Fully Bayesian Blind Source Separation of Astrophysical Images Modelled by Mixture of Gaussians

Simon P. Wilson

Trinity College Dublin, Ireland

Ercan E. Kuruoğlu, *Senior Member, IEEE* and Emanuele Salerno

ISTI-CNR, Pisa, Italy

Abstract

In this work, we address the problem of source separation in the presence of prior information. We develop a fully Bayesian source separation technique which assumes a generic model for the sources, namely Gaussian mixtures with a priori unknown number of components and utilise Markov chain Monte Carlo techniques for model parameter estimation. The development of this methodology is motivated by the need to bring an efficient solution to the separation of components in the microwave radiation maps to be obtained by the satellite mission PLANCK which has the objective of uncovering cosmic microwave background radiation. The proposed algorithm successfully incorporates a rich variety of prior information available to us in this problem in contrast to most of the previous work which assume completely blind separation of the sources. We report results on realistic simulations of expected Planck maps and on WMAP 3rd year results. The technique suggested is easily applicable to other source separation applications by modifying some of the priors.

Index Terms

Cosmic microwave background (CMB) radiation, PLANCK satellite mission, Bayesian source separation, Markov chain Monte Carlo (MCMC), Gibbs sampling

Contact author: Simon Wilson, Department of Statistics, Lloyd Institute, Trinity College Dublin, Dublin 2, Ireland. Phone: +353 1 896 1759. Fax: +353 1 677 0711. E-mail: simon.wilson@tcd.ie. Ercan E. Kuruoğlu and Emanuele Salerno are with Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, via G. Moruzzi 1, Pisa, Italy.

I. Introduction

One of the most important discoveries of the past century was undoubtedly the observation by Penzias and Wilson of the cosmic microwave background (CMB) radiation in 1964. Their accidental discovery gave the much searched for proof of the hot Big Bang theory which was proposed by Gamov two decades earlier. The hot big bang model, which aims to provide an explanation to the formation of our universe, has the main thesis that the universe evolved into its current state expanding and cooling from an initially much more dense and hotter universe. According to the theory, initially the universe was so dense that even the light was trapped and could not propagate. However, the universe then faced a phase of very rapid extension which in turn means less denseness and weaker forces. There arrived a certain point (about 380,000 years after the initial singularity) when the forces were weak enough that the light was released. Cosmic microwave background radiation is this ”relic” radiation. Initially, this light had a temperature of about 3000 Kelvin but as the universe continued to expand the temperature dropped. About 14 billion years after, that is today, CMB is still detectable but at a much lower temperature, around 2.7 K.

CMB carries immensely important information about of the universe and therefore it is of tremendous importance to make a full sky measurement of CMB. Firstly, it is the picture of the universe shortly after it has started and houses vital information on the early universe. Secondly, it has something to tell about the current universe: the inhomogeneities in the CMB are the seeds of the galaxies that exist today, and it provides us information about the formation of the galaxies and the topology of the universe. Thirdly, CMB not only informs us about the past and today of the universe but also about the future of the universe. The precise measurement of CMB and the calculation of angular spectrum and certain cosmological constants from it will enable us to choose between competing theories for the future evolution of our universe. In particular, we will be able to tell whether the universe will continue to expand, ending up in a zero entropy state, whether its expansion will stop and the dimension will stabilise or whether it will be followed by a phase of shrinking that will end with a ”big crunch”.

In order to be able to answer such important questions and others, many attempts have been made to measure the CMB accurately. These attempts range from ground based measurements such as DASI to balloon missions such as BOOMERANG and to satellite missions such as COBE. The COBE satellite mission of NASA has detected small-amplitude fluctuations, that is anisotropies in CMB. Considering COBEs limited angular resolution (7 degrees) it became evident that more accurate measurements were needed. The Wilkinson Microwave Anisotropy Probe (WMAP) [1] was launched in 2001 with the aim of obtaining much higher resolution maps of the universe. The third year results of WMAP are already published and are available [1]. Finally, another satellite mission, namely the Planck surveyor, which will provide images with even higher resolution and higher sensitivity on a wider frequency span, will be launched in July 2008 by ESA [2].

An important problem is that the signals measured by these satellites do not contain only radiation from CMB but also contributions from a number of other sources, namely foreground radiations and extragalactic sources. Foreground sources include synchrotron, galactic dust and free-free emission. Extragalactic sources include point sources and the Sunyaev-Zeldovich effect.

The recovery of the CMB signal from this mixture of various sources is a major task and is the subject of this paper. One line of approach is to consider all signals other than CMB as contaminants and try to recover only CMB. Another approach, which we adopt in this work, is to try to recover all sources separately and hence make a picture of other sources as well as CMB. An advantage of this approach is that the other sources also carry important information about the universe and more specifically the Milky Way.

Various work over the last decade has addressed the problem in a source separation framework. In particular, Baccigalupi et al. [3] modelled the problem as a noiseless linear mixture and solved this source separation problem using a gradient descent algorithm. Later, this work was extended to the full sky by Maino et al. [4] who used a fast fixed point-algorithm, namely FastICA [5].

Snoussi et al. [6] and Dellabrouille et al. [7] concentrate on separation in the frequency domain with the motivation of obtaining the angular spectrum of CMB directly and propose a method based on spectral matching. They write a spectral decomposition for each source in the angular spectrum domain and they fit the observations to this decomposition using the Expectation-Maximisation (EM) method. Despite similarity in the technique, in [6] a Bayesian approach has been adopted with Gaussian and entropic priors on the spatial specturm of the components. Moudden et al. later extended this work to wavelet transform domain to be able to deal with data with missing patches [8].

The common drawback of these models is that they lack of a statistical model for the sources and the mixing matrix and therefore are ”blind” in the full sense of the term. To provide a partial solution to the problem, Kuruoglu et al. assumed a generic model, namely Gaussian mixtures for the sources [9] and proposed solution with Independent Factor Analysis where they used two alternative methods, nthe EM algorithm and simulated annealing, for the estimation of source parameters and the mixing matrix. They also employed some known astrophysical information to reduce the number of unknown matrix entries and to restrict the possible values. The results of this work shows clearly the importance of the prior information and the need for a much more flexible framework for incorporating the prior information.

The use of Bayesian methods for this problem were first discussed in [10], where an entropic prior was proposed for the sources derived for sampled images by Skilling. This prior regularizes the inversion such that the most plausible (highest entropy) image is obtained among plausible ones. Stolyarov et al. expand this work to the case of spatially varying noise and spectral properties of foreground components [11]. However, the prior information introduced by this general image prior is minimal and does not exploit any astrophysical information. Kuruoglu and Comparetti, in [12], proposed a Bayesian separation method using Gibbs sampling where a mixture of Gaussians model was adopted and report preliminary results.

Our work can be seen as a continuation of [12] and otherwise is closest to that of [13], who also adopt an independent components model, implemented through Bayesian inference. That work concentrates on obtaining a reconstruction of CMB across the whole sky from WMAP data, and to avoid computational difficulties they use a two-stage inference approach, where the mixing matrix is estimated first by an MCMC procedure on a low resolution map, and then the sources are reconstructed through another MCMC run using a point estimate of the mixing matrix computed from the first stage MCMC output. Priors on the sources and other parameters are assumed uniform, hence the results are based on the likelihood only. In this work we jointly estimate all unknowns in the model — sources, mixing matrix and other parameters — and propose to use Gaussian mixtures to model the sources. This flexible model for the sources allows information about the marginal properties of each source to be incorporated into the analysis.

The rest of the paper is structured as follows. In Section II we give a brief description of the sources and the antenna noise. Based on this information, Section III gives the model for the mixing problem and describes the hierarchical Bayesian model that we use, including the priors we assume for the sources. Section IV describes the MCMC approach we use for the inferring on the model we proposed. Section V provides simulation results on both synthetic

Planck images and real WMAP images. Finally, we provide a discussion of the results in Section VI.

II. SOURCES

The various sources of radiation in the universe can be divided into three main groups. The first is the cosmic microwave background radiation, which is present at every point of the universe. The second are sources originating from inside the galaxy and the third are sources originating outside the galaxy known as extragalactic sources.

A. CMB

The physics of CMB is very well studied and understood theoretically [14], [15]. It is widely accepted that CMB is Gaussian although recently there has been some debated on the deviation of CMB from Gaussianity. In this work, we assume, as does most of the literature, that CMB is Gaussian. It is also widely expected to be stationary, as validated on WMAP data in [16], and that CMB anisotropies can be represented as the multiplication of a spatial template with a nonlinear function of the frequency.

For the recovery of certain cosmological parameters it is convenient to work in the angular spectrum domain and the CMB angular spectrum has a well known structure with a number of peaks [15]. This structure also suggests a long range dependence [17]. The CMB sources used in this work for the simulation example in Section V are generated synthetically using the HEALPIX software package, which assumes a spatially flat standard inflationary CDM model with a Gaussian realisation.

