Independent Factor Analysis for Component Separation from Planck Channel Maps

Ercan E. Kuruoğlu, Luigi Bedini, Maria T. 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,
Pisa, 56124, ITALY
kuruoglu@iei.pi.cnr.it

Corresponding author

ABSTRACT

A microwave sky map results from a combination of signals from various astrophysical sources, such as cosmic microwave background radiation, synchrotron radiation and galactic dust radiation. To derive information about these sources, one needs to separate them from the measured maps on different frequency channels. This task is made difficult by our insufficient knowledge of the weights to be given to the individual signals at different frequencies. Recent work on the problem led to only limited success due to ignoring the noise and to the lack of a suitable statistical model for the sources. In this paper, we derive the statistical distribution of some source realizations, and check the appropriateness of a Gaussian mixture model for them. A source separation technique, namely independent factor analysis, had been suggested recently in the literature for Gaussian mixture sources in the presence of noise. This technique employs a three layered neural network architecture which allows a simple, hierarchical treatment of the problem. We modify the algorithm proposed in the literature to accommodate for space-varying noise and test its performance on simulated astrophysical maps. We also compare the performances of the expectation-maximization and the simulated annealing learning algorithm in the estimation of the mixing parameters. The simulation results demonstrate the success of the independent factor analysis approach with simulated-annealing learning, which proves better than the expectation-maximization learning especially for higher noise levels and when the independence of the performance from starting points is considered.

KEYWORDS

Astrophysical image processing, independent factor analysis, independent component analysis, blind source separation, cosmic microwave background radiation.

I. Introduction

The microwave signal coming from the celestial sphere contain very important information about our Universe. Unfortunately, this information cannot be obtained in an easily interpretable way. A radiometric image taken at any working frequency is a superposition of radiations coming from different sources, corrupted by the detector noise, which is normally space-varying and difficult to remove. The classical components of the microwave sky radiation are the cosmic microwave backgorund (CMB), the galactic dust radiation, the synchrotron radiation, extra galactic radio sources and the free-free radiation. Each of these radiations has its own interest in astrophysics. For example, an important open problem in cosmology is the measurement of the CMB anisotropies at high angular resolution. The CMB radiation is a strong affirmation of the hot big bang model. According to this model, the Universe was once hot and dense and has been expanding and cooling since then. After the discovery of CMB it was realized that the presence of density fluctuations at the last scattering epoch would induce angular anisotropy in the CMB intensity. A high-resolution and high-sensitivity measurement of CMB anisotropies would enable cosmologists to assess the validity of the present competing cosmological theories. Similarly, maps of the synchrotron emission of the galaxy carry information about the structure of the magnetic field, the spatial and energy distribution of relativistic electrons and the variations of electron density, of electron energy and of magnetic field induced by supernova shocks. To explore correlations between magnetic fields and the matter distribution, it is also vital to obtain the snychrotron map. Since each source has its own importance, rather than just looking for a CMB map as clean as possible, we should try to extract each source from the mixture maps. At each measurement frequency, the data map is a superposition of the individual sources, multiplied by mixing coefficients depending on the instrument features and on the radiative properties of the source processes. The mixing coefficients are normally unknown, and this is the main difficulty in solving our separation problem: we need to perform a blind source separation.

One classical approach to blind source separation is based on the independent component analysis (ICA) concept [13]. Assuming that all the source processes are statistically independent, any ICA technique transforms the observed data in order to ensure that the

output signals are independent. It can be shown that these outputs provide an estimate for the initial sources. Most of the ICA formulations have been derived in the case where the number of sources and the number of sensors are equal and when the observations do not contain noise. Unfortunately, the assumption of negligible noise is invalid for our problem. There exist attempts in the ICA literature to include the noisy case in the analysis. One of them led to the so-called noisy fastICA algorithm [6] [7]. Unfortunately, our earlier work on source separation of astrophysical images has shown that this algorithm has significantly deteriorating performance when the noise level is increased [10].

The fact that ICA has a completely blind estimation approach is another reservation we have about this technique. Indeed, we possess some crucial information about the statistical distributions of the sources and the noise and we believe that all the available prior knowledge should be exploited to reach a good solution to the separation problem. In this paper, we present a method that incorporates prior information about the sources in a very generic way. The method is called independent factor analysis (IFA) and has been introduced recently by two papers, one in signal processing [11] and one in neural networks literature [1]. The novelty of the technique is in proposing a generic model for the source densities, namely the Gaussian mixture model, and in providing a neural network architecture that is specially convenient for learning through an expectation maximization (EM) algorithm. The noise is also taken into account in the mixture generation model in a very natural way. Therefore, IFA provides an alternative to noisy ICA for solving our particular problem.

Despite this attractive appearance of IFA, the numerical studies 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. In this paper, we study this technique in the context of simulated but realistic sky radiation maps and try to identify its potentials. The data maps we used for our experiments simulate the ones expected from the *Planck Surveyor Satellite* mission, which will be launched in 2007 by the European Space Agency. The aim of this mission is to accurately map the CMB 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 extrapo-

lated 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.

In our work, we also addressed the points where IFA, as suggested in [11] and [1], has clear drawbacks. In particular, we extended the IFA algorithm to accommodate for the location-dependent noise. Moreover, we developed a globally convergent simulated annealing alternative to the EM learning algorithm, which in our case demonstrated some difficulties with its locally convergent behavior.

The rest of the paper is organised as follows: in Section 2 the statistical distributions of some important radiation sources are studied, and the suitability of a Gaussian mixture model is shown. We also show a typical noise pattern for high-resolution radiometric sky maps. Section 3 briefly describes the IFA algorithm and provides details about its implementation. Section 4 provides simulation studies first on synthetic data and then on simulated astrophysical data, using the EM learning algorithm. In Section 5, the alternative simulated annealing (SA) learning algorithm is presented. Section 6 summarizes our remarks and our research program for the future.

II. Source Models

As mentioned above, our data images contain radiation from various sources. The important ones include cosmic microwave background radiation, synchrotron, galactic dust, radiation from extra galactic radio sources, free-free emission and thermal clusters. In this section we will look into the amplitude distributions of the most important ones, to see whether we can suggest a generic statistical model for all of them.

A. CMB Radiation

The standard inflationary cosmological theory tells us that the CMB anisotropies are Gaussian distributed. In Figure 1.a, we show a synthetic CMB map, generated following the standard theory, on a square sky patch with a 15^o aperture.

