FASTGEO - A HISTOGRAM BASED APPROACH TO LINEAR GEOMETRIC ICA

Andreas Jung[*,1]*, Fabian J. Theis*[2] *, Carlos G. Puntonet*[3]*, Elmar W. Lang*[2]

[1]Institute for Theoretical Physics, University of Regensburg
[2]Institute of Biophysics, University of Regensburg, 93040-Regensburg, Germany
[3]Dept. of Architecture and Computer Technology, University of Granda (Spain)
email: *Andreas.Jung@physik.uni-regensburg.de*

ABSTRACT

Guided by the principles of neural geometric ICA, we present a new approach to linear geometric ICA based on histograms rather than basis vectors. Considering that the learning process converges to the medians and not the maxima of the underlying distributions restricted to the receptive fields of the corresponding neurons, we observe a considerable improvement in separation quality of different distributions and a sizable reduction in computational cost by a factor of 100 at least. We further explore the accuracy of the algorithm depending on the number of samples and the choice of the mixing matrix. Finally we discuss the problem of high dimensions and how it can be treated with geometrical ICA algorithms.

1. INTRODUCTION

Independent component analysis (ICA) describes the process of finding statistically independent data within a given random vector. In blind source separation (BSS) one furthermore assumes that the given vector has been mixed using a fixed set of independent sources. The advantage of applying ICA algorithms to BSS problems in contrast to correlation-based algorithms is the fact that ICA tries to make the output signals as independent as possible by also including higher-order statistics.

Since the first introduction of the ICA method by Jutten and Herault [1] various algorithms have been proposed to solve the blind source separation problem [2] [3] [4]. Most of them are based on information theory, but recently geometrical algorithms have gained some attention due to their relatively easy implementation. They were first proposed in [5] [6], and since then have been successfully used for separating real world data [7] [8].

As shown in [9] the geometric update step requires the signum function in the following way:

$$w_i(t) = w_i(t-1) + \eta(t)\,\mathrm{sgn}(x^\nu - w_i(t-1)). \qquad (1)$$

*Thanks to the Deutsche Forschungsgemeinschaft (DFG) for financial support.

Then the w_i converge to the medians in their receptive field. Note that the medians do not necessarily coincide with the maxima (α_1, α_2) of the mixed density distribution on the sphere as shown in figure 1. Therefore, in general, any algorithm searching for the maxima of the distribution will not find the medians, which are the images of the unit vectors after mixing [9]. However with special restrictions to the sources (identical super-gaussian distribution of each component, as for example speech signals), the medians correspond to the maxima [10].

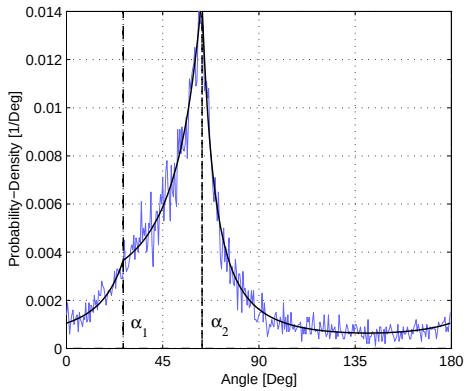

Fig. 1. Projected density distribution on the sphere of a mixture of two Laplacian signals with different variances, with the mixture matrix mapping the unit vectors e_i to $(\cos\alpha_i, \sin\alpha_i)$ for $i = 1, 2$. (dark line = theoretical density function, gray line = histogram of a mixture of 10.000 samples)

2. THEORY

Given an independent random vector $S : \Omega \longrightarrow \mathbb{R}^n$, which will be called **source vector** with zero mean and symmetric distribution, where Ω is a fixed probability space, and $A \in \mathrm{Mat}(n \times n; \mathbb{R})$ is a quadratic invertible matrix, we call the random variable $X := A \cdot S$ the **mixed vector**. The goal

of linear ICA is to recover the sources and the **mixing matrix** A from the given mixture X. At first, we will restrict ourselves to the two-dimensional case, so let $n = 2$.