B. Galactic components

1) Synchrotron: Synchrotron radiation is generated by the electrons spiralling (hence accelerating) through magnetic fields. Although synchrotron radiation originates in the galaxy, it extends also to outside the galactic plane and is less concentrated in the galacitic plane when compared to other galactic foreground components. In the simulation studies, the synchrotron map we use is the commonly used one in the literature, i.e. the template taken from 408 MHz Haslam survey extrapolated and scaled to Planck frequencies [18]. A large number of patches from this map were analysed in [9] and was shown that a Gaussian mixture with 3 to 5 components describes well the histogram of the patches.

2) Galactic dust: Galactic dust is made up of small particles which range from the order of nanometers to micrometers. They are made of various materials including silicate and carbon. The shapes vary and they can be of crystalline or amorphous structure. Their radiation is dominant especially in very high frequency channels. In [9], it was shown that a Gaussian mixture with 3 to 5 components provides a good fit to Galactic dust histograms.

3) Free-free emission: Free-free emission or ”Bremsstrahlung” emission is caused by the collision of free electrons with heavy ions in the ionised medium. Electrons loose energy in these collisions and emit photons.

C. Extragalactic sources

1) Sunyaev-Zeldovich effect: The Sunyaev-Zeldovich (SZ) effect is generated with the inverse Compton scattering of photons from CMB on electrons. There are two versions: kinetic SZ effect and thermal SZ effect. They can be observed in the presence of a cluster of galaxies although they can also be caused by any large body with hot ionised gas.

2) Point sources: Point sources are caused by distant stars or galaxies which appear as localised, impulsive bursts of radiation. Unlike the sources discussed above, they are not diffuse and their impulsive behaviour cannot be modelled with a mixture of Gaussian model [19]. It is not possible to consider templates that scale in different frequencies and each channel needs to be considered separately. Due to these properties, the general approach in the literature is to detect and remove them from radiation maps before starting the component separation task [20]. Therefore, in this work we do not consider point sources.

III. MODEL

We observe the sky over J pixels at n_f frequencies. The data are $d_j \in \mathbb{R}^{n_f}$, $j = 1, \ldots, J$. Following the standard independent components analysis model, we assume that d_j can be represented as a linear combination of explanatory variables $s_j \in \mathbb{R}^{n_s}$:

$$d_j \;=\; As_j + e_j, \tag{1}$$

where A is an $n_f \times n_s$ “mixing” matrix and $e_j \sim N(0, \mathrm{diag}(\tau_1^{-1}, \ldots, \tau_{n_f}^{-1}))$. For convenience, define $\underline{\tau} = (\tau_1^{-1}, \ldots, \tau_{n_f})$, $S_k = \{s_{kj}, \,|\, j = 1, \ldots, J\}$ to be the values of the kth component of the s_j, and

$$D \;=\; \{d_{ij} \,|\, i = 1, \ldots, n_f,\, j = 1, \ldots, J\}; \tag{2}$$

$$S \;=\; \{s_{kj} \,|\, k = 1, \ldots, n_s,\, j = 1, \ldots, J\} \tag{3}$$

to represent all data and explanatory variables.

For the application that motivates this work, each component in an observation is the microwave amplitude at a specific frequency, and each component in an s_j is the amplitude of a distinct physical source of those microwaves. The n_f frequencies at which observations are made are denoted $(\nu_1, \ldots, \nu_{n_f})$.

For the application to CMB separation, it is reasonable to parameterize A with a vector θ of considerably smaller dimension; see Section III-B. We write the mixing matrix as $A(\theta)$ to emphasize this point. We assume that each source is independent, defined by a prior distribution $p(S_k \mid \psi_k)$ with parameters ψ_k.

The goal is to estimate the S, the parameters $\underline{\psi} = (\psi_1, \ldots, \psi_{n_s})$ associated with the models for S, θ and the noise variances $\underline{\tau}$, given observation of D. Our model is an example of factor analysis, and in particular we propose to use mixture models for the sources, in which case it is an example of a mixture of factor analysers [21]. In contrast to [21], in this work the mixtures are defined over the sources, rather than over both the mixing matrix and sources, since for this application it is reasonable to assume a common mixing matrix and error over all mixture components. Like in [21], we adopt a Bayesian approach to the data fitting but implemented by MCMC rather than the variational Bayes approach used there. We also assume that the number of components for the mixture of each source is unknown.

Bayesian inference will be based on the posterior distribution, which following the above description can be factorized as:

$$\begin{aligned} p(S, \underline{\psi}, \theta, \underline{\tau} \mid D) &\propto p(D \mid S, A(\theta), \underline{\tau})\, P(S \mid \underline{\psi})\, p(\underline{\psi})\, p(\underline{\tau})\, p(\theta) \\ &= \left[\prod_{j=1}^{J} p(d_j \mid s_j, A(\theta), \underline{\tau})\right] \left[\prod_{k=1}^{n_s} p(S_k \mid \psi_k)\, p(\psi_k)\right] p(\underline{\tau})\, p(\theta). \quad (4) \end{aligned}$$

In the next few subsections, we define each component of this distribution in turn (see Equations 5, 10, 11, 12 and 13).

A. *Noise Structure*

The error e_i primarily rises from detector noise and so can be reasonably assumed to be independent and identically distributed Gaussian. This has been adopted in all source separation studies, where noise is modeled, that we have come across. For most CMB data, and certainly for Planck, the noise variances are known well from extensive simulation. The terms $p(d_j \mid s_j, A(\theta), \underline{\tau})$ are therefore a product of Gaussian density functions.

Although stochastic spatial independence for the noise is reasonable, it is pointed out in [9] that noise is dependent on the pixel index j because the number of observations at each point in the sky may differ, according to the scanning schedule adopted by the detector. We define r_j to be the known number of measurements at the observation indexed by j. Where $r_j \geq 2$, d_j is taken to be the component-wise mean of the measurements and standard statistical theory of the Gaussian distribution implies that the observation precisions of d_j are $r_j \underline{\tau}$. The likelihood term is therefore generalized to:

$$p(d_j \mid s_j, A(\theta), \underline{\tau}) \;=\; \prod_{i=1}^{n_f} (0.5\tau_i r_j/\pi)^{0.5} \exp\left(-0.5\tau_i r_j \left(d_{ij} - A_{i\cdot} s_j\right)^2\right), d_{ij} \in \mathbb{R}, \tag{5}$$

where $A_{i\cdot}$ is the ith row of $A(\theta)$.

B. Mixing Matrix Structure for Astrophysical Microwave Sources

Each column of $A(\theta)$ pertains to a source, and each row to a microwave frequency. Some restrictions are usually placed on $A(\theta)$ in order to force a unique solution; factor analysis can only estimate $A(\theta)$ and S up to a rotation and scaling. Typically in source separation studies for this application, this is achieved by setting one row of $A(\theta)$ to be ones e.g. [22]. We arbitrarily choose the first row to be ones.

The considerable theory and observation about the sources, and what should be their contribution at each frequency, leads to a parameterization of each column of $A(\theta)$. These parameterizations are approximations that come from the current state of knowledge about how the microwaves for each source are generated. Here we merely state the parameterization that we are going to use, and refer to [13], [23] and the may references therein for a more detailed exposition on the background to them.

The CMB is indexed as the first source and so takes the first column of $A(\theta)$. It is modelled as a black body at a temperature $T_0 = 2.725$ (see [24]), and hence its contribution is a known constant at each frequency of the form:

$$A_{i1}(\theta) \;=\; \frac{g(\nu_i)}{g(\nu_1)}, \tag{6}$$

where

$$g(\nu_i) \;=\; \left(\frac{h\nu_i}{k_B T_0}\right)^2 \frac{e^{h\nu_i/k_B T_0}}{(e^{h\nu_i/k_B T_0} - 1)^2},$$

$T_0 = 2.725K$ is the average CMB temperature, h is the Planck constant and k_B is Boltzmann's constant. The ratio $g(\nu_i)/g(\nu_1)$ is taken merely to ensure that $A_{11} = 1$, in line with our policy of constraining the first row of $A(\theta)$ to be ones.

Next we define $A(\theta)$ for the synchrotron radiation column to be:

$$A_{i2}(\theta) = \left(\frac{\nu_i}{\nu_1}\right)^{\theta_s}, \tag{7}$$

where θ_s is an unknown parameter known as the spectral index. Studies such as [25] put a range of $(2.3, 3.0)$ on its value.

The column of $A(\theta)$ for galactic dust is of a slightly more complicated form [23]:

$$A_{i3}(\theta) = \frac{e^{h\nu_1/k_B T_1} - 1}{e^{h\nu_i/k_B T_1} - 1} \left(\frac{\nu_i}{\nu_1}\right)^{\theta_d}, \tag{8}$$

where $T_1 = 18.1K$ and θ_d is the unknown spectral index for dust.

The column of $A(\theta)$ for free-free emission is next and like that for synchrotron [23]:

$$A_{i4}(\theta) = \left(\frac{\nu_i}{\nu_1}\right)^{\theta_f}, \tag{9}$$

where θ_f is the spectral index, thought to lie in the range $(-2.3, -2.1)$.

