BLIND SEPARATION USING A CLASS OF NEW INDEPENDENCE MEASURES

Yang Chen   Zhenya He

Department of Radio Engineering, Southeast University, Nanjing, China

ABSTRACT

Quasi-entropy (QE), a class of new independence measures, is proposed. By optimizing QE, blind separation of signals with arbitrary continuous distributions is achieved. Simulations verify these results.

1. INTRODUCTION

Blind signal separation (BSS), or blind separation, is the task to recover independent source signals from their mixtures without knowing the mixing coefficients. In instantaneous BSS, one has samples of d sensor signals (mixtures) collected in $\boldsymbol{x}(t) = [x_1(t), \cdots, x_d(t)]^T$ and described by

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t), \qquad t = 1, \cdots, N \tag{1}$$

where $\boldsymbol{s}(t) = [s_1(t), \cdots, s_n(t)]^T$ comprises n unknown independent source signals with at most one being Gaussian, $d \geq n$, and $\boldsymbol{A}$ is an unknown $d \times n$ full-rank mixing matrix. To recover $\boldsymbol{s}(t)$ from $\boldsymbol{x}(t)$, an $n \times d$ de-mixing matrix $\boldsymbol{B}$ is used to give the output signals $\boldsymbol{y}(t) = [y_1(t), \cdots, y_n(t)]^T$ by

$$\boldsymbol{y}(t) = \boldsymbol{B}\boldsymbol{x}(t), \qquad t = 1, \cdots, N \tag{2}$$

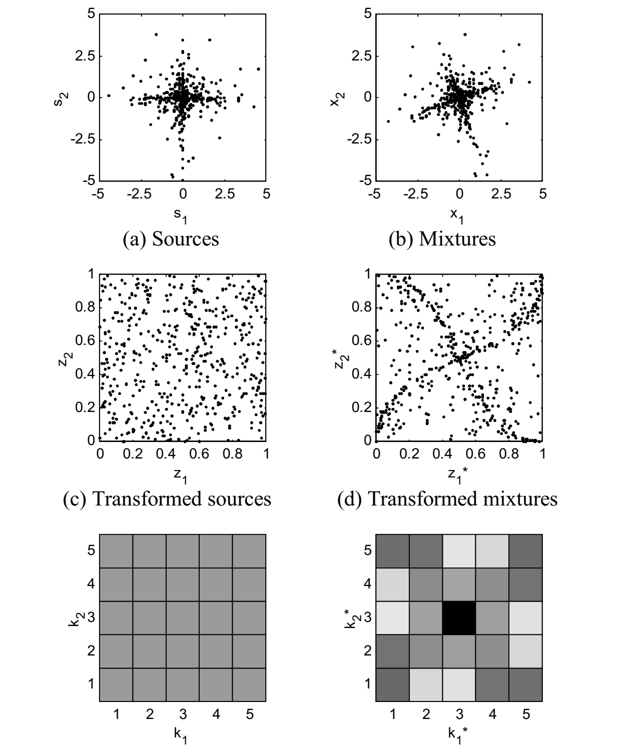
The components of $\boldsymbol{y}(t)$, *if and only if* (iff) they are independent, will become shuffled and scaled versions of those in $\boldsymbol{s}(t)$.

As we know, *n random variables* (RVs) $r_1, \cdots, r_n$ are independent iff their joint *probability density function* (PDF) can be factored as the product of their individual (marginal) PDFs, i.e.,

$$p_{r_1 \cdots r_n}(u_1, \cdots, u_n) = \prod_{i=1}^{n} p_{r_i}(u_i), \qquad \forall (u_1, \cdots, u_n) \tag{3}$$

In practice however, these PDFs are usually unknown. So BSS is generally achieved by optimizing a criterion regarding the independence of the output signals.

Due to its wide applications, BSS was intensively studied recently. Various criteria and algorithms were proposed. But many of them have restrictions on source distributions. For instance, maximum likelihood algorithms require a rough fitness to some distribution model [1]. And the problem with the rather popular kurtosis-based algorithms is they cannot deal with two or more zero-kurtosis sources, since the kurtosis of the mixture of zero-kurtosis sources is still zero. Designing algorithms that can separate zero-kurtosis sources is important. Because we may encounter some signals, such as images, whose kurtoses are exactly or very near zeros.



(a) Sources (b) Mixtures

(c) Transformed sources (d) Transformed mixtures

(e) Quantized transformed sources (f) Quantized transformed mixtures

Fig. 1. Illustration of quasi-entropy (QE).

