Stochastic and Physical Monitoring Systems
SPMS 2018
June 18 - June 22, 2018
Dobřichovice, Czech Republic

Heuristics in blind source separation

Václav Kautský[1], Jakub Štěch[2]

[1,2]Department of Mathematics, FNSPE Czech Technical University in Prague,
[1]Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec, Czech Republic,
[2]Institute of Information Theory and Automation, 182 08 Prague 8, Czech Republic

Email: kautsvac@fjfi.cvut.cz, stechjak@fjfi.cvut.cz

Abstract. This paper deals with application of heuristic algorithms (DEBR, MCRS) in blind source separation (BSS). BSS methods focus on a separation of the (source) signal from a linear mixture. The idea of using heuristic algorithms is introduced on the independent component extraction (ICE) model. The motivation for considering heuristics is to obtain an initial guess needed by many ICE algorithms. Moreover, the comparison of this initialization, and other algorithms accuracy is performed.

Key words: Blind Source Separation; DEBR; Independent Component Extraction, MCRS.

1 Introduction

Independent component analysis (ICA) is a blind source separation (BSS) problem. The aim of this approach is to separate all components of the instantaneous linear mixing model

$$\mathbf{x} = \mathbf{A}\mathbf{u}. \tag{1}$$

Here, $\mathbf{x}$ is a $d \times 1$ vector of d mixed signals, $\mathbf{A}$ is a $d \times d$ non-singular mixing matrix, and $\mathbf{u}$ is a $d \times 1$ vector of the original signals that are assumed to be *mutually independent*. The jth signal u_j (the jth element of $\mathbf{u}$) is modeled as a random variable with the probability density function (pdf) $p_j(\cdot)$. The separation of the components is to estimate $\mathbf{A}^{-1}$ from $\mathbf{x}$ through finding a square de-mixing matrix $\mathbf{W}$ such that $\mathbf{y} = \mathbf{W}\mathbf{x}$ are as independent as possible. In this paper, we will assume real-valued signals and parameters.

Many algorithms to solve the ICA problem have been developed; see, e.g., [1]. It is known that if, at most, one original signal has Gaussian pdf while the other signals are non-Gaussian, then $\mathbf{A}^{-1}$ can be identified up to the order and scales of its rows [6]. It means that the de-mixing matrix $\mathbf{W}$ can be estimated as such that $\mathbf{G} = \mathbf{W}\mathbf{A} \approx \mathbf{P}\boldsymbol{\Lambda}$, where $\mathbf{P}$ and $\boldsymbol{\Lambda}$ is, respectively, a permutation and a diagonal matrix. The elements of $\mathbf{G}$ determine the accuracy of the separation. Its ijth element, G_{ij}, determines the presence of u_j in the ith separated signal y_i. There is a lower bound on the variance of G_{ij}, the Cramér-Rao Lower Bound (CRLB). The CRLB provides an algorithm-independent bound for the estimation accuracy (for unbiased estimators). Using the CRLB theory, it is known, for the non-Gaussian ICA, that

$$\mathrm{E}[G_{ij}^2] \geq \frac{1}{N} \frac{\kappa_j}{\kappa_i \kappa_j - 1}, \qquad i \neq j, \tag{2}$$

where E[·] stands for the expectation operator, N is the number of samples of $\mathbf{x}$ (assuming identically and independently distributed samples), and $\kappa_i = \mathrm{E}[\psi_i^2]$ where $\psi_i(x) = -\partial/\partial x \log p_i(x)$, which is called the score function of p_i where p_i is pdf of ith signal. For normalized variables with unit variance it holds that $\kappa_i \geq 1$ where $\kappa_i = 1$ if and only if the ith pdf is Gaussian; see, e.g., [7].

