BLIND SEPARATION USING ABSOLUTE MOMENTS BASED ADAPTIVE ESTIMATING FUNCTION

Juha Karvanen and Visa Koivunen

Signal Processing Laboratory
Helsinki University of Technology
P.O. Box 3000, FIN-02015 HUT, Finland
Tel. +358 9 451 2455 Fax +358 9 460 224
{juha.karvanen, visa.koivunen}@hut.fi

ABSTRACT

We propose new absolute moment based estimating functions for blind source separation purposes. Absolute moments are a computationally simple choice that can also adapt to the skewness of source distributions. They have lower sample variance than cumulants employed in many widely used ICA (Independent Component Analysis) methods. The complete estimating function consists of two parts that are sensitive to peakedness and asymmetry of the distribution, respectively. Expression for optimal weighting between the parts is derived using an efficacy measure. The performance of the proposed contrast and employed efficacy measure are studied in simulations.

1. INTRODUCTION

This paper deals with the problem how skewness of source distributions can be exploited in Blind Source Separation (BSS) or Independent Component Analysis (ICA). Traditionally, BSS methods assume implicitly that source distributions are symmetric [1, 2, 3]. However, in many application areas, such as in biomedical signal processing and telecommunications, the source distributions may be skewed. As an example, fast and slow fading encountered in mobile digital communication systems are often characterized using Rayleigh and log-normal distributions that are both asymmetric.

In this paper we show how skewness information may be used to improve the estimator needed in finding the independent components and consequently improve the quality of separation. More dramatically, we demonstrate that in some cases ignoring the skewness information may lead to total failure in separation. We propose an estimating function based on absolute moments. It is composed of two parts: one associated with lack of symmetry (skewness) and the other characterizing peakedness (kurtosis) property of the distributions. We also derive expressions for optimal weighting of these properties in the estimating function based on an efficacy measure. Absolute moments are computationally simple and have lower sample variance compared to cumulants. The source distributions can be sub-Gaussian or super-Gaussian or even have a zero kurtosis. The skewness may be positive or negative or zero provided that only one source has both zero kurtosis and zero skewness. The properties of the contrast are studied by simulations. The usage of the proposed contrast is particularly easy since we can directly incorporate it to widely used blind separation algorithms.

This paper is organized as follows. In Section 2 contrast functions composed of symmetric and asymmetric parts are introduced. A contrast based on absolute moments is then proposed. Optimal weighting for combining symmetric and asymmetric part is derived in general case and for the proposed contrast. Simulation results demonstrating the performance gain obtained by using skewed contrasts are presented in Section 3. The performance of the weighting between symmetric and asymmetric parts is studied as well.

This work was funded by the Academy of Finland.

2. ICA CONTRASTS COMBINING SYMMETRIC AND ASYMMETRIC NON-LINEARITIES

We consider the ICA model with instantaneous mixing

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{1}$$

where the sources $\mathbf{s} = [s_1, s_2, \ldots, s_m]^T$ are mutually independent random variables and $\mathbf{A}_{m\times m}$ is an unknown invertible mixing matrix. The goal is to find only from the observations, $\mathbf{x} = [x_1, x_2, \ldots, x_m]^T$, a matrix $\mathbf{W}_{m\times m}$ such that the output

$$\mathbf{y} = \mathbf{W}\mathbf{x} \tag{2}$$

is an estimate of the possibly scaled and permutated source vector $\mathbf{s}$. The components of $\mathbf{y}$ are denoted y_i =

$\sum_{j=1}^{m} w_{ij} x_j$ where (w_{ij}) is the i, j element of matrix W. An ICA method consists of three parts: a theoretical measure of independence; an estimator (contrast, objective function) for the chosen theoretical measure, and finally an algorithm for minimizing or maximizing the selected objective function (see, [4]).

Because we are interested in separating source distributions that may also be asymmetric, we consider contrast functions that can be presented as a sum of absolute values of symmetric (even function) and asymmetric contrast (odd function). In general, these contrast functions can be presented as follows

$$\Phi(y_i; \omega_1, \omega_2) = \omega_1 |\Phi_{symmetric}(y_i)| + \omega_2 |\Phi_{asymmetric}(y_i)|, \quad (3)$$

where ω_1, ω_2 are weighting parameters between the symmetric and the asymmetric contrast. Optimal values of the parameter ω_1, ω_2 are data dependent. In this section we propose a choice for the contrast function and a method to determine the optimal ω's iteratively.

