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. The BSSD Model

1.1 The BSSD Equations

The BSSD (Blind Subspace Deconvolution) model is

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

That is, only casual FIR filtered mixture of hidden, independent, multidimensional *components* is available for observation. Shortly,

$$\mathbf{x}=\mathbf{H}[z]\mathbf{s}, \quad (2)$$

where $\mathbf{H}[z]=\mathbf{H_0}+\mathbf{H}_1z+\ldots+\mathbf{H}_Lz^L \overset{1:1}{\longleftrightarrow}$ convolutive mixing. Task: estimate $\mathbf{s}(t)$ by using observation $\mathbf{x}(t)$ only.

1.2 uBSSD Assumptions

Notation: $\mathbf{s}=[\mathbf{s}^1;\ldots;\mathbf{s}^M]$, $\mathbf{s}^m\mathbf{s}\in\mathbb{R}^{d_m}$ are the components.

- for a given m, $\mathbf{s}^m(t)$ is i.i.d. in time t (for notational simplicity, $\forall d_m=d$),
- there is at most a single Gaussian component among $\mathbf{s}^m$s,
- independent components: $I(\mathbf{s}^1,\ldots,\mathbf{s}^M)=0$ where I is the mutual information,
- undercomplete task: $dim(\mathbf{x})>dim(\mathbf{s})$,
- $\mathbf{H}[z]\in\mathbb{R}^{dim(\mathbf{x})\times dim(\mathbf{s})}$ has left inverse ($\exists\mathbf{W}[z]\in\mathbb{R}^{dim(\mathbf{s})\times dim(\mathbf{x})}$: $\mathbf{W}[z]\mathbf{H}[z]=\mathbf{I}$) $\rightarrow\exists$ with prob. 1, if undercompleteness + matrix $[\mathbf{H}_0;\ldots;\mathbf{H}_L]$ is drawn from a continuous, non-degenerate distribution.
- without loss of generality: $E[\mathbf{s}]=\mathbf{0}\Rightarrow E[\mathbf{x}]=\mathbf{0}$.

1.3 Special Cases

- $d_m=1(\forall m)$: Blind Source Deconvolution (BSD),
- $L=0$: Independent Subspace Analysis (ISA),
- $L=0$ and $d_m=1(\forall m)$: Independent Component Analysis (ICA).

2. Decomposition Principles in the BSSD Problem Family

Former techniques:

- uBSD $\xrightarrow{\text{temporal concatenation}}$ ISA [2],
- uBSD $\xrightarrow{\text{spat.-temporal decorr. or linear prediction}}$ ICA [3, 4],
- uBSSD $\xrightarrow{\text{temporal concatenation}}$ ISA} =:TCC [1],
- ISA $\xrightarrow{\text{Separation Theorem}}$ ICA + permutation search [5, 6, 1],
- uBSSD $\xrightarrow{\text{linear prediction}}$ ISA, now (alternative for TCC):

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 $\mathbf{W}_{\rm AR}[z]\in\mathbb{R}[z]^{D_x\times D_x}$ such that $\mathbf{W}_{\rm AR}[z]\mathbf{x}=\mathbf{H}_0\mathbf{s}$.*

Thus,

- uBSSD can be solved by applying ARfit (Linear Predictive Approximation, LPA) + ISA. For pseudocode, see Table 1.
- Recovery of the hidden components $\mathbf{s}^m$:
  - only up to ambiguities of the ISA task [7],
  - whiteness assumption [$E[\mathbf{s}]=0, cov(\mathbf{s})=\mathbf{I}$] $\Rightarrow$ ambiguity up to permutation (for components with equal dimension) 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 x estimate $\hat{\mathbf{W}}_{\rm AR}[z]$
  Estimate innovation: $\tilde{\mathbf{x}}=\hat{\mathbf{W}}_{\rm AR}[z]\mathbf{x}$
  Reduce uISA to ISA and whiten: $\tilde{\mathbf{x}}'=\hat{\mathbf{W}}_{\rm PCA}\tilde{\mathbf{x}}$
  Apply ISA for $\tilde{\mathbf{x}}'$: separation matrix is $\hat{\mathbf{W}}_{\rm ISA}$
Estimation
  $\hat{\mathbf{W}}_{\rm uBSSD}[z]=\hat{\mathbf{W}}_{\rm ISA}\hat{\mathbf{W}}_{\rm PCA}\hat{\mathbf{W}}_{\rm AR}[z]$
  $\hat{\mathbf{s}}=\hat{\mathbf{W}}_{\rm uBSSD}[z]\mathbf{x}$

3. Illustrations

3.1 Databases

Four databases ($\mathbf{s}$; as in [1]) to study our LPA algorithm:

- *3D-geom*, *letters* test: (i) uniformly distributed variables on geometric shapes/letters, (ii) $d=3, M=6$ ($d=2, M=10$).
- *celebrities* test: (i) distribution according to pixel intensities on cartoons of celebrities, (ii) $d=2, M=10$.
- *Beatles* database: (i) non-i.i.d., (ii) 8 kHz sampled portions of two songs (A Hard Day's Night, Can't Buy Me Love), (iii) $d=2$, $M=2$.

