Acta Cybernetica **18** (2007) 213–221.

Independent Subspace Analysis can Cope with the 'Curse of Dimensionality'

Zoltán Szabó* and András Lőrincz†

Abstract

We search for hidden independent components, in particular we consider the independent subspace analysis (ISA) task. Earlier ISA procedures assume that the dimensions of the components are known. Here we show a method that enables the non-combinatorial estimation of the components. We make use of a decomposition principle called the ISA separation theorem. According to this separation theorem the ISA task can be reduced to the independent component analysis (ICA) task that assumes one-dimensional components *and then* to a grouping procedure that collects the respective non-independent elements into independent groups. We show that non-combinatorial grouping is feasible by means of the non-linear f-correlation matrices between the estimated components.

Keywords: independent subspace analysis, non-combinatorial solution

1 Introduction

The technique called independent component analysis (ICA) and its independent subspace analysis (ISA) extension are in the focus of research interest for signal processing tasks. ICA applications include, among others: (i) feature extraction [4], (ii) denoising [6], (iii) processing of financial [11] and neurobiological data, e.g. fMRI, EEG, and MEG [12,26]. The ISA model is frequently applied for the analysis of EEG-fMRI signals [1].

Originally, ICA is one-dimensional in the sense that all sources are assumed to be independent real valued stochastic variables. The typical example of ICA is the so-called *cocktail-party problem*, where there are D sound sources and D microphones and the task is to separate the original sources from the observed mixed signals. Clearly, applications where not all, but only certain groups of the sources are independent may have high relevance in practice. In this case, independent sources can be multidimensional. For example, there could be independent groups

*Department of Information Systems, Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary; E-mail: `szzoli@cs.elte.hu`

†Corresponding author; Department of Information Systems, Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary; E-mail: `andras.lorincz@elte.hu`

of people talking about independent topics at a conference or more than one group of musicians may be playing at a party. This is the independent subspace analysis (ISA) extension of ICA.[1] Strenuous efforts have been made to develop ISA algorithms [1,3,5,7–9,13–15,18,19,22,24,25,27], where the theoretical problems concern mostly (i) the estimation of the entropy or of the mutual information, or (ii) joint block diagonalization.

Earlier ISA methods were constrained by assuming that the dimensions of the hidden components are known. Here, we show a non-combinatorial solution to the estimation of the dimensions. In the ISA problem one assumes temporally i.i.d. (independent and identically distributed) hidden sources. For the non i.i.d case, one may try the autoregressive assumption (see, e.g., [16] and references therein). This problem family is called independent process analysis (IPA). The method that we present here can be extended to IPA tasks by applying the innovation trick of [17].

The paper is built as follows: Section 2 formulates the problem domain. The estimation of the dimensions of the ISA components is described in Section 3. We illustrate our method in Section 4. Conclusions are drawn in Section 5.

2 The ISA Model

First, we define the ISA model. Assume that we have M hidden independent multidimensional and i.i.d random variables and that only the mixture of these M *components* is available for observation:

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

where $\mathbf{s}(t) := \left[\mathbf{s}^1(t); \ldots; \mathbf{s}^M(t)\right]$ is the vector concatenated form of the components $\mathbf{s}^m \in \mathbb{R}^{d_m}$. We assume that (i) for a given m, $\mathbf{s}^m(t)$ is i.i.d. in time t, (ii) there is at most a single Gaussian component amongst $\mathbf{s}^m$s, and (iii) $I(\mathbf{s}^1, \ldots, \mathbf{s}^M) = 0$, where I stands for the mutual information of the arguments. The total dimension of the components is $D := \sum_{m=1}^{M} d_m$. $\mathbf{A} \in \mathbb{R}^{D \times D}$ is the so-called *mixing matrix* that, according to our assumptions, is invertible. The goal of the ISA task is to uncover hidden components $\mathbf{s}^m$ (and the *separation matrix* $\mathbf{W} = \mathbf{A}^{-1}$) using the observations $\mathbf{x}(t)$ only. The ICA task is recovered when every components is of one-dimensional, i.e., if $d_m = 1$ $(m = 1, \ldots, M)$.

