Nonlinear Blind Source Separation by Self-Organizing Maps

Petteri Pajunen, Aapo Hyvärinen, and Juha Karhunen

Helsinki University of Technology
Laboratory of Computer and Information Science
Rakentajanaukio 2 C, FIN-02150, Espoo, Finland
e-mail: Petteri.Pajunen@hut.fi

***Abstract*— In neural blind source separation most approaches have considered the linear source separation problem where the input data consist of unknown linear mixtures of unknown independent source signals. The solution is a linear transformation which makes the output vector components statistically independent. More generally we can consider nonlinear mixtures of sources. Then we can try to separate the sources by constructing mappings that make the components of the output vectors independent. We show that such a mapping can be approximately realized using self-organizing maps with rectangular map topology. We apply these mappings to the separation of nonlinear mixtures of sub-Gaussian sources.**

1 Introduction

In linear blind source separation (BSS), the goal is to separate independent sources signals from their linear mixtures using a minimum of a priori information. Such blind techniques have applications for example in array processing, communications, speech processing, and medical signal processing. Recently, BSS has become an active research area of unsupervised neural learning. Several neural algorithms that use nonlinearities instead of explicit higher-order statistics have been proposed; see [1, 2, 3, 4]. These algorithms try to find a linear transformation that makes the components of the output vectors statistically independent. The underlying data model is essentially the same as in linear Independent Component Analysis (ICA), which is studied and defined formally in the fundamental paper [5].

One natural generalization of the linear ICA and BSS is to consider the problem of finding more general, nonlinear transformations that would yield statistically independent outputs. This extension has already been considered in some papers [6, 7, 8, 9]. However, the proposed algorithms are not completely satisfying from a neural network point of view, because they are either complicated and/or require explicit computation of higher-order moments.

In this paper, we introduce a new neural approach for nonlinear blind source separation and independent component analysis. It uses self-organizing maps to construct mappings that make output vectors statistically independent. These mappings are used to separate nonlinearly mixed sub-Gaussian sources.

2 The Blind Source Separation Problem

We consider the problem of blind separation of statistically independent source signals $s_k^i,\ i = 1, \ldots, M$, where k is a time index. It is assumed that we only observe the instantaneous mixtures

$$\mathbf{x}_k = F(\mathbf{s}_k) \tag{1}$$

of the sources $\mathbf{s}_k = [s_k^1, s_k^2, \ldots, s_k^M]^T$, where $F : \mathbb{R}^m \to \mathbb{R}^m$ is an unknown mixing function. The problem consists of finding a mapping $G : \mathbb{R}^m \to \mathbb{R}^m$ which would give estimates of the source signals as

$$\mathbf{y}_k = G(\mathbf{x}_k). \tag{2}$$

In the linear case, the mixing function F is assumed to be an invertible linear transformation. The mixtures are obtained as a matrix multiplication $\mathbf{x}_k = \mathbf{A}\mathbf{s}_k$. The solution G exists and is also a linear transformation. The estimated sources are $\mathbf{y}_k = \mathbf{B}\mathbf{x}_k$. It can be shown that the matrix $\mathbf{B}$ cannot be uniquely determined by the independence assumption only. The ordering and the variances of the sources are arbitrary without additional constraints. However, their waveforms can be recovered in practice.

The nonlinear case, where we do not assume that the mixing function F is linear, requires obviously some constraints to make the problem tractable. Intuitively it is natural to require that the mixing function is one-to-one. If we also require continuity of F and G, it follows that F is a *homeomorphism*, i.e. a topology-preserving mapping. Now the problem is essentially geometric; the mixing function F transforms the joint density of the sources into mixture density and we should determine an inverse mapping G which does the opposite. In the nonlinear case, the indeterminacy in the mapping G is even more serious. Any componentwise function

$$\mathbf{x} = [H^1(s^1), H^2(s^2), \ldots, H^m(s^m)]^T. \tag{3}$$

of the source vector $\mathbf{s}$ keeps the components statistically independent. Thus the waveforms are in general not recoverable. However, they can be approximately recovered in practice in many situations.

3 Self-Organizing Maps and Statistical Independence

A self-organizing map [10] performs a mapping from input space to an array of nodes in the output space. The positions of the nodes are fixed and each node is represented by a reference vector in the input space. Input vectors are mapped by finding the closest reference vector with respect to some distance function. The image of the input vector is the corresponding node on the map.

The self-organizing map can be made continuous if we apply some suitable interpolation method. Then we can think of SOM as a homeomorphism from input space to the map.

The distribution of the reference vectors in the input space can be roughly described using the density

$$c \cdot p(\mathbf{x})^\alpha \tag{4}$$

where p is the density of the input vector $\mathbf{x}$ and α is the magnification factor. If $\alpha = 1$, the weight vectors are distributed approximately according to the input space density. In this case we have the important property that each weight vector on the map is equally likely the winner, that is, it lies closest to a random input vector. This implies that the joint density on the map is uniformly distributed. Since we consider rectangular maps, the final implication is that the density on the map is factorizable, i.e. the coordinates are statistically independent. Learning rules that can make the magnification factor equal to one have been recently proposed [11, 12].