In this paper, a class of new independence measures named quasi-entropy (QE) is proposed, algorithms based on which can separate signals with arbitrary continuous distributions and favor digital implementation.

2. A CLASS OF NEW INDEPENDENCE MEASURES

Hereafter, we denote the *cumulative distribution function* (CDF) of RV r by $q_r(\cdot)$, $q_r(u) = \text{Prob}(r \leq u) = \int_{-\infty}^{u} p_r(v)\,dv$.

This work was supported by the National Natural Science Foundation of China under Grant No. 60202014.

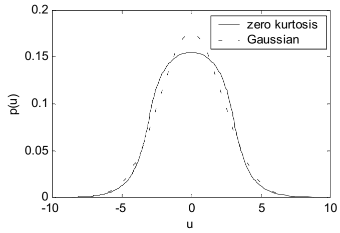
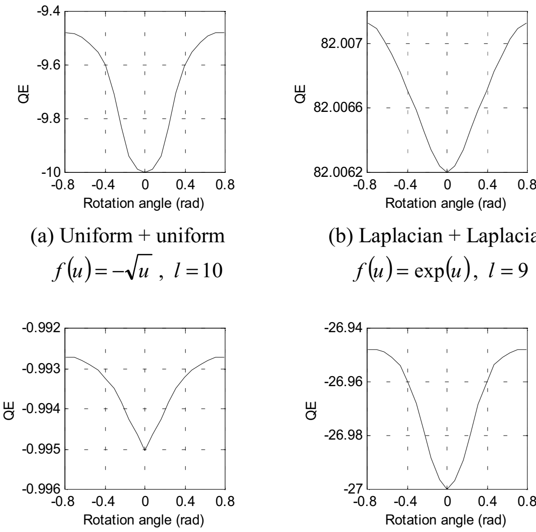


Fig. 2. A zero-kurtosis PDF.



(a) Uniform + uniform $f(u) = -\sqrt{u}$, $l = 10$

(b) Laplacian + Laplacian $f(u) = \exp(u)$, $l = 9$

(c) Sinusoid + Rayleigh $f(u) = -\ln(1+u)$, $l = 10$

(d) Zero-kurtosis + zero-kurtosis $f(u) = -\sqrt[4]{u}$, $l = 9$

Fig. 3. QE of rotational mixtures of two equal-variance sources versus the rotation angle.

Consider two continuous RVs r_1, r_2. Transform them by their individual CDFs respectively to get two RVs

$$z_1 \equiv q_{r_1}(r_1), \qquad z_2 \equiv q_{r_2}(r_2) \tag{4}$$

Clearly $(z_1, z_2) \in [0,1] \times [0,1]$. And from

$$p_{z_1 z_2}(u,v) = \frac{p_{r_1 r_2}(w,h)}{p_{r_1}(w) p_{r_2}(h)} \tag{5}$$

where $u = q_{r_1}(w)$ and $v = q_{r_2}(h)$, we know that

Lemma 1: r_1 and r_2 are independent iff $p_{z_1 z_2}(u,v) = 1$, $\forall (u,v) \in [0,1] \times [0,1]$.

Let l be an integer greater than 1. Define an l-level uniform quantization operator $D_l(\cdot)$ on $[0,1]$ as

$$D_l(u) = \begin{cases} \lceil ul \rceil & 0 < u \leq 1 \\ 1 & u = 0 \end{cases} \tag{6}$$

where $\lceil v \rceil$ denotes the least integer not less than v. Clearly $D_l(u) \in \{1, 2, \cdots, l\}$ for $u \in [0,1]$.

Now let us define discrete RVs

$$k_1 \equiv D_l(z_1), \qquad k_2 \equiv D_l(z_2) \tag{7}$$

Let $p_{k_1 k_2}(i,j) = \text{Prob}\,(k_1 = i \text{ and } k_2 = j)$ be the joint probability of k_1 and k_2 where $(i,j) \in \{1, \cdots, l\} \times \{1, \cdots, l\}$. Then the new independence measure, which is called *quasi-entropy* (QE) due to its some similarity to the entropy in information theory [4],

$$\beta(r_1, r_2) = \sum_{i=1}^{l} \sum_{j=1}^{l} f\left(p_{k_1 k_2}(i,j)\right) \tag{8}$$

where $f(\cdot)$ is a strictly convex function on $[0,1]$.

