ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 217-222

Geometric Overcomplete ICA

Fabian J. Theis, Elmar W. Lang

Institute of Biophysics
University of Regensburg, D-93040 Regensburg, Germany
email: fabian.theis@mathematik.uni-regensburg.de

Abstract. In independent component analysis (ICA), given some signal input the goal is to find an independent decomposition. We present an algorithm based on geometric considerations [11] to decompose a linear mixture of more sources than sensor signals. We present an efficient method for the matrix-recovery step in the framework of a two-step approach to the source separation problem. The second step — source-recovery — uses the standard maximum-likelihood approach.

1 Introduction

In overcomplete ICA more sources are mixed to less signals, and the goal is to recover the original signals. The ideas used in overcomplete ICA originally stem from coding theory, where the task is to find a representation of some signals in a given set of generators which often are more numerous than the signals, hence the term overcomplete basis. Olshausen and Fields first put these ideas into an information theoretic context decomposing natural images into an overcomplete basis [10]. Later, Harpur and Prager [4] and, independently, Olshausen [9] presented a connection between sparse coding and ICA in the quadratic case. Lewicki and Sejnowski [8] then were the first to apply these terms to overcomplete ICA, which was further studied and applied by Lee et al [7]. De Lathauwer et al [6] provided an interesting algebraic approach to overcomplete ICA of 3 sources and 2 mixtures by solving a system of linear equations in the third- and fourth-order cumulants, whereas Taleb [12] reduced the $n \times 2$ case to solving a partial differential equation in the second-order cumulants. Bofill and Zibulevsky [2] treated a special case ('delta-like' source distributions) of source signals after Fourier transformation. In this paper we generalize their approach for arbitrary supergaussian source distributions using a geometric matrix-recovery algorithm.

For $m, n \in \mathbb{N}$ let $\mathrm{Mat}(m \times n)$ be the $\mathbb{R}-$vectorspace of real $m \times n$ matrices, and $\mathrm{Gl}(n) := \{W \in \mathrm{Mat}(n \times n) \mid \det(W) \neq 0\}$ be the general linear group of $\mathbb{R}^n$. In the general case of linear blind source separation (BSS), a random vector $X : \Omega \to \mathbb{R}^m$ composed of **sensor signals** is given; it originates from an

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 217-222

independent random vector $S : \Omega \to \mathbb{R}^n$, which is composed of **source signals**, by mixing with a **mixing matrix** $A \in \mathrm{Mat}(m \times n)$, i.e. $X = AS$. Here Ω denotes a fixed probability space. Only the sensor signals are known, and the task is to recover both the mixing matrix A and the source signals S. We will assume that the mixing matrix A has full rank and any two different columns of A are linearly independent. The problem stated like this is ill-posed for the **overcomplete** case $(m > n)$, hence further restrictions will have to be made.

2 A Two Step Approach to the Separation

In the quadratic case it is sufficient to recover the mixing matrix A in order to solve the separation problem, because the sources can be reconstructed from A and X by inverting A. For the overcomplete case as presented here, however, after finding A in a similar fashion as in quadratic ICA (**matrix-recovery step**), the sources will be chosen from the $n - m$-dimensional affine vector space of the solutions of $AS = X$ using a suitable boundary condition (**source-recovery step**). Hence with this algorithm we follow a two step approach to the separation of more sources than mixtures; this two-step approach has been proposed recently by Bofill and Zibulevsky [2] for delta distributions. It contrasts to the single step separation algorithm by Lewicki and Sejnowski [8], where both steps have been fused together into the minimization of a single complex energy function. We show that our approach resolves the convergence problem induced by the complicated energy function, and, moreover, it reflects the quadratic case as special case in a very obvious way.

2.1 Matrix-Recovery Step

In the first step, given only the mixtures X, the goal is to find a matrix $A' \in \mathrm{Mat}(m \times n)$ with full rank and pairwise linearly independent columns such that there exists an independent random vector S' with $X = A'S'$.

For geometric matrix-recovery, we use a generalization of the geometric ICA algorithm [11]. Later on, we will restrict ourselves to the case of two-dimensional mixture spaces for illustrative purposes mainly. With high dimensional problems, however, geometrical algorithms need very many samples [5], hence seem less practical; but for now let $m > 1$ be arbitrary.