C. The Sources

We assume that each source $S_k = \{s_{kj} \mid j \in J\}$ is a Gaussian mixture with an unknown number of components m_k. This provides a very flexible but tractable class of models for the sources. Define $\mu_k = (\mu_{k1}, \ldots, \mu_{km_k})$, $t_k = (t_{k1}, \ldots, t_{km_k})$ and $p_k = (p_{k1}, \ldots, p_{km_k})$ to be the mixture component means, precisions and weights for the kth source. Hence the parameters ψ_k of the kth source are $\psi_k = (\mu_k, t_k, p_k, m_k)$ and

$$p(S_k \mid \psi_k) = \prod_{j=1}^{J} \sum_{a=1}^{m_k} p_{ka} \sqrt{\frac{t_{ka}}{2\pi}} \exp\left(-0.5 t_{ka}(s_{kj} - \mu_{ka})^2\right), s_{kj} \in \mathbb{R}. \tag{10}$$

Let $\underline{\mu} = (\mu_1, \ldots, \mu_{n_s})$, $\underline{t} = (t_1, \ldots, t_{n_s})$, $\underline{p} = (p_1, \ldots, p_{n_s})$ and $\underline{m} = (m_1, \ldots, m_{n_s})$ denote the vectors of all mixture means, precisions, weights and number of components for all the sources, so that $\underline{\psi} = (\underline{\mu}, \underline{t}, \underline{p}, \underline{m})$.

D. Priors

The remaining terms in Equation 4 are $p(\psi_k)$, $p(\underline{\tau})$ and $p(\theta)$.

For $\psi_k = (\mu_k, t_k, p_k, m_k)$, we use conjugate prior distributions that facilitate the computation of the posterior and yet are flexible enough to incorporate good prior information. Given m_k, the means $\mu_k = (\mu_{k1}, \ldots, \mu_{km_k})$ are assigned independent and identical normal priors with mean ξ_k and precision κ_k, the precisions $t_k = (t_{k1}, \ldots, t_{km_k})$ are assigned independent and identical gamma priors with shape α_k and scale β_k and the weights p_k are assigned a

Dirichlet prior with equal parameters $(\delta_k, \dots, \delta_k)$. The number of components m_k is assigned a geometric prior with mean λ_k, hence:

$$p(\psi_k) = \left(\prod_{a=1}^{m_k} (0.5\kappa_k/\pi)^{0.5} \exp(-0.5\kappa_k(\mu_{ka} - \xi_k)^2)\, \beta_k^{\alpha_k} t_{ka}^{\alpha_k - 1} e^{-\beta_k t_{ka}} / \Gamma(\alpha_k) \right) \times \frac{\Gamma(m_k \delta_k)}{\Gamma(\delta_k)^{m_k}} p_{ka}^{\delta_k - 1} \, (1 - \lambda^{-1})^{m_k - 1} \lambda^{-1}. \quad (11)$$

This prior specification follows [26].

For $\underline{\tau}$, the conjugate prior is a gamma distribution and in general we use

$$p(\underline{\tau}) \;=\; \prod_{i=1}^{n_f} \frac{\alpha_{\tau,i}^{\beta_{\tau,i}}}{\Gamma(\beta_{\tau,i})} \tau_i^{\beta_{\tau,i} - 1} \, e^{-\alpha_{\tau,i}\tau_i}, \quad (12)$$

for prior scale parameters $\alpha_{\tau,i}$ and shape parameters $\beta_{\tau,i}$. However, for microwave studies the noise precision is known accurately, and in the examples we assume that it is known, so that $p(\underline{\tau})$ is a point mass of 1 at the specified precisions.

Section III-B defines θ to be a vector of spectral indices, for which various studies have established rough bounds on their values. We use independent normal distributions for each θ,

$$p(\theta_k) = \frac{1}{\sqrt{2\pi\sigma_{\theta,k}^2}} \exp(-0.5(\theta_k - m_{\theta,k})^2 / \sigma_{\theta,k}^2). \quad (13)$$

with mean $m_{\theta,k}$ and standard deviation $\sigma_{\theta,k}$ so that 95% of the prior probability lies within the rough bounds. For example, the synchrotron spectral index is believed to lie in the range (2.3,3.0), which leads to a normal prior with mean 2.65 and standard deviation 0.18.

IV. Implementing the Inference

The posterior distribution of Equation 4 is sampled by MCMC. An easy to implement scheme is the Gibb's sampler, with some block updating e.g. we repeatedly sample from the full conditional distributions of blocks of the unknown variables. In many cases the full conditional distributions are of known form, while in other cases they are sampled from by a Metropolis move.

We sample from $p(S, \underline{\psi}, \theta, \underline{\tau} \,|\, D)$ using the following full conditional distributions. One iteration of the MCMC algorithm consists of one sample from each.

a) Sampling s_{kj}: The full conditional distribution of each s_{kj} is (from Equation 4):

$$p(s_{kj} \,|\, \text{everything else}) \;\propto\; \exp\left(-0.5 r_j \sum_{i=1}^{n_f} \tau_i (d_{ij} - A_{i\cdot} s_j)^2\right) \times \sum_{a=1}^{m_k} p_{ka} \sqrt{t_{ka}} \, \exp\left(-0.5 t_{ka} (s_{kj} - \mu_{ka})^2\right), \quad (14)$$

which after some manipulation is seen to be a mixture of m_k Gaussians with precisions $t_{ka} + r_j \sum_{i=1}^{n_f} \tau_i A_{ik}^2$, means

$$\frac{t_{ka}\mu_{ka} + r_j \sum_{i=1}^{n_f} \tau_i A_{ik} \left(d_{ij} - \sum_{\substack{l=1\\ l \neq k}}^{n_s} A_{il} s_{lj}\right)}{t_{ka} + r_j \sum_{i=1}^{n_f} \tau_i A_{ik}^2}$$

and weights p_{ka}, for $a = 1, \ldots, m_k$. It is therefore easily sampled. The s_{kj} are conditionally independent over pixels j for each source, so we sample each s_{kj} separately, for $k = 1, \ldots, n_s$ and $j = 1, \ldots, J$.

b) Alternative block sampling of $s_{\cdot j}$: The vector of sources at pixel j can also be updated jointly because is a mixture of multivariate normal distributions. The number of components in the mixture is $\prod_{k=1}^{n_s} m_k$ e.g. one for each combination of components of the mixtures for each source. It is easiest to label the components by these combinations as $(a_1, \ldots, a_{n_s})$ for $a_k = 1, \ldots, m_k$. Component $(a_1, \ldots, a_{n_s})$ of the mixture has weight $\prod_{k=1}^{n_s} p_{ka_k}$, precision (inverse covariance) matrix denoted by $\mathcal{T}_{(a_1,\ldots,a_{n_s})}$ with elements

$$\left(\mathcal{T}_{(a_1,\ldots,a_{n_s})}\right)_{xy} \;=\; I(x = y) t_{x a_x} + r_j \sum_{i=1}^{n_f} \tau_i A_{ix} A_{iy}, \;\; x, y = 1, \ldots, n_s, \quad (15)$$

and mean vector

$$\mu_{(a_1,\ldots,a_{n_s})} = \mathcal{T}^{-1}_{(a_1,\ldots,a_{n_s})} \left(\begin{pmatrix} t_{1a_1}\mu_{1a_1} \\ \vdots \\ t_{n_s a_{n_s}} \mu_{n_s a_{n_s}} \end{pmatrix} + r_j A^T \begin{pmatrix} \tau_1 d_{1j} \\ \vdots \\ \tau_{n_f} d_{n_f j} \end{pmatrix} \right). \quad (16)$$

The vector $s_{\cdot j}$ is easily sampled from this mixture.

c) Joint sampling of (θ_k, S_k): Experience has shown that attempting to update θ_k on its own by a Metropolis move is difficult because its value is highly correlated with its corresponding source S_k. A better idea is to jointly update (θ_k, S_k) from the full conditional by a Metropolis move. Given the current value θ_k, θ_k^* is proposed from a normal distribution with mean θ_k; all other components of θ are kept constant. The new proposed mixing matrix is denoted $A(\theta_{-k}, \theta_k^*)$ to reflect the fact that only θ_k, and hence the kth column of $A(\theta)$, is changed by the proposal. Then S_k^* is sampled from the full conditional distribution of

Equation 14 using $A(\theta_{-k}, \theta_k^*)$. In the Appendix it is shown that the accept probability for such a move is the minimum of 1 and

$$\frac{\int_{S_k} p(D \mid S, A(\theta_{-k}, \theta_k^*), \underline{\tau})\, p(S_k \mid \mu_k, t_k, p_k)\, dS_k}{\int_{S_k} p(D \mid S, A(\theta), \underline{\tau})\, p(S_k \mid \mu_k, t_k, p_k)\, dS_k}. \tag{17}$$

Equation 23 in the Appendix gives the expression for the integrals in this expression.

d) Sampling τ_i: The full conditional distribution of τ_i is a gamma distribution with scale parameter

$$\alpha_{\tau,i} + 0.5 \sum_{j \in J} r_j (d_{ij} - A_{i\cdot}(\theta) s_j)^2$$

and shape parameter $\beta_{\tau,i} + 0.5|J|$, for $i = 1, \ldots, n_f$. In the CMB application, the τ_i are assumed known and so this sampling is ignored.

