Bayesian Factor Analysis using Gaussian Mixture Sources, with Application to Separation of the Cosmic Microwave Background

Simon P. Wilson  
School of Computer Science and Statistics  
Trinity College Dublin  
Dublin, Ireland  
E-mail: simon.wilson@tcd.ie

Ercan E. Kuruoğlu  
Istituto di Scienza e Tecnologie dell’Informazione  
Consiglio Nazionale delle Ricerche  
Pisa, Italy  
E-mail: ercan.kuruoglu@isti.cnr.it

Alicia Quirós Carretero  
Departamento de Estadística e Investigación Operativa  
Universidad Rey Juan Carlos  
Madrid, Spain  
E-mail: alicia.quiros@urjc.es

I. INTRODUCTION

In this paper a fully Bayesian factor analysis model is developed that assumes a very general model for each factor, namely the Gaussian mixture. We discuss the cases where factors are both independent and dependent. In the statistical literature, factor analysis has been used principally as a dimension reduction technique, with little interest in a priori modelling of the factors, but here the application is source separation where the factors may have a direct interpretation and the usual Gaussian model for a factor may not be appropriate. That is the case for the application that illustrates our work, which is that of identifying different sources of extra-terrestrial microwaves from all-sky images taken at different frequencies. In particular there is interest in separating out the cosmic microwave background (CMB) signal from the other sources.

The posterior distribution is computed by Monte Carlo sampling, and the separated sources are estimated as averages of the samples from the posterior distribution. Beyond this, further information can be extracted from the samples if desired, such as: estimates of uncertainty in the separation, like the standard deviation point-wise of the source samples, or functions of interest like the mean of the spectral density of the samples. The ability to do this is one of the principal benefits of the Bayesian approach.

Various work over the last decade has addressed the CMB problem in a factor analysis or source separation framework. In particular, [1] modelled the problem as a noiseless linear mixture and solved this separation problem using a gradient descent algorithm. Later, this work was extended to the full sky in [2] who used FastICA [3], a fast fixed point-algorithm. A Bayesian approach was discussed in [4], where an entropic prior was proposed for the sources. Such a prior regularizes the separation such that the most plausible (highest entropy) sources are obtained from the data among all plausible ones. This work was generalized to the case of spatially varying noise and spectral properties of foreground components in [5]. However, the problem in this case was considered as pure source reconstruction and not model learning, since the factor loading or mixing matrix was assumed to be perfectly known. In [6] a generic model was proposed, namely GMM for the sources and suggested Independent Factor Analysis for the solution. This approach has assumed fixed but unknown model parameters which were learned using either EM algorithm or simulated annealing. This was extended in [7] to adopt prior distributions for model parameters hence moving to a fully Bayesian model. The resulting model was learned with Gibbs sampling. The work in this paper can thus be considered as an ideal logical continuation of [6], for the data model, and [7], for the sampling approach.

In this work we jointly estimate by sampling all unknowns in the model — sources, mixing matrix and other parameters — and propose to use Gaussian mixtures with an unknown number of factors to model the sources. This prior distribution is a highly flexible prior model that can model any continuous probability distribution to arbitrary accuracy, thus at least in principal allowing any information about the marginal distribution of the sources over the sky to be incorporated.

The rest of the paper is structured as follows. Section II 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 III describes the MCMC approach we use for the inferring on the model we proposed. Section IV provides results on real images. Finally, we provide a discussion of the results in Section V.

II. MODEL

The model description is defined in terms of the microwave source separation problem, where there are $n_f$ maps of the sky at frequencies $(\nu_1, \ldots, \nu_{n_f})$, each map consisting of $J$ pixels. The data are denoted $d_{j} \in \mathbb{R}^{n_f}$, $j = 1, \ldots, J$. The source model consists of $n_s$ sources and is represented by the vectors $s_j \in \mathbb{R}^{n_s}$, with each component representing the amplitude of a distinct physical source of those microwaves. We assume that the $d_j$ follow a standard statistical independent components analysis model, so that they can be represented...
as a linear combination of the $s_j$:

$$d_j = A s_j + e_j,$$

where $A$ is an $n_f \times n_s$ “mixing” matrix and $e_j$ is a vector of $n_f$ independent Gaussian error terms with precisions (inverse variances) $\tau_1, \ldots, \tau_{n_f}$. For convenience, define $\tau = (\tau_1, \ldots, \tau_{n_f})$. $S_k = \{ s_{kj} | j = 1, \ldots, J \}$ to be the values of the $k$th source over all pixels, and

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

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

(2) (3)

to represent all data and sources.