For illustration (*3D-geom-letters-celebrities*), see Fig. 1.

Figure 1: *Illustration of the* 3D-geom *(left), (b)* celebrities *(center) and (c)* letters *(right) databases.*

3.2 Performance Index: The Amari-index

- Section 2 $\Rightarrow$ ideally: $\mathbf{G}:=\hat{\mathbf{W}}_{\rm ISA}\hat{\mathbf{W}}_{\rm PCA}\mathbf{H}_0\in\mathbb{R}^{dim(\mathbf{s})\times dim(\mathbf{s})}$ is a block-permutation matrix made of $d\times d$ blocks.
- To compare TCC and LPA the Amari-error adapted to ISA [8] was normalized [6] to take values in $[0,1]$ *independently from d and $dim(\mathbf{s})$*: Amari index (r).
- Definition of the Amari-index:

$$r(\mathbf{G}):=\frac{1}{2M(M-1)}\left[\sum_{i=1}^{M}\left(\frac{\sum_{j=1}^{M}g^{i,j}}{\max_j g^{i,j}}-1\right)+\sum_{j=1}^{M}\left(\frac{\sum_{i=1}^{M}g^{i,j}}{\max_i g^{i,j}}-1\right)\right], \quad (3)$$

where

  - $\mathbf{G}$ is decomposed into $d\times d$ blocks: $\mathbf{G}=\left[\mathbf{G}^{i,j}\right]_{i,j=1,\ldots,M}$, ($\mathbf{G}^{i,j}\in\mathbb{R}^{d\times d}$),
  - $g^{i,j}$: sum of the absolute values of the elements of $\mathbf{G}^{i,j}$.

3.3 Simulations

Two questions:

- comparison: TCC and LPA.
- performance of LPA as a funcion of convolution length.

Simulation parameters: $dim(\mathbf{x})=2dim(\mathbf{s})$, coordinates of $\mathbf{H}_l\sim$ standard normal, order of ARfit $\le 2(L+1)$, ISA method = JFD [9], performance index (average of 50 runs).

3.4 Comparison: TCC and LPA

Simulation domain: (i) $1,000\le T\le 100,000$, $1\le L\le 5$ (*3D-geom*, *celebrities*), (ii) $1,000\le T\le 75,000$, $1\le L\le 30$ (*letters*, *Beatles*). Results are summarized in Fig. 2.

- For the *3D-geom*, *celebrities* and *letters* tests (1^{th} row of Fig. 2):
  - power law decline of the Amari-index: $r(T)\propto T^{-c}$ $(c>0)$, $T\ge 2,000$,
  - $1.1-88$, $1.0-87$, $1.2-110$-times increase in precision.
- For the *Beatles* database (2^{nd} row of Fig. 2):
  - LPA gives reliable estimations for $T\ge 30,000$ (TCC: $T\ge 50,000$),
  - more pronounce improvement for increasing L, namely $1.50, 2.24, 4.33, 4.42, 9.03, 11.13$-times ($L=1,2,5,10,20,30$).