e) Sampling (μ_k, t_k, p_k, m_k): To sample these we follow the reversible jump MCMC method of [26]. Each source is updated separately. Associated with each source is a vector of mixture component indices $Z_k = \{z_{kj} \mid j = 1, \ldots, J\}$, $z_{kj} \in \{1, \ldots, m_k\}$, that assign each pixel to a particular component of the Gaussian mixture. The sampling is then done on $(Z_k, \mu_k, t_k, p_k, m_k)$, which turns out to lead to full conditionals of known form. Briefly, the means μ_{ka} are sampled from a normal with means $(t_{ka} \sum_{j:\, z_{kj}=a} s_{kj} + \kappa_k \xi_k)/(t_{ka} n_{ka} + \kappa_k)$ and precisions $t_{ka} n_{ka} + \kappa_k$, where n_{ka} is the number of pixels assigned to component a in Z_k e.g. $n_{ka} = |\{z_{kj} \mid z_{kj} = a\}|$. To avoid the label-switching problem associated with fitting mixture models, the means are constrained to be ordered $\mu_{k1} < \mu_{k2} < \cdots < \mu_{km_k}$. The sample is regarded as a Metropolis proposal and is rejected if this ordering is not maintained. The precisions t_{ka} are sampled from gamma distributions with scale $\alpha_k + 0.5 \sum_{j:\, z_{kj}=a} (s_{kj} - \mu_{ka})^2$ and shape $\beta_k + 0.5 n_{ka}$. The weights p_k are updated jointly from a Dirichlet with parameters $\delta_k + n_{k1}, \ldots, \delta_k + n_{km_k}$. The allocations z_{kj} are sampled from the discrete distribution:

$$p(z_{kj} = a \mid \text{everything else}) \propto p_{ka} \sqrt{t_{ka}} \exp\left(-0.5 t_{ka} (s_{kj} - \mu_{ka})^2\right), \; a = 1, \ldots, m_k,$$

for $j = 1, \ldots, J$. Finally, the number of components m_k can change by a Metropolis move that proposes to either split or merge a component, then either create a new or delete an existing empty component (one for which $n_{ka} = 0$). For these proposals, re-allocation of some existing pixels to new components, and proposals for new means, variances and weights, must take place. We refer to [26] for the details of these, and the associated accept probabilities for the proposals.

V. Examples

A. Simulation Study

The first example is a 512×512 patch of realistic data that have been syntetically generated (CMB) or generated from previously reconstructed sources of synchrotron and dust, as discussed in Section II. These sources are shown in 1. Observations were generated at 9 channels at the frequencies to be observed by Planck ([$30 \quad 44 \quad 70 \quad 100 \quad 143 \quad 217 \quad 353 \quad 545$ 857] GHz), using the mixing matrix model of Section III-B with spectral indices $\theta_s = 2.9$ and $\theta_d = 2$ and Gaussian noise was added with standard deviations [$0.00126 \quad 0.00120 \quad 0.00113$ $0.00028 \quad 0.00018 \quad 0.00018 \quad 0.00018 \quad 0.00018 \quad 0.00018$] at each channel respectively. The measurement precisions τ_i are those that are expected to be attained by the Planck detectors. These maps are shown in 2. The algorithm of the previous section was then implemented. Figures 3, 4 and 5 summarize the analysis for each of the 3 sources. Each figure shows the posterior mean reconstruction, obtained by taking the sample mean of the MCMC samples of each source at each pixel, following a suitable burn-in, and then a scatter plot of the posterior mean at each pixel against the value of the source used to generate the data. In all cases we see that the method has recovered the sources very well. It also recovered the values of the spectral indices of both synchrotron and dust.

B. Analysis of a WMAP patch

The Wilkinson Microwave Anisotropy Probe (WMAP) was launched in 2001 and at the time of writing is still operational. It observes 5 frequencies from 22 to 90 GHz. Figure 6 shows some WMAP data on a 20° square patch of the sky outside the galactic plane, consisting of 512×512 pixels. The images are contaminated by 3 point sources, which are most clearly visible as brights spots on the top 3 images in Figure 6. The algorithm of Section IV was implemented with 4 sources (CMB, synchrotron, dust and free-free emission). The noise precisions $\underline{\tau}$ were assumed known and the spectral density for free-free emission was fixed at -2.14 (following [13]). The algorithm was run for 250,000 iterations and the results are based on the last 150,000 of these.

Figures 7 and 8 show the posterior means and standard deviations of the sources. Figure 9 shows the posterior distribution of CMB at pixel $(200, 20)$ in the patch. Figure 10 is a scatter plot of the posterior expected value of the d_{jk}, defined as:

$$\mathbb{E}(d_{jk} \mid D) = \mathbb{E}(A_{k\cdot}(\theta) s_j \mid D),$$

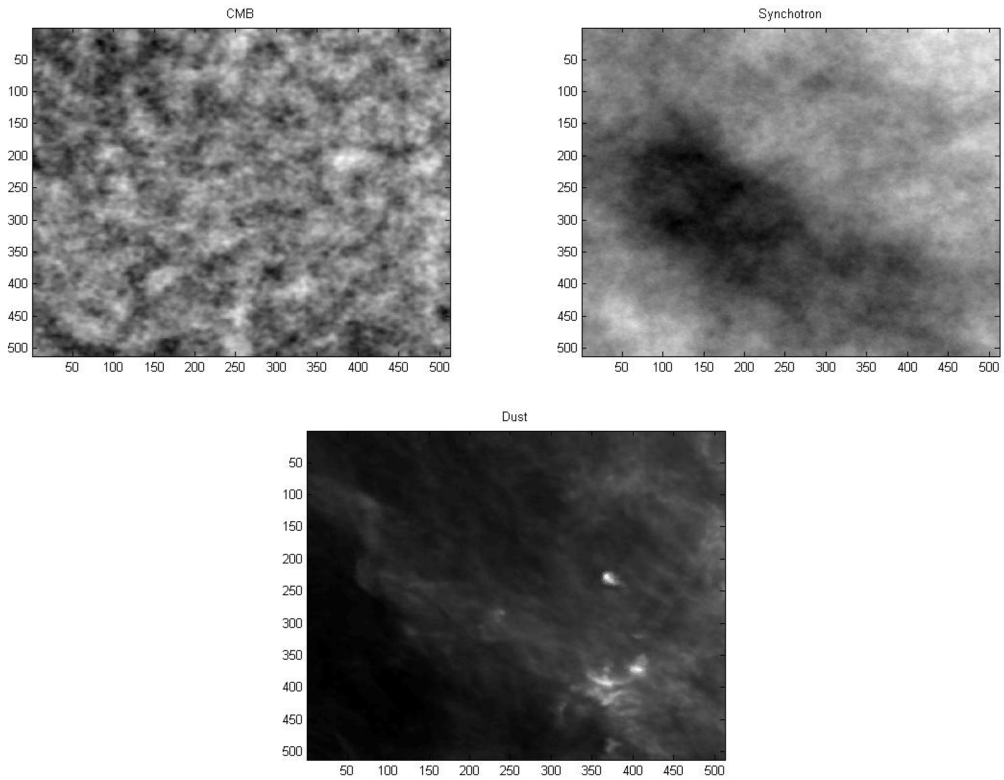

Fig. 1. Simulated Example. Original CMB, synchrotron and galactic dust patches.

obtained by taking the average of $A_{k\cdot}(\theta)s_j$ over the MCMC samples, with the standardised residuals

$$e_{kj} = \sqrt{\tau_i}(d_{jk} - \mathbb{E}(d_{jk} \,|\, D)),$$

with one figure for each frequency $k = 1, \ldots, 5$. These plots show a satisfactory fit of the data to the model; a separation consistent with the data has been produced, with no significantly large residuals or systematic unexplained trend in the residuals being displayed. The point source contamination seems to have affected synchrotron most, and does not appear in the CMB.

VI. Discussion

We have presented a fully Bayesian source separation algorithm for multi-channel image data, where sources are modelled as Gaussian mixtures. The algorithm performs very well on simulated Planck data and has been applied to data from WMAP with success.

