# Numerical Experiments on Blind Separation of Astrophysical Maps by Independent Factor Analysis

*Ercan E. Kuruoglu, Luigi Bedini, Maria Teresa Paratore, Emanuele Salerno and Anna Tonazzini*

Istituto di Elaborazione della Informazione
Consiglio Nazionale delle Ricerche
Area della Ricerca di CNR Pisa
via G. Moruzzi 1
56124 Pisa, ITALY

**Abstract**

The celestial microwave radiation is generated by various astrophysical sources. Measuring this radiation does not enable us to immediately extract information about the sources, since the individual signals are superimposed to one another all over the measurement bandwidth. Each source process radiates in accordance with a typical frequency-emission law, which, unfortunately, is often unknown. Some blind technique should thus be used to separate the individual components from the total radiation. In this paper, we apply a recently introduced approach, called Independent Factor Analysis (IFA), which models the sources as mixtures of Gaussians. After checking the appropriateness of this model to our case, we briefly describe the IFA approach and show some results from simulated but realistic data.

Keywords: ACM: I.4.9 Image processing and computer vision: applications. I.2.6 Artificial intelligence: learning

## 1. Introduction

Astronomical microwave images carry important information about the Universe. Unfortunately, a radiometric image taken at any working frequency is a superposition of radiations coming from different sources, and always corrupted by measurement noise. The classical components of the microwave sky radiation are the cosmic microwave background (CMB), the galactic dust radiation, the synchrotron radiation and the free-free radiation. Each of these radiations has its own interest in cosmology or astrophysics. For example, the CMB temperature map over the celestial sphere is a picture of the Universe at last scattering and would yield invaluable information to estimate fundamental quantities, thus enabling cosmologists to assess competing theories. The problem is thus to separate the individual radiation components from the total measured field. Since the spectral features of the individual source emissions are normally not known, they should be separated by a blind technique. The problem is further complicated by the presence of the sensor noise, which can be very strong and location-dependent. One technique for blind source separation, namely, independent component analysis (ICA), has been exploited for astrophysical image separation [Baccigalupi *et al.*, 2000]. Owing to the particular data model adopted, blind separation by ICA techniques has only been shown to be feasible and reliable when noise is negligible. Strong noise components affect both the output noise level and the quality of the separation. The ICA procedure does not assume any other information than mutual independence of the source signals, it implicitly assumes some particular form for their statistical distributions and normally does not permit the introduction of any additional information. This forces one to neglect useful information when, as in our case, something is known on the source distributions, the mixing coefficients and the noise process.

In this paper, we adopt a method that is able to incorporate prior information about the sources in a very generic way. This method is called independent factor analysis (IFA), and has been introduced recently [Moulines, *et al.*, 1997, Attias, 1999]. Attias proposes a Gaussian mixture model for the source densities (with the mixture coefficients to be estimated), and provides a neural network architecture with an expectation-maximization (EM) learning algorithm [Dempster *et al.*, 1977]. Since noise is also taken into account in the data model, IFA offers a promising alternative to ICA. Nevertheless, the basic IFA approach, as described in [Attias, 1999], is computationally expensive, and this would complicate the problem when a considerable number of sources are to be separated. Moreover, the fact that the learning algorithm is not guaranteed to converge to a global optimum suggests us that the effectiveness of this approach could depend on the particular problem. The numerical studies on IFA reported in the literature are only limited to some simple toy problems, and the potentials and the drawbacks of the technique are not well understood yet. Our aim is thus to analyze the IFA features in the case of realistic data. Our data maps simulate the ones expected from the *Planck Surveyor Satellite*, a mission that will be launched in 2007 by the European Space Agency. The aim of this mission is to map, with unprecedented accuracy, the CMB radiation anisotropies over the entire celestial sphere and on nine measurement channels in the millimeter and submillimeter-wave range, with working frequencies from 30 to 857 GHz. Our data are totally synthetic or extrapolated from other data sets, with different frequency ranges or spatial resolutions, but are considered realistic for the Planck application, especially as far as the location-dependent noise maps are concerned. We constructed our mixture data on this basis and tested the IFA technique against them.