In the ISA model, we can assume without any loss of generality, that both the hidden source $\mathbf{s}$ and the observation $\mathbf{x}$ are white, that is, their expected values and covariances are $\mathbf{0}$ and $\mathbf{I}_D$, respectively. Here $\mathbf{I}_D$ denotes the D-dimensional identity matrix. Then:

- The $\mathbf{s}^m$ components are determined up to permutation and orthogonal transformation [23].

[1] ISA is also called multidimensional independent component analysis (MICA) [5] and group ICA [24] in the literature.

- One may assume that the separation matrix $\mathbf{W}$ is orthogonal: $\mathbf{W} \in \mathcal{O}^D := \{\mathbf{W} \in \mathbb{R}^{D \times D} | \, \mathbf{W}\mathbf{W}' = \mathbf{I}_D\}$ where $\mathcal{O}^D$ denotes orthogonal matrices of size $D \times D$ and $'$ stands for transposition.

3 Dimension Estimation of the Components in the ISA Task

Here we put forth a non-combinatorial solution that can uncover the the dimensions of the ISA components. We build our method onto (i) the ISA separation theorem [21, 22] and (ii) the ISA cost function introduced in [19].

The ISA separation theorem, which was conjectured by Jean-François Cardoso [5], allows one to decompose the solution of the ISA problem, under certain conditions, into 2 steps: In the first step, ICA estimation is executed. In the second step, the ICA elements are grouped by finding an optimal permutation. Formally:

Theorem 1 (Separation Theorem for ISA)**.** *Let* $\mathbf{y} = [y_1; \ldots; y_D] = \mathbf{W}\mathbf{x}$*, where* $\mathbf{W} \in \mathcal{O}^D$*,* $\mathbf{x} \in \mathbb{R}^D$ *is the whitened observation of the ISA model. Let* $\mathcal{S}^{d_m}$ *denote the surface of the* d_m*-dimensional unit sphere, that is* $\mathcal{S}^{d_m} := \{\mathbf{w} \in \mathbb{R}^{d_m} : \sum_{i=1}^{d_m} w_i^2 = 1\}$*.* H *is Shannon's differential entropy.*

Presume that the $\mathbf{u} := \mathbf{s}^m$ *sources* $(m = 1, \ldots, M)$ *of the ISA model satisfy condition*

$$H\left(\sum_{i=1}^{d_m} w_i u_i\right) \geq \sum_{i=1}^{d_m} w_i^2 H\left(u_i\right), \forall \mathbf{w} \in \mathcal{S}^{d_m}, \tag{2}$$

and that the ICA cost function $J_{ICA}(\mathbf{W}) = \sum_{i=1}^{D} H(y_i)$ *has minimum over the orthogonal matrices in* $\mathbf{W}_{\mathrm{ICA}}$*. Then it is sufficient to search for the solution of the ISA task as a permutation of the solution of the ICA task. Using the concept of separation matrices, it is sufficient to explore forms*

$$\mathbf{W}_{\mathrm{ISA}} = \mathbf{P}\mathbf{W}_{\mathrm{ICA}},$$

where $\mathbf{P} \in \mathbb{R}^{D \times D}$ *is a permutation matrix to be determined, and* $\mathbf{W}_{\mathrm{ISA}}$ *is the ISA separation matrix.*

Sufficient conditions for Eq. (2) were eventually found by Szabó et al. (see [22] and references therein). Further, one can group the ICA components and can find the optimal permutation efficiently by means of the joint f-decorrelation (JFD) technique introduced in [19]. Roughly speaking, the JFD technique performs decorrelation over an $\mathcal{F}$ set of functions. In particular, the method aims the simultaneous block-diagonalization of covariance matrices $\mathbf{C}_f(\mathbf{W}) := cov\left(f\left[\hat{\mathbf{s}}(\mathbf{W})\right], f\left[\hat{\mathbf{s}}(\mathbf{W})\right]\right)$ of all functions $f \in \mathcal{F}$, where blocks are d_m-dimensional.

