Model structure selection in convolutive mixtures

Mads Dyrholm[1], Scott Makeig[2] and Lars Kai Hansen[1]

[1]Informatics and Mathematical Modelling
Technical University of Denmark, 2800 Lyngby, Denmark
mad,lkh@imm.dtu.dk
[2]Swartz Center for Computational Neuroscience
University of California, San Diego 0961, La Jolla CA 92093-0961
scott@sccn.ucsd.edu

Abstract. The CICAAR algorithm (convolutive independent component analysis with an auto-regressive inverse model) allows separation of white (i.i.d) source signals from convolutive mixtures. We introduce a source color model as a simple extension to the CICAAR which allows for a more parsimoneous representation in many practical mixtures. The new filter-CICAAR allows Bayesian model selection and can help answer questions like: 'Are we actually dealing with a convolutive mixture?'. We try to answer this question for EEG data.

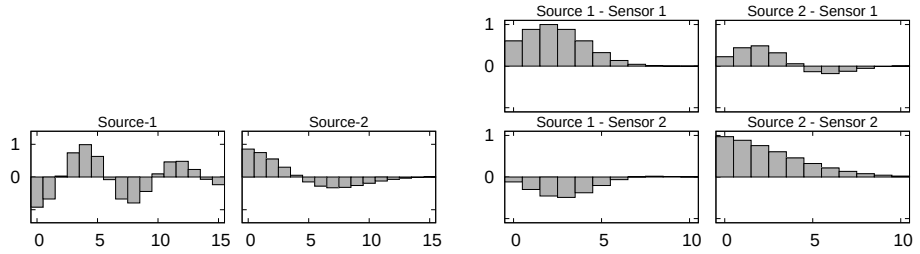
1 Introduction

Convolutive ICA (CICA) is a topic of high current interest and several schemes are now available for recovering mixing matrices and sources signals from convolutive mixtures, see e.g., [4]. Convolutive models are more complex than conventional instantaneous models, hence, the issue of model optimization is important. Convolutive ICA in its basic form concerns reconstruction of the L+1 mixing matrices $\mathbf{A}_\tau$ and the N source signal vectors $\mathbf{s}_t$ of dimension K, from a D-dimensional convolutive mixture

$$\mathbf{x}_t = \sum_{\tau=0}^{L} \mathbf{A}_\tau \mathbf{s}_{t-\tau} \tag{1}$$

Here we focus, for simplicity, on the case where the number of sources equals the number of sensors, $D = K$.

We have earlier proposed the CICAAR approach for convolutive ICA [3] as a generalization of Infomax [2] to convolutive mixtures. The CICAAR exploits the relatively simple structure of the un-mixing system resulting when the inverse mixing is represented as an autoregressive process. In the original derivation we were forced to assume white (i.d.d) sources, i.e., that all temporal correlation in the mixture signals appeared through the convolutive mixing process. A more economic representation is obtained, however, if we explicitly introduce filters to represent possible auto-correlation of sources. This added degree of freedom also



(a) Source color filters, $M = 15$. (b) Convolutive mixing system, $L = 10$.

Fig. 1. Filters for generating synthetic data. First, two i.i.d. signals are colored through their respective filters (a). Then, the colored signals are convolutively mixed using a distinct filter for each source-sensor path (b).

carries another benefit, it allows for optimizing the model structure: How much correlation should be accounted for by the source filters, and how much should be accounted for by the convolutive mixture? Explicit source auto-correlation modeling using filtered white noise has been proposed earlier by several authors, see e.g., [1, 7, 8].

2 Modelling convolutive ICA with auto-correlated sources

We introduce a model for each of the sources

$$s_k(t) = \sum_{\lambda=0}^{M} h_k(\lambda) z_k(t-\lambda) \tag{2}$$

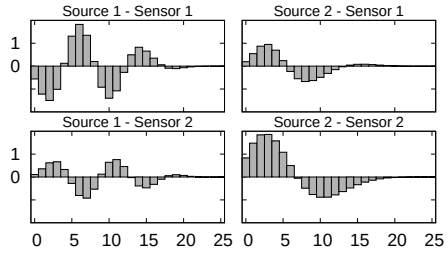
where $z_k(t)$ represents a whitened version of the source signal. The negative log likelihood for the model combining (1) and (2) is given by

$$\mathcal{L} = N \log |\det \mathbf{A}_0| + N \sum_k \log |h_k(0)| - \sum_{t=1}^{N} \log \mathrm{p}(\hat{\mathbf{z}}_t) \tag{3}$$

