Chapter 2

Independent component analysis and blind source separation

Erkki Oja, Juha Karhunen, Ella Bingham, Maria Funaro, Johan Himberg, Antti Honkela, Aapo Hyvärinen, Alexander Ilin, Karthikesh Raju, Tapani Ristaniemi, Jaakko Särelä, Harri Valpola, Ricardo Vigário

2.1 Introduction

Erkki Oja

What is Independent Component Analysis? Independent Component Analysis (ICA) is a computational technique for revealing hidden factors that underlie sets of measurements or signals. ICA assumes a statistical model whereby the observed multivariate data, typically given as a large database of samples, are assumed to be linear or nonlinear mixtures of some unknown latent variables. The mixing coefficients are also unknown. The latent variables are nongaussian and mutually independent, and they are called the independent components of the observed data. By ICA, these independent components, also called sources or factors, can be found. Thus ICA can be seen as an extension to Principal Component Analysis and Factor Analysis. ICA is a much richer technique, however, capable of finding the sources when these classical methods fail completely.

In many cases, the measurements are given as a set of parallel signals or time series. Typical examples are mixtures of simultaneous sounds or human voices that have been picked up by several microphones, brain signal measurements from multiple EEG sensors, several radio signals arriving at a portable phone, or multiple parallel time series obtained from some industrial process. The term blind source separation is used to characterize this problem.

Our contributions in ICA research. In our ICA research group, the research stems from some early work on on-line PCA, nonlinear PCA, and separation, that we were involved with in the 80's and early 90's. Since mid-90's, our ICA group grew considerably. This earlier work has been reported in the previous Triennial and Biennial reports of our laboratory from 1994 to 2001. A notable achievement from that period was the textbook "Independent Component Analysis" (Wiley, May 2001) by A. Hyvärinen, J. Karhunen, and E. Oja. It has been very well received in the research community; according to the latest publisher's report, over 3500 copies have been sold by August, 2003. The book has been extensively cited in the ICA literature and seems to have evolved into the standard text on the subject worldwide. Another tangible contribution has been the FastICA software package (`http://www.cis.hut.fi/projects/ica/fastica/`) which during the reporting period was downloaded by 8900 registered users. This is one of the few most popular ICA algorithms used by the practitioners and a standard benchmark in algorithmic comparisons in ICA literature.

In the reporting period 2002 - 2003, ICA research stayed as a core project in the laboratory. It was extended to several new directions. It is no more possible to report all this work under a single ICA Chapter. The most advanced developments are now presented in their separate Chapters in this report. They are: "Variational Bayesian learning of generative models" (a project led by prof. Juha Karhunen and Dr. Harri Valpola), "Analysis of independent components in biomedical signals" (a project led by Dr. Ricardo Vigario) and "Computational neuroscience" (a project led by Doc. Aapo Hyvärinen).

This Chapter starts by introducing some theoretical advances undertaken during the reporting period. Also comparisons on post-nonlinear mixtures are reported. Then, several smaller application-oriented ICA projects are covered. The applications range from text mining and astronomical data analysis to telecommunications. Note that more extensive applications are covered especially in the Chapter on biomedical signal analysis. Finally, in the present Chapter, the EU project BLISS is reviewed, which ended during 2003 with highly favorable reviews.

2.2 Theoretical advances

Erkki Oja, Ella Bingham, Aapo Hyvärinen, Jaakko Särelä, Harri Valpola

ICA in a regression problem

In this project it was shown how independent component analysis (ICA) can be used in the context of regression. In a regression problem, one has a set of predictor variables and a set of predicted variables. The task is to generate a mapping between these sets so that given the values of the predictor variables, the values of the predicted variables can be estimated.

The regression problem can be cast into the ICA framework as follows. Using the training data of predictor and predicted variables, the independent components in the data are estimated. Using them we can estimate future values of predicted variables, given the observations of the predictor variables only. The problem is discussed in detail in [1], where it is also shown that regression by ICA is closely related to regression by a multilayer perceptron (MLP) network, which is a widely used neural network. However, regression by ICA is more straightforward and better defined in terms of choosing the number of units and the nonlinear transformations in the hidden layer of the MLP network, and finding the weights of the layers of the network.

