A PARAMETER-FREE CAUCHY-SCHWARTZ INFORMATION MEASURE FOR INDEPENDENT COMPONENT ANALYSIS

Lei Sun[†,⁎], *Badong Chen*[‡], *Kar-Ann Toh*[⋆], *Zhiping Lin*[⁎]

[†] School of Information and Electronics, Beijing Institute of Technology,
Beijing 100081, PR China. (e-mail: sunlei@bit.edu.cn)
[†] Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University,
Xi'an 710049, PR China. (e-mail: chenbd@mail.xjtu.edu.cn)
[⋆] School of Electrical and Electronic Engineering, Yonsei University,
Seoul 120-749, Republic of Korea. (e-mail: katoh@ieee.org)
[⁎] School of Electrical and Electronic Engineering, Nanyang Technological University,
Singapore 639798, Singapore. (e-mail: ezplin@ntu.edu.sg)

ABSTRACT

Independent component analysis (ICA) by an information measure has seen wide applications in engineering. Different from traditional probability density function based information measures, a probability survival distribution based Cauchy-Schwartz information measure for multiple variables is proposed in this paper. Empirical estimation of survival distribution is parameter-free which is inherited by the estimation of the new information measure. This measure is proved to be a valid statistical independence measure and is adopted as an objective function to develop an ICA algorithm which is validated by an experiment. This work shows promising potential regarding the use of survival distribution based information measure for ICA.

Index Terms— Probability Survival Distribution, Information Measure, Independent Component Analysis, Blind Signal Separation

1. INTRODUCTION

Consider the estimation of D latent variables from a $N \times D$ observation matrix $\mathbf{X}$ representing a set of D variables each with N observations. The observations are assumed with linear but unknown combinations of the latent variables. The estimation goal is to find an $D \times D$ matrix $\mathbf{W}$ to recover the latent signals by

$$\hat{\mathbf{S}} = \mathbf{X}\mathbf{W}, \tag{1}$$

where $\hat{\mathbf{S}}$ is the recovered signal matrix with each column being estimations for one of the D latent variables. To find the matrix $\mathbf{W}$, the method of independent component analysis (ICA) assumes that the latent random variables are statistically independent and are non-Gaussian distributed (or at most one latent Gaussian signal) [1, 2]. In other words, the task of ICA is to recover signals by a contrast objective function based on statistical independence.

Statistical independence can be evaluated by different statistical measures. Besides the second order statistics which provides a weak independence measure, higher order statistics such as information measures are used to provide robust independence evaluation. For example, Shannon's entropy was proposed as an objective function in information maximization (InfoMax) [3] which has been an important ICA approach [2]. Another example is using Rényi's entropy as an ICA objective function [4]. However, numerical calculation of these information measures is nontrivial due to the dependency of underlying probability density estimation [5].

Different from the traditional *probability density* based information measures, several information measures defined on *probability distributions* have been recently proposed: cumulative residual entropy [6], generalized survival exponential entropy [7] and survival information potential (SIP) [8]. It is noted that empirical estimation of the distribution based SIP is parameter-free [8, 9] and can be directly computed from given samples [10, 11]. Inspired by this attractive parameter-free property in estimation, we extend the work in [11] from two random variables to multiple random variables to match the ICA application where latent variables are generally more than two.

2. METHODOLOGY AND ALGORITHM

2.1. Parameter-free survival distribution estimation

For a random variable X, its survival distribution (a.k.a. the cumulative residual distribution or the tail distribution) is the complementary of its probability cumulative distribution giv-

978-1-4799-9988-0/16/$31.00 ©2016 IEEE

en by

$$\bar{F}_X(x) = \int_x^\infty f_X(u)du \tag{2}$$

where $f_X(\cdot)$ denotes the probability density function of X and x denotes an evaluation point. Given a sample sequence $\{u_n\}, n = 1, \dots, N$, the *empirical* survival distribution can be obtained by [7, 8] $\bar{F}_N(x) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}(u_n > x)$, where $\mathbb{I}(\cdot)$ denotes the indicator function ($\mathbb{I}(A)$ is 1 if event A occurs and is 0 otherwise).

