30th Annual International IEEE EMBS Conference
Vancouver, British Columbia, Canada, August 20-24, 2008

Cumulant Matching for Independent Source Extraction

Ronald Phlypo, Vicente Zarzoso, Pierre Comon and Ignace Lemahieu

***Abstract*— In this work we show how one can make use of priors on signal statistics under the form of cumulant guesses to extract an independent source from an observed mixture. The advantage of using statistical priors on the signal lies in the fact that no specific knowledge is needed about its temporal behavior, neither about its spatial distribution. We show that these statistics can be obtained either by reasoning on the theoretical values of a supposed waveform, either by using a subset of the observations from which we know that their statistics are merely hindered by interferences. Results on an electro-cardiographic recording confirm the above assumptions.**

I. INTRODUCTION

The use of blind source separation (BSS) algorithms, and more specifically the independent component analysis (ICA) [1] algorithms, are nowadays widespread in biomedical signal processing. Being able to separate the observed signals into a set of independent random variables is close to reality, when the considered signals are generated by physiologically distinct entities [2], [3]. One of the examples is the reduction of eye blinking artifacts in electroencephalographic recordings, where the electrical source of the eye is physiologically distinct from the electrical processes which take place in the brain. In this case, the application of traditional ICA methods gives rather good results concerning the separation [2] but the identifiability of the signal of interest requires human interaction or well-adapted methods [4], [5].

A. Conventions

In this contributions all scalars, vectors and matrices will be represented by a lower case light face, lower case bold face and upper case bold face, respectively. Constants will be represented by an uppercase light face. All random variables are supposed i.i.d. Marginal cumulants of order r for a random scalar w will be denoted by $\mathcal{C}_r^w$. Likewise, $\mathcal{C}_r^\gamma$ denotes the r-th order cumulant of a distribution characterized by the parameter(vector) γ. The fourth order cumulant is also referred to as the excess kurtosis.

B. Independent Component Analysis with priors

Our linear, noiseless ICA model is defined as:

$$\mathbf{y} = \mathbf{A}\mathbf{s} \ , \quad p_{\mathbf{s}} = \Pi_{i=1}^n p_{s_i} \ , \tag{1}$$

This work was supported by FWO contract number G.0095.05

R. Phlypo and I. Lemahieu are with Ghent University - IBBT, IbiTech-MEDISIP, Ghent, Belgium {ronald.phlypo, ignace.lemahieu}@ugent.be

V. Zarzoso and P. Comon are with I3S, University of Nice-Sophia Antipolis - CNRS, Sophia Antipolis, France {zarzoso, pcomon}@i3s.unice.fr

where the equation on the left denotes the linear mixing model which transforms the source signals $\mathbf{s} \in \mathbb{R}^n$ through a linear instantaneous mixture $\mathbf{A} \in \mathbb{R}^{m\times n}$ into the observations $\mathbf{y} \in \mathbb{R}^m, m \geq n$. The equation on the right in (1) shows the independence of the source signals s_i. The independence is characterized by the fact that the joint probability density function (pdf) $p_{\mathbf{s}}$ can be factorized into the product of marginal pdf's p_{s_i}. ICA - given our model - aims at estimating the mixing process $\mathbf{A}$ and the sources $\mathbf{s}$ as $\mathbf{x} \triangleq \mathbf{B}^{-1}\mathbf{y} = \mathbf{B}^{-1}\mathbf{A}\mathbf{s} \approx \mathbf{G}\mathbf{s}$, given only the observations. Any scaling and/or permutation of the variables s_i in $\mathbf{s}$ will not affect their independence. Additional information is thus needed in order to extract the source(s) of interest without *a posteriori* variable selection, which is known to be prone to errors [6].

To extract a source (say s_1, without loss of generality) from a mixture $\mathbf{y}$ by using the above model (1) and *a priori* source information, the literature mentions a range of algorithms based either on reference signals [7], [8], [9], [10] or on the distribution over the (physical) sensors [11], [4], [5], [12], [6]. However, all of these techniques stand or fall with the availability of an accurate enough reference signal (respectively a reference spatial distribution). In this contribution, we follow a quasi maximum likelihood (ML) approach to extract the source of interest among a set of independent sources when only few of its (higher order) statistics can be estimated.