Recently, a novel approach called Independent Component Extraction (ICE) have been introduced, see [5]. In ICE, the goal is to separate only one independent signal from $\mathbf{x}$ using a priori knowledge such as an initial guess (to determine which signal should be extracted). Without any loss on generality, let the signal of interest be $s = u_1$. The motivation for ICE is that ICE algorithms could solve the simpler problem (to extract only one signal) faster that ICA methods, since in ICE the minimum number of parameters needed for extraction of the target signal is estimated. In ICE, the mixing model (1) is re-parameterized for the extraction of s in the way that the rest of the mixture is not object of any particular decomposition, as compared to ICA. Following the ideas of ICE, the signal of interest s is assumed to be non-Gaussian while the rest of the mixture is modeled as Gaussian. The latter is motivated by the fact that the other signals are never separated from each other (up to very special cases), so their joint distribution is close to Gaussian even if the pdfs of $u_2, \ldots, u_d$ are non-Gaussian.

In [5], an ICE algorithm, called OGICE, was introduced. In this paper, we study some performances of the OGICE algorithm and compare and combine it with selected heuristics algorithms: Differential Evolution (DEBR) and Modifed Controlled Random Search (MCRS). Accuracies of all methods are than compared to the CRLB.

The ICE mixing model parametrization and the statistical model of signals are described in Section 2. Section 3 is devoted to the introduction of DEBR and MCRS algorithms, which are used in Section 4 when performing numerical simulations. Conclusions are drawn in Section 5.

The following notation will be used throughout the article. Plain letters denote scalars, bold letters denote vectors, and bold capital letters denote matrices. The Matlab convention for matrix/vector concatenation and indexing will be used, e.g., $[1;\, \mathbf{g}] = [1,\, \mathbf{g}^T]^T$, and $(\mathbf{A})_{j,:}$ is the jth row of $\mathbf{A}$.

2 Problem Statement

2.1 Mixing Model

ICE is based on a re-parameterization of (1) mixing model. Let the mixing matrix $\mathbf{A}$ be partitioned as $\mathbf{A} = [\mathbf{a},\, \mathbf{A}_2]$, and let $\mathbf{x}$ be written as $\mathbf{x} = \mathbf{A}\mathbf{u} = \mathbf{a}s + \mathbf{y}$, where $\mathbf{y} = \mathbf{A}_2\mathbf{u}_2$ and $\mathbf{u}_2 = [u_2, \ldots, u_d]^T$. Since the scales of s and of $\mathbf{a}$ are ambiguous (s and $\mathbf{a}$ can be substituted, respectively, by αs and $\alpha^{-1}\mathbf{a}$ with any $\alpha \neq 0$), we can fix the first element of $\mathbf{A}_{\text{ICE}}$ equal to one, i.e. $(\mathbf{A}_{\text{ICE}})_{1,1} = 1$. This also corresponds to the image of that source on the first sensor [3]. Hence, the new mixing matrix is $\mathbf{A}_{\text{ICE}} = [\mathbf{a}, \mathbf{Q}]$. Now, the ICE mixing model can be written as

$$\mathbf{x} = \mathbf{A}_{\text{ICE}}\mathbf{v}, \tag{3}$$

where $\mathbf{v} = [s; \mathbf{z}]$, and $\mathbf{z} = \mathbf{B}\mathbf{x}$. The de-mixing, $\mathbf{W}_{\text{ICE}} = \mathbf{A}_{\text{ICE}}^{-1}$ can be partitioned as $\mathbf{W}_{\text{ICE}} = [\mathbf{w}^H; \mathbf{B}]$, where $\mathbf{w}^H\mathbf{x} = s$ and $\mathbf{B}\mathbf{x} = \mathbf{z}$.

Let $\mathbf{a}$ be partitioned as $\mathbf{a} = [1; \mathbf{g}]$. It is required to $\mathbf{B}$ be orthogonal to $\mathbf{a}$, which ensures that the signals separated by the lower part of $\mathbf{W}_{\text{ICE}}$, that is $\mathbf{B}\mathbf{x}$, do not contain any contribution by s. A straightforward selection is $\mathbf{B} = [\mathbf{g} \quad -\gamma\mathbf{I}_{d-1}]$ where $\mathbf{I}_d$ denotes the $d \times d$ identity matrix. The free variables of $\mathbf{W}_{\text{ICE}}$ are therefore represented by the elements of $\mathbf{a}$ and $\mathbf{w}$; let $\mathbf{w} = [\beta; \mathbf{h}]$. Hence

$$\mathbf{W}_{\text{ICE}} = \begin{pmatrix} \mathbf{w}^T \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} \beta & \mathbf{h}^T \\ \mathbf{g} & -\gamma\mathbf{I}_{d-1} \end{pmatrix}. \tag{4}$$