Non-negative ICA

The basic linear ICA model can be considered to be solved, with a multitude of practical algorithms and software. However, if one makes some further assumptions which restrict or extend the model, then there is still ground for new theoretical analysis and solution methods. One such assumption is *positivity or non-negativity* of the sources and perhaps the mixing coefficients. Non-negativity is a very natural condition for many practical real-world applications, for example in the analysis of images, text, or spectral data. The constraint of non-negative sources, perhaps with an additional constraint of non-negativity on the mixing matrix, is often known as *positive matrix factorization* or *non-negative matrix factorization*. We refer to the combination of non-negativity and independence assumptions on the sources as *non-negative independent component analysis*.

It was suggested by the co-author of [3] that a suitable cost function for actually finding the rotation could be constructed as follows: denote the estimates for the positive sources by $y_i,\ i = 1, ..., n$, and suppose we have an output truncated at zero, $\mathbf{y}^+ = (y_1^+, ..., y_n^+)$ with $y_i^+ = \max(0, y_i)$. Let us construct a reestimate of whitened but not zero-mean observation data $\mathbf{z} = \mathbf{W}^T\mathbf{y}$, given by $\hat{\mathbf{z}} = \mathbf{W}^T\mathbf{y}^+$, where $\mathbf{W}$ is the (orthogonal) parameter matrix of the ICA problem. Then a suitable cost function would be

$$J(\mathbf{W}) = E\{\|\mathbf{z} - \hat{\mathbf{z}}\|^2\} = E\{\|\mathbf{z} - \mathbf{W}^T\mathbf{y}^+\|^2\} \tag{2.1}$$

because obviously its value will be zero if $\mathbf{W}$ is such that all the y_i are positive, or $\mathbf{y} = \mathbf{y}^+$.

We have considered the minimization of this cost function by on-line learning algorithms. In [3], the cost function (2.1) was taken as a special case of "nonlinear PCA" for which an algorithm was earlier suggested by one of the authors [2]. However, a rigorous convergence proof for the nonlinear PCA method could not be constructed except in some special cases. The general convergence seems a very challenging problem.

In another paper [4] we showed that the cost function (2.1) has very desirable properties in the Stiefel manifold of rotation (orthogonal) matrices: the function has no local minima and it is a Lyapunov function for its gradient matrix flow. A gradient algorithm, suggested in the paper, is therefore monotonically converging and is guaranteed to find the absolute minimum of the cost function. The minimum is zero, giving positive components y_i, which must be a positive permutation of the original unknown sources s_j, as proven in the paper. Some preliminary results along these lines were given by the authors in [5].

Semi-blind source separation by denoising

Many algorithms have been invented to perform blind source separation. There, it is assumed that the sources are completely unknown. However, in many cases of time series separation there exists some prior knowledge on the behaviour of the sources. This is the case especially, when the data is collected in a controlled experiment.

In [8] we have introduced a fast semi-blind source separation algorithm (SBSS) that allows an easy incorporation of such prior knowledge. In one iteration of SBSS, the essential step denoises the current source estimate from part of the noise as well as the interference of the other sources. Then, the projection is sought that gives a source estimate closest to this denoised source estimate in least mean squares sense.

SBSS introduces a full continuum of algorithms, where the denoising can vary from a very detailed matched filter to looser, more general denoising principles, such as the non-Gaussianity in ICA.

Overlearning in ICA

The research on overlearning in ICA has been continued. We have elaborated the discussion as well as suggested several solutions to solve both the problems of spikes and bumps. We have published a comprehensive article on the findings [6]. Further study on a Bayesian approach to overcome the problem has been conducted as well [7].

References

[1] Aapo Hyvärinen and Ella Bingham. Connection between multilayer perceptrons and regression using independent component analysis. *Neurocomputing*, 50(C):211–222, January 2003.

[2] Oja, E.: The nonlinear PCA learning rule in Independent Component Analysis. *Neurocomputing 17*, pp. 25 - 45 (1997)

[3] Plumbley, M. and Oja, E.: A "Non-negative PCA" algorithm for independent component analysis. *IEEE Trans. on Neural Networks 15*, no. 1, pp. (2004)

[4] Oja, E. and Plumbley, M.: Blind separation of positive sources by globally convergent gradient search. *Neural Computation*, to appear.