C. Likelihoods in ICA

The normalized log-likelihood [13] of our observations $\mathbf{y}$ can be expressed with respect to a distribution parameter θ as

$$\begin{aligned} \mathcal{L}_\theta(\mathbf{x}) &= \int_{\mathbb{R}^n} p_{\mathbf{x}}(\mathbf{u}) \log p_\theta(\mathbf{u}) d\mathbf{u} \\ &\stackrel{s}{=} -KL(p_{\mathbf{x}}(\mathbf{x})||p_\theta(\mathbf{x})) - H(p_{\mathbf{x}}(\mathbf{x})) \ , \end{aligned} \tag{2}$$

where KL is the Kullback-Leibler divergence defined as $KL(x,y) = \int_{\mathbb{R}} x \log(x/y)$, H the Shannon (differential) entropy, $\stackrel{s}{=}$ denotes the sample equivalent and $p_\theta(\mathbf{x})$ is the distribution of $\mathbf{x}$ conditioned on θ. In this work, the parameter vector θ contains the prior information α_i, the respective cumulant approximations of s_i, characterizing the source distributions.

Finding the maximum of $\mathcal{L}_\theta(\mathbf{x})$ is done by maximizing $-KL(\Pi_i p_{x_i}(x_i)||\Pi_i p_{\alpha_i}(x_i))$, because the differential entropy in (2) remains constant under an invertible basis change $\mathbf{B}$ [1], i.e. $H(\mathbf{x}) = H(\mathbf{y})$. As already noted in [13], the ML (up to a constant) can be subdivided into two separate

978-1-4244-1815-2/08/$25.00 ©2008 IEEE.

contributing terms, namely the mutual information and the marginal mismatch.

1) The Mutual Information (MI) term: $MI(\mathbf{x})$ is given by

$$MI(\mathbf{x}) = KL(p_{\mathbf{x}}(\mathbf{x})||\Pi_i p_{x_i}(x_i)) \ , \tag{3}$$

and is a direct measure of independence, since $MI(\mathbf{x}) \geq 0$ with equality if and only if the entries x_i are mutually independent. Note that this term does not depend on the priors α_i.

2) The marginal mismatch (MM) term: A measure for the gap between the hypothesized source distribution and the marginal distributions of the (supposedly independent) estimates. It is the likelihood term $\mathcal{L}_\theta(\mathbf{x})$ when the entries of $\mathbf{x}$ are supposed to be independent and is equivalent to

$$MM(\mathbf{x}, \theta) = \sum_{i=1}^{n} KL(p(x_i)||p(x_i|\theta_i)) \ . \tag{4}$$

In what follows, we combine both criteria when guesses of some output statistic(s) of only a single s_i is available, such that x_i is an appropriate estimate for s_i. In that case, one can not use the maximum likelihood approach from [14] which supposes all marginal cumulants to be known, neither can ordinary ICA algorithms [1], [15] alleviate the permutation ambiguity.

II. METHODS

A. Orthogonal Contrasts

Since independence is equivalent to the cancellation of all cross-cumulants, we subdivide our problem into a two-stage process. At first, we cancel all second order cross-cumulants and subsequently we will minimize the cross-cumulants of a given order. This reduces the computational load since it allows for a reduced search space in the second stage.

