Undercomplete Blind Subspace Deconvolution via Linear Prediction*

Zoltán Szabó, Barnabás Póczos, and András Lőrincz

Department of Information Systems, Eötvös Loránd University,
Pázmány P. sétány 1/C, Budapest H-1117, Hungary
szzoli@cs.elte.hu, pbarn@cs.elte.hu, andras.lorincz@elte.hu

Abstract. We present a novel solution technique for the blind subspace deconvolution (BSSD) problem, where temporal convolution of multidimensional hidden independent components is observed and the task is to uncover the hidden components using the observation only. We carry out this task for the undercomplete case (uBSSD): we reduce the original uBSSD task via linear prediction to independent subspace analysis (ISA), which we can solve. As it has been shown recently, applying temporal concatenation can also reduce uBSSD to ISA, but the associated ISA problem can easily become 'high dimensional' [1]. The new reduction method circumvents this dimensionality problem. We perform detailed studies on the efficiency of the proposed technique by means of numerical simulations. We have found several advantages: our method can achieve high quality estimations for smaller number of samples and it can cope with deeper temporal convolutions.

1 Introduction

There is a growing interest in independent component analysis (ICA) and blind source deconvolution (BSD) for signal processing and hidden component searches. For recent reviews on ICA and BSD see [2] and [3], respectively. The traditional example of ICA is the so-called *cocktail-party problem*, where there are D pieces of *one-dimensional* 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. For example, there could be *independent rock bands* playing at a party. This is the independent subspace analysis (ISA) extension of ICA [4]. Strenuous efforts have been made to develop ISA algorithms, where the theoretical problems concern mostly (i) the estimation of the entropy or of the mutual information, or (ii) joint block diagonalization. A recent list of possible ISA solution techniques can be found in [1].

Another extension of the original ICA task is the BSD problem [3], where the observation is a temporal mixture of the hidden components (*echoic cocktail*

* J.N. Kok et al. (Eds.): ECML 2007, LNAI 4701, pp. 740–747, 2007.
 The original publication is available at http://dx.doi.org/10.1007/978-3-540-74958-5_75.

party). A novel task, the blind subspace deconvolution (BSSD) [1] arises if we combine the ISA and the BSD assumptions. One can think of this task as the separation problem of the pieces played simultaneously by independent rock bands in an echoic stadium. One of the most stringent applications of BSSD could be the analysis of EEG or fMRI signals. The ICA assumptions could be highly problematic here, because some sources may depend on each other, so an ISA model seems better. Furthermore, the passing of information from one area to another and the related delayed and transformed activities may be modeled as echoes. Thus, one can argue that BSSD may fit this important problem domain better than ICA or even ISA. It has been shown in [1] that the undercomplete BSSD task (uBSSD)—where in terms of the cocktail-party problem there are more microphones than acoustic sources—can be reduced to ISA by means of temporal concatenation. However, the reduction technique may lead to 'high dimensions' in the associated ISA problem. Here, an alternative reduction method is introduced for uBSSD and this solution avoids the increase of ISA dimensions. Namely, we show that one can apply linear prediction to reduce the uBSSD task to ISA such that the dimension of the associated ISA problem equals to the dimension of the original hidden sources. As an additional advantage, we shall see that this reduction principle is more efficient on problems with deeper temporal convolutions.

The paper is built as follows: Section 2 formulates the problem domain. Section 3 shows how to reduce the uBSSD task to an ISA problem. Section 4 contains the numerical illustrations. Section 5 contains a short summary.

2 The BSSD Model

We define the BSSD task in Section 2.1. Earlier BSSD reduction principles are reviewed in Section 2.2.

2.1 The BSSD Equations

Here, we define the BSSD task. Assume that we have M hidden, independent, multidimensional *components* (random variables). Suppose also that only their casual FIR filtered mixture is available for observation:

$$\mathbf{x}(t) = \sum_{l=0}^{L} \mathbf{H}_l \mathbf{s}(t-l), \tag{1}$$

where $\mathbf{s}(t) = \left[\mathbf{s}^1(t); \ldots; \mathbf{s}^M(t)\right] \in \mathbb{R}^{Md}$ is a vector concatenated of components $\mathbf{s}^m(t) \in \mathbb{R}^d$. Here, for the sake of notational simplicity we used identical dimension for each component. For a given m, $\mathbf{s}^m(t)$ is i.i.d. (independent and identically distributed) in time t, there is at most a single Gaussian component in $\mathbf{s}^m$s, and $I(\mathbf{s}^1, \ldots, \mathbf{s}^M) = 0$, where I stands for the mutual information of the arguments. The total dimension of the components is $D_s := Md$, the dimension of the observation $\mathbf{x}$ is D_x. Matrices $\mathbf{H}_l \in \mathbb{R}^{D_x \times D_s}$ $(l = 0, \ldots, L)$ describe the convolutive mixing. Without any loss of generality it may be assumed that $E[\mathbf{s}] = \mathbf{0}$, where E denotes the expectation value. Then $E[\mathbf{x}] = \mathbf{0}$ holds, as well. The goal of the BSSD problem is to estimate the original source $\mathbf{s}(t)$ by using observations $\mathbf{x}(t)$ only. The case $L = 0$ corresponds to the ISA task, and if $d = 1$ also holds then the ICA task is recovered. In the BSD task $d = 1$ and L is a non-negative integer. $D_x > D_s$ is the *undercomplete*, $D_x = D_s$ is the *complete*, and $D_x < D_s$ is the *overcomplete* task. Here, we treat the undercomplete BSSD (uBSSD) problem.

For consecutive reductional steps we rewrite the BSSD model using operators. Let $\mathbf{H}[z] := \sum_{l=0}^{L} \mathbf{H}_l z^{-l} \in \mathbb{R}[z]^{D_x \times D_s}$ denote the $D_x \times D_s$ polynomial matrix corresponding to the convolutive mixing, in a one-to-one manner, where z is the time-shift operation. Now, the BSSD equation (1) can be written as $\mathbf{x} = \mathbf{H}[z]\mathbf{s}$. In the uBSSD task it is assumed that $\mathbf{H}[z]$ has a polynomial matrix left inverse: there exists polynomial matrix $\mathbf{W}[z] \in \mathbb{R}[z]^{D_s \times D_x}$ such that $\mathbf{W}[z]\mathbf{H}[z]$ is the identity mapping. It can be shown [5] that for $D_x > D_s$ such a left inverse exists with probability 1, under mild conditions: coefficients of polynomial matrix $\mathbf{H}[z]$, that is, the random matrix $[\mathbf{H}_0; \ldots; \mathbf{H}_L]$ is drawn from a continuous distribution. For the ISA task it is supposed that mixing matrix $\mathbf{H}_0 \in \mathbb{R}^{D_x \times D_s}$ has full column rank, i.e., its rank is D_s.

2.2 Existing Decomposition Principles in the BSSD Problem Family

There are numerous reduction methods for the BSSD problem in the literature. For example, its special case, the undercomplete BSD task can be reduced (i) to ISA by temporal concatenation of the observations [6], or (ii) to ICA by means of either spatio-temporal decorrelation [7], or by linear prediction (autoregressive (AR) estimation), see e.g., [8]. As it was shown in [1], the uBSSD task can also be reduced to ISA by temporal concatenation. In Section 3, we show another route and describe how linear prediction can help to transcribe the uBSSD task to ISA. According to the ISA Separation Theorem [1, 9], under certain conditions, the solution of the ISA task requires an ICA preprocessing step followed by a suitable permutation of the ICA elements. This principle was conjectured in [4] on basis of numerical simulations. Only sufficient conditions are available in [1,9] for the ISA Separation Theorem.

3 Reduction of uBSSD to ISA by Linear Prediction

Below, we reduce the uBSSD task to ISA by means of linear prediction. The procedure is similar to that of [8], where it was applied for undercomplete BSD (i.e., for $d = 1$).

Theorem. *In the uBSSD task, observation process* $\mathbf{x}(t)$ *is autoregressive and its innovation* $\tilde{\mathbf{x}}(t) := \mathbf{x}(t) - E[\mathbf{x}(t)|\mathbf{x}(t-1), \mathbf{x}(t-2), \ldots]$ *is* $\mathbf{H}_0\mathbf{s}(t)$*, where* $E[\cdot|\cdot]$ *denotes the conditional expectation value. Consequently, there is a polynomial matrix* $\mathbf{W}_{\mathrm{AR}}[z] \in \mathbb{R}[z]^{D_x \times D_x}$ *such that* $\mathbf{W}_{\mathrm{AR}}[z]\mathbf{x} = \mathbf{H}_0\mathbf{s}$.

