Complete Blind Subspace Deconvolution*

Zoltán Szabó

Department of Information Systems, Eötvös Loránd University,
Pázmány P. sétány 1/C, Budapest H-1117, Hungary
szzoli@cs.elte.hu
http://nipg.inf.elte.hu

Abstract. In this paper we address the blind subspace deconvolution (BSSD) problem; an extension of both the blind source deconvolution (BSD) and the independent subspace analysis (ISA) tasks. While previous works have been focused on the undercomplete case, here we extend the theory to complete systems. Particularly, we derive a separation technique for the complete BSSD problem: we solve the problem by reducing the estimation task to ISA via linear prediction. Numerical examples illustrate the efficiency of the proposed method.

Key words: complete blind subspace deconvolution, separation principle, linear prediction, independent subspace analysis

1 Introduction

Recently, research on independent component analysis (ICA) [1, 2] and its extensions has gained much attention. One can think of ICA as a cocktail-party problem, where there are D *one-dimensional* sound sources and D microphones and the task is to estimate the original sources from the observed mixed signals. Nonetheless, applications in which only certain groups of sources are independent may be highly relevant in practice. In this case, the independent sources can be multidimensional. For instance, consider the generalization of the cocktail-party problem, where *independent groups* of people are talking separately about independent topics or more than one group of musicians are playing at the party. This problem is referred to as independent subspace analysis (ISA) [3].[1] Another extension of the original ICA task is the blind source deconvolution (BSD) problem. This problem emerges, for example, when a cocktail-party takes place in an *echoic* room. Several theoretical questions of ICA, ISA and BSD have already been addressed (see, e.g., [4], [5] and [6] for recent reviews, respectively), and numerous application areas show the potential of these fields including (i) remote sensing applications: passive radar/sonar processing, (ii) image-deblurring,

* T. Adali et al. (Eds.): ICA 2009, LNCS 5441, pp. 138–145, 2009. © Springer-Verlag Berlin Heidelberg 2009. The original publication is available at http://dx.doi.org/10.1007/978-3-642-00599-2_18.

[1] ISA is also called multidimensional independent component analysis and group ICA in the literature.

image restoration, (iii) speech enhancement using microphone arrays, acoustics, (iv) multi-antenna wireless communications, sensor networks, (v) financial, gene and biomedical signal—EEG, ECG, MEG, fMRI—analysis, (vi) face view recognition, (vii) optics, (viii) seismic exploration.

The simultaneous assumption of the two extensions, that is, ISA combined with BSD, has recently emerged in the literature. For example, at the cocktail-party, groups of people or groups of musicians may form *independent source groups* and *echoes* may be present. The task is called blind subspace deconvolution (BSSD). Probably one of the most exciting and fundamental hypotheses of the ICA research has been formed by [3]: the solution of the ISA problem can be *separated* to ICA and then clustering the ICA elements into statistically dependent subspaces. This ISA *separation principle* has been rigorously proven for some distribution types in [5], and forms the basis of the state-of-the-art ISA algorithms. Similar separation based techniques can be derived for the solution of the *undercomplete* BSSD task (uBSSD), where in terms of the cocktail-party problem there are more microphones than acoustic sources. It has been shown that the uBSSD problem can be reduced to ISA by means of temporal concatenation [5]. However, the associated ISA problem can easily become 'high dimensional'. The dimensionality problem can be circumvented by applying a linear predictive approximation (LPA) based reduction [7]. Here, we show that it is possible to extend the LPA idea to the *complete* BSSD task.[2] In the undercomplete case, the LPA based solution was based on the observation that the polynomial matrix describing the temporal convolution had, under rather general conditions[3], a polynomial matrix left inverse. In the complete case such an inverse doesn't exist in general. However, provided that the convolution can be represented by an infinite order autoregressive process, one can construct an efficient estimation method for the hidden components via an asymptotically consistent LPA procedure. This thought is used here to extend the technique of [7] to the complete case.

The paper is structured as follows: Section 2 formulates the problem domain. Section 3 shows how to transform the complete BSSD task into an ISA task via LPA. Section 4 contains numerical illustrations. Conclusions are drawn in Section 5.