The next condition is that $\mathbf{W}_{\text{ICE}}$ should be the inverse matrix of $\mathbf{A}_{\text{ICE}}$, which guarantees that $s = \mathbf{w}^T\mathbf{x}$. This way $\mathbf{Q}$ and $\mathbf{z}$ can be determined. The reader can verify that the choice

$$\mathbf{A}_{\text{ICE}} = [\mathbf{a} \quad \mathbf{Q}] = \begin{pmatrix} 1 & \mathbf{h}^T \\ \mathbf{g} & \mathbf{g}\mathbf{h}^T - \mathbf{I}_{d-1} \end{pmatrix}, \tag{5}$$

where β is constrained to satisfy $\beta = 1 - \mathbf{h}^T\mathbf{g}$, guarantees that $\mathbf{W}_{\text{ICE}}\mathbf{A}_{\text{ICE}} = \mathbf{I}_d$.

s

By adopting the idea of ICA, that is, taking the assumption that s and $\mathbf{z}$ are *independent*, ICE can be formulated as follows: *Find vectors* $\mathbf{g}$ *and* $\mathbf{h}$ *such that* $\mathbf{w}^T\mathbf{x}$ *and* $\mathbf{B}\mathbf{x}$*, where* $\mathbf{w} = [1 - \mathbf{h}^T\mathbf{g}; \mathbf{h}]$ *and* $\mathbf{B} = [\mathbf{g}, \quad -\mathbf{I}_{d-1}]$*, are independent (or as independent as possible).*

2.2 Statistical Model

The most popular ICA model (1) is non-Gaussian, i.e. all (but one) signals are non-Gaussian i.i.d. sequences. In ICE, there are assumed only two variables: the target signal s and $\mathbf{z}$, the background, which is a vector variable having unspecified structure (it is a mixture of $u_2, \ldots, u_d$). As in [5], we will assume that (1) s has a non-Gaussian pdf denoted as $p(s)$, while (2) $\mathbf{z}$ has multivariate Gaussian pdf with covariance $\mathbf{C}_\mathbf{z}$. The latter assumption can be justified by the fact that, $\mathbf{z}$ is a mixture of sources in $\mathbf{u}_2$. Even if $u_2, \ldots, u_d$ are non-Gaussian, their mixture tends to have distribution close to Gaussian for large number of signals (due to the Central Limit Theorem [4]).The ICE model coincides with ICA model (1) when $u_2, \ldots, u_d$ are Gaussian. Hence, from (3), the pdf of $\mathbf{x}$ is

$$p_\mathbf{x}(\mathbf{x}) = p_s(\mathbf{w}^T\mathbf{x})p_\mathbf{z}(\mathbf{B}\mathbf{x})|\det(\mathbf{W}_{\text{ICE}})|, \tag{6}$$

where $\mathbf{W}_{\text{ICE}}$, $\mathbf{w}$, and $\mathbf{B}$ depend on $\mathbf{g}$ and $\mathbf{h}$ as described by (4), and $p_\mathbf{z}$ corresponds to $\mathcal{N}(\mathbf{0}, \mathbf{C}_\mathbf{z})$. A straightforward calculus, not shown here to save space, can show that $|\det(\mathbf{W}_{\text{ICE}})| = 1$. Hence, the log-likelihood function, for one signal sample, is equal to

$$\mathcal{L}(\mathbf{g}, \mathbf{h}|\mathbf{x}) = \log p_s(\mathbf{w}^T\mathbf{x}) - \frac{1}{2}\mathbf{x}^T\mathbf{B}^T\mathbf{C}_\mathbf{z}^{-1}\mathbf{B}\mathbf{x} - \frac{1}{2}\log(|\mathbf{C}_\mathbf{z}|) - (d-1)\log\sqrt{2\pi}, \tag{7}$$

where $|\mathbf{C_z}|$ denotes the determinant of $\mathbf{C_z}$.

