Image Separation Using Particle Filters

Mauro Costagli, Ercan Engin Kuruoğlu
ISTI-CNR, v. Moruzzi 1, 56124 Pisa, Italy

September 27, 2004

Abstract

In this work, we will analyze the problem of source separation in the case of superpositions of different source images, which need to be extracted from a set of noisy observations. This problem occurs, for example, in the field of astrophysics, where the contributions of various Galactic and extra-Galactic components need to be separated from a set of observed noisy mixtures. Most of the previous work on the problem performed blind source separation, assuming noiseless models, and in the few cases when noise is taken into account assumed Gaussianity and space-invariance. We present a novel technique, namely particle filtering, for the solution of the source separation problem: it is an advanced Bayesian estimation method which can deal with non-Gaussian and non-linear models, and additive space-varying noise, in the sense that it is an extension of the Kalman filter. Our simulations on realistic astrophysical data show that the particle filter provides significantly better results in comparison with one of the most widespread algorithms for source separation (FastICA), especially in the case of low SNR.

Keywords: Particle Filtering, Sequential Markov Chain Monte Carlo, Blind Source Separation, Bayesian Source Separation, Independent Component Analysis, Image Separation, Non-Gaussian Models, Non-Stationary Noise, Cosmic Microwave Background, A-priori Information.

1 Introduction

Signal separation has been applied with a certain degree of success in applications ranging from speech processing to fMRI and from financial time series analysis to telecommunications. In these applications, generally a classical blind source separation technique, namely ICA [13], has been employed in a simple form assuming equal number of sources and observations, stationary mixing, and in the few cases where noise is taken into account, a Gaussian stationary model is adopted. In most such work, the time (or space) structure in the signal (or image) is completely ignored, turning the technique into an ensemble data analysis technique rather than a time-series analysis one. Moreover, ICA is blind and does not consider any prior information.

However, in some applications these assumptions are highly questionable. In most real life environments noise is present, and usually both the sources and the noise are non-Gaussian and space-varying. The mixing matrix might change in time or space as well as the sources which might be highly non-stationary. There may exist a wealth of prior information regarding the space structure of the data and its statistical distribution as well as the mixing matrix. We will present a relatively novel technique, namely particle filtering, which can take into account all these different features, and deal with non-Gaussian and non-linear models, and additive space-varying noise. The algorithm has been implemented and tested for the solution of the independent component separation problem in astrophysical radiation maps. This approach is extremely flexible as it is possible to exploit the available a-priori information about the statistical properties of the sources through the Bayesian theory.

The following section is about particle filters, from the description of the algorithm to its implementation. The third section is a brief introduction to the practical astrophysical scenario in which our algorithm has been tested. The fourth part describes the numerical experiments to compare the particle filter to the FastICA algorithm, while the conclusions are presented in the last section of this work.

2 Particle Filtering

Filtering is the problem of estimating the hidden variables (states) of a system, as a set of observations becomes available on-line.

In many real-world data analysis applications, prior knowledge about the unknown quantities to be estimated is available, and this information can be exploited to formulate Bayesian models: prior distributions for the unknown quantities and likelihood functions that relate these quantities to the observations. Then, all inference on the unknown quantities is based on the posterior distribution obtained from Bayes' theorem.

It is possible to express the model in terms of a state equation and an observation equation:

$$\boldsymbol{\alpha}_t = g_t(\boldsymbol{\alpha}_{t-1}, \boldsymbol{\nu}_t); \tag{1}$$

$$\mathbf{y}_t = h_t(\boldsymbol{\alpha}_t, \mathbf{w}_t). \tag{2}$$

The state equation (1) evaluates the state sequence: $\boldsymbol{\alpha}_t$ is the state at current step t, g_t is a possibly nonlinear function, $\boldsymbol{\alpha}_{t-1}$ is obviously $\boldsymbol{\alpha}_t$ at the previous step, and $\boldsymbol{\nu}_t$ is the so-called dynamic noise process. The observation equation (2) is characterized by a possibly nonlinear function h_t, and both the current state $\boldsymbol{\alpha}_t$ and the observation noise realisation $\mathbf{w}_t$ at time step t are taken into account to generate the observation $\mathbf{y}_t$.

The *Kalman filter* (KF) is an extension of the Wiener filter, and it was presented by R. E. Kalman in 1961 [15]: this filter derives an exact analytical expression to compute the evolving sequence of the posterior distributions, when the data are modelled by a linear Gaussian state-space model. The obtained posterior density at every time step is Gaussian, hence parametrized by a mean and a covariance.

The best known algorithm that allows a non-Gaussian and nonlinear model is the *Extended Kalman filter* (EKF) [2], based upon the principle of linearising the measurements and evolution models using Taylor series expansions. Unfortunately, this procedure may lead to poor representations of both the non-linear functions and the probability distributions of interest, so the filter can diverge. Also the more recent *Unscented Kalman Filter* (UKF), which provides Gaussian approximations of the probability density functions, has however a limitation, that is it does not apply to general non-Gaussian distributions [14].