## 2. Data and Noise Analysis

Before starting our numerical experimentation, we analyzed the amplitude distributions of our test signals in order to see whether the mixture of Gaussians model proposed by the IFA approach is justified, and how many significant Gaussians are to be expected for each source process.

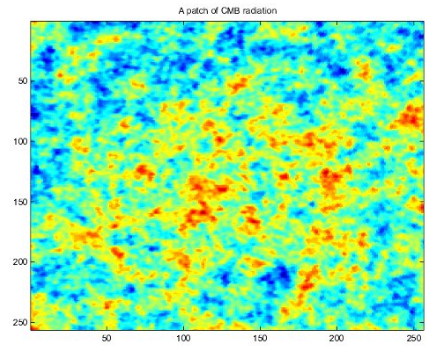


**Figure 1:** Simulated CMB radiation map

The analysis we present here is related to the CMB, galactic dust and synchrotron radiations. According to the standard cosmological theories, the CMB radiation should have a Gaussian distribution. Figure 1 shows a typical CMB image generated synthetically. Much less is known about the statistical distribution of the other radiations we are considering here. For galactic dust, existing sky maps obtained from different frequency channels have been used as spatial templates, from which the specific emission values have been generated according to the hypothesized dust emission process. A map of galactic dust radiation is shown in Figure 2.a. The related histogram is provided in Figure 2.b, solid curve.

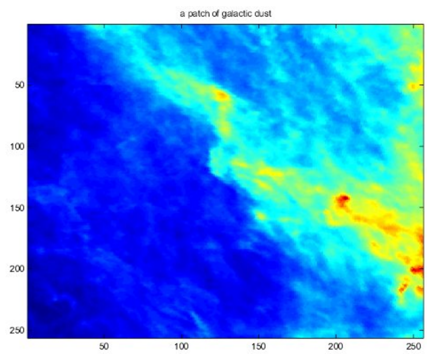


**Figure 2 a:** Simulated galactic dust radiation map

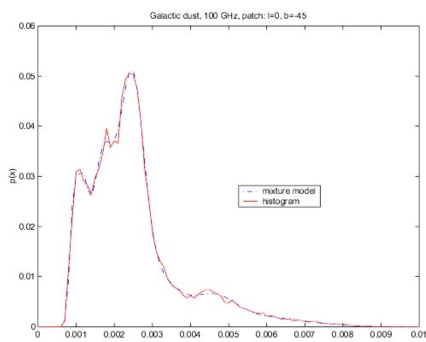


**Figure 2 b:** Histogram for Fig. 2 a and related Gaussian mixture model fit

It is clear from the histogram that the galactic dust exhibits a non-Gaussian behavior: the curve is multimodal and unsymmetric. We checked the feasibility of a mixture of Gaussians model for this spatial distribution, that is, we tried to fit the density

$$p_X(x) = \sum_i a_i \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right) \qquad (1)$$

to the existing histogram using an EM algorithm. The resulting density is shown in Figure 2.b, dashed curve. We repeated our experiments on 15 different images of the same size, and in all cases we have seen that the Gaussian mixture model provides very good fits using less than five components. The simulated galactic synchrotron maps have been obtained by extrapolating existing data, both for spatial resolution and for spectral emission. Figure 3.a shows one such map. The related histogram is given in Figure 3.b, solid curve. Again, we tried to fit the curve by a Gaussian mixture using the EM algorithm. The result obtained by fitting a mixture of only four components is given in Figure 3.b, dashed curve. These observations were repeated for many other radiation maps, showing that the Gaussian mixture density is an efficient generic model for images of the type considered here. The measurement noise is often assumed to be white, Gaussian and space-invariant. This is not always the case: in satellite radiometric images, noise may not be space-invariant, since the antenna does not scan the sky uniformly. A typical noise map is shown in Figure 4. For each measurement channel, our test maps have been constructed by combining some of the available source maps and then adding one such noise realization. Each map has been considered as the product of a frequency-independent spatial template and a specified function of the frequency, which is assumed unknown in the separation process. The measured data have been simulated at the four working frequencies of the "low-frequency instrument" that will be onboard the Planck spacecraft, i.e., 30 GHz, 44 GHz, 70 GHz and 100 GHz.