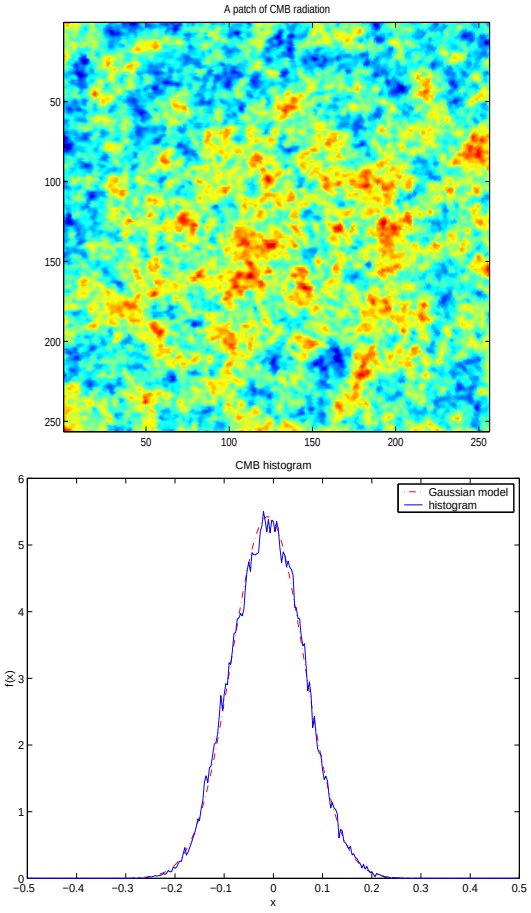

Fig. 1. a) A patch of CMB data, b)Histogram of a typical CMB patch

B. Galactic Dust

The galactic dust emission is expected to be significant in the high-frequency region of our measurement band. The simultaneous study of dust and free-free emission would be useful to investigate the relationships between different phases of the interstellar medium.

A map of the galactic dust radiation is shown in Figure 2.a. The sky-patch assumed is the same as the one used for the CMB radiation shown above, and the temperature values have been extrapolated from available measurements at frequencies outside our range. The histogram of this dust map is shown in Figure 2.b.

It is clear from the histogram that the amplitude distribution of galactic dust is non-Gaussian. The curve actually is multimodal and unsymmetric. Motivated by the fact that radial basis function networks are universal approximators [12], we attempted to model

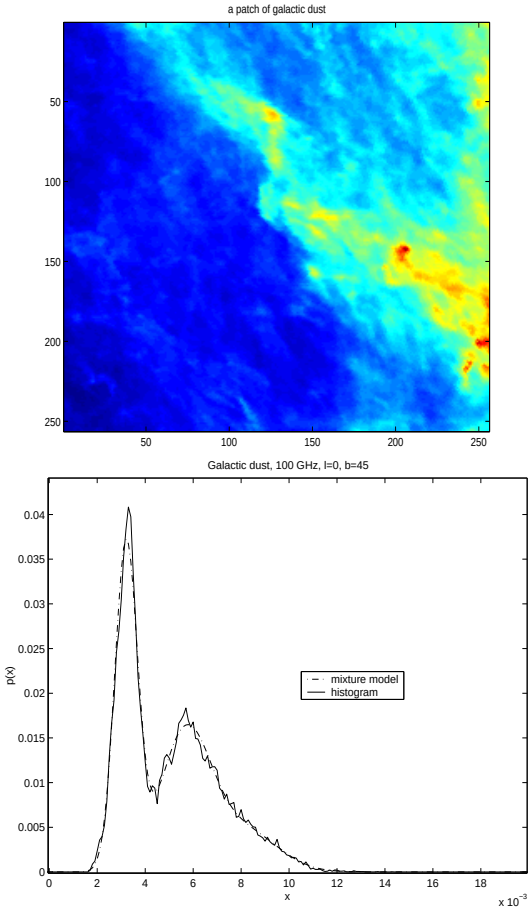

Fig. 2. a) A patch image showing galactic-dust intensity, b) Histogram and the Gaussian mixture model fit for the galactic-dust intensity

the histogram by a mixture of Gaussian densities. The parametric model used to fit the experimental curve is thus the following:

$$p_X(x) = \Sigma_i w_i \frac{1}{\sqrt{2\pi}\nu_i} \exp(-\frac{(x-\mu_i)^2}{2\nu_i^2}). \quad (1)$$

where the parameters to be estimated are the weights w_i, whose sum over i must be equal to one, the mean values μ_i, and the standard deviations ν_i. Although Gaussian densities are symmetric functions, it is possible to represent unsymmetric curves by choosing the mixture parameters carefully [12].

We used an expectation-maximization (EM) algorithm to fit a Gaussian mixture to the histogram. The details of the EM algorithm can be found in [5]. The result is shown in

Figure 2.b, plotted with the dashed curve. As can be seen, the fit is very close to the histogram. It is also interesting to note that we used only three Gaussian components in the mixture, which is far less than we expected. We repeated our experiments on about 15 different sky patches 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. We conclude that Gaussian mixture models are very efficient for modelling galactic dust intensity distribution. Since the dust emission is only related to our galaxy, its distribution varies sensibly in those regions that are far from the galactic plane, and the mixture parameters are normally different for every different sky patch.

C. Synchrotron

The galactic synchrotron emission carries information about the structure of the magnetic field, the spatial and energy distribution of relativistic electrons and the variations of electron density, of electron energy and of magnetic field induced by supernova shocks. Since this radiation has a power-law spectrum, its influence on the total measured field is significant in the low-frequency measurement channels. In Figure 3.a a synchrotron map extrapolated from already available observations is shown. The related histogram is given in Figure 3.b, plotted with solid curve.

Again, we tried to fit the curve by a Gaussian mixture using the EM algorithm. The result obtained by fitting a Gaussian mixture of only four components is given in Figure 3.b, plotted with the dashed curve. As can be seen, the fit is very good. We tried this procedure for 15 different patches of the same size from different locations in the sky, and observed that in all cases the Gaussian mixture model provides a good fit with less than five components. Hence, we conclude that Gaussian mixtures are good representatives of the amplitude distribution of synchrotron emission.

D. Noise distribution

Although noise is significant in our application, in our earlier research we assumed a noiseless model and adopted the ICA approach [2]. Even when noise is taken into account, it is normally assumed to be Gaussian and space-invariant. In our problem, however, noise

