# Blind source separation based on minimum description length

Elena G. Revunova

*Dept. of Neural Information Processing Technologies, International Research and Training Centre of Information Technologies and Systems, National Academy of Sciences of Ukraine, prospect Academica Glushkova 40, Kiev, Ukraine*

helab@i.com.ua

**Abstract.** *Separation of signal mixtures using blind source separation (BSS) approach is considered. Objective function for BSS based on minimum description length principle is developed. Testing has shown a better robustness to additive noise than that of PCA and FastICA.*

## Keywords

Blind source separation, minimum description length, sparse approximation.

## 1 Method BSSMDL

In *BSS* problems data $D_L$ are represented by the set of $L$ *mixture vectors* $\mathbf{y}_i \in \Re^N$: $D_L=\{\mathbf{y}_i\}_{i=1,L}$. The mixture vector $\mathbf{y}$ is connected with the source vector $\mathbf{x}$ by a linear model $\mathbf{y}=\mathbf{Ax}$, where the full rank mixture matrix $\mathbf{A}\in\Re^{N\times N}$ is unknown. The problem is to obtain $L$ unobservable vectors $\mathbf{x}$ (matrix $\mathbf{X}$) using $L$ observable vectors $\mathbf{y}$ (matrix $\mathbf{Y}\in\Re^{L\times N}$) by estimating *un-mixture* matrix $\mathbf{W}=(\mathbf{w}_1,\ldots,\mathbf{w}_N)^T\in\Re^{N\times N}$ and then calculating $\mathbf{X}=\mathbf{YW}$. Well-known approaches to solution of *BSS* problem use a priori information about the sources: the sources are supposed to be statistically independent (independent component analysis *ICA* [1, 2]), sparse (sparse component analysis *SCA* [3]), uncorrelated (principal component analysis *PCA* [2]). The methods of *BSS* are sensitive to additive noise and require large data sets so that their usage in non-stationary applications is rather problematic [4].

To increase performance of *BSS* methods for noisy mixtures we propose [5] to estimate independence of sources using objective function based on algorithmic mutual information [6]:

$$K(x_1{:}x_2) = K(x_1) - K(x_1|x_2), \quad (1)$$

where $K(x_1)$ is algorithmic complexity of $x_1$, $K(x_1|x_2)$ is complexity of $x_1$ conditioned upon $x_2$.

We propose to calculate $K(\cdot)$ using minimum description length (*MDL*) [6, 7] based on a universal model *nMDL* proposed by *Bin Yu* [8]:

$$nMDL = (L/2)\log(RSS/(L-k)) + 0.5\,k\log(F) + \log(L-k) - 3/2\log(k),$$
$$F = (\mathbf{x}^{\mathrm{T}}\mathbf{x} - RSS)/(k\,RSS/(L-k)),\ RSS = \|\,\mathbf{x} - \mathbf{\Phi\theta}\,\|^2. \quad (2)$$

Here $\mathbf{x}$ is described by a linear model $M$ as $\mathbf{x}=\sum_{i=1,k}\theta_i\boldsymbol{\varphi}_i$, where $\boldsymbol{\theta}\in\Re^k$ are model parameters, $\boldsymbol{\varphi}\in\Re^L$ are basis functions selected in process of constructing $M$ and used to approximate $\mathbf{x}\in\Re^L$, $L$ is the number of data samples used to construct the model, $k$ is the number of parameters, $\varphi_{ij}$ compose matrix $\mathbf{\Phi}\in\Re^{L\times k}$.

The optimization problem of finding $\mathbf{w}$ to separate 2-component mixture (denoting mixtures $(\mathbf{y}_1,\mathbf{y}_2)$ as $\mathbf{Y}\in\Re^{L\times 2}$) is as follows

$$\mathbf{w} = \arg\min_{\mathbf{w}}(nMDL(\mathbf{Yw}_1) - nMDL(\mathbf{Yw}_1|M_2)). \quad (3)$$

At each step of optimization procedure we construct linear model $M_1$ to approximate current estimate of $\mathbf{x}_1^*=\mathbf{Yw}_1^*$ and linear model $M_2$ to approximate current estimate of $\mathbf{x}_2^*=\mathbf{Yw}_2^*$, where $\mathbf{w}_1^*$ and $\mathbf{w}_2^*$ are estimates of $\mathbf{w}_1$ and $\mathbf{w}_2$ obtained at the current step of optimization. To construct $M_1$ and $M_2$ we use sparse approximation methods with Gribonval test [9] that allows for noise tolerance. Then we calculate $nMDL(\mathbf{Yw}_1)$ using (2). To calculate $nMDL(\mathbf{Yw}_1|M_2)$ we propose to approximate $\mathbf{Yw}_1$ again using only the basis functions $\mathbf{\Phi}_2$ obtained in $M_2$ and to calculate (2) using $RSS=\|\mathbf{Yw}_1-\mathbf{\Phi}_2\mathbf{\theta}_2\|^2$.