Particle filtering is a relatively novel technique that provides a solution to the general nonlinear, non-Gaussian filtering problem using numerical Bayesian (sequential Monte Carlo) techniques. Although known since late 60's [10], this technique has received interest only in the latest years, finding successful applications especially in tracking problems (see [8] for examples). Very recently it has also been applied to solve source separation problems ([1, 3, 9]). In particular, Everson and Roberts [9] considered a linear instantaneous mixing in which the sources and the noise are stationary while the mixing matrix is non-stationary. They assumed generalised Gaussian models for the sources which are fixed but unknown. Andrieu and Godsill [3] considered the problem of convolutional mixing instead and adopted a parametric model (time-varying AR) for the sources which were assumed to be Gaussian. The mixing was also assumed to be evolving according to a time-varying AR Gaussian process.

In our problem, the non-stationarity is in the sources and in the noise rather than in the mixing and we consider the source model parameters as random to fully exploit the potentials of Bayesian modelling. Moreover, in order to allow general, multi-modal distributions, we assume the sources to be modelled with mixtures of Gaussians, rather than with Gaussian or generalised Gaussian densities. Therefore, in this work we basically follow a formulation similar to that in [1].

2.1 Theoretical Fundamentals

Our objective is to estimate the expectations

$$\begin{aligned} I(f_t) &= E_{p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})}\{f_t(\boldsymbol{\alpha}_{0:t})\} \\ &\triangleq \int f_t(\boldsymbol{\alpha}_{0:t})p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})\mathrm{d}\boldsymbol{\alpha}_{0:t} \end{aligned} \tag{3}$$

for some function of interest f_t, like the mean of the sources, or their covariance.

The basis of particle filtering is the representation of continuous pdfs with discrete points (*particles*): for example, the posterior distribution $p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$ can be approximated by

$$p_N(d\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) = \frac{1}{N}\sum_{i=1}^{N}\delta_{\boldsymbol{\alpha}_{0:t}^{(i)}}(d\boldsymbol{\alpha}_{0:t}) \ , \tag{4}$$

where $\delta_{\boldsymbol{\alpha}_{0:t}^{(i)}}$ denotes the delta-Dirac mass and N is the number of points where the continuous pdf is discretised. Thus, it is possible to obtain the estimate of $I(f_t)$ by means of N particles drawn from the posterior distribution:

$$\begin{aligned} I_N(f_t) &= \int f_t(\boldsymbol{\alpha}_{0:t})p_N(d\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) \\ &= \sum_{i=1}^{N} f_t\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right) \end{aligned} \tag{5}$$

Unfortunately, it is usually impossible to generate samples from the posterior distribution since it is, in general, multivariate and non-standard. A classical solution is to use the *importance sampling* method, in which the true posterior distribution is replaced by an *importance function* $\pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$ which is easier to sample from. Provided that the support of $\pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$ includes the support of $p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$, from (3) we get the identity

$$I(f_t) = \frac{\int f_t(\boldsymbol{\alpha}_{0:t})w(\boldsymbol{\alpha}_{0:t})\pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})d\boldsymbol{\alpha}_{0:t}}{\int w(\boldsymbol{\alpha}_{0:t})\pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})d\boldsymbol{\alpha}_{0:t}}, \tag{6}$$

where

$$w(\boldsymbol{\alpha}_{0:t}) = \frac{p(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})}{\pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})} \tag{7}$$

is known as the *importance weight*. Consequently, it is possible to obtain an estimate of $I(f_t)$ using N particles $\{\boldsymbol{\alpha}_{0:t}^{(i)};\ i = 1, \cdots, N\}$ sampled from $\pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t})$:

$$\begin{aligned} I_N(f_t) &= \frac{\frac{1}{N}\sum_{i=1}^{N} f_t\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right) w\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right)}{\frac{1}{N}\sum_{j=1}^{N} w\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right)} \\ &= \sum_{i=1}^{N} f_t\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right) \tilde{w}_t^{(i)}, \end{aligned} \tag{8}$$

where the *normalised* importance weights $\tilde{w}_t^{(i)}$ are given by

$$\tilde{w}_t^{(i)} = \frac{w\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right)}{\sum_{j=1}^{N} w\left(\boldsymbol{\alpha}_{0:t}^{(i)}\right)}. \tag{9}$$