[5] Oja, E. and Plumbley, M.: Blind separation of positive sources using non-negative PCA. *Proc. 4th Int. Symp. on Independent Component Analysis and Blind Source Separation*, April 1 - 4, 2003, Nara, Japan, pp. 11 - 16 (2003).

[6] J. Särelä and R. Vigário, "Overlearning in marginal distribution-based ICA: analysis and solutions," *Journal of machine learning research*, vol. 4 (Dec), pp. 1447–1469, 2003.

[7] J. Särelä and R. Vigário. A Bayesian approach to overlearning in ICA. Tech. Rep A 70, Lab of Computer and Information Science, Helsinki University of Technology, Finland, 2003.

[8] H. Valpola and J. Särelä. A fast semi-blind source separation algorithm. Tech. Rep A 66, Lab of Computer and Information Science, Helsinki University of Technology, Finland, 2002.

2.3 Comparison studies on blind separation of post-nonlinear mixtures

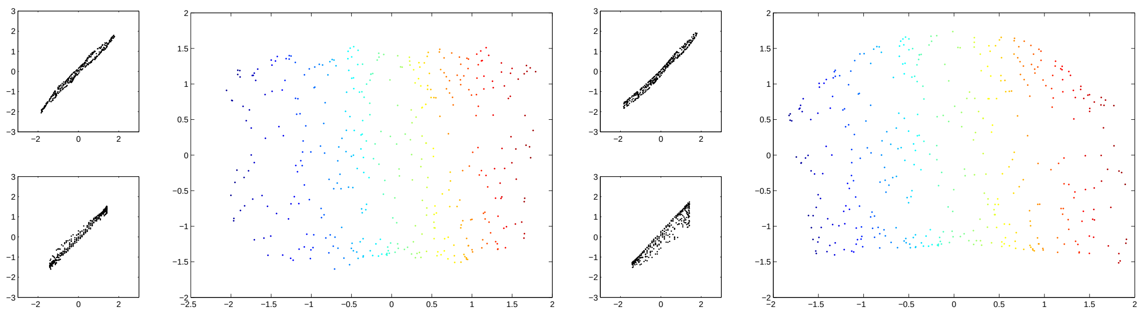
Alexander Ilin, Juha Karhunen

Different approaches proposed for nonlinear independent component analysis (ICA) and blind source separation (BSS) have been recently reviewed in [1]. However, their limitations and domains of preferable application have been studied only a little, and there do not exist hardly any comparisons of the proposed methods. We have experimentally compared two approaches introduced for nonlinear BSS: the Bayesian methods developed at the Neural Network Research Centre (NNRC) of Helsinki University of Technology, and the BSS methods introduced for the special case of post-nonlinear (PNL) mixtures at Institut National Polytechnique de Grenoble (INPG) in France. This comparison study took place within the framework of the European joint project BLISS on blind source separation and its applications.

The Bayesian method developed at NNRC for recovering independent sources consists of two phases: Applying the general nonlinear factor analysis (NFA) [3] to obtain Gaussian sources; and their further rotation with a linear ICA technique such as the FastICA algorithm [2]. The compared BSS method, developed at INPG for post-nonlinear mixtures, is based on minimization of the mutual information between the sources. It uses a separating structure consisting of nonlinear and linear stages [4].

Both approaches were applied to the same ICA problems with artificially generated post-nonlinear mixtures of two independent sources. The sources were a sine wave and uniformly distributed white noise.

Figure 2.1 shows some of the experimental results. The INPG method based on the independence criterion obtained a good signal-to-noise ratio for the recovered sources. However, it failed to cope with non-invertible post-nonlinearities in the mixtures. The more general Bayesian NFA+FastICA approach was able to recover the sources as well. Moreover, it was able to process mixtures with non-invertible distortions.



(a) (b) (a) (b)

PNL method of INPG: SNR 17.73 dB Bayesian NLFA+FastICA: SNR 13.57 dB

Figure 2.1: The two sources recovered from their PNL mixture by the two alternative approaches: (a) – the scatter plots showing the values of one original source signal plotted against the found sources, (c) – the distribution of the found sources. The INPG method used only mixtures with invertible post-nonlinear distortions. The Bayesian approach was able to process the mixture with one non-invertible post-nonlinearity.