We are currently working on obtaining full sky maps for different sources using the WMAP 3rd year data so that angular spectrum of CMB will be constructed as well, the results of

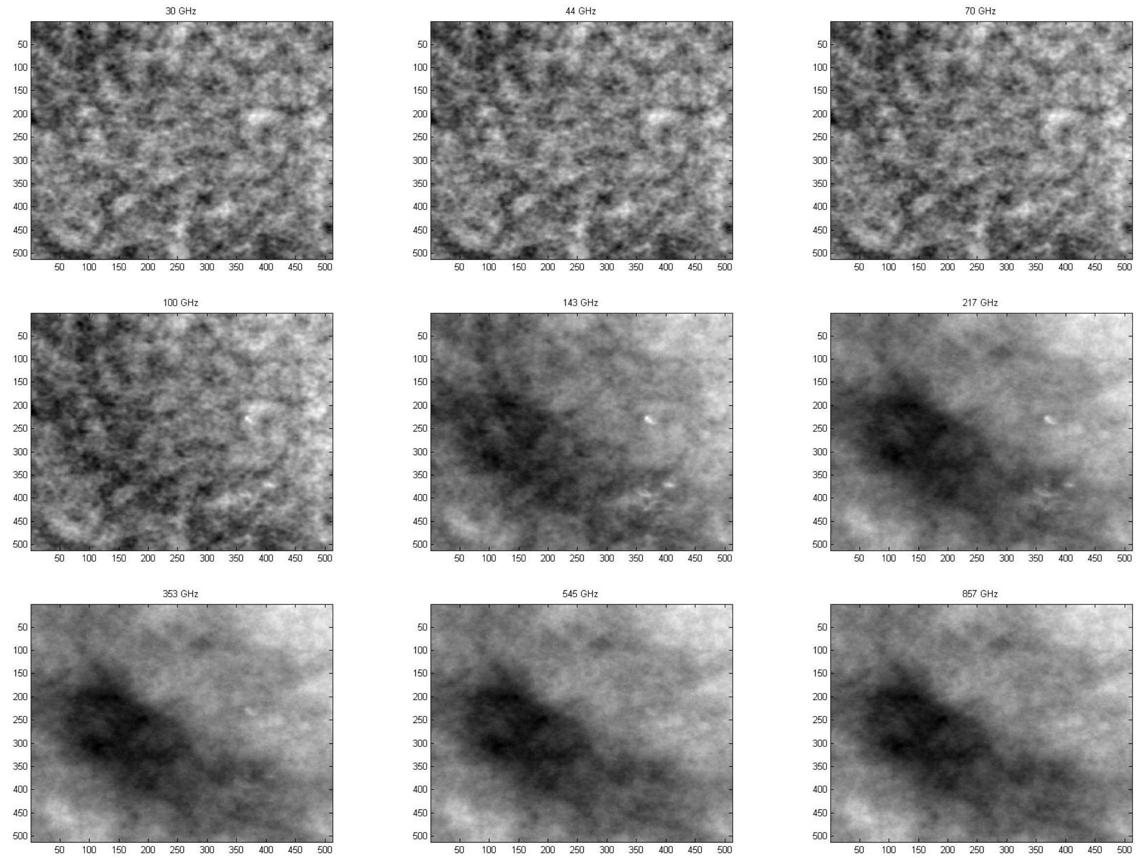

Fig. 2. Simulated Example. Observations on the 30Ghz, 44GHz, 70GHz, 100GHz, 143GHz, 217GHz, 353GHz, 545GHz, 857GHz channels (Planck channels).

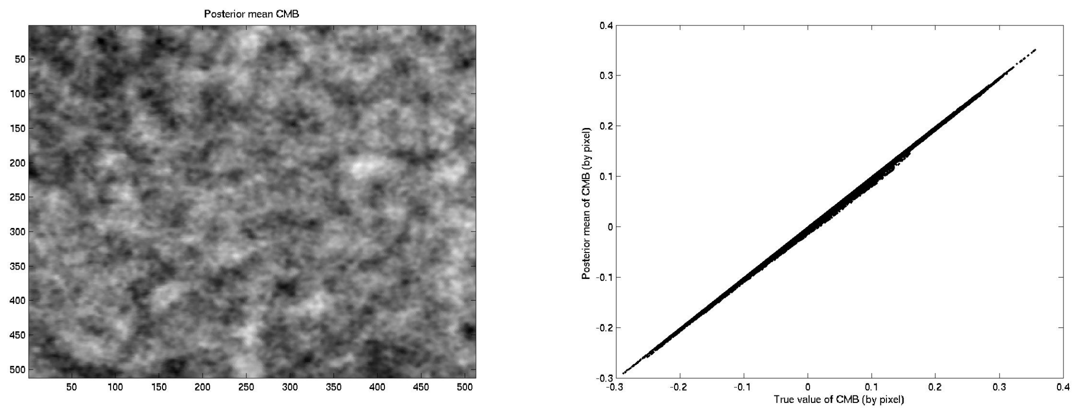

Fig. 3. Simulated Example. The posterior mean reconstruction of the CMB (left) with a scatter plot of true vs posterior mean (right).

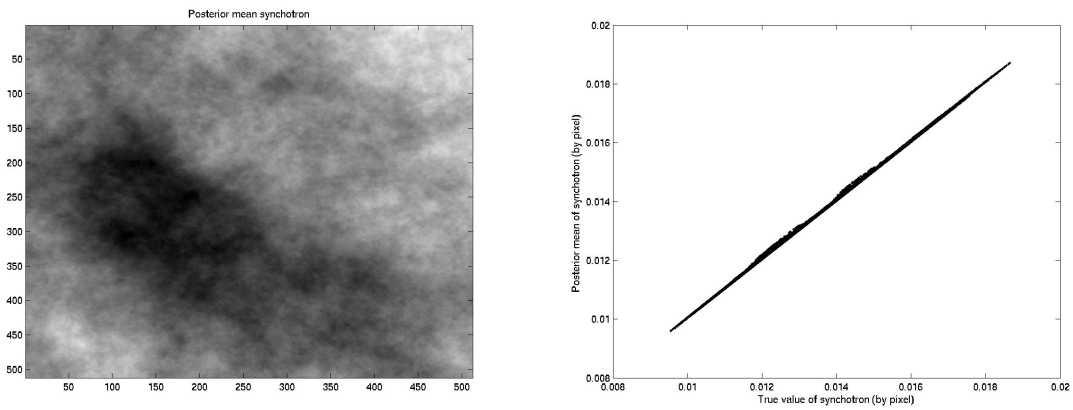

Fig. 4. Simulated Example. The posterior mean reconstruction of synchotron (left) with a scatter plot of true vs posterior mean (right).

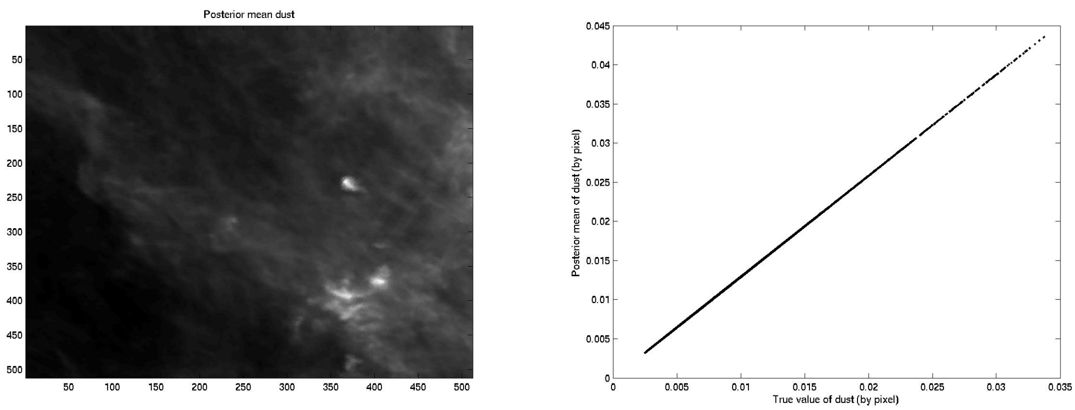

Fig. 5. Simulated Example. The posterior mean reconstruction of dust (left) with a scatter plot of true vs posterior mean (right).

this study will be presented in a followup publication.

Several generalizations are possible. First, there are dependencies between sources, particularly those originating in the galaxy. Dependent component analysis is a relatively straightforward extension of the model, by allowing the source priors to be mixtures of multivariate Gaussians.

Another source of dependence is spatial, that is within a source; clearly sources display spatial smoothness. Generalizing the source prior to a Gaussian Markov random field will allow such dependence to be incorporated into the model.

The technique was developed for instantaneous mixing; however, extension to convolutional mixing is straightforward which can be an efficient way of modelling the finite size beam

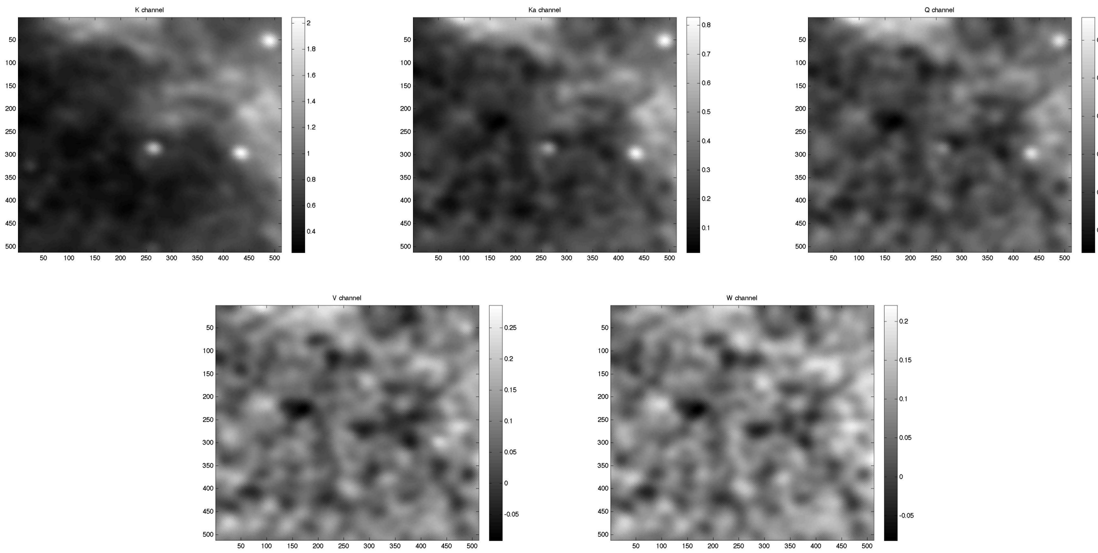

Fig. 6. Data on a 20° square patch of the sky from WMAP at 5 microwave frequencies.