When the importance function is restricted to be of the general form:

$$\begin{aligned} \pi(\boldsymbol{\alpha}_{0:t}|\mathbf{y}_{1:t}) &= \pi(\boldsymbol{\alpha}_{0:t-1}|\mathbf{y}_{1:t-1})\pi(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{0:t-1}, \mathbf{y}_{1:t}) \\ &= \pi(\boldsymbol{\alpha}_0) \prod_{k=1}^{t} \pi(\boldsymbol{\alpha}_k|\boldsymbol{\alpha}_{0:k-1}, \mathbf{y}_{1:k}) \end{aligned} \tag{10}$$

the importance weights and hence the posterior can be evaluated recursively. Unfortunately, for the importance distributions of the form specified before, a degeneracy phenomenon may occur: after a few iterations, all but one of the normalised importance weights are very close to zero. This happens because the variance of the importance weights can only increase (stochastically) over time, as demonstrated in [1]. It is thus indispensable to include one more step (called *selection*) in the particle filter algorithm. The purpose of this procedure is to discard the particles with low importance weights, and to multiply the particles having high importance weights: the idea is that of associating with each particle (say $\tilde{\boldsymbol{\alpha}}_{0:t}^{(i)} : i = 1, \cdots, N$) a number of offsprings $N_t^{(i)}$, such that $\sum_{i=1}^{N} N_t^{(i)} = N$, in order to obtain N new particles $\{\boldsymbol{\alpha}_{0:t}^{(i)} : i = 1, \cdots, N\}$. After the selection step, all the importance weights are divided by N (normalization); since they do not depend on any past values of the normalised importance weights, all information regarding the old importance weights is discarded.

2.2 Model Specification

In the practical case of a linear (though possibly non-stationary) mixing system, the model expressed in Eq. (2) can be written as:

$$\mathbf{y}_{1:n,t} = \mathbf{H}_t \boldsymbol{\alpha}_{1:m,t} + \mathbf{w}_{1:n,t} \tag{11}$$

where $\mathbf{y}_{1:n,t}$, $\boldsymbol{\alpha}_{1:m,t}$ and $\mathbf{w}_{1:n,t}$ are column vectors, representing the n observations, the m sources and the n additive noise samples at step t respectively. The $n \times m$ real valued mixing matrix $\mathbf{H}_t$ is allowed to vary in t, and can be re-arranged it into a vector $\mathbf{h}_t = vec\{\mathbf{H}_t\}$ so that $[\mathbf{h}_t]_{n(j-1)+1} = h_{i,j,t}$. It is then possible to express the mixing model in terms of a new state equation and observation equation, in this way:

$$\mathbf{h}_t = \mathbf{A}_t \mathbf{h}_{t-1} + \mathbf{v}_t \tag{12}$$

$$\mathbf{y}_{1:n,t} = \mathbf{C}_t \mathbf{h}_t + \mathbf{w}_{1:n,t} \tag{13}$$

where $\mathbf{A}_t$ and $\mathbf{C}_t$ are $(nm \times nm)$ and $(n \times nm)$ real valued matrices respectively. The distributions of the dynamic noise $\mathbf{v}_t$ and the observation noise $\mathbf{w}_t$ are assumed to be i.i.d. and mutually independent: $\mathbf{v}_t \sim \mathcal{N}(0, \boldsymbol{\sigma}_v)$ and $\mathbf{w}_t \sim \mathcal{N}(0, \boldsymbol{\sigma}_w)$, with obvious notation. $\mathbf{C}_t$ can be expressed in terms of the source signal vector $\boldsymbol{\alpha}_{1:m,t}$, as $\mathbf{C}_t = \boldsymbol{\alpha}^T_{1:m,t} \bigotimes \mathbf{I}_n$. In absence of further prior information, we assume $\mathbf{A}_t = \mathbf{I}_{nm}$. Our goal is to estimate $\mathbf{C}_t$ (i.e. the source signal vector $\boldsymbol{\alpha}_{1:m,t}$) and $\mathbf{h}_t$. The introduction of the state equation allows to deal with non-stationary mixing matrices, as the coefficients of $\mathbf{h}$ can be updated at every step.

We model the probability density function $p(\alpha_{i,t})$ of source i at step t by a finite mixture of Gaussians:

$$p(\alpha_{i,t}) = \sum_{j=1}^{q_i} \rho_{i,j} \mathcal{N}(\alpha_{i,t}; \mu_{i,j,t}, \sigma^2_{i,j,t}); \tag{14}$$

$$\sum_{j=1}^{q_i} \rho_{i,j} = 1 \ , \tag{15}$$

where $\rho_{i,j}$ is the weight of the j^{th} Gaussian component of the i^{th} source, q_i is the number of Gaussian components for the i^{th} source, while $\mu_{i,j,t}$ and $\sigma_{i,j,t}$ are the parameters which describe each Gaussian component. We define an index variable z_i which takes on a finite set of values $Z_i = \{1, \cdots, q_i\}$ and determines the active Gaussian component in the mixture at time t, so that $p(\alpha_{i,t}|z_{i,t} = j) = \mathcal{N}(\alpha_{i,t}; \mu_{i,j}, \sigma^2_{i,j})$ and $p(z_{i,t} = j) = \rho_{i,j}$.

