Digitale Bibliothek Braunschweig
Der Publikationsserver der TU Braunschweig

# Independent Component Analysis using a Geometric Approach

Thomas Pitschel

Braunschweig: Institut für Mathematische Stochastik, 2013

**Elektronisch veröffentlicht am 13. 06.2013 in der**
**Digitalen Bibliothek Braunschweig**
**Publikationsserver des Wissenschaftsstandortes Braunschweig**

**unter: http://www.digibib.tu-bs.de/?docid=** 00051759

UB Braunschweig

# Independent Component Analysis using a Geometric Approach

Thomas Pitschel
8th June 2013

***Abstract*—For the problem of independent component analysis we propose an alternative solution to determining $n$ independent sources from $n$ channels using a geometric approach. The resulting algorithm inherits many properties of the then applicable simplex algorithm, namely computational tractability even in high dimensions ($n \geq 100$), and yields demixing estimates comparable to standard algorithms (e.g. FastICA) for extremely small sample sizes ($N = 50$). Our results are corroborated by demixing of a real world ECG data set from the PhysioNet database.**

***Index Terms*—independent components, blind source separation, linear programming**

## I. Introduction

Given $N$ samples of $n$ observed signals, $(y_{i,j})_{j=1..n,i=1..N}$, or $(\mathbf{y}_i)_{i=1..N}$, "Independent component analysis" (ICA) describes the problem of determining usually up to $n$ source signals $(x_{i,j})_{j=1..n,i=1..N}$, or $(\mathbf{x}_i)_{i=1..N}$, such that the observed signals represent a linear combination of the source signals, i.e.

$$y = Ax,$$

and the source signals are independent. The matrix $W$ transforming the observed signals to the source signals, i.e. $x = Wy$, is called demixing matrix. For obvious reason, source and observed signals can be assumed to have zero mean. One can show (see appendix) that a possible solution $W$ is determined up to scaling (including sign choice) of the source signals only. Further, no two source signals having Gaussian distribution with the same variance can be separated based on statistical properties; thus, it is often, and also here, assumed that at most one of the source signals is Gaussian. While ICA can be formulated also for complex signals, here we consider only real-valued signals.

The problem setting has received thorough attention by the signal processing community over the past two decades, with applications ranging from biomedical diagnosis/analysis [1], astrophysical image analysis [2] to acoustic engineering [3], e.g. in the automotive domain. The solution was successfully embarked upon with a series of papers, among them for example Bell and Sejnowksi [4]. The first computationally efficient algorithm for ICA was given by Hyvärinen and Oja [5], applying a neural network inspired fixed point scheme targetting a value comparable to an inherent entropy, and applying an approximation to the expectation of non-linear expression. Further solutions around that time were given by [6] (based on an information maximization principle). The problem was then examined in the setting of non-linear mixing [7], and in the scenario of extracting more sources than observed signals are available. More recent papers in the area in the past decade have focused on specialized settings such as positive sources ( [8] [9]). In even more recent years, the problem was examined in the more general framework of latent variable analysis ("LVA"), and recovering source signals has extended its attention to using signal-inherent properties (moment statistics and correlations) [10]. The algorithm seen as the standard method for tackling the problem of signal separation, at least before the advent of moment-based methods, is FastICA [11] (based on the fixed point scheme).

Here we present a novel algorithm for the original problem setting. Our algorithm is distinguished from algorithms such as FastICA in using different elements of information from the set of observed signals. Concretely, while FastICA employs expectations taken over the wholity of observed samples, our approach will effectively use only information contained in samples constituting yet to be described convex hulls ("outer samples"). The geometric approach further has the consequence of making a choice of a non-linearity unnecessary. Further, pre-whitening can be omitted since the method will determine suitable coefficients (with implicit whitening) also on non-white observed signal sets.