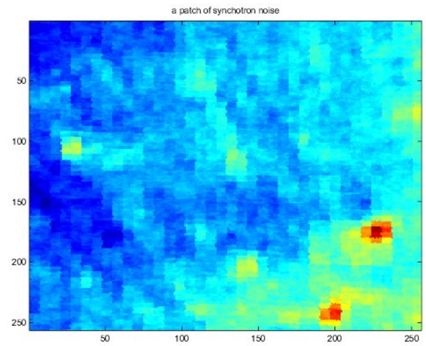


**Figure 3 a:** Simulated galactic synchrotron radiation map

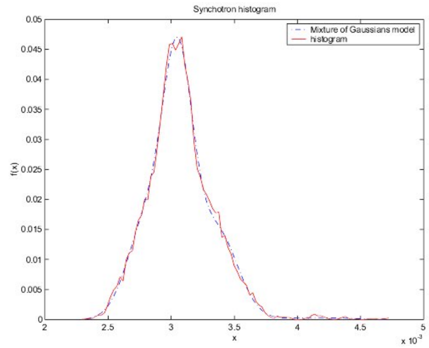


**Figure 3 b:** Histogram for Fig. 3 a and related Gaussian mixture model fit

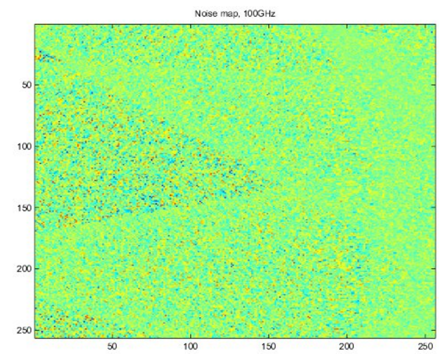


**Figure 4:** Typical location-dependent noise map

## 3. Independent Factor Analysis

For the case considered here, we can restrict our attention to a linear mixture model, described by:

$$y_i = \sum_{j=1}^{L} \mathbf{H}_{ij} x_j + n_i \,, \quad i = 1,\dots,N \tag{2}$$

where $\mathbf{H}_{ij}$ is the *ij*-th entry of a mixing matrix **H**, $y_i$ are the observations, $x_j$ are the sources and $n_i$ are the noise realizations. In our application, the *N*-vector **y** is made of the instrument output maps at the *N* different channels, the *L*-vector **x** collects the source spatial maps, the *N*-vector **n** contains the instrumental noise maps on all the channels, which are assumed to be Gaussian and space-varying. The mixing coefficients in the $L \times N$ matrix **H** depend on the frequency laws characteristic of the individual sources and on the frequency responses of the measuring instrument on all the channels considered. We would like to obtain **H** and **x** from **y**. In the last decade, various efforts have been made for the solution of this blind source separation (BSS) problem. In particular, the ICA approach assumes statistical independence among the source functions, and solves the separation problem by optimizing some criterion involving a separable joint density for the estimated sources. Studies on the related class of techniques have been widely reported in the literature. However, ICA considers a highly idealized problem, and its performance deteriorates as the noise increases [Attias, 1999]. Efforts have been made to include noise into the analysis:

Hyvarinen suggested employing a special class of noise-insensitive contrast functions [Hyvarinen, 1998]. However, we still observed a deteriorating behavior when the noise level increases [Maino *et al.*, 2001].

To remove these drawbacks, in [Moulines *et al.*, 1997] it is suggested to model the sources with mixtures of Gaussians, and employ an EM-based technique to estimate the mixing matrix and the source distribution parameters. Attias named this formulation *independent factor analysis* [Attias, 1999]. An IFA algorithm is performed in two steps: in the first one, the mixing matrix is learned, along with the noise covariance matrix and the source density parameters. The adaptation of a Gaussian mixture model for the source densities makes the model analytically tractable and yet flexible, and enables one to use the EM algorithm for the estimation of the parameters. In the second step the sources are estimated using the posterior source densities obtained in the first step.