2 The BSSD Model

Here, we define the BSSD task [5]. Assume that we have M hidden, independent, multidimensional *components* (random variables). Suppose also that only their

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

[2] The overcomplete BSSD task is challenging and as of yet no general solution is known.

[3] If the coefficients of the undercomplete polynomial matrix are drawn from a non-degenerate continuous distribution, such an inverse exists with probability one.

convolutive mixture is available for observation, where $\mathbf{x}(t) \in \mathbb{R}^{D_x}$ and $\mathbf{s}(t)$ is the concatenation of the components $\mathbf{s}^m(t) \in \mathbb{R}^{d_m}$, that is $\mathbf{s}(t) = [\mathbf{s}^1(t); \ldots; \mathbf{s}^M(t)] \in \mathbb{R}^{D_s}$ ($D_s = \sum_{m=1}^M d_m$). Denoting the time-shift operation by z, one may write Eq. (1) compactly as

$$\mathbf{x} = \mathbf{H}[z]\mathbf{s}, \tag{2}$$

where the mixing process is described by the polynomial matrix $\mathbf{H}[z] := \sum_{l=0}^{L} \mathbf{H}_l z^l \in \mathbb{R}[z]^{D_x \times D_s}$. We assume that the components $\mathbf{s}^m$ are

1. independent: $I(\mathbf{s}^1, \ldots, \mathbf{s}^M) = 0$, where I denotes the mutual information,
2. i.i.d. (independent identically distributed) in t, and
3. there is at most one Gaussian component among $\mathbf{s}^m$s.

The goal of the BSSD problem is to estimate the original source $\mathbf{s}(t)$ by using observations $\mathbf{x}(t)$ only. While $D_x > D_s$ is the *undercomplete* case , $D_x = D_s$ is the *complete* one. The case $L = 0$ corresponds to the ISA task, and if $d_m = 1$ ($\forall m$) also holds, then the ICA task is recovered. In the BSD task $d_m = 1$ ($\forall m$) and L is a non-negative integer.

3 Method

Contrary to previous works [5, 7] focusing on the undercomplete BSSD problem, in the present paper we address the complete task ($D = D_x = D_s$). We assume that the polynomial matrix $\mathbf{H}[z]$ is *invertible*, that is $\det(\mathbf{H}[z]) \neq 0$, for all $z \in \mathbb{C}, |z| \leq 1$. Let $E(\cdot)$ and $cov(\cdot)$ denote the expectation value, and the covariance of a random variable, respectively. Because the mean can be subtracted from the process and the transformation $\mathbf{x} = (\mathbf{H}[z]\mathbf{B}^{-1})(\mathbf{B}\mathbf{s})$ leads to the same observation, one may presume, without any loss of generality, that $\mathbf{s}$ is white:

$$E(\mathbf{s}) = \mathbf{0}, \quad cov(\mathbf{s}) = \mathbf{I}, \tag{3}$$

where $\mathbf{I}$ is the identity matrix. The invertibility of $\mathbf{H}[z]$ implies that the observation process $\mathbf{x}$ can be represented as an infinite order autoregressive (AR) process [8]:

$$\mathbf{x}(t) = \sum_{j=1}^{\infty} \mathbf{F}_j \mathbf{x}(t-j) + \mathbf{F}_0 \mathbf{s}(t). \tag{4}$$

By applying a finite order LPA approximation (fitting an AR process to $\mathbf{x}$), the innovation process $\mathbf{F}_0\mathbf{s}(t)$ can be estimated. The innovation can be seen as the observation of an ISA problem because components of $\mathbf{s}$ are independent: ISA techniques can be used to identify components $\mathbf{s}^m$. Choosing the order of the fitted AR process to $\mathbf{x}$ as $p = o(T^{\frac{1}{3}}) \xrightarrow{T \to \infty} \infty$, where T denotes the number of samples, guarantees that the AR approximation for the MA model is asymptotically consistent [9].

4 Illustrations

Here, we illustrate the efficiency of the proposed complete BSSD estimation technique. Test cases are introduced in Section 4.1. To evaluate the solutions we use the performance measure given in Section 4.2. Numerical results are presented in Section 4.3.

