BLIND DETECTION OF INDEPENDENT DYNAMIC COMPONENTS

Lars Kai Hansen, Jan Larsen and Thomas Kolenda

Technical University of Denmark
Department for Mathematical Modelling
Richard Petersens Plads, Bldg. 321
2800 Kgs. Lyngby, Denmark
Emails: lkh,jl,thko@imm.dtu.dk

ABSTRACT

In certain applications of independent component analysis (ICA) it any of interest to test hypotheses concerning the number of components or simply to test whether a given number of components is significant relative to a "white noise" null hypothesis. We estimate probabilities of such competing hypotheses for ICA based on dynamic decorrelation. The probabilities are evaluated in the so-called Bayesian information criterion approximation, however, they are able to detect the content of dynamic components as efficient as an unbiased test set estimator.
Keywords: Blind Source Separation (BSS), Dynamic Components, Independent Component Analysis (ICA), BIC detection

1. INTRODUCTION

Blind separation of linear mixtures is an extremely active research area [1, 2]. Despite the obvious relevance in many applications, remarkably little effort has been devoted to blind *signal detection*. In medical applications, e.g., there is often an interest in quantifying the statistical significance of independent component representations. In this contribution we develop a scheme for testing competing hypotheses about the content of independent dynamic components in a multi-channel signal. We use an approximate Bayesian framework for computing relative probabilities of the relevant hypotheses, hence, obtain control of both type I and type II errors.

Independent component analysis (ICA) is typically based on non-Gaussianity [3, 4] or temporal correlations [5, 6]. Attias and Schreiner proposed a very rich ICA framework based on higher order statistics and decorrelation [7, 8], allowing for completely general and learnable source distributions, however at the price of significant computation [9].

Molgedey and Schuster proposed an approach based on dynamic decorrelation which can be used if the independent source signals have different autocorrelation functions [5, 10, 11]. The main advantage of this approach is that the solution is simple and constructive, and can be implemented in a fashion that requires minimal user intervention (parameter tuning). In [11] we applied the Molgedey-Schuster algorithm to image mixtures and proposed a symmetrized version of the algorithm that relieves a problem of the original approach, namely that it occasionally produces complex mixing coefficients and source signals.

This work is funded by the Danish Research Councils through the THOR Center for Neuroinformatics and the Center for Multimedia.

2. PROBABILISTIC MODELING

We start by reviewing Bayesian estimation of probabilities over sets of hypotheses. Let such a set of hypotheses (models) be indexed by $m = 0, ..., M$ (we use $m = 0$ to signify a null-hypothesis, corresponding to no non-trivial independent components in the data). The probability of a specific model given the observed data X is denoted by $P(m|X)$, using Bayes' relation this can be written as,

$$P(m|X) = \frac{P(X|m)P(m)}{P(X)}. \quad (1)$$

The prior probability $P(m)$ reflects our prior beliefs in the specific model in relation to the other models in the set, if no specific belief is relevant we will use a uniform distribution over the set.

A model will typically be defined in terms of a set of parameters θ so that we have a so-called generative model density $P(X|\theta, m)$, this density is often given by the observation model. We then have the relation

$$P(X|m) = \int d\theta P(X, \theta|m) = \int d\theta P(X|\theta, m)P(\theta|m). \quad (2)$$

The $P(\theta|m)$ distribution carries possible prior beliefs on the level of parameters, often we will assume so-called vague priors that have no or little influence on the above integral, except making it finite in the case X is empty (i.e., $P(\theta|m)$ is normalizable).

The integral in equation (2) is often too complicated to be evaluated analytically. Various approximation schemes have been suggested, here we will use the Bayesian Information Criterion (BIC) approximation [12]. This approximates the integral by a Gaussian in the vicinity of the parameters that maximize the integrant (the so-called maximum posterior parameters θ^*). With this approximation the integral becomes

$$P(X|m) \approx P(X|\theta^*, m)P(\theta^*, m)T^{-d/2}, \quad (3)$$

where d is the dimension of the parameter vector and T is the number of training cases. Observe that high-dimensional models (large d) are exponentially penalized, hence, can only be accepted if they provide highly likely descriptions of data.