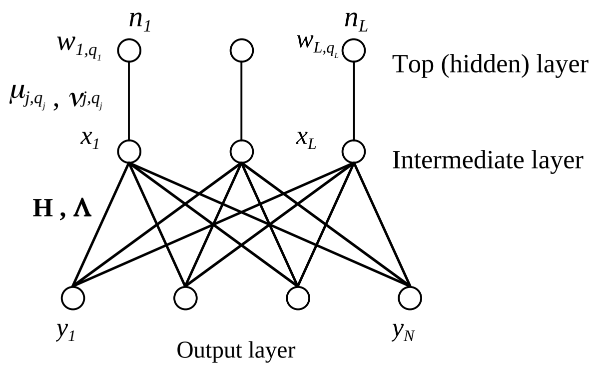


**Figure 5:** *IFA* data model

The data generation model assumed is depicted in Figure 5 [Attias, 1999]. The top layer of the network generates the independent sources **x** following a mixture of Gaussians model, with parameters $\theta_i = \{w_{i,q_i}, \mu_{i,q_i}, \nu_{i,q_i}\}$ for any element $x_i$ of the source vector, where $w_{i,q_i}$ is the probability of generating $x_i$ from the $q_i$-th Gaussian density, whose mean and variance are $\mu_{i,q_i}$ and $\nu_{i,q_i}$, respectively. The joint density of vector **x** is thus:

$$p(\mathbf{x} \mid \theta) = \prod_{i=1}^{L} p(x_i \mid \theta_i) = \sum_{\mathbf{q}} w_{\mathbf{q}} G(\mathbf{x} - \mu_{\mathbf{q}}, \mathbf{V}_{\mathbf{q}}) \tag{3}$$

where $w_{\mathbf{q}}$, is a vector collecting all the probabilities $w_{i,q_i}$ and function $G$ is an $L$-dimensional Gaussian density function with mean and variance vectors $\mu_{\mathbf{q}}$, $\mathbf{V}_{\mathbf{q}}$, respectively. The $i$-th element of the $L$-vector **q** is the index that determines which component in the $i$-th mixture has generated the sample.

The sensor signals **y** are generated from the intermediate layer of the network, through the mixing matrix **H** and the additive noise, which is zero-mean Gaussian with covariance matrix **Λ**. The probability of generating a particular sensor vector **y** given a source vector **x** is

$$p(\mathbf{y} \mid \mathbf{x}) = G(\mathbf{y} - \mathbf{H}\mathbf{x}, \mathbf{\Lambda}) \tag{4}$$

where $G$ is now an $N$-dimensional Gaussian and, in general, **Λ** may depend on any particular sample of the data vector realization; a space-varying noise can thus be easily modeled.

With some manipulations, on the basis of the data generation model and the Gaussian form of the source densities, it is possible to derive the density $p(\mathbf{y}|W)$ of the data vector conditioned to the model parameters **H**, $\theta$ and **Λ**, synthetically denoted by the vector $W$. The estimation of $W$ is then performed by minimizing the Kullback-Leibler divergence between $p(\mathbf{y}|W)$ and the measured data density $p^0(\mathbf{y})$:

$$J(W) = \int d\mathbf{y}\; p^0(\mathbf{y}) \log \frac{p^0(\mathbf{y})}{p(\mathbf{y} \mid W)} = -E[\log p(\mathbf{y} \mid W)] - H_{p^0} \tag{5}$$

where $E$ is the averaging operator over the observed data.

The model parameters are learned through a modified EM algorithm:
Maximization step:

$$\mathbf{H} = E\left[\mathbf{y}\langle\mathbf{x} \mid \mathbf{y}\rangle\right]\left(E\left\langle\mathbf{x}\mathbf{x}^{\mathbf{T}} \mid \mathbf{y}\right\rangle\right)^{-1} \tag{6}$$

$$\mathbf{\Lambda} = E\left[\mathbf{y}\mathbf{y}^{\mathbf{T}}\right] - E\left[\mathbf{y}\left\langle\mathbf{x}^{\mathbf{T}} \mid \mathbf{y}\right\rangle\mathbf{H}^{\mathbf{T}}\right] \tag{7}$$