For the application to CMB separation, it is reasonable to parameterize $A$ with a vector $\theta$ of considerably smaller dimension; see Section II-0b. We write the mixing matrix as $A(\theta)$ to emphasize this point. We assume that each source is independent, defined by a prior distribution $p(S_k | \psi_k)$ with parameters $\psi_k$.

The goal is to estimate the $S$, the parameters $\psi = (\psi_1, \ldots, \psi_{n_s})$ associated with the models for $S$, and $\theta$, given observation of $D$. The noise variances $\tau$ are assumed known. GMMs are used to represent the non-Gaussian sources, in which case it is an example of a model known as a mixture of GMMs. Two models are proposed. In the first, there is independence both between sources and across a source spatially. For this model, define $\mu_k = (\mu_{k1}, \ldots, \mu_{km_s})$, $t_k = (t_{k1}, \ldots, t_{km_s})$ and $p_k = (p_{k1}, \ldots, p_{km_s})$ to be the mixture factor means, precisions and weights for the $k$th source. Hence the parameters $\psi_k$ of the $k$th source are $\psi_k = (\mu_k, t_k, p_k, m_k)$ and

$$p(S | \psi) = \prod_{j=1}^n \prod_{k=1}^m \sum_{p_k} \frac{1}{\sqrt{2\pi}} \exp \left( -0.5 t_{ka} (s_{kj} - \mu_{ka})^2 \right),$$

(4)

Each element of this distribution is defined next in turn.

a) Noise Structure: Gaussian error $e_j$ is assumed, independent within and between pixels $j$ and frequency, which gives

$$p(D | S, A(\theta), \tau) = \prod_{j=1}^{n_h} (0.5\tau_j/\pi)^{0.5} \exp \left( -0.5\tau_j (d_{ij} - A_i s_j)^2 \right), d_{ij} \in \mathbb{R},$$

(5)

where $A_i$ is the $i$th row of $A(\theta)$.

b) Mixing Matrix Structure for Astrophysical Microwave Sources: 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 permutation. Typically, in source separation studies for this application, this is achieved by setting a row of $A(\theta)$ to be ones [9]; here the first row is set to 1. There is a considerable amount of physical theory and observation about the sources [10], [11] that define $A(\theta)$. For example, CMB is modelled as a black body at a temperature $T_0 = 2.725$ [12], and so:

$$A_{11}(\theta) = \frac{g(\nu_1)}{g(\nu_1)},$$

(6)

where

$$g(\nu) = \left( \frac{\hbar \nu}{k_B T_0} \right)^2 \frac{\exp \left( \frac{\hbar \nu}{k_B T_0} - 1 \right)}{\left( \frac{\hbar \nu}{k_B T_0} - 1 \right)^2},$$

$T_0 = 2.725K$ is the average CMB temperature, $\hbar$ is the Planck constant and $k_B$ is Boltzmann’s constant. The ratio $g(\nu_1)/g(\nu_1)$ is taken merely to ensure that $A_{11}(\theta) = 1$. For another source, synchrotron radiation, the relationship is

$$A_{12}(\theta) = \left( \frac{\nu_2}{\nu_1} \right)^{\theta_s},$$

(7)

where the spectral index, $\theta_s$ is typically in the range ($-2.3, -3.0$) [13]. Similar relationships exist for other sources; see [11].

c) The Sources: The distribution of $s_{kj}$, source $k$ at pixel $j$, is modeled as a GMM with an unknown number of factors $m_k$. This provides a very flexible but tractable class of models for the sources. Two models are proposed. In the first, there is independence both between sources and across a source spatially. For this model, define $\mu_k = (\mu_{k1}, \ldots, \mu_{km_s})$, $t_k = (t_{k1}, \ldots, t_{km_s})$ and $p_k = (p_{k1}, \ldots, p_{km_s})$ to be the mixture factor means, precisions and weights for the $k$th source. Hence the parameters $\psi_k$ of the $k$th source are $\psi_k = (\mu_k, t_k, p_k, m_k)$ and

$$p(S | \psi) = \prod_{j=1}^n \prod_{k=1}^m \sum_{p_k} \frac{1}{\sqrt{2\pi}} \exp \left( -0.5 t_{ka} (s_{kj} - \mu_{ka})^2 \right),$$

(8)