Since the m sources are statistically independent of one another, we can factorize the probability density function of all the sources at step t as:

$$p(\boldsymbol{\alpha}_{1:m,t}) = \prod_{i=1}^{m} p(\alpha_{i,t}). \tag{16}$$

At time t let $\mathbf{z}_{1:m,t} \triangleq [z_{1,t} \cdots z_{m,t}]^T$. It is possible to describe the discrete probability distribution of $\mathbf{z}_{1:m,t}$ using the i.i.d. model: in this case, the indicators of the states $z_{i,t}$ have identical and independent distributions. If we want to introduce temporal correlation between the samples of a particular source, we have to consider the first-order Markov model case, where the vector of the states evolves as a homogeneous Markov chain for $t > 1$:

$$\begin{aligned} & p(\mathbf{z}_{1:m,t} = \mathbf{z}_l | \mathbf{z}_{1:m,t-1} = \mathbf{z}_j) = \\ & = \prod_{i=1}^{m} p(z_{i,t} = [\mathbf{z}_l]_i | z_{i,t-1} = [\mathbf{z}_j]_i) = \prod_{i=1}^{m} \tau_{j,l}^{(i)} \end{aligned} \tag{17}$$

where $\tau_{j,l}^{(i)}$ is an element of the $q_i \times q_i$ real valued *transition matrix* for the states of the i^{th} source, denoted by $\boldsymbol{\tau}^{(i)}$. The state transition can be thus parametrised by a set of m transition matrices $\boldsymbol{\tau}^{(i)}$, $i \in \{1, \cdots, m\}$.

Given the observations $\mathbf{y}_t$ (assuming that the number of sources m, the number of Gaussian components q_i for the i^{th} source, and the number of sensors n are known), we would like to estimate the mixing matrix $\mathbf{H}$ and all the following unknown parameters of interest, grouped together:

$$\boldsymbol{\theta}_{0:t} = [\boldsymbol{\alpha}_{1:m,0:t}, \mathbf{z}_{1:m,0:t}, \{\mu_{i,j,0:t}\}, \{\sigma^2_{i,j,0:t}\}, \{\boldsymbol{\tau}^{(i)}_{0:t}\}] \tag{18}$$

where we recall that $\boldsymbol{\alpha}_{1:m,0:t}$ are the sources, $\mathbf{z}_{1:m,0:t}$ is the matrix of the indicator variables which determines which Gaussian component is active at a particular time for each source, $\{\mu_{i,j,0:t}\}$ and $\{\sigma^2_{i,j,0:t}\}$ are the means and the variances of the j^{th} Gaussian component of the i^{th} source and $\{\boldsymbol{\tau}^{(i)}_{0:t}\}$ is the transition matrix for the evolution of $z_{i,0:t}$.

The particles we should deal with will be thus $\{(\mathbf{h}^{(i)}_{0:t}, \boldsymbol{\theta}^{(i)}_{0:t}) : i = 1, \cdots, N\}$, generated according to $p(\mathbf{h}_{0:t}, \boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$. An empirical estimate of this distribution is given by

$$\bar{p_N}(\mathbf{h}_{0:t}, \boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\boldsymbol{h_{0:t}}^{(i)}}, \delta_{\boldsymbol{\theta_{0:t}}^{(i)}}(d\mathbf{h}_{0:t}, d\boldsymbol{\theta}_{0:t}), \tag{19}$$

and, as a corollary, an estimate of $p(\mathbf{h}_t, \theta_t|\mathbf{y}_{1:t})$ is

$$\bar{p_N}(\mathbf{h}_t, \boldsymbol{\theta}_t|\mathbf{y}_{1:t}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\boldsymbol{h_t}^{(i)}}, \delta_{\boldsymbol{\theta_t}^{(i)}}(d\mathbf{h}_t, d\boldsymbol{\theta}_t). \tag{20}$$

It is possible to reduce the problem of estimating $p(\mathbf{h}_t, \boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$ to a simpler one of sampling from $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$. In fact,

$$p(\mathbf{h}_t, \boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}) = p(\mathbf{h}_t|\boldsymbol{\theta}_{0:t}, \mathbf{y}_{1:t})p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t}). \tag{21}$$

Given an approximation of $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$, $p(\mathbf{h}_t|\boldsymbol{\theta}_{0:t}, \mathbf{y}_{1:t})$ is a Gaussian distribution whose parameters can be recursively estimated in closed form using the Kalman filter in Eq. (12)-(13). This technique, called *Rao-Blackwellisation* [6], leads to better results, as we are reducing the size of the parameter set to be estimated by marginalising out the mixing coefficients $\mathbf{h}_t$ using the Kalman filter, so that the only distribution we have to estimate by particle filtering is $p(\boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$, and not $p(\mathbf{h}_t, \boldsymbol{\theta}_{0:t}|\mathbf{y}_{1:t})$.