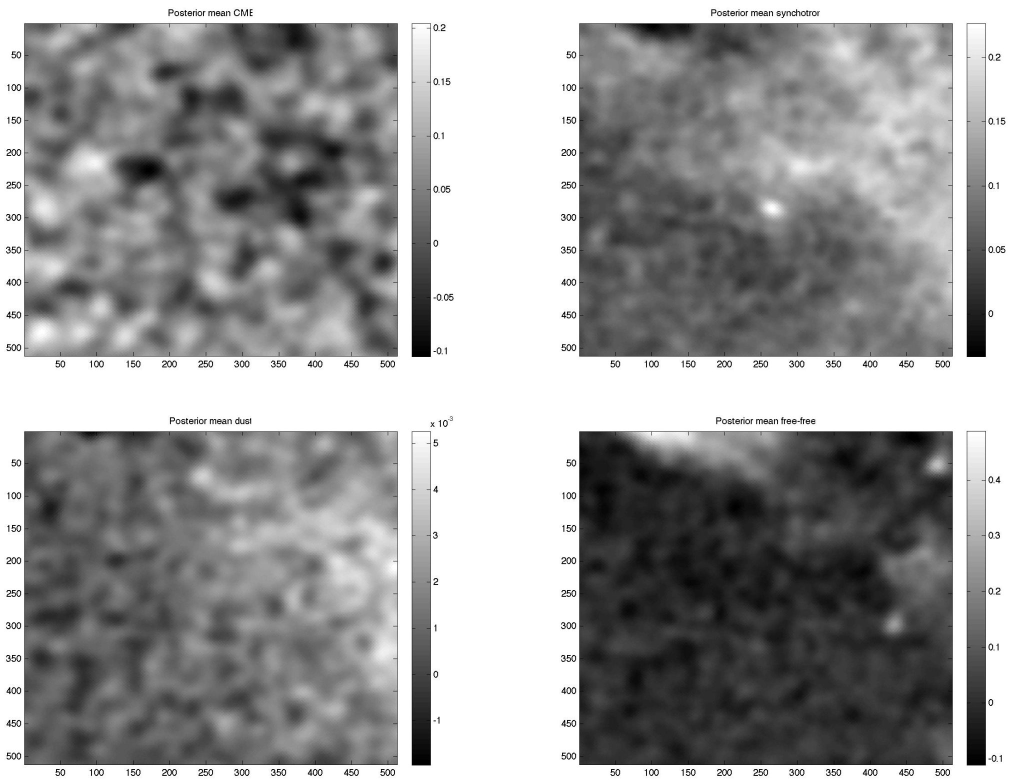

Fig. 7. Posterior means of 4 sources. Clockwise from top left: CMB, synchrotron, free-free emission and dust.

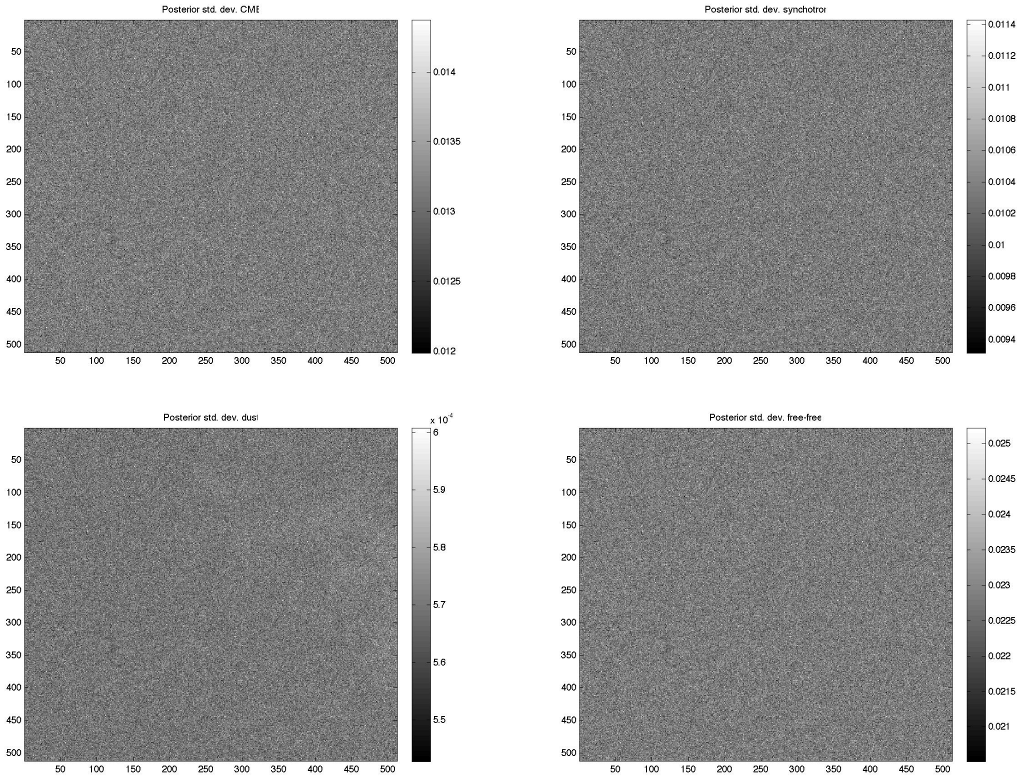

Fig. 8. Posterior standard deviations of 4 sources. Clockwise from top left: CMB, synchrotron, free-free emission and dust.

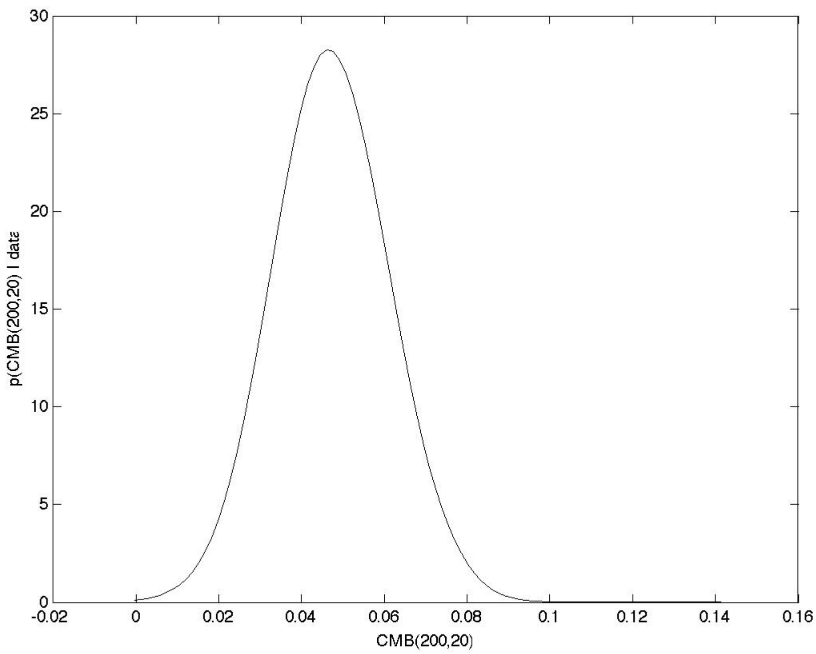

Fig. 9. The posterior distribution of CMB at pixel (200,20) in the patch.

