COMPONENT SEPARATION IN ASTRONOMICAL IMAGES USING INDEPENDENT FACTOR ANALYSIS

Ercan E. Kuruoğlu, 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,
Pisa, 56124, ITALY
e-mail: kuruoglu@iei.pi.cnr.it

ABSTRACT

Astronomical microwave images carry information about radiation from various sources, including cosmic microwave background radiation, galactic dust, synchrotron, etc. Moreover, the observations are corrupted with sensor noise, which is normally space varying. All of these components carry important information about the Universe and need to be recovered separately. In this paper, we study the problem of component separation in astronomical images using a recently introduced technique called independent factor analysis (IFA). We briefly analyse the source distributions and suggest a Gaussian mixture model. We then introduce IFA and discuss how the developed source and noise models are incorporated in the IFA algorithm. We present simulation results obtained by IFA on realistic data, which simulate the ones expected from the Planck Surveyor Satellite mission, to be launched by the European Space Agency.

1 INTRODUCTION

Astronomical microwave images carry important information about the Universe. For example, the recovery of cosmic microwave background (CMB) radiation would help us in making the picture of the early Universe. Unfortunately, astronomical images contain mixtures of radiations from various sources such as CMB, galactic dust, synchrotron, free-free emission, extra-galactic radio sources, and noise arising from the sensor. One needs to isolate these sources and clean up the noise in order to study and understand them.

One technique for source separation, namely independent component analysis (ICA), has been explored for the astronomical image separation problem in [2]. This work showed that ICA has limited success in the presence of noise, and falls short of the needs of our problem, where the detector noise is not negligible. Moreover, ICA is a completely blind model estimation technique while we have some crucial prior information to be exploited about the sources and the noise. In this paper, we adopt a method that incorporates prior information about the sources in a very generic way. This method is called independent factor analysis (IFA), and has been introduced recently in [5], [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 which is specially convenient for learning through an expectation-maximization algorithm. The noise is also taken into account in the mixture generation model in a very natural way and hence IFA provides an important alternative to ICA.

Despite this attractive appearance, the numerical studies on IFA are limited only to a couple of simple synthetic toy problems, and the potentials and the drawbacks of the technique are not well understood. In this paper, we study this technique in the context of simulated but realistic astronomical data, and try to identify its potentials, its weak and strong points.

2 SOURCE DISTRIBUTIONS

In this section we will look into the amplitude distributions of the most important components in astronomical images, to see whether we can suggest a common statistical model for all of them. These maps simulate the ones that are expected from the Planck mission, and are of the same type of those used in our experiments.

2.1 CMB Radiation

The theory tells us that CMB is Gaussian distributed. Figure 1 shows a typical CMB image generated synthetically.

2.2 Galactic Dust

A map of galactic dust radiation is shown in Figure 2.a. The related histogram is provided in Figure 2.b, solid curve. It is clear from the histogram that the galactic dust exhibits a non-Gaussian behaviour. The curve is multimodal and unsymmetric. We suggest modelling it with a mixture of Gaussian densities, given by

$$p_X(x) = \sum_i a_i \exp(-\frac{(x-\mu_i)^2}{2\sigma^2}). \qquad (1)$$

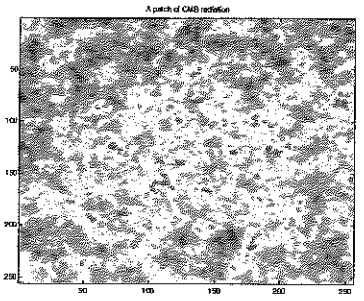

Figure 1: A map of CMB data

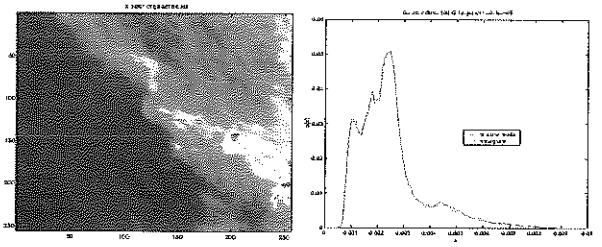

Figure 2: a) An image showing galactic dust intensity, b)Histogram and the Gaussian mixture model fit for the galactic dust intensity

We used an expectation-maximization (EM) algorithm [3] for fitting a Gaussian mixture to the histogram. The resulting density is shown in Figure 2.b, dashed curve. We repeated our experiments on about 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.