4 Application of SOM to Blind Source Separation

The self-organizing map can be used to estimate the inverse of the mixing function G by taking the set of observed mixture vectors $\mathbf{x}_k$ as the input vectors to the SOM. Then the coordinates of the winner neuron on the map define the estimated source vector $\mathbf{y}_k$ for each $\mathbf{x}_k$. Using interpolation on the map this function can be made continuous.

Even though the self-organizing map can produce statistically independent output vectors, this property does not guarantee that the components of the output vectors are good estimates for the source signals. With some heuristic constraints, however, it seems that the sources can be separated at least roughly. The mixture density should be of such a shape that a rectangular map can naturally adapt to it. Furthermore this natural adaptation should provide the correct separation. One reasonable class of problems for which this description holds consists of sub-Gaussian sources that are first mixed linearly and then mildly nonlinearly distorted. Typically, sub-Gaussian signals have a probability density which is flatter than the Gaussian density. Formally stated, sub-Gaussian signals have the property $E[x^4] - 3(E[x^2])^2 < 0$. The expression on the left is called the *kurtosis* of x, and it equals zero for Gaussian signals.

Due to the flat shape of sub-Gaussian densities the linear mixture of sub-Gaussian sources gives often rise to a roughly rectangular density. The converged map in Figure 1 demonstrates this. Mild nonlinearities do not distort too much the rectangular form of the density, allowing fitting of a self-organizing map.

5 Experiments

In each experiment two sub-Gaussian source signals are considered. The source signals consisted of a sinusoid and uniformly distributed white noise.

First, the sources $\mathbf{s}_k$ were linearly mixed using a mixing matrix

$$\mathbf{A} = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix} \tag{5}$$

The resulting vectors $\mathbf{v}_k = \mathbf{A}\mathbf{s}_k$ were distorted by a nonlinearity f applied separately to each component. This yields the mixture vectors

$$\mathbf{x}_k = [f(v_k^1), f(v_k^2), \ldots, f(v_k^m)]^T. \tag{6}$$

The resulting vectors were first whitened so that their covariance matrix becomes the identity matrix.

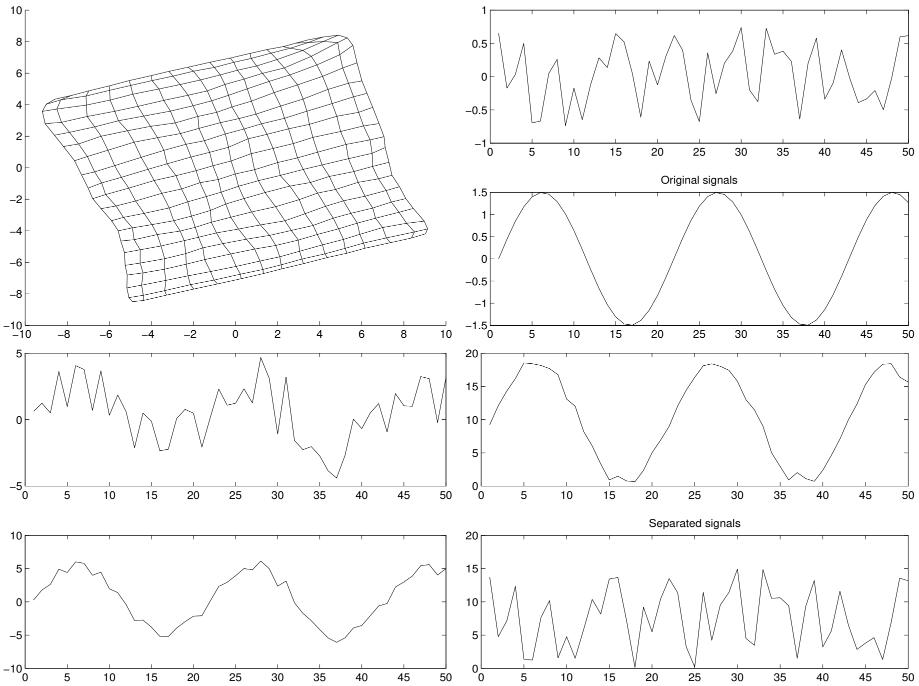

Figure 1: Results using linear mixtures; Top left: converged map. Top right: original source signals. Bottom left: mixture signals. Bottom right: separated source signals.

Fig. 1 shows the results for the linear case where $f(x) = x$ and Fig. 2 the results for $f(x) = x^3 + x$. Note that the signs of separated sources have been changed in some cases to make comparisons with original source signals easier.

In these experiments the basic SOM learning rule has been used. It is possible that the learning rules proposed in [11, 12] could improve the results since the density on the map would be closer to uniform density. Further experiments are currently being made.