http://www.digibib.tu-bs.de/?docid=00051759 13/06/2013

## II. The algorithm

The core of the approach starts with the observation that in a plot representation of two independent uniformly distributed sources, the sample set constitutes a set of points enveloped by a square. Similarly, a likewise plot for $m$ independent uniformly distributed sources will result in a point set resembling an $m$-dimensional hypercube. Assume we had already found the correct demixing matrix $W$, i.e. all components of the vector $x$ in

$$x = Wy$$

are independent, a suitable transformation of those components with their respective marginal distributions would result in $m$ $[0,1]$-uniformly distributed signals. We can suppose that as long as a current estimate of the demixing matrix is not fully, but at least closely, resembling the true demixing matrix, the obtained demarginalized distributions will be close to the aforementioned uniform distributions. Our approach is thus motivated by the endeavour to successively refine the demixing matrix by alternatingly (1) determining demarginalized signals based on the current demixing estimate, and (2) refine the demixing estimate based on the obtained demarginalized sample set, using the above geometrical interpretation. For (2) we inherently use the assumption that the distortion of the $m$-dimensional sample set evoked by the still remaining (linear) mixing is translating itself into the demarginalized sample set, which is qualitatively reasonable due to the monotonicity of the demarginalization.

Our algorithm thus proceeds as follows:

1) Initialize the current demixing matrix with $W^{[0]} = 1$.
2) Based on the current estimate $W^{[l]}$ of the demixing, determine the demixed signals: $x^{[l]} = W^{[l]}y$
3) Demarginalize the current demixed signals[1]: $\hat{x}^{[l]} = \tilde{F}^m(x^{[l]})$
4) Determine a suitable correction to the demixing matrix using the $\hat{x}^{[l]}$. Call the new demixing matrix $W^{[l+1]}$.
5) If a chosen maximum number of cycles is reached, or another criterion on the independence of the demixing result indicates, stop the iteration. Otherwise loop continuing at 2.

[1]Here $F$ shall denote the joint distribution function of $x^{[l]}$, the $F_j^m$ the $n$ marginal distributions of the $n$ elements of the vector $x^{[l]}$, the $\tilde{F}_j^m := 2F_j^m - 1$ (i.e. the marginal distributions scaled to $[-1,1]$), and $\tilde{F}^m(x^{[l]})$ the element-wise application of these marginal distributions.

The remaining problem is thus to determine the correction to the linear transformation posed by the remaining mixing "component" ("residual mixing matrix"), given a distorted-hypercube resembling object. This object is in fact resembling a parallelepiped.

The approach taken is to determine the bounding hyperplane for each dimension (each coordinate) separately. Concretely, letting $j$ be the coordinate of the direction in which to find the hyperplane, and $p$ being its normal vector pointing out of the parallelepiped, the solution for $p$ is given by the LP (linear program)

$$\begin{aligned} &\max p_j \\ \text{s.t. } &x_i^T p \leq 1 \\ -&x_i^T p \leq 1 \\ &p \text{ free.} \end{aligned}$$

where the conditions in $x_i$ are understood to hold for all samples $i = 1..N$, and the equation of the hyperplane is given by $p^T x = \pm 1$ (see appendix). Since $p_j = e_j^T p$ and $p$ is free, and letting $A^T := [x_1, ..., x_N, -x_1, ..., -x_N]$, the dual problem (DP) is given by

$$\begin{aligned} &\min 1^T u \\ \text{s.t. } &A^T u = e_j \quad (**) \\ &u \geq 0 \ . \end{aligned}$$

A suitable method for finding an initial solution for DP is to first solve $(**)$ using the first $n$ independent columns in $A^T$, then proceed with basis changes such that successively any component of $u$ still being negative is increased to at least zero (others being held non-negative). (Since a solution is guaranteed to exist, this must always be possible.)

Finally, the desired normal vector $p$ is calculated as follows: For the indices in $B$ for which $u_i$ are non-zero, the corresponding inequalities in (LP) hold with equality, so in compact notation $p^T(A^T)_B = 1^T$ (where subscript $B$ selects the basis columns of $A^T$). Then follows $p^T = 1^T(A^T)_B^{-1}$, so the normal vector is simply the sum of the rows of the inverse of the basis matrix in (DP) at the optimal solution (which is available from the Simplex algorithm).