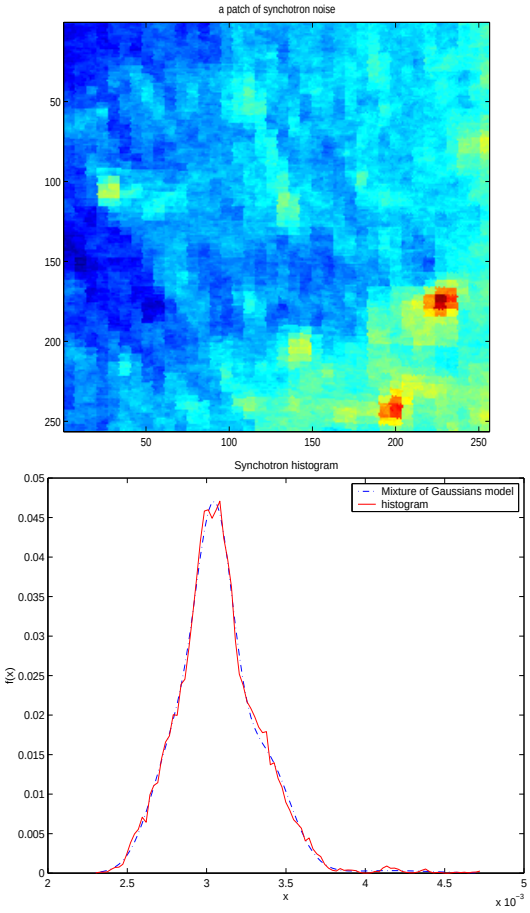

Fig. 3. a) The synchrotron intensity from the same patch, b) histogram and the Gaussian mixture model fit of the synchrotron intensity image

is considered Gaussian but space-varying, since the value given to each pixel of the data maps is obtained from multiple measurements in the same direction, and the number of measurements is normally different for different pixels. To clarify this characteristics, we provide a plot of the noise samples (Figure 4.a) and the RMS map (Figure 4.b) from the same sky patch we have been considering for various sources in the previous sections. The pattern due to non-uniform scanning of the sensor is apparent.

III. Independent Factor Analysis

The model we assumed to describe the relationship between the astrophysical source processes and our data maps is a simple instantaneous linear combination, which holds

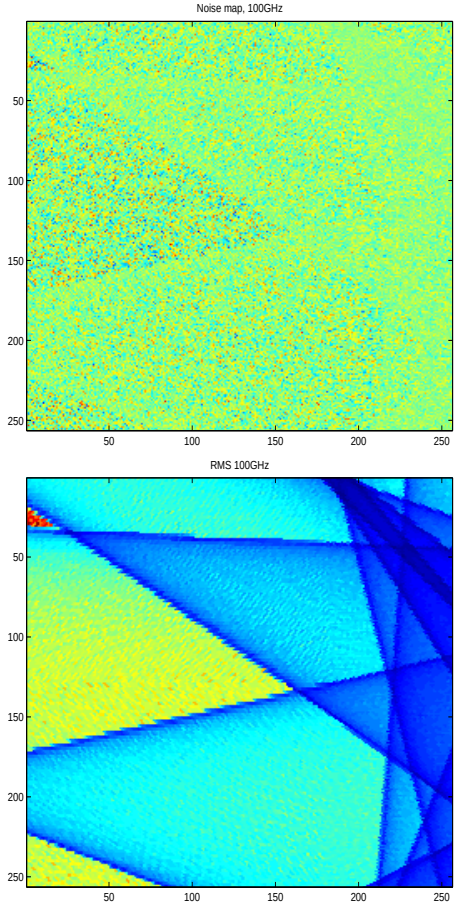

Fig. 4. a) Noise samples (one realization), b) RMS map of the underlying space-varying Gaussian distribution.

true for any pixel:

$$y_i = \Sigma_{j=1}^{L} H_{ij} x_j + n_i, \qquad i = 1, \ldots, N \tag{2}$$

For the generic pixel considered, y_i is the value of the mixture signal over the i-th of the N channels, x_j is the contribution of the j-th source process to the i-th channel measurement, n_i is the noise over the i-th channel. We would like to obtain H_{ij} and x_j from observations y_i only; this is the classical blind source separation problem. Various efforts have been made to solve this problem in the last decade. In particular, an important approach has been independent component analysis [3], [4], which aims at decomposing the observations into independent sources and has been studied widely in the literature. The ICA model is highly idealized: the mixing is often assumed to be square ($L = N$) and invertible, and

the possible presence of noise is not considered. However, in real situations noise is always present, and the number of observations is often different from the number of sources. Efforts have been made to include noise into the analysis: in his negentropy maximization approach, Hyvarinen suggested employing suboptimal nonlinearities to approximate the negentropy function [6], [7]. He suggested using a simple preprocessing strategy and particular contrast functions that are not sensitive to Gaussian noise in the data. However, his formulation is not optimal and we observed a deteriorating behavior when the noise level increases [10].

Moreover, existing algorithms usually employ a source model that is either fixed or has limited flexibility. One should instead use a flexible model, whose parameters can be estimated with the maximum likelihood estimation method. Also, most of the ICA algorithms use gradient maximization methods, which result in slow convergence.

Towards removing these drawbacks, Moulines *et al.* (1997) [11] and, later, Attias (1999) [1] suggested modeling the sources with mixtures of Gaussians and employed EM-based techniques to estimate the mixing and the noise-covariance matrices, and the source distribution parameters. Attias named this formulation *independent factor analysis* (IFA). IFA is performed in two steps: in the first one, the IF model is trained to learn the mixing matrix, the noise covariance and the source density parameters. To make the model analytically tractable and yet flexible, a Gaussian mixture model is adopted for the source densities. Note that this is well in accordance with our problem, since in the previous section we demonstrated that the main source components are well approximated by Gaussian mixture models. This formulation enables one to use the EM algorithm for the estimation of the parameters. In the second step the sources are estimated by using the posterior source densities obtained in the first step.

The IFA mixture model can be represented as in Figure 5. Following the figure, we provide here a brief intuitive description of IFA rather than a lengthy derivation, for which the readers are referred to [11] and [1]. The top level of the neural network provides a collection of Gaussian kernels, each parametrized by a mean μ_{i,q_i} and a standard deviation ν_{i,q_i}, and weighted by the coefficient w_{i,q_i}. To have a normalized probability density function for each node, the sum of the w_{i,q_i} for each i must be equal to one. The hidden middle level

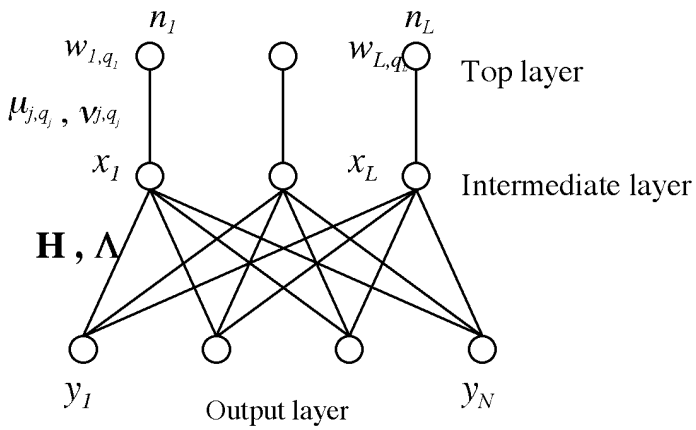

Fig. 5. IFA model

nodes contain the independent sources x_i, which are assumed to have Gaussian mixture distributions.

$$p(\mathbf{x}|\theta) = \Pi_{i=1}^{L} p(x_i|\theta_i) = \Sigma_{\mathbf{q}} w_{\mathbf{q}} \mathcal{G}(\mathbf{x} - \mu_{\mathbf{q}}, \mathbf{V_q}) \tag{3}$$

where $\theta_i = \{w_{i,q_i}, \mu_{i,q_i}, \nu_{i,q_i}\}$, $w_{\mathbf{q}}$ is the vector of the active weights for each mixture, and $\mathcal{G}$ represents the L-dimensional Gaussian density with mean vector $\mu_{\mathbf{q}}$ and covariance matrix $\mathbf{V_q}$.

