Adaptive Underdetermined ICA for Handling an Unknown Number of Sources

Andreas Sandmair, Alam Zaib, and Fernando Puente León

Karlsruhe Institute of Technology
Institute of Industrial Information Technology
Hertzstr. 16, 76187 Karlsruhe, Germany
{sandmair,puente}@kit.edu
http://www.iiit.kit.edu

Abstract. Independent Component Analysis is the best known method for solving blind source separation problems. In general, the number of sources must be known in advance. In many cases, previous assumption is not justified. To overcome difficulties caused by an unknown number of sources, an adaptive algorithm based on a simple geometric approach for Independent Component Analysis is presented. By adding a learning rule for the number of sources, the complete method is a two-step algorithm, adapting alternately the number of sources and the mixing matrix. The independent components are estimated in a separate source inference step as required for underdetermined mixtures.

Keywords: Underdetermined blind source separation, independent component analysis.

1 Introduction

Since its inception, Independent Component Analysis (ICA) has become a fundamental tool for solving Blind Signal Separation (BSS) problems [2]. BSS involves extracting the source signals from multiple sensor observations which are (linear) mixtures of unobserved source signals. Based on the principle of statistical independence, ICA renders output signals as independent as possible by evaluating e.g. higher order statistics (HOS). Originally, ICA was designed to solve determined linear systems of equations (number of sources is equal to number of sensors). If there are fewer sensors than sources, the problem is referred to as underdetermined or overcomplete and more difficult to solve. Therefore, several methods based on classical [5] and geometric approaches [9], [10] have been proposed. In order to solve the BSS problems, whether in the determined or in the underdetermined case, the number of sources should be known in advance. The problem of unknown number of sources in BSS has received little attention in the past, and current approaches are still less developed. Therefore, the problem of unknown source numbers will be addressed in this paper. In the following, a new approach based on a geometric algorithm is presented.

V. Vigneron et al. (Eds.): LVA/ICA 2010, LNCS 6365, pp. 181–188, 2010.

2 Independent Component Analysis

2.1 Basic Principles

In ICA, it is assumed that the observed m-dimensional (sensor) data $\mathbf{x}(t) = [x_1(t), \ldots, x_m(t)]^{\mathrm{T}}$ has been generated from the model

$$\mathbf{x}(t) = \mathbf{A}\,\mathbf{s}(t), \tag{1}$$

where $\mathbf{A}$ is some unknown mixing matrix of dimensions $m \times n$ and $\mathbf{s}(t) = [s_1(t), \ldots, s_n(t)]^{\mathrm{T}}$ is the n-dimensional source data. In terms of column vectors $\mathbf{a}_i$ of the matrix $\mathbf{A}$, eq. (1) can be rewritten as

$$\mathbf{x}(t) = \sum_{i=1}^{n} \mathbf{a_i}\, s_i(t). \tag{2}$$

The goal of ICA is to estimate both the mixing matrix $\mathbf{A}$ and the independent components $s_i(t)$ given only the observed data $\mathbf{x}(t)$. In the determined case, where the number of sources is equal to the number of sensors ($n = m$), this problem can be rephrased as finding an inverse transformation $\mathbf{W}$ such that the original signals $s_i(t)$ can be reconstructed as $\mathbf{s}(t) = \mathbf{W}\,\mathbf{x}(t)$. To apply Independent Component Analysis, the signals must be statistically independent and non-Gaussian distributed. Based on these assumptions, the mixing matrix can be estimated by means of measures describing the independence of the components [3].

To sum up, ICA essentially represents a linear transformation of multivariate data, that captures the underlying structure in the data. This is exploited in many applications including BSS and feature extraction.

2.2 Underdetermined ICA

For underdetermined systems, the estimation of the source signals is more complex, because the source recovery problem is ill-posed ($n > m$). However, after estimating the mixing matrix, the original signals can be recontructed by exploiting the underlying statistical structure. Unfortunately, finding the 'best' representation in terms of an overcomplete basis is a challenging problem because the signal representation is not a unique combination of the basis functions (vectors). The problem of estimating original sources from sensor observations now involves two separate problems. One is to estimate the mixing matrix, referred to as matrix recovery step, and the other is to estimate the original sources also called source inference step. This is in sheer contrast with the determined case, where source inference is trivially done by inverting the mixing matrix. It is also worth mentioning that even if the mixing matrix is perfectly estimated, original sources cannot be recovered perfectly, because some information is permanently lost in representation.