3. DYNAMIC COMPONENT LIKELIHOOD FUNCTION

Let the multi-channel signal be represented as a *data matrix* X with a time row index,

$$X = (X)_{l,t} = \sum_{k=1}^{K} A_{l,k} S_{l,t} = A \cdot S, \tag{4}$$

where $l = 1, ..., L$ represent measurements (e.g., microphones) $t = 1, ..., T$, are the sampling time points, and A is the $L \times K$ real mixing matrix. The dynamic components S are assumed to be given by unknown independent, unit variance, white noise signals, $U_{k,t}$, filtered by the unknown, and source specific filters h_k,

$$S_{k,t} = \sum_{\tau=0}^{N_k} h_{k,\tau} U_{k,t-\tau} \tag{5}$$

This leads to the following model

$$P(X|A, K) = \int dS \delta(X - AS) P(S). \tag{6}$$

The source distribution is given by,

$$\begin{aligned} P(S) &= \prod_k \frac{1}{\sqrt{|2\pi\Sigma_k|}} \\ &\exp\left(-\frac{1}{2}\sum_{t,t'} S_{k,t}(\Sigma_k^{-1})_{t,t'} S_{k,t'}\right), \end{aligned} \tag{7}$$

where the source covariance matrix is given by

$$\Sigma_{k;t,t'} = \sum_{\tau} h_{k;\tau} h_{k,t'-t+\tau}. \tag{8}$$

These matrices are Toeplitz under the model assumptions. Evaluating the integral in equation (6) provides the expression

$$\begin{aligned} P(X|A, m) &= \prod_k \frac{1}{\sqrt{|2\pi\Sigma_k|}} \left(\frac{1}{\|A\|}\right)^T \\ &\exp\left(-\frac{1}{2}\sum_{t,t'} \widehat{S}_{k,t}(\Sigma_k^{-1})_{t,t'} \widehat{S}_{k,t'}\right). \end{aligned} \tag{9}$$

with $\|A\|$ being the absolute value of the determinant of A, while we use the notation $\widehat{S}_{k,t}(X)$, for the sources estimated from A, X $\widehat{S}_{k,t} = \sum_l (A^{-1})_{k,l} X_{l,t}$.

4. MOLGEDEY SCHUSTER SEPARATION

Let X_τ be the time shifted data matrix. The delayed correlation approach is based on solving the simultaneous eigenvalue problem for the correlation matrices $X_\tau X^\top$ and $XX^\top$, see [11] for a more detailed derivation. This is implemented by solving the eigenvalue problem for the *quotient* matrix $Q \equiv X_\tau X^\top \left(XX^\top\right)^{-1}$. From equation (4) we have

$$XX^\top = ASS^\top A^\top, \qquad X_\tau X^\top = AS_\tau S^\top A^\top \tag{10}$$

If the sources furthermore are independent, we obtain in the limit $\lim_{N\to\infty} N^{-1} SS^\top = C(0)$, the diagonal source crosscorrelation matrix at lag zero. Similarly, $\lim_{N\to\infty} N^{-1} S_\tau S^\top = C(\tau)$ produces the diagonal crosscorrelation matrix at lag τ. Hence, to zero'th order in $1/N$,

$$\begin{aligned} X_\tau X^\top \left(XX^\top\right)^{-1} &\approx AC(\tau)A^\top (A^\top)^{-1} C(0)^{-1} A^{-1} \\ &= AC(\tau)C(0)^{-1} A^{-1} \end{aligned} \tag{11}$$

with $C(\tau)C(0)^{-1}$ being a diagonal matrix. If we solve the eigenvalue problem for the quotient matrix $Q \equiv X_\tau X^\top \left(XX^\top\right)^{-1}$ we have a direct scheme for estimating A, S. Let

$$Q\Phi = \Phi\Lambda, \tag{12}$$

and identify $\Phi = A$ and $\Lambda = C(\tau)C(0)^{-1}$ up to scaling factors.