Remark 1 In probability density function estimation, for example by the Parzen window method [12], the estimator is $\hat{p}_X(x) = \frac{1}{N}\sum_{n=1}^{N} k(x, u_n)$ where $k(x, \cdot)$ denotes a kernel function. When Gaussian kernel is adopted in this estimator, the kernel width parameter needs to be properly set. It can be observed that the calculation for $\bar{F}_N(x)$ is free of parameter.

2.2. SCS-MI is a valid Statistical Independence measure

In this paper, for convenience and without loss of generality, we restrict our discussions to nonnegative random variables. This would not restrict its practical applications since data sets are mostly bounded in real-world situations and the non-negative constraint can be satisfied by data translation.

For two random variables X and Y (both in $\mathbb{R}_+$), Cross Survival Information Potential (CSIP) of the two random variables was defined in [8, 11] as

$$S_c(X, Y) = \int_{\mathbb{R}_+} \bar{F}_X(x)\bar{F}_Y(x)dx. \tag{3}$$

Survival Cauchy-Schwartz mutual information (SCS-MI) for two random variables is defined to evaluate the Cauchy-Schwartz divergence between the joint survival function $\bar{F}_{(X,Y)}(X, Y)$ and the product of the marginal survival functions $\bar{F}_X(X)\ \bar{F}_Y(Y)$ [11]:

$$\mathcal{I}_{\text{SCS}}(X; Y) \stackrel{\text{def}}{=} -\log \frac{S_c((X,Y), XY)}{\sqrt{S_c((X,Y),(X,Y))}\sqrt{S_c(XY, XY)}}, \tag{4}$$

where (X, Y) and XY denote $\bar{F}_{(X,Y)}(X, Y)$ and $\bar{F}_X(X)\bar{F}_Y(Y)$, respectively. The augment terms in (4) are based on the definition shown in (3) and the nominator term is explicitly given as an example

$$S_c((X,Y), XY) = \int_{\mathbb{R}_+} \bar{F}_{(X,Y)}(x, y)\bar{F}_X(x)\bar{F}_Y(y)dxdy. \tag{5}$$

It was mentioned in [11] without proof that the SCS-MI was a valid statistical independence evaluation measure. We re-state this result in the following proposition with a proof.

Proposition 1 *[11] $\mathcal{I}_{SCS}(X; Y) \geq 0$ and the equality holds if and only if X and Y are mutually independent.*

Proof: By substituting the survival distribution definition (2) and the CSIP definition (3) into the arguments in (4) with adoption of Cauchy-Schwartz inequality, the following inequality can be subsequently obtained as $\sqrt{S_c((X, Y), (X, Y))}\sqrt{S_c(XY, XY)} \geq S_c((X, Y), XY)$ which implies that $\mathcal{I}_{\text{SCS}}(X; Y) \geq 0$. $\mathcal{I}_{\text{SCS}}(X; Y) = 0$ holds if and only if $\bar{F}_{(X,Y)}(X, Y) = \bar{F}_X(X)\bar{F}_Y(Y)$ which corresponds to $\int\int_{\mathbb{R}_+} \mathbb{I}(X > x, Y > y)[f_{(X,Y)}(X, Y) - f_X(X)f_Y(Y)]dXdY = 0$. Since this equality holds for any evaluation values (x, y) and the indicator function $\mathbb{I}(\cdot)$ is non-negative, it implies $f_{(X,Y)}(X, Y) = f_X(X)f_Y(Y)$ holds almost everywhere which means X and Y are mutually independent. The above interpretations are reversible and therefore the proof is completed. ■

2.3. Empirical SCS-MI estimator for multiple variables

Let the capital letter with superscript X^d, $d = 1, \cdots, D$, denote the d-th random variable. We first generalize the CSIP definition (3) to D random variables by defining

$$\begin{aligned}
&S_c\big((X^1,\dots,X^D),(X^1,\dots,X^D)\big) = \int \Big[\bar{F}_{(X^1,\dots,X^D)}(x^1,\dots,x^D)\Big]^2 dx^1\cdots dx^D \\
&S_c\big(X^1\cdots X^D, X^1\cdots X^D\big) = \int \Big[\bar{F}_{X^1}(x^1)\times\cdots\times\bar{F}_{X^D}(x^D)\Big]^2 dx^1\cdots dx^D \\
&S_c\big((X^1,\dots,X^D), X^1\cdots X^D\big) = \\
&\int \Big[\bar{F}_{(X^1,\dots,X^D)}(x^1,\dots,x^D)\Big]\times\Big[\bar{F}_{X^1}(x^1)\times\cdots\times\bar{F}_{X^D}(x^D)\Big] dx^1\cdots dx^D,
\end{aligned} \tag{6}$$

where $(X^1, \dots, X^D)$ denotes the joint survival distribution $\bar{F}_{(X^1,\dots,X^D)}(X^1, \dots, X^D)$ and $X^1 \cdots X^D$ denotes the product of marginal survival distributions $\bar{F}_{X^1}(X^1)\cdots\bar{F}_{X^D}(X^D)$.