The transition from the top hidden level to the middle hidden level is characterized by the probability distribution $p(\mathbf{q})$, where $\mathbf{q}$ is a random vector indicating the index of the Gaussian kernel which the current sample of the source x_i comes from. The marginal probability of the i-th element of $\mathbf{q}$ is

$$p(q_i) = w_{i,q_i}, \qquad i = 1, \ldots, N, \tag{4}$$

As already said, to form a proper probability function, the coefficients μ_{i,q_i} should sum up to 1 for any i.

$$\sum_{q_i} w_{i,q_i} = 1, \qquad i = 1, \ldots, N. \tag{5}$$

Summarizing, the collective hidden states are given by:

$$\mathbf{q} = (q_1, \ldots, q_L) \tag{6}$$

$$\mathbf{w}_q = \Pi_{i=1}^{L} w_{i,q_i} = w_{1,q_1} \times \ldots \times w_{L,q_L} \tag{7}$$

$$\mu_{\mathbf{q}} = (\mu_{1,q_1}, \ldots, \mu_{L,q_L}) \quad (8)$$

$$\mathbf{V_q} = diag(\nu_{1,q_1}, \ldots, \nu_{L,q_L}). \quad (9)$$

The middle hidden layer is characterised by the conditional probability of generating a particular source vector $\mathbf{x}$ given $\mathbf{q}$:

$$p(\mathbf{x}|\mathbf{q}) = \mathcal{G}(\mathbf{x} - \mu_{\mathbf{q}}, \mathbf{V_q}). \quad (10)$$

The observations are simply a linear mixture of the sources summed with noise.

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (11)$$

In our case, we assume that the noise is zero-mean Gaussian and location dependent. Its probability density function is

$$p(\mathbf{n}) = \mathcal{G}(\mathbf{n}, \mathbf{\Lambda}), \quad (12)$$

The transition from the middle level to the bottom level, which contains the observations, is thus characterized by the mixing matrix $\mathbf{H}$ and the noise covariance matrix $\mathbf{\Lambda}(.,.)$, which is assumed to be space dependent. This means that Equation 11, as Equation 2, holds true for each pixel of the data maps, but in general each pixel is affected by a different noise amount. Denoting the collective parameters as $W = (\mathbf{H}, \mathbf{\Lambda}, \theta)$, the resulting density for the observations is

$$p(\mathbf{y}|W) = \int d\mathbf{x}\ p(\mathbf{y}|\mathbf{x})\ p(\mathbf{x}) \quad (13)$$

$$= \int d\mathbf{x}\ \mathcal{G}(\mathbf{y} - \mathbf{H}\mathbf{x}, \mathbf{\Lambda})\ \Pi_{i=1}^{L} p(x_i|\theta_i) \quad (14)$$

and can be calculated conveniently from

$$p(\mathbf{y}|W) = \Sigma_{\mathbf{q}} p(\mathbf{q}) \int d\mathbf{x}\ p(\mathbf{x}|\mathbf{q}) p(\mathbf{y}|\mathbf{x}) = \Sigma_{\mathbf{q}}\ p(\mathbf{q})\ p(\mathbf{y}|\mathbf{q}) \quad (15)$$

where

$$p(\mathbf{y}|\mathbf{q}) = \mathcal{G}(\mathbf{y} - \mathbf{H}\mu_{\mathbf{q}}, \mathbf{H}\mathbf{V_q}\mathbf{H}^T + \mathbf{\Lambda}) \quad (16)$$

Next step is to define an error function that measures the difference between the data density in Equation 15 and the density calculated from the actual data maps. The Kullback-Leibler (KL) divergence has been chosen in both [11] and [1] to evaluate this measure:

$$J(W) = \int d\mathbf{y}\ p^0(\mathbf{y}) \log \frac{p^0(\mathbf{y})}{p(\mathbf{y}|W)} = -E[\log p(\mathbf{y}|W)] - H_{p^0}. \quad (17)$$

To minimize this divergence, Moulines *et al.* [11] and Attias [1] have suggested techniques which are extensions of the EM algorithm. The details can be found in the respective papers. Although attractive, this algorithm shares the important drawbacks of the EM algorithm, that is, the convergence is local and slow. Additionally, their algorithm was developed without taking special care of the time (or space)-varying noise.

We developed a simple extension of their derivation to take the space-varying noise into account [8]. This extension, however, further slows convergence down, since the pixels must be considered one by one.

A. Source estimation

The next step is the estimation of the source maps, using the parameters estimated by minimizing the KL-divergence in Equation (17). Attias suggests two schemes to perform this step [1], namely, least squares and MAP estimation.

A.1 Least squares estimation of the sources

This estimate is given by the expectation of the sources, given the observation data. If $< \cdot|\cdot >$ denotes the conditional expectation operator, we have

$$\mathbf{x}^{LS}(\mathbf{y}) =< \mathbf{x}|\mathbf{y} >= \int d\mathbf{x}\ \mathbf{x}\ p(\mathbf{x}|\mathbf{y}, W) \tag{18}$$

where

$$p(\mathbf{x}|\mathbf{y}, W) = \Sigma_{\mathbf{q}} p(\mathbf{q}|\mathbf{y}) p(\mathbf{x}|\mathbf{q}, \mathbf{y}) \tag{19}$$

Evaluating the integral, we obtain:

$$\mathbf{x}^{LS}(\mathbf{y}) = \Sigma_{\mathbf{q}}(\mathbf{A}_{\mathbf{q}}\mathbf{y} + \mathbf{b}_{\mathbf{q}}) \tag{20}$$

where

$$\mathbf{A}_{\mathbf{q}} = (\mathbf{H}^T \mathbf{\Lambda}^{-1} \mathbf{H} + \mathbf{V}_{\mathbf{q}})^{-1} \mathbf{H}^T \mathbf{\Lambda}^{-1} \tag{21}$$

$$\mathbf{b}_{\mathbf{q}} = (\mathbf{H}^T \mathbf{\Lambda}^{-1} \mathbf{H} + \mathbf{V}_{\mathbf{q}})^{-1} {\mathbf{V}_{\mathbf{q}}}^{-1} \mu_{\mathbf{q}} \tag{22}$$

A.2 MAP Estimator

The posterior density of the sources given the observation data is maximized

$$\mathbf{x}^{MAP}(\mathbf{y}) = \mathrm{a}rgmax_{\mathbf{x}}[\log p(\mathbf{y}|\mathbf{x}) + \Sigma_{i=1}^{L} \log p(x_i)]. \tag{23}$$

Attias suggests solving this equation by a gradient ascent technique, where the estimate is incremented at each iteration by

$$\delta\hat{\mathbf{x}} = \eta \mathbf{H}^T \lambda^{-1}(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}) - \eta\phi(\mathbf{x}) \tag{24}$$

where η is the learning rate and

$$\phi(x_i) = -\frac{\partial \log p(x_i)}{\partial x_i} = -\Sigma_{q_i=1}^{n_i} p(q_i|x_i) \frac{x_i - \mu_{i,q_i}}{\nu_{i,q_i}} \tag{25}$$

IV. Simulation Results with EM Learning

In this section, we report our IFA performance study, first with synthetic data and then with our simulated astrophysical data.

