Independent Component Analysis using Gaussian Mixture Models

By T. Becker and P. R. White

ISVR, University of Southampton, Highfield, Hants, SO17 1SX

Email ts.becker@gmx.net, prw@isvr.soton.ac.uk

Abstract

This paper discusses a method for performing independent component analysis exploiting Gaussian mixture models (GMMs). Previously most techniques that combine these methods have used GMMs to model the source signals. This paper considers a parsimonious method for modelling the observed signals. The GMM is fitted to the observed data using a modified version of the expectation maximisation algorithm.

1. Introduction

Independent Component Analysis (ICA) is one method of blind source separation. A set of observed signals is generated through a mixing process applied to the source signals. The goal of ICA algorithms is to reconstruct the source signals based only on the observed signals, i.e. without detailed knowledge of the mixing process. This is evidently an ill-posed problem and to render it tractable some additional information is required (e.g. Roberts 2001a). In this particular paper we shall concentrate on the linear, instantaneous mixing problem, and exploit an assumption of independence of the source signals, specifically we assume

$$\mathbf{X} = \mathbf{AS} \tag{1.1}$$

The $P \times N$ matrix of observed signals is denoted $\mathbf{X}$, where N is the number of samples in each of the channels of the observation and P is the number of observations. Similarly $\mathbf{S}$ is the $Q \times N$ matrix of source signals**,** where Q is number of sources. $\mathbf{A}$ is the $P \times Q$ mixing matrix. In this paper we specifically restrict our attention to the "square" case where $P=Q=D$.

Gaussian mixture models (GMMs), and other mixture models, have been applied to ICA in various guises (Choudrey, 2002; Miskin, 2000; Roberts 2001b; Todros 2004). ICA relies on the source signals being non-Gaussian; hence the ability of GMMs to model a wide range of probability density functions (pdfs) makes them an enticing tool with which to attempt ICA.

Previous work has primarily considered the use of GMMs to model the source signals, the parameters of this model, along with the unknown mixing matrix, are then optimised to best explain the observed data (e.g. Choudrey, 2002; Miskin, 2000; Roberts 2001b). Modelling the sources using a GMM is simplified because of their assumed independence, which means that the GMM can be constructed without the need to accommodate cross-correlations between sources. An alternative philosophy, which is closer to the methodology described herein, was proposed in (Todros 2004). In this approach the GMM is fitted to the observed data, rather than the sources. This ICA algorithm proceeds by attempting to find the rotation matrix that approximately diagonalises all of the correlation matrices resulting from the GMM. Such a model, as we shall demonstrate, lacks parsimony, since it fails to capture the independence of the sources and relies on the approximate joint diagonalisation to impose this *a posteriori*.

This paper considers the construction of an ICA algorithm in which a GMM is applied to the observed signals using a parsominious model. Section 2 describes the algorithm and the principles underlying it, section 3 presents some results of applying the algorithm to synthetic data and conclusions are drawn in section 4.

2. Algorithmic details

Throughout the following we shall assume that the observed signals have been pre-processed using Principle Component Analysis (PCA), i.e. the data has been decorrelated, a process sometimes refered to as sphering, which is common practice in ICA (Robert 2001a). The solution to the ICA problem is obtained by identifying a rotation (unitary) matrix which renders the estimated source signals independent, as measured via some metric.

The observed signals have a particular structure as a result of the generative model (1.1) and the assumed independence of the sources. In particular one can show (Becker 2006) that a GMM of the (independent) source signals should satify the following properties.

1) The centres (means) in the GMM lie on a rectangular grid.

2) The correlation matrices of the components in the GMM are diagonal.

These properties translate to equivalent properties of the observed signals. The first property means that the centres of a GMM modelling the observations should lie on a rotated grid. The second property ensures that the GMM of the observations should consist of components whose correlation matrices share common eigenvectors, i.e. have the form

$$C_k = U \Lambda_k U^t \tag{2.1}$$

where Λ_k is the diagonal eigenvalue matrix for the k^{th} component in the GMM, U is the matrix of eigenvectors and ^{t} denotes transposition.

Property 1) can be used to generate an efficient method for ICA, but the method is only effective when the components of the GMMs are well separated. This is a significant limitation which makes the algorithm only suitable for a very limited range of applications.