However, the hidden components can be determined without knowing their dimensions, provided that the separation theorem holds. In this case, the estimated ICA elements correspond to the ISA components up to permutation. In other words, matrices $\mathbf{C}_f$ are block-diagonal with block size d_m apart from a *common* permutation. Thus, the coupled components can be found by the following procedure. We say that two coordinates i and j are $\mathbf{C}^{\mathcal{F}}$-'connected' ($\mathbf{C}^{\mathcal{F}} := \sum_{f \in \mathcal{F}} |\mathbf{C}_f|$, $|\cdot|$ denotes absolute values for all coordinates) if $\max(C^{\mathcal{F}}_{ij}, C^{\mathcal{F}}_{ji}) > \epsilon$, where $\epsilon \geq 0$ and in the ideal case $\epsilon = 0$. Then we group the $\mathbf{C}^{\mathcal{F}}$-'connected' coordinates into separate subspaces as follows: (1) Choose an arbitrary coordinate i and group all $j \neq i$ coordinates to it which are $\mathbf{C}^{\mathcal{F}}$-'connected' with it. (2) Choose an arbitrary and not yet grouped coordinate. Find its connected coordinates. Group them together. (3) Continue until all components are grouped. This is the *gathering procedure* and it is fast. In the worst case, it is quadratic in the number of the coordinates.

4 Illustration

Here we illustrate how our method works. Test cases are introduced in Section 4.1. The quality of the solutions will be measured by the normalized Amari-error, the Amari-index (Section 4.2). Numerical results are presented in Section 4.3.

4.1 Databases

We define three databases to study our identification algorithm. The databases are illustrated in Fig. 1. In the *3D-geom* test $\mathbf{s}^m$s were random variables uniformly distributed on 3-dimensional geometric forms ($d = 3$). We chose 6 different components ($M = 6$) and, as a result, the dimension of the hidden source $\mathbf{s}$ is $D = 18$. The *celebrities* test has 2-dimensional source components generated from cartoons of celebrities ($d = 2$).[2] Sources $\mathbf{s}^m$ were generated by sampling 2-dimensional coordinates proportional to the corresponding pixel intensities. In other words, 2-dimensional images of celebrities were considered as density functions. $M = 10$ was chosen ($D = 20$). In the *ABC* database, hidden sources $\mathbf{s}^m$ were uniform distributions defined by 2-dimensional images ($d = 2$) of the English alphabet. The number of components was $M = 10$, thus the dimension of the source D was 20.

4.2 Normalized Amari-error, the Amari-index

The optimal estimation provides matrix $\mathbf{G} := \mathbf{WA}$, a block-permutation matrix made of $d \times d$ sized blocks. This block-permutation property can be measured by the Amari-index. Namely, let matrix $\mathbf{G} \in \mathbb{R}^{D \times D}$ be decomposed into $d \times d$ blocks: $\mathbf{G} = \left[\mathbf{G}^{ij}\right]_{i,j=1,\ldots,M}$. Let $g^{i,j}$ denote the sum of the absolute values of the elements of matrix $\mathbf{G}^{i,j} \in \mathbb{R}^{d \times d}$. Then the normalized version of the Amari-error [2] adapted to the ISA task [24] is defined as [20]:

$$r(\mathbf{G}) := \frac{1}{2M(M-1)} \left[\sum_{i=1}^{M} \left(\frac{\sum_{j=1}^{M} g^{ij}}{\max_j g^{ij}} - 1 \right) + \sum_{j=1}^{M} \left(\frac{\sum_{i=1}^{M} g^{ij}}{\max_i g^{ij}} - 1 \right) \right].$$

[2]http://www.smileyworld.com

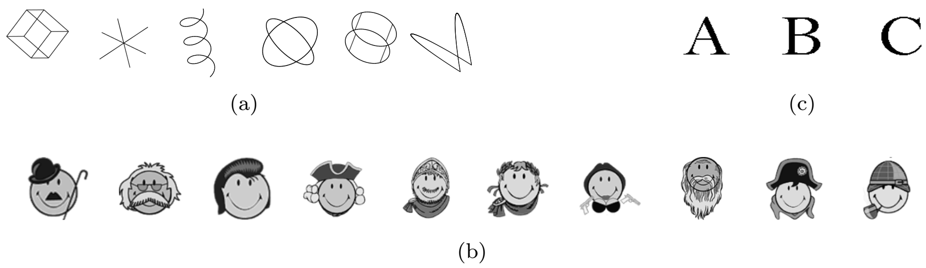


Figure 1: Illustration of the *3D-geom*, *celebrities* and ABC databases. (a): database *3D-geom*, 6 pieces of 3-dimensional components ($M = 6$, $d = 3$). Hidden sources are uniformly distributed variables on 3-dimensional geometric objects. (b): database *celebrities*. Density functions of the hidden sources are proportional to the pixel intensities of the 2-dimensional images ($d = 2$). Number of hidden components: $M = 10$. (c): database ABC. Here, the hidden sources $\mathbf{s}^m$ are uniformly distributed on images ($d = 2$) of letters. Number of components M was 10 (A-J).