2.1. Optimal weighting in combined ICA contrast

The performance analysis of contrast functions is considered in [4], [1], [5] and [6]. It is usually assumed in the analysis that all sources are identically distributed. Local stability is found to depend on the following non-linear moments

$$\vartheta_i = E\{\varphi'(s_i)\} - E\{s_i \varphi(s_i)\} \quad (4)$$

and the variance of separation solution is found to depend on

$$\xi_i = E\{\varphi(s_i)^2\} - E\{s_i \varphi(s_i)\}^2. \quad (5)$$

In [5] it is proposed that the following measure can be used as a performance criterion

$$\Upsilon = \frac{\vartheta_i^2}{\xi_i}. \quad (6)$$

This measure is called BSS efficacy and it is independent of the scaling of estimating function φ. The BSS efficacy gives us an analytical way to compare contrast functions. By maximizing the BSS efficacy we can find the optimal values for the weighting parameters.

Now we consider efficacy maximization in the case where the estimating function is weighted sum of two estimating functions

$$\varphi(s_i) = \omega_1 \varphi_1(s_i) + \omega_2 \varphi_2(s_i). \quad (7)$$

When performance criterion (6) is maximized under the constraint

$$\vartheta_i = \omega_1 \vartheta_{i,1} + \omega_1 \vartheta_{i,2} = 1, \quad (8)$$

where

$$\vartheta_{i,k} = E\{\varphi_k'(s_i)\} - E\{s_i \varphi_k(s_i)\} \quad (9)$$

we obtain

$$\begin{aligned}\omega_1 = &\Big(2E\{\varphi_2^2(s_i)\}\vartheta_{i,1} - 2E\{\varphi_1(s_i)\varphi_2(s_i)\}\vartheta_{i,2} + \\ &\vartheta_{i,2}^2 E\{s_i\varphi_1(s_i)\} - \vartheta_{i,1}\vartheta_{i,2}E\{s_i\varphi_2(s_i)\}\Big)\Big/ \\ &\Big(2(E\{\varphi_2^2(s_i)\}\vartheta_{i,1}^2 - 2E\{\varphi_1(s_i)\varphi_2(s_i)\}\vartheta_{i,1}\vartheta_{i,2} + \\ &E\{\varphi_1^2(s_i)\}\vartheta_{i,2}^2(s_i))\Big) \quad (10)\\ \omega_2 = &\Big(2E\{\varphi_1^2(s_i)\}\vartheta_{i,2} - 2E\{\varphi_1(s_i)\varphi_2(s_i)\}\vartheta_{i,1} + \\ &\vartheta_{i,1}^2 E\{s_i\varphi_2(s_i)\} - \vartheta_{i,1}\vartheta_{i,2}E\{s_i\varphi_1(s_i)\}\Big)\Big/ \\ &\Big(2(E\{\varphi_2^2(s_i)\}\vartheta_{i,1}^2 - 2E\{\varphi_1(s_i)\varphi_2(s_i)\}\vartheta_{i,1}\vartheta_{i,2} + \\ &E\{\varphi_1^2(s_i)\}\vartheta_{i,2}^2(s_i))\Big). \quad (11)\end{aligned}$$

2.2. Contrast based on absolute moments

The simplest ICA contrasts are based on cumulants. Cumulant based skewness and kurtosis are defined as follows

$$\kappa_3^\circ(y_i) = \frac{\kappa_3(y_i)}{\kappa_2(y_i)^{3/2}} = E\left\{\left(\frac{y_i - \mu_{y_i}}{\sigma_{y_i}}\right)^3\right\} \quad (12)$$

$$\kappa_4^\circ(y_i) = \frac{\kappa_4(y_i)}{\kappa_2(y_i)^2} = E\left\{\left(\frac{y_i - \mu_{y_i}}{\sigma_{y_i}}\right)^4\right\} - 3, \quad (13)$$