It is straight forward to generalize the constructive scheme to the under-determined case (more microphones L than sources K). We use a subspace projection scheme based on the SVD of the data matrices. First we note $XX^\top \approx X_\tau X_\tau^\top$, hence,

$$X = UDV^\top, \qquad X_\tau = UDV_\tau^\top. \tag{13}$$

In other words, the correlation matrices become,

$$XX^\top = UDV^\top VD^\top U^\top, \qquad X_\tau X^\top = UDV_\tau^\top VD^\top U^\top. \tag{14}$$

The appropriate subspace (Moore-Penrose) inversion of $XX^\top$, is then

$$\left(XX^\top\right)^{-1} = UD^{-1}(V^\top V)^{-1} D^{-1} U^\top = UD^{-2}U^\top, \tag{15}$$

and we find that the quotient matrix for the eigenvalue problem becomes[1],

$$X_\tau X^\top \left(XX^\top\right)^{-1} = UDV_\tau^\top VDU^\top UD^{-2}U^\top \tag{16}$$

$$= UDV_\tau V^\top D^{-1} U^\top. \tag{17}$$

We have an option here for *regularization* by reducing the number of sources, i.e., by reducing the dimension of the SVD representation.

5. EXPERIMENTAL EVALUATION

We have established a simple synthetic data experiment to evaluate the proposed Bayesian hypothesis testing framework. Three sinusoidal source signals of unit variance and different periods are mixed and projected into $L = 10$ dimensions, and white noise added to all channels. A set of $M = 8$ hypotheses are evaluated with $m = 0$ representing a model without dynamic components, and $m = 1 - 7$ representing $K = 1 - 7$ dynamic components.

The number of component was controlled by an initial SVD projection from the original ten dimensional measurements to K dimensions.

In figure 1 we show the recovery of the source signals in the largest model $K = 7$ using the Molgedey-Schuster scheme.

The component autocorrelation functions were estimated from the reconstructed source signals, forming the Toeplitz source covariance matrix. Note that both the determinants and the inverse matrices can be computed in $\sim T^2$ operations.

In figure 2 we show the value of a "cost function" defined as the average negative log likelihood $-\log P(X|m)$, estimated on a test set, by Akaike's information criterion (AIC) [13], and by

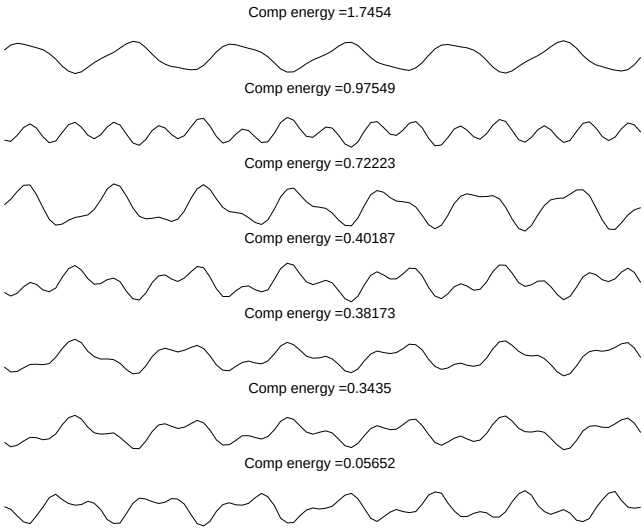

Fig. 1. Decomposition of a synthetic data set comprising three spatially non-orthogonal independent (periodic) components. These component were embedded in $L = 10$ dimensions and degraded by additive white noise. We show the result of projecting data onto a seven dimensional subspace before doing the independent component analysis, hence forcing seven independent components. The components are found by the Molgedey-Schuster decorrelation method. The components are ordered according to variance of the reconstructed signal. The probabilistic analysis reject this hypothesis in favor of the true hypothesis of $K = 3$ component, with a relative probability of 1/1000 as seen in figure 3.