We refer to the normalized Amari-error as the Amari-index. One can see that $0 \leq r(\mathbf{G}) \leq 1$ for any matrix $\mathbf{G}$, and $r(\mathbf{G}) = 0$ if and only if $\mathbf{G}$ is a block-permutation matrix with $d \times d$ sized blocks.

4.3 Simulations

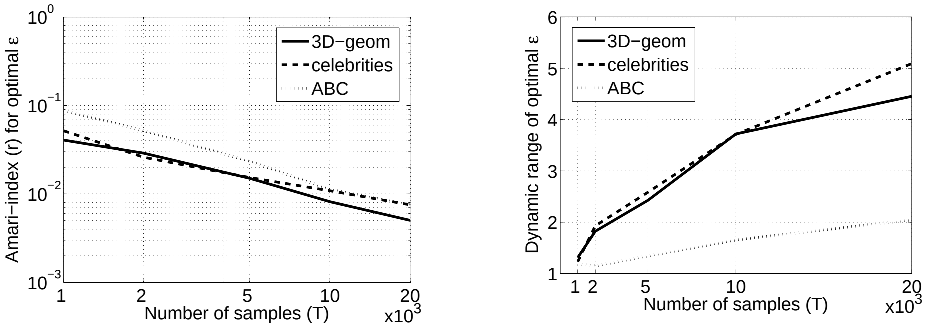
Results on databases *3D-geom*, *celebrities*, and ABC are provided here. Our gauge to measure the quality of the results is the Amari-index (Section 4.2) that we computed by averaging over 50 random runs.[3] These experimental studies concerned the following problems:

1. The quality of the gathering procedure depends on the threshold parameter ε. We studied the estimation error, the Amari-index, as a function of sample number. The ε values were preset to reasonably good values.

2. We studied the optimal domain for the ε values. We looked for the dynamic range, i.e., the ratio of the highest and lowest 'good ε values': We divided interval $[0, C^{\mathcal{F}}_{max}]$ ($C^{\mathcal{F}}_{max} := \max_{i,j} C^{\mathcal{F}}_{ij}$) into 200 equal parts. For different sample numbers in all databases at each division point we used the gathering procedure to group the ICA elements. For each of the 50 random trials we have computed the Amari-indices separately. For the smallest Amari-index, we determined the corresponding interval of ε's, these are the 'good ε values'. Then we took the ratio of the largest and smallest ε values in this set and averaged the ratios over the 50 runs. The average is called the dynamic range.

In our simulations, sample number T of observations $\mathbf{x}(t)$ was varied between $1,000$ and $20,000$. Mixing matrix $\mathbf{A}$ was generated randomly from the orthogonal

[3]Random run means random choice of quantities $\mathbf{A}$ and $\mathbf{s}$.



(a) (b)

Figure 2: Amari-index on log-log scale (a) and dynamic range (b) as a function of sample number for the *3D-geom*, *celebrities*, and *ABC* databases.

group. The fastICA [10] algorithm was chosen to perform the ICA computation. In the JFD technique, we chose manifold $\mathcal{F}$ as $\mathcal{F} := \{\mathbf{u} \mapsto \cos(\mathbf{u}), \mathbf{u} \mapsto \cos(2\mathbf{u})\}$, where the functions operated on the coordinates separately [19]. We computed correlations for matrices $\mathbf{C}_f$ $(f \in \mathcal{F})$ (instead of covariances) because it is normalized.

Our results are summarized in Fig. 2. According to Fig. 2(a), there are good ε parameters for the $\mathbf{C}^{\mathcal{F}}$-'connectedness' already for $1,000 - 2,000$ samples: our method can find the hidden components with high precision. Figure 2(a) also shows that by increasing the sample number the Amari-index decreases. For $20,000$ samples, the Amari-index is 0.5% for the *3D-geom*, 0.75% for the *celebrities*, and 0.75% for the *ABC* database, respectively on the average. The decline of the Amari-index follows power law $(r(T) \propto T^{-c}$ $(c > 0))$ manifested by straight line on log-log scale. Figure 2(b) demonstrates that for larger sample numbers threshold parameter ε that determines the $\mathbf{C}^{\mathcal{F}}$-'connected' property can be chosen from a broader domain; the dynamic range grows. For the *3D-geom*, the *celebrities* and the *ABC* databases the measured dynamic ranges are 4.45, 5.09 and 2.05 for $20,000$ samples and for the different databases, respectively on the average.