With the above definitions in (6), the SCS-MI definition (4) can be generalized to multiple (D, $D \geq 2$) random variables (all in $\mathbb{R}_+$) as

$$\mathcal{I}_{\text{SCSM}}\left(X^1; \dots; X^D\right) \stackrel{\text{def}}{=} -\log \frac{S_c\big((X^1,\dots,X^D), X^1\cdots X^D\big)}{\sqrt{S_c\big((X^1,\dots,X^D),(X^1,\dots,X^D)\big)}\sqrt{S_c\big(X^1\cdots X^D, X^1\cdots X^D\big)}}, \tag{7}$$

where subscript "SCSM" denotes the SCS-MI for multiple variables.

Proposition 2 $\mathcal{I}_{SCSM}\left(X^1; \dots; X^D\right) \geq 0$ *and the equality holds if and only if $X^1, \dots, X^D$ are mutually independent.*

The proof is similar to the one for Proposition 1 and the details are omitted here for brevity. By Proposition 2, SCS-MI can therefore be used for independence evaluation. However, empirical estimation of SCS-MI depends on the estimation of the CSIP augments therein. Here we present only the empirical estimator for the numerator part in (7) for brevity, and the estimators for the denominator can be obtained in a similar routine. Assume there are N samples for each of the D random variables, we get

$$
\begin{aligned}
&S_c\left((X^1,\ldots,X^D),X^1\cdots X^D\right)\\
&=\int [\bar{F}_{(X^1,\ldots,X^D)}(x^1,\ldots,x^D)]\times[\bar{F}_{X^1}(x^1)\times\cdots\times\bar{F}_{X^D}(x^D)]dx^1\cdots dx^D\\
&=\int [\frac{1}{N}\sum_{n=1}^{N}\mathbb{I}\left((X_n^1,\ldots,X_n^D)>\left(x^1,\ldots,x^D\right)\right)]\times\\
&[\frac{1}{N}\sum_{\alpha_1=1}^{N}\mathbb{I}\left(X_{\alpha_1}^1>x^1\right)]\times\cdots\times[\frac{1}{N}\sum_{\alpha_D=1}^{N}\mathbb{I}\left(X_{\alpha_D}^D>x^D\right)]dx^1\cdots dx^D\\
&=\frac{1}{N^{D+1}}\int[\sum_{n,\alpha_1,\ldots,\alpha_D=1}^{N}\mathbb{I}\left((X_n^1,\ldots,X_n^D)>\left(x^1,\ldots,x^D\right)\right)\times\\
&\quad\mathbb{I}\left(X_{\alpha_1}^1>x^1\right)\times\cdots\times\mathbb{I}\left(X_{\alpha_D}^D>x^D\right)]dx^1\cdots dx^D\\
&=\frac{1}{N^{D+1}}\sum_{n,\alpha_1,\ldots,\alpha_D=1}^{N}\int[\mathbb{I}\left(X_n^1>x^1\right)\times\mathbb{I}\left(X_{\alpha_1}^1>x^1\right)\\
&\quad\times\cdots\times\mathbb{I}\left(X_n^D>x^D\right)\times\mathbb{I}\left(X_{\alpha_D}^D>x^D\right)]dx^1\cdots dx^D\\
&=\frac{1}{N^{D+1}}\sum_{n,\alpha_1,\ldots,\alpha_D=1}^{N}[\int\mathbb{I}\left(\min(X_n^1,X_{\alpha_1}^1)>x^1\right)dx^1]\\
&\quad\times\cdots\times[\int\mathbb{I}\left(\min(X_n^D,X_{\alpha_D}^D)>x^D\right)dx^D]\\
&=\frac{1}{N^{D+1}}\sum_{n,\alpha_1,\ldots,\alpha_D=1}^{N}[\min(X_n^1,X_{\alpha_1}^1)\times\cdots\times\min(X_n^D,X_{\alpha_D}^D)],
\end{aligned}
$$

where $\min(X_n^d,X_{\alpha_1}^d)$ denotes the minimum function returning the minimal value of X_n^d and $X_{\alpha_1}^d$. Here, X_n^d denotes the n-th sample of the random variable X^d.

