35 Extensions of the Basic Source Separation Problem

Petteri Pajunen, Juha Karhunen, Aapo Hyvärinen, Harri Lappalainen, and Simona Mălăroiu

35.1 Nonlinear mixing of the sources

In the basic Independent Component Analysis (ICA) model

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{96}$$

it is assumed that the M unknown source signals are linearly mixed into M different known mixtures. Here $\mathbf{s}(t)$ denotes the M-vector containing the M source signals at time t. The matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_M]$ is a constant $M \times M$ mixing matrix whose elements are the unknown coefficients of the mixtures. The columns $\mathbf{a}_i$ of $\mathbf{A}$ are the basis vectors of ICA, and $\mathbf{x}(t)$ is the M-dimensional tth data vector made up of the mixtures at discrete time (or point) t.

In realistic applications the linearity assumption of the simple basic model (96) is not necessarily valid. Since ICA defines a linear transformation $\mathbf{B}$ which makes the components of the random vector $\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$ as independent as possible, it is natural to consider more general transformations which have the same effect.

The Self-Organizing Map (SOM) can be used to define a nonlinear transformation which approximately estimates the probability density of the input data. The weight vectors of SOM are distributed proportionally to the input vector density. Using certain learning rules this relationship is accurate, and the distribution of the SOM weight vectors is asymptotically the same as the distribution of the input vectors. This forces each weight vector to have the same probability of 'winning' and therefore the distribution on the converged map is uniform. By using a rectangular map, the output vector coordinates become approximately statistically independent [1]. We have successfully applied this property of the SOM to the blind source separation problem when the source signals have a flat (sub-Gaussian) distribution, and the nonlinear mixing function is not too nonlinear [2]. The restriction of using SOM is that the source densities are implicitly modeled as uniform densities. If there is prior knowledge of the source densities, this can be used to improve the results. The generative topographic map (GTM) allows to do this. We applied this to the nonlinear ICA problem with improved results compared to SOM [7].

Even though this method has some limitations, its advantage is that it is truly neural, contrary to the few other existing approaches to the generally very difficult nonlinear blind separation problem.

We have also studied the theoretical questions that arise when considering nonlinear ICA. Especially we have shown that the solution to nonlinear ICA always exists but is highly non-unique. We have also developed a set of conditions which lead to a unique solution [9].

35.2 Binary sources

Another restriction of the basic ICA model is the assumption that the number of available mixtures equals to the number of source signals. If the source signals are continuous, it is generally impossible to separate more sources than mixtures, because the solution of the blind separation problem becomes highly nonunique. However, assuming that all the source vectors are binary and all the mixtures are different, the mixing transformation is one-to-one, and in theory it then becomes possible to separate the sources. In the special case of two mixtures the separation can be achieved by computing the convex hull of the observed mixtures [3]. It can be shown that the convex hull uniquely determines the basis vectors of ICA (and the number of them) under mild assumptions. The sources can then be easily separated when the basis vectors are known.

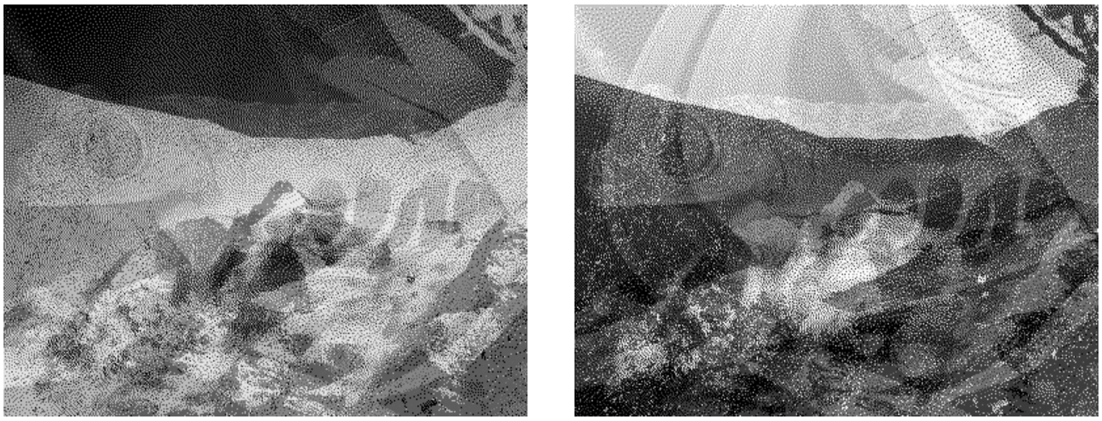

Figure 57: Noisy mixture images.

An example of blind separation of binary sources is given in figures 57 and 58 highlighting the possibility of separating binary signals from less mixtures than sources. In figure 57, two noisy linear mixtures of four binary images are shown. A binary source separation algorithm developed by us was applied to these mixtures producing the separated images shown in figure 58.