Based on the experimental results, the following conclusions were drawn on the applicability of the INPG and Bayesian NFA+FastICA approaches to post-nonlinear blind source separation problems:

- The INPG methods perform better in PNL mixtures with the same number of sources and observed mixtures when all the post-nonlinearities are invertible.
- The performance of both methods can be improved by exploiting more mixtures than the number of sources.
- The advantage of the Bayesian methods in post-nonlinear BSS problems is that they can separate overdetermined post-nonlinear mixtures with non-invertible post-nonlinearities while the existing INPG methods cannot do this.

The results of this comparison study will be presented in a forthcoming international joint paper.

References

[1] C. Jutten and J. Karhunen. Advances in nonlinear blind source separation. In *Proc. of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 915–920, Nara, Japan, 2003. Invited paper in the special session on nonlinear ICA and BSS.

[2] A. Hyvärinen and E. Oja. A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[3] Harri Lappalainen and Antti Honkela. Bayesian nonlinear independent component analysis by multi-layer perceptrons. In Mark Girolami, editor, *Advances in Independent Component Analysis*, pages 93–121. Springer-Verlag, Berlin, 2000.

[4] A. Taleb and C. Jutten. Source Separation in Post-Nonlinear Mixtures. *IEEE Transactions on Signal Processing*, 47(10):2807–2820, 1999.

2.4 Text mining

Ella Bingham

Independent component analysis (ICA) was originally developed for signal processing applications. Recently it has been found out that ICA is a powerful tool for analyzing text document data as well, if the text documents are presented in a suitable numerical form. This opens up new possibilities for automatic analysis of large textual data bases: finding the topics of documents and grouping them accordingly.

First approaches of using ICA in the context of text data considered the data static. In our recent study, we concentrated on text data whose topic changes over time. Examples of dynamically evolving text are chat line discussions or newsgroup documents. The dynamical text stream can be seen as a time series, and methods of time series processing may be used to extract the underlying characteristics — here the topics — of the data.

The project is described in detail in [1]. As an example of dynamically changing textual data, we used chat line discussions where several discussions are going on simultaneously, and the topics of the discussions change dynamically as participants enter and leave the chat room. We were able to find meaningful topics which can be visualized both by different sets of keywords for each topic, and by their behaviour through time (some topics are more or less persistent during the whole period of time; some topics die but will come up again later; and some topics are only active at a certain period of time).

To conclude, our method finds meaningful topics inherent in the data, and the experimental results suggest the applicability of the method to query-based retrieval from a temporally changing text stream.

References

[1] Ella Bingham, Ata Kabán, and Mark Girolami. Topic identification in dynamical text by complexity pursuit. *Neural Processing Letters*, 17(1):69–83, 2003.

2.5 ICA for astronomical data

Erkki Oja, Harri Valpola, Maria Funaro

One of the main research directions in modern astrophysics is to understand the dark matter in the universe, especially the baryonic component, supposed to be formed by compact objects with substellar mass called Massive Astrophysical Compact Halo Objects (MACHO). Possible candidates include small black holes, dwarf stars, or exoplanets. When such an object passes near the line of sight of a star, the luminosity of the star will increase – an effect called gravitational lensing, predicted by the general theory of relativity. In studying other galaxies than our own, individual stars cannot be resolved, but a whole group of unresolved stars is registered in a single pixel element of a telescope CCD camera. In a new technique called pixel lensing, the pixel luminosity variations over time are monitored, and using these time series the lensing events can be detected even in the case of unresolved stars.

A severe problem in the analysis of the images and luminosity variations is the presence of artefacts. One possible cause for artefacts are the individual stars between the far-out galaxy and the camera, which emerge sharply from the luminosity background. Other artefacts are cosmic rays, atmospheric events, and noise in the CCD camera. Separating these artefacts from the interesting astrophysical events is one of the necessary steps in the analysis of pixel lensing data. Artefact removal techniques have to be fast and highly acurate to avoid the interesting phenomena from being eliminated from the data along with the artefacts.