$s_{kj} \in \mathbb{R}$. Let $\mu = (\mu_1, \ldots, \mu_{m_s})$, $t = (t_1, \ldots, t_{m_s})$, $p = (p_1, \ldots, p_{m_s})$ and $m = (m_1, \ldots, m_{m_s})$ denote the vectors of all mixture means, precisions, weights and number of factors for all the sources, so that $\psi = (\mu, t, p, m)$. The second model allows for between source dependence; the vector of sources at a pixel is a mixture of multivariate Gaussians:

$$p(S | \psi) = \prod_{j=1}^n \prod_{a=1}^m p_a \frac{|Q_a|^{0.5}}{2^{n_p} 2^{m_s}} \times \exp \left( -0.5 (s_j - \mu_a)^T Q_a (s_j - \mu_a) \right),$$

(9)

for mixture component weights $p_a$, mean vectors $\mu_a$ and precision matrices $Q_a$. In this case $\psi$ is all the $p_a$, $\mu_a$ and $Q_a$.

d) Priors: The remaining terms in Equation 4 are $p(\psi)$ and $p(\theta)$. For $\psi$, we use the conjugate prior distributions [14, Chapter 2] that facilitate the computation of the posterior and yet are flexible enough to incorporate good prior information: Gaussians for the component means, Dirichlet for the component weights, gamma for the component precisions $t_{ka}$ and Wishart for precision matrices $Q_a$. In the microwave source application, background knowledge about the magnitude of the sources can be incorporated through specifying values of
the parameters of these prior distributions. Rough bounds on
the values of the spectral indices $\theta$ are also known. We use
independent normal distributions for each $\theta$ with mean and
standard deviation 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
0.175.

III. IMPLEMENTING THE SOURCE SEPARATION

The source separation is implemented in two stages as
follows. The first stage is Monte Carlo sampling of the
posterior distribution of Equation 4, specifically by a Markov
chain Monte Carlo method (MCMC) [15]. We do this rather
than approximate Equation 4 by a variational approach, as
in [8], because of concerns about the underestimation of
posterior variance with variational methods [16]; since one
of the principal advantages of the Bayesian approach is the
availability of estimates of uncertainty, we opted for the more
computationally intensive MCMC. Once this is done, the
second stage is to compute the average of the samples of the
sources; this average is taken to be the estimated source.

An easy to implement MCMC scheme to sample from
Equation 4 is the Gibbs 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 by a Metropolis step within the Gibbs
sampler [17], [18]. The $\psi_k = (\mu_k, t_k, p_k, m_k)$ are sampled
by a reversible jump MCMC method of [19]. The details of
the sampling distributions can be found in [20].

IV. EXAMPLE: ANALYSIS OF A WMAP 5 YEAR 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 1 shows
some recently released 5-year WMAP data at $20^\circ$ north, $60^\circ$
west, consisting of $128 \times 128$ pixels.

The algorithm of Section III was implemented with 4
sources (CMB, synchrotron, dust and free-free emission). The
noise precisions $\tau$ were assumed to be the published values for
WMAP’s detectors. The spectral density for free-free emission
was fixed at $-2.14$ (following [11]). The priors for the
synchrotron and dust spectral indices were the same as those
for the simulated example, as were the priors for the number
of factors in the GMM source models. Informative priors were
placed on the GMM parameters, based on discussions on the
expected marginal properties of the sources. For example, the
prior distributions of the GMM parameters for CMB were
$\xi_1 = 0, \kappa_k = 100, \alpha_1 = 0.005$ and $\beta_1 = 3$. This corresponds
to a Gaussian prior distribution for the CMB mean $\mu_1$ with
mean 0 and variance 0.01, and a gamma prior distribution on
the CMB precision $t_1$ with mean 0.05 and standard deviation
0.05.

The MCMC algorithm was run for 50,000 iterations for the
independent source case. To check for adequate convergence
and mixing, trace plots, auto-correlation functions and the
Gelman-Rubin diagnostic [21] were computed over 2 runs of
the MCMC from different starting values. Trace plots showed
that the algorithm appeared to have converged in 20,000
– 30,000 iterations for the values of the sources, although
the spectral indices were taking a longer time to converge.
Autocorrelations were also low for the sampled values of the
sources.