This procedure will have yielded the normal vector for the hyperplane in one direction. Let $p_j, j = 1..n$, denote the normal vectors for all $n$ coordinates, respectively. The matrix $P$ with $P = (p_1^T, ..., p_n^T)^T$ (i.e. the normal vectors assembled

http://www.digibib.tu-bs.de/?docid=00051759 13/06/2013

in rows) is related to the distorting transformation matrix (= residual mixing matrix) $A$ by[2] $PA = 1$.

## References


[1] Tzyy-Ping Jung. http://cnl.salk.edu/~jung/ica.html.

[2] Maria Funaro, Erkki Oja, and Harri Valpola. Independent component analysis for artefact separation in astrophysical images. *Neural Networks*, 16(3–4):469–478, 2003.

[3] David K. Campbell. Adaptive beamforming using a microphone array for hands-free telephony. Technical report, Virginia Polytechnic Institute and State University, 1999.

[4] A. J. Bell and T. J. Sejnowski. An information maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[5] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[6] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems*, pages 757–763. MIT Press, 1996.

[7] A. Hyvärinen and P. Pajunen. Nonlinear independent component analysis: Existence and uniqueness results. *Neural Networks*, 12(3):429–439, 1999.

[8] M. D. Plumbey. Algorithms for nonnegative independent component analysis. *IEEE Trans. Neural Networks*, 14(3):534–543, 2003.

[9] E. Oja and M. D. Plumbey. Blind separation of positive sources by globally convergent gradient search. *Neural Comput.*, 16(9):1811–1825, 2004.

[10] A. L. F. de Almeida, X. Luciani, and P. Comon. Fourth-order CONFAC decomposition approach for blind identification of underdetermined mixtures. *EUSIPCO*, pages 290–294, 2012.

[11] A. Hyvärinen. FastICA algorithm. http://www.cs.helsinki.fi/u/ahyvarin/papers/fastica.shtml.


## Appendix

*Scale, sign and permutation invariance*

Let $A$ be the true mixing matrix, and thus $y = Ax$, where $x$ are the true source signals. Let the variances of the source signals be $\sigma_j^2$, and let $D = \text{diag}(\sigma_j)$. Let $P$ be an arbitrary permutation matrix. Then $y = (ADP)(P^{-1}D^{-1}x) =: A'x'$, where now $x'$ are signals with normalized variance and $A'$ is another valid mixing matrix. This means that given $y$ alone, it is impossible to determine which of all possible mixing matrices (consistent with the observed signals) was the true mixing matrix.

*Determining the parallelepiped*

Let $x_i \in \mathbb{R}^n$, $i = 1..N$, be a given point cloud with $x_i \in [-1, 1]^n$. We want to find hyperplanes $(p_j, c_j)$, $j = 1..n$, such that

$$\begin{aligned} p_j^T x_i &\leq c_j \quad \text{and} \\ -p_j^T x_i &\leq c_j \end{aligned}$$

for all $i$, and the $c_j \geq 0$ are minimal. (The $p_j$ point *out of* the described parallelepiped. The point cloud is enveloped from both a "bottom" and a "top" hyperplane with common normal vector $p_j$.) Note that for each $j$ this problem can be solved independently.

Solution using Linear programming: First note that the above can only be well-defined when fixing the length of the directional vector $p_j$ to counter the invariance under simultaneous scaling of $p_j$ and $c_j$. Assume this to be done by requiring $p_{j,j} = 1$, which is always possible since we choose to yield an *outward* pointing normal vector, and a corresponding feasible solution must exist with $p_{j,j} > 0$. At the optimal solution, for the rescaled normal vector $\tilde{p}_j := p_j / c_j$ the $j$-th component is maximal among all feasible solutions. Thus, by dividing the inequalities, the optimization problem may be restated as

$$\begin{aligned} \max \tilde{p}_{j,j} \\ \text{s.t. } x_i^T \tilde{p}_j &\leq 1 \\ -x_i^T \tilde{p}_j &\leq 1 \\ \tilde{p}_j &\text{ free.} \end{aligned}$$

The tilde notation and the coordinate index $j$ will be omitted in the sequel.