where μ_{y_i} and σ_{y_i} are the expected value and the standard deviation of y_i, respectively. Cumulant based contrasts can be easily modified to contrast functions that possess more complicated theoretical properties but may in some cases have better performance in practice. Instead of cumulants we can use measures derived from the absolute moments [7] defined by

$$\nu_q(y_i) = E\{|y_i - \mu|^q\}, \quad (14)$$

where μ is the expected value of the distribution. The even absolute moments are equal to conventional central moments of the same order but the odd absolute moments cannot be directly written in terms of central moments. In addition, we may define the qth skewed absolute moment by

$$\begin{aligned}\nu_q^*(y_i) = &E\{(y_i - \mu)|y_i - \mu|^{q-1}\} = \\ &E\{\text{sign}(y_i - \mu)|y_i - \mu|^q\}. \quad (15)\end{aligned}$$

Analogously to the absolute moments, the odd skewed absolute moments are equal to usual central moments of the same order but the even skewed absolute moments cannot be directly written in terms of the central moments.

The kurtosis of a distribution with unit variance can be measured by the third absolute moment

$$\nu_3(y_i) = E\left\{|y_i - \mu|^3\right\}. \tag{16}$$

As a measure for skewness we can use the second skewed absolute moment

$$\nu_2^*(y_i) = E\left\{|y_i - \mu|(y_i - \mu)\right\}. \tag{17}$$

Exploiting ν_3 and ν_2^* we may construct an ICA contrast. First, we find that for Gaussian random variable y_i with $\mu = 0$ and $\sigma^2 = 1$

$$\nu_3(y_i) = \int_{-\infty}^{\infty} |y_i|^3 \frac{1}{\sqrt{2\pi}} e^{-y_i^2/2} dy_i = 2\sqrt{\frac{2}{\pi}} \approx 1.59577 \tag{18}$$

and $\nu_2^*(y_i) = 0$. Furthermore, we define measures resembling the cumulant based kurtosis and skewness

$$\nu_3^\circ(y_i) = \nu_3(\frac{y_i - \mu}{\sigma}) - 2\sqrt{\frac{2}{\pi}} \tag{19}$$

$$\nu_2^\circ(y_i) = \nu_2^*(\frac{y_i - \mu}{\sigma}). \tag{20}$$

The behavior of ν_3° and ν_2° appears to be analogous to the behavior of cumulant based kurtosis and skewness, respectively. Usually, the sign of ν_3° equals to the sign of κ_4° and the sign of ν_2° equals to the sign of κ_3°. The polynomial orders of ν_3° and ν_2° are lower than the polynomial orders of κ_4° and κ_3°. This suggests that estimators of ν_3° and ν_2° have lower variance than estimators of κ_4° and κ_3° and smaller sample sizes are needed. We propose the contrast

$$\Phi_\nu(y_i; \omega_1, \omega_2) = \omega_1|\nu_3^\circ(y_i)| + \omega_2|\nu_2^\circ(y_i)|. \tag{21}$$

The estimating function related to contrast (21) and the derivative of the estimating function are given by

$$\varphi_\nu(y_i; \omega_1, \omega_2) = \omega_1 \text{sign}(\nu_3^\circ)3|y_i|y_i + \omega_2 \text{sign}(\nu_2^\circ)2|y_i| \tag{22}$$

$$\varphi_\nu'(y_i; \omega_1, \omega_2) = \omega_1 \text{sign}(\nu_3^\circ)6|y_i| + \omega_2 \text{sign}(\nu_2^\circ)2\text{sign}(y_i). \tag{23}$$

Based on the efficacy measure (6) the optimal weighting parameter for the absolute moment based contrast function can be given as follows

$$\begin{aligned}\omega_1 =& \Big(2\vartheta_{i,1}^2 - 2\,\text{sign}(\nu_3^\circ)\text{sign}(\nu_2^\circ)2\mu_3\vartheta_{i,2} + \\ & \text{sign}(\nu_3^\circ)\nu_3\vartheta_{i,2}^2 - \text{sign}(\nu_2^\circ)\nu_2^*\vartheta_{i,1}\vartheta_{i,2}\Big)\Big/ \\ & \Big(2(\vartheta_{i,1}^2 - 2\,\text{sign}(\nu_3^\circ)\text{sign}(\nu_2^\circ)2\mu_3\vartheta_{i,1}\vartheta_{i,2} + \\ & mu_4\vartheta_{i,2}^2)\Big)\end{aligned} \tag{24}$$