Consider the centered version of the observations $\tilde{\mathbf{y}} = \mathbf{y} - \mathcal{E}\{\mathbf{y}\}$ and its associated covariance matrix $\mathbf{C}_{\tilde{\mathbf{y}}} = \mathcal{E}\{\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T\}$, where $\mathcal{E}\{\cdot\}$ denotes mathematical expectation. If both $\mathcal{E}\{\mathbf{y}\}$ and $\mathcal{E}\{\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T\}$ are perfectly known we can standardize $\tilde{\mathbf{y}}$, yielding decorrelated random variables $\mathbf{z} = \mathbf{C}_{\tilde{\mathbf{y}}}^{-1/2}\tilde{\mathbf{y}}$, with $\mathcal{E}\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}_n$, the identity matrix in $\mathbb{R}^n$. Transformations of $\mathbf{z}$ are guaranteed to preserve the decorrelation if they are restricted to $\mathbf{x} = \mathbf{Q}^T\mathbf{z}$, with $\mathbf{Q} \in SO(n)$, the special orthogonal group in $\mathbb{R}^n$. Our system then reduces to $\mathbf{x} = \mathbf{Q}^T\mathbf{z} = \mathbf{Q}^T\mathbf{C}_{\tilde{\mathbf{y}}}^{-1/2}\tilde{\mathbf{y}}$, where the unknown $\mathbf{Q}$ is now in $SO(n) \subset GL(n)$.

B. The Connection Between Orthogonal ML and Standardized Cumulants

The ML term in (2) is the divergence between the marginal pdf's of the estimates and those of the envisaged sources. We thus need to compare individually each $p_{x_i}(x_i)$ to its respective hypothesized distribution $p_{\alpha_i}(x_i)$. Since estimating a distribution from observed variables is a highly complicated task, we will compare two distributions by comparing the respective expansions around their closest normal distribution, in terms of their Edgeworth Type A series [16], [1], [14]. The Edgeworth expansion for a density function of a standardized variable (zero-mean, unit variance) is given by

$$p(u|\gamma) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) \sum_{j=1}^{\infty} k_j(\gamma) H_j(u) \ , \tag{5}$$

where $k_j(\gamma)$ are functions of the cumulants of the distribution characterized by γ and $H_j(u)$ are the Chebychev-Hermite polynomials [16, §6.20]. If we further suppose that the distributions are symmetric (odd cumulants vanish) and standardized ($\mathcal{C}_1 = 0, \mathcal{C}_2 = 1$) and we consider our expansion until the term in u^4 (the parameters associated to higher orders of u are less accurate with increasing order when estimated from population samples of a distribution), then all polynomial coefficients of $k_j(x_i)$ $[k_j(\alpha_i)]$ depend solely on the cumulants $\mathcal{C}_4^{x_i}$ $[\mathcal{C}_4^{\alpha_i}]$. If we further consider that the expansion is valid in the neighbourhood of the origin ($|f(u)| \ll 1$), then $\log 1 + f(u) \approx f(u)$,. By putting (5) in (4), we then get

$$ML(\mathbf{x}, \theta) = \prod_i \frac{\mathcal{C}_4^{x_i} - \mathcal{C}_4^{\alpha_i}}{24} \left(\int_{\mathbb{R}} p_{x_i}(u_i) u_i^4 \mathrm{d}u_i - 6 \right) \ . \tag{6}$$

The distance between the cumulant of our outputs x_i and the hypothesized distributions parameterized by α_i is minimized (cumulant matching) through maximizing

$$\phi_{ML} = \sum_{i=1}^{p} -(\mathcal{C}_4^{x_i})^2 + \alpha_i \mathcal{C}_4^{x_i} \ . \tag{7}$$

C. Mutual Information

To express the mutual information (3) between variables, one can fall back onto the information theoretic equality that the mutual information is minimized when the KL-divergence of the marginal distributions with respect to a normal with equal first and second moments are maximized [1].

As above in the case of the ML, we can show that under the same assumptions of standardized distributions and truncation of the polynomials at u^4, there is a connection between the optimization of the mutual information and the optimization of $\phi_{MI} = \sum_{i=1}^{n} (\mathcal{C}_4^{x_i} - \mathcal{C}_4^{\mathcal{N}})^2 = \sum_{i=1}^{n} (\mathcal{C}_4^{x_i})^2$, where $\mathcal{N}$ denotes the standardized normal distribution, where we use the fact that all higher order cumulants of a normal distribution vanish, i.e. $\mathcal{C}_4^{\mathcal{N}}) = 0$.

