

NORMALIZED INFORMATION THEORETIC CRITERIA FOR BLIND SIGNAL EXTRACTION

Sergio Cruces and Iván Durán

Teoría de la Señal y Comunicaciones
University of Seville
41092-Seville, SPAIN
sergio@us.es

ABSTRACT

In this paper we present new normalized criteria for the extraction of the scaled sources whose density have the minimum support measure or the minimum entropy. Both criteria are part of a more general entropy minimization principle based on Renyi's entropies. However, the proposed approach (based on Renyi's entropies or orders zero and one) have some special advantages, which allow to relax the assumption of having identically distributed source signals.

1. INTRODUCTION

In the last decade, independent component analysis (ICA) has been revealed as an important tool in signal processing, data mining, biomedicine and digital communications. The ICA model considers the observations as a mixture of a certain number of underlying independent components called sources.

An elegant presentation of the ICA criteria has been done in terms of existing information theoretical and statistical results. Good examples of this methodology in approaching to the problem were the manuscript of Comon [1], which formally established the definition of ICA in information theoretic terms, and the classic manuscript of Donoho [2], which did a similar work with the minimum entropy criterion for blind deconvolution. More recently, in [3], a criterion based on the statistical range of the outputs was proposed for blind signal separation. In [4] we presented an extension of this criterion, to the case of blind signal extraction of one signal, based on the support of its density, or on the convex-hull of the support. The convex-hull of the support was also used as criterion in [5] where the authors showed that it does not have local deceptive solutions.

In the present paper we will obtain novel practical normalizations of the minimum entropy and minimum support criteria for blind signal extraction. We will also show how the normalization may influence and favor the extraction of certain sources.

Part of this research was supported by the MCYT Spanish project TEC2004-06451-C05-03.

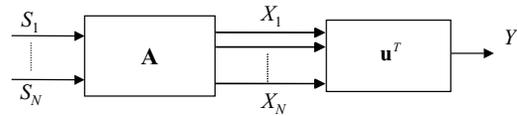


Fig. 1. Model for blind signal extraction.

2. SIGNAL MODEL AND NOTATION

Let us first consider the standard linear mixing model in a noiseless situation. Consider N mutually independent sources grouped in the random vector $\mathbf{S} = [S_1, \dots, S_N]^T$. The probability density function (p.d.f.) of the i -th source is denoted by $f_{S_i}(s_i)$ and its support is given by

$$\mathcal{S}_{S_i} = \{s_i : f_{S_i}(s_i) > 0\}. \quad (1)$$

The Renyi's differential entropy of order r of the S_i is

$$h_r(S_i) = \frac{r}{1-r} \log \|f_{S_i}\|_r, \quad r \in \{(0, 1) \cup (1, \infty)\}. \quad (2)$$

where $\|\cdot\|_r$ denotes the r -th order norm of the density. These entropies can be extended by continuity to the cases

$$h_r(S_i) = \begin{cases} \log(\mu\{\mathcal{S}_{S_i}\}) & \text{for } r = 0, \\ -\int f_{S_i}(s_i) \log f_{S_i}(s_i) dS_i & \text{for } r = 1, \end{cases} \quad (3)$$

where $\mu\{\cdot\}$ denotes the Lebesgue measure of the support set of the density.

The random vector observations $\mathbf{X} = [X_1, \dots, X_N]^T$ follows a linear relationship with the vector of sources

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \quad (4)$$

where $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_N] \in \mathbf{R}^{N \times N}$ is the mixing matrix.

In the linear model (4), the Darmois-Skitovitch theorem guarantees the identifiability of the original non-Gaussian sources from the observations up to a permutation and scaling of them [1]. The scaling and permutation indeterminacy in ICA means that, for any non-singular diagonal matrix

$$\mathbf{D} = \text{diag}(d_1, \dots, d_N), \quad (5)$$

and for any permutation matrix $\mathbf{P} \in \mathbf{R}^{N \times N}$, any transformation of the form $\mathbf{S} \rightarrow (\mathbf{P}^{-1}\mathbf{D}^{-1}\mathbf{S})$ preserves the independence of the resulting components while still verifying the linear decomposition of the observations

$$\mathbf{X} = \mathbf{A}'\mathbf{S}' = (\mathbf{ADP})(\mathbf{P}^{-1}\mathbf{D}^{-1}\mathbf{S}), \quad (6)$$