$$\begin{aligned}\omega_2 =& \Big(2\mu_4\vartheta_{i,2}^2 - 2\,\text{sign}(\nu_3^\circ)\text{sign}(\nu_2^\circ)2\mu_3\vartheta_{i,1} + \\ & \text{sign}(\nu_2^\circ)\nu_2^*\vartheta_{i,1}^2 - \text{sign}(\nu_3^\circ)\nu_3\vartheta_{i,1}\vartheta_{i,2}\Big)\Big/ \\ & \Big(2(\vartheta_{i,1}^2 - 2\,\text{sign}(\nu_3^\circ)\text{sign}(\nu_2^\circ)2\mu_3\}\vartheta_{i,1}\vartheta_{i,2} + \\ & mu_4\vartheta_{i,2}^2)\Big),\end{aligned} \tag{25}$$

where μ_q is the qth central moment, ν_q is the qth absolute moment, ν_q^* is the qth skewed absolute moment and

$$\vartheta_{i,1} = 2\,\text{sign}(\nu_3^\circ)\nu_1 - \text{sign}(\nu_3^\circ)\nu_3 \tag{26}$$

$$\vartheta_{i,2} = \text{sign}(\nu_2^\circ)(\nu_0^* - \nu_2^*). \tag{27}$$

Instead of trying to estimate the optimal weighting for sources, we estimate the optimal weighting parameters for the current data $\mathbf{y}$. Because $\mathbf{y}$ is known, the statistics in (24) and (25) can be replaced by their sample counterparts. It is clear that this approach does not always lead to optimal estimates but in the weighting problem an approximate solution provides sufficiently good performance and the computation is straightforward.

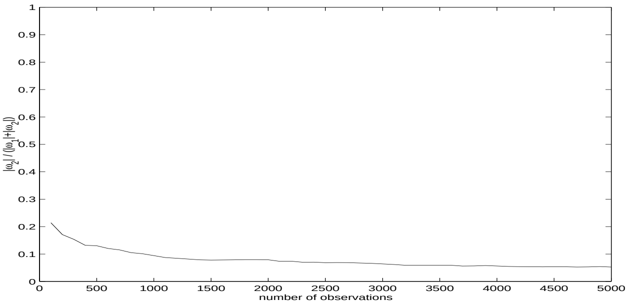
3. EXAMPLES

3.1. Weighting between symmetric and asymmetric contrasts

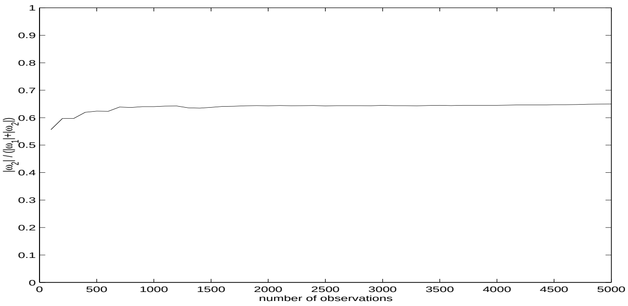
Examples on weighting between symmetric and asymmetric parts in the combined contrast are presented next. The converge of (24) and (25) is demonstrated in three different situations: Uniform distributed source (i.e. symmetric case) and two cases with different values of skewness. The convergence of the weighting parameters are presented as a function of the sample size in Figure 1.