4.1 Databases

We define three databases to study our identification algorithm. The *smiley* test has 2-dimensional source components representing the 6 basic facial expressions ($d_m = 2$, $M = 6$). Sources $\mathbf{s}^m$ were generated by sampling 2-dimensional coordinates proportional to the corresponding pixel intensities (see Fig. 1(a)). In the *3D-geom* test $\mathbf{s}^m$s were random variables uniformly distributed on 3-dimensional geometric forms ($d_m = 3$). We chose 4 different components ($M = 4$) and, as a result, the dimension of the hidden source $\mathbf{s}$ is $D = 12$ (see Fig. 1(b)). Our *Beatles* test [5] is a non-i.i.d. example. Here, hidden sources are stereo Beatles songs.[4] 8 kHz sampled portions of two songs (A Hard Day's Night, Can't Buy Me Love) made the hidden $\mathbf{s}^m$s. Thus, the dimension of the components d_m was 2, the number of the components M was 2, and the dimension of the hidden source D was 4.

(a) (b)

Fig. 1: Illustration of the *smiley* (a) and the *3D-geom* databases (b).

4.2 Performance Measure, the Amari-index

Recovery of components $\mathbf{s}^m$ are subject to the ambiguities of the ISA task. Namely, components of equal dimension can be recovered up to permutation and invertible transformation within the subspaces [10]. For this reason, in the ideal case, the linear transformation $\mathbf{G}$ that optimally approximates the relation $\mathbf{s} \mapsto \hat{\mathbf{s}}$, where $\hat{\mathbf{s}}$ denotes the estimated hidden source, resides also within the subspaces and so it is a *block-permutation matrix*. This block-permutation structure can be measured by the ISA adapted version [11] of the Amari-error [12] normalized to the interval $[0, 1]$ [13]. Namely, let us suppose that the source components

[4] See http://rock.mididb.com/beatles/.

are d-dimensional[5], and let us decompose matrix $\mathbf{G} \in \mathbb{R}^{D\times D}$ into blocks of size $d \times d$: $\mathbf{G} = [\mathbf{G}_{ij}]_{i,j=1,\ldots,M}$. Let g_{ij} denote the sum of the absolute values of matrix $\mathbf{G}_{ij} \in \mathbb{R}^{d\times d}$. Now, the following term

$$r(\mathbf{G}) := \frac{1}{2M(M-1)}\left[\sum_{i=1}^{M}\left(\frac{\sum_{j=1}^{M} g_{ij}}{\max_j g_{ij}} - 1\right) + \sum_{j=1}^{M}\left(\frac{\sum_{i=1}^{M} g_{ij}}{\max_i g_{ij}} - 1\right)\right] \qquad (5)$$

denotes the *Amari-index* that takes values in [0,1]: for an ideal block-permutation matrix $\mathbf{G}$ it takes 0; for the worst case it takes 1.

4.3 Simulations

Results on databases *smiley*, *3D-geom* and *Beatles* are provided here. The Amari-index was used to measure the performance of the proposed complete BSSD method. For each individual parameter, 20 random runs were averaged. Our parameters are: T, the sample number of observations $\mathbf{x}(t)$, L, the parameter of the length of the convolution (the length of the convolution is $L+1$), and λ, parameter of the invertible $\mathbf{H}[z]$. It is expected that if the roots of $\mathbf{H}[z]$ are close to the unit circle then our estimation will deteriorate, because the invertibility of $\mathbf{H}[z]$ comes to question. We investigated this by generating the polynomial matrix $\mathbf{H}[z]$ as follows:

$$\mathbf{H}[z] = \left[\prod_{l=0}^{L}(\mathbf{I} - \lambda\mathbf{O}_i z)\right]\mathbf{H}_0 \quad (|\lambda| < 1, \lambda \in \mathbb{R}), \qquad (6)$$

where matrices $\mathbf{H}_0$ and $\mathbf{O}_i \in \mathbb{R}^{D\times D}$ were random orthogonal and the $\lambda \to 1$ limit was studied. 'Random run' means random choice of quantities $\mathbf{H}[z]$ and $\mathbf{s}$. The AR fit to observation $\mathbf{x}$ was performed by the method detailed in [14]. To study how the $o(T^{1/3})$ AR order (see Section 3) is exploited, the order of the estimated AR process was limited from above by $p_{max}(T) = 2\lfloor T^{\frac{1}{3}-\frac{1}{1000}}\rfloor$, and we used the Schwarz's Bayesian Criterion to determine the optimal p_{opt} order from the interval $[1, p_{max}(T)]$. The ISA subtask on the estimated innovation was carried out by the joint f-decorrelation method [15].