D. ICA with Cumulant Priors

Define now a function $\phi_\xi(\mathbf{Q}) = \xi \cdot \phi_{ML}(\mathbf{Q}) + (1 - \xi)\phi_{MI}(\mathbf{Q})$, where $\xi \in [0, 1]$ is a weighting factor. If we have no priors on the cumulants of the sources, we can set $\xi = 0$ and obtain $\phi_0 = \phi_{MI}$, which is the standard squared cumulants ICA contrast (COM2) [1]. However, if all sources would have a prior on their cumulants we obtain by setting $\xi = 1/2$ the contrast $\phi_1 = \sum_{i=1}^{n} \alpha_i \mathcal{C}_4^{x_i}$, the contrast as proposed in [17]. Setting $\xi = 1/2$, we obtain $\phi_{1/2} = 1/2 \sum_{i=1}^{n} \alpha_i \mathcal{C}_4^{x_i} + (\mathcal{C}_4^{x_i})^2$ which is equivalent to the orthogonal ML contrast [14].

Consider now the function

$$\phi = \alpha \mathcal{C}_4^{x_1} + \sum_{j=2}^{n} (\mathcal{C}_4^{x_i})^2 \ , \tag{8}$$

which is a hybrid function of the ML (for x_1) and the mutual information $x_{j\neq 1}$. We thus have a function that can be used when prior knowledge is available for a single source only.

E. Optimizing the Objective Function by Planar Rotations

Since the optimization of the contrast ϕ is over SO(n), we can define the optimal rotation matrix $\mathbf{Q}^\star$ through a chain of rotations $\mathbf{Q}^\star = \lim_{p\to\infty} \mathbf{Q}^{\star(p)}$ where $\mathbf{Q}^{\star(p)} = \Pi_{r=1}^{p} \mathbf{q}^{\star(r)}$. Optimization can be done iteratively by taking at each iteration (k) a pair $(x_i^{(k-1)}, x_{j\neq i}^{(k-1)})$ of the output $\mathbf{x}^{(k-1)} = (\mathbf{Q}^{\star(k-1)})^T \mathbf{z}$ and rotating it in its plane as $\mathbf{x}_{ij}^{(k)} = \mathbf{q}^{(k)} \mathbf{x}_{ij}^{(k-1)}$. If the matrix $\mathbf{q}^{(k)}$ is taken as $\mathbf{q}_{ii}^{(k)} = \mathbf{q}_{jj}^{(k)} = 1/\sqrt{1+t^2}$, $\mathbf{q}_{ij}^{(k)} = -\mathbf{q}_{ji}^{(k)} = t/\sqrt{1+t^2}$ (Givens rotation), the sole parameter on which $\phi_p(\mathbf{Q}^{(k)})$ depends, is $t = \tan\varphi$.

An interesting by-product of an optimization scheme using planar rotations is that it can easily be adapted to source extraction by defining the update sequence as an iteration over all pairs including x_i, where x_i is the source of interest with prior α_i [**?**]. Without loss of generality we can take $i = 1$. In this paper we will only make use of this source extraction mode. The contrast for each of these pairs can thus be reduced to $\phi_{(x_1,x_j)} = 2\alpha \mathcal{C}_4^{x_1} + (\mathcal{C}_4^{x_j})^2$. Rooting the first derivative of this contrast (a polynomial in t) and retaining the real root $t^\star = \arg\max_t \phi_p(\mathbf{q}^{(k)})$ gives the optimal solution $\mathbf{q}^{\star(k)} = \mathbf{q}^{(k)}(t^\star)$ at each iteration. The polynomial of degree 8 to be rooted is given in A.

III. RESULTS

We applied our method onto an ECG fragment with known atrial fibrillation of which the high-pass, centered version is shown in figure 1. Two experiments are set up where we envisage the extraction of the source associated to 1)the QRS complex and 2) the atrial activity. To have an estimator α_{QRS} (respectively α_{AA}) for the cumulant of the envisaged source, we use an empirical estimate. For the estimator α_{AA} we also introduce two theoretical values based on prior knowledge of the waveform/distribution for comparison.