3.2. Two source example

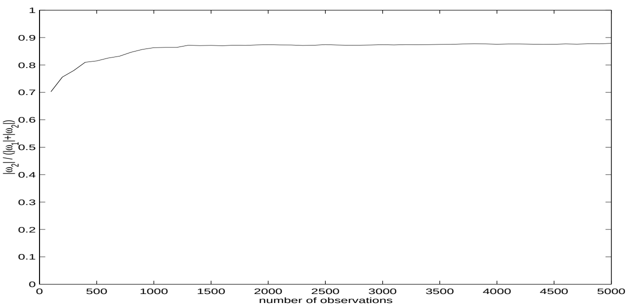
Quality of separation of the proposed ICA criterion is studied in simulation experiments. The absolute moments based contrast is straightforwardly implemented to FastICA algorithm [3]. Other gradient type algorithms are suitable, too. Comparisons are made with the standard FastICA contrasts. Differences in performance between symmetric contrast and its asymmetric generalization are illustrated. As a



(a) Uniform distribution



(b) GLD, $\kappa_3 = 0.3$, $\kappa_4 = -0.5$



(c) GLD, $\kappa_3 = 0.5$, $\kappa_4 = 1$

Fig. 1. An example on convergence of the optimal weighting in the case of absolute moment contrast. Parameters (24) and (25) are estimated from a single source and the ratio $|\omega_2|/(|\omega_1| + |\omega_2|)$ is plotted. Uniform distribution is symmetric and thus the ratio of the weighting parameters converges towards zero when the number of observations increases. The other examples generated from Generalized Lambda Distribution (GLD) [8, 9] are asymmetric and consequently the asymmetric estimating functions have large weights. The values presented are means over 100 realizations.

benchmark we use a simple case where two skewed sources are mixed. The overall goal of the simulations is to demonstrate that the proposed contrast separates reliably both symmetric and asymmetric sources. In the case of symmetric sources the the proposed asymmetric contrast reduces to the symmetric contrast and it is expected that its performance is approximately equal to that of conventional contrasts. When the source distributions are asymmetric, it is expected that the asymmetric contrast will outperform the conventional contrasts. Both sub-Gaussian and super-Gaussian sources are used in simulations. The details of source signal statistics are given in Table 1 and the density functions of the theoretical distributions are presented in Figure 2. In each simulation the length of signals is 10000 and the number of realizations is 1001. A full rank random mixing matrix is generated for each realization.

Simulation	Source 1			Source 2		
	pdf	κ_3	κ_4	pdf	κ_3	κ_4
A	Laplace	0	3	Laplace	0	3
B	Uniform	0	-1.2	Uniform	0	-1.2
C	GLD	0.5	1	GLD	0.3	0.5
D	GLD	0.5	1	GLD	0.3	-0.5
E	GLD	-0.5	-0.1	GLD	0.3	-0.5
F	GLD	0.3	0	GLD	0.3	0
G	GLD	0.8	0	GLD	0.8	0
H	GLD	0.8	0.1	GLD	0.8	0.1

Table 1. The theoretical third and fourth cumulants of source signals for the simulation experiments. The sources have zero mean and unit variance. In simulations from C to H, sources are generated from the Generalized Lambda Distribution (GLD) [8, 9] with the corresponding theoretical cumulants. The density functions are visualized in in Figure 2.

The quality of the separation is measured by Signal to Interference Ratio *SIR(dB)*= $-10\log_{10}(\text{MSE})$, where MSE stands for Mean Square Error *MSE*= $E\left\{(s(t) - y(t))^2\right\}$). To eliminate scaling differences both original signals and extracted signals are normalized to have zero mean and unit variance before the calculation of the SIR. After that source signals are matched to the extracted signals so that the resulting MSE values are as small as possible.

Boxplots are used to describe the SIR-values in simulations. The boxplot is graphical presentation tool for sample distributions and it is widely utilized in applied statistics. In a boxplot the box defines the quintile range (from 25% percentile to 75% percentile). The line inside the box is the median. The 'whiskers' are lines extending from each end of the box to show the extent of the rest of the data. The length of a whisker is defined as 1.5 times the length of the quintile range. Nevertheless, the whiskers are always bound by sample minimum and maximum. The possible outliers, outside the whiskers area, are marked by crosses.