Finally, we illustrate the quality and the working of our method in Fig. 3. The figure depicts the *3D-geom* test and we used $T = 20,000$ samples. According to this figure, the algorithm was able to uncover the hidden components up to the ambiguities of the ISA task.

5 Conclusions

We have introduced a non-combinatorial solution to the estimation of the dimension of the hidden components in the ISA task. We build our method onto the ISA separation theorem and solve the ISA task in 2 steps. First, we perform ICA and then we group the ICA components. The grouping step utilizes a set of non-linear

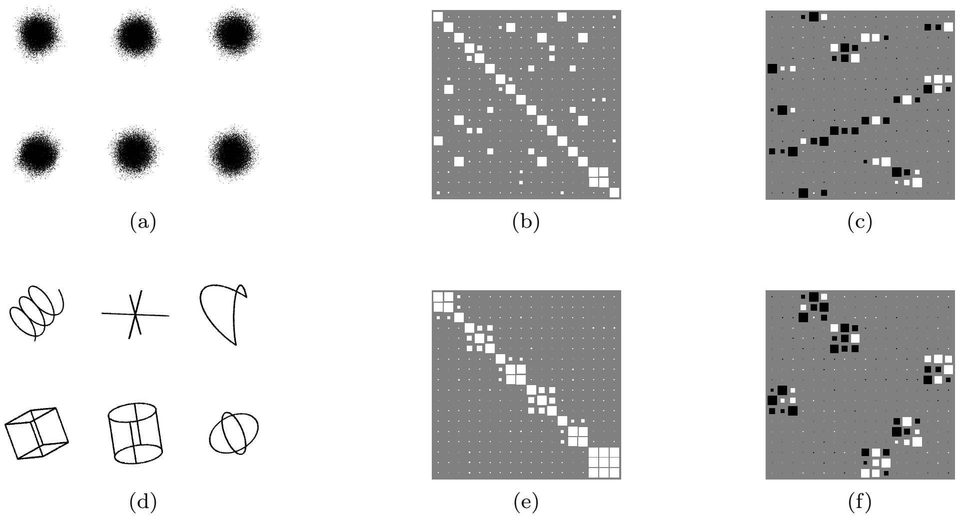


Figure 3: **Illustrations**. (a): observed mixed signal $\mathbf{x}(t)$, (b) $\mathbf{C}^{\mathcal{F}}$ - the sum of absolute values of the elements of the non-linear correlation matrices used for the grouping of the ICA coordinates, (c): the product of the ICA separation matrix and the mixing matrix, (d): estimated components $\mathbf{s}(t)$–up to ambiguities of the ISA problem–, based on (e): $\mathbf{C}^{\mathcal{F}}$ after grouping, (f) product of the estimated ISA separation matrix and the mixing matrix: with high precision, it is a block-permutation matrix made of 3×3 blocks.

correlations between the coordinates of the estimated components. Our simulations indicate that the presently known sufficient conditions of the separation theorem may be extended considerably. This remains to be shown.

References

[1] Akaho, Shotaro, Kiuchi, Yasuhiko, and Umeyama, Shinji. MICA: Multimodal independent component analysis. In *Proceedings of International Joint Conference on Neural Networks (IJCNN '99)*, pages 927–932, 1999.

[2] Amari, Shun-ichi, Cichocki, Andrzej, and Yang, Howard H. A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems*, 8:757–763, 1996.

[3] Bach, Francis R. and Jordan, Michael I. Beyond independent components: Trees and clusters. *Journal of Machine Learning Research*, 4:1205–1233, 2003.

[4] Bell, Anthony J. and Sejnowski, Terrence J. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37:3327–3338, 1997.

[5] Cardoso, Jean-François. Multidimensional independent component analysis. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, volume 4, pages 1941–1944, Seattle, WA, USA, 1998.

[6] Hyvärinen, Aapo. Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, 11:1739–1768, 1999.