First we studied the Amari-index as a function of the sample size. For the *smiley* and *3D-geom* databases the sample number T varied between $1,000$ and $20,000$. The length of convolution varied as $L = 1, 2, 5, 10$. The λ parameter of $\mathbf{H}[z]$ was chosen as $0.4, 0.6, 0.7, 0.8, 0.85, 0.9$. Results are shown in Fig. 2(a)-(b). The estimation errors indicate that for $L = 10$ and about $\lambda = 0.85$ the estimation is still efficient, see Fig. 3 for an illustration of the estimated source components. The Amari-indices follow the power law $r(T) \propto T^{-c}$ $(c > 0)$. The power law decline is manifested by straight line on log-log scale. The slopes of these straight lines are very close to each other. Numerical values for the estimation errors are

[5] The $d = d_m$ $(\forall m)$ constraint was used only at the performance measurements (i.e., for the Amari-index).

given in Table 1. The estimated optimal AR orders are provided in Fig. 2(c). The figure demonstrates that as $\lambda \to 1$ the maximal possible order $p_{max}(T)$ is more and more exploited.

On the *Beatles* database the λ parameter was increased to 0.9, and the sample number T varied between $1,000$ and $100,000$. Results are presented in Fig. 2(d). According to the figure, for $L = 1, 2, 5$ the error of estimation drops for sample number $T = 10,000 - 20,000$, and for $L = 10$ the 'power law' decline of the Amari-index, which was apparent on the *smiley* and the *3D-geom* databases, also appears. Numerical values for the estimation errors are given in Table 1. On the *Beatles* test, the maximal possible AR order $p_{max}(T)$ was fully exploited on the examined parameter domain.

	$L = 1$	$L = 2$	$L = 5$	$L = 10$
smiley	0.99% (±0.11%)	1.04% (±0.09%)	1.22% (±0.15%)	1.69% (±0.26%)
3D-geom	0.42% (±0.06%)	0.54% (±0.05%)	0.88% (±0.14%)	1.15% (±0.24%)
Beatles	0.72% (±0.12%)	0.75% (±0.11%)	0.90% (±0.23%)	6.64% (±7.49%)

Table 1: Amari-index in percentages on the *smiley*, *3D-geom* ($\lambda = 0.85, T = 20,000$) and the Beatles dataset ($\lambda = 0.9, T = 100,000$) for different convolution lengths: mean± standard deviation. For other sample numbers, see Fig. 2.

5 Conclusions

In this paper we focused on the complete case of the blind subspace deconvolution (BSSD) problem, a common extension of the independent subspace analysis (ISA) and the blind source deconvolution (BSD) tasks. We presented a separation technique for the solution of the complete BSSD task: the estimation task has been reduced to ISA via linear predictive approximation (LPA). We also demonstrated the efficiency of the algorithm on different datasets. Our simulations revealed that the error of the estimation of the hidden sources decreases in a power law fashion as the sample size increases. Interestingly, our algorithm recovered the sources when the assumptions of the BSSD problem were violated; that is in the case of the *Beatles* songs with non-i.i.d. sources. This result points to the ISA separation principle; one expects that it may be valid for a larger domain. Similar conjecture exists for joint block diagonalization [16] about the *global* minima.

Acknowledgments. This work has been supported by the National Office for Research and Technology.