In Figure 3, the SIR values of the first extracted signals are presented. In simulations A-E all contrast functions did

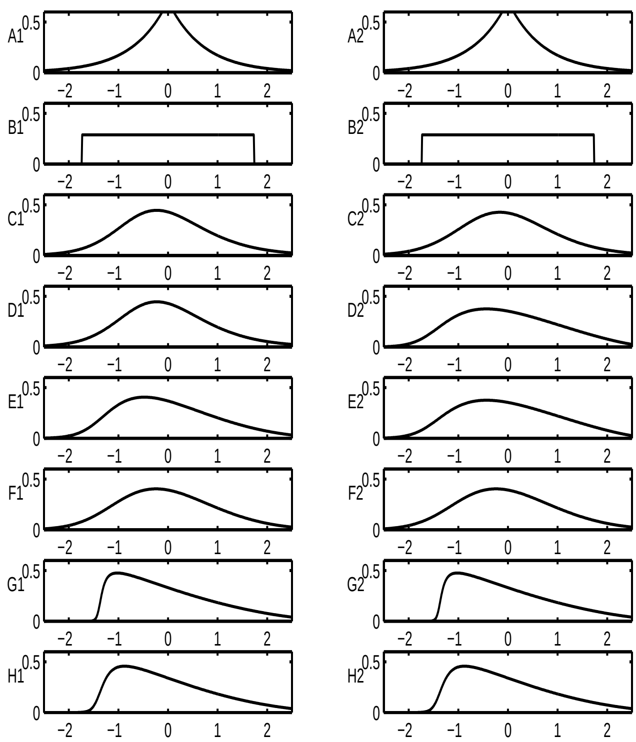


Fig. 2. The theoretical density functions of source distributions. The letter of the simulation and the number of source are given on the left side of each plot. The cumulants of distributions are summarized in Table 1.

well: median SIR values are over 30 dB. In simulations F, G and H the symmetric contrasts are outperformed by the skewed absolute moment. The overall results indicate that the skewed absolute third moment contrast (3) are the most reliable among the methods considered. This is a reasonable result because even if the weighting parameter estimators (24) and (25) are not simple, they use only statistics up to the fourth order. These simulations are only simple special cases but they do, nonetheless, strongly support the idea that skewness information may significantly improve the quality of separation.

4. CONCLUSION

We considered blind separation using absolute moments as estimating functions. Absolute moments are computationally simple and have lower sample variance compared to cumulants. The proposed estimating function combines absolute moments (symmetric estimating function) and skewed absolute moments (asymmetric estimating function). Consequently, the separation remains good even if the source distributions are skewed. The optimal weighting between the symmetric and asymmetric estimating function is obtained using the concept of BSS efficacy. In practice, the weighting is adaptively estimated from data. Simulation examples demonstrate the reliable performance of the proposed method for various different sources.

5. REFERENCES

[1] Jean-Francois Cardoso and Beate Hvam Laheld, “Equivariant adaptive source separation,” *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.

[2] Shun-Ichi Amari, Andrzej Cichocki, and H. Yang, “A new learning algorithm for blind signal separation,” in *Advances in Neural Information Processing Systems*, vol. 8, pp. 757–763. MIT Press, Cambridge MA, 1996.

[3] Aapo Hyvärinen, “Fast and robust fixed-point algorithms for independent component analysis,” *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[4] Jean-Francois Cardoso, “Blind signal separation: Statistical principles,” *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, 1998.