Let $S : \Omega \longrightarrow \mathbb{R}^2$ be an independent n-dimensional Lebesgue-continuous random vector describing the source pattern distribution; its density function is denoted by $\rho : \mathbb{R}^n \longrightarrow \mathbb{R}$. As S is independent, ρ factorizes into $\rho(x_1, \ldots, x_n) = \rho_1(x_1) \ldots \rho_n(x_n)$, with the marginal source density functions $\rho_i : \mathbb{R} \longrightarrow \mathbb{R}$.

As above, let X denote the vector of sensor signals and A the mixing matrix such that $X = AS$. A is assumed to be of full rank and to have pairwise linearly independent columns. Since we are not interested in dealing with scaling factors, we can assume that the columns in A have Euclidean norm 1. The **geometric learning algorithm** for symmetric distributions in its simplest form then goes as follows:

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 217-222

Pick $2n$ starting elements $w_1, w'_1, \ldots, w_n, w'_n$ on the unit sphere $S^{m-1} \subset \mathbb{R}^m$ such that w_i and w'_i are opposite each other, i.e. $w_i = -w'_i$ for $i = 1, \ldots, n$, and such that the w_i are pairwise linearly independent vectors in $\mathbb{R}^m$. Often, these w_i are called **neurons** because they resemble the neurons used in clustering algorithms and in Kohonen's self-organizing maps. If $m = 2$, one usually takes the unit roots $w_i = \exp(\frac{n-1}{n}\pi i)$. Furthermore fix a learning rate $\eta : \mathbb{N} \longrightarrow \mathbb{R}$ such that $\eta(n) > 0$, $\sum_{n\in\mathbb{N}} \eta(n) = \infty$ and $\sum_{n\in\mathbb{N}} \eta(n)^2 < \infty$. Then iterate the following step until an appropriate abort condition has been met:

Choose a sample $x(t) \in \mathbb{R}^m$ according to the distribution of X. If $x(t) = 0$ pick a new one – note that this case happens with probability zero since the probability density function ρ_X of X is assumed to be continuous. Project $x(t)$ onto the unit sphere to yield $y(t) := \frac{x(t)}{|x(t)|}$. Let i be in $\{1, \ldots, n\}$ such that w_i or w'_i is the neuron closest to y with respect to the Euclidean metric. Then set

$$w_i(t+1) := \pi(w_i(t) + \eta(t)\,\mathrm{sgn}(y(t) - w_i(t))),$$

where $\pi : \mathbb{R}^m \setminus \{0\} \longrightarrow S^{(m-1)}$ denotes the projection onto the $(m-1)$-dimensional unit sphere $S^{(m-1)}$ in R^m, and

$$w'_i(t+1) := -w_i(t+1).$$

All other neurons are not moved in this iteration.

Similar to the quadratic case, this algorithm may be called **absolute winner-takes-all learning**. It resembles Kohonen's competitive learning algorithm for self-organizing maps with a trivial neighbourhood function (**0-neighbour algorithm**) but with the modification that the step size along the direction of a sample does not depend on distance, and that the learning process takes place on $S^{(m-1)}$ not in $\mathbb{R}^{(m-1)}$.

2.2 Source-Recovery Step

Using the results given above, we can assume that an estimate of the original mixing matrix A has been found. We are therefore left with the problem of reconstructing the sources using the sensor signals X and the estimated matrix A. Since A has full rank, the equation $x = As$ yields the $n-m$-dimensional affine vectorspace $A^{-1}\{x\}$ as solution space for s. Hence, if $n > m$ the source-recovery problem is ill-posed without further assumptions. An often used [8] [2] assumption can be derived using a maximum likelihood approach, as will be shown next.