The most known measure of separation algorithm accuracy is called Interference-to-Signal Ration (ISR). For ICE the ISR is defined as

$$\texttt{ISR} = \frac{\mathrm{E}[(\widehat{\mathbf{w}}^T \mathbf{y})^2]}{\mathrm{E}[(\widehat{\mathbf{w}}^H \mathbf{a}s)^2]}. \tag{8}$$

As shown in [2], the Cramér-Rao Lower Bound for the mean ISR, called Cramér-Rao Induced Bound (CRIB), was derived

$$\mathrm{E}\left[\texttt{ISR}\right] \geq \frac{1}{N}\frac{d-1}{\kappa\sigma_s^2 - 1}, \tag{9}$$

where $\kappa = \mathrm{E}\left[\left(\frac{\partial \log p(s)}{\partial s}\right)^2\right]$. Numerical simulations compare the results with the bound (9) to verify algorithms efficiency.

3 Heuristic Algorithms

We will deal with the following optimization problem: For a given objective function f the point x^* is to be found such that:

$$f(x^*) = \min_{x \in D} f(x). \tag{10}$$

The point x^* is defined as global minimum point and D is the search space. The space D is closed compact set. The objective function we want to optimize is the log-likelihood (7).

3.1 Modifed Controlled Random Search (MCRS)

MCRS, introduced in [9], is basically random search controlled by procedure called Reflection. This technique is used for generating the next trial point. First, let us take N random points in D from a population P. The new trial point x is generated from a simplex S (one point from each dimension) and perform the Reflection of the point z by the formula

$$y = g - Y(z - g), \tag{11}$$

where g is the centroid of the d poles of the simplex, Y is a random multiplication factor and z is one (random) pole of the simplex S. Denote x_{max} the point with the largest function value in the population. The procedure runs as shown in Alg. 1.

There is plenty of choices of setting the multiplication factor Y. We have considered:

- Y = (constant),
- Y is a random variable with uniform distribution on the interval $(0; t)$, where t is a positive real number.

The value of the parameter t is discussed in [9].

Algorithm 1 MCRS Algorithm

```
procedure MCRS
    P:= population of N uniformly distributed points in $D$
    while stopping condition is False do
        y:=Reflection(P; y) until $y \in D$
        if $f(y) < f(x_{max})$ then return $x_{max} := y$;
```

3.2 Differential Evolution (DEBR)

As already mentioned, we deal with the optimization problem (10). Let us consider two population P and Q of the same size N. The basic differential evolution (DE) [8] is a method which iteratively tries to improve the condidate solution instead of guessing a new one as in random search. A new trial point y is composed of the current point x_i of old population and the point u obtained by using mutation. If $f(y) < f(x_i)$ the point y is accepted and inserted into the new population Q instead of x_i. After completion of the new population Q the old population P is replaced by Q and the search continues until stopping condition is fulfilled. The pseudocode is shown in Alg. 2

Algorithm 2 DEBR Algorithm

```
procedure DEBR
    P:= population of N random points in $D$
    while stopping condition is False do
        for $i = 1 : N$ do
            compute a mutant vector $u$;
            create a trial point $y$ by the crossover of $u$ and $x_i$;
            if $f(y) < f(x_{max})$ then insert $y$ into Q;
            else: insert $x_i$ into Q;
        $P := Q$
```

There are many variants of DE, however main differences are only in the methodology of generating a new trial point. We focused on two main cases:

1. DE

$$u = r_1 - F(r_2 - r_3), \tag{12}$$

2. DEBEST

$$u = x_{min} + F(r_1 + r_2 - r_3 - r_4), \tag{13}$$

where F is a multiplication factor. The combination of these two techniques is called DEBR [8].

The elements y_j, $j = 1, 2, \ldots, d$, of a trial point y are built up by the crossover of its parents x_i and u using the following rule

$$y_j = \begin{cases} u_j, & \text{if } U_j \le C \text{ or } j = l \\ x_{ij}, & \text{if } U_j > C \text{ and } j \neq l \end{cases}$$

where l is a randomly chosen integer from $\{1, 2, \ldots, d\}$, and $U_1, U_2, \ldots, U_d$ are independent random variables uniformly distributed in $[0; 1)$, and $C \in [0; 1]$ is an input parameter influencing the number of elements to be exchanged by crossover.