where $\langle\cdot\rangle$ denotes the expectation operator.
Expectation step:

$$\mu_{i,q_i} = \frac{Ep(q_i \mid \mathbf{y})\langle x_i \mid q_i, \mathbf{y}\rangle}{Ep(q_i \mid \mathbf{y})} \tag{8}$$

$$\nu_{i,q_i} = \frac{Ep(q_i \mid \mathbf{y})\left\langle x_i^2 \mid q_i, \mathbf{y}\right\rangle}{Ep(q_i \mid \mathbf{y})} - \mu_{i,q_i}^2 \tag{9}$$

$$w_{i,q_i} = Ep(q_i \mid \mathbf{y}) \tag{10}$$

All the quantities in steps (6)-(10) can be calculated from the probability densities evaluated on the basis of the data model. No scheme is suggested in [Moulines *et al.*, 1997] for the estimation of the sources, while in [Attias, 1999] two schemes are suggested, namely, least squares and MAP estimation. For our experiments, we used the least squares estimation scheme:

$$\mathbf{x}^{LS}(\mathbf{y}) = \langle\mathbf{x} \mid \mathbf{y}\rangle = \int d\mathbf{x}\, \mathbf{x} p(\mathbf{x} \mid \mathbf{y}, W) \tag{11}$$

where the posterior $p(\mathbf{x}|\mathbf{y},W)$ can be calculated from the parameters and the density functions already evaluated.

## 4. Numerical Experiments

We ran various simulations on ideal data (the source samples were extracted from Gaussian mixture densities and uniform Gaussian noise was added), and found that the results in [Attias, 1999] were partly confirmed. When some of the parameters to be estimated are fixed at their correct values, and thus the number of unknowns is small, we observed a fast convergence to the optimal values. However, when all the parameters are left unknown, we observed a significant degradation in performance.

To test the technique in more realistic situations, we used data maps constructed from source maps of the type shown in Section 2. The first simulations we ran considered two measurement channels and two sources. We formed mixtures of CMB and synchrotron and of CMB and galactic dust, with the frequency coefficients related to the 70 and 100 GHz channels, and added space-varying noise at a level of 3% of the CMB radiation values. We observed that, when the mixture model parameters are fixed, for good starting points the algorithm finds the optimal **H**, and the sources are recovered successfully, despite the presence of noise. In Figure 6, it is shown the worst case examined, where the data contained CMB and dust radiation sources. When the mixture model parameters are also unknown, the algorithm fails in convergence and gets stuck in a local minimum, due to the complicated error-function surface.

This means that not all the promised improvements are achieved by IFA for the problem of separating radiometric astrophysical images. We should note, however, that this technique allows us to include in the problem any prior information on the source densities and any noise variance spatial configuration, thus avoiding some of the rigidities of the ICA approaches.

Further study is needed to assess the performance of this technique. We are now starting to apply optimization with simulated annealing instead of EM, to ensure global convergence.


## 5. Acknowledgments

We are indebted to the Planck teams in Bologna and Trieste, Italy, for supplying us with the maps used for our simulations.