2.3 Synchrotron

Figure 3.a shows a synchrotron map extrapolated from 408 MHz observations. The related histogram is given in

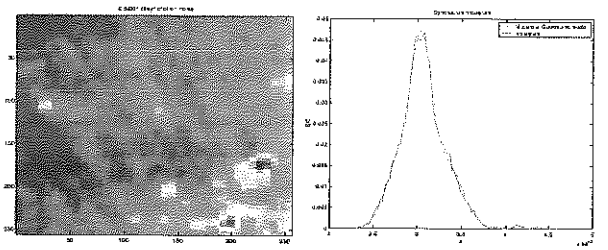

Figure 3: a) A synchrotron intensity map, b) Histogram and the Gaussian mixture model fit of the synchrotron intensity

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 Gaussian mixture of only four components is given in Figure 3.b, dashed curve. These observations were repeated for many other radiation sources. We suggest the Gaussian mixture density as an efficient generic model for the spatial images of astronomical radiation sources.

2.4 Noise distribution

Generally, it is assumed that the noise is space-invariant Gaussian. However, in satellite images the noise may not be space-invariant, as can be seen in Figure 4. The reason for this non-stationarity is that the satellite antenna does not scan the sky uniformly.

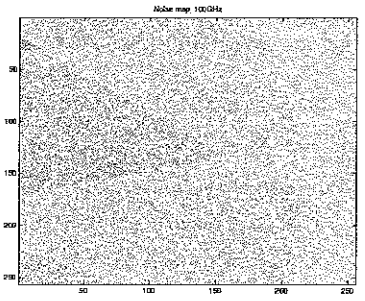

Figure 4: A typical noise map expected from Planck data

3 INDEPENDENT FACTOR ANALYSIS

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

$$y_i = \sum_{j=1}^{L} \mathbf{H}_{ij} x_j + n_i, \quad i = 1, \ldots, N \qquad (2)$$

where $\mathbf{H}$ is the mixing matrix, y_i are the observations, x_j are the sources and n_i are the noise samples which are assumed to be Gaussian distributed and space-varying. We would like to obtain H_{ij} and x_j from observations y_i. In the last decade, various efforts have been made for the solution of this blind source separation (BSS) problem, which arises in many practical applications. In particular, ICA aims at decomposing the observations into independent sources, and has been studied widely in the literature. However, ICA considers a highly idealised problem where the mixing is square ($L = N$), invertible, instantaneous and noiseless. In many real situations, and specifically in the problem we are considering here, the observations are noisy and the available sensors are not as many as the sources to be separated. As the noise level increases, the performance of ICA algorithms deteriorate and the separation quality decreases [1]. Efforts have been made to include noise into the

analysis: Hyvarinen suggested employing a special class of noise-insensitive contrast functions [4]. However, we still observed a deteriorating behaviour when the noise level increases.

Towards removing these drawbacks, Moulines *et al.* [5] suggested modelling the sources with mixtures of Gaussians, and employed EM-based techniques to estimate the mixing matrix and the source distribution parameters. Later, Attias named this formulation *independent factor analysis* (IFA) [1]. IFA is performed in two steps: in the first one, the IF model is trained to learn the mixing matrix, noise covariance and 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.

3.1 Source Model

Let us assume for each source a model in the form of a mixture of Gaussians:

$$p(x_i|\theta_i) = \sum_{q_i}^{n_i} w_{i,q_i} \mathcal{G}(x_i - \mu_{i,q_i}, \nu_{i,q_i}), \theta_i = \{w_{i,q_i}, \mu_{i,q_i}, \nu_{i,q_i}\} \tag{3}$$

Hence, assuming independence of the sources it is

$$p(\mathbf{x}|\theta) = \prod_{i=1}^{L} p(x_i|\theta_i) = \sum_{\mathbf{q}} w_{\mathbf{q}} \mathcal{G}(\mathbf{x} - \mu_{\mathbf{q}}, \mathbf{V_q}) \tag{4}$$

where $w_{\mathbf{q}}, \mu_{\mathbf{q}}, \mathbf{V_q}$ are matrices that contain amplitudes, means and standard deviations of the Gaussian kernels. $\mathbf{q}$ contains the indices that determine which component in the mixture the observation comes from.