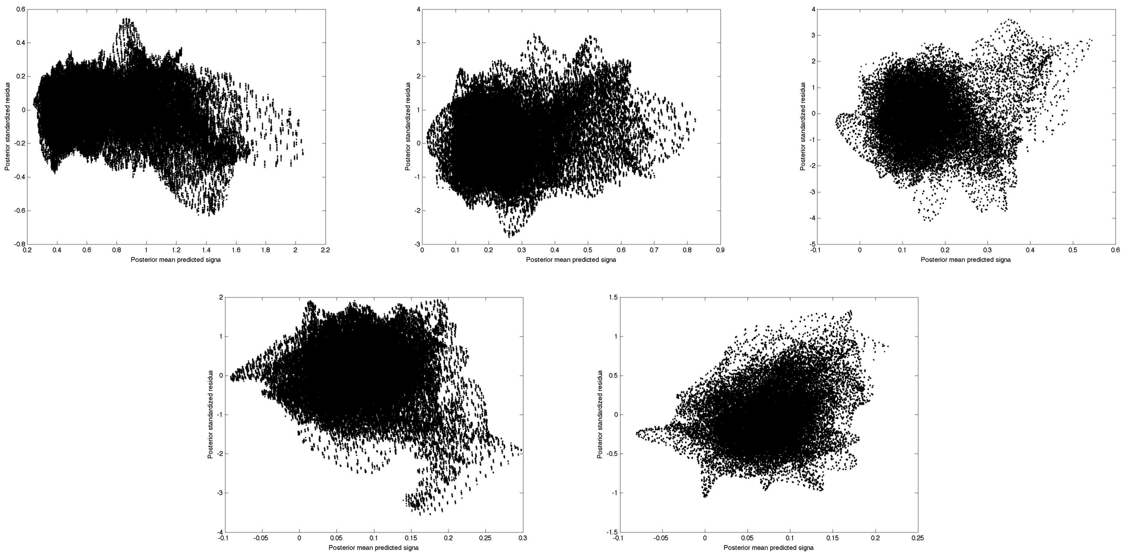

Fig. 10. Assessment of model fit. Scatter plot of the posterior predicted value of d_j against the standardized residual over all pixels at 5 microwave frequencies.

effect.

Generalizing the relationship between sources and data to be non-linear presents no difficulties either, at least in principle; it merely implies a replacement of the linear term $A_i.s_j$ in Equation 5 with the desired non-linear function $f(s_j;\theta)$. However, depending on the nature of the non-linearity, satisfactory convergence and mixing of the MCMC for the parameters θ can become slow and difficult to confirm. Such a generalization would be useful in the galactic plane, where the relationship between sources and observation is not well understood but is certainly not linear because of the interaction, such as absorption and re-emission, by the different causes of the galactic sources.

Finally, although the technique was developed for the astrophysical source separation problem in mind, it is general and is applicable to other source separation problems as well.

Acknowledgements

This work has been made possible by the Network of Excellence MUSCLE, http://www.muscle-noe.org, contract number FP6-507752, funded by the European Union.

References

[1] NASA, “Wilkinson Microwave Anisotropy Probe mission homepage,” http://map.gsfc.nasa.gov/.

[2] ESA, "Planck surveyor homepage," http://www.rssd.esa.int/index.php?project=PLANCK.

[3] C. Baccigalupi, L. Bedini, C. Burigana, G. D. Zotti, A. Farusi, D. Maino, M. Maris, F. Perotta, 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*, vol. 318, pp. 769–780, 2000.

[4] D. Maino, A. Farusi, C. Baccigalupi, F. Perotta, A. J. Banday, L. Bedini, C. Burigana, G. D. Zotti, K. M. Gorski, and E. Salerno, "All-sky astrophysical component separation with fast independent component analysis," *Monthly Notices of the Royal Astronomical Society*, vol. 334, no. 1, pp. 53–68, 2002.

[5] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[6] H. Snoussi, G. Patanchon, J. Macias-Peres, A. Mohammad-Djafari, and J. Delabrouille, "Bayesian blind component separation for cosmic microwave background observations," in *AIP Proceedings of Maximum Entropy and Bayesian Inference Conference*, 2001, pp. 125–140.

[7] J. Delabrouille, J.-F. Cardoso, and G. Patanchon, "Multidetector multicomponent spectral matching and applications for cosmic microwave background data analysis," *Mon. Not. Royal Astronomical Society*, vol. 300, pp. 1–29, 1998.

[8] Y. Moudden, J. F. Cardoso, J. L. Starck, and J. Delabrouille, "Blind component separation in wavelet space: Application to cmb analysis," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 15, pp. 2437–2454, 2005.

[9] E. E. Kuruoğlu, L. Bedini, M. T. Paratore, E. Salerno, and A. Tonazzini, "Source separation in astrophysical maps using independent factor analysis," *Neural Networks*, vol. 16, pp. 479–491, 2003.

[10] M. P. Hobson, A. W. Jones, A. N. Lasenby, and F. R. Bouchet, "Foreground separation methods for satellite observations of the cosmic microwave background," *Mon. Not. Royal Astronomical Society*, vol. 300, pp. 1–29, 1998.

[11] V. Stolyarov, M. P. Hobson, A. N. Lasenby, and R. B. Barreiro, "All-sky component separation in the presence of anisotropic noise and dust temperatute variations," *Mon. Not. Royal Astronomical Society*, vol. 336, pp. 97–111, 2005.

[12] E. E. Kuruoğlu and P. M. Comparetti, "Bayesian separation of independent sources in astrophysical radiation maps using mcmc," in *Proceedings of PHYSTAT (Statistical problems in Particle Physics, Astrophysics and Cosmology)*, L. Lyons, Ed., 2003, pp. 321–323.

[13] H. K. Eriksen, C. Dickinson, C. R. Lawrence, C. Baccigalupi, A. J. Banday, K. M. Górski, F. K. Hansen, P. B. Lilje, E. Pierpaoli, K. M. Smith, and K. Vanderlinde, "CMB component separation by parameter estimation," *Astrophysical Journal*, vol. 641, pp. 665–682, 2006.

[14] M. White and J. D. Cohn, "The theory of anisotropies in the cosmic microwave background," *American Journal of Physics*, vol. 70, no. 2, pp. 106–118, 2002.

[15] W. Hu and S. Dodelson, "Cosmic microwave background anisotropies," *Annual Review of Astronomy and Astrophysics*, vol. 70, no. 2, pp. 106–118, 2002.

[16] E. Komatsu, A. Kogut, M. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, L. Verde, E. Wollack, and E. L. Wright, "First year Wilkinson Microwave Anisotropy Probe (wmap) observations: tests of Gaussianity," *Astrophysical Journal Supplement Series*, vol. 148, pp. 119–134, 2003.

[17] C. Caiafa, A. N. Proto, and E. E. Kuruoglu, "Long-correlated gaussian random fields parameter estimation and noise reduction," *Digital Signal Processing*, vol. 17, no. 4, pp. 819–834, 2007.

[18] C. G. T. Haslam, C. J. Salter, H. Stoffel, and W. E. Wilson, "?" *Astronomy and Astrophysics Supplement Series*, vol. 47, pp. 1–143, 2006.

[19] D. Herranz, E. E. Kuruoglu, and L. Toffolatti, "An alpha-stableapproach to the study of the p(d) distribution of unresolved point sources in cmb sky maps," *Astronomy and Astrophysics*, vol. 424, pp. 1081–1096, 2004.

[20] M. Tegmark, D. J. Eisenstein, W. Hu, and A. de Oliveira-Costa, "Accurate removing point sources from cosmic microwave background maps," *Astrophysical Journal*, 1998.

[21] Z. Ghahramani and M. Beal, "Variational inference for bayesian mixtures of factor analysers," in *Advances in Neural Information Processing Systems*, T. K. L. S. A. Solla and K.-R. Müller, Eds. MIT Press, 2000, vol. 12, pp. 449–455.

[22] L. Bedini, D. Herranz, E. Salerno, C. Baccigalupi, E. E. Kuruoğlu, and A. Tonazzini, "Separation of correlated astrophysical sources using multiple-lag data covariance matrices," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 15, pp. 2400–2412, 2005.

[23] M. Tegmark, D. J. Eisenstein, W. Hu, and A. de Oliveira-Costa, "Foregrounds and forecasts for the cosmic microwave background," *Astrophysical Journal*, vol. 530, pp. 133–165, 2000.

[24] J. C. Mather, D. J. Fixsen, R. A. Shafer, C. Mosier, and D. T. Wilkinson, "Calibrator design for the cobe far infrared absolute spectrophotometer (firas)," *Astrophysics Journal*, vol. 512, pp. 511–520, 1999.

[25] P. Reich, W. Reich, and J. C. Testori, "Spectral index variations of galactic emission," in *The Magnetised Interstellar Medium*, B. Uyaniker, W. Reich, and R. Wielebinski, Eds., 2003, pp. 63–70.

[26] S. Richardson and P. Green, "On Bayesian analysis of mixtures with an unknown number of components (with discussion)," *Journal of the Royal Statistical Society, Series B*, vol. 59, pp. 731–792, 1997.

APPENDIX

In this appendix we derive the accept probability for the joint proposal of (θ_k, S_k) in the MCMC method described in Section IV.