We examined that for our special case the efficiency of the DE search for the global minimum is not really sensitive to the setting of values F and C. The recommended values are F = 0.6 and C = 0.5 (see [8]).

3.3 Comparison

Running the simulations with different setting of parameters in MCRS and DEBR shows, that both algorithms reach nearly equivalent performance. Moreover, both, MCRS and DEBR, are not very sensitive to the changes of parameters. As you can see in Tab. 1, the performance of MCRS does not depend on the multiplication factor Y.

Y = rand(0,x)	Cost function	# steps
2	0.7116	3450
3	0.7116	3526
4	0.7116	3681
5	0.7116	3381

Table 1: MCRS results.

The Tab. 2 shows, that the DEBR algorithm requires less steps when the probability of crossover is set to one. The other parameters F and C does not affect the results. Although Tab. 2 and Tab. 1 show, that MCRS needs more steps to converge, the running time of both algorithms is almost equal, since DEBR is more computationally expensive.

F	P_{cross}	Cost function	# steps
0.6	0.2	0.7041	537
0.6	0.6	0.7040	532
0.6	0.8	0.7040	526
0.6	1	0.7040	470

Table 2: DEBR results.

4 Simulations

The CRIB (9) is the lower bound for ISR, i.e for the separation accuracy. Only efficient methods reach the CRIB in ISR. In [5] was introduced the OGICE algorithm and ISR of this method was compared to other methods. We perform some other examples of ISR vs. CRIB comparisons. The aim is to compare the accuracy of OGICE with that of MCRS (DEBR has equivalent performance). We generate $d = 5$ signals as follows: one non-Gaussian signal and four Gaussian signals. Used signals are generated with the same length N, zero mean and unit variance.

Figure 1 shows a comparison of CRIB and ISRs of three different settings depending on the length of the signal N for separation of a signal s, where s is drawn according to the Laplacian distribution with zero mean and the unit variance. The dashed line with circle points ('no-ini' in the legend) shows the achieved ISR by OGICE when randomly initialized, dashed-and-dotted line with circles ('MCRS' in the legend) corresponds to the ISR of the MCRS algorithm. The dashed-and-dotted line with crosses is then the ISR of OGICE when initialized by the results of MCRS. As can be seen from Figure 1, the performance of OGICE is significantly dependent on the initialization. On the other hand, MCRS reaches even lower separation accuracy than randomly initialized OGICE. When OGICE is initialized by MCRS, the accuracy is more stable and the method almost reaches the lower bound.

Figure 2 shows a comparison of convergence rate of used methods. As can be seen, when OGICE is initialized randomly, only about 60% of runs converges. Despite the accuracy of MCRS is low, this algorithm converges to the desired signal in all runs, which allows us to consider results from MCRS to be suitable candidate for the initial guess for other algorithms. When OGICE is initialized this way, it converges also in all runs.

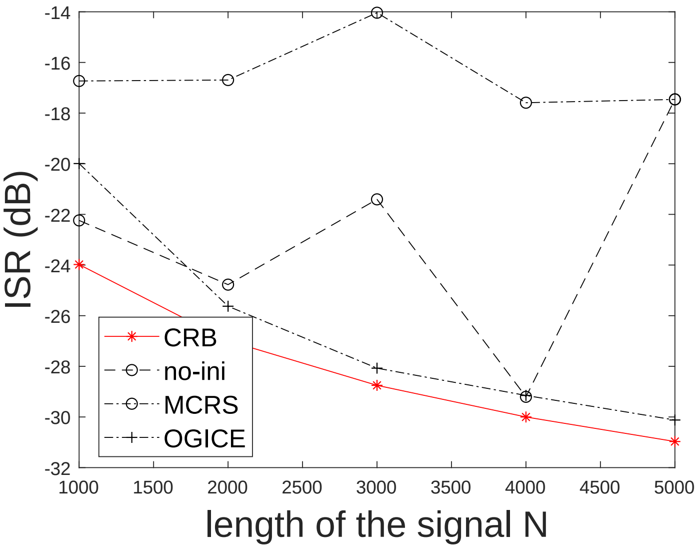

Figure 1: Accuracy comparison of selected separation algorithms: dashed line with circle points is used for the accuracy of OGICE with random initialization, dashed-and-dotted line with circles corresponds to the MCRS accuracy and dashed-and-dotted line with crosses to OGICE initialized by the results of MCRS.