Once the estimators for the CSIP augments in (7) are obtained, the SCS-MI empirical estimator of (7) can be subsequently obtained by using the CSIP estimators and we finally get

$$
\begin{aligned}
&\hat{\mathcal{I}}_{\text{SCSM}}\left(X^1;\ldots;X^D\right)=\\
&-\log\frac{\sum_{n,\alpha_1,\alpha_2,\ldots,\alpha_D=1}^{N}\min(X_n^1,X_{\alpha_1}^1)\times\cdots\times\min(X_n^D,X_{\alpha_D}^D)}{\sqrt{\sum_{n,m=1}^{N}\left(\prod_{d=1}^{D}\min(X_n^d,X_m^d)\right)}\sqrt{\prod_{d=1}^{D}\left(\sum_{n,m=1}^{N}\min(X_n^d,X_m^d)\right)}}.
\end{aligned}
\tag{8}
$$

Remark 2 It is noted that the SCS-MI empirical estimator (8) provides a means for statistical independence evaluation in terms of survival distribution with no free parameter but just the observed samples.

2.4. Proposed ICA algorithm based on SCS-MI estimator

As stated in Proposition 2, the SCS-MI is a valid statistical independence measure. Therefore, it can be used as an objective function and the optimal matrix $\mathbf{W}^*$ for ICA in (1) is obtained by

$$
\mathbf{W}^*=\arg\min_{\mathbf{W}}\hat{\mathcal{I}}_{\text{SCSM}}\left(\hat{S}^1;\ldots;\hat{S}^D\right), \tag{9}
$$

where $\hat{S}^1,\ldots,\hat{S}^D$ denote the D estimated variables.

In order to get the optimal solution in (9), an orthogonal constraint $\mathbf{W}^T\mathbf{W}=\mathbf{I}$ can be used to simplify the optimization procedure. Here, we use the method of Givens' rotation [13, 14] to impose an orthogonal constraint on $\mathbf{W}$ by products of $D(D-1)/2$ parameterized rotations based on

$$
\mathbf{W}=\prod_{i=1}^{D-1}\prod_{j=i+1}^{D}G_{ij} \tag{10}
$$

where G_{ij} is the $D\times D$ identity matrix with its (i,j)-th element determined by θ_{ij} given by

$$
\begin{bmatrix}G_{ij}(i,i) & G_{ij}(i,j)\\ G_{ij}(j,i) & G_{ij}(j,j)\end{bmatrix}=\begin{bmatrix}\cos(\theta_{ij}) & \sin(\theta_{ij})\\ -\sin(\theta_{ij}) & \cos(\theta_{ij})\end{bmatrix}
$$

with θ_{ij} being the ij-th rotation angle parameter. Note that the $D(D-1)/2$ angle parameters completely determine the orthogonal matrix $\mathbf{W}$ [13].

Here, we use a gradient based optimization procedure to obtain the solution in (9). The SCS-MI in (9) (see (8)) is firstly abbreviated as

$$
\begin{aligned}
&\hat{\mathcal{I}}_{\text{SCSM}}\left(\hat{S}^1;\ldots;\hat{S}^D\right)=\frac{1}{2}\log S_c\left((\hat{S}^1,\ldots,\hat{S}^D),(\hat{S}^1,\ldots,\hat{S}^D)\right)\\
&+\frac{1}{2}\log S_c((\hat{S}^1\cdots\hat{S}^D),(\hat{S}^1\cdots\hat{S}^D))\\
&-\log S_c\left((\hat{S}^1,\ldots,\hat{S}^D),(\hat{S}^1\cdots\hat{S}^D)\right)=\frac{1}{2}\log I_a+\frac{1}{2}\log I_b-\log I_c
\end{aligned}
$$

where the logarithmic arguments are elliptically represented. With these abbreviations, the gradient of the SCS-MI is written as