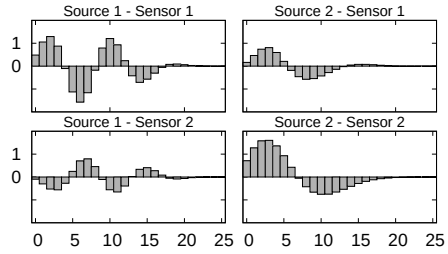
where $\hat{\mathbf{z}}_t$ is a vector of whitened source signal estimates at time t using an operator that represents the inverse of (2). We can without loss of generality set $h_k(0) = 1$, then

$$\mathcal{L} = N \log |\det \mathbf{A}_0| - \sum_{t=1}^{N} \log \mathrm{p}(\hat{\mathbf{z}}_t) \tag{4}$$

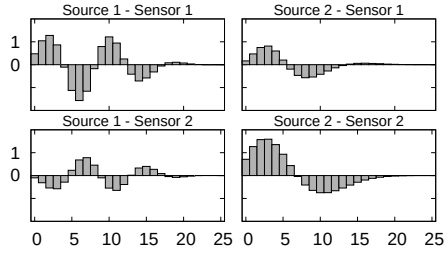
The number of parameters in this model is $D^2(L+1) + DM$, and it can thus be minimized if M is increased so as to explain the source auto-correlations allowing L to be reduced in return. An algorithm for convolutive ICA which includes the source model can be derived by making a relative straight forward modification to the equations of the CICAAR algorithm found in [3], see appendix A.



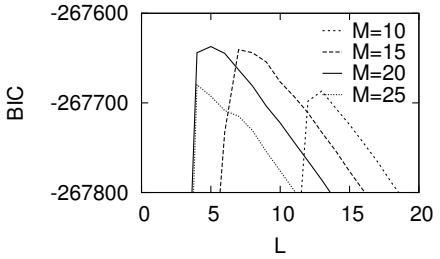
(a) Generating model $(L, M) = (10, 15)$



(b) Estimated by algorithm $(L, M) = (10, 15)$



(c) Estimated by algorithm $(L, M) = (5, 20)$



(d) BIC optimal is $(L, M) = (5, 20)$.

Fig. 2. Mixing filters convolved with respective color filters. (a) for the generating model. (b) for an estimated model with the 'true' L and M. (c) for the Bayes optimal model with $(L, M) = (5, 20)$. (d) shows the BIC for various models, and (L,M)=(5,20) is found optimal.

3 Model Selection Protocol

Let $\mathcal{M}$ represent a specific choice of model structure (L, M). The Bayes Information Criterion (BIC) is given by $\log \mathrm{p}(\mathcal{M}|\mathbf{X}) \approx \log \mathrm{p}(\mathbf{X}|\boldsymbol{\theta}_0, \mathcal{M}) - \frac{\dim \boldsymbol{\theta}}{2} \log N$ where $\dim \boldsymbol{\theta}$ is the number of parameters in the model, and $\boldsymbol{\theta}_0$ are the maximum likelihood parameters [9].

We propose a simple protocol for the dimensions (L, M) of the convolutional- and source-filters. First, expand the convolution length L without a source model (i.e. keeping $M = 0$). This will model the total temporal dependency structure of the system. The optimal L, denote it $L_{\max}$, is found by monitoring BIC. Next, expand the dimensions M of the source model filters while keeping the temporal dependency constant, i.e. keeping $(L + M) = L_{\max}$.