So far in geometric ICA, mostly "neural" algorithms have been applied [5] [6]. The idea is to consider the projected random variable $Y := \pi \circ X = \pi(X)$ where

$$\begin{aligned} \pi : \mathbb{R}^2 \setminus \{0\} &\longrightarrow S^1 := \{x \in \mathbb{R}^2 \,||x| = 1\} \\ &\longrightarrow [0, 2\pi) \longrightarrow [0, \pi) \end{aligned} \tag{2}$$

denotes the projection of $\mathbb{R}^2$ onto the unit sphere S^1, then taking the angle and finally mapping modulo π . Two arbitrary starting values $w_1(0), w_2(0) \in [0, \pi)$ are chosen (usually the angles of the two unit vectors e_1 and e_2) and are then subjected to the following update rule:

$$w_i(t) = w_i(t-1) + \eta(t)\,\mathrm{sgn}(y_t - w_i(t-1)) \tag{3}$$

Here y_t denotes a sample of the distribution of Y, and i is chosen such that the distance of $w_i(t)$ to y_t modulo π is smaller than the distance for any other $w_j(t)$, $j \neq i$. $\eta(t)$ is the learning rate parameter which has to converge to 0. Hence it makes sense to define the **receptive field** of the "neuron" $w_i(t)$ to be

$$F_i(t) := \{x \in [0, \pi) | x \text{ closest (modulo } \pi) \text{ to } w_i(t)\}. \tag{4}$$

Note that since we are in two-dimensions the length of $F_i(t)$ is always $\frac{\pi}{2}$.

As shown in [9], after convergence the neurons $w_i(\infty)$ satisfy the geometric convergence condition:

Definition 2.1. *Two angles $l_1, l_2 \in [0, \pi)$ satisfy the **Geometric Convergence Condition (GCC)** if l_i is the median of Y restricted to its receptive field F_i for $i = 1, 2$.*

Let the mixing matrix A be given as follows:

$$A := \begin{pmatrix} \cos(\alpha_1) & \cos(\alpha_2) \\ \sin(\alpha_1) & \sin(\alpha_2) \end{pmatrix}. \tag{5}$$

Then the vectors $(\cos(\alpha_i), \sin(\alpha_i))^\top$ satisfy the GCC; hence we may assume that the above algorithm converges to these two solutions, and therefore we have recovered A and the sources S. The distribution on the sphere together with the angles α_i is depicted in figure 2.

3. HISTOGRAM BASED ALGORITHM: FASTGEO

Using the geometric convergence condition (GCC), we know that the neurons will converge to the medians in their receptive fields. This enables us to compute these positions directly using a search on the histogram of Y, which reduces the computation time by a factor of about 100 or more. In the **FastGeo-algorithm** we scan through all different receptive fields and test GCC. In practice this means discretizing

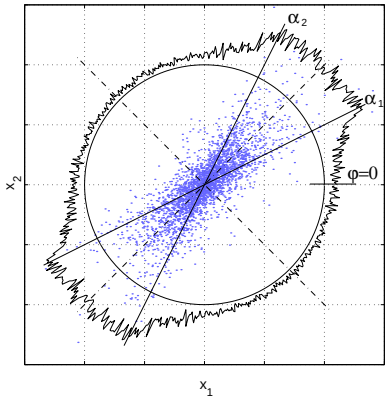

Fig. 2. Example of a two-dimensional scatter plot of a mixture of two Laplacian signals with identical variances. Dash-dotted lines show borders of the receptive fields.

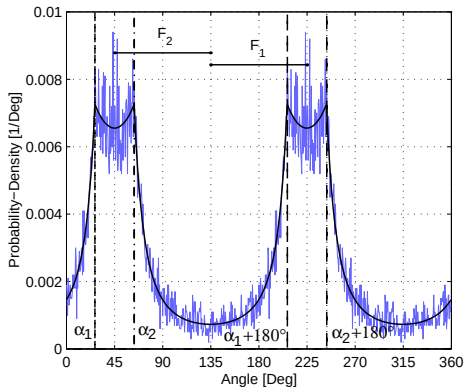

Fig. 3. Probability density function f_Y of Y from figure 2 with the mixing angles α_i and their receptive fields F_i for $i = 1, 2$.

the distribution f_Y of Y using a given bin-size $\beta > 0$ and then testing the π/β different receptive fields. The algorithm will be formulated more precisely in the following:

For simplicity let us assume that the cumulative distribution F_Y of Y is invertible – this means that F_Y is nowhere constant. Define a function

$$\begin{aligned} \mu : [0, \pi) &\longrightarrow \mathbb{R} \\ \varphi &\longmapsto \frac{l_1(\varphi) + l_2(\varphi)}{2} - (\varphi + \frac{\pi}{2}) \end{aligned} \tag{6}$$

where

$$l_i(\varphi) := F_Y^{-1}\left(\frac{F_Y(\varphi + i\frac{\pi}{2}) + F_Y(\varphi + (i-1)\frac{\pi}{2})}{2}\right) \tag{7}$$

is the median of $Y | [\varphi + (i-1)\frac{\pi}{2}, \varphi + i\frac{\pi}{2}]$ in $[\varphi + (i-1)\frac{\pi}{2}, \varphi + i\frac{\pi}{2}]$ for $i = 1, 2$.

Lemma 3.1. *Let φ be a zero of μ in $[0,\pi)$. Then the $l_i(\varphi)$ satisfy the GCC.*

Proof. By definition,

$$\left[\frac{l_1(\varphi)+l_2(\varphi)}{2}-\frac{\pi}{2},\frac{l_1(\varphi)+l_2(\varphi)}{2}\right] \tag{8}$$

is the receptive field of $l_1(\varphi)$. Since $\mu(\varphi)=0$, the starting point of the above interval is φ, because

$$\varphi=\frac{l_1(\varphi)+l_2(\varphi)}{2}-\frac{\pi}{2}. \tag{9}$$

Hence we have shown that the receptive field of $l_1(\varphi)$ is $[\varphi,\varphi+\frac{\pi}{2}]$, and by construction $l_1(\varphi)$ is the median of Y restricted to the above interval. The claim for $l_2(\varphi)$ then follows. □

Algorithm 3.2 (FastGeo). *Find the zeros of μ.*

μ always has at least two zeros which represent the stable and the unstable fixpoint of the neural algorithm. In practice we extract the fixpoint which then gives the proper demixing matrix A^{-1} by picking φ_0 such that

$$f_Y(l_1(\varphi_0))+f_Y(l_2(\varphi_0)) \tag{10}$$

is maximal. For unimodal and super-gaussian source distributions this results in a the stable fixpoint (see conjecture 6.7 in [9]). For sub-gaussian sources choosing φ_0, with

$$f_Y(l_1(\varphi_0))+f_Y(l_2(\varphi_0)) \tag{11}$$

minimal, induces the demixing matrix. Alternatively, using the result for the super-gaussian case, note that for unimodal distributions

$$\frac{l_1(\varphi_0))+l_2(\varphi_0))}{2}=\varphi_0+\frac{\pi}{2} \tag{12}$$

and

$$\frac{l_1(\varphi_0))+l_2(\varphi_0))-\pi}{2}=\varphi_0 \tag{13}$$

are the desired α_i.

4. ACCURACY

In this section we want to consider the dependence of the FastGeo-algorithm on the number of samples after the bin-size β has been fixed. As seen in the previous section, the accuracy of the histogram based algorithm then *only* depends on the distribution of the samples X respectively Y i.e. we can estimate the error made by approximating the mixing matrix A by a finite number of samples. In the following we will give some results of test-runs made with this algorithm.

When choosing two arbitrary angles $\alpha_i \in [0,\pi), i=1,2$ for the mixing matrix A, we define α as the distance between these two angles modulo $\frac{\pi}{2}$. This will give us an angle in the range between 0 and $\frac{\pi}{2}$ respectively $0°$ and $90°$.

First let us consider the accuracy of the recovered angles $\Delta\alpha=|\alpha_i-\alpha_i^{\text{recovered}}|$, when varying the angle α for a fixed number of samples. Choosing a mixture of two Laplacian source signals with identical variances, figure 4 shows a nearly linear decrease of the error $\Delta\alpha$ with decreasing α. The 95% confidence interval decreases similar to the standard deviation, i.e. both parameters provide a good estimate for the error. Note that the mean of the error $\Delta\alpha$ is very close to zero.