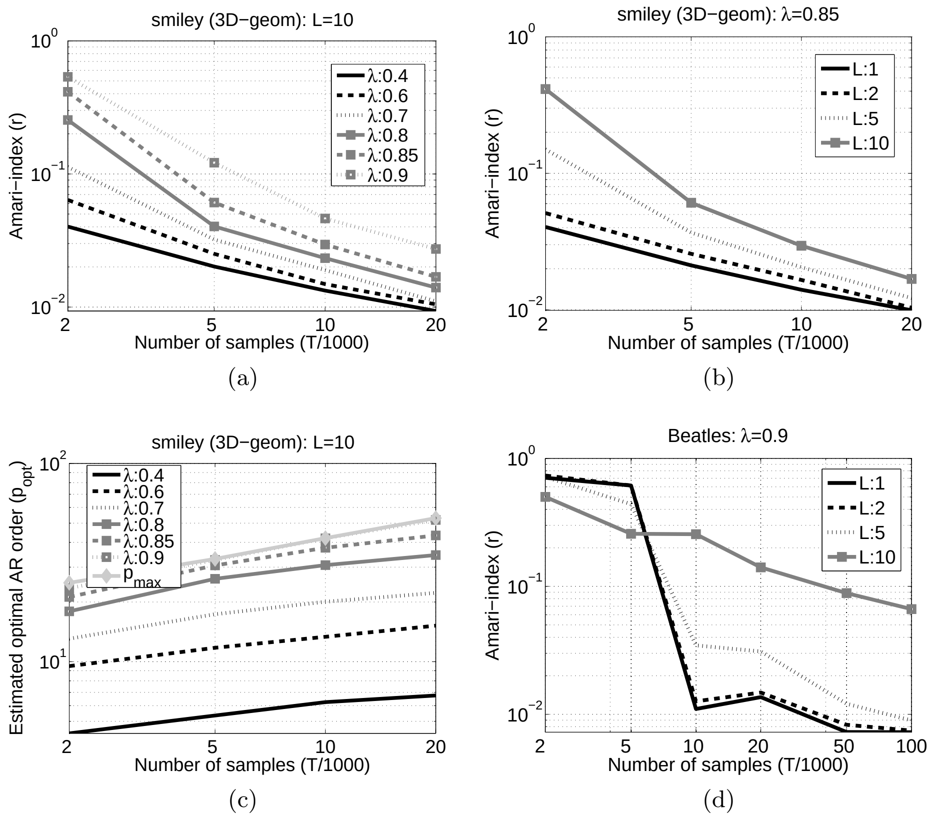
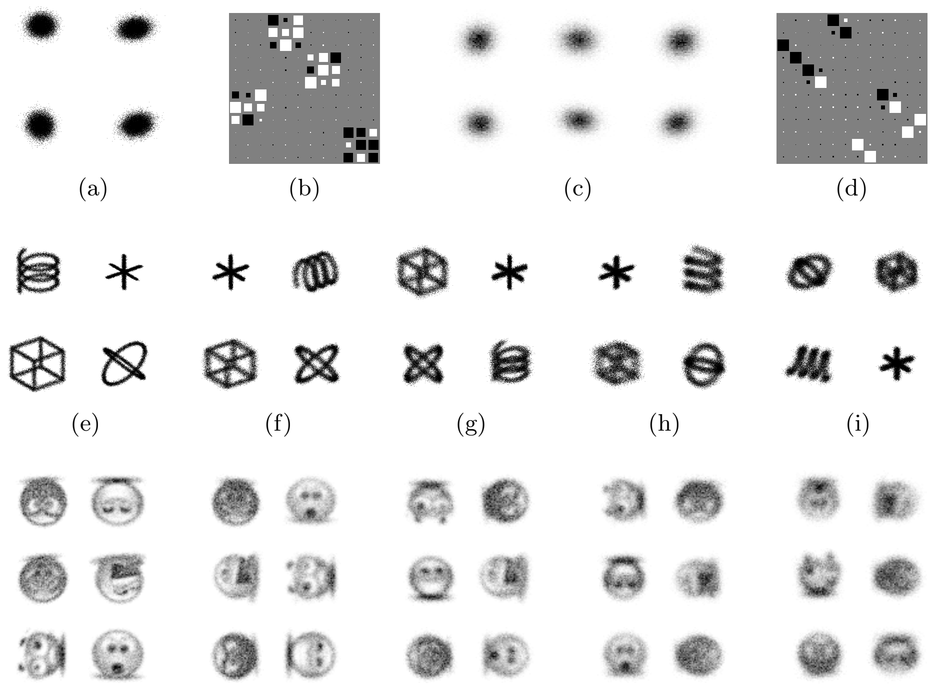


Fig. 2: Precision of the estimations and the estimated optimal AR orders. The plots are on log-log scale. (a), (b): on the *smiley* (*3D-geom*) database the Amari-index as a function of the sample number for different $\lambda \to 1$ parameter values of $\mathbf{H}[z]$ and convolution lengths, respectively. In (a): $L = 10$, in (b): $\lambda = 0.85$. (c): on the *smiley* (*3D-geom*) database the estimated AR order as a function of the sample number with $L = 10$ for different λ values. (d): the same as (b), but for the *Beatles* dataset with $\lambda = 0.9$. For graphical illustration, see Fig. 3. For numerical values, see Table 1.