As it was introduced before, it is not easy to sample directly from the optimal importance distribution: this is the reason why a sub-optimal method will be employed throughout taking the importance distribution at step t to be the prior distribution $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$ of the sources to be estimated. This choice offers an appealing solution to overcome the analytic untractability of the optimal importance function: in fact the mixture of Gaussians model allows for an easy factorization of the prior distribution into several easy-to-sample distributions related to the parameters which describe the model itself. Thus, we write the prior importance function in the following way:

$$\begin{aligned} p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) &= p(\boldsymbol{\alpha}_{1:m,t}, \mathbf{z}_{1:m,t}, \mu_{i,j,t}, \sigma^2_{i,j,t}, \boldsymbol{\tau}_t| \\ &\quad \boldsymbol{\alpha}_{1:m,t-1}, \mathbf{z}_{1:m,t-1}, \mu_{i,j,t-1}, \sigma^2_{i,j,t-1}, \boldsymbol{\tau}_{t-1}) \\ &= p(\boldsymbol{\alpha}_{1:m,t}|\mathbf{z}_{1:m,t}, \{\mu_{i,j,t}\}, \{\sigma^2_{i,j,t}\}) \times \\ &\quad p(\{\mu_{i,j,t}\}|\{\mu_{i,j,t-1}\}, \mathbf{z}_{i,t}) \times \\ &\quad p(\{\sigma^2_{i,j,t}\}|\{\sigma^2_{i,j,t-1}\}, \mathbf{z}_{i,t}) \times \\ &\quad p(\mathbf{z}_{1:m,t}|\mathbf{z}_{1:m,t-1}, \boldsymbol{\tau}_t) \times \\ &\quad p(\boldsymbol{\tau}_t|\boldsymbol{\tau}_{t-1}). \end{aligned} \tag{22}$$

This hierarchical structure allows us to obtain an approximation of the distribution of the sources exploiting the particles generated from the distributions of the other parameters, sampling subsequently from $p(\boldsymbol{\tau}_t|\boldsymbol{\tau}_{t-1})$,

$p(\mathbf{z}_{1:m,t}|\mathbf{z}_{1:m,t-1}, \boldsymbol{\tau}_t)$, $p(\{\sigma^2_{i,j,t}\}|\{\sigma^2_{i,j,t-1}\}, \mathbf{z}_{i,t})$, $p(\{\mu_{i,j,t}\}|\{\mu_{i,j,t-1}\}, \mathbf{z}_{i,t})$, and finally obtain the particles of $p(\boldsymbol{\alpha}_{1:m,t}|\mathbf{z}_{1:m,t}, \{\mu_{i,j,t}\}, \{\sigma^2_{i,j,t}\})$. More detailed information can be found in [7].

2.3 The Algorithm

According to what has been said above, for every t, the classical algorithm can be summarised as:

- Sequential Importance Sampling Step:
  - For $i = 1, \cdots, N$, sample $\tilde{\boldsymbol{\alpha}}_t^{(i)}$ from the importance function $\pi(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}^{(i)}_{0:t-1}, \mathbf{y}_{1:t})$ and set $\tilde{\boldsymbol{\alpha}}^{(i)}_{0:t} = (\boldsymbol{\alpha}^{(i)}_{0:t-1}, \tilde{\boldsymbol{\alpha}}^{(i)}_t)$;
  - For $i = 1, \cdots, N$, evaluate (and then normalize) the importance weights.
- Selection Step:
  - Discard / multiply particles with low / high normalised importance weights to obtain N particles $\{\boldsymbol{\alpha}^{(i)}_{0:t} : i = 1, \cdots, N\}$.

This classical one-dimensional particle filtering formulation in t now needs to be adapted to the two-dimensional signal case in (k, l). Our technique uses a combination of two classical 1-D particle filtering schemes as a pre-block, and then manipulates the information stored in the parameters of each pixel to reflect the 1-D scheme to a 2-D one.

- In the 1-D pre-block, each observed image is converted into two vectors: the first one keeps the vertical correlation among adjacent pixels (by reading the images column by column, one column from the top to the bottom, and the following one from the bottom to the top, and so on), while the second vector keeps the horizontal correlation. Then the classical 1-D algorithm is run twice, once using the vector which keeps the vertical correlation, and then using the other one, so that two sets of estimates of the parameters related to each pixel (i.e. the active Gaussian component of each source with the respective mean and variance) and the MMSE estimate of each unknown source vectors become available. At this point a new variable $\boldsymbol{\theta}$, structured as shown in Eq. (18), and significant in describing the vertical and horizontal evolution of the image, is obtained by averaging the two previously calculated parameter sets (Fig. 1).