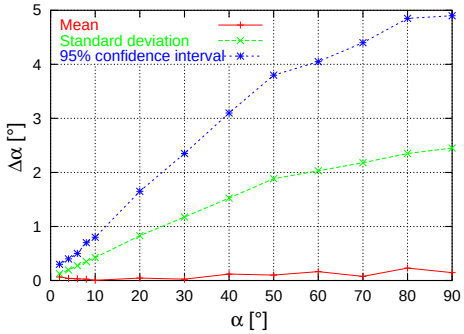

Fig. 4. Mixture of 1.000 samples of two Laplacian source signals with identical variances. Plotted is the mean, standard deviation and 95% confidence interval of $\Delta\alpha$ calculated from 100 runs for each angle α.

In order to show more precisely that $\Delta\alpha$ is proportional to α, we have plotted $\Delta\alpha/\alpha$ versus α in figure 5, which should result in a horizontal line. Obviously the estimate of the α_i with respect to α is good for a wide range of α ($\alpha > 10°$) and gets only slightly worse for smaller values.

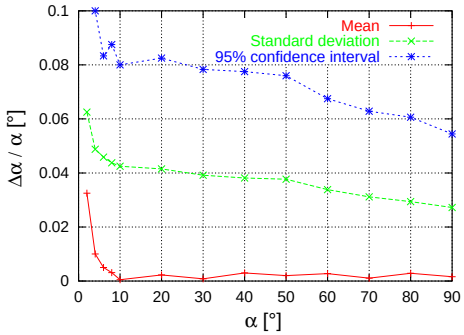

Fig. 5. Same mixture as in figure 4, plotting $\Delta\alpha/\alpha$ versus α.

Varying the number of samples used for estimating the angles α_i shows that with increasing number of samples the

estimate enhances as denoted in figure 6. This fact is well know in statistics and signal processing.

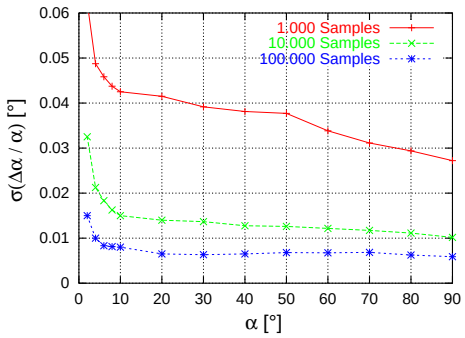

Fig. 6. Mixture of two Laplacian source signals with identical variances for different number of samples. The standard deviation of $\Delta\alpha/\alpha$ calculated from 100 runs for each angle α is plotted.

To get a relation between the error $\Delta\alpha$ and the number of samples, we plot for different α, the standard deviation of $\Delta\alpha$ versus the number of samples, see figure 7. From the slope of the curve we see that the standard deviation of $\Delta\alpha$ is roughly proportional to $n^{-\frac{1}{3}}$, where n denotes the number of samples.

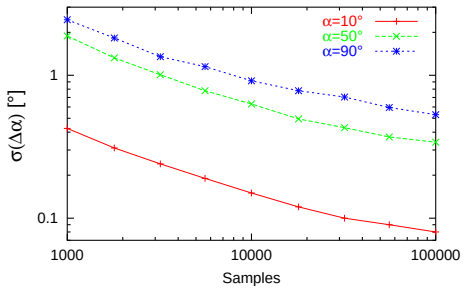

Fig. 7. Dependence of standard deviation of $\Delta\alpha$ with the number of samples for estimating the α_i for three different angles $\alpha = 10°$, $50°$ and $90°$. Note the logarithmic scale on both axes.

The above results have been put together numerically in table 1, where we chose following mixing matrix A:

$$A := \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}. \tag{14}$$

Given is the standard deviation of the non-diagonal terms after normalizing each column of the mixing matrix, so that the diagonal elements are unity. For comparison, we also calculated the performance index E_1 as proposed by Amari [11]

$$E_1 = \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{|p_{ij}|}{\max_k |p_{ik}|} - 1\right) + \sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{|p_{ij}|}{\max_k |p_{kj}|} - 1\right) \tag{15}$$

where $P = (p_{ij}) = A^{-1}_{recon} \cdot A$.

number of samples	standard deviation	index E_1
1.000	0.033	0.18
10.000	0.013	0.07
100.000	0.007	0.038

Table 1. Standard deviations of the non-diagonal terms and the performance index E_1 with different number of samples.