[7] Hyvärinen, Aapo. Topographic independent component analysis. *Neural Computation*, 13(7):1527–1558, 2001.

[8] Hyvärinen, Aapo and Hoyer, Patrik O. Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation*, 12:1705–1720, 2000.

[9] Hyvärinen, Aapo and Köster, Urs. FastISA: A fast fixed-point algorithm for independent subspace analysis. In *Proceedings of European Symposium on Artificial Neural Networks (ESANN 2006)*, Bruges, Belgium, 2006.

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

[11] Kiviluoto, Kimmo and Oja, Erkki. Independent component analysis for parallel financial time series. In *Proceedings of International Conference on Neural Information Processing (ICONIP '98)*, volume 2, pages 895–898, 1998.

[12] Makeig, Scott, Bell, Anthony J., Jung, Tzyy-Ping, and Sejnowski, Terrence J. Independent component analysis of electroencephalographic data. In *Proceedings of Neural Information Processing Systems (NIPS '96)*, volume 8, pages 145–151, 1996.

[13] Nolte, Guido, Meinecke, Frank C., Ziehe, Andreas, and Müller, Klaus-Robert. Identifying interactions in mixed and noisy complex systems. *Physical Review E*, 73(051913), 2006.

[14] Póczos, Barnabás and Lőrincz, András. Independent subspace analysis using geodesic spanning trees. In *Proceedings of International Conference on Machine Learning (ICML 2005)*, pages 673–680, Bonn, Germany, 2005.

[15] Póczos, Barnabás and Lőrincz, András. Independent subspace analysis using k-nearest neighborhood distances. *Artificial Neural Networks: Formal Models and their Applications - ICANN 2005, pt 2, Proceedings*, 3697:163–168, 2005.

[16] Póczos, Barnabás and Lőrincz, András. Non-combinatorial estimation of independent autoregressive sources. *Neurocomputing Letters*, 2006.

[17] Póczos, Barnabás, Takács, Bálint, and Lőrincz, András. Independent subspace analysis on innovations. In *European Conference on Machine Learning (ECML 2005)*, volume 3720 of *LNAI*, pages 698–706. Springer Verlag, 2005.

[18] Stögbauer, Harald, Kraskov, Alexander, Astakhov, Sergey A., and Grassberger, Peter. Least dependent component analysis based on mutual information. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 70(066123), 2004.

[19] Szabó, Zoltán and Lőrincz, András. Real and complex independent subspace analysis by generalized variance. In *ICA Research Network International Workshop (ICARN 2006)*, pages 85–88, Liverpool, U.K., 2006. http://arxiv.org/abs/math.ST/0610438.

[20] Szabó, Zoltán, Póczos, Barnabás, and Lőrincz, András. Cross-entropy optimization for independent process analysis. In *Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume 3889 of *LNCS*, pages 909–916. Springer, 2006.

[21] Szabó, Zoltán, Póczos, Barnabás, and Lőrincz, András. Separation theorem for $\mathbb{K}$-independent subspace analysis with sufficient conditions. Technical report, Eötvös Loránd University, Budapest, 2006. http://arxiv.org/abs/math.ST/0608100.

[22] Szabó, Zoltán, Póczos, Barnabás, and Lőrincz, András. Undercomplete blind subspace deconvolution. *Journal of Machine Learning Research*, 8:1063–1095, 2007.

[23] Theis, Fabian J. Uniqueness of complex and multidimensional independent component analysis. *Signal Processing*, 84(5):951–956, 2004.

[24] Theis, Fabian J. Blind signal separation into groups of dependent signals using joint block diagonalization. In *Proceedings of International Society for Computer Aided Surgery (ISCAS 2005)*, pages 5878–5881, Kobe, Japan, 2005.

[25] Theis, Fabian J. Towards a general independent subspace analysis. In *Proceedings of Neural Information Processing Systems (NIPS 2006)*, 2006.

[26] Vigário, Ricardo, Jousmäki, Veikko, Hämäläinen, Matti, Hari, Riitta, and Oja, Erkki. Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In *Proceedings of Neural Information Processing Systems (NIPS '97)*, volume 10, pages 229–235, 1997.

[27] Vollgraf, Roland and Obermayer, Klaus. Multi-dimensional ICA to separate correlated sources. In *Proceedings of Neural Information Processing Systems (NIPS 2001)*, volume 14, pages 993–1000, 2001.