- In the 2-D processing block the parameter set is resized to re-convert the images stored in the previously created vectors into matrices. At this point, for every position of coordinates (k, l), all the source parameters $\boldsymbol{\alpha}_{1:m}$, $\mathbf{z}_{1:m}$, $\{\mu_{i,j}\}$, $\{\sigma^2_{i,j}\}$, $\{\boldsymbol{\tau}^{(i)}\}$ are available. The accuracy of this set of estimates will be now improved by means of the following technique, which emphasizes the two-dimensional correlation between pixels in the source images. A new value $\mathbf{z}^*_{1:m,k,l}$ for the parameter $\mathbf{z}_{1:m,k,l}$ is created for every position (k, l) by combining $\mathbf{z}_{1:m,k,l}$ with the respective parameters of a set of N adjacent pixels:

$$\mathbf{z}^*_{1:m,k,l} = p * \mathbf{z}_{1:m,k,l} + \frac{1-p}{N} * \sum_{q=1}^{N} \mathbf{z}_{q^{th}\text{neighbour}} \tag{23}$$

  where $\mathbf{z}_{q^{th}\text{neighbour}}$ is the value of $\mathbf{z}_{1:m}$ of each q^{th} neighbour, and where $p \in (0, 1)$ is an arbitrary weight, in the sense that a small value of p emphasizes the correlation among pixels. Subsequently, for every (k, l), we similarly combine the parameters which describe the mean value and the variance of the active Gaussian component (indicated by $\mathbf{z}^*_{1:m,k,l}$) with the respective parameters of the adjacent pixels, to obtain a new set of particles $\boldsymbol{\alpha}^*_{1:m,k,l}$ according to the procedure described at the end of section 2.2. At last, this new set of particles $\boldsymbol{\alpha}^*_{1:m}$ is used to generate the final estimates of the m source images.

3 The Astrophysical Context

Our algorithm has been tested on astrophysical images: this new application has been motivated by the need of analysing vast amounts of astrophysical radiation maps that will be available after the launch of the Planck satellite in 2007 by the European Space Agency (ESA) [12]. This mission will provide measurements in nine different frequency channels ranging from 30 GHz to 857 GHz, with a spectral and spatial resolution much higher than the previous NASA missions COBE and WMAP. These data are the superposition of various independent astrophysical sources, among which the most important one is the Cosmic Microwave Background (CMB): in fact it is the relic radiation remaining from the first light radiation in the universe released at the Big Bang. Therefore, the CMB provides a picture of the universe shortly after it has started. Moreover, it houses vital information to determine the values of certain cosmological parameters, the high-sensitivity calculation of which would help us decide between competing theories for

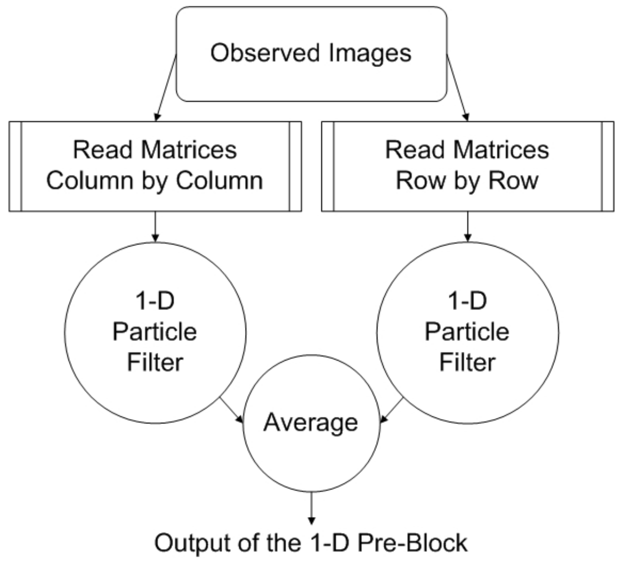

Figure 1: Scheme of the 1-D pre-block.

the evolution of the Universe. The signal measured in CMB experiments is however contaminated not only by the intrinsic noises due to the satellite microwave detectors but also by astrophysical contaminants (the so-called *foregrounds*). The most relevant foregrounds are the dust emission, synchrotron (caused by the interaction of the electrons with the magnetic field of the galaxy) and the free-free radiation (due to the the interaction of hot electrons with the interstellar gas), while other contaminations come from extragalactic microwave sources and from the so-called Sunyaev-Ze'ldovich effect [12]. Before achieving cosmological information from the statistical analysis of the CMB anisotropies, all these components must be separated from the intrinsic CMB signal. The problem, therefore, is conveniently formulated as the source separation from linear instantaneous mixtures, as in Eq. (11). The problem has been dealt with using other methods by several researchers in the literature including Baccigalupi et al. [4] and Maino et al. [18] who implemented the FastICA algorithm and its noisy version which had limited success in the presence of significant noise. A source model was introduced by Kuruoglu et al. in [16] implementing the Independent Factor Analysis (IFA) technique which also included the noise in the mixing model. Despite this added flexibility, IFA uses a fixed source model which lacks freedom in modelling source model parameters and moreover could not deal