35.3 The effect of noise, correlation, and various network structures to separation results

In a joint project with the laboratory of Artificial Brain Systems, RIKEN research institute, Japan, we have considered some other in practice interesting extensions of the basic blind source separation problem. Such extensions have been outlined and discussed in an invited tutorial review paper [4]. We have in particular considered what happens in different neural network structures when the number of source signals is different from the number of sources and/or outputs of the network, and proposed various methods for handling such situations. We have also studied the ability of the networks to separate correlated sources, and the effect and removal or suppression of noise in context with blind separation. Two journal papers [5,6] summarize the results achieved in this joint project on these topics.

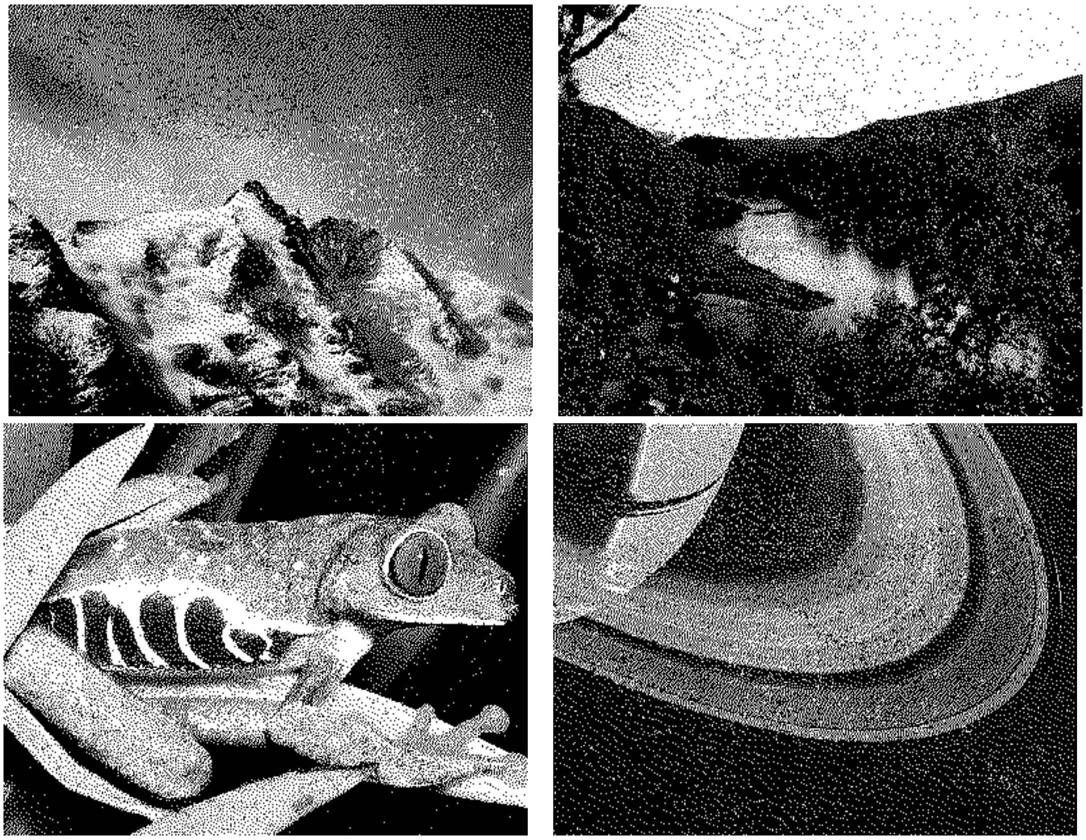

Figure 58: Separated images.

Later on, we have shown [16] that if there is additive noise present in the basic ICA/BSS model (96), the optimal solution of the problem in fact depends nonlinearly on the observed mixture vectors $\mathbf{x}(t)$. Computationally efficient approximations to the optimal maximum likelihood solution of this problem have been derived in various situations in [16].

35.4 Local ICA methods

In standard ICA, a linear data model (96) is used for a global description of the data. Even though linear ICA yields meaningful results in many cases, it can provide a crude approximation only for nonlinear data distributions. In [10], a new structure is proposed, where local ICA models are used in connection with a suitable clustering algorithm grouping the data. The clustering part is responsible for an overall coarse nonlinear representation of the underlying data, while linear ICA models of each cluster are used for describing local features of the data. The goal is to represent the data better than in linear ICA while avoiding computational difficulties associated with nonlinear ICA. In first experiments with such a local ICA method, we have used simple K-means clustering. The proposed method performs well for natural image data, yielding meaningful local features with suitable preprocessing [10].