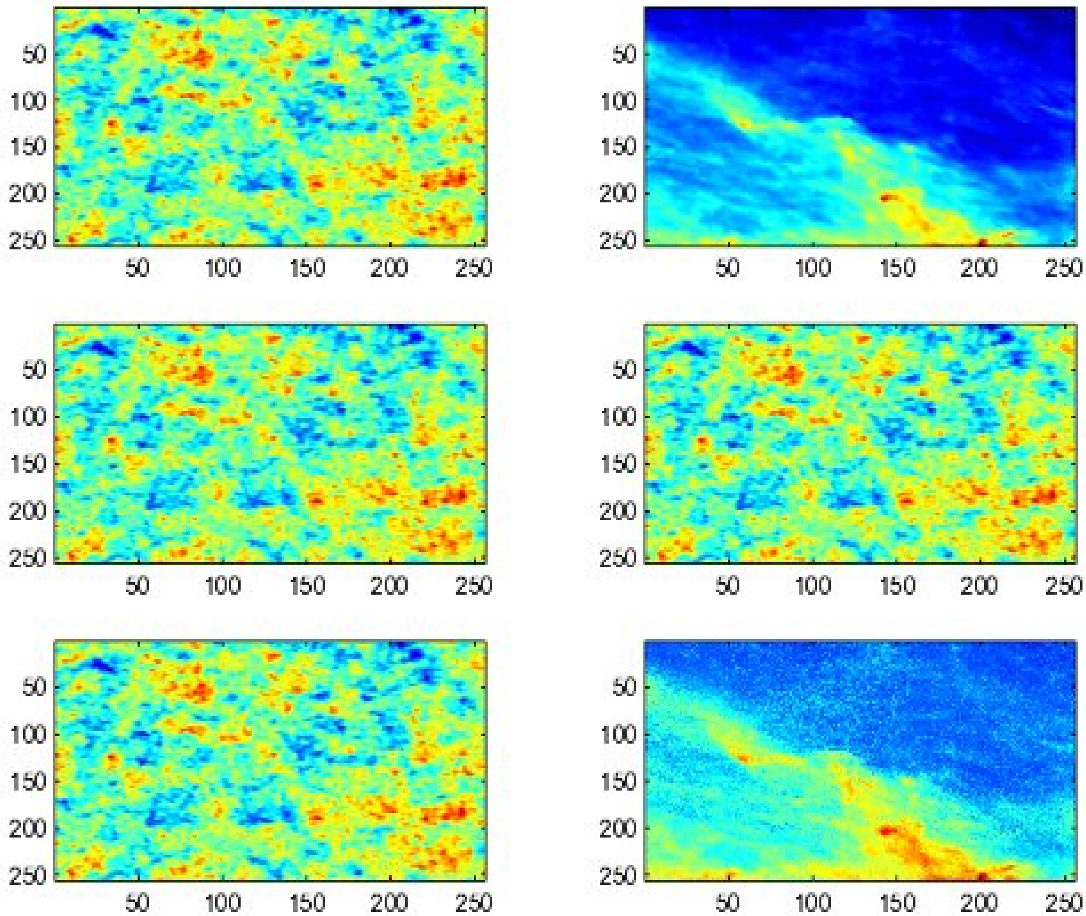


**Figure 6:** *IFA* experiment with two sources and two measurement channels, 3% noise. Left to right, top row, original source maps; second row, measurement maps at 70 and 100 GHz; bottom row, estimated sources.

**References**

Attias, H. 1999. Independent Factor Analysis. *Neural Computation*, 11:803-851.

Baccigalupi, C. Bedini, L. Burigana, C. De Zotti, G. Farusi, A. Maino, D. Maris, M. Perrotta, F. Salerno, E. Toffolatti, L. and Tonazzini, A. 2000. Neural Networks and the Separation of Cosmic Microwave Background and Astrophysical Signals in Sky Maps. *Monthly Notices of the Royal Astronomical Society*, 318:769-780.

Dempster, E. J. Laird, N. M. and Rubin, D. B. 1977. Maximum Likelihood from Incomplete Data via EM Algorithm. *Annals of the Royal Statistical Society*, 39:1-38.

Hyvarinen, A. 1998. Noisy Independent Component Analysis, Maximum Likelihood Estimation, and competitive Learning. In *1998 IEEE International Joint Conference on Neural Networks Proceedings, IEEE World Congress on Computational Intelligence*, Vol. 3, pp. 2282-2287.

Maino, D. Farusi, A. Baccigalupi, C. Perrotta, F. Banday, A. J. Bedini, L. Burigana, C. De Zotti, G. Górski, K. M. and Salerno, E. 2001. All-sky astrophysical component separation with Fast Independent Component Analysis (FastICA). *Monthly Notices of the Royal Astronomical Society*, to appear (astro-ph/0108362 22 Aug. 2001).

Moulines, E. Cardoso, J. F. and Gassiat E. 1997. Maximum Likelihood for Blind Separation and Deconvolution of Noisy Signals Using Mixture Models. In *Proceedings of ICASSP'97*, Vol. 5, pp. 3617-3620.