A. Experiments with Synthetic Data

As a first step, we aimed at understanding the behavior of IFA on the recovery of synthetic sources sampled from Gaussian mixture densities. We generated two synthetic source sequences of length 30000 with the following parameters:

$$\begin{aligned} \mu_1 &= [0 \quad 1 \quad -1], & \mu_2 &= [0 \quad 2 \quad -2], \\ \sigma_1 &= [0.010 \quad 0.200 \quad 0.200], & \sigma_2 &= [0.020 \quad 0.300 \quad 0.300], \\ \mathbf{w}_1 &= [0.5385 \quad 0.23080.2308], & \mathbf{w}_2 &= [0.5714 \quad 0.2143 \quad 0.2143]. \end{aligned} \tag{26}$$

The noise in both channels generated from a Gaussian distribution with mean $\mu_n = 0$ and standard deviation $\sigma_n = 0.01$.

We tried to estimate all parameters using the EM algorithm suggested by Attias [1]. The mixing matrix was

$$\mathbf{H}_0 = \begin{bmatrix} 0.7000 & 0.5000 \\ 0.6000 & 0.8000 \end{bmatrix}, \tag{27}$$

q	channel	$\mathbf{w}$	$\hat{\mathbf{w}}$
1	1	0.5385	0.1403
	2	0.5714	0.0000
2	1	0.2308	0.7105
	2	0.2143	0.0000
3	1	0.2308	0.1492
	2	0.2143	1.0000

TABLE I

COMPARISON OF REAL AND ESTIMATED MIXTURE COEFFICIENTS

and the result obtained

$$\mathbf{H} = \begin{bmatrix} 0.4686 & 0.7347 \\ 0.3622 & 1.1483 \end{bmatrix}. \tag{28}$$

Although the estimated mixture matrix is different from the actual one, one should keep in mind that IFA can estimate only up to a scale factor. If we take the ratio of the elements of $\mathbf{H}$ and $\mathbf{H}_0$ we obtain:

$$\frac{\mathbf{H}}{\mathbf{H}_0} = \begin{bmatrix} 0.6695 & 1.4695 \\ 0.6037 & 1.4354 \end{bmatrix}. \tag{29}$$

Note that the elements in the same column are scaled with approximately the same factors. This good performance is not repeated in the estimation of the source parameters. The results are provided in Tables I, II, and III. These estimates are clearly very poor.

Although the above results could seem encouraging as far as the estimation of $\mathbf{H}$ is concerned, we observe that the noise level was in this case very low. Increasing σ_n would also result in an unsatisfactory estimate for the mixing matrix.

In real situations, some prior knowledge could be available on the source distribution. This is the case, for example, with the particular application we are considering here. Then, we assumed to have a perfect knowledge of the coefficients in the source model, and assessed

q	channel	μ	$\hat{\mu}$
1	1	0.0	1.6902
	2	0.0	3.6228
2	1	1.0	-0.0321
	2	2.0	3.6228
3	1	-1.0	-1.7308
	2	-2.0	0.4244

TABLE II

COMPARISON OF REAL AND ESTIMATED MEANS OF THE MIXTURE COMPONENTS

q	channel	σ	$\hat{\sigma}$
1	1	0.01	0.1382
	2	0.02	0.0014
2	1	0.2	0.1537
	2	0.3	0.0014
3	1	0.2	0.1664
	2	0.3	1.0000

TABLE III

COMPARISON OF REAL AND ESTIMATED STANDARD DEVIATIONS OF THE MIXTURE COMPONENTS

the performance of IFA with EM learning in the case where the only mixing matrix $\mathbf{H}$ is left unknown. We thus fixed the source parameters at the values shown in Equation (26), calculated the data maps by using the same mixing matrix as in Equation (27), and finally tried to estimate $\mathbf{H}$ by means of the EM algorithm. In Figure 6, we give the source histograms before and after mixing, and the histograms of the estimates after 100 EM iterations, again, with $\sigma_n = 0.01$. The estimated mixing matrix was

$$\mathbf{H} = \begin{bmatrix} 0.6690 & 0.4600 \\ 0.5692 & 0.7515 \end{bmatrix}. \tag{30}$$

$$\frac{\mathbf{H}}{\mathbf{H}_0} = \begin{bmatrix} 0.9557 & 0.9200 \\ 0.9487 & 0.9394 \end{bmatrix}. \tag{31}$$

It is clear that this is a very good estimate. The performance of this algorithm as a function of noise was studied by defining the following matrix

$$J = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{H}_0. \tag{32}$$

The ratio between the mean square value of the off-diagonal elements to the mean square value of the diagonal elements in J is assumed as an indicator of the estimation error. In Figure 7, we show the behavior of this indicator as a function of the noise level.

B. Experiments with realistic Data

We also investigated the performance of the EM learning algorithm on our simulated astrophysical images. We started with experiments on mixtures of two sources only, first on mixtures of CMB and galactic dust, then on mixtures of CMB and synchrotron. The noise was assumed space-variant, with an RMS map of the type shown in Figure 4.b. When all the parameters are assumed unknown, the technique is not able to separate the source maps. Figure 8 gives one particular result as an example. We have also seen that the mixing matrices were not estimated correctly. We believe that the reason for this behavior is that EM algorithm got stuck at a local minimum.

We then ran simulations for the case where all mixture parameters were fixed and the mixing matrix was left unknown. The mixture parameters were set to the values suggested by our prior knowledge about the sources. We observed that the convergence of the EM

learning algorithm depends on the initial points. Although the convergence is fast, it might be to a local optimum. The original mixing matrix was as follows:

$$\mathbf{H}_0 = \begin{bmatrix} 1.0000 & 1.0000 \\ 1.1400 & 0.6800 \end{bmatrix}. \tag{33}$$

And, for a good initial estimate, the algorithm converged to the following matrix

$$\mathbf{H} = \begin{bmatrix} 1.1623 & 1.2985 \\ 1.3251 & 0.8109 \end{bmatrix}. \tag{34}$$

The ratio of the estimated to the original matrix is

$$\frac{\mathbf{H}}{\mathbf{H}_0} = \begin{bmatrix} 1.1623 & 1.2985 \\ 1.1624 & 1.1925 \end{bmatrix}. \tag{35}$$

Figure 9 shows the CMB and dust sources, their mixtures, and the source estimates obtained by IFA. One should also note that we modified the IFA algorithm given in [1] to take non-stationary noise into account.