6 Discussion

A method for separating sub-Gaussian sources from nonlinear mixtures is presented. First experiments show that at least in simple cases the sources can be recovered. However, the resulting estimates of source signals are often noisy. The noise in the estimated sources is partly due to the quantization error caused by a finite number of neurons. Using interpolation this can be reduced but not eliminated completely. Also the indeterminacy in the separating function G mentioned earlier can cause distortion to separated sources.

The complexity of SOM grows exponentially with the number of dimensions of the map. Thus it is not practical to consider mixtures of a large number of sources using this method.

The main problem of our approach is the requirement of sub-Gaussian sources. This requirement is mainly due to the inherent rectangular topology of the self-organizing map. Further research is currently carried out to extend this method to a larger class of problems.

References

[1] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*. John Wiley, 1994. (new revised and improved edition).

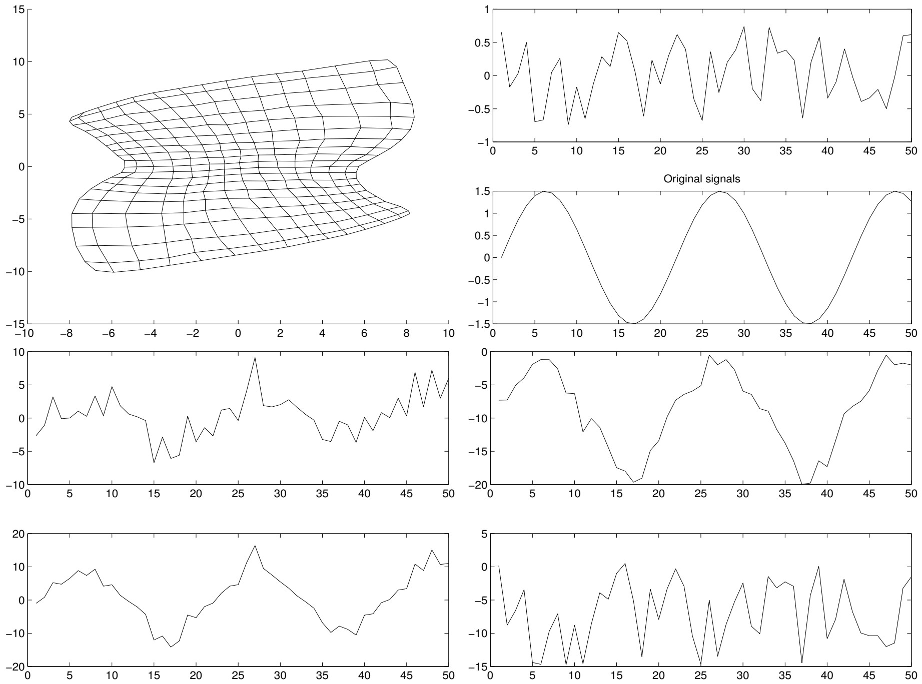

Figure 2: Results using linear mixtures distorted by $f = x^3 + x$; Top left: converged map. Top right: original source signals. Bottom left: mixture signals. Bottom right: separated source signals.

[2] J. Karhunen, E. Oja, L. Wang, R. Vigario, and J. Joutsensalo, "A class of neural networks for independent component analysis," Report A28, Helsinki Univ. of Technology, Lab. of Computer and Information Science, October 1995. Manuscript submitted to a journal.

[3] J. Karhunen, "Neural approaches to independent component analysis and source separation," in *Proc. of ESANN'96 (4th European Symposium on Artificial Neural Networks)*, (Bruges, Belgium), April 1996. to appear, invited paper.

[4] A. Bell and T. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1004–1034, 1995.

[5] P. Comon, "Independent component analysis – a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[6] G. Burel, "Blind separation of sources: A nonlinear neural algorithm," *Neural Networks*, vol. 5, no. 6, pp. 937–947, 1992.

[7] G. Deco and W. Brauer, "Nonlinear higher-order statistical decorrelation by volume-conserving neural architectures," *Neural Networks*, vol. 8, no. 4, pp. 525–535, 1995.

[8] L. Parra, "Symplectic nonlinear component analysis," in *NIPS-95,Advances in Neural Information Processing Systems 8*, (Cambridge, MA), MIT Press, 1996. (in press).

[9] L. Parra, G. Deco, and S. Miesbach, "Statistical independence and novelty detection with information preserving nonlinear maps," *Neural Computation*, vol. 8, pp. 260–269, 1996.

[10] T. Kohonen, *Self–Organizing Maps*, vol. 30 of *Springer Series in Information Sciences*. Springer–Verlag, 1995.

[11] M. Herrmann, H.-U. Bauer, and R. Der, "Optimal magnification factors in self-organizing feature maps," in *Proc. of the Int. Conf. on Artificial Neural Networks*, (Paris, France), pp. 75–80, October 9-13 1995.

[12] M. Van Hulle, "Globally-ordered topology-preserving maps with a learning rule performing local weight updates only," in *Neural Networks for Signal Processing V*, pp. 95–104, IEEE Press, 1995.