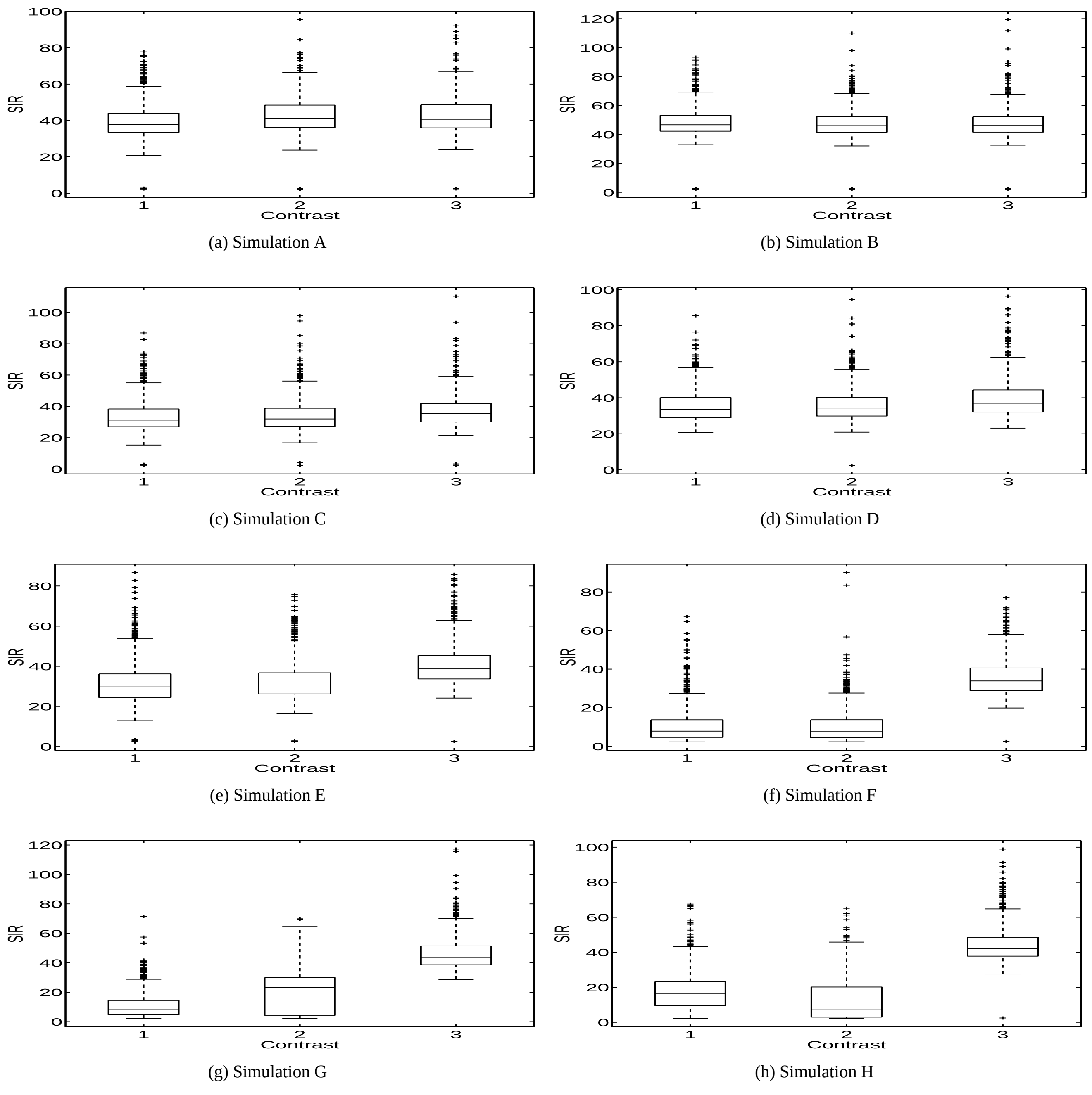
[5] Saleem A. Kassam, Yinglu Zhang, and George V. Moustakides, “Some results on a BSS algorithm under non-standard conditions,” in *Proc. of the 33rd Annual Conference on Information Sciences and Systems*, 1999.

[6] Jean-Francois Cardoso, “On the performance of orthogonal source separation algorithms,” in *Proc. EUPISCO*, 1994, pp. 776–779.

[7] Alan Stuart and J. Keith Ord, *Kendall’s Advanced Theory of Statistics: Distribution Theory*, vol. 1, Edward Arnold, sixth edition, 1994.

[8] Zaven A. Karian, Edward J. Dudewicz, and Patrick McDonald, “The extended generalized lambda distribution system for fitting distributions to data: History, completion of theory, tables, applications, the ”final word” on moment fits,” *Communications in Statistics: Simulation and Computation*, vol. 25, no. 3, pp. 611–642, 1996.

[9] Jan Eriksson, Juha Karvanen, and Visa Koivunen, “Source distribution adaptive maximum likelihood estimation of ICA model.,” in *ICA2000. Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation*, 2000, pp. 227–232.



(a) Simulation A (b) Simulation B

(c) Simulation C (d) Simulation D

(e) Simulation E (f) Simulation F

(g) Simulation G (h) Simulation H

Fig. 3. Boxplots of the SIR-values of first separated signal. The non-linearities in comparison are kurtosis (1), tanh (2) and skewed absolute moment (3). The source distributions used in simulations are summarized in Table 1 and in Figure 2.