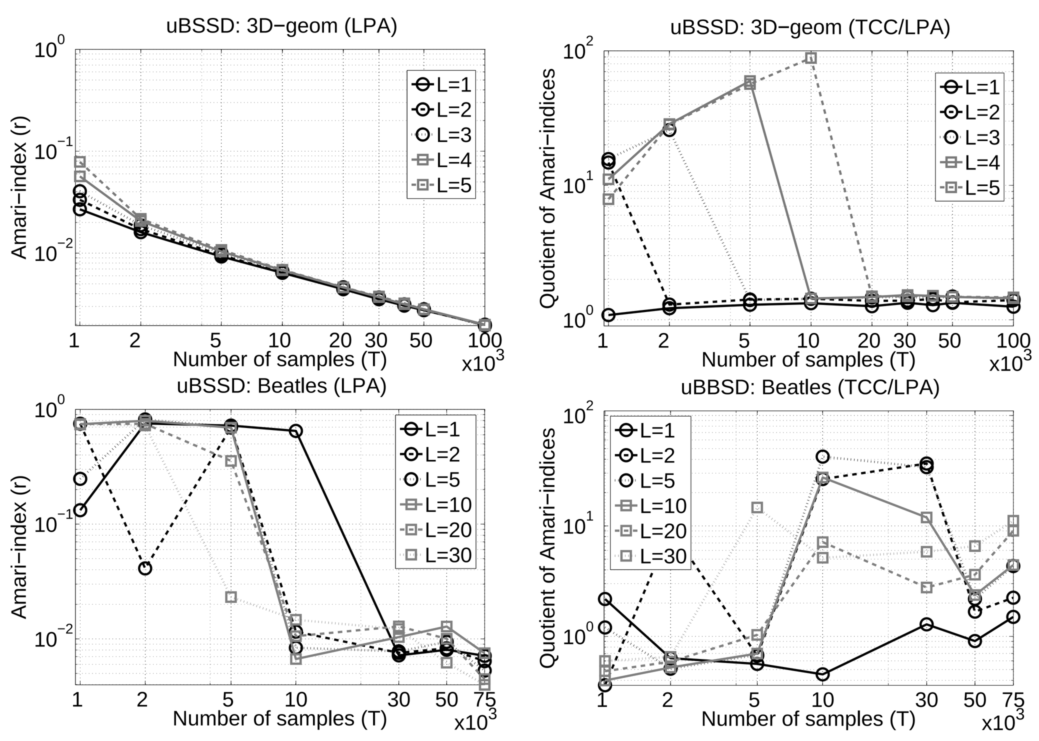

Figure 2: *Estimation error: LPA compared to TCC, database* 3D-geom *(celebrities and letters) and* Beatles*, 'log log' plots, different convolution lengths ($L+1$). 1^{st} column: Amari-index as a function of the sample number. 2^{nd} column: Quotients of the Amari-indices of the TCC and the LPA methods: for quotient value $q>1$, the LPA method is q times more precise than the TCC method.*

3.5 Performance of LPA vs. L

Acceptable estimations up to

- $L=20$ (*3D-geom*, *celebrities*; $T=20,000$) $\rightarrow$ right-most of the 1^{th} row in Fig. 3,
- $L=230$ (*letters*, *Beatles*; $T=15,000$) $\rightarrow$ right-most of the 2^{nd} row in Fig. 3.

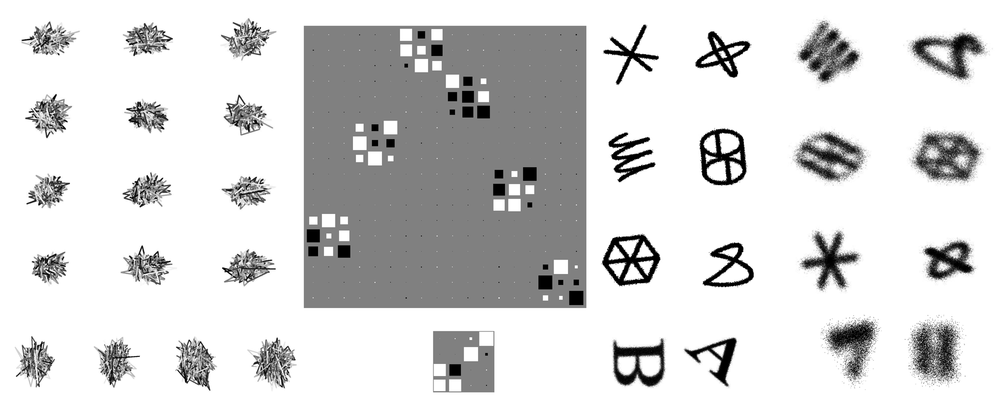

Figure 3: *Illustration of the LPA method on the uBSSD task for the* 3D-geom *(*letters*) database. First 3 columns: sample number $T=100,000$ [75,000], convolution length $L+1=6$ [31]. 1^{th} column: observed convolved signals $\mathbf{x}(t)$. 2^{nd} column: Hinton-diagram of $\mathbf{G}$, ideally block-permutation matrix with 3×3 [2×2] blocks. 3^{rd} column: estimated components ($\hat{\mathbf{s}}^m$), Amari-index: 0.2% [0.3%]. 4^{th} column: estimation of hidden components ($\hat{\mathbf{s}}^m$) for sample number $T=20,000$ [$15,000$] and convolution parameter $L=20$ [230].*

References

[1] Szabó, Z., Póczos, B., Lőrincz, A.: Undercomplete blind subspace deconvolution. Journal of Machine Learning Research **8** (2007) 1063–1095

[2] Févotte, C., Doncarli, C.: A unified presentation of blind source separation for convolutive mixtures using block-diagonalization. In: ICA '03. (2003) 349–354

[3] Choi, S., Cichocki, A.: Blind signal deconvolution by spatio-temporal decorrelation and demixing. Neural Networks for Signal Processing **7** (1997) 426–435

[4] Gorokhov, A., Loubaton, P.: Blind identification of MIMO-FIR systems: A generalized linear prediction approach. Signal Processing **73** (1999) 105–124

[5] Cardoso, J.: Multidimensional independent component analysis. In: ICASSP '98. (Volume 4.) 1941–1944

[6] Szabó, Z., Póczos, B., Lőrincz, A.: Cross-entropy optimization for independent process analysis. In: ICA '06. Volume 3889 of LNCS., Springer (2006) 909–916

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

[8] Theis, F.J.: Blind signal separation into groups of dependent signals using joint block diagonalization. In: ISCAS '05. (2005) 5878–5881

[9] Szabó, Z., Lőrincz, A.: Real and complex independent subspace analysis by generalized variance. In: ICARN '06. (2006) 85–88