Proof. We assumed that $\mathbf{H}[z]$ has left inverse, thus the hidden $\mathbf{s}$ can be expressed from observation $\mathbf{x}$ by causal FIR filtering, i.e., $\mathbf{s} = \mathbf{H}^{-1}[z]\mathbf{x}$, where $\mathbf{H}^{-1}[z] = \sum_{n=0}^{N} \mathbf{M}_n z^{-n} \in \mathbb{R}[z]^{D_s \times D_x}$ and N denotes the degree of the $\mathbf{H}^{-1}[z]$ polynomial. Thus, terms in observation $\mathbf{x}$ that differ from $\mathbf{H}_0\mathbf{s}(t)$ in (1) belong to the linear hull of the finite history of $\mathbf{x}$: $\mathbf{x}(t) = \mathbf{H}_0\mathbf{s}(t) + \sum_{l=1}^{L} \mathbf{H}_l(\mathbf{H}^{-1}[z]\mathbf{x})(t - l) \in \mathbf{H}_0\mathbf{s}(t) + \langle \mathbf{x}(t-1), \mathbf{x}(t-2), \ldots, \mathbf{x}(t-L+N)\rangle$. Because $\mathbf{s}(t)$ is independent of $\langle \mathbf{x}(t-1), \mathbf{x}(t-2), \ldots, \mathbf{x}(t-L+N)\rangle$, we have that observation process $\mathbf{x}(t)$ is autoregressive with innovation $\mathbf{H}_0\mathbf{s}(t)$.

Thus, the AR fit of $\mathbf{x}(t)$ can be used for the estimation of $\mathbf{H}_0\mathbf{s}(t)$. This innovation corresponds to the observation of an undercomplete ISA model[1], which can be reduced to a complete ISA model using principal component analysis (PCA). Finally, the solution can be finished by any ISA procedure. We will call the above uBSSD method linear predictive approximation (LPA). The LPA pseudocode is given in Table 1. The reduction procedure implies that hidden components $\mathbf{s}^m$ can be recovered only up to the ambiguities of the ISA task [10]: that is, assuming (without any loss of generality) that both the hidden source ($\mathbf{s}$) and the observation are white – their expectation values are zeroes and the covariance matrices are identities – the $\mathbf{s}^m$ components are determined up to permutation and orthogonal transformation.

Table 1: Linear predictive approximation (LPA): Pseudocode

Input of the algorithm
Observation: $\{\mathbf{x}(t)\}_{t=1,\ldots,T}$
Optimization
AR fit: for observation $\mathbf{x}$ estimate $\hat{\mathbf{W}}_{\mathrm{AR}}[z]$
Estimate innovation: $\tilde{\mathbf{x}} = \hat{\mathbf{W}}_{\mathrm{AR}}[z]\mathbf{x}$
Reduce uISA to ISA and whiten: $\tilde{\mathbf{x}}' = \hat{\mathbf{W}}_{\mathrm{PCA}}\tilde{\mathbf{x}}$
Apply ISA for $\tilde{\mathbf{x}}'$: separation matrix is $\hat{\mathbf{W}}_{\mathrm{ISA}}$
Estimation
$\hat{\mathbf{W}}_{\mathrm{uBSSD}}[z] = \hat{\mathbf{W}}_{\mathrm{ISA}}\hat{\mathbf{W}}_{\mathrm{PCA}}\hat{\mathbf{W}}_{\mathrm{AR}}[z]$
$\hat{\mathbf{s}} = \hat{\mathbf{W}}_{\mathrm{uBSSD}}[z]\mathbf{x}$

4 Illustrations

We show the results of our studies concerning the efficiency of the algorithm of Table 1.[2] We compare the LPA procedure with the uBSSD method described in [1], where temporal concatenation was applied to transform the uBSSD task

[1] Assumptions made for $\mathbf{H}[z]$ in the uBSSD task implies that $\mathbf{H}_0$ is of full column rank and thus the resulting ISA task is well defined.

[2] Further details can be found in our accompanying technical report [11].

to a 'high-dimensional' ISA task. We shall refer to that method as the method of temporal concatenation, or TCC for short. Test problems are introduced in Section 4.1. The performance index that we use to measure the quality of the solutions is detailed in Section 4.2. Numerical results are presented in Section 4.3.