In statistics, an often used technique for estimating the error made by approximating a probability density function (pdf) by a finite number of samples is the so-called **confidence interval**, which is the interval around the estimate obtained from a given number of samples such that the probability that the real value lies outside this interval is less than a fixed **error probability** ε:

$$P(|X - \hat{x}| < c) = 1 - \varepsilon \tag{16}$$

For estimating the median of a pdf, we refer to [12]: Let $x_1, \ldots, x_n$ be independent identically distributed (i.i.d.) samples of the random variable X, such that $x_i < x_{i+1}$, then the **estimated median** $\hat{x}$ of the samples is defined as $x_{\frac{n+1}{2}}$ if n is odd or $\frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$ if n is even. For large n ($n > 36$), let z be the inverse of the cumulative standard distribution of $1 - \frac{\varepsilon}{2}$ and

$$k = -0,5 + \frac{n}{2} + \frac{\sqrt{n}}{2}z. \tag{17}$$

Then the confidence interval of $\hat{x}$ is given by

$$[x_{1+k}, x_{n-k}]. \tag{18}$$

In our case, an approximated confidence interval can be calculated by first running the algorithm and thus getting approximate angles α_i. Then use the above equations to get confidence intervals for the samples of Y restricted to the corresponding receptive fields F_i.

5. HIGHER DIMENSIONS

In the above sections we have explicitly used two-dimensional data. In real world problems, however, the mixed data is

usually high-dimensional (for example EEG-data with 21 dimensions). Therefore it would be satisfactory to generalize geometric algorithms to higher dimensions. The neural geometric algorithm can be easily translated to higher dimensional cases, but a serious problem occurs in the explicit calculation: In order to approximate higher dimensional pdfs it is necessary to have an exponentially growing number of samples, as will be shown in the following.

The number of samples in a ball B^{d-1} of radius ϑ on the unit sphere $S^{d-1} \subset \mathbb{R}^d$ divided by the number of samples on the whole S^{d-1} can be calculated as follows, if we assume a uniformly distributed random vector.

Let $B^d := \{x \in \mathbb{R}^d || x| \leq 1\}$ and $S^{d-1} := \{x \in \mathbb{R}^d || x| = 1\}$ – referring to [13], the volume of B^d can be calculated by

$$vol(B^d) = \frac{\pi^{\frac{d}{2}}}{\left(\frac{d}{2}\right)!} = c_d. \tag{19}$$

It follows for $d > 3$:

$$\begin{aligned} \frac{\text{Number of Samples in Ball}}{n} &= \frac{n\frac{\text{vol}(B^{d-1})\vartheta^{d-1}}{\text{vol}(S^{d-1})}}{n} \\ &= \frac{\vartheta^{d-1}c_{d-1}}{c_d} \\ &\leq \frac{\vartheta^{d-1}c_{d-1}}{c_{d+1}} \\ &= \frac{\vartheta^{d-1}d}{\pi} \end{aligned} \tag{20}$$

Obviously the number of samples in the Ball decreases by $\vartheta^{d-1}d$ if $\vartheta < 1$, which is the interesting case. To have the same accuracy when estimating the medians, the decrease must be compensated by an exponential growth in the number of samples. For three dimensions, we have found a good approximate for the demixing matrix by using 100.000 samples, in four dimensions the reconstructed mixing matrix couldn't be reconstructed correctly, even with a larger number of samples.

A different approach for higher dimensions has been taken by [14] and [8], where A has been calculated by using $\frac{d(d-1)}{2}$ projections of X from $\mathbb{R}^d$ onto $\mathbb{R}^2$ along the different coordinate axises and reconstructed the multidimensional matrix from the two-dimensional solutions. However, this approach works only satisfactory if the mixing matrix A is close to the unit matrix up to permutation and scaling. Otherwise, even in three dimensions, this projection approach won't give the desired results, as can be seen in figure 8, where the mixing matrix has been chosen as:

$$A = \begin{pmatrix} 1 & 0 & 0.65 \\ 0 & 1 & 0.3 \\ 0 & 0 & 0.7 \end{pmatrix} \tag{21}$$

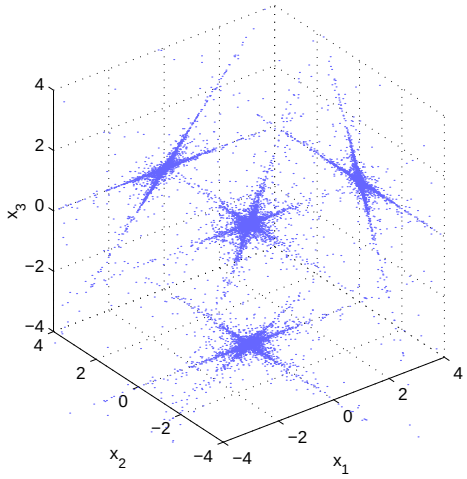

Fig. 8. Projection of a three dimensional mixture of Laplacian signals onto the three coordinate planes. Note that the projection into the x_1-x_2-plane does not have two distinct lines which are needed for the geometric algorithms.