In order to extract one non-Gaussian source from the mixture, one can compute the inner product of the observations with the vector \mathbf{u} , to obtain the output random variable or estimated source

$$Y = \mathbf{u}^T \mathbf{X} = \mathbf{g}^T \mathbf{S}, \quad (7)$$

where $\mathbf{g}^T = \mathbf{u}^T \mathbf{A}$ denotes the global mixture from the sources to the output. Note that the i th source will be extracted when \mathbf{u} be proportional to one of the columns of the inverse transpose of the mixing system $\mathbf{A}^{-T} = [\mathbf{a}_1^-, \dots, \mathbf{a}_N^-]$.

3. CONVOLUTION INEQUALITIES AND RENYI'S ENTROPIES

We will assume in this section, without loss of generality, that there are two sources S_1 and S_2 whose densities have respective bounded norms $f_{S_1} \in L^p(\mathbf{R})$ and $f_{S_2} \in L^q(\mathbf{R})$. The p.d.f. of their mixture $Y = g_1 S_1 + g_2 S_2$ is given by the convolution of the two original densities

$$f_{g_1 S_1} * f_{g_2 S_2}(y) = \int f_{g_1 S_1}(y - \tau) f_{g_2 S_2}(\tau) d\tau. \quad (8)$$

The classical convolution inequality by Young relates the norm of the three involved densities

$$\|f_{g_1 S_1} * f_{g_2 S_2}\|_r \leq \|f_{g_1 S_1}\|_p \|f_{g_2 S_2}\|_q, \quad (9)$$

where $p, q, r \geq 1$ and $\frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1$. A reverse form of this inequality also exist for $0 < p, q, r \leq 1$ and $\frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1$

$$\|f_{g_1 S_1} * f_{g_2 S_2}\|_r \geq \|f_{g_1 S_1}\|_p \|f_{g_2 S_2}\|_q. \quad (10)$$

Noting that $\|f_{g_i S_i}\|_1 = 1$ and attending to the sign of $\frac{r}{(1-r)}$, for $r = p > 0$ and $q = 1$ the classical Young's inequality and its reverse form can both be rewritten as

$$\|f_{g_1 S_1} * f_{g_2 S_2}\|_{r^{\frac{r}{(1-r)}}} \geq \max_i \|f_{g_i S_i}\|_{r^{\frac{r}{(1-r)}}}, \quad (11)$$

Thus, after taking logarithms, one can express the result in terms of Renyi's entropies

$$h_r(Y) \geq \max_i h_r(g_i S_i). \quad (12)$$

Then, when assuming i.i.d. sources, one obtains a lower bound for the entropy of the output

$$h_r(Y) \geq h_r(\|\mathbf{g}\|_\infty S). \quad (13)$$

Two recent papers [6, 7] independently have arrived to this expression, which can be used as criterion for the blind extraction of i.i.d. sources. However, the resulting criterion depends on the infinite norm of the output, which is not easy to control since one does not have direct access to \mathbf{g} .

In the late 1970's several mathematics improved inequalities (9) and (10) with sharp constants [8], as the following theorem shows.

Theorem 1 (Strengthened convolution inequalities)

Let $p, q, r > 0$ satisfy $\frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1$, and let $f_{g_1 S_1} \in L^p(\mathbf{R})$ and $f_{g_2 S_2} \in L^q(\mathbf{R})$ be non-negative functions. Then, for $p, q, r \geq 1$

$$\|f_{g_1 S_1} * f_{g_2 S_2}\|_r \leq C^{1/2} \|f_{g_1 S_1}\|_p \|f_{g_2 S_2}\|_q, \quad (14)$$

and for $0 < p, q, r \leq 1$

$$\|f_{g_1 S_1} * f_{g_2 S_2}\|_r \geq C^{1/2} \|f_{g_1 S_1}\|_p \|f_{g_2 S_2}\|_q, \quad (15)$$

where $C = C_p C_q / C_r$ and

$$C_\eta = \frac{|\eta|^{1/\eta}}{|\eta'|^{1/\eta'}}, \quad (16)$$

for $1/\eta + 1/\eta' = 1$.