4.1 Databases

We define four databases ($\mathbf{s}$) to study our LPA algorithm[3]. These are the databases used in [1], too. In the *3D-geom*, *celebrities* and *letters* data sets, the d-dimensional hidden components $\mathbf{s}^m$ are 3,2,2-dimensional random variables, respectively. They are distributed (a) uniformly on geometric forms, (b) according to pixel intensities on cartoons of celebrities, and (c) uniformly on images of letters A and B. We have $M = 6, 10, 2$ components, thus the dimension of the hidden source $\mathbf{s}$ is $D_s = 18, 20, 4$. Databases are illustrated in Figs. 1(a)-(c). Our *Beatles* database is a non-i.i.d. example. Here, hidden sources are stereo Beatles songs. 8 kHz sampled portions of two songs (A Hard Day's Night, Can't Buy Me Love) made the hidden $\mathbf{s}^m$s ($d = 2, M = 2, D_s = 4$). In the *letters* and *Beatles* test the number of components and their dimensions were minimal ($d = 2$, $M = 2$).

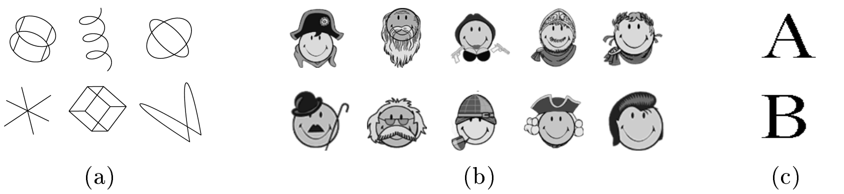

Fig. 1: Illustration of the (a) *3D-geom*, (b) *celebrities* and (c) *letters* databases.

4.2 The Amari-index

According to Section 3, in the ideal case, the product of matrix $\hat{\mathbf{W}}_{\text{ISA}}\hat{\mathbf{W}}_{\text{PCA}}$ and matrix $\mathbf{H}_0$, that is matrix $\mathbf{G} := \hat{\mathbf{W}}_{\text{ISA}}\hat{\mathbf{W}}_{\text{PCA}}\mathbf{H}_0 \in \mathbb{R}^{D_s \times D_s}$ is a block-permutation matrix made of $d \times d$ blocks. To measure this block-permutation property, the Amari-error adapted to the ISA task [12] was normalized [9] to take values in $[0, 1]$ *independently from d and D_s*. This performance measure, the Amari-index, was used to compare the TCC and LPA techniques.

[3] *Smiley*: http://www.smileyworld.com, *Beatles*: http://rock.mididb.com/beatles/ .

4.3 Simulations

The experimental studies concern two questions: (i) the TCC and the LPA methods are compared on uBSSD tasks, (ii) the performance as a function of convolution length is studied for the LPA technique.

We studied the $D_x = 2D_s$ case, like in [1]. Both the TCC and the LPA method reduce the uBSSD task to ISA problems and we use the Amari-index (Section 4.2) to measure and compare their performances. For all values of the parameters (sample number: T, convolution length: $L+1$), we have averaged the performances upon 50 random initializations of $\mathbf{s}$ and $\mathbf{H}[z]$. The coordinates of matrices $\mathbf{H}_l$ were chosen independently from standard normal distribution. We used the Schwarz's Bayesian Criterion to determine the optimal order of the AR process. The criterion was constrained: the order Q of the estimated AR process (see Table 1) was limited from above, the upper limit was set to twice the length of the convolution, i.e., $Q \leq 2(L+1)$. The AR process and the ISA subtask of TCC and LPA were estimated by the method detailed in [13], and by joint f-decorrelation (JFD) [14], respectively.

We studied the dependence of the precision versus the sample number. In the *3D-geom* and *celebrities* (*letters* and *Beatles*) tests, the sample number T varied between $1,000$ and $100,000$ ($1,000$ and $75,000$), the length of the convolution $(L+1)$ changed between 2 and 6 (2 and 31). Comparison with the TCC method and the estimations of the LPA technique are illustrated in Figs. 2(a)-(b) (Figs. 2(c)-(d)) on the *3D-geom* (*Beatles*) database. According to Fig. 2(a), the LPA algorithm is able to uncover the hidden components with high precisions on the *3D-geom* database. We found that the Amari-index r decreases according to power law $r(T) \propto T^{-c}$ $(c > 0)$ for sample numbers $T > 2000$. The power law is manifested by straight lines on $\log\log$ scales. According to Fig. 2(b) the LPA method is superior to the TCC method (i) for all sample numbers $1,000 \leq T \leq 100,000$, moreover (ii) LPA can provide reasonable estimates for much smaller sample numbers. on the *3D-geom* database. This behavior is manifested by the initial steady increase of the quotients of the Amari indices of the TCC and LPA methods as a function of sample number followed by a sudden drop when the sample number enables reasonable TCC estimations, too. Similar results were found on the *celebrities* and the *letters* databases, too. The LPA method resulted in $1.1-88$, $1.0-87$, $1.2-110$-times increase of precision for the *3D-geom*, *celebrities* and *letters* database, respectively. For the *3D-geom* (*celebrities*, *letters*) dataset the Amari-index for sample number $T = 100,000$ ($T = 100,000$, $T = 75,000$) is $0.19-0.20\%$ ($0.33-0.34\%$, $0.30-0.36\%$) with small $0.01-0.02$ (0.02, $0.11-0.15$) standard deviations.