$$
\begin{aligned}
&\frac{\partial}{\partial\theta_{ij}}\hat{\mathcal{I}}_{\text{SCSM}}=\sum_{i,j=1}^{D}\frac{\partial\hat{\mathcal{I}}_{\text{SCSM}}}{\partial w_{ij}}\frac{\partial w_{ij}}{\partial\theta_{ij}}\\
&=\sum_{i,j=1}^{D}\left(\frac{1}{2I_a}\frac{\partial I_a}{\partial w_{ij}}+\frac{1}{2I_b}\frac{\partial I_b}{\partial w_{ij}}-\frac{1}{I_c}\frac{\partial I_c}{\partial w_{ij}}\right)\left(\prod_{u=1}^{D-1}\prod_{v=u+1}^{D}\tilde{G}_{uv}\right)_{ij}
\end{aligned}
\tag{11}
$$

where $\tilde{G}_{uv}=\frac{\partial G_{uv}}{\partial\theta_{uv}}$ when $u=i$ and $v=j$, and $\tilde{G}_{uv}=G_{uv}$ otherwise. The derivative of the matrix G_{uv} with respect to the angle parameter θ_{uv} is given by

$$
\frac{\partial}{\partial\theta_{uv}}\begin{bmatrix}G_{uv}(u,u) & G_{uv}(u,v)\\ G_{uv}(v,u) & G_{uv}(v,v)\end{bmatrix}=\begin{bmatrix}-\sin(\theta_{uv}) & \cos(\theta_{uv})\\ -\cos(\theta_{uv}) & -\sin(\theta_{uv})\end{bmatrix}. \tag{12}
$$

By some strait forward calculations, the partial derivatives in (11) can be obtained as

$$
\frac{\partial I_a}{\partial w_{ij}}=\sum_{u,v=1}^{N}\left(\left(\frac{\partial Ad}{\partial w_{dj}}\right)_{uv}\circ\bigodot_{k=1,k\neq d}^{D}(Ak)_{uv}\right) \tag{13}
$$

$$
\frac{\partial I_b}{\partial w_{ij}}=\left(\sum_{u,v=1}^{N}\left(\frac{\partial Ad}{\partial w_{dj}}\right)_{uv}\right)\times\prod_{k=1,k\neq d}^{D}\left(\sum_{u,v=1}^{N}(Ak)_{uv}\right) \tag{14}
$$

$$
\frac{\partial I_c}{\partial w_{ij}}=\sum_{u,v=1}^{N}\left(\left(\frac{\partial Ad}{\partial w_{dj}}\right)_{uv}\circ\left(\bigodot_{k=1,k\neq d}^{D}\left(\sum_{v=1}^{N}(Ak)_{uv}\right)\times\mathbf{1}_{1\times N}\right)\right) \tag{15}
$$

where "$\circ$" denotes the Hadamard product, "$\bigodot$" denotes series of the Hadamard products and $\mathbf{1}_{1\times N}$ is an N-length vector with all elements being 1. Here we use two abbreviations to denote matrices with their respective (u,v)-th element.

Algorithm 1: SCS-MI based ICA Algorithm

Input: $\{\tilde{x}_n^d\}, n = 1 \dots N, d = 1, \dots, D$

Output: $\mathbf{W}^*, \hat{\mathbf{S}}$

Initialization

$\eta = 10^{-2}, tol = 10^{-4}, t = 2 \times D\,\mathbf{X} \leftarrow h\left(\{\tilde{x}_n^d\}\right)^*$; `// h(·) denotes whitening`

$\boldsymbol{\theta} = [\dots \theta_u \dots] = [1, \dots, 1]^T, u = 1, \dots, D(D-1)/2$; `// angle parameters`

while *tol and iteration k are within valid range* † **do**

$\theta_u^{(k+1)} = \theta_u^{(k)} - \eta \nabla_{\theta_u} \hat{\mathcal{I}}_{\text{SCSM}}\left(\hat{S}^1; \dots; \hat{S}^D\right)^{\ddagger}$;

$\mathbf{W}^{(k)} \leftarrow \prod_{i=1}^{D-1} \prod_{j=i+1}^{D} G_{ij}(\boldsymbol{\theta})$; `// ref.` (10)

$\hat{\mathbf{S}}^{(k)} \leftarrow \mathbf{X} \times \mathbf{W}^{(k)}$; `// ref.` (1)

$tol \leftarrow \hat{\mathcal{I}}_{\text{SCSM}}(\hat{\mathbf{S}}^{(k)} + t)$; `// ref.` (8)

$\theta_u^{(k)} = \theta_u^{(k+1)}$;

return $\mathbf{W}^*$, $(\hat{\mathbf{S}} = \mathbf{X}\mathbf{W}^*)$; `// ref.` (1)

$*$: observations are firstly whitened to ensure $E[\mathbf{X}^T\mathbf{X}] = \mathbf{I}$ where $E[\cdot]$ denotes the expectation operator, ref. [15, 3].