35.5 Generalization of ICA using complexity and coding

It is possible to generalize independent component analysis by considering representations of the observed mixtures which can be coded using as few bits as possible. Equivalently, we can look for representations that have a minimum complexity. Choosing a linear representation and measuring the complexity using entropy, we obtain the same approach as in ICA where mutual information is minimized [8]. The generalization here is conceptual and is of fundamental importance. It allows principled application of ICA to *any* data instead of noiseless data containing linear mixtures of strictly independent sources. The general form of the complexity measure serves as a framework for true extensions. The entropy can be replaced by any other coding measure, which can be chosen quite freely. Using compression algorithms to approximately measure the codelength yields improved results compared to standard ICA algorithms [8]. Using principal component analysis to measure the complexity leads to new algorithms as well [11, 12].

35.6 Bayesian learning

In modeling there is a trade-off between the flexibility of models and robustness against overfitting. Too simple a model is not able to capture all the regularities and structure of the data, but too complex a model overfits, i.e., learns also the coincidental noise always present in real data.

Bayesian approach to learning solves the trade-off by finding the most probable model. It is closely related to information theoretically motivated approaches which minimize the description length of the data, because the description length is defined to be the minus logarithm of the probability. Minimal description length thus means maximal probability.

In practice, Bayesian learning involves approximating the posterior density of the models. This has been done using a recently developed method called ensemble learning, where a simple parametric approximation is fitted to the posterior density by minimizing the Kullback-Leibler distance. The method has been applied to linear ICA in [13]. A nonlinear extension, where the nonlinear mapping from the sources to the observations is modeled by a multi-layer perceptron (MLP) network, has been studied in [14]. The methods for using ensemble learning with MLP networks have been developed in [15] using an information theoretically motivated approach.

References

[1] P. Pajunen. Nonlinear independent component analysis by self-organizing maps. In C. von der Malsburg et al., editors, *Artificial Neural Networks – ICANN'96*, pages 815–819. Springer, 1996.

[2] P. Pajunen, A. Hyvärinen, and J. Karhunen. Nonlinear blind source separation by self-organizing maps. In S. Amari et al., editors, *Progress in Neural Information Processing (ICONIP-96)*, pages 1207–1210. Springer, 1996.

[3] P. Pajunen. An algorithm for binary blind source separation. Technical Report A36, Lab. of Computer and Information Science, Helsinki University of Technology, 1996.

[4] J. Karhunen. Neural approaches to independent component analysis and source separation. In *Proc. of the 4th European Symposium on Artificial Neural Networks (ESANN'96)*, pages 249–266, Bruges, Belgium, April 1996.

[5] J. Karhunen, A. Cichocki, W. Kasprzak, and P. Pajunen. On neural blind separation with noise suppression and redundancy reduction. *Int. J. of Neural Systems*, vol. 8, no. 2, pp. 219-237, 1997.

[6] A. Cichocki, J. Karhunen, W. Kasprzak, and R. Vigario. On neural blind separation with unequal numbers of sensors, sources, and outputs. *Neurocomputing*, vol. 24, pp. 55-93, February 1999.

[7] P. Pajunen and J. Karhunen. A maximum likelihood approach to nonlinear blind source separation. In *Proc. Int. Conf. on Artificial Neural Networks (ICANN'97)*, pages 541–546, Lausanne, Switzerland, October 1997.

[8] P. Pajunen. Blind source separation using algorithmic information theory. *Neurocomputing*, vol. 22, pp. 35-48, 1998.

[9] A. Hyvärinen and P. Pajunen. Nonlinear independent component analysis: Existence and uniqueness results. *Neural Networks*, vol. 12, no. 3, pp. 429-439, 1999.

[10] J. Karhunen and S. Mălăroiu. Local independent component analysis using clustering. In *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pp. 43–48, Aussois, France, January 1999.

[11] P. Pajunen. Blind source separation of natural signals based on approximate complexity minimization. In *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pp. 267-270, Aussois, France, January 1999.

[12] A. Ypma and P. Pajunen. Rotating machine vibration analysis using second-order independent component analysis. In *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pp. 37-42, Aussois, France, January 1999.

[13] H. Lappalainen. Ensemble learning for independent component analysis. In *Proceedings of the ICA'99*, pages 7–12, Aussois, France, January 1999.

[14] H. Lappalainen and X. Giannakopoulos. Multi-layer perceptrons as nonlinear generative models for unsupervised learning: a Bayesian treatment. Submitted for ICANN'99.

[15] H. Lappalainen. Using an MDL-based cost function with neural networks. In *Proceedings of the IJCNN'98*, pages 2384–2389, Anchorage, Alaska, May 1998.

[16] A. Hyvärinen. Independent component analysis in the presence of Gaussian noise by maximizing joint likelihood. *Neurocomputing*, vol. 22, pp. 49–67, 1998.