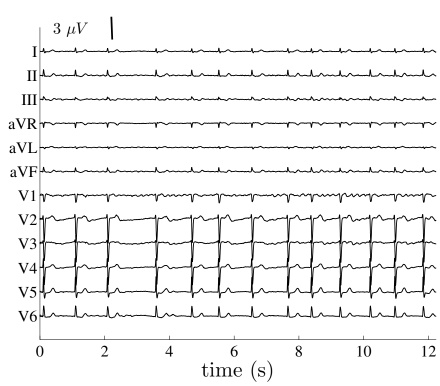

Fig. 1. The original ECG fragment after high pass filtering at 1Hz.

1) Extracting the QRS component: Hereto, α_{QRS} has been chosen larger than the largest cumulant of $\mathbf{s}$ accordingly to the assumption that there is no waveform with a larger cumulant in the set. Since we have no knowledge of the kurtosis values in the set $\mathbf{s}$, we set α_{QRS} arbitrarily high, namely $\alpha_{QRS} = 100$. Figure 2 shows the so extracted source x_1 together with the manually selected estimate of an ordinary ICA algorithm (COM2). The absolute value of the correlation coefficient between both is 1.0000.

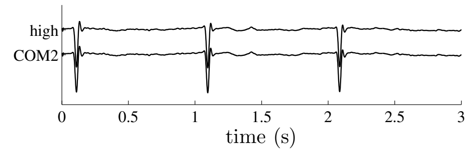

Fig. 2. The estimates x_1 of the proposed algorithm at order 4 ($\alpha_{QRS} = 100$) and of a regular ICA algorithm (COM2), where the estimate with largest kurtosis has been selected *a posteriori*.

2) Extracting the AA component: In figure 3 we show the results when the AA source is envisaged. We tried three different values of α: The empirical prior α_{AA} was obtained by calculating the statistics on isolated AA-waves from the observed dataset ($\alpha_{AA}^{AA} = -0.2213$). Theoretical estimates take the value of the cumulant of a stationary sinusoidal ($\alpha_{AA}^{sin} = -3/2$) or triangular ($\alpha_{AA}^{tri} = -6/5$) waveform. The so obtained estimated source x_1 has a kurtosis value of -0.2038, respectively -0.2091 and -0.2138. As a comparison, we give the estimate for a spatiotemporal BSS method (STBSS [18], $\mathcal{C}_4^{x_1} = -0.0000$) and a regular ICA method, where the kurtosis value was -0.1483.

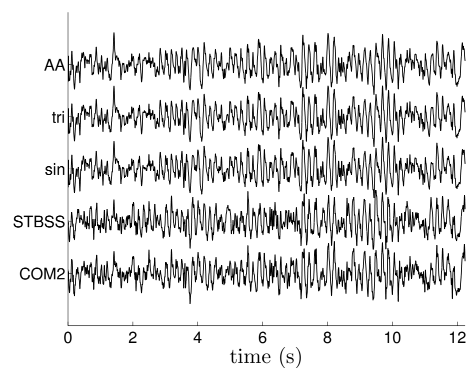

Fig. 3. The estimated source x_1 when using the prior derived from the dataset (AA) or a prior on the waveform (α_{AA}^{tri} (tri) or α_{AA}^{sin} (sin)), and the estimate from a spatio-temporal BSS (ST-BSS), where *a posteriori* variable selection has been performed.

IV. DISCUSSION AND CONCLUSION

From the above results, we can see that the estimated source is in line with the independent sources estimated by more general ICA algorithms, such as COM2 [1], and with more specific signal *extraction* techniques, such as ST-BSS [18] for temporally correlated, non-impulsive signals. Whereas in the latter two, selection takes place *a posteriori*, our method solves for the permutation ambiguity within the algorithm itself, reducing it to a single stage, and thus making it less prone to errors in the previous stages of the algorithm.

This is especially the case in deflation algorithms where the source of interest is not the first source to be extracted and the algorithm thus propagates and accumulates the errors made in previous stages [19].