Due to the Jensen's inequality, we have

$$\beta(r_1, r_2) \geq l^2 f\left(1/l^2\right) \tag{9}$$

with equality iff (k_1, k_2) is uniform in $\{1, \cdots, l\} \times \{1, \cdots, l\}$.

On the other hand, if r_1 and r_2 are independent, then (z_1, z_2) is uniform in $[0,1] \times [0,1]$ (Lemma 1) and thus (k_1, k_2) is uniform in $\{1, \cdots, l\} \times \{1, \cdots, l\}$. Therefore we have

Lemma 2: If r_1 and r_2 are independent, then $\beta(r_1, r_2)$ reaches its minimum [R.H.S. of (9)].

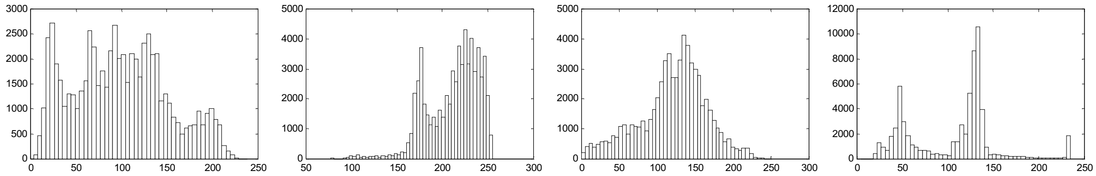
Indeed, we may also prove [2]

Lemma 3: If r_1 and r_2 are not independent, then there exists l_0 such that $\beta(r_1, r_2)$ cannot reach its minimum for any $l > l_0$.

Fig. 1 illustrates QE. Fig. 1 (a) and (b) are observed points of two independent RVs, s_1 and s_2, and two *de*pendent RVs, x_1 and x_2, respectively. Fig. 1 (c) and (d) plot observations of (z_1, z_2) and (z_1^*, z_2^*), respectively, where $z_1 \equiv q_{s_1}(s_1)$, $z_2 \equiv q_{s_2}(s_2)$, $z_1^* \equiv q_{x_1}(x_1)$, and $z_2^* \equiv q_{x_2}(x_2)$. The points in Fig. 1 (c) are uniformly distributed while those in Fig. 1 (d) are not, manifesting Lemma 1. According to Fig. 1 (c, e) and (d, f), $k_1 \equiv D_5(z_1)$, $k_2 \equiv D_5(z_2)$, $k_1^* \equiv D_5(z_1^*)$, $k_2^* \equiv D_5(z_2^*)$. The uniform $p_{k_1 k_2}$ and the nonuniform $p_{k_1^* k_2^*}$ are shown by grayscales in Fig. 1 (e) and (f) respectively. (The darker the color, the greater the probability.) Therefore, $\beta(s_1, s_2) = l^2 f\left(1/l^2\right) = 25 f(1/25) < \beta(x_1, x_2)$.

Since there are infinitely many strictly convex functions on $[0,1]$, QE is actually a class of infinitely many independence measures. Some of the common convex functions were listed in [4].

In practice, analytical forms of CDFs are usually unknown. But this does not matter. Using the definition of CDF, we may directly map samples of (r_1, r_2) to estimates of samples of (k_1, k_2) and no floating-point operation is involved. For details, see [3]. Then $p_{k_1 k_2}$, and thus QE, are easily estimated. Note the input range of the convex function $f(\cdot)$ is limited to $[0,1]$, so it is easily implemented by a look-up table. All these facilitate digital realization.



(a) Lena, kurtosis –0.7492 (b) Tiffany, kurtosis 0.7155 (c) couple, kurtosis 0.0231 (d) cla0, kurtosis 0.0668

Fig.4. Histograms and kurtoses of the source images.

(a) Source images, from left to right are Lena, Tiffany, couple, and cla0, respectively.

(b) Mixtures

(c) Recovered sources by QE-based algorithm with $f(u) = -\sqrt{u}$ and $l = 100$

(d) Output signals by fastICA

(e) Output signals by extended infomax

Fig. 5. Blind separation of image signals.