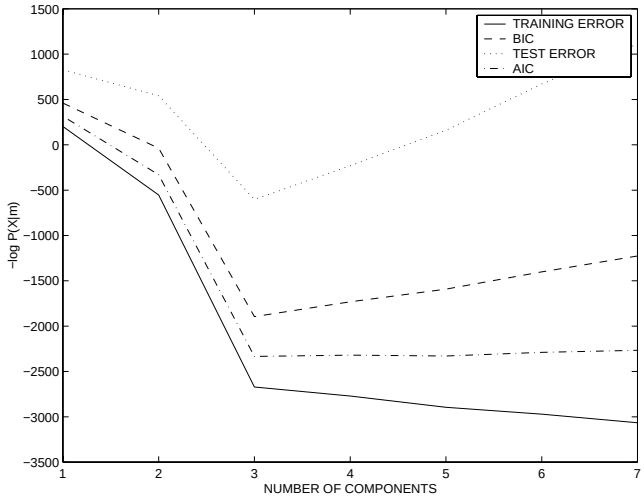

Fig. 2. Performance indices for the experiment described in figure 1. In this case both the empirical test error (using an additional test set for evaluation) and the AIC and BIC agree and point to the true $K = 3$ model. The error is computed as the negative log-probability.

the Bayesian Information Criterion approximation. In figure 3 we show the BIC approximation probabilities over the set of hypotheses evaluated, see equation (3). We next repeated the experiment

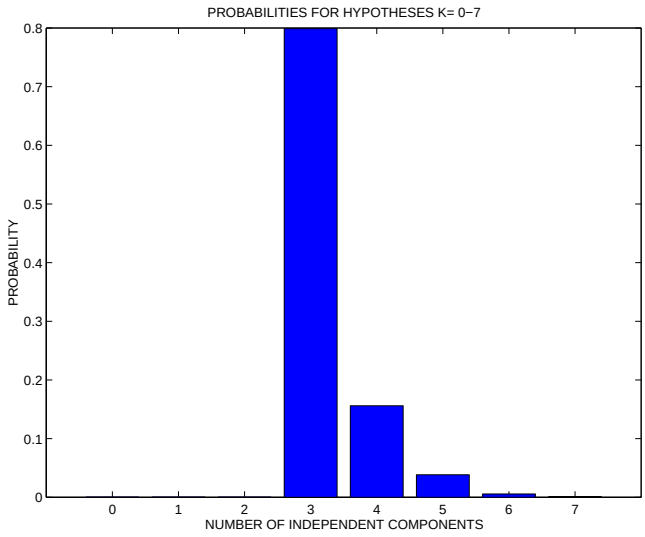

Fig. 3. Probabilities for the set of hypotheses formed by increasing the number of independent components from $K = 0$ (null-hypothesis) to $K = 7$ on the data set described in figure 1. The probability of the "null hypothesis" is $P(0) < 10^{-10}$ while the probability of the $K = 7$ hypothesis is $P(7) \sim 0.001$

100 times counting how often each of the methods succeeded in picking up the correct model. Figures 4-5 summarize the results, and show that the probabilistic approach is very efficient for picking the correct hypothesis.

6. CONCLUSION

We have formulated a probabilistic analysis of ICA that allows the evaluation probabilities across a set of competing hypotheses. The probabilities showed very efficient in selecting the number of independent components in a small simulation study.

7. ACKNOWLEDGEMENT

LKH wishes to thank Terry Sejnowski for hosting a visit to the Salk Institute, July 2000, where this work was initiated.

8. REFERENCES

[1] T.-W. Lee, *Independent Component Analysis: Theory and Applications*, Kluwer Academic Publishers, New York, 1998.

[2] M. Girolami (Ed.), *Advances in Independent Component Analysis.*, Springer-Verlag, New York, 2000.

[3] P. Comon, "Independent component analysis: A new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.

[4] A. Bell and T.J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[1] Note that V is quadratic as $L = T$.

