random Forest Regression [11], which is a recent multidimensional interpolation technique with good performance across different applications (see [12] and the references within).

In the next section we present the analytical derivation of the new proposed estimators (MAP-ICAMM and LMSE-ICAMM). Then Section 3 is devoted to the mentioned EEG data application. Conclusion section ends the paper.

## 2. Estimators based on ICAMM

Let us consider an observation vector $\mathbf{x}$ of dimension $(M \times 1)$. Without loss of generality, this vector can be defined as being composed by a vector of known components, $\mathbf{y}$, and a vector of unknown components, $\mathbf{z}$, in the form

$$\mathbf{x} = \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} = \mathbf{P}_y \mathbf{y} + \mathbf{P}_z \mathbf{z}, \tag{1}$$

* Corresponding author.
*E-mail address:* lvergara@dcom.upv.es (L. Vergara).

https://doi.org/10.1016/j.sigpro.2018.10.003
0165-1684/

where $\mathbf{z}$ is a vector of size $(M_{unk} \times 1)$ and therefore $\mathbf{y}$ is a vector of size $((M - M_{unk}) \times 1)$. $\mathbf{P}_y$ and $\mathbf{P}_z$ are rectangular diagonal matrices respectively equal to the first $M - M_{unk}$ columns and the last $M_{unk}$ columns of the identity matrix of dimension $(M \times M)$, so that $\mathbf{P}_y\mathbf{y} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_{M_{unk}} \end{bmatrix}$ and $\mathbf{P}_z\mathbf{z} = \begin{bmatrix} \mathbf{0}_{M-M_{unk}} \\ \mathbf{z} \end{bmatrix}$, where $\mathbf{0}_i$ is a zero vector of size $(i \times 1)$. The goal is to estimate $\mathbf{z}$ from $\mathbf{y}$. Let us assume that the MPDF $p(\mathbf{x})$ is modeled with a $K$-class ICAMM. If $\mathbf{x}$ belongs to class $C_k$ $k = 1...K$, then

$$\mathbf{x} = \mathbf{A}_k\,\mathbf{s}_k + \mathbf{b}_k, \tag{2}$$

where $\mathbf{s}_k$ is a vector of size $(M \times 1)$ that contains statistically independent components (also known as "sources"); $\mathbf{A}_k$ is the mixing matrix; and $\mathbf{b}_k$ is a bias vector. Notice that $\mathbf{A}_k$ is a square matrix that can be inverted to obtain $\mathbf{W}_k = {\mathbf{A}_k}^{-1}$, the de-mixing matrix of class $k$. Hence $p(\mathbf{x})$ can be expressed as a mixture of $K$ components respectively corresponding to the $K$ classes.

$$p(\mathbf{x}) = p(\mathbf{z},\mathbf{y}) = \sum_{k=1}^{K} p(\mathbf{z},\mathbf{y}|C_k)P(C_k) = \sum_{k=1}^{K} |\det \mathbf{W}_k|\, p(\mathbf{s}_k)P(C_k), \tag{3}$$

where $P(C_k)$ is the prior probability of class $k$.

Given a training set of observation vectors $\mathbf{x}^{(n)}$ $n = 1, ..., N$, we can estimate the model parameters $\mathbf{W}_k$, $\mathbf{b}_k$ and $P(C_k)$, using one of the many existing methods (see [13] for a general procedure).

### 1.1. MAP estimator (MAP-ICAMM)

Let us consider the MAP estimator of $\mathbf{z}$ from $\mathbf{y}$

$$\mathbf{z}_{MAP} = \underbrace{\max}_{\mathbf{z}} \log p(\mathbf{z},\mathbf{y}) = \underbrace{\max}_{\mathbf{z}} L(\mathbf{z},\mathbf{y}). \tag{4}$$