These inequalities are interesting because they bring a common way to proof the entropy power and Brunn-Minkowski inequalities. Choosing $p = \frac{r}{r+\lambda(1-r)}$ and $q = \frac{r}{r+(1-\lambda)(1-r)}$ where

$$\lambda = \frac{e^{(1+r)h_r(g_1 S_1)}}{e^{(1+r)h_r(g_1 S_1)} + e^{(1+r)h_r(g_2 S_2)}}, \quad (17)$$

and taking each of the following limits $r \rightarrow 0$ and $r \rightarrow 1$, by the continuity of the entropies, the super-additivity of the following function is obtained [9]

$$e^{(1+r)h_r(Y)} \geq e^{(1+r)h_r(g_1 S_1)} + e^{(1+r)h_r(g_2 S_2)}, \quad (18)$$

for $r \in \{0, 1\}$. Note that these are the two only possible values of the order r which can satisfy $r = p = q$ and simultaneously $\frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1$.

4. NORMALIZED MINIMUM ENTROPY AND MINIMUM SUPPORT CRITERIA

The desired blind criteria for the extraction of one source are grounded on the result of next lemma, which lower-bounds entropy of order zero and of order one of the output in terms of the respective entropies of the sources.

Lemma 1 (Entropy's lower-bound) Let $Y = \mathbf{g}^T \mathbf{S}$, then for $r \in \{0, 1\}$, $\forall \beta \geq (1+r)$, a lower bound of the r -order entropy of the output is given by

$$h_r(Y) \geq \log \|\mathbf{Dg}\|_\beta + \sum_{j=1}^N \left| \frac{g_j d_j}{\|\mathbf{Dg}\|_\beta} \right|^\beta h_r(d_j^{-1} S_j) \quad (19)$$

The tighter lower-bound is obtained for $\beta = 1 + r$, while the loosest bound is for $\beta = \infty$.

Proof: The outline of the proof of this result is based on the following chain of inequalities

$$\begin{aligned} h_r(Y) &\stackrel{(a)}{\geq} \frac{1}{1+r} \log \sum_{j=1}^N e^{(1+r)h_r((g_j d_j)(d_j^{-1} S_j))}, \quad r \in \{0, 1\} \\ &\stackrel{(b)}{\geq} \frac{1}{\beta} \log \sum_{j=1}^N |g_j d_j|^\beta e^{\beta h_r(d_j^{-1} S_j)}, \quad \beta \geq (1+r) \\ &\stackrel{(c)}{\geq} \log \|\mathbf{Dg}\|_\beta + \sum_{j=1}^N \left| \frac{g_j d_j}{\|\mathbf{Dg}\|_\beta} \right|^\beta h_r(d_j^{-1} S_j) \end{aligned}$$

The idea of the proof consist in exploiting: a) the super-additivity of the function $\exp((1+r)h_r(Y))$ for the considered orders, b) the convexity of the power function for exponents greater or equal to the unity, c) the strict concavity of the logarithm. \square

The normalization of the inequality in the previous lemma is obtained by subtracting the term $\log \|\mathbf{Dg}\|_\beta$ from both sides. In this way, the left-hand-part of the equation becomes scale invariant. However, since the β -norm of the vector \mathbf{Dg} is unknown for us, in practice, we need to upper-bound it in a tight way by another normalizing function which depends only on the available information: the considered extraction system \mathbf{u} and the vector of observations \mathbf{X} . This is what the next theorem proposes.