3 Adaptive Geometrical Approach

A geometric approach to ICA was first proposed by Puntonet [8] and since that time successfully used for separating real-world data in determined and underdetermined cases. Because the geometrical approaches do not require estimation of HOS, both the matrix recovery and source inference steps are decoupled and independent of each other. Source inference can be obtained by maximum likelihood approaches or linear programming [10], but will not be described here in detail. In the following, the matrix recovery step for a geometric algorithm [10] is presented, before a extension of the algorithm to an unknown number of sources is illustrated. Following considerations are restricted to two sensors ($m = 2$.)

3.1 Geometrical Approach

The basic idea of geometric approaches is to use the concept of independence from a geometrical point of view. As the mixing process can be regarded as a geometrical transformation of a rectangular area into a parallelogram, the angle of rotation can be identified either in the mixture or in the whitened space, in order to recover the original sources using ordinary geometric algorithms [7]. The theoretical background for geometric ICA has been studied in detail and a convergence criterion has been derived, which resulted in a faster geometric algorithm [9]. Ideas of geometric algorithms have been successfully generalized to overcomplete and higher-dimensional systems [10].

The first step of the geometrical algorithm is to project the mixture or the observed data onto the unit sphere. The task is then to locate the axis of the maxima of the distributions on the unit sphere, which correspond to the original basis vectors, thus solving the separation problem. The idea of identifying the axis of maximum distributions is implemented as an unsupervised neural net with competitive learning which contains $2\,n$ elements (neurons).

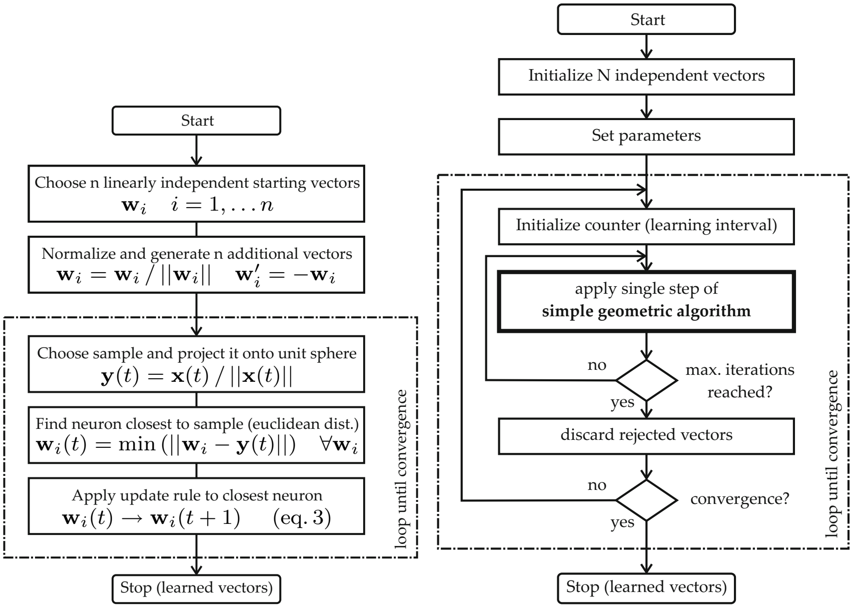
As follows, the method will be introduced according to the steps in the flow diagram shown in figure 1. The key elements of the algorithm are: initialization (random) of $2\,n$ elements, calculating the proximity of the input data sample from each element w.r.t. the Euclidean metric and then applying following update rule to the closest or winning neuron:

$$\begin{aligned} \mathbf{w}_i(t+1) &= \mathrm{Pr}[\mathbf{w}_i(t) + \eta(t)\mathrm{sgn}(\mathbf{y}(t) - \mathbf{w}_i(t))] \\ \mathbf{w}_i^{'}(t+1) &= -\mathbf{w}_i(t+1) \end{aligned} \tag{3}$$

where ‘Pr’ denotes the projection onto the unit sphere. All other neurons are not moved in this iteration. A frequency f_i is assigned to each element (neuron), which counts the number of times each neuron ($\mathbf{w}_i$) has won. The step size is then modified according to:

$$\eta(t+1) = \eta_0\, \mathrm{e}^{f_i(t)/\tau}. \tag{4}$$

To prevent the network from becoming stuck in a meta-stable state, the learning rate is maintained at a certain low level η_f.