Figure 2 shows the average of the last 10,000 samples of the
CMB from one of the runs, along with an estimate of CMB from this patch obtained by part of the WMAP team
[22], while Figure 3 shows the average of the samples of the
other sources. We see that in this patch, the algorithm has
inferred that dust and free-free are relatively constants,
and most variation is explained by CMB and synchrotron. We
have obtained a result that is in agreement with [22]. The
posterior mean of $\theta_3$ and $\theta_4$ are $-3.0$ and 0.8 respectively; the
latter figure reflecting the fact that there appears to be little
contribution from dust in this patch. This could also explain
the slow mixing of $\theta_3$ in the MCMC.

For model assessment, Figure 4 is a scatter plot of the
observed values $d_{ij}$ against the standardised residuals
$$R_{ij} = \sqrt{\tau_i}(d_{ij} - \mathbb{E}(d_{ij} | D)),$$
where
$$\mathbb{E}(d_{ij} | D) = \mathbb{E}(A_i(\theta)s_j | D),$$
is the posterior expectation of $d_{ij}$ under the model, obtained
by taking the average of $A_i(\theta)s_j$ over the MCMC samples.
A good model fit is indicated by a lack of trend in the plot,
and values of the residual consistent with the standard normal
distribution e.g. rarely outside the interval $(3, -3)$. There is
one figure for each frequency $i = 1, \ldots, 5$. 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.

V. DISCUSSION AND FURTHER WORK

A fully Bayesian factor analysis algorithm has been pre-
presented and applied to a multi-channel image source separation
problem, where sources are modelled as GMMs. The algo-
program performs very well on simulated Planck data and has
been applied to data from WMAP.

Current work is on obtaining full sky maps for different
sources using the WMAP 5th year data, so that angular
spectrum of CMB will be constructed as well. The results
of this study will be presented in a follow up publication. For
this reason inference for the CMB power spectrum [23] is not
discussed. However, since such a map is sampled by MCMC
then it is a straightforward matter to compute the posterior
distribution of the power spectrum and summary statistics of
it as follows. For each MCMC sample of the CMB, the power
spectrum is computed. Then from this set of “samples” of
power spectra, a pointwise sample mean can be taken as the
best posterior estimate of the power spectrum. Uncertainty
Fig. 1: Temperatures (in mKs) a 20° square patch of the sky from WMAP at 5 microwave frequencies.

Fig. 2: Average of 10,000 samples of CMB (left) and a reconstruction from the WMAP team (right) in mKs.

Fig. 3: Posterior means of synchrotron (top left), dust (top right) and free-free emission (bottom) in mKs.

Fig. 4: Assessment of model fit. Scatter plots of the observed value of $d_{ij}$ against the standardized residual over all pixels at the 5 frequencies.

in the power spectrum can be quantified by computing, for example, the (2.5, 97.5) sample percentiles pointwise.

Several generalizations are possible. An important model assumption, that is assumed in all model-based source separation approaches to date, is that *a priori* sources are independent from each other. While our results show that such dependencies clearly exist in the posterior distribution, due to the stochastic linear constraint that $s_j \approx Ad_j$, prior modelling of them should help to produce a more realistic separation. As we stated in the Introduction, this is a relatively straightforward extension of the model by allowing the source priors to be mixtures of multivariate Gaussians; the posterior distribution of 4 remains the same except the term $p(S | \psi)$ is now a mixture of multivariate Gaussian distributions for each pixel.

The data of Section IV were also analysed with the dependent source prior. In this case almost identical results were obtained, although MCMC convergence took considerably longer to achieve. This is due to the fact that the problem is still well-posed, in the sense that the number of images exceeds the number of sources to be separated, hence the result is not sensitive to the prior. It is expected that the prior will be more important when the number of sources exceeds the number of images. Another type of dependence is that a source is spatially uncorrelated. Spatial dependence is most conveniently modelled by a Gaussian Markov random field and we have done some preliminary work on this idea...
Combining with cross source correlations, one might ultimately consider a mixture of multivariate Gaussian Markov random fields as a prior for the sources. Implementing the analysis with such a prior would be a significant challenge computationally; we hypothesize that it will be difficult to derive a well-behaved MCMC approach. Other functional approximations, such as that of [25], offer feasible alternative to computing the posterior distribution in this case.

ACKNOWLEDGEMENTS

This work was started under the Network of Excellence MUSCLE, http://www.muscle-noe.org, contract number FP6-507752, funded by the European Union. Simon Wilson’s work is currently under the STATICA programme, funded by Science Foundation Ireland, contract number IN.1/I1879. Ercan Kuruoglu’s work is partly supported by the Italian Space Agency (ASI) under the contract Planck LFI Phase E2 activity.

REFERENCES