(a) (b) (c) (d)

(e) (f) (g) (h) (i)

(j) (k) (l) (m) (n)

Fig. 3: Illustration of the estimations on the *3D-geom*[(a),(b),(e)-(i)] and *smiley*[(c),(d),(j)-(n)] datasets. Number of samples: $T = 20,000$. Length of the convolution: $L = 10$. In the first row: $\lambda = 0.4$. (a), (c): observed convolved signal $\mathbf{x}(t)$. (b), (d): Hinton-diagram of $\mathbf{G}$, ideally a block-permutation matrix with 2×2 and 3×3 blocks, respectively. (e)-(i), (j)-(n): estimated components $\hat{\mathbf{s}}^m$, recovered up to the ISA ambiguities from left to right for $\lambda = 0.4, 0.6, 0.7, 0.8, 0.85$. All the plotted estimations have average Amari-indices, see Fig. 2(a).

References

1. Jutten, C., Hérault, J.: Blind separation of sources: An adaptive algorithm based on neuromimetic architecture. Signal Processing **24** (1991) 1–10
2. Comon, P.: Independent component analysis, a new concept? Signal Processing **36** (1994) 287–314
3. Cardoso, J.: Multidimensional independent component analysis. In: International Conference on Acoustics, Speech, and Signal Processing (ICASSP'98). Volume 4. (1998) 1941–1944
4. Choi, S., Cichocki, A., Park, H., Lee, S.: Blind source separation and independent component analysis. Neural Information Processing - Letters and Reviews **6** (2005) 1–57
5. Szabó, Z., Póczos, B., Lőrincz, A.: Undercomplete blind subspace deconvolution. Journal of Machine Learning Research **8** (2007) 1063–1095

6. Pedersen, M.S., Larsen, J., Kjems, U., Parra, L.C.: A survey of convolutive blind source separation methods. In: Springer Handbook of Speech Processing. Springer Press (November 2007)
7. Szabó, Z., Póczos, B., Lőrincz, A.: Undercomplete blind subspace deconvolution via linear prediction. In et al., J.K., ed.: European Conference on Machine Learning (ECML'07). Volume 4701 of LNAI., Berlin Heidelberg, Springer-Verlag (2007) 740–747
8. Fuller, W.A.: Introduction to Statistical Time Series. Wiley-Interscience (1995)
9. Galbraith, J., Ullah, A., Zinde-Walsh, V.: Estimation of the vector moving average model by vector autoregression. Econometric Reviews **21**(2) (2002) 205–219
10. Theis, F.J.: Uniqueness of complex and multidimensional independent component analysis. Signal Processing **84**(5) (2004) 951–956
11. Theis, F.J.: Blind signal separation into groups of dependent signals using joint block diagonalization. In: Proceedings of International Society for Computer Aided Surgery (ISCAS'05), Kobe, Japan (2005) 5878–5881
12. Amari, S., Cichocki, A., Yang, H.H.: A new learning algorithm for blind signal separation. Advances in Neural Information Processing Systems **8** (1996) 757–763
13. Szabó, Z., Póczos, B., Lőrincz, A.: Cross-entropy optimization for independent process analysis. In: Independent Component Analysis and Blind Signal Separation (ICA'06). Volume 3889 of LNCS., Springer (2006) 909–916
14. Neumaier, A., Schneider, T.: Estimation of parameters and eigenmodes of multivariate autoregressive models. ACM Transactions on Mathematical Software **27**(1) (2001) 27–57
15. Szabó, Z., Lőrincz, A.: Real and complex independent subspace analysis by generalized variance. In: ICA Research Network International Workshop (ICARN'06). (2006) 85–88 http://arxiv.org/abs/math.ST/0610438.
16. Abed-Meraim, K., Belouchrani, A.: Algorithms for joint block diagonalization. In: European Signal Processing Conference (EUSIPCO 2004). (2004) 209–212