Visual inspection of Fig. 2(c) shows that on the *Beatles* database the LPA method found the hidden components for sample number $T \geq 30,000$. The TCC method gave reliable solutions for sample number $T = 50,000$ or so. According to Fig. 2(d) the LPA method is more precise than TCC for $T \geq 30,000$. The increase in precision becomes more pronounced for larger convolution parameter L. Namely, for sample number $75,000$ and for $L = 1,2,5,10,20,30$ the ratios of precision are $1.50, 2.24, 4.33, 4.42, 9.03, 11.13$, respectively on the average. For sample number $T = 75,000$ the Amari-index stays below 1% on average ($0.4 - 0.71\%$) and has $0.02 - 0.08$ standard deviation for the *Beatles* test.

According to our simulations, the LPA method may provide acceptable estimations for sample number $T = 20,000$ ($T = 15,000$) up to convolution length $L = 20$ ($L = 230$) for the *3D-geom* and *celebrities* (*letters* and *Beatles*) datasets. Such estimations are shown in Fig. 3(d) and Fig. 3(h) for the *3D-geom* and *letters* tests, respectively.

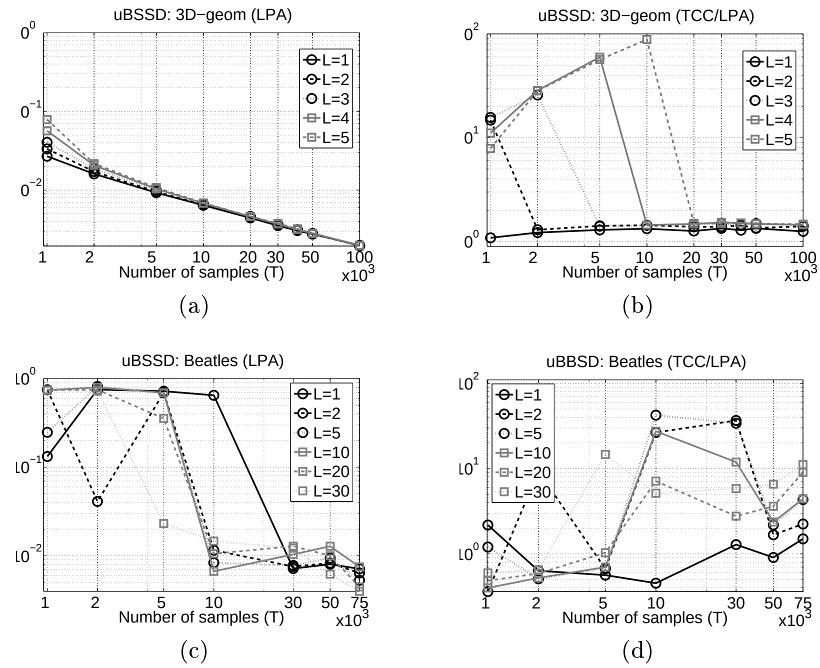

Fig. 2: Estimation error of the LPA method and comparisons with the TCC method for the *3D-geom* and *Beatles* databases. Scales are 'log log' plots. Data correspond to different convolution lengths $(L + 1)$. (a) and (c): Amari-index as a function of the sample number. (b) and (d): Quotients of the Amari-indices of the TCC and the LPA methods: for quotient value $q > 1$, the LPA method is q times more precise than the TCC method. In the *celebrities* and *letters* tests, we found similar results as on the *3D-geom* data set.