Another advantage of our method, this time with respect to algorithms using sampled references (either temporal or spatial), is its robustness to prior mismatches. Since generally a constraint is put on the maximally allowed distance between the reference and the source to extract, the waveforms (both in phase and frequency) have to be aligned to be non-orthogonal, a problem that is not present in our proposed extraction algorithm. The approximation of a distribution by its higher order statistics also omits the difficulty of adaptation of the score function (a nonlinear function that is ideally the cumulative density function of the source to extract) to the source distribution [20], [21] or to solve a maximum a posteriori problem through stochastic optimization [22]. We reduce the prior information on the source distribution to a single set of cumulants.

Remarks: 1) When the excess kurtosis is not discriminative over the set (i.e. the kurtosis is not a sufficient statistic), then the algorithm can be extended to priors on a set of cumulants of different order. 2) Since we do not restrict ourselves to ϕ_{MI}, the polynomial $\partial\phi_p(\mathbf{Q}^{(k)})/\partial t = 0$ has no symmetry and hence the roots cannot be found algebraically (the roots of a fourth order polynomial can be found by the method of Ferrari, but this is also the highest order for which an algebraic solution exists). Therefore, we need to turn to numerical algorithms to solve for the roots, e.g. by calculating the eigenvalues of its associated matrix [23].

V. ACKNOWLEDGMENTS

The authors gratefully acknowledge F. Castells and the Electrophysiology Laboratory of the University Clinical Hospital of Valencia, Spain for sharing the data with us.

APPENDIX

A. Optimization of $\alpha\mathcal{C}_4^{x_1} + \sum_{j=2}^{n}\mathcal{C}_4^{x_j}$

The coefficients of the polynomial $\partial\phi/\partial t$ are listed here. Define the output cumulants of order 4 of the pair $(x_1, x_{j\neq 1})$ as κ_{pq} where p and q denote the occurences of x_1, respectively x_j in $\mathcal{C}_4(x_1 x_1 \ldots x_j x_j \ldots)$ and the priors on the marginal cumulant of x_1 as $\alpha = \hat{\mathcal{C}}_4^{s_1}$. The polynomial can then be given in the form $\sum_{i=1}^{8} a_i t^i$ with the a_i given by

$$
\begin{aligned}
a_8 &= \alpha\kappa_{13} - \kappa_{40}\kappa_{31}\\
a_7 &= \kappa_{40}^2 + \alpha\kappa_{04} - 4\kappa_{31}^2 - 3\kappa_{22}(\kappa_{40} + \alpha)\\
a_6 &= (3\kappa_{31} - \kappa_{13})\alpha + (7\kappa_{31} - 3\kappa_{13})\kappa_{40} - 18\kappa_{31}\kappa_{22}\\
a_5 &= c_1 + 9\kappa_{40}\kappa_{22} + 12\kappa_{31}^2 + \alpha(2\kappa_{04} - \kappa_{40} - 3\kappa_{22})\\
a_4 &= 5(6\kappa_{22} + \alpha)(\kappa_{31} - \kappa_{13}) + 5(\kappa_{13}\kappa_{40} - \kappa_{31}\kappa_{04})\\
a_3 &= -c_1 - 9\kappa_{04}\kappa_{22} - 12\kappa_{13}^2 - \alpha(2\kappa_{40} - \kappa_{04} - 3\kappa_{22})\\
a_2 &= (\kappa_{31} - 3\kappa_{13})\alpha + (3\kappa_{31} - 7\kappa_{13})\kappa_{04} + 18\kappa_{13}\kappa_{22}\\
a_1 &= -\kappa_{04}^2 - \alpha\kappa_{40} + 4\kappa_{13}^2 + 3\kappa_{22}(\kappa_{04} + \alpha)\\
a_0 &= -\alpha\kappa_{31} + \kappa_{04}\kappa_{13}\ ,
\end{aligned}
$$

where $c_1 = -18\kappa_{22}^2 - \kappa_{40}\kappa_{04}$.