3.1 Simulation example

The first experiment is designed to illustrate the protocol for determining the dimensions of the convolution and the source filters. We create a 2×2 system with known source filters $M = 15$ and known convolution $L = 10 \ldots$

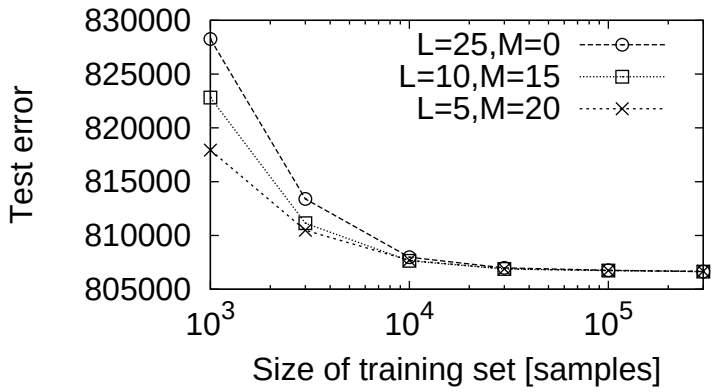


Fig. 3. Learning curves for three models: The generating model $(L, M) = (10, 15)$, a model with $(L, M) = (25, 0)$ which is more complex but fully capable of 'imitating' the first model, and the model $(L, M) = (5, 20)$ which was found Bayes optimal according to BIC. The generalization error is estimated as the likelihood of a test set ($N_{\text{test}} = 300000$). The uniform improvements in generalization of the 'optimal model' further underlines the importance of model selection in the context

Data — Two signals are generated by filtering temporally white signals using the filters shown on Figure-1(a). The signals are then mixed using the $2 \times 2 \times 10$ system shown on Figure-1(b). The generating model has thus $(L, M) = (10, 15)$.

Result — First we note, the model is in itself ambiguous; an arbitrary filter can be applied to a color filter if the inverse filter is applied to the respective column of mixing filters. Therefore, to compare results we inspect the system as a whole, i.e. source color convolved with a column of mixing filters.

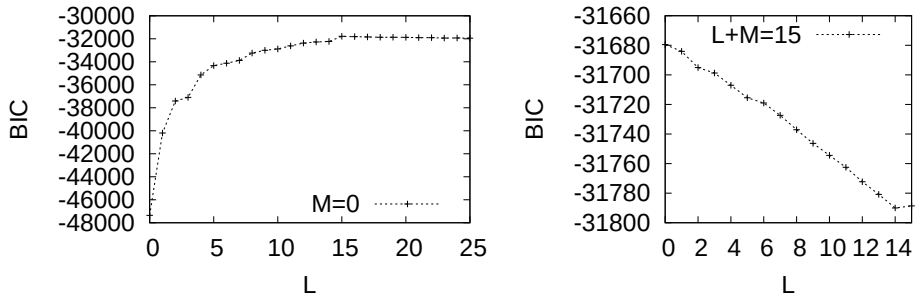
Figure-2 displays convolutive mixing systems where each mixing channel has been convolved with the respective color filter; (a) for the true generating model; (b) a run with the algorithm using $N = 300000$ training samples and using the (L, M) of the generating model. The result is perfect up to sign and scaling ICA ambiguities; (c) shows a run with the algorithm using $N = 100000$ and the Bayes optimal choice of $(L, M) = (5, 20)$ c.f. (d), in the finite data the protocol has found a parsimonious model with similar overall transfer function. We first study the learning curves, i.e., how does the training set dimension N, influence learning. We use the likelihood evaluated on a test set to measure the learning of different models. We now compare learning curves for three models; one which is the generating model $(L, M) = (10, 15)$, one $(L, M) = (25, 0)$ which is more complex but fully capable of imitating the first model, and $(L, M) = (5, 20)$ which is optimal according to BIC. Figure-3 shows learning curves of the three models, the test set is $N_{\text{test}} = 300000$ samples. The uniform improvements in generalization of the 'optimal model' further underlines the importance of model selection in the context of convolutive mixing.

3.2 Rejecting convolution in an instantaneous mixture

We will now illustrate the importance of the source color filters when dealing with the following fundamental question: 'Do we learn anything by using Convolutive ICA instead of instantaneous ICA?'—or put in another way: 'should L be larger than zero?'.

Data — To produce an instantaneous mixture we now mix the two colored sources from before using a random matrix.

Result — Figure-4(a) shows the result of using Bayesian model selection without allowing for a filter ($M = 0$). This corresponds to model selection in a conventional convolutive model. Since the signals are non-white L is detected and the model BIC simply increases as function of L up to the maximum which is attained at a value of $L = 15$. Next, in Figure-4(b) we fix $L + M = 15$. Models with a greater L have at least the same capability as a model with a lower L; but as expected lower L are preferable because the models has fewer parameters. Thus, thanks to the filters, we now get the correct answer: 'There is no evidence of convolutive ICA'.