(a) geometric algorithm (b) modified geometric algorithm

Fig. 1. Flowchart for two different geometric algorithms

3.2 Handling an Unknown Number of Sources

A limitation of most ICA algorithms is that the number of sources n must be known in advance. To become independent of this constraint, an extension of the geometric algorithm presented in previous section is proposed. It combines the source number estimation and the geometric learning procedure to recover the mixing matrix. Therefor, the independence of the matrix recovery step and the source inference step is needed.

Recall that in the geometric algorithm the observed data is first projected onto the unit sphere, which results in an asymmetric distribution. But unlike the previous case, not only the locations of the axis of maximum distribution that correspond to the true basis vectors has to be estimated, but also the number of maxima, which correspond to the number of sources.

The idea is to start with a large number of independent basis vectors, or neurons N that span the whole data space. A basic assumption is that N must be greater than the actual number of unknown sources n, which can be easily satisfied because for m sensors. The maximum number of signals that can be separated using m sensors can be assessed knowing the nature of the distributions of the sources and the degree of their sparseness. After N is fixed, the geometric algorithm is applied as usual and the network is gradually pruned iteratively by comparing the neuron frequencies with a predefined threshold. The key point is that frequencies greater than the threshold suggests that the maximum of a data distribution is in its direction, which is also an indication of the presence of a source signal. If two parameters – interval length and threshold – are appropriately chosen, the algorithm would stabilize with no further pruning of the network. At that point the algorithm would not only have converged to the true number of sources but also would have learned the true directions of the distributions. The mixing matrix can be recovered in a single step by unifying the source number estimation and the learning of the basis vectors.

3.3 Description of Algorithm

For applicability of our algorithm, the following assumption is made: $N > n\,(unknown)$. The algorithm will be presented by explaining the separate steps of the blockdiagram in figure 1.

- Initialize N independent vectors symmetrically according to $\mathbf{w}_i = \mathrm{e}^{\mathrm{j}\,\frac{\pi\, i}{N}}$, $i = 1, 2, \ldots, N$
- Set the parameters: Interval length Δ_{win}, threshold p_{thr}, learning rate parameters $(\eta_0,\, \eta_f,\, \tau)$
- Execute iteratively:
  - Start outer loop and initialize the counter $(n_C = 0)$
  - Apply the geometric algorithm to each data sample and increment the counter
  - Exit inner loop if $n_C = \Delta_{\text{win}}$
  - Discard vectors $\mathbf{w}_i$ if $p_i = \frac{f_i}{\sum f_i} \leq p_{\text{thr}}$
  - Abort outer loop if convergence is reached (and no more vectors discarded)

The interval length Δ_{win} and the threshold p_{thr} are key parameters of the algorithm, which are responsible for controlling the accuracy and the stability. Appropriate values can be found empirically. The algorithm delivers good results with a low initial value. It is slightly linearly increased after every iteration step as given by following relation:

$$p_{\text{thr}}(i) = p_{\text{thr}}(i-1) + (i-1)\Delta\, p_{\text{thr}}, \qquad i = 2, 3, \ldots \tag{5}$$

where $p_{\text{thr}}(i-1)$ is the threshold value of the previous iteration.

4 Simulation and Results

4.1 Performance Measure

To evaluate the proposed algorithm and to test the quality of the reconstruction, two measures are defined. To compare the quality of the matrix recovery, the *generalized crosstalking error* $E(\mathbf{A}, \mathbf{B})$ is used [10]. To analyze the source recovery step, the crosstalking error $E_1(\mathbf{C})$ of the correlation matrix $(\mathbf{C} = \mathrm{Cor}(\mathbf{s}(t), \hat{\mathbf{s}}(t)))$ of the original signals $\mathbf{s}(t)$ and the recovered signals $\hat{\mathbf{s}}(t)$, defined in [1] is calculated.