The new idea proposed by us [1] is to use Independent Component Analysis for artefact detection and removal. We can assume that the astrophysical images consist of three additive components: first, images revealing the interesting physical effects like pixel lensing; second, images of artefacts such as atmospheric events, cosmic rays, or the resolved stars; and third, additive noise mainly due to the camera system. All of these are guaranteed to be independent, so the ICA model holds very well. In the image processing that we propose here, we first apply PCA to the images. This has the effect of strongly reducing the additive noise, leaving the images with physical effects and artefacts. Next, we apply ICA on these images, with the result that the artefacts are separated and can be removed. What is left are clean artefact-free images that can be analyzed further for the possible physical phenomena.

Using image data on the M31 Galaxy, it was shown in [1] that several clear artefacts can be detected and recognized based on their temporal pixel luminosity profiles and independent component images. Once these are removed, it is possible to concentrate on the real physical events like gravitational lensing.

References

[1] Funaro, M., Oja, E. and Valpola, H.: Independent component analysis for artefact separation in astrophysical images. *Neural Networks 16*, no. 3-4, pp. 469-478 (2003).

2.6 ICA in CDMA communications

Karthikesh Raju, Tapani Ristaniemi, Juha Karhunen, Erkki Oja

In wireless communication systems, like mobile phones, an essential issue is division of the common transmission medium among several users. A primary goal is to enable each user of the system to communicate reliably despite the fact that the other users occupy the same resources, possibly simultaneously. As the number of users in the system grows, it becomes necessary to use the common resources as efficiently as possible.

During the last years, various systems based on CDMA (Code Division Multiple Access) techniques [1, 2] have become popular, because they offer several advantages over the more traditional FDMA and TDMA schemes based on the use of non-overlapping frequency or time slots assigned to each user. Their capacity is larger, and it degrades gradually with increasing number of simultaneous users who can be asynchronous. On the other hand, CDMA systems require more advanced signal processing methods, and correct reception of CDMA signals is more difficult because of several disturbing phenomena [1, 2] such as multipath propagation, possibly fading channels, various types of interferences, time delays, and different powers of users.

Direct sequence CDMA data model can be cast in the form of a linear independent component analysis (ICA) or blind source separation (BSS) data model [3]. However, the situation is not completely blind, because there is some prior information available. In particular, the transmitted symbols have a finite number of possible values, and the spreading code of the desired user is known. The project started with application of ICA and BSS methods to various problems in multiuser detection [1, 2], trying to take into account the available prior information whenever possible. We showed that ICA based methods can yield considerably better performances than more conventional methods based on second-order statistics. The work carried out during this stage is reviewed together with the necessary background in Chapter 23 of the book [3].

In the second stage of the project in 2001-2003, we have applied independent component analysis to blind suppression of various interfering signals appearing in direct sequence CDMA communication systems. First we studied bit-pulsed jamming, which constitutes an important problem in practical CDMA communication systems. We have taken into account both data modulation and temporally uncorrelated jamming, improving and extending earlier preliminary work on the same problem. Computer simulations show that the proposed method performs better than the well-known RAKE method, which is the standard choice for suppressing jammer signals. The results have been reported in more detail in the conference papers [4, 5].

In papers [6, 7], ICA-RAKE Pre-Switch and ICA-RAKE Post-Switch structures were introduced. They switch between the ICA portion and the RAKE portion depending on the signal-to-jammer ratio. If the jammer signal is weak or absent, preprocessing by ICA is not advisable, because it might even cause additional interference.

Fig. 2.2 shows the distribution of correct bits using the post-switched ICA-RAKE and plain RAKE methods for a coherent 5 path channel. In the case of a single path (left subfigures), post-switched ICA (upper left subfigure) is able to separate the jammer completely since about 95% of the blocks are correct, while the results provided by the conventional RAKE receiver (shown in the lower left subfigure) are poor. The situation is qualitatively similar in the case of 5 paths, as shown by the right subfigures of Fig. 2.2. These results are summarized in the paper [10].