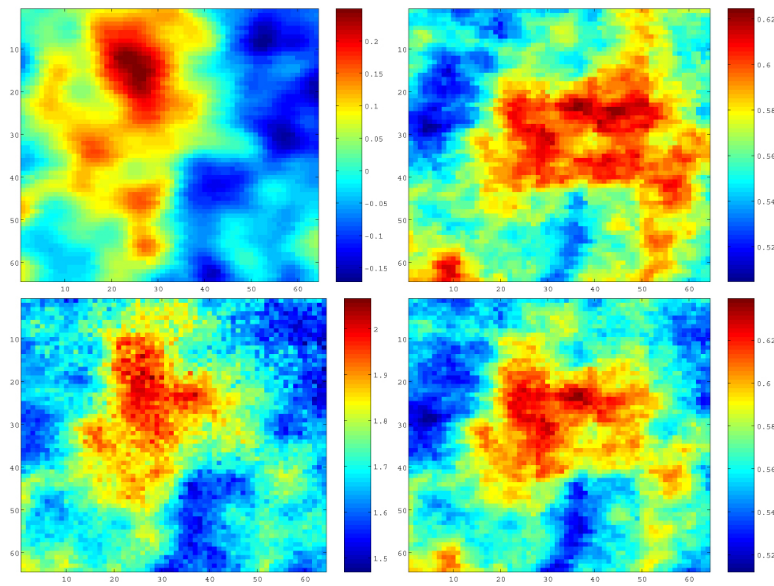

Figure 2: From top to bottom: original CMB and Synchrotron signals and realistic mixtures at 100 GHz and 30 GHz.

with non-stationary noise which is the case in our problem.

Snoussi et al. [19] utilise an EM algorithm in the spectral domain, making use of some generic priors for the sources. Cardoso et al. [5] perform blind source separation via spectral matching. Both of these works assume stationary noise and signals, which is not the case for the astrophysical image separation problem, and they both suffer from common drawbacks of the EM algorithm, i.e. local optimality and computational complexity. All of these approaches are blind or semi-blind techniques which do not exploit a wealth of information about the sources the astrophysics theory provides us with. To be able to incorporate these prior information, a full Bayesian formulation was derived in [17], which utilises MCMC techniques but does not address non-stationarity and does not consider the auto-correlation structure in images.

Our algorithm based on particle filtering avoids all of these problems. In fact it deals with the non stationarity of the noise, allows very flexible modelling of non-Gaussian sources and conveniently enables the utilisation of available prior information including dependence structure in the images.

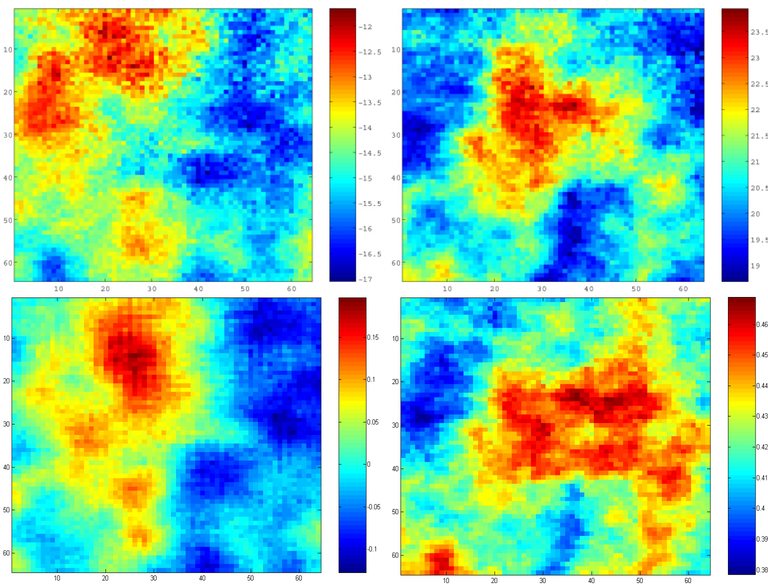

Figure 3: From top to bottom: FastICA estimates and particle filter estimates.

4 Numerical Experiments

We tested our algorithm on two 64×64 mixtures of CMB and synchrotron radiation, at 100 GHz and 30 GHz. Synthetic but realistic maps of the sources have been provided by the Planck Technical Working Group 2.1 on Diffuse Component Separation [12]: in particular, the CMB map is generated synthetically to follow a Gaussian distribution as it is implied by the cold dark matter model widely accepted for CMB. The synchrotron template was obtained by extrapolating the 408 MHz radio map of Haslam et al. [11] to Planck frequency channels and resolution. The expected antenna noise RMS maps are used to generate the additive space-varying noise samples. The average SNR is 10 dB. We use a mixture of three Gaussian components to approximate the synchrotron posterior distribution, whose starting values can be initialized randomly, and 300 particles are generated at each step, for each parameter of interest. It is obvious that better approximations could be obtained by increasing the number of particles and Gaussian components for each source, albeit an increase in computational cost. A Dirichlet prior is adopted for the index distribution of the Gaussian components in the mixture, while means and variances are drawn from Gaussian distributions