Thus the method *BSSMDL* of blind source separation using optimization problem (3) includes the following stages at each step of optimization procedure.

1. Form basis function matrix $\mathbf{\Psi}\in\Re^{L\times N}$, $L<<N$. Column values $\mathbf{\Psi}(\cdot,i)$ are values of radial basis functions $\mathbf{\varphi}_i(z)$ $i=1,\ldots,N$, calculated for $z=1,\ldots,L$ and normalized as $\|\mathbf{\varphi}_i\|_2=1$. Initialize $\mathbf{w}_1=(1, 1)$.

2. Using $\mathbf{\Psi}$, select linear model for $\mathbf{Yw}_1$ (i.e. current estimate of $\mathbf{x}_1$) and obtain $\mathbf{\Phi}_1 = \{\mathbf{\varphi}_1, \mathbf{\varphi}_2,\ldots, \mathbf{\varphi}_{k+1}\}$ and $\mathbf{\theta}_1$.

3. Select linear model for $\mathbf{Yw}_2$ (current estimate of $\mathbf{x}_2$) and obtain $\mathbf{\Phi}_2 = \{\mathbf{\varphi}'_1, \mathbf{\varphi}'_2,\ldots, \mathbf{\varphi}'_{k'+1}\}$.

4. Calculate $RSS=\|\mathbf{Yw}_1 - \mathbf{\Phi}_1\mathbf{\theta}_1\|^2$ and $nMDL(\mathbf{Yw}_1)$ using (2).

5. Using $\mathbf{\Phi}_2 = \{\mathbf{\varphi}'_1, \mathbf{\varphi}'_2,\ldots, \mathbf{\varphi}'_{k'+1}\}$ to approximate $\mathbf{Yw}_1$, obtain $\mathbf{\theta}_2$ as $\mathbf{\theta}_2=(\mathbf{\Phi}_2^T\mathbf{\Phi}_2)^{-1}\mathbf{\Phi}_2^T(\mathbf{Yw}_1)$.

6. Calculate $RSS=\|\mathbf{Yw}_1 - \mathbf{\Phi}_2\mathbf{\theta}_2\|^2$ and $nMDL(\mathbf{Yw}_1|M_2)$ using (2).

7. Calculate objective function value as: $nMDL(\mathbf{Yw}_1)–nMDL(\mathbf{Yw}_1|M_2)$.

8. Calculate the new value $\mathbf{w}_1$ using *Nelder-Mead* optimization procedure.

9. Check if the objective function *OF* minimum is reached using the standard tools of *Nelder-Mead* optimization procedure. If the minimum is not reached, go to step 2, else Finish.

*BSSMDL* method is schematically represented in figure 1.

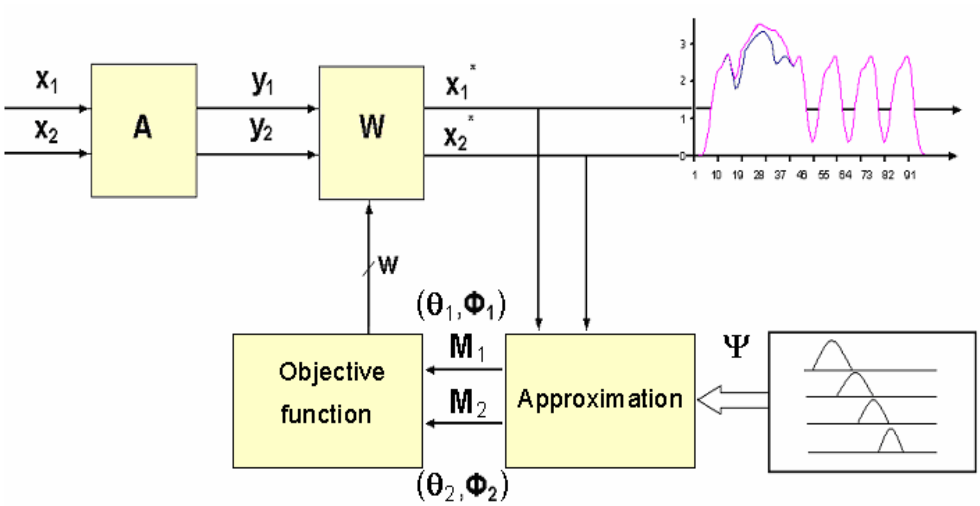


**Fig. 1.** *BSSMDL* method

## 2 Experimental investigation

In the first experiment the mixtures were formed by multiplication of sources by mixing matrix $\mathbf{A}$ = [1 1.5] [3 2.2]. Signals of sources were (a) pulse of 15 unit width and amplitude 1 and (b) 100 realizations of a stochastic random process with normal distribution of zero mean. The standard deviation *std* of noise signal (b) varied from 5 up to $10^5$. Gaussian *RBF*s $\mathbf{\varphi}_i(z)$ $i=1,\ldots,L$ centered in $z=1,\ldots,L$ and each having dispersion of 0.2 have been used as $\mathbf{\Psi}$.