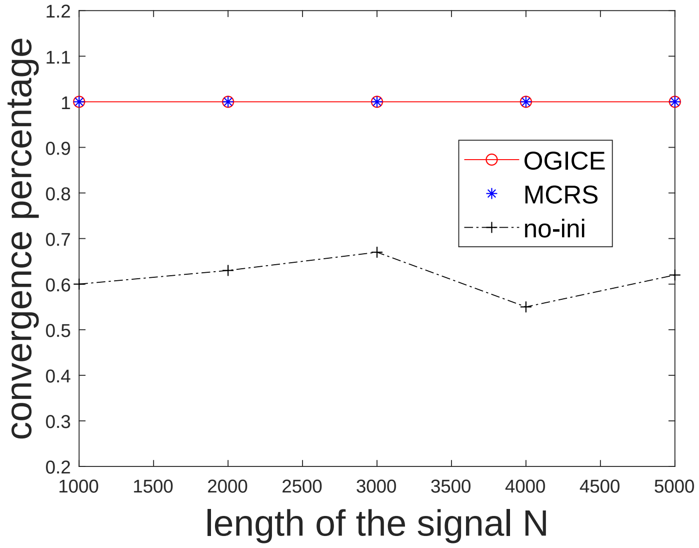

Figure 2: Convergence rate of selected separation algorithms: dashed-and-dotted line is used for OGICE with random initialization, stars correspond to the MCRS algorithm and full line to OGICE initialized by the results of MCRS.

5 Conclusions

Studied heuristic algorithms, MCRS and DEBR, reached equivalent accuracy in almost the same time. In comparison to the existing method OGICE, the heuristics are less accurate and the duration of one run is higher that for OGICE. However, MCRS and/or DEBR require no initialization and converge in all runs to the desired signal. The best separation accuracy was reached when OGICE was initialized by the result of MCRS, since then the algorithm attained the CRLB.

Acknowledgements

V. Kautský acknowledges: The Czech Science Foundation Project No. 17-00902S, and the United States Department of the Navy, Office of Naval Research Global, Project No. N62909-18-1-2040, and the Grant SGS18/188/OHK4/3T/14 provided by the Ministry of Education, Youth, and Sports of the Czech Republic (MŠMT ČR).
J. Štěch acknowledges: The Czech Science Foundation Project GA18-15970S.

References

[1] P. Tichavský and Z. Koldovský, *Fast and accurate methods of independent component analysis: A survey,* Kybernetika, vol. 47, no. 3, pp. 426438, 2011.

[2] V. Kautský, Z. Koldovský, and P. Tichavský, *Cramer-Rao-induced bound for interference-to-signal ratio achievable through non-gaussian independent component*

extraction, in 2017 IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Dec 2017, pp. 9497.

[3] Z. Koldovský and F. Nesta, *Performance analysis of source image estimators in blind source separation,* IEEE Transactions on Signal Processing, vol. 65, no. 16, pp. 41664176, Aug. 2017.

[4] J. F. Cardoso, *Multidimensional independent component analysis,* in Proceedings of IEEE International Conference on Audio, Speech and Signal Processing, vol. 4, May 1998, pp. 19411944 vol.4.

[5] Z. Koldovský, P. Tichavský, and V. Kautský, *Orthogonally constrained independent component extraction: Blind MPDR beamforming,* in Proceedings of European Signal Processing Conference, Sep. 2017, pp. 11951199.

[6] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis.* John Wiley & Sons, 2001.

[7] P. Tichavský, Z. Koldovský, and E. Oja, *Performance analysis of the FastICA algorithm and Cramer-Rao bounds for linear independent component analysis,* IEEE Transactions on Signal Processing, vol. 54, no. 4, pp. 11891203, April 2006.

[8] Josef Tvrdík, *Competitive Differential Evolution*, Mendel. 2006.

[9] I. Křivý, J. Tvrdík, Radek Krpec, *Stochastic algorithms in nonlinear regression*, Computational Statistics & Data Analysis 33 (2000) 277-290.