*Example demixing using ECG data*

Figure 1 and 2 show the demixing using the presented algorithm on electro cardiogram (ECG) signals taken from the PhysioNet database. The recording is comprised of channels from fifteen electrodes sampled at 1kHz with a resolution of 12bits, and has a length of 20secs, of which 10secs were used. Figure 1 shows the original signals, figure 2 the demixing result after 500 iterations.

[2]Consider, as example, the $A$-transformation of points on the plane $x_1 = 1$. It must hold $p_1^T Ax = 1$ for all these points. On the other hand, the vector $e_1$ is transformed such that for all $j \neq 1$, $p_j Ae_1 = 0$. Thus, analogously, $p_j^T Ae_{j'} = \delta_{j,j'}$.

http://www.digibib.tu-bs.de/?docid=00051759 13/06/2013

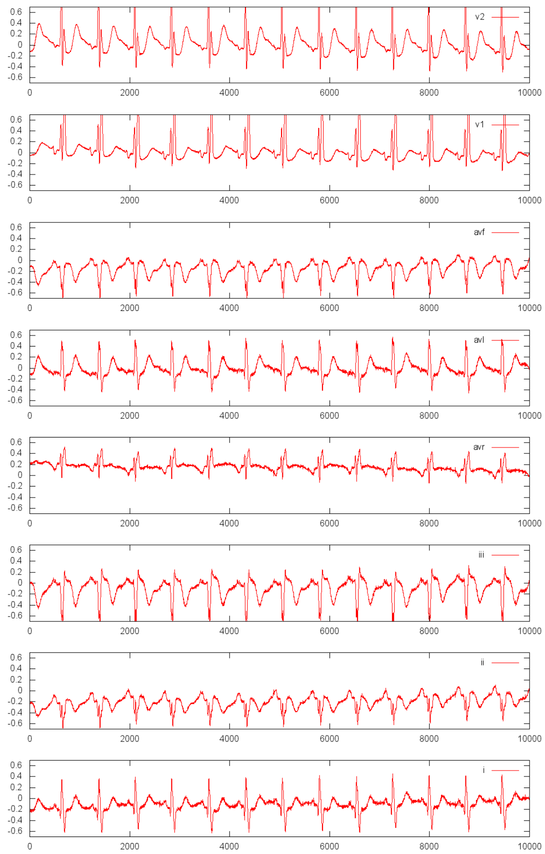


Fig. 1: Voltage at eight of 15 electrodes taken over the period of 10 seconds.

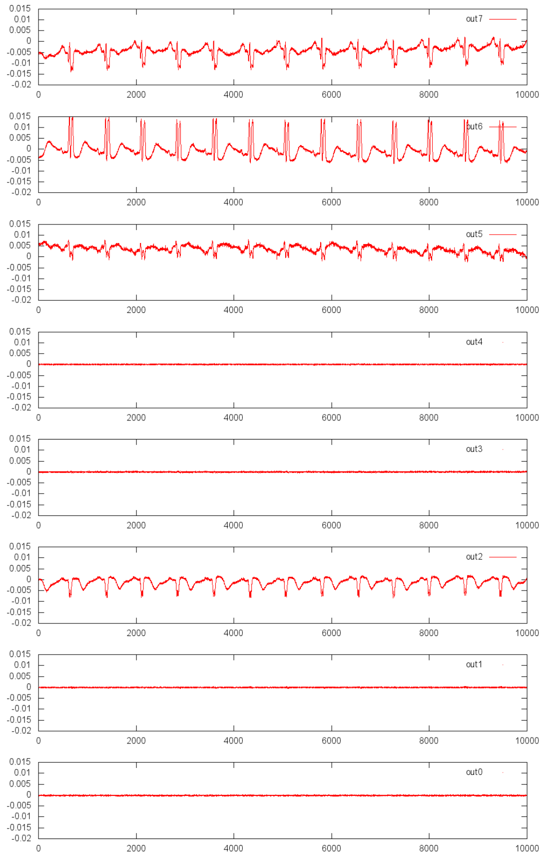


Fig. 2: The independent components derived by applying the presented algorithm.

http://www.digibib.tu-bs.de/?docid=00051759 13/06/2013