† : tolerance tol and iterations k can be set when running an application

‡ : ref. (9), (11), (12), (13), (14) and (15).

With the above results (11)-(15), the SCS-MI gradient is obtained and subsequently used for a gradient based ICA search as shown in Algorithm 1.

Algorithm 1 is a demo to show the effectiveness of using the CSIP esitmator (8) for independence evaluation. The computational complexity in Algorithm 1 is dominated by calculating $\hat{\mathcal{I}}_{\text{SCSM}}(\hat{\mathbf{S}}^{(k)} + t)$ at about $\mathcal{O}(3 \times (D+1) \times N^2)$ which could be large with increment of sample number N and iterations. We leave the development of a computationally more efficient algorithm as our future work.

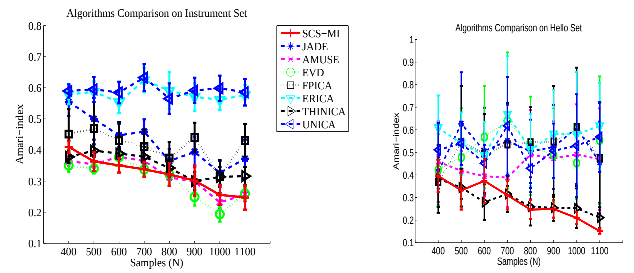
3. EXPERIMENT

In this experiment, the proposed ICA Algorithm 1 (SCS-MI) is compared with several existing well-known ICA techniques namely, the method of Joint Approximate Diagonalization of the Eigen-matrices (JADE) [16], the Algorithm for Multiple Unknown Source Extraction (AMUSE) [17], the method employing EigenValue Decomposition (EVD) [18], the Fixed-Point ICA method (FPICA) [5], the Equivariant Robust ICA (ERICA) [19], the Thin algorithm for ICA (ThinICA) [19] and the Unbiased quasi Newton algorithm for ICA (UNICA) [20]. Here, all the selected ICA methods deal with the samples in the input data space and we exclude those methods involving kernel mapping of input data (such as the kernel ICA [21] and the Hilbert-Schmidt independence criterion [22]).

Amari-index is used for de-mixing matrix quality assessment which was defined in [23] as

$$\text{Amari-index}(\mathbf{W}^*, \mathbf{M}) = \frac{1}{2D}\sum_{i=1}^{D}\left(\frac{\sum_{j=1}^{D}|r_{ij}|}{\max_j |r_{ij}|} - 1\right) + \frac{1}{2D}\sum_{j=1}^{D}\left(\frac{\sum_{i=1}^{D}|r_{ij}|}{\max_i |r_{ij}|} - 1\right) \tag{16}$$

where $r_{ij} = (\mathbf{W}^* \times \mathbf{M})_{ij}$ and $\mathbf{W}^*$ denotes the recovered demixing matrix. The Amari-index is equal to zero when two matrices represent the same components.



(a) Instrument (b) Hello

Fig. 1. Performance index comparison among algorithms. The experimental results implemented on the Instrument data set and the Hello data set are shown in (a) and (b), respectively. The number of samples in experiments is varying in $[400, 1100]$ (denoted in x-axes) with the samples randomly selected from a region of the data sets. The mixing matrix is a 2×2 matrix with elements randomly selected from the region of $(0, 1)$. The performance of each method is obtained from 50 separate runs.

Two real-world data sets are adopted in our experiment: the first data set includes recordings of two musical instruments (*violin* and *guitar*), and the second data set includes recordings of two speakers saying the English word “hello” [23]. All compared algorithms are tested on the two data sets and the experimental results are shown in Fig. 3 (a) and (b), respectively. It can be observed that the performance of the proposed SCS-MI method is among the best of all the competing methods in this experiment.