This paper concentrates on methods that explicitly exploit property 2). The corresponding GMM leads to a model for the pdf of the observations, $p(\mathbf{x})$, of the form

$$p(\mathbf{x}) = \sum_{k=1}^{K} p_k \frac{1}{\sqrt{(2\pi)^D |C_k|}} e^{-(\mathbf{x}-\mu_k)^t C_k^{-1} (\mathbf{x}-\mu_k)/2} = \sum_{k=1}^{K} p_k \frac{1}{\sqrt{(2\pi)^D \prod_{m=1}^{D} \lambda_m}} e^{-(\mathbf{x}-\mu_k)^t U^t \Lambda_k^{-1} U (\mathbf{x}-\mu_k)/2} \tag{2.2}$$

where p_k are the label probabilities, μ_k are the vectors of means (centres) for the GMM and K is the number of components in the GMM. We refer to this model as an aligned GMM.

The matrix of eigenvectors represents the rotation matrix that is required to perform ICA, so by fitting an align GMM one essentially solves the ICA problem. The aligned GMM can be fitted to the observed data using a modification of the familiar Expectation Maximisation (EM) algorithm (Bilmes, 1998).

The modified EM algorithm is implemented by iterating the following equations over i:

$$p_k^{i+1} = \frac{1}{N} \sum_{n=1}^{N} p\left(k \mid \mathbf{x}_n, \Theta^i\right) \tag{2.3a}$$

$$\mathbf{m}_k^{i+1} = \frac{1}{p_k^{i+1}} \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n p\left(k \mid \mathbf{x}_n, \Theta^i\right) \tag{2.3b}$$

$$\begin{aligned} &U^{i+1}: \quad \text{solution to} \quad \sum_{k=1}^{K} \left[U^{i+1} \left(\Lambda_k^i\right)^{-1} \left(U^{i+1}\right)^t, \mathbf{K}_k^{i,i+1} \right] = 0 \\ &\mathbf{K}_k^{i,i+1} = \sum_{n=1}^{N} p\left(k \mid \mathbf{x}_n, \Theta^i\right) \left(\mathbf{x}_n - \mathbf{m}_k^{i+1}\right) \left(\mathbf{x}_n - \mathbf{m}_k^{i+1}\right)^t \end{aligned} \tag{2.3c}$$

$$\Lambda_k^{i+1} = \frac{1}{Np_k^{i+1}} diag\left\{\left(U^{i+1}\right)^t \mathbf{K}_k^{i,i+1} U^{i+1}\right\} \tag{2.3d}$$

where [,] denotes the Lie bracket, Θ^i is the set of GMM parameters and $p\left(k \mid \mathbf{x}_n, \Theta^i\right)$ denotes the probability of measured signals being generated by the k^{th} component of the GMM with parameters Θ^i.

The algorithm described by (2.3) shares much in common with the classical EM algorithm for GMMs (Bilmes, 1998). The primary difference is that the correlation matrices are updated by updating a common eigenvector matrix and individual eigenvalues. The update for the matrix U, equation (2.3c), needs to take into account the fact that this matrix must be orthogonal. In fact we are seeking solutions on the Stiefel manifold. In general, the implementation of (2.3) is not trivial, in particular, the solution of (2.3c) requires some care. Equation (2.3c) derives from the maximisation step of the EM algorithm and one has to ensure that the solution found does indeed represent a maximum, as opposed to a minimum. One approach to approximating the solution (2.3c) is update U as the eigenvector matrix associated with the matrix

$$\mathbf{K}^{\%} = \sum_{k=1}^{K} \mathbf{K}_k^{i,i+1} \tag{2.4}$$

In the case where all of the component matrices $\mathbf{K}_k^{i,i+1}$ share a common set of eigenvectors, which in turn will the eigenvectors of $\mathbf{K}^{\%}$, then that eigenvector matrix provides an exact solution to (2.3c). Adoption of this approximation robs the EM algorithm of the useful property that the likelihood is guaranteed to increase monotonically.

In the special case of 2 sources and 2 measurements it is straightforward to obtain an exact solution to (2.3c), see (Becker 2006) for further details. For this reason we shall concentrate on such problems in the remainder of this paper.

3. Results

To demonstrate the performance of this algorithm we present results on a simulated data set. Comparisons are made to the GMM algorithm most closely related to ours, namely that of (Todros, 2004), which we shall refer to this as the full covariance GMM algorithm.