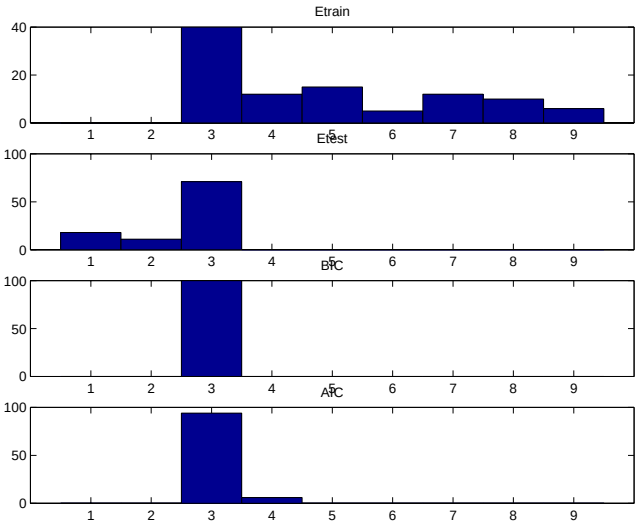

Fig. 4. The synthetic data experiment involving three components mixed in ten dimensions was repeated 100 times. The 10×3 mixing matrices where chosen at random with i.i.d. unit variance matrix elements. The noise variance was set to $\sigma^2 = 0.01$. For each instance we let the training error, the test error, the BIC and the AIC pick the best fitting model. We show the histograms of occurring choices. The empirical test error on a test set is at times too conservative and selects models with only one or two components. It is surprising that the training error is not minimal for the larger models with $K = 7$, this may be related to the fact that the Molgedey Schuster algorithm is only approximately maximum likelihood.

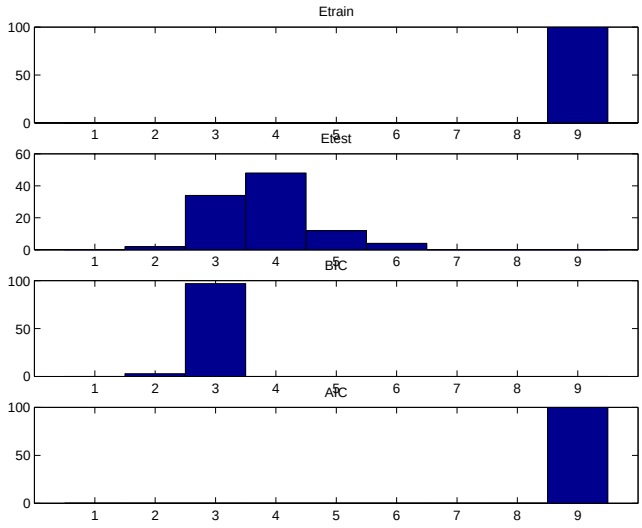

Fig. 5. The synthetic data experiment involving three components mixed in ten dimensions was repeated 100 times. The 10×3 mixing matrices where chosen at random with i.i.d. unit variance matrix elements. The noise variance was set to $\sigma^2 = 1.0$. For each instance we let the training error, the test error, the BIC and the AIC pick the best fitting model. We show the histograms of occurring choices. Note, that the overfit is quite dramatic at these high noise levels. The training error is optimistic as is the AIC criterion.

[5] L. Molgedey and H. Schuster, "Separation of independent signals using time-delayed correlations," *Physical Review Letters*, vol. 72, no. 23, pp. 3634–3637, 1994.

[6] J-F. Cardoso A. Belouchrani, K. Abed-Meraim and E. Moulines, "Blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[7] H. Attias and C.E. Schreiner, "Blind source separation and deconvolution by dynamic component analysis," *Neural Networks for Signal Processing VII: Proceedings of the 1997 IEEE Workshop, 456-465 (1997*, pp. 456–465, 1997.

[8] H. Attias and C.E. Schreiner, "Dynamic component analysis," *Neural Computation*, vol. 10, pp. 1373–1424, 1998.

[9] K.S. Petersen, L.K. Hansen, T. Kolenda, E. Rostrup, and S. Strother, "On the independent components in functional neuroimages," *In proceedings of ICA-2000, Finland, June 2000*, 2000.

[10] L.K. Hansen and J. Larsen, "Source separation in short image sequences using delayed correlation," *Proceedings of the IEEE Nordic Signal Processing Symposium, Vigsø, Denmark 1998. Eds. P. Dalsgaard and S.H. Jensen*, pp. 253–256, 1998.

[11] J. Larsen L.K. Hansen and T. Kolenda, "On independent component analysis for multimedia signals," *L. Guan, S.Y. Kung and J. Larsen (eds.) Multimedia Image and Video Processing, CRC Press*, vol. Chapter 7, pp. 175–200, 2000.

[12] D.J.C. MacKay, "Bayesian model comparison and backprop nets," *Proceedings of Neural Information Processing Systems 4*, pp. 839–846, 1992.

[13] H. Akaike, "Statistical predictor identification," *Annals. Inst. Stat. Math.*, vol. 22, pp. 203, 1970.