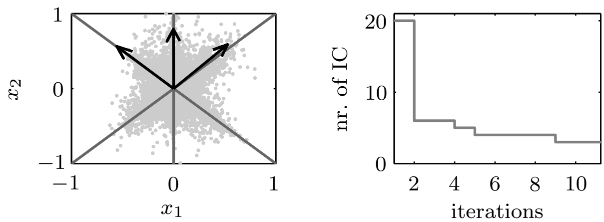
4.2 Test Setup

To demonstrate the algorithm, the simulation example of [4] with three speech sources and two mixtures is presented, assuming that the number of sources is not known a priori. The algorithm is initialized with following parameters: $N = 20$; $\Delta_{\text{win}} = 3833$; $\eta_0 = 0.1$, $\eta_f = 0.0002$, $\tau = 1000$ (learning rate parameters); $p_{\text{thr}} = 0.05$, $\Delta p_{\text{thr}} = 0.013$ (threshold parameters).

The original mixing matrix was:

$$\mathbf{A} = \begin{pmatrix} 0 & 0.7071 & 0.7071 \\ 1 & 0.7071 & -0.7071 \end{pmatrix}$$

The simulation result is shown in figure 2 after 46000 samples were presented to the algorithm. As it besomes evident from figure 2(b) the number of vectors are gradually reduced and the algorithm stabilized to the actual number of sources after some iterations. The learned basis vectors are shown in figure 2(a), matching almost exactly the underlying structure.



(a) Estimated vectors (b) Learning of vectors

Fig. 2. Evaluation of extended geometric algorithm

The recovered mixing matrix was:

$$\mathbf{B} = \begin{pmatrix} -0.0034 & 0.6970 & -0.7116 \\ 1.0000 & 0.7171 & 0.7026 \end{pmatrix}$$

which is very close to the original matrix. Also the cross-talking error $E(\mathbf{A}, \mathbf{B})$ is very close to zero, showing good estimation quality. Additionally, the sources were reconstructed using linear programming. The obtained correlation matrix was:

$$\mathbf{C} = \mathrm{Cor}(\mathbf{s}, \hat{\mathbf{s}}) = \begin{pmatrix} 0.8572 & 0.1574 & 0.1513 \\ 0.2216 & 0.9364 & -0.1456 \\ -0.2084 & 0.1210 & -0.9458 \end{pmatrix}$$

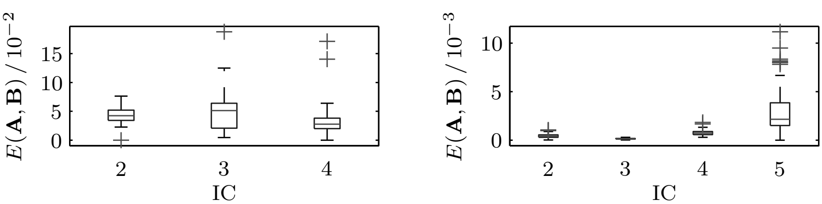
with $E_1(\mathbf{C}) = 2.2135$, which shows high correlation values between the original and the estimated sources and low cross-correlation values indicating a high signal independence. The complexity of the algorithm still remains the same. The approach requires little more samples, because of the additional learning process.

4.3 Results

In order to verify the effectiveness of the proposed method, it is necessary to assess the performance of the algorithm with several simulations under different conditions. In this section, the performance of our algorithm is evaluated with: (a) real speech signals, so that a comparison can be drawn with previous approaches using known source number; (b) sparse sources having nearly delta-like distributions, so that our algorithm can be compared with the results of [6].

For real speech signals, the values of the error index $E(\mathbf{A}, \mathbf{B})$ for 100 independent trials are shown as box plots in figure 3(a) for a varying number of sources. As we can recover up to four speech sources from two mixtures, the maximum number of sources or independent components (ICs) is taken as 4. The error index values were calculated after the algorithm stabilizes to the true number of sources. The median values of $E(\mathbf{A}, \mathbf{B})$ are also tabulated in table 1. Furthermore, it is possible that the algorithm does not stabilize exactly to the actual number of sources. This accounts for errors in the source number estimation. The error values (absolute) are shown in table 1 over 100 trials in each case. Even though the algorithm might have recovered some of the sources successfully, it is treated strictly as an error.