3.2 Sensor Model

To generate sensor signals $\mathbf{y}$, first pick a unit q_i for each source i with probability $p(\mathbf{q}) = w_{\mathbf{q}}$. The probability of generating a particular source vector $\mathbf{x}$ given $\mathbf{q}$ is

$$p(\mathbf{x}|\mathbf{q}) = \mathcal{G}(\mathbf{x} - \mu_{\mathbf{q}}, \mathbf{V_q}) \tag{5}$$

and the probability of generating a particular sensor vector $\mathbf{y}$ given a source vector $\mathbf{x}$ is

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{G}(\mathbf{y} - \mathbf{H}\mathbf{x}, \mathbf{\Lambda}) \tag{6}$$

The joint density of the sensor vector, the source vector and the top hidden layer is given by

$$p(\mathbf{q}, \mathbf{x}, \mathbf{y}|W) = p(\mathbf{q})\ p(\mathbf{x}|\mathbf{q})\ p(\mathbf{y}|\mathbf{x}) \tag{7}$$

where W is the vector of the parameters to be estimated: $W = (\mathbf{H}, \theta, \mathbf{\Lambda})$. Then, the sensor vector probability is

$$p(\mathbf{y}|W) = \sum_{\mathbf{q}} \int d\mathbf{x}\ p(\mathbf{q}) p(\mathbf{y}|\mathbf{x}) = \sum_{\mathbf{q}} p(\mathbf{q})\ p(\mathbf{y}|\mathbf{q}) \tag{8}$$

$$p(\mathbf{y}|\mathbf{q}) = \mathcal{G}(\mathbf{y} - \mathbf{H}\mu_{\mathbf{q}}, \mathbf{H}\mathbf{V_q}\mathbf{H}^T + \mathbf{\Lambda}) \tag{9}$$

3.3 Learning the model

Next step is to define an error function that measures the distance between the sensor model density and the measured density. The following Kullback-Leibler distance function has been chosen

$$J(W) = \int d\mathbf{y} p^0(\mathbf{y}) \log \frac{p^0(\mathbf{y})}{p(\mathbf{y}|W)} = -E[\log p(\mathbf{y}|W)] - H_{p^0}. \tag{10}$$

where E is the averaging operator over the observed data.

3.3.1 Expectation-Maximization Algorithm

Then the learning algorithm is obtained [1] as a modified EM algorithm:

Maximization step:

$$\mathbf{H} = E[\mathbf{y} < \mathbf{x}|\mathbf{y} >](E < \mathbf{x}\mathbf{x}^\mathbf{T}|\mathbf{y} >)^{-1} \tag{11}$$

$$\mathbf{\Lambda} = E[\mathbf{y}\mathbf{y}^T] - E[\mathbf{y} < \mathbf{x}^T|\mathbf{y} > \mathbf{H}^T] \tag{12}$$

where $< \cdot >$ denotes the expectation operator.

Expectation step:

$$\mu_{i,q_i} = \frac{Ep(q_i|\mathbf{y})\ < x_i|q_i, \mathbf{y} >}{Ep(q_i|\mathbf{y})} \tag{13}$$

$$\nu_{i,q_i} = \frac{Ep(q_i|\mathbf{y})\ < x_i^2|q_i, \mathbf{y} >}{Ep(q_i|\mathbf{y})} - \mu_{i,q_i}^2 \tag{14}$$

$$w_{i,q_i} = Ep(q_i|\mathbf{y}) \tag{15}$$

where

$$< \mathbf{x}|\mathbf{y} > = \sum_{\mathbf{q}} p(\mathbf{q}|\mathbf{y}) < \mathbf{x}|\mathbf{q}, \mathbf{y} > \tag{16}$$

$$< \mathbf{x}|\mathbf{q}, \mathbf{y} > = \rho_{\mathbf{q}}(\mathbf{y}), \tag{17}$$

$$< \mathbf{x}\mathbf{x}^T|\mathbf{q}, \mathbf{y} >= (\mathbf{H}^T\mathbf{\Lambda}^{-1}\mathbf{H} + \mathbf{V_q}^{-1})^{-1} + \rho_{\mathbf{q}}(\mathbf{y})\rho_{\mathbf{q}}(\mathbf{y})^T, \tag{18}$$

$$\rho_{\mathbf{q}}(\mathbf{y}) = (\mathbf{H}^T\mathbf{\Lambda}^{-1}\mathbf{H} + \mathbf{V_q}^{-1})^{-1}(\mathbf{H}^T\mathbf{\Lambda}^{-1}\mathbf{y} + \mathbf{V_q}^{-1}\mu_{\mathbf{q}}) \tag{19}$$