We have applied ICA also to cancellation of interferences due to adjacent cells in a DS-CDMA system. The gain in performance is about $5-8$dB when the interfering source

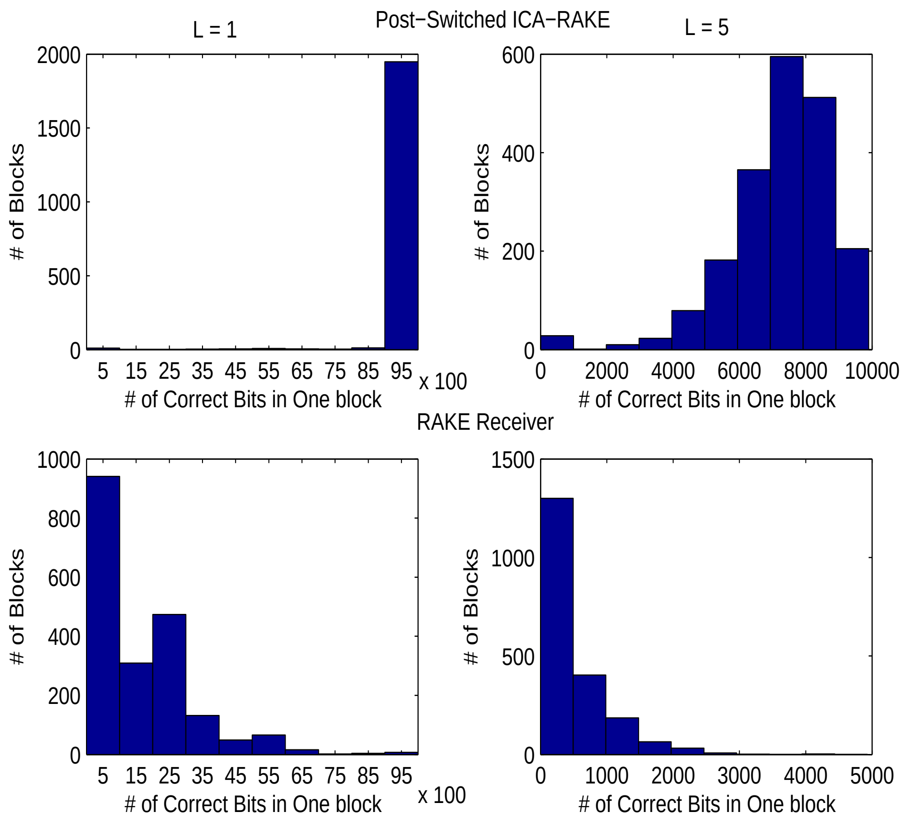


Figure 2.2: Distribution of correct bits per block for post-switched ICA-RAKE (upper subfigures) and plain RAKE (lower subfigures) methods for $L = 1$ (left subfigures) path and $L = 5$ (right subfigures) paths. ICA-RAKE has more than 95% correct bits in most blocks for $L = 1$ while most blocks have at least 70% correct symbols under $L = 5$ coherent paths.

has been suppressed with ICA [8, 9].

References

[1] S. Verdu, *Multiuser Detection.* Cambridge Univ. Press, 1998.

[2] J. Proakis, *Digital Communications.* McGraw-Hill, 3rd edition, 1995.

[3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis.* Wiley, 2001, 481+xxii pages.

[4] T. Ristaniemi, K. Raju, and J. Karhunen, Jammer mitigation in DS-CDMA array systems using independent component analysis. In *Proc. of the 2002 IEEE Int. Conf. on Communications (ICC2002)*, New York City, NY, USA, April 28–May 2, 2002.

[5] K. Raju, T. Ristaniemi, J. Karhunen, and E. Oja, Suppression of bit-pulsed jammer signals in DS-CDMA array systems using independent component analysis. In *Proc. of the 2002 IEEE Int. Symp. on Circuits and Systems (ISCAS2002)*, Phoenix, Arizona, USA, May 26-29, 2002, pp. I-189/I-192.

[6] K. Raju and T. Ristaniemi, ICA-RAKE switch for jammer cancellation in DS-CDMA array systems. In *Proc. of the 2002 IEEE Int. Symp. on Spread Spectrum Techniques and Applications (ISSSTA2004)*, Prague, Czech Republic, September 2-5, 2002, pp 638-642.