The results shown in this section suggest us that EM learning is not an ideal technique for realising IFA estimation when the number of unknown parameters is too large. On one hand, EM is known only to assure convergence to a stationary point. This means that, when a large number of unknowns increases the number of local minima, the choice of the starting point is crucial for a correct convergence. As mentioned above, this consideration has been confirmed by our experiments. Based on these observations, we decided to attempt minimizing the error function using a global optimization technique, such as simulated annealing.

V. Simulation Results with SA Learning

The simulated annealing (SA) optimization strategy was suggested in the 80's as a global optimization algorithm and obtained success in wide application areas [9]. The IFA formulation specifically lends itself to be implemented by a simulated annealing algorithm. We are carrying out simulations on synthetic data, but at present we are still having problems to tune the annealing schedule for the case where all the parameters are left unknown. We observed, however, that SA is more robust than EM against noise when the

only the mixing matrix is to be estimated. As an example, when the noise RMS value was set at 20% of the RMS measured signal, EM did not obtain any useful result. Conversely, a fairly good estimate of $\mathbf{H}$ was obtained by SA.

For one case, with CMB-synchrotron mixtures, the true matrix was

$$\mathbf{H}_0 = \begin{bmatrix} 1.0000 & 1.0000 \\ 1.1400 & 2.8100 \end{bmatrix}, \tag{36}$$

and the estimation result was

$$\mathbf{H} = \begin{bmatrix} 1.0266 & 0.7122 \\ 1.1616 & 2.4025 \end{bmatrix}, \tag{37}$$

and

$$\frac{\mathbf{H}}{\mathbf{H}_0} = \begin{bmatrix} 1.0266 & 0.7122 \\ 1.0189 & 0.8550 \end{bmatrix}. \tag{38}$$

For one case with CMB-dust mixtures, we had

$$\mathbf{H}_0 = \begin{bmatrix} 1.0000 & 1.0000 \\ 1.1400 & 0.6800 \end{bmatrix}, \tag{39}$$

$$\mathbf{H} = \begin{bmatrix} 0.9630 & 0.7450 \\ 1.1035 & 0.3834 \end{bmatrix}, \tag{40}$$

$$\frac{\mathbf{H}}{\mathbf{H}_0} = \begin{bmatrix} 0.9630 & 0.7450 \\ 0.9680 & 0.5638 \end{bmatrix}. \tag{41}$$

As is easily seen, the first column of $\mathbf{H}$, which contains the mixing coefficients for CMB is always estimated fairly well. The second column contains the mixing coefficients for the synchrotron and the dust maps, respectively, in the two cases. These radiations are sensibly weaker, and thus more affected by noise, than the CMB radiation. For this reason, the errors in the estimation of the second column of $\mathbf{H}$ are larger than the ones for the first column.

VI. Discussion

In this paper, we studied the performance of the independent factor analysis technique for component separation in astrophysical images. We first studied the statistical distributions of some typical astrophysical sources and showed that the Gaussian mixtures provide a very suitable model. Then, we studied the performance of IFA with an EM learning algorithm on synthetic data and showed that, although it is very successful when the parameter space is small, it shows a severe performance degradation when the size of the problem increases. When all the parameters are left unknown, this observation has been confirmed by our experience with images that realistically simulate astrophysical data. In this case, choosing good starting points could improve the estimation, but our experience showed that the error surface is too complicated to make feasible the search for such starting points. However, if the mixture parameters are fixed to *a priori* known values the convergence is reasonably fast and accurate, even though it results highly degraded when the noise level increases.

We are not yet able to say whether an SA learning is capable to avoid these drawbacks. As far as noise is concerned, we noted that SA performs better than EM, independently of the starting point chosen.

Another side issue is that, in our discussions, we ignored the spatial correlation in the images. Actually, the correlation between neighboring cells potentially carries important information, and the inclusion of spatial dependence in the source model is one of our future research directions.

Acknowledgements

The authors are indebted to the Planck teams in Bologna and Trieste, Italy, for supplying the maps used for our simulations.

References

[1] H. Attias. Independent factor analysis. *Neural computation*, 11:803–851, 1999.

[2] C. Baccigalupi, L. Bedini, C. Burigana, G. De Zotti, A. Farusi, D. Maino, M. Maris, F. Perrotta, E. Salerno,

L. Toffolatti, and A. Tonazzini. 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, 2000.

[3] A. J. Bell and T.J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[4] J. F. Cardoso and B.H. Laheld. Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, 44:3017–3030, 1996.

[5] E. J. Demspter, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via em algorithm. *Annals of Royal Statistical Society*, 39:1–38, 1977.

[6] A. Hyvarinen. 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*, volume 3, pages 2282–7, 1998.

[7] A. Hyvarinen. Gaussian moments for noisy independent component analysis. *IEEE Signal Processing Letters*, 6:145–147, 1999.

[8] E. E. Kuruo˘ glu, M. T. Paratore, L. Bedini, E. Salerno, A. Tonazzini, and A. Farusi. Independent factor analysis for component separation from planck channel maps. *IEI-CNR, Pisa, Technical Report*, TR-43-01, 2001.

[9] P. J. M. Laarhoven and E. H. L. Aarts. *Simulated Annealing: Theory and Applications.* D. Reidel Publishing, 1987.

[10] D. Maino, A. Farusi, C. Baccigalupi, F. Perrotta, A.J. Banday, L. Bedini, C. Burigana, G. De Zotti, K.M. Grski, and E. Salerno. All-sky astrophysical component separation with fast independent component analysis (fastica). *IEI-CNR, Pisa, Technical Report*, TR-31-01 (astro-ph/0108362 22 Aug. 2001), 2001.

[11] E. Moulines, J. F. Cardoso, and E. Gassiat. Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In *Proceedings of ICASSP'97*, volume 5, pages 3617–20, 1997.

[12] J. Park and I. W. Sandberg. Universal approximation using radial basis functions networks. *Neural Computation*, 3:246–257, 1991.

[13] K. J. Pope and R. E. Bogner. Blind signal separation. i. linear, instantaneous combinations. *Digital Signal Processing*, 6:5–16, 1996.