The simulated signal used to generate the following set of results is generated using a six component independent GMM of dimension 2. The data is rotated by an angle -9.445º. The result shown depicts histograms of the rotation computed by ICA algorithms to render the data independent, i.e. +9.445º. The algorithm is run on 2000 realisations of the data, the length of each time series is 2000 samples. The angles computed by the ICA algorithm have been recorded and then histograms are computed. This is repeated for various GMM model orders (K). These results are shown in Figure 1.

From Figure 1 one can see that the aligned algorithm offers significantly more consistent results than the full covariance scheme. A further observation one can make is that the performance of the aligned GMM algorithm is significantly less sensitive to the selection of the number of Gaussians in the GMM. It is evident from Figure 1a) that the performance of full covariance algorithm is best when the number of Gaussians matches that of the synthesised signal, specifically six. There is little evidence from Figure 1 that similar factors affect the aligned algorithm.

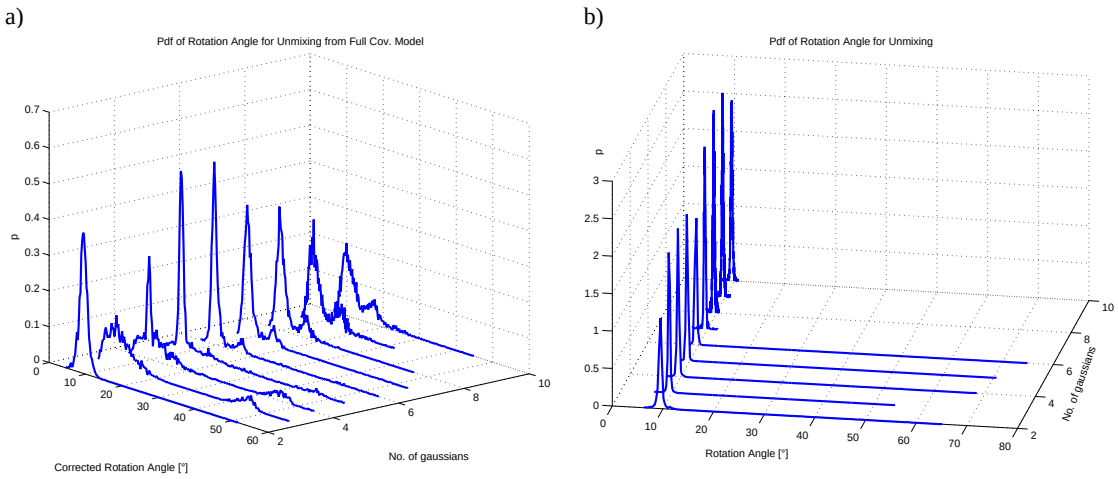

FIGURE 1 Histograms of estimated angles. a) Full covariance GMM b) Aligned GMM

4. Conclusions

Aligned GMMs provide a potentially powerful method for performing ICA. They show several distinct advantages over the full covariance method. Their performance is generally superior to that of the full covariance method, they are more robust to the selection of model order for the GMM.

REFERENCES

BECKER T. 2006. *Independent component analysis with Gaussian mixture models*. MSc Thesis, University of Southampton.

BILMES J.A. 1998. *A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models.* Technical Report TR-97-021, International Computer Science Institute, Berkeley, California.

CHOUDREY R. 2002. *Variational Methods for Bayesian Independent Component Analysis.* PhD Thesis, University of Oxford, Sommerville College.

MISKIN, J.W. 2000. *Ensemble learning for independent component analysis*, PhD Thesis, University of Cambridge.

MOULINES E., CARDOSO J. & GASSIAT E. 1997. *Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models.* Proc. ICASSP 1997, Munich, Germany, pages 3617–3620.

ROBERTS S. & EVERSON R. 2001. *Independent Component Analysis, Principles and Practice*. Cambridge University Press, Cambridge.

ROBERTS S. & PENNY W. 2001. *Mixtures of independent component analyzers.* Proc. of ICANN 2001, Vienna.

TODROS K. & TABRIKIAN J. 2004. Application of Gaussian Mixture Models for Blind Separation of Independent Sources. *Proc. of the 5th Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA2004)*, Granada, Spain, 382–389.