3.4 Source estimation

For the estimation of the sources Moulines *et al.* do not suggest any scheme [5], while Attias [1] suggests two schemes, namely, least squares and MAP estimation. For our experiments, we used the least squares estimation scheme:

$$\mathbf{x}^{LS}(\mathbf{y}) =< \mathbf{x}|\mathbf{y} >= \int d\mathbf{x}\ \mathbf{x}\ p(\mathbf{x}|\mathbf{y}, W) \tag{20}$$

where $p(\mathbf{x}|\mathbf{y}, W) = \sum_{\mathbf{q}} p(\mathbf{q}|\mathbf{y})\ p(\mathbf{x}|\mathbf{q}, \mathbf{y})$. Evaluating the integral, we obtain:

$$\mathbf{x}^{LS}(\mathbf{y}) = \sum_{\mathbf{q}} p(\mathbf{q}|\mathbf{y})(\mathbf{A_q}\mathbf{y} + \mathbf{b_q}) \tag{21}$$

where

$$\mathbf{A_q} = (\mathbf{H}^T\mathbf{\Lambda}^{-1}\mathbf{H} + \mathbf{V_q})^{-1}\mathbf{H}^T\mathbf{\Lambda}^{-1} \tag{22}$$

$$\mathbf{b_q} = (\mathbf{H}^T\mathbf{\Lambda}^{-1}\mathbf{H} + \mathbf{V_q})^{-1}\mathbf{V_q}^{-1}\mu_{\mathbf{q}}. \tag{23}$$

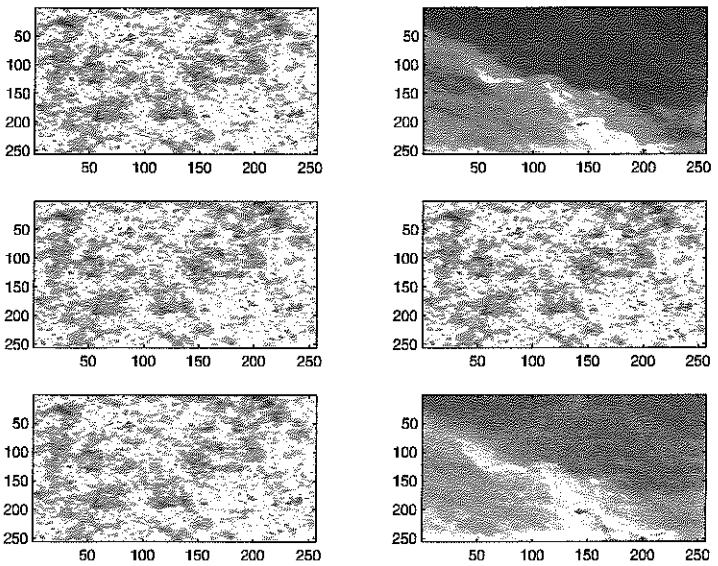

Figure 5: Source separation of CMB+galactic dust mixture. Top row: original CMB and galactic dust sources, respectively. Second row: mixtures. Third row: source estimates.

4 SIMULATION RESULTS AND CONCLUSIONS

We ran various simulations on synthetic and realistic data. Our results with synthetic data (two Gaussian mixture sources and three mixtures embedded in Gaussian noise) partly confirmed the results of [1]. When the number of unknown variables is small (the mixture model parameters are fixed and assumed known, while the mixing matrix is unknown), we observed a fast convergence to the optimal values. However, when the number of unknowns is increased (mixture model parameters are assumed to be unknown), we observed a significant degradation in the performance.

As for more realistic data, we formed mixtures of CMB and synchrotron and of CMB and galactic dust, from maps of the type shown in Section 2 and added %3 (of the CMB) space-varying noise. When the mixture model parameters are fixed, we observed that for good starting points the algorithms finds the optimal $\mathbf{H}$, and the sources are recovered successfully despite the presence of some noise, as shown in Figure 5. 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.

For the continuation of this work we will consider optimization with simulated annealing as opposed to 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.

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] 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.

[4] 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.

[5] 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.