[7] T. Ristaniemi, K. Raju, J. Karhunen, and E. Oja, Jammer cancellation in DS-CDMA arrays: pre and post switching of ICA and RAKE. In *Proc. of the 2002 IEEE Workshop on Neural Networks for Signal Processing (NNSP2002)*, Martigny, Switzerland, September 4-6, 2002, pp. 495–504.

[8] T. Ristaniemi, K. Raju, and J. Karhunen, Inter-cell interference cancellation in CDMA array systems by independent component analysis. In *Proc. of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, Japan, April 1-4, 2003, pp. 739–744.

[9] K. Raju and T. Ristaniemi, Exploiting independences to cancel interferences due to adjacent cells in a DS-CDMA system. In *Proc. of the Personal, Indoor, and Mobile Radio Communications (PIMRC 2003)*, Beijing, China, September 7-10, 2003.

[10] K. Raju, T. Ristaniemi, and J. Karhunen, Semi-blind interference suppression on coherent multipath environments. Submitted to a conference.

2.7 Explorative investigation of the reliability of independent component estimates

Johan Himberg, Aapo Hyvärinen

FastICA [2] is a fast, fixed point algorithm for estimating independent components[1]. However, FastICA, as most ICA algorithms, finds a local minimum of its objective function. Even with algorithms which are deterministic and always find the global optimum of their objective function, the valid interpretation of the results need some analysis of the statistical reliability or significance of the components. There are two different reasons for this. Firstly, as real data never exactly follows the ICA model, the contrast function used in the estimation may have many local minima which are all equally good. The independent components are simply not well-defined in this case. Secondly, even in the extreme where the data is exactly generated according to the ICA model, the finite sample size induces statistical errors in the estimation—this is the case where classical analysis of statistical significance and confidence intervals would be needed.

As such, the bootstrapping method for analyzing the statistical reliability of independent components in [3] seems applicable only in the case of deterministic algorithms. Some additional development is required for stochastic algorithms, and consequently, we have developed an interactive visualization method and software package[2] for reliability assessment for FastICA.

Icasso estimates a large number of independent components with changing initial conditions and/or bootstrapping, and visualizes their clustering in the signal space. Each estimated independent component is one point in the space. If an independent component is stable, (almost) every run of the algorithm should produce a point that is very close to the ideal component corresponding to the cluster center. Reliable independent components correspond to tight clusters, and unreliable ones correspond to points which do not belong to any cluster. Preliminary results with a biomedical benchmarking data set are reported in [1]. See also Fig. 2.3.

References

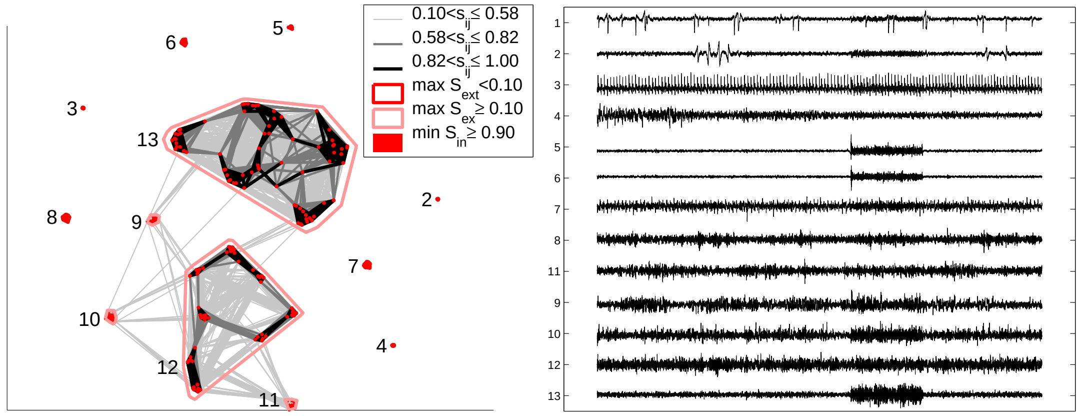
[1] J. Himberg and A. Hyvärinen. (2003). Icasso: software for investigating the reliability of ICA estimates by clustering and visualization. In *Proc. Int. workshop of Neural Networks for Signal Processing (NNSP2003)*, Toulouse, France.

[2] A. Hyvärinen. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions of Neural Networks*, 10(3):626–635.