REFERENCES

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

[2] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Advances in Neural Information Processing Systems*, vol. 8, 1996, pp. 145 – 151.

[3] C. J. James and C. W. Hesse, "Independent component analysis for biomedical signals," *Physiol Meas*, vol. 26, pp. R15–R39, 2005.

[4] N. Ille, P. Berg, and M. Scherg, "Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies." *J Clin Neurophysiol*, vol. 19, no. 2, pp. 113–124, Apr 2002.

[5] R. Phlypo, Y. D'Asseler, and I. Lemahieu, "Removing ocular movement artefacts by a joint smoothened subspace estimator (JSSE)," *Computational Intelligence and Neuroscience*, vol. 2007, no. Article ID 75079, 13 pages, doi:10.1155/2007/75079, 2007.

[6] C. W. Hesse and C. J. James, "On semi-blind source separation using spatial constraints with applications in eeg analysis." *IEEE Trans Biomed Eng*, vol. 53, no. 12 Pt 1, pp. 2525–2534, Dec 2006.

[7] W. Lu and J. C. Rajapakse, "Approach and applications of constrained ica," *IEEE Trans on Neural Networks*, vol. 16, no. 1, pp. 203–212, 2005.

[8] C. J. James and O. Gibson, "ICA with a reference: extracting desired electromagnetic brain signals," *Medical Applications of Signal Processing*, 2002, dr C J James, Biomedical Information Engineering Research Group, Aston University, Aston Triangle, Birmingham B4 7ET, UK.

[9] J. Lee, K. L. Park, and K. J. Lee, "Temporally constrained ICA-based foetal ecg separation," *Electronics Letters*, vol. 41, no. 21, pp. 1158–1160, 2005.

[10] A. Adib, E. Moreau, and D. Aboutajdine, "Source separation contrasts using a reference signal," *IEEE Sig Proc Letters*, vol. 11, no. 3, pp. 312–315, 2004.

[11] M. Knaak, S. Araki, and S. Makino, "Geometrically constrained independent component analysis," *IEEE Trans On Audio, Speech and Language Processing*, vol. 15, no. 2, pp. 715–726, 2007.

[12] C. W. Hesse and C. J. James, "The FastICA algorithm with spatial constraints," *IEEE Sign Proc Letters*, vol. 12, no. 11, pp. 792–795, 11 2005.

[13] J. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 90, pp. 2009–2026, Oct 1998.

[14] ——, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, jan 1999.

[15] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neur Comp*, vol. 9, pp. 1483–1492, 1997.

[16] A. Stuart and K. Ord, *Distribution Theory*, ser. Kendall's advanced theory of statistics. Hodder Arnold, 2006, vol. 1.

[17] V. Zarzoso, P. Comon, and R. Phlypo, "ICA permutation ambiguity is fixed with rough guesses on source kurtoses," UNSA/CNRS I3S, http://www.i3s.unice.fr/∼mh/RR/2008/liste-2008.html, Tech. Rep., 2008.

[18] F. Castells, J. Rieta, J. Millet, and V. Zarzoso, "Spatiotemporal blind source separation approach to atrial activity estimation in atrial tachyarrhythmias," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 2, pp. 258–267, Feb. 2005.

[19] N. Delfosse and P. Loubaton, "Adaptive blind source separation of independent sources: A deflation approach," *Signal Processing*, vol. 45, pp. 59–83, 1995.

[20] L. Zhang, A. Cichocki, and S. ichi Amari, "Self-adaptive blind source separation based on activation functions adaptation," *IEEE Trans on Neural Networks*, vol. 15, no. 2, pp. 233–244, 2004.

[21] T.-W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Computation*, vol. 11, pp. 417 – 441, 1999.

[22] K. K. Knuth, "A bayesian approach to blind source separation," in *ICA 99*, Aussois, France, 1 1999.

[23] E. W. Weisstein, "Polynomial roots," MathWorld–A Wolfram Web Resource, http://mathworld.wolfram.com/PolynomialRoots.html, 04 2008.