The target distribution is the joint full conditional distribution of θ_k and S_k, which depends only on D, S_{-k} (all the other sources except for S_k), $\underline{\tau}$ and the normal mixture parameters for S_k. Given the current value θ_k, θ_k^* is proposed from a normal distribution with mean θ_k; all other components of θ are kept constant. The new proposed mixing matrix is denoted $A(\theta_{-k}, \theta_k^*)$ to reflect the fact that only θ_k, and hence the kth column of $A(\theta)$, is changed by the proposal. Then S_k^* is sampled from the full conditional distribution of Equation 14 using $A(\theta_{-k}, \theta_k^*)$. The accept probability for such a move is the usual target distribution ratio times proposal distribution ratio:

$$\min\left\{1, \frac{p(\theta_k^*, S_k^* \,|\, \text{everything else})}{p(\theta_k, S_k \,|\, \text{everything else})} \, \frac{p(S_k \,|\, \text{everything else with } \theta_k)}{p(S_k^* \,|\, \text{everything else with } \theta_k^*)}\right\}. \tag{18}$$

The full conditional of (θ_k, S_k) is, from the relevant elements of Equation 4, given by $p(D \,|\, S, A(\theta), \underline{\tau}) \, p(S_k \,|\, p_k, \mu_k, t_k) p(\theta_k)$. For the ratio of full conditionals of S_k^* to S_k, we must be careful since these are full conditionals that are conditional on different values of θ_k. Since $p(\theta, S_k \,|\, \text{everything else}) \propto p(D \,|\, S, A(\theta), \underline{\tau}) \, p(S_k \,|\, p_k, \mu_k, t_k) p(\theta_k)$, so

$$p(S_k \,|\, \text{everything else with } \theta_k) \propto \frac{p(D \,|\, S, A(\theta), \underline{\tau}) \, p(S_k \,|\, p_k, \mu_k, t_k) p(\theta_k)}{\int p(D \,|\, S, A(\theta), \underline{\tau}) \, p(S_k \,|\, p_k, \mu_k, t_k) p(\theta_k) \, dS_k}.$$

Thus the accept probability is:

$$\min\left\{1, \frac{p(D \mid S^*, A(\theta_{-k}, \theta_k^*), \underline{\tau})\, p(S_k^* \mid p_k, \mu_k, t_k)}{p(D \mid S, A(\theta), \underline{\tau})\, p(S_k \mid p_k, \mu_k, t_k)}\right.$$

$$\left. \times \frac{p(D \mid S, A(\theta), \underline{\tau})\, p(S_k \mid p_k, \mu_k, t_k) / \int p(D \mid S, A(\theta), \underline{\tau})\, p(S_k \mid p_k, \mu_k, t_k)\, dS_k}{p(D \mid S^*, A(\theta_{-k}, \theta_k^*), \underline{\tau})\, p(S_k^* \mid p_k, \mu_k, t_k) / \int p(D \mid S, A(\theta_{-k}, \theta_k^*), \underline{\tau})\, p(S_k \mid p_k, \mu_k, t_k)\, dS_k}\right\}$$

$$= \min\left\{1, \frac{\int_{S_k} p(D \mid S, A(\theta_{-k}, \theta_k^*), \underline{\tau})\, p(S_k \mid \mu_k, t_k, p_k)\, dS_k}{\int_{S_k} p(D \mid S, A(\theta), \underline{\tau})\, p(S_k \mid \mu_k, t_k, p_k)\, dS_k}\right\} \quad (19)$$

where we have also substituted the uniform prior for θ_k is uniform.

The integrand is the product of a Gaussian likelihood term $p(D \mid S, A(\theta), \underline{\tau})$ (see Equation 5) and an iid Gaussian mixture $p(S_k \mid p_k, \mu_k, t_k)$ (see Equation 10).

After rearranging the exponent term to isolate the s_{kj}, $p(D \mid S, A(\theta), \underline{\tau})$ can be written:

$$p(D \mid S, A(\theta), \underline{\tau})$$

$$= \left(\prod_{i=1}^{n_f} \sqrt{\frac{\tau_i}{2\pi}}\right)^{0.5|J|} \prod_{j=1}^{J} \exp\left(-0.5 \sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2 r_j \left(s_{kj} - d_{ij}/A_{ik}(\theta) + \sum_{\substack{l=1\\l\neq k}}^{n_s} A_{il}(\theta) s_{lj}/A_{ik}(\theta)\right)^2\right), \quad (20)$$

which, after expanding the exponent and completing the square, we can write as

$$p(D \mid S, A(\theta), \underline{\tau})$$

$$\propto \prod_{j=1}^{J} \exp\left\{-0.5\left(\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2 r_j\right)\left(s_{kj} - \frac{\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2 r_j \left(d_{ij}/A_{ik}(\theta) - \sum_{\substack{l=1\\l\neq k}}^{n_s} A_{il}(\theta) s_{lj}/A_{ik}(\theta)\right)}{\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2 r_j}\right)^2\right.$$

$$\left. + 0.5 \frac{\left(\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)\left(d_{ij} - \sum_{\substack{l=1\\l\neq k}}^{n_s} A_{il}(\theta) s_{lj}\right)\right)^2}{\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2}\right\}$$

$$= \prod_{j=1}^{J} \exp\left\{-0.5\left(r_j \sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2\right)\left(s_{kj} - \frac{\sum_{i=1}^{n_f} \tau_i \left(A_{ik}(\theta) d_{ij} - \sum_{\substack{l=1\\l\neq k}}^{n_s} A_{ik}(\theta) A_{il}(\theta) s_{lj}\right)}{\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2}\right)^2\right.$$

$$\left. + 0.5 \frac{\left(\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)\left(d_{ij} - \sum_{\substack{l=1\\l\neq k}}^{n_s} A_{il}(\theta) s_{lj}\right)\right)^2}{\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2}\right\}, \quad (21)$$

where the constants that we have ignored do not concern us as they do not depend on θ_k, and so will cancel out when we take the ratio in the accept probability.

Thus, both terms in the integrand of Equation 19 factorize into functions of each s_{kj}, hence our integral is the product of 1-dimensional integrals. Subsituting Equation 20 into the integrand, we have:

$$\begin{aligned}
&\int p(D \,|\, S, A(\theta), \underline{\tau})\, p(S_k \,|\, p_k, \mu_k, t_k)\, dS_k \\
&\qquad \propto \prod_{j=1}^{J} \sum_{a=1}^{m_k} p_{ka} \exp \left\{ 0.5 \frac{\left(\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta) \left(d_{ij} - \sum_{\substack{l=1 \\ l \neq k}}^{n_s} A_{il}(\theta) s_{lj} \right) \right)^2}{\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2} \right\} \\
&\times \int_{-\infty}^{\infty} \exp \left(-0.5 \left(r_j \sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2 \right) \left(s_{kj} - \frac{\sum_{i=1}^{n_f} \tau_i \left(A_{ik}(\theta) d_{ij} - \sum_{\substack{l=1 \\ l \neq k}}^{n_s} A_{ik}(\theta) A_{il}(\theta) s_{lj} \right)}{\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2} \right)^2 \right) \\
&\qquad \times \sqrt{\frac{t_{ka}}{2\pi}} \exp\left(-0.5 t_{ka} (s_{kj} - \mu_{ka})^2 \right)\, ds_{kj}
\end{aligned} \tag{22}$$

We now see that the integral is that of a Gaussian mixture over a Gaussian mean (treating s_{kj} as the "mean"); standard results show this to be another Gaussian with variance as the sum of the two Gaussians in the integral (e.g. the sum t_{ka}^{-1} and $\left(r_j \sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2 \right)^{-1}$) and mean as the mean of the mixing Gaussian (e.g. in this case, s_{kj} will be replaced by μ_{ka}).

Thus:

$$
\begin{aligned}
&\int p(D \mid S, A(\theta), \underline{\tau})\, p(S_k \mid p_k, \mu_k, t_k)\, dS_k \\
&\qquad\qquad \propto \prod_{j=1}^{J} \sum_{a=1}^{m_k} p_{ka} \exp\left\{ 0.5 \frac{\left(\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta) \left(d_{ij} - \sum_{\substack{l=1 \\ l \neq k}}^{n_s} A_{il}(\theta) s_{lj} \right) \right)^2}{\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2} \right\} \\
&\qquad\qquad\quad \times \sqrt{\frac{2\pi}{r_j \sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2}} \times \frac{1}{\sqrt{2\pi(t_{ka}^{-1} + (r_j \sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2)^{-1})}} \\
&\times \exp\left(- \frac{1}{2(t_{ka}^{-1} + (r_j \sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2)^{-1})} \left(\mu_{ka} - \frac{\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta) \left(d_{ij} - \sum_{\substack{l=1 \\ l \neq k}}^{n_s} A_{il}(\theta) s_{lj} \right)}{\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2} \right)^2 \right) \\
&= \prod_{j=1}^{J} \sum_{a=1}^{m_k} p_{ka} \frac{1}{\sqrt{1 + r_j \sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2 / t_{ka}}} \exp\left\{ 0.5 \frac{\left(\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta) \left(d_{ij} - \sum_{\substack{l=1 \\ l \neq k}}^{n_s} A_{il}(\theta) s_{lj} \right) \right)^2}{\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2} \right\} \\
&\qquad\qquad \times \exp\left(-0.5 \frac{\left(\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta) \left(d_{ij} - A_{ik}(\theta)\mu_{ka} - \sum_{\substack{l=1 \\ l \neq k}}^{n_s} A_{il}(\theta) s_{lj} \right) \right)^2}{(t_{ka}^{-1} + (r_j \sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2)^{-1}) \left(\sum_{i=1}^{n_f} \tau_i A_{ik}(\theta)^2 \right)^2} \right). \qquad (23)
\end{aligned}
$$

Substituting Equation 23 into Equation 19 gives the required accept probability.