This maximization requires the calculation of the derivative of $L(\mathbf{z}, \mathbf{y})$ with respect to $\mathbf{z}$. By taking the log of (3) and considering that the components of $\mathbf{s}_k$ are independent, we arrive to

$$\begin{aligned} \frac{\partial L(\mathbf{z},\mathbf{y})}{\partial \mathbf{z}} &= \frac{1}{p(\mathbf{z},\mathbf{y})} \sum_{k=1}^{K} |\det \mathbf{W}_k| P(C_k) \sum_{m=1}^{M} \frac{\partial p(\mathbf{s}_k)}{\partial s_{km}} \frac{\partial s_{km}}{\partial \mathbf{z}} \\ &= \sum_{k=1}^{K} p(C_k|\mathbf{z},\mathbf{y}) \sum_{m=1}^{M} \frac{\partial \log p(s_{km})}{\partial s_{km}} \frac{\partial s_{km}}{\partial \mathbf{z}}, \end{aligned} \tag{5}$$

where $s_{km}$ is the $m$-th source of class $k$. The value $s_{km}$ can be obtained as $s_{km} = \mathbf{w}_{km}^T(\mathbf{x} - \mathbf{b}_k)$, with $\mathbf{w}_{km}^T$ being the $m$-th row of $\mathbf{W}_k$. Thus, its derivative is equal to

$$\frac{\partial s_{km}}{\partial \mathbf{z}} = \left({\mathbf{w}_{km}}^T \mathbf{P}_z\right)^T = \mathbf{P}_z^T \mathbf{w}_{km}. \tag{6}$$

The derivative of $\log p(s_{km})$, also known as the score function, can be calculated explicitly for many common probability density functions. This requires prior knowledge of the source PDF which could limit the applicability of the algorithm. To reach general applicability, we will assume that the probability density of each source is to be estimated using a nonparametric kernel density estimator with a Gaussian kernel. This requires labelling of the training samples which can be made, once the model parameters have been estimated, by selecting the class $k$ which maximizes the posterior probability $P(C_k|\mathbf{x}_n) = P(\mathbf{x}_n|C_k)P(C_k) / \sum_{k'=1}^{K} P(\mathbf{x}_n|C_{k'})P(C_{k'})$ [3]. Let us call $\mathbf{x}_k^{(l)}$ $l = 1...L_k$ to the subset of the training samples assigned to class $k$, and $\mathbf{s}_k^{(l)} = \mathbf{W}_k(\mathbf{x}_k^{(l)} - \mathbf{b}_k)$, the corresponding source vectors. The non-parametric estimator of $p(s_{km})$ is given by

$$p(s_{km}) = \frac{1}{a_0} \sum_{l=1}^{L_k} e^{-\frac{1}{2h^2}\left(s_{km} - s_{km}^{(l)}\right)^2}, \tag{7}$$

where $s_{km}^{(l)}$ is the $m$-th component of vector $\mathbf{s}_k^{(l)}$, $h$ is called the bandwidth of the nonparametric estimator, and $a_0$ is a scaling constant calculated so that $\int_{-\infty}^{\infty} p(s_{km})ds_{km} = 1$. For the Gaussian kernel of (7), this scaling constant is $a_0 = \sqrt{2\pi}\,h\,L_k$. The derivative of $\log p(s_{km})$ is

$$\frac{\partial \log p(s_{km})}{\partial s_{km}} = \frac{\sum_{l=1}^{L_k} \left(\frac{s_{km} - s_{km}^{(l)}}{h}\right) e^{-\frac{1}{2h^2}\left(s_{km} - s_{km}^{(l)}\right)^2}}{\sum_{l=1}^{N_T} e^{-\frac{1}{2h^2}\left(s_{km} - s_{km}^{(l)}\right)^2}}. \tag{8}$$

Hence, the derivative of $L(\mathbf{z}, \mathbf{y})$ can be calculated by replacing (6) and (8) into (5). Then, maximization of $L(\mathbf{z}, \mathbf{y})$ could be achieved by using classical gradient algorithms, although Newton methods are preferable to get fast convergence. The problem of the Newton methods is that computation of the second derivative is required. To overcome this inconvenience, a family of quasi-Newton methods have been proposed. In particular, the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [14] is generally accepted as the quasi-Newton method yielding the best compromise between fast convergence and computational payload, even for non-smooth optimization. That is the option that we have selected to obtain $\mathbf{z}_{MAP} = \underbrace{\max}_{\mathbf{z}} L(\mathbf{z},\mathbf{y})$. We have named this estimator as MAP-ICAMM. A pseudocode summary of the algorithm is included in the following:

Algorithm 1. *Computation of the MAP-ICAMM.*

**Input**: Training set of complete vectors $\mathbf{x}^{(n)}$ $n = 1...N$, vector of known components $\mathbf{y}$, initial value for the unknown components $\mathbf{z}(0)$
**Output**: MAP prediction, $\mathbf{z}_{MAP}$
1: Estimate ICAMM parameters $\mathbf{b}_k\,\mathbf{W}_k\,P(C_k)$, $k = 1...K$ from the training set [13]
2: Label training samples [3] and compute $\mathbf{s}_k^{(l)} = \mathbf{W}_k(\mathbf{x}_k^{(l)} - \mathbf{b}_k)$ $l = 1...L_k$ $k = 1...K$
3: **for** $i = 1...I$
4: Compute $\frac{\partial L(\mathbf{z}(i-1),\mathbf{y})}{\partial \mathbf{z}(i-1)}$ (5)–(8)
5: **end for**
6: $\mathbf{z}_{MAP} = \mathbf{z}(I)$

### 1.2. LMSE estimator (LMSE-ICAMM)

The general solution to the LMSE criterion is the conditional expectation of unknown data with respect to known data, that is,

$$\mathbf{z}_{LMSE} = E[\mathbf{z}|\mathbf{y}] = \int \mathbf{z}\, p(\mathbf{z}|\mathbf{y})d\mathbf{z}. \tag{9}$$

Considering the mixture model in (3) and using the chain rule, $\mathbf{z}_{LMSE}$ can be expressed as

$$\begin{aligned} \mathbf{z}_{LMSE} &= \int \mathbf{z}\, p(\mathbf{z}|\mathbf{y})d\mathbf{z} = \sum_{k=1}^{K} \int \mathbf{z}\, p(\mathbf{z}|\mathbf{y},C_k)d\mathbf{z} \cdot P(C_k|\mathbf{y}) \\ &= \sum_{k=1}^{K} E[\mathbf{z}|\mathbf{y},C_k] \cdot P(C_k|\mathbf{y}). \end{aligned} \tag{10}$$

So we need to compute $E[\mathbf{z}|\mathbf{y}, C_k]$ and $P(C_k|\mathbf{y})$. Regarding $E[\mathbf{z}|\mathbf{y}, C_k]$, let us first compute the conditional expectation of the sources $E[\mathbf{s}_k|\mathbf{y}, C_k]$. Considering (1) and (2) we may write

$$\begin{aligned} E[\mathbf{s}_k|\mathbf{y},C_k] &= E[\mathbf{W}_k(\mathbf{P}_y\mathbf{y} + \mathbf{P}_z\mathbf{z} - \mathbf{b}_k)|\mathbf{y},C_k] \\ &= \mathbf{W}_k\mathbf{P}_y\,\mathbf{y} + \mathbf{W}_k\mathbf{P}_z\,E[\mathbf{z}|\mathbf{y},C_k] - \mathbf{W}_k\mathbf{b}_k, \end{aligned} \tag{11}$$

then we can solve for $E[\mathbf{z}|\mathbf{y}, C_k]$ by using the pseudoinverse $(\mathbf{W}_k\mathbf{P}_z)^+$

$$E[\mathbf{z}|\mathbf{y},C_k] = (\mathbf{W}_k\mathbf{P}_z)^+ \cdot [E[\mathbf{s}_k|\mathbf{y},C_k] - \mathbf{W}_k\mathbf{P}_y\mathbf{y} + \mathbf{W}_k\mathbf{b}_k]. \tag{12}$$

Application of (12) to estimate $E[\mathbf{z}|\mathbf{y}, C_k]$ requires knowledge of $E[\mathbf{s}_k|\mathbf{y}, C_k]$. This later can be estimated using a variety of existing