6. CONCLUSION

We presented a new algorithm for linear geometric ICA based on histograms, which is both stable and more efficient compared to the neural geometric ICA algorithm. The accuracy of the algorithm concerning the estimation of the relevant medians of the underlying data distributions, when varying both the mixing matrices and the sample numbers, has been explored quantitatively, showing a rather good performance of the algorithm. We also considered the problem of high dimensional data sets with respect to the geometrical algorithms and discussed how projections to low-dimensional subspaces could solve this problem for a special class of mixing matrices.

Simulations with non-symmetrical and non-unimodal distributions have shown promising results so far, indicating that the new algorithm will perform as well with almost any distribution. This is the subject of ongoing research in our group.

In the future, the histogram based algorithm could as well be extended to the non-linear case similar to [8], using multiple centered spheres for projection on the surface on which the projected data histograms could then be evaluated.

7. ACKNOWLEDGMENTS

This research was supported by the Deutsche Forschungsgemeinschaft (DFG) in the Graduiertenkolleg "Nonlinearity and Nonequilibrium in Condensed Matter". We also would like to thank Christoph Bauer and Tobias Westenhuber for

suggestions and comments on the neural algorithm.

8. REFERENCES

[1] C. Jutten and J. Herault, "Blind separation of sources - an adaptive algorithm based on neuromimetics architecture," *Signal Processing*, pp. 1–10, 1991.

[2] P. Common, "Independent component analysis - a new concept ?," *Signal Processing*, vol. 36, pp. 287–314, 1994.

[3] A.J. Bell and T.J. Sejnowski, "An information-maximization approach to blind seperation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[4] J.-F. Cardoso, "Blind signal separation: Statistical principals," *Proc. IEEE*, vol. 86, pp. 2009–2025, 1998.

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

[6] C. G. Puntonet and A. Prieto, "Neural net approach for blind separation of sources based on geometric properties," *Neurocomputing*, vol. 18, pp. 141–164, 1998.

[7] Ch. Bauer, M. Habl, E.W. Lang, C.G. Puntonet, and M.R. Alvarez, "Probabilistic and geometric ICA applied to the separation of EEG signals," *M.H.Hamza, ed., Signal Processing and Communication (Proc.SPC'2000), IASTED/ACTA Press, Anaheim, USA*, pp. 339 – 346, 2000.

[8] C.G. Puntonet, Ch. Bauer, E.W. Lang, M.R. Alvarez, and B. Prieto, "Adaptive-geometric methods: application to the separation of EEG signals," *P. Pajunen and J. Karhunen, eds., ICA 2000 Proceedings*, pp. 273 – 277, 2000.

[9] Fabian J. Theis, Andreas Jung, and Elmar W. Lang, "A theoretic model for linear geometric ICA," *preprint*, 2001.

[10] A. Prieto, B. Prieto, C.G. Puntonet, A. Canas, and P. Martin-Smith, "Geometric separation of linear mixtures of sources: Application to speech signals," *J.F.Cardoso, Ch.Jutten, Ph.Loubaton, eds., Independent Component Analysis and Signal Separation (Proc. ICA'99)*, pp. 295–300, 1999.

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

[12] Bosch, "Statistik-Taschenbuch," *Oldenbuch Verlag*, pp. 697–701, 1993.

[13] R.K. Pathria, "Statistical mechanics," *Butterworth, Heinemann*, pp. 504–505, 1998.

[14] Ch. Bauer, C.G. Puntonet, M. Rodriguez-Alvarez, and E.W.Lang, "Separation of EEG signals with geometric procedures," *C. Fyfe, ed., Engineering of Intelligent Systems (Proc. EIS'2000)*, pp. 104–108, 2000.