[3] F. Meinecke, A. Ziehe, M. Kawanabe, and K.-R. Müller (2002). A resampling approach to estimate the stability of one-dimensional or multidimensional independent components. *IEEE Transactions on Biomedical Engineering*, 49(12):1514–1525.

[4] R. Vigário, V. Jousimäki, M. Hämäläinen, R. Hari, and E. Oja. (1998). Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In *Advances in Neural Information Processing Systems*, vol. 10, pp. 229–235. MIT Press.

[1] A popular public domain software package is available at `www.cis.hut.fi/research/software.shtml`
[2] at `www.cis.hut.fi/jhimberg/icasso`



(a) (b)

Figure 2.3: A MEG data set from [4] is analyzed with FastICA and Icasso. Panel (a) shows the correlations of all ICA estimates; 13 clusters have been selected and labeled #1–#13. One estimate from each is selected, the one that is nearest to the centroid of the cluster. The estimates are presented in panel (b) ordered according to the dispersion of the clusters. From the previous studies, we know that source estimates corresponding to clusters #1 and #2 correspond to eye movements, #3 to heart and #9 to the digital watch. Sources #5 and #6 are related to muscular activities due to biting. Source #4 is interesting since it is clearly well estimated but the physiological explanation is not yet known.

2.8 The European joint project BLISS

Juha Karhunen, Erkki Oja, Harri Valpola, Ricardo Vigario, Antti Honkela, Jaakko Särelä

Our laboratory has been one of the five participants in a large European joint project on Blind Source Separation and Applications, abbreviated BLISS. The project originally covered three years between June 2000 and June 2003, but it was later on extended by five months up to October 2003. The total funding of the BLISS project was 1.2 million euros, and it belonged to the "Information Society Technologies" programme (1998-2002) funded by the European Community. The other participating institutes and the leaders of the BLISS project there were:

- INESC, Lisbon, Portugal (Prof. Luis Almeida, coordinator);
- INPG (Inst. Nat. Polytechnique de France), Grenoble, France (Profs. Christian Jutten and Dinh-Tuan Pham);
- GMD First (Fraunhofer Institute), Berlin, Germany (Prof. Klaus-Robert Müller);
- McMaster University, Hamilton, Canada (Prof. Simon Haykin; adjunct member, getting its funding from Canada).

INESC withdrew from the project in summer 2002, and after that INPG has acted as a new coordinator. In addition, the project had both industrial and scientific advisory board members.

The project divided into two major parts: Theory and Algorithms, and Applications. The first part Theory and Algorithms consisted of three subprojects, which are Linear ICA, Nonlinear Separation, and Nonlinear BSS (Blind Source Separation). Our laboratory has formally been involved in the last two subprojects, but we contributed also to the subproject on linear ICA. The second major part Applications had two subprojects, Biomedical Applications and Acoustic Mixtures, and our laboratory participated in the first subproject on biomedical applications.

Meetings of the participants of the project have been arranged mainly in context with conferences and other events. The second official review meeting of the BLISS project was held in Brussels, Belgium, in July 2002. There it was decided to extend the project by five months for achieving its goals, and the workplan was revised appropriately. The final official review meeting was held in Paris at the end of October 2003. The reviewers were quite satisfied with the final results of the project, stating that all the major goals of the BLISS project have been achieved.

Especially during the later part of the project, practical co-operation between the participating laboratories has been deepened by researcher visits and writing of joint publications. Two major events intended to European researchers and graduate students have been organized largely within the BLISS project: the European Meeting on Independent Component Analysis in Vietri sul Mare, Italy, in February 2002; and the European Summer School on ICA in Berlin, Germany, in June 2003. Both events were quite successful with tens of interested participants.

The research carried out in our laboratory using the BLISS project funding is described in the chapters "Independent component analysis and blind source separation", "Variational Bayesian learning of generative models", and "Analysis of independent components in biomedical signals", as well as in the many associated publications. The results have been reported also in the deliverables and reports of the project. More information on the

BLISS project is available on its homepage [1] where one can find final reports and other deliverables, data sets, as well as software.

References

[1] Homepage of the BLISS project `http://www.lis.inpg.fr/pages_perso/bliss/`.