We have separated the (useful) pulse signal (a) from the mixture using *PCA*, *ICA* and *BSSMDL* and calculated the corresponding *signal to noise ratio SNR*=10 $\log_{10}(\|\mathbf{s}\|^2/\|\mathbf{s}-\mathbf{s}^*\|^2)$, where $\mathbf{s}\in\Re^L$ is the true signal vector, $\mathbf{s}^*\in\Re^L$ is the signal vector obtained at the output of mixture separation algorithm.

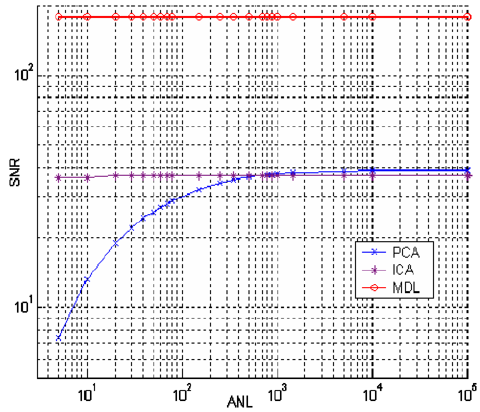


**Fig. 2.** Dependence *SNR=f(ANL)*.

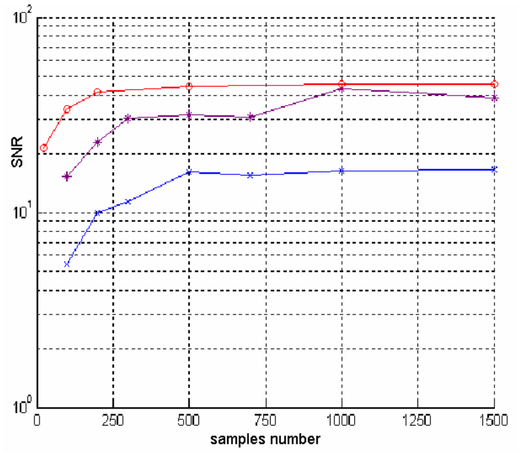


**Fig. 3.** *SNR* vs the number of samples.

In figure 2 dependence *SNR=f(ANL)* is given, where *ANL* is the *std* of gaussian noise (b). Additional noise was not present in the mixtures. For *PCA* the *SNR=f(ANL)* first grows with the *std* of (b) and then becomes constant. We connect this with the nature of *PCA* that tries to discover the direction of maximal dispersion. When amplitudes of sources are comparable, the choice of the maximal dispersion direction is sensitive to amplitude ratio of source signals. When the amplitude of one of sources is much more than the amplitude of the other, the maximal amplitude defines the maximal dispersion direction.

For *ICA* and *BSSMDL* the resulting *SNR* does not depend on the *ANL* value. We consider this as the consequence of their assumption of sources' independence. Since increasing amplitude of one source appears in both mixtures, the mutual information between mixtures does not change, and the *SNR* does not depend on the amplitude of one of the sources.

In the second experiment we have investigated dependence of the *SNR* on the number of samples. Signals of sources were (a) a "saw-tooth" signal with amplitude 1 and (b) realization of normally distributed random variable of zero mean and dispersion 1. The mixtures were formed by multiplication of sources by mixing matrix **A** = [1 1.5] [3 2.2]. We changed the number of samples from 20 up to 1500.

In figure 3 dependence *SNR* on the number of samples is given. The *SNR* for *BSSMDL* exceeds the *SNR* for *ICA* for the number of samples less than $10^3$. The *SNR* for *BSSMDL* and *ICA* exceed the *SNR* for *PCA* in all investigated range of the samples' number. Therefore *BSSMDL* is effective for source separation for small number of data samples.

Blind source separation is useful for many application areas – radio communications, navigation, radar, etc. [10]. Example is jamming cancellation in the homogeneous antenna system. The signal in all antenna channels is a mix of three components – a useful signal, jamming (active noise), and additive inherent noise of the channel. We investigated dependence of *SNR* on inherent noise level.