SIR	FastICA	Particle Filtering
CMB	1.14 dB	16.81 dB
Synchrotron	1.62 dB	20.41 dB

Table 1: Signal to Interference Ratio (SIR) values for FastICA and Particle Filtering

centered at the value of the previous particles and with variance determined by drift parameters. Noise is assumed to be Gaussian and space-varying, in accordance to the fact that the value given to each pixel of the maps is obtained from multiple measurements by the Planck antenna in the same direction, and the number of multiple measurements is not the same for pixels with different location. RMS noise maps are available, because the satellite scanning scheme of the sky is known, and the noise variance of each pixel is given as an input, for all the observed images. Since we are not aware of any other work which considers separation of non-stationary sources under non-stationary noise environment, we will compare our results with those obtained by the FastICA algorithm [13], which is one of the most widespread methods in source separation. As seen in Fig. 3, FastICA provides estimates with high level of interference, while the particle filter succeeds in recovering the original maps. Signal-to-interference results in Table 1 quantify this performance.

5 Conclusions

We have presented a relatively novel technique, namely particle filtering, for the separation of the independent components in image processing. In contrast with the other work in literaure, this method provides a very flexible framework which can successfully account for the non-stationarity in the noise and in the sources, as well as the prior knowledge about the sources and the mixing matrix. This technique, in addition to providing point estimates, gives us the posteriors for the sources and the mixing matrix out of which inference can be made on various statistical measures. We demonstrated on realistic astrophysical data that our algorithm provides significantly better results in comparison with one of the most widespread algorithms for source separation (FastICA), especially in the case of low SNR.

References

[1] Ahmed A., Andrieu C., Doucet A., Rayner P. J. W.: On-line Non-stationary ICA Using Mixture Models. Proc. IEEE ICASSP, (2000) Vol. **5**, 3148–3151.

[2] Anderson B. D., Moore J. M.: Optimal Filtering. Prentice-Hall, New Jersey (1979).

[3] Andrieu C., Godsill S.J.: A Particle Filter for Model Based Audio Source Separation. Int. Work. on ICA and Blind Signal Separation, ICA 2000, Helsinki, Finland.

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

[5] Cardoso J.-F., Snoussi H., Delabrouille J., Patanchon G.: Blind Separation of Noisy Gaussian Stationary Sources. Application to Cosmic Microwave Background Imaging. Proc. EUSIPCO, **1** (2002), 561–564.

[6] Casella G., Robert C. P.: Monte Carlo Statistical Methods. Springer, (1999).

[7] Costagli M., Kuruoğlu E. E., Ahmed A.: Source Separation of Astrophysical Images Using Particle Filters. ISTI-CNR Pisa, Italy – Technical Report 2003-TR-54.

[8] Doucet A., De Freitas J. F. G., Gordon N. J.: Sequential Monte Carlo Methods in Practice. Springer-Verlag (2001).

[9] Everson R. M., Roberts S. J.: Particle Filters for Non-stationary ICA. Advances in Independent Components Analysis, M. Girolami (Ed.) 23–41, Springer (2000).

[10] Handschin J. E., Mayne D. Q.: Monte Carlo Techniques to Estimate the Conditional Expectation in Multi-Stage Non-linear Filtering. Int. Journal Control **9** (1969), 547–559.

[11] Haslam C. G. T., Salter C. J., Stoffel H., Wilson W. E.: A 408 MHz All-Sky Continuum Survey. II - The Atlas of Contour Maps. Astronomy & Astrophysics, **47** (1982), 1.

[12] http://astro.estec.esa.nl/planck/: The Home Page of Planck.

[13] Hyvärinen A., Oja E.: A Fast Fixed-point Algorithm for Independent Component Analysis. Neural Computation, **9 (7)** (1997), 1483–1492.

[14] Julier S. J., Uhlmann J. K.: A New Extension of the Kalman Filter to Nonlinear Systems. Proceedings of AeroSence: the 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls, Vol. Multi Sensor Fusion, Tracking and Resource Management II (1997).

[15] Kalman R. E.: A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME, Journal of Basic Engineering, **83** (1961), 95–107.

[16] Kuruoğlu E. E., Bedini L., Paratore M. T., Salerno E., Tonazzini A.: Source Separation in Astrophysical Maps Using Independent Factor Analysis. Neural Networks, **16** (2003), 479–491.

[17] Kuruoglu, E. E., Comparetti, P. M.: Bayesian Source Separation of Astrophysical Images Using Markov Chain Monte Carlo. Proc. PHYSTAT (Statistical Problems in Particle Physics, Astrophysics and Cosmology), September 2003.

[18] Maino D., Farusi A., Baccigalupi C., Perrotta F., Banday A. J., Bedini L., Burigana C., De Zotti G., Grski K. M., Salerno E.: All-Sky Astrophysical Component Separation with Fast Independent Component Analysis (FastICA). Monthly Notices of the Royal Astronomical Society, **334** (2002), 53–68.

[19] Snoussi H., Patanchon G., Macias-Perez J., Mohammad-Djafari A., Delabrouille J.: Bayesian blind component separation for cosmic microwave background observation, AIP Proceedings of MaxEnt, (2001), 125–140.