Additionally, the validity of our algorithm for sparse delta-like distributed sources treated in [6], is shown. For simulation purpose, artificial source signals with high sparse distributions were generated. Following the same simulation setup as above, the results illustrated in figure 3(b) were achieved, which represent the accuracy of the algorithm over 100 independent trials with a varying number of sources respectively. The median values of the generalized crosstalking error are also tabulated in table 1, which can be compared with other algorithms.



(a) Quality of estimation (b) Quality of estimation

Fig. 3. Evaluation of signals, 100 trials

Table 1. Generalized crosstalking error of speech signals (left table) and sparse signals (right table)

sources	$E(\mathbf{A}, \mathbf{B})$	errors
2	$4.246 \cdot 10^{-2}$	2
3	$5.129 \cdot 10^{-2}$	0
4	$2.792 \cdot 10^{-2}$	8
–	–	–

sources	$E(\mathbf{A}, \mathbf{B})$	errors
2	$4.213 \cdot 10^{-4}$	0
3	$1.507 \cdot 10^{-4}$	0
4	$7.310 \cdot 10^{-4}$	0
5	$2.153 \cdot 10^{-3}$	8

5 Conclusion

In this paper, a modification of a geometrical ICA algorithm was presented in order to address the problem of unknown number of in underdetermined BSS. The simulation demonstrates the efficiency of our algorithm to handle different types of sources. We can see that the variance of the generalized cross-talking error is quite small with no outliers in the real sense, which shows that once the algorithm stabilizes to actual source numbers, it always converges to the original mixing matrix. The number of errors in the estimation of the number of sources can be reduced further by judicious choice of parameters, thus increasing the reliability of our method. For sparse sources, the accuracy of our algorithm is extremely high with error index values almost zero, which means that the mixing matrix is perfectly recovered. We conclude that for sparsely distributed sources, our algorithm stabilizes quickly to actual number of sources in few iterations which extends the learning time for the vectors and improves accuracy and convergence. The algorithm is essentially a top-down approach where a large network is initialized and gradually pruned based on some criterion. The final structure of the network after convergence is what truly defines the actual model.

References

1. Amari, S., Cichocki, A., Yang, H.: A new learning algorithm for blind signal separation. In: Advances in neural information processing systems (1996)
2. Comon, P., et al.: Independent component analysis, a new concept? Signal processing 36(3), 287–314 (1994)
3. Hyvarinen, A., Karhunen, J., Oja, E. (eds.): Independent Component Analysis. Wiley Interscience, Hoboken (2001)
4. Lee, T.-W., Lewicki, M., Girolami, M., Sejnowski, T.: Blind source separation of more sources than mixtures using overcomplete representations. IEEE Signal Processing Letters 6(4), 87–90 (1999)
5. Lewicki, M., Sejnowski, T.: Learning overcomplete representations. Neural computation 12(2), 337–365 (2000)
6. Li, R., Tan, B.: Estimation of Source Signals Number and Underdetermined Blind Separation Based on Sparse Representation. In: Wang, Y., Cheung, Y.-m., Liu, H. (eds.) CIS 2006. LNCS (LNAI), vol. 4456, p. 943. Springer, Heidelberg (2007)
7. Mansour, A., Puntonet, C., Ohnishi, N.: A simple ICA algorithm based on geometrical approach. In: Sixth International Symposium on Signal Processing and its Applications (ISSPA 2001), pp. 9–12 (2001) (Citeseer)
8. Puntonet, C., Prieto, A.: An adaptive geometrical procedure for blind separation of sources. Neural Processing Letters 2(5), 23–27 (1995)
9. Theis, F., Jung, A., Puntonet, C., Lang, E.: Linear geometric ICA: Fundamentals and algorithms. Neural Computation 15(2), 419–439 (2003)
10. Theis, F., Lang, E., Puntonet, C.: A geometric algorithm for overcomplete linear ICA. Neurocomputing 56, 381–398 (2004)