The mixture was composed of (a) an impulse of length 10 and amplitude 1, (b) realization of a random process with uniform distribution, and (c) inherent noise of varied *std*. The models for *BSSMDL* were constructed using *RBF*s as basis functions.

We observed (figure 4) that the *SNR* for *BSSMDL* exceeds the *SNR* for *FastICA* at small values of inherent noise and exceeds the *SNR* for *PCA* in the whole range of inherent noise values. For example, as inherent noise increased from 0.005 to 2.56 the *SNR* for *BSSMDL* decreased from 70.84 dB to 16.8 dB, whereas the *SNR* for *FastICA* decreased from 59.27 dB to 16.8 dB, and the *SNR* for *PCA* decreased from 12.6 dB to 11.1 dB.

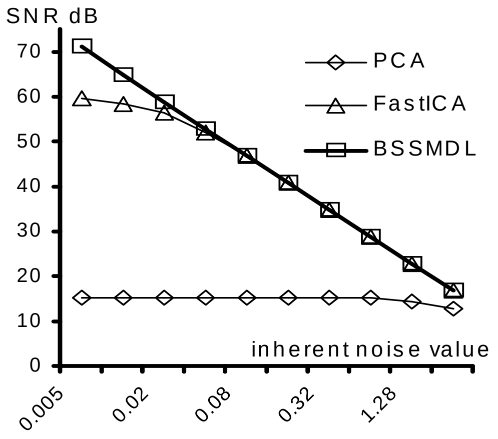


**Fig. 4.** *SNR* vs inherent noise value.

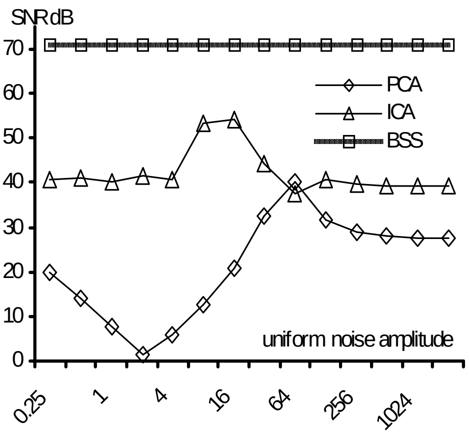


**Fig. 5.** *SNR* vs uniform noise amplitude.

Another set of experiments has been conducted to investigate *SNR* vs the amplitude of uniform noise (b). We observed (figure 5) that the *SNR* for *BSSMDL* is constant and higher than the *SNR* for *FastICA* and *PCA*. For example, at the inherent noise level of 0.01 and increasing level of (b) source from 0.25 to 2048 the *SNR* for *BSSMDL* was 70dB, whereas the *SNR* for *FastICA* was less than 54 dB and equal to 40 dB for the most part of amplitude range of source (b). The *SNR* of *PCA* is non-uniform and worse than the *SNR* of *FastICA*.

# 3 Conclusion

In conclusion, the developed objective function *BSSMDL* using minimum description length based on a universal model *nMDL* demonstrated robust performance in *BSS* tasks in the presence of noise as well as a higher signal to noise ratio compared to *FastICA* (at small level of noise) and *PCA* (in the whole range of tested noise values).

# References

[1] Comon P.: Independent component analysis: a new concept. *Signal Processing*, 36:287-314, 1994.
[2] Hyvarinen A., Oja E.: Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5):411–430, 2000.
[3] Zibulevsky M., Pearlmutter B.A.: Blind source separation by sparse decomposition in a signal dictionary. *Neural Computations*, 13(4):863-882, 2001.
[4] Cichocki A., Amari S.: Adaptive Blind Signal and Image Processing. *New York*: *Wiley*, 586 pp, 2002.
[5] Revunova E.G.: Two approaches to separation of signal mixtures based on linear models. *System Technologies* 6(41):142-148, 2005 (in Russian).
[6] Rissanen J.: Lectures on statistical modeling theory. http://www.cs.tut.fi/~rissanen/, 2002.
[7] Li M., Vitanyi P.M.B.: An Introduction to Kolmogorov Complexity and Its Applications, 2nd Edition. *New York*: *Springer-Verlag*, 637 pp, 1997.
[8] Hansen M., Yu B.: Model selection and minimum description length principle. *J. Amer. Statist. Assoc*., 96: 746-774, 2001.
[9] Gribonval R., Figueras i Ventura R. M., Vandergheynst P.: A simple test to check the optimality of sparse signal approximations. *Tech. Rep., IRISA*: PI-1661, 2004.
[10] Shirman J.D.: Radio-electronic systems: bases of construction and the theory. *Moscow*: *MAKVIS*, 826 pp, 1998 (in Russian).