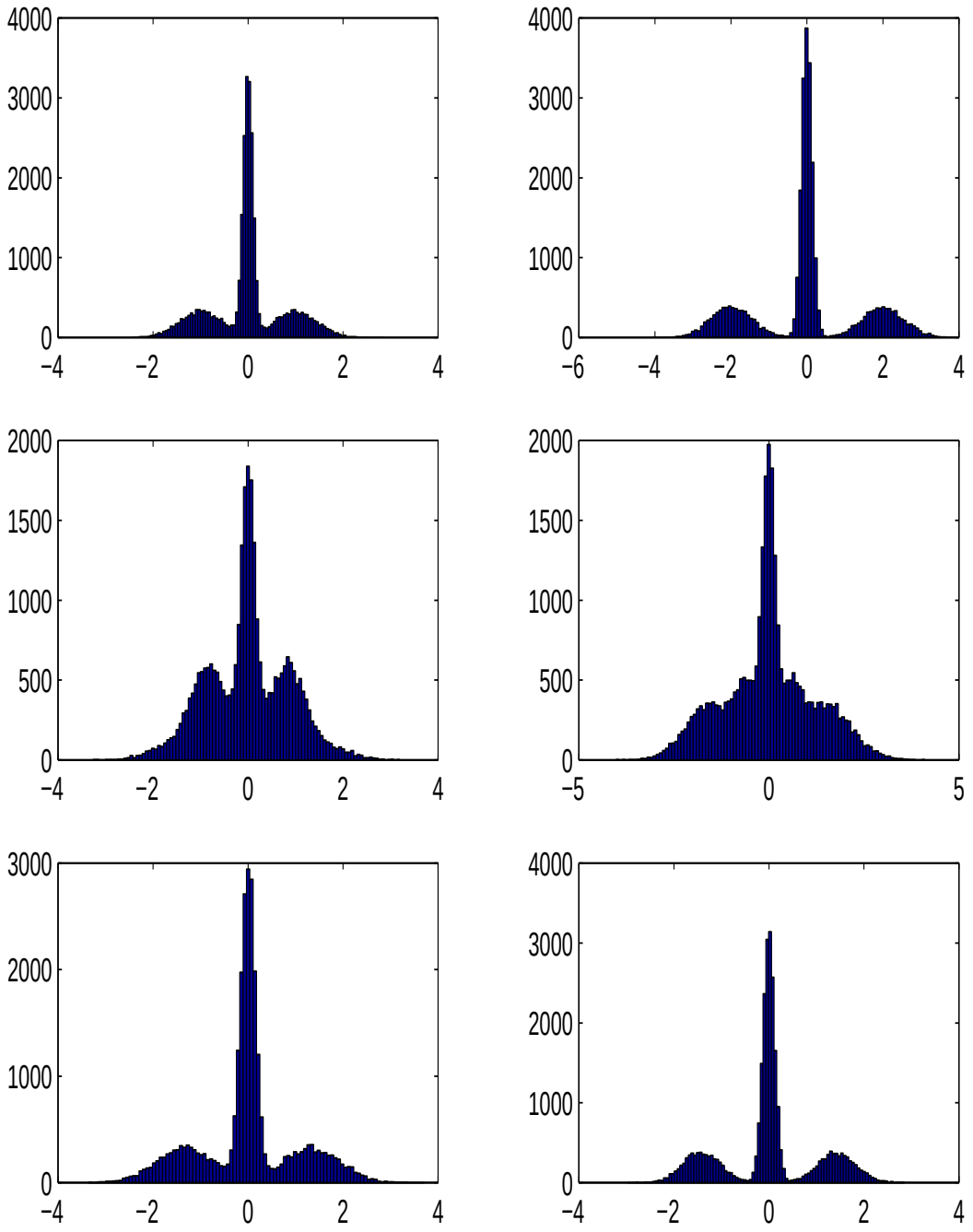

Fig. 6. Experiments with synthetic Gaussian mixture sources. Top row: Densities of two Gaussian mixture sources; middle row: densities of the observations after mixing and addition of noise; bottom row: density of the estimated sources.

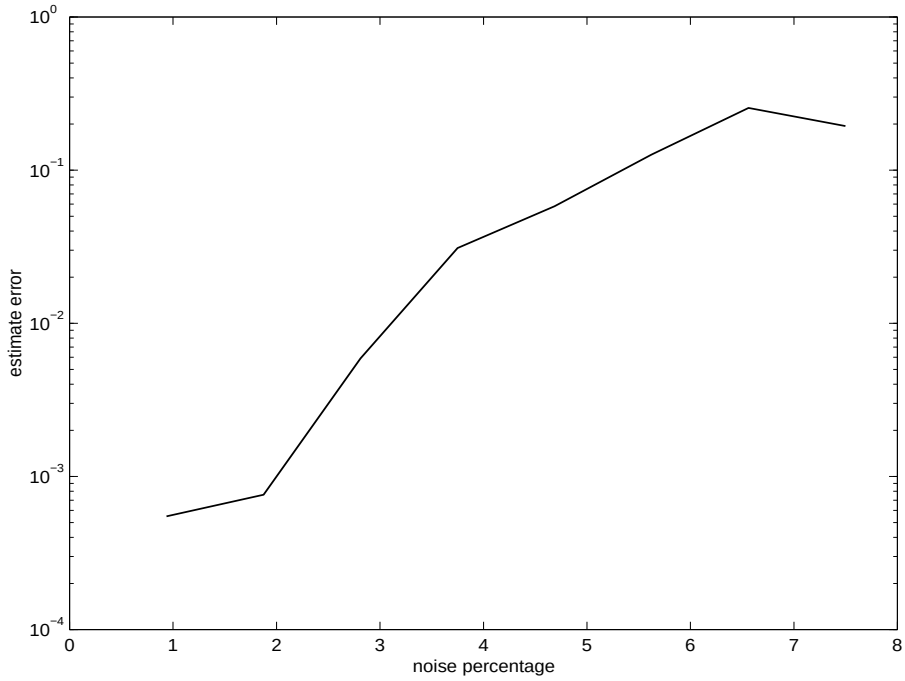

Fig. 7. Noise-dependence of the IFA algorithm performance.