4. CONCLUSION

The newly defined multiple variable SCS-MI (7) is a generalization of the two variable SCS-MI (4). The SCS-MI was proved to be a valid statistical independence measure and a parameter-free SCS-MI estimator was given in (8). Based on the SCS-MI estimator, a novel independent component analysis algorithm was proposed and was validated by an experiment. Comparison with several well-known ICA methods showed relatively good performance of the developed method. This work shows promising potential regarding the use of survival distribution in ICA.

5. ACKNOWLEDGMENT

This work was supported by 973 Program (No. 2015CB351703) and National NSF of China (No. 61372152).

6. REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?," *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.

[2] T. Adali, M. Anderson, and G. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 18–33, 2014.

[3] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[4] D. Erdogmus, K. E. Hild, J. C. Principe, M. Lazaro, and I. Santamaria, "Adaptive blind deconvolution of linear channels using Renyi's entropy with Parzen window estimation," *IEEE Trans. on Signal Processing*, vol. 52, no. 6, pp. 1489–1498, 2004.

[5] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[6] M. Rao, Y. Chen, B. C. Vemuri, and F. Wang, "Cumulative residual entropy: a new measure of information," *IEEE Trans. on Information Theory*, vol. 50, no. 6, pp. 1220–1228, 2004.

[7] K. Zografos and S. Nadarajah, "Survival exponential entropies," *IEEE Trans. on Information Theory*, vol. 51, no. 3, pp. 1239–1246, 2005.

[8] B. Chen, P. Zhu, and J. C. Principe, "Survival information potential: a new criterion for adaptive system training," *IEEE Trans. on Signal Processing*, vol. 60, no. 3, pp. 1184–1194, 2012.

[9] B. Chen, N. Zheng, and J.C. Principe, "Survival kernel with application to kernel adaptive filtering," in *The 2013 International Joint Conference on Neural Networks*. IEEE, 2013, pp. 1–6.

[10] L. Sun, B. Chen, K. Toh, and Z. Lin, "Sequential extreme learning machine incorporating survival error potential," *Neurocomputing*, vol. 155, pp. 194–204, 2015.

[11] B. Chen, X. Yang, H. Qu, J. Zhao, N. Zheng, and J. C. Principe, "Feature selection based on survival Cauchy-Schwartz mutual information," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2014, pp. 6711–6715.

[12] J. C. Principe, *Information Theoretic Learning: Rényi's Entropy and Kernel Perspectives*, Springer Science & Business Media, 2010.

[13] L. Faivishevsky and J. Goldberger, "ICA based on a smooth estimation of the differential entropy," in *Advances in neural information processing systems*, 2009, pp. 433–440.

[14] U. Shalit and G. Chechik, "Coordinate-descent for learning orthogonal matrices through Givens rotations," in *Proceedings of the 31st International Conference on Machine Learning*, 2014, pp. 548–556.

[15] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4, pp. 411–430, 2000.

[16] J. F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM journal on matrix analysis and applications*, vol. 17, no. 1, pp. 161–164, 1996.

[17] L. Tong, Y. Inouye, and R. Liu, "Waveform-preserving blind estimation of multiple independent sources," *IEEE Trans. on Signal Processing*, vol. 41, no. 7, pp. 2461–2470, 1993.

[18] P. Georgiev and A. Cichocki, "Robust blind source separation utilizing second and fourth order statistics," in *Artificial Neural Networks - ICANN 2002*, pp. 1162–1167. Springer, 2002.

[19] S. Cruces and A. Cichocki, "Combining blind source extraction with joint approximate diagonalization: Thin algorithms for ICA," in *Proceedings of 4th International Symposium on Independent Component Analysis and Blind Signal Separation*, 2003, pp. 463–468.

[20] S. Cruces, A. Cichocki, and S. Amari, "On a new blind signal extraction algorithm: different criteria and stability analysis," *IEEE Signal Processing Letters*, vol. 9, no. 8, pp. 233–236, 2002.

[21] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *The Journal of Machine Learning Research*, vol. 3, pp. 1–48, 2002.

[22] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two-sample-problem," in *Advances in neural information processing systems*, 2006, pp. 513–520.

[23] A. Cichocki, S. Amari, K. Siwek, and T. Tanaka, "ICALAB toolboxes for signal and image processing, [http://www.bsp.brain.riken.jp/icalab]," 2002-2014.