Tab. 1 SNRs and PSNRs of the recovered images by QE-based algorithms (l=100), fastICA, and extended infomax (dB)

$f(u)$	Lena		Tiffany		couple		cla0		Average	
	SNR	PSNR	SNR	PSNR	SNR	PSNR	SNR	PSNR	SNR	PSNR
$-\sqrt{u}$	24.9	36.7	19.9	31.0	15.0	28.8	30.7	41.7	22.6	34.6
$\exp(u)$	26.7	38.2	20.8	32.2	14.3	28.0	52.6	60.8	28.6	39.8
$-\ln(1+u)$	27.0	38.1	20.7	31.9	13.9	27.8	45.6	53.9	26.8	37.9
$-\sqrt[4]{u}$	26.4	37.7	19.1	30.4	16.0	29.7	28.4	39.3	22.5	34.3
fastICA	14.1	26.2	15.6	26.6	3.5	18.6	1.9	16.4	8.8	21.9
extended infomax	21.8	30.7	2.0	14.5	6.0	20.5	3.0	16.4	8.2	20.5

3. SIMULATIONS AND CONCLUSION

First we verify our results with some standard distributions. We mix two equal-variance independent sources s_1, s_2 by a 2×2 rotation matrix with rotation angle θ to get r_1, r_2 [3]. Thus we can plot QE of r_1, r_2 as a function of θ, i.e. $\beta(\theta)=\beta(r_1,r_2)$. Uniform, Laplacian, sinusoid, and Rayleigh sources are used. Besides, we generate a zero-kurtosis source like this: Suppose two independent sources s^+ and s^- with positive and negative kurtoses k^+ and k^- respectively, then the source $s=as^+ + bs^-$ will have zero kurtosis when the two coefficients satisfy $a^4/b^4=\left|k^-/k^+\right|$. If we choose Laplacian and Uniform sources as s^+ and s^- respectively, the generated zero-kurtosis distribution is as the solid curve in Fig. 2, which is apparently different from the Gaussian distribution with same variance plotted with dashed curve. Fig. 3 shows $\beta(\theta)$ using different combinations of the above sources and several different convex functions. All the curves reach minima at $\theta=0$, where the signals are separated and independent. If calculating the minima according to (9), we get –10, 82.0062, –0.9950, and –27 for Fig. 3 (a) through (d) respectively, which are just the minima in the plots.

The pairwise iterative structure of the BSS algorithms based on QE is the same as that based on the *grid occupancy* (GO) [4], which was described in [3], only that we optimize QE instead of GO. We do not list it here due to lack of space. We show the efficacy of the QE-based algorithms by an image separation example. Fig. 4 shows the histograms and kurtoses of the four 256×256 images with negative, positive, and two near-zero kurtoses, respectively. Fig. 5 shows the results. Fig. 5 (a) are the source images, (b) the mixtures obtained using a 4×4 mixing matrix, and (c) the separated images by our algorithm with $f(u)=-\sqrt{u}$ and $l=100$. Comparisons are made with two widely used algorithms, fastICA [5] and extended infomax [6], output signals by which are shown in Fig. 5 (d) and (e) respectively. The kurtosis-based fastICA cannot separate the last two near-zero kurtosis images and the effect of extended infomax is not good for the last three images. Tab. 1 summarizes the SNRs and *peak SNRs* (PSNRs) of our algorithms with several convex functions and those of the two other algorithms.

Moreover, the robustness of QE-based algorithms has been fully tested with mixtures of 20 sources as those in [6].

Finally we mention some related works. In [4], the new independence measures GO, QE, and *generalized mutual information* (GMI) were briefly summarized. The novel BSS algorithm based on GO was proposed in [3]. Interestingly, the expectation of GO is also a kind of QE [4]. A recursive algorithm for computing GMI, which includes the famous one for computing mutual information (MI) [7] as a special case, was developed in [4]. All these new measures are good criteria beyond the classical MI for choosing delay in strange attractor reconstruction [4].

REFERENCES

[1] J. -F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE,* vol. 86, no. 10, pp. 2009-2025, 1998.

[2] Y. Chen, "On theory and methods for blind information extraction," Ph.D. dissertation, Southeast University, China, Mar. 2001 (in Chinese).

[3] Y. Chen, Z. He, and L. Yang, "A novel grid-box-counting method for blind separation of sources," in *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing (ISIMP 2001),* pp. 243-246, Hong Kong, May 2-4, 2001.

[4] Y. Chen, "New methods for measuring independence with an application to chaotic systems," in *NOLTA 2002,* pp. 523-526, Xi'an, PRC, Oct. 7-11, 2002.

[5] A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.,* vol. 9, pp. 1483-1492, 1997.

[6] T. -W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Computation,* vol. 11, pp. 417-441, 1999.

[7] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A,* vol. 33, no. 2, pp. 1134-1140, 1986.