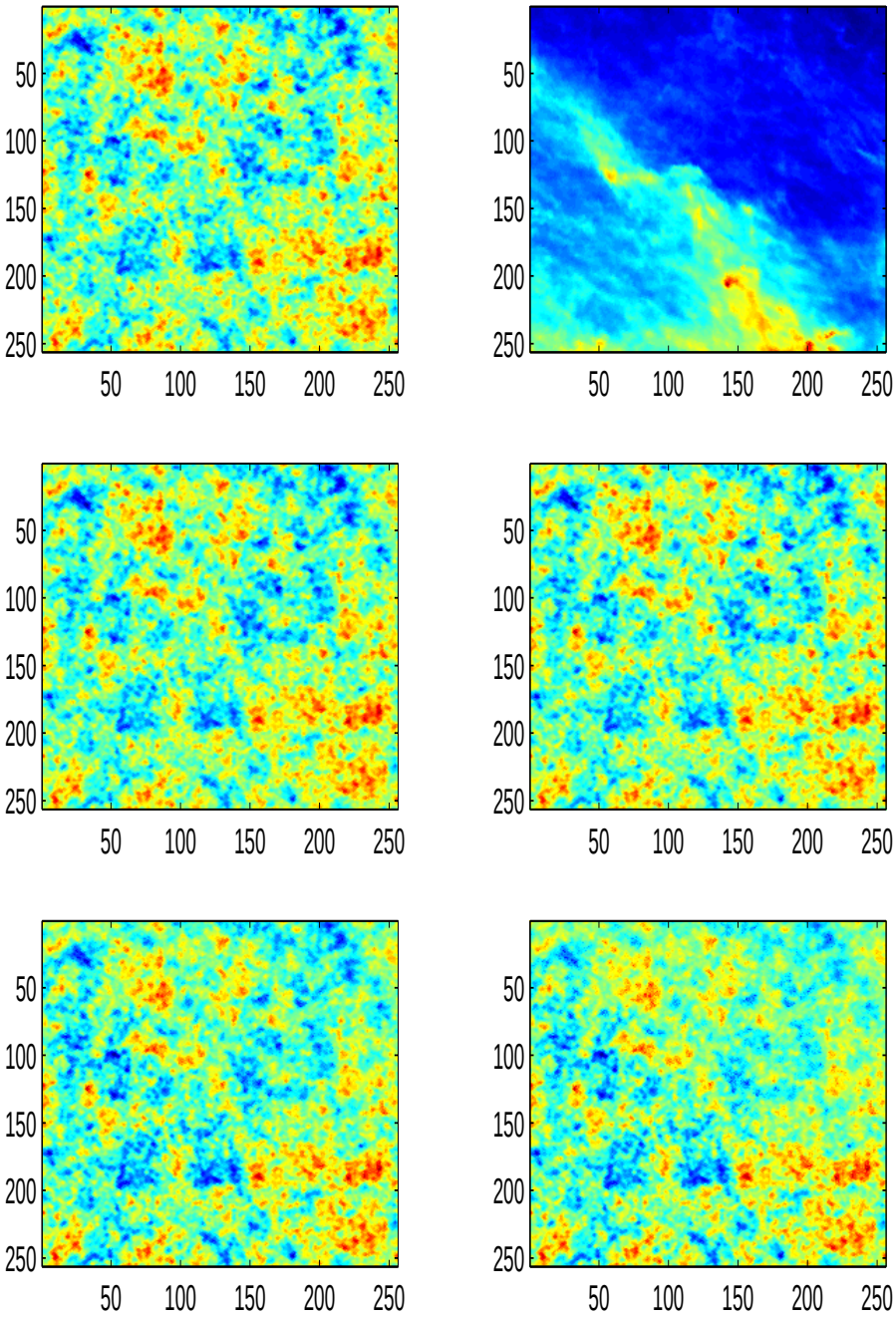

Fig. 8. Separation of CMB+Dust mixture with all the IFA parameters unknown. Noise as in Fig. 4, average SNR=30dB. Top: original sources. Middle: Mixtures. Bottom: source estimates.

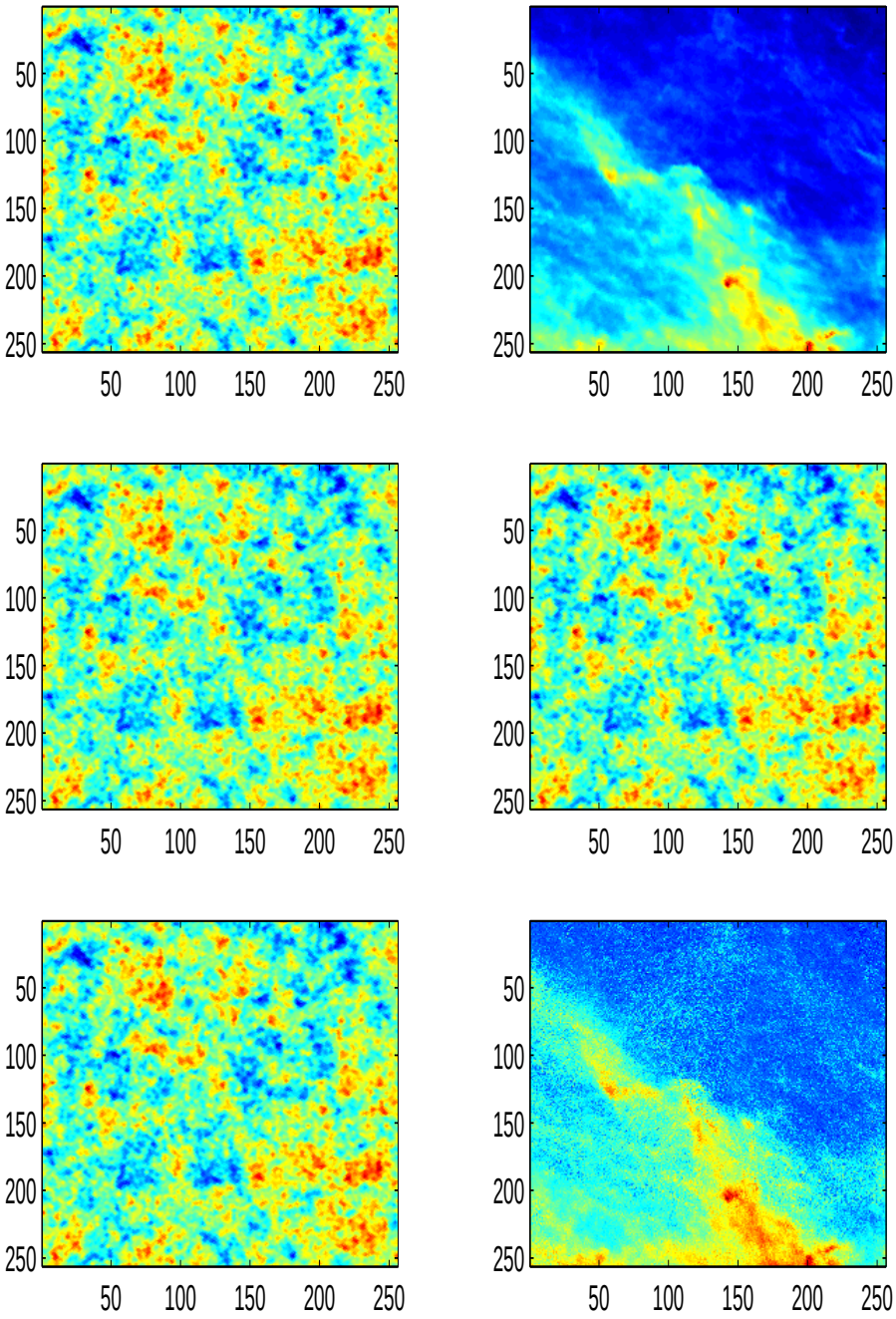

Fig. 9. Separation of CMB+Dust mixture with fixed mixture parameters. Noise as in Fig. 8. Top: original sources. Middle: Mixtures. Bottom: source estimates.