The problem of the source-recovery step can be formulated as follows: Given a random vector $X : \Omega \longrightarrow \mathbb{R}^m$ and a matrix A as above, find an independent vector $S : \Omega \longrightarrow \mathbb{R}^n$ satisfying an assumption yet to be found such that $X = AS$. Considering $X = AS$, i.e. neglecting any additional noise, X can be imagined to be determined by A and S. Hence the probability of observing X given A and S can be writen as $P(X|S, A)$. Using Bayes Theorem the

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 217-222

posterior probability of S is then

$$P(S|X,A) = \frac{P(X|S,A)P(S)}{P(X)},$$

the probability of an event of S after knowing X and A. Given some samples of X, a standard approach for reconstructing S is the **maximum-likelihood algorithm** which means maximizing this posterior probability after knowing the **prior probability** $P(S)$ of S. Using the samples of X one can then find the most probable S such that $X = AS$. In terms of representing the observed sensor signals X in a basis $\{Ae_i\}$ this is called the most probable decomposition of X in terms of the overcomplete basis of $\mathbb{R}^m$ given by the columns of A.

Using the posterior of the sources $P(S|X,A)$, we can obtain an estimate of the unknown sources by solving the following relation

$$\begin{aligned} S &= \arg\max_{X=AS} P(S|X,A) \\ &= \arg\max_{X=AS} P(X|S,A)P(S) \\ &= \arg\max_{X=AS} P(S). \end{aligned}$$

In the last equation, we use that X is fully determined by S and A, and hence $P(X|S,A)$ is trivial. Note that of course the maximum under the constraint $X = AS$ is not necessarily unique.

If $P(S)$ is assumed to be Laplacian that is $P(S_i)(t) = a\exp(-|t|)$, then we get $S = \arg\min_{X=AS}|S|_1$, where $|v|_1 := \sum_i |v_i|$ denotes the 1-norm. We can show that the solution S is unique, which may not be the case for other norms; The general algorithm for the source-recovery step therefore is the maximization of $P(S)$ under the constraint $X = AS$. This is a linear optimization problem which can be tackled using various optimization algorithms [3].

In the following we will assume a Laplacian prior distribution of S which is characteristic of a sparse coding of the observed sensor signals. In this case, the minimization has a nice visual interpretation, which suggests an easy to perform algorithm: The source-recovery step consists of minimizing the 1-norm $|s_\lambda|_1$ under the constraint $As_\lambda = x_\lambda$ for all samples x_λ. Since the 1-norm of a vector can be pictured as the length of a path with parallel steps to the axes, Bofill and Zibulevsky call this search **shortest-path decomposition** — indeed, one can show that s_λ represents the shortest path to x_λ in $\mathbb{R}^m$ along the lines given by the matrix columns Ae_i of A.

3 Experimental Results

In this section, we give a demonstration of the algorithm. The calculations have been performed on a AMD Athlon 1 GHz computer using Matlab and took no more than one minute at most.

We mixed three speech signals to two sensor signals as shown in figure 1, left side and middle. After 10^5 iterations, we found a mixing matrix with satisfactorily small minimal column distance 0.1952 to the original matrix, and

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 217-222

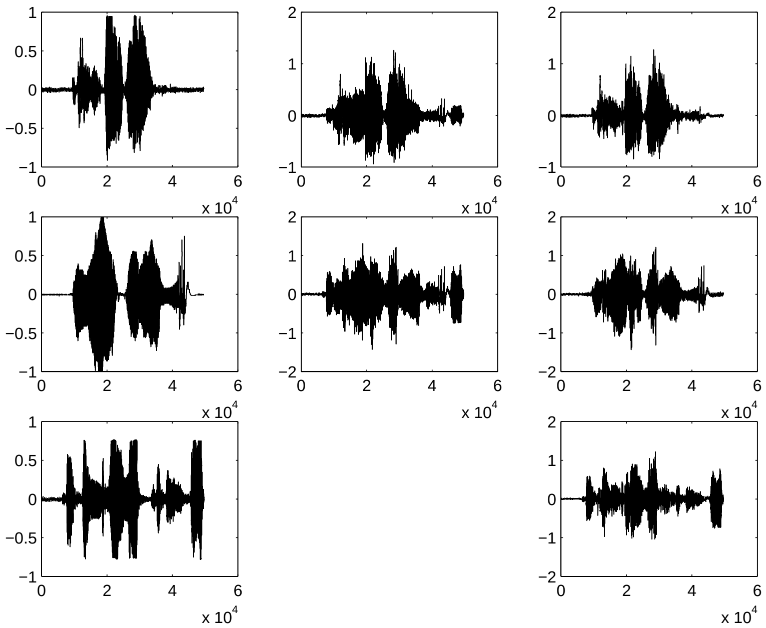

Figure 1: Example: The three sources to the left, the two mixtures in the middle, and the recovered signals to the right. The speech texts were 'californication', 'peace and love' and 'to be or not to be that'. The signal kurtosis were 8.9, 7.9 and 7.4.

after source-recovery, we calculate a correlation of estimated and original source signals with a crosstalking error [1] $E_1(\mathrm{Cor}(S, S')) = 3.7559$. In figure 1 to the right, the estimated source signals are shown. One can see a good resemblance to the original sources, but the crosstalking error is still rather high.