5 Summary

We presented a novel solution method for the undercomplete case of the blind subspace deconvolution (uBSSD) task. We used a stepwise decomposition principle and reduced the problem with linear prediction to independent subspace

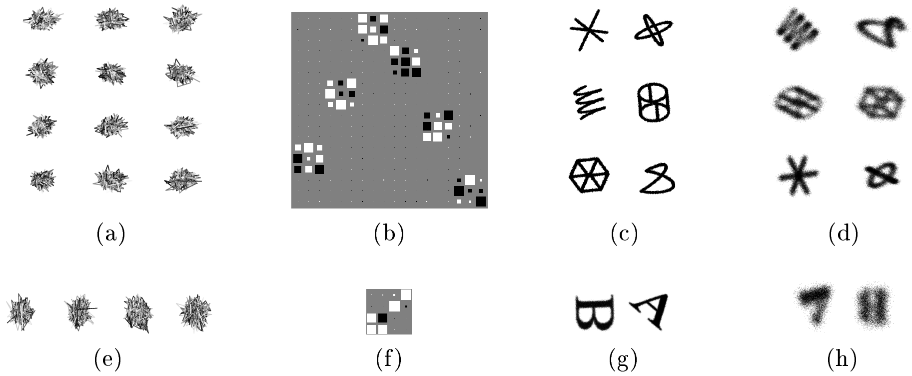

Fig. 3: Illustration of the LPA method on the uBSSD task for the *3D-geom* (*letters*) database. (a)-(c) [(e)-(g)]: sample number $T = 100,000$ [75,000], convolution length $L + 1 = 6$ [31]. (a), (e): observed convolved signals $\mathbf{x}(t)$. (b) [(f)]: Hinton-diagram of $\mathbf{G}$, ideally block-permutation matrix with 3×3 [2×2] blocks. (c) [(g)]: estimated components ($\hat{\mathbf{s}}^m$), Amari-index: 0.2% [0.3%]. (d) [(h)]: estimation of hidden components ($\hat{\mathbf{s}}^m$) for sample number $T = 20,000$ [15,000] and convolution parameter $L = 20$ [230].

analysis (ISA) task. We illustrated the method on different tests. Our method supersedes the temporal concatenation based uBSSD method, because (i) it gives rise to a smaller dimensional ISA task, (ii) it produces similar estimation errors at considerably smaller sample numbers, and (iii) it can treat deeper temporal convolutions.

References

1. Szabó, Z., Póczos, B., Lőrincz, A.: Undercomplete blind subspace deconvolution. Journal of Machine Learning Research **8** (2007) 1063–1095
2. Cichocki, A., Amari, S.: Adaptive blind signal and image processing. John Wiley & Sons (2002)
3. Pedersen, M.S., Larsen, J., Kjems, U., Parra, L.C.: A survey of convolutive blind source separation methods. In: Springer Handbook of Speech (to appear). Springer Press (2007) (http://www2.imm.dtu.dk/pubdb/p.php?4924).
4. Cardoso, J.: Multidimensional independent component analysis. In: ICASSP '98. (Volume 4.) 1941–1944
5. Rajagopal, R., Potter, L.C.: Multivariate MIMO FIR inverses. IEEE Transactions on Image Processing **12** (2003) 458 – 465
6. Févotte, C., Doncarli, C.: A unified presentation of blind source separation for convolutive mixtures using block-diagonalization. In: ICA '03. (2003) 349–354
7. Choi, S., Cichocki, A.: Blind signal deconvolution by spatio-temporal decorrelation and demixing. Neural Networks for Signal Processing **7** (1997) 426–435
8. Gorokhov, A., Loubaton, P.: Blind identification of MIMO-FIR systems: A generalized linear prediction approach. Signal Processing **73** (1999) 105–124

9. Szabó, Z., Póczos, B., Lőrincz, A.: Cross-entropy optimization for independent process analysis. In: ICA '06. Volume 3889 of LNCS., Springer (2006) 909–916
10. Theis, F.J.: Uniqueness of complex and multidimensional independent component analysis. Signal Processing **84** (2004) 951–956
11. Szabó, Z., Póczos, B., Lőrincz, A.: Undercomplete blind subspace deconvolution via linear prediction. Technical report, Eötvös Loránd University, Budapest (2007) (http://arxiv.org/abs/0706.3435).
12. Theis, F.J.: Blind signal separation into groups of dependent signals using joint block diagonalization. In: ISCAS '05. (2005) 5878–5881
13. Neumaier, A., Schneider, T.: Estimation of parameters and eigenmodes of multivariate AR models. ACM Trans. on Mathematical Software **27** (2001) 27–57
14. Szabó, Z., Lőrincz, A.: Real and complex independent subspace analysis by generalized variance. In: ICARN '06. (2006) 85–88