(a) $M = 0$ (b) $M + L = 15$

Fig. 4. (a) the result of using Bayesian model selection without allowing for a filter ($M = 0$). Since the signals are non-white L is detected at a value of $L = 15$. (b) we fix $L + M = 15$, and now get the correct answer: $L = 0$ — 'There is no evidence of convolutive ICA'.

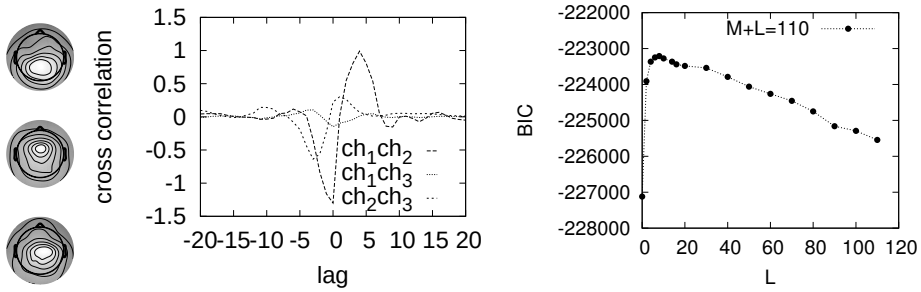
4 Is convolutive ICA relevant for EEG?

The EEG signals from the entire brain superimpose onto every EEG electrode instantaneously; there are no delays or echoes, hence, the mixing of the electromagnetic activity is definitely not a convolutive process. However, the question is whether the convolutive mixing model is relevant as a model for the brain

activity itself. It is well known that EEG activity exhibits rich spatio-temporal dynamics and that different tasks of the brain combine different regions in different frequency bands, and so, we expect the Bayes optimal model to potentially include some convolutive mixing $L > 0$.

Data — 20 minutes of a 71-channel human EEG recording downsampled to a 50-Hz sampling rate after filtering between 1 and 25 Hz with phase-indifferent FIR filters. First, the recorded (channels-by-times) data matrix ($\mathbf{X}$) was decomposed using extended infomax ICA [2, 5] into 71 maximally independent components whose ('activation') time series were contained in (components-by-times) matrix $\mathbf{S}^{\mathrm{ICA}}$ and whose ('scalp map') projections to the sensors were specified in (channels-by-components) mixing matrix $\mathbf{A}^{\mathrm{ICA}}$, assuming instantaneous linear mixing $\mathbf{X} = \mathbf{A}^{\mathrm{ICA}}\mathbf{S}^{\mathrm{ICA}}$. Three of the resulting independent components were selected for further analysis on the basis of event-related coherence results that showed a transient partial collapse of component independence following the subject button presses [6]. Their scalp maps (the relevant three columns of $\mathbf{A}^{\mathrm{ICA}}$) are shown on Figure 5(a).

Convolutive ICA analysis — Next, convolutive ICA decomposition was applied to the three component activation time series (relevant three rows of $\mathbf{S}^{\mathrm{ICA}}$) which we shall refer to as channels ch_1, ch_2 and ch_3. Following our proposed protocol, we find $L_{\max} = 110$, then $L = 9$ as shown on Figure-5(c) — so, we are in fact dealing with a convolutive mixture. Figure-5(b) shows, for one of the resulting convolutive ICA components, cross correlation functions between its contribution to the channels (with each a scalp map associated). Clearly, there are delayed correlation between the different brain regions, and this is not possible to model with an instantaneous ICA model, hence the need for convolutive mixing.



(a) (b) (c) We find $L = 9$

Fig. 5. (a) Scalp maps for the three ICA components. (b) for one of the resulting convolutive ICA components, cross correlation functions between its contribution to the channels. (c) Finding $L = 9$ for the EEG data.

5 Conclusion

We have incorporated filters for modelling possible source auto-correlations into an existing algorithm for convolutive ICA. We have proposed a protocol for determining the dimension L of a convolutive mixture utilizing the filters. We have shown that convolutive ICA is relevant for real EEG data.