We suggest that this is a fundamental problem of the source-recovery step, which, to our knowledge, using the above probabilistic approach cannot be improved any further. To explore this aspect further, we performed an experiment using the source recovery algorithm to recover three Laplacian signals mixed with

$$A_\alpha = \begin{pmatrix} 1 & \cos(\alpha) & \cos(2\alpha) \\ 0 & \sin(\alpha) & \sin(2\alpha) \end{pmatrix},$$

where we started the algorithm already with the correct mixing matrix. We then compared the crosstalking error $E(\mathrm{Cor}(S, S'))$ of the correlation matrix of the recovered signals S' with the original ones (S) for different angles $\alpha \in [0, \frac{\pi}{2}]$. We found that the result is nearly independent of the angle, which makes sense because one can show that the shortest-path-algorithm is invariant under coordinate transformations like A_α. This experiment indicates that there might be a general border on how good sources can be recovered in overcomplete settings.

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 217-222

4 Conclusion

We have presented a two-step approach to overcomplete blind source separation. First, the original mixing matrix is approximated using the geometry of the mixture space in a similar fashion as geometric algorithms do this in the quadratic case. Then the sources are recovered by the usual maximum-likelihood approach with a Laplacian prior.

For further research, two issues will have to be dealt with. On the one hand, the geometric algorithm for matrix-recovery will have to be improved and tested, especially for higher mixing dimensions m. We currently experiment with an overcomplete generalization of the quadratic 'FastGeo' algorithm [5], a histogram-based geometric algorithm, which looks more stable and also faster. On the other hand, the question if there is a natural information theoretic barrier of how well data can be recovered has to be treated.

References

[1] S. Amari, A. Cichocki, and H.H. Yang. A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems*, 8:757–763, 1996.

[2] P. Bofill and M. Zibulevsky. Blind separation of more sources than mixtures using sparsity of their short-time fourier transform. *Proc. of ICA 2000*, pages 87–92, 2000.

[3] S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *Technical report, Dept. Stat., Stanford Univ, Stanford, CA*, 1996.

[4] G.F. Harpur and R.W. Prager. Development of low-entropy coding in a recurrent network. *Network*, 7:277–284, 1996.

[5] A. Jung, F.J. Theis, C.G. Puntonet, and E.W. Lang. Fastgeo - a histogram based approach to linear geometric ICA. *Proc. of ICA 2001*, pages 418–423, 2001.

[6] L. De Lathauwer, P. Comon, B. De Moor, and J. Vandewalle. ICA algorithms for 3 sources and 2 sensors. *IEEE Sig. Proc. Workshop on Higher-Order Statistics, June 14-16, 1999, Caesarea, Israel*, pages 116–120, 1999.

[7] T. Lee, M.S. Lewicki, M. Girolami, and T.J. Sejnowski. Blind source separation of more sources than mixtures using overcomplete representations. *IEEE Signal Processing Letters*, 6(4):87–90, 1999.

[8] M.S. Lewicki and T.J. Sejnowski. Learning overcomplete representations. *Neural Computation*, 1998.

[9] B.A. Olshausen. Learning linear, sparse, factorial codes. Technical Report AIM-1580, 1996.

[10] B.A. Olshausen and D. Field. Sparse coding of natural images produces localized. Technical Report CCN-110-95, 1995.

[11] C.G. Puntonet and A. Prieto. An adaptive geometrical procedure for blind separation of sources. *Neural Processing Letters*, 2, 1995.

[12] A. Taleb. An algorithm for the blind identification of n independent signals with 2 sensors. *To appear in ISSPA 2001, Kuala Lumpur (Malaysia)*, 2001.