Appendix A: Source modeling with the CICAAR algorithm

For notational convenience we introduce the following matrix notation instead of (2), handling all sources in one matrix equation

$$\mathbf{s}_t = \sum_{\lambda=0}^{M} \mathbf{H}_\lambda \mathbf{z}_{t-\lambda} \tag{5}$$

where the $\mathbf{H}_\lambda$'s are diagonal matrices defined by $(\mathbf{H}_\lambda)_{ii} = h_i(\lambda)$.

Given a current estimate of the mixing matrices $\mathbf{A}_\tau$ and the source filter coefficients $h_k(\lambda)$, First apply equation 7 of [3] to obtain $\hat{\mathbf{s}}_t$. Then apply the inverse source coloring operator

$$\hat{\mathbf{z}}_t = \hat{\mathbf{s}}_t - \sum_{\lambda=1}^{M} \mathbf{H}_\lambda \hat{\mathbf{z}}_{t-\lambda} \tag{6}$$

which must replace $\hat{\mathbf{s}}_t$ in [3] (in equations 6,8,9 and 11). This involves the following partial derivatives which in turn uses the result from [3] (from equations 7,10,12)

$$\frac{\partial(\hat{\mathbf{z}}_t)_k}{\partial(\mathbf{B}_\tau)_{ij}} = \frac{\partial(\hat{\mathbf{s}}_t)_k}{\partial(\mathbf{B}_\tau)_{ij}} - \sum_{\lambda=1}^{M} \mathbf{H}_\lambda \frac{\partial(\hat{\mathbf{z}}_{t-\lambda})_k}{\partial(\mathbf{B}_\tau)_{ij}} \tag{7}$$

where $\mathbf{B}_\tau = \mathbf{A}_\tau$ for $\tau > 0$ and $\mathbf{B}_0 = \mathbf{A}_0^{-1}$. Furthermore

$$\frac{\partial(\hat{\mathbf{z}}_t)_k}{\partial(\mathbf{H}_\lambda)_{ii}} = -\delta(k-i)(\hat{\mathbf{z}}_{t-\lambda})_i - \left(\sum_{\lambda'=1}^{M} \mathbf{H}_{\lambda'} \frac{\partial \hat{\mathbf{z}}_{t-\lambda'}}{\partial(\mathbf{H}_\lambda)_{ii}} \right)_k \tag{8}$$

The work involved in this plug-in is minimal due to the diagonal structure of the $\mathbf{H}_\lambda$ matrices. Finally,

$$\frac{\partial \mathcal{L}}{\partial(\mathbf{H}_\lambda)_{ii}} = -\sum_{t=1}^{N} \boldsymbol{\psi}_t^T \frac{\partial \hat{\mathbf{z}}_t}{\partial(\mathbf{H}_\lambda)_{ii}} \tag{9}$$

where $(\boldsymbol{\psi}_t)_k = p'((\hat{z}_t)_k)/p((\hat{z}_t)_k)$.

References

1. Hagai Attias and C. E. Schreiner. Blind source separation and deconvolution: the dynamic component analysis algorithm. *Neural Computation*, 10(6):1373–1424, 1998.
2. Tony Bell and Terrence Sejnowski. An information maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.
3. M. Dyrholm and L. K. Hansen. CICAAR: Convolutive ICA with an auto-regressive inverse model. In Carlos G. Puntonet and Alberto Prieto, editors, *Independent Component Analysis and Blind Signal Separation*, volume 3195, pages 594–601, sep 2004.
4. A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons., 2001.
5. Scott Makeig, Anthony J. Bell, Tzyy-Ping Jung, and Terrence J. Sejnowski. Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*, 8:145–151, 1996.
6. Scott Makeig, Arnaud Delorme, Marissa Westerfield, Tzyy-Ping Jung, Jeanne Townsend, Eric Courchesne, and Terrence J. Sejnowski. Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biology*, 2004.
7. L. Parra, C. Spence, and B. Vries. Convolutive source separation and signal modeling with ml. In *International Symposioum on Inteligent Systems*, 1997.
8. Barak A. Pearlmutter and Lucas C. Parra. Maximum likelihood blind source separation: A context-sensitive generalization of ICA. In Michael C. Mozer, Michael I. Jordan, and Thomas Petsche, editors, *Advances in Neural Information Processing Systems*, volume 9, page 613. The MIT Press, 1997.
9. G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.