Theorem 2 (Normalization) Assume that for a given function $F(\mathbf{u}, \mathbf{X})$ the values of the scaling coefficients are assigned in the following manner

$$d_j = \begin{cases} \epsilon & \text{if } F(\mathbf{a}_j^-, \mathbf{X}) = 0, \\ F(\mathbf{a}_j^-, \mathbf{X}) & \text{otherwise.} \end{cases} \quad (20)$$

where $j = 1, \dots, N$ and $\epsilon \neq 0$ is arbitrary small. Define set of indices Ω'_r as

$$\Omega'_r = \{i : i = \arg \min_{j=1, \dots, N} h_r(d_j^{-1} S_j)\}. \quad (21)$$

If the normalizing function satisfies

$$\begin{cases} F(\mathbf{u}, \mathbf{X}) = \|\mathbf{Dg}\|_\beta & \text{if } \mathbf{u} \propto \mathbf{a}_i^-, i \in \Omega'_r, \\ F(\mathbf{u}, \mathbf{X}) \leq \|\mathbf{Dg}\|_\beta & \text{otherwise.} \end{cases} \quad (22)$$

then, for $r \in \{0, 1\}$ and for a properly chosen norm $\beta \geq 1+r$, the following inequality applies

$$h_r(Y) - \log F(\mathbf{u}, \mathbf{X}) \geq \sum_{j=1}^N \left| \frac{g_j d_j}{\|\mathbf{Dg}\|_\beta} \right|^\beta h_r(d_j^{-1} S_j) \quad (23)$$

Table 1 presents some normalizing functions $F(\mathbf{u}, \mathbf{X})$ which satisfy the conditions of the theorem. The minimization of the right-hand-side of (23), with respect to the vector \mathbf{g} , yields two interesting corollaries (whose proof is omitted).

Table 1. Some of the normalizing functions satisfying the conditions of equation (22).

	$F(\mathbf{u}, \mathbf{X})$	$F(\mathbf{a}_i^-, \mathbf{X})$
$r = 1, \beta = 2$	σ_Y	σ_{S_i}
$r = 1, \beta > 2$	$ Cum_\beta(Y) ^{\frac{1}{\beta}}$	$ Cum_\beta(S_i) ^{\frac{1}{\beta}}$
$r = 0, \beta = 1$	$E[Y]$	$E[S_i]$
$r = 0, \beta = 1$	$ E[Y] $	$ E[S_i] $
$r = 0, \beta > 1$	$ Cum_\beta(Y) ^{\frac{1}{\beta}}$	$ Cum_\beta(S_i) ^{\frac{1}{\beta}}$
$r = 0, \beta \geq 1$	$\ \mathbf{u}\ _\beta$	$\ \mathbf{a}_i^-\ _\beta$

Corollary 1 (Normalized minimum entropy criterion)

When $F(\mathbf{u}, \mathbf{X})$ satisfies the conditions of equation (22) for $\beta \geq 2$, the following inequality holds true

$$h_1(Y) - \log F(\mathbf{u}, \mathbf{X}) \geq h_1(S_i / F(\mathbf{a}_i^-, \mathbf{X})), \quad i \in \Omega'_1. \quad (24)$$

Moreover, the minimum of the left-hand-side of the inequality is only reached at the extraction of any of the scaled sources of minimum entropy, i.e., when $i \in \Omega'_1$, otherwise the inequality is strict.

When the normalizing function is equal to the standard deviation of the output $F(\mathbf{u}, \mathbf{X}) = \sigma_Y$, the previous corollary reduces to the well known minimum entropy criterion for extraction (see [2, 10] and references therein). However, other normalizations based on higher order cumulants, like the one presented Table 1, could result preferable when there is additive Gaussian noise in the mixture.

Corollary 2 (Normalized minimum support criterion)

When $F(\mathbf{u}, \mathbf{X})$ satisfies the conditions of equation (22) for $\beta \geq 1$, the following inequality holds true

$$h_0(Y) - \log F(\mathbf{u}, \mathbf{X}) \geq h_0(S_i / F(\mathbf{a}_i^-, \mathbf{X})), \quad i \in \Omega'_0. \quad (25)$$

When $\beta > 1$, or when $\beta = 1$ and the cardinality of the set $|\Omega'_0| = 1$, the minimum of the left-hand-side of the inequality is only reached at the extraction of any of the scaled sources of minimum support.

Some of the normalized criteria exhibit robustness to certain kinds of noise and outliers, but this analysis will not be considered here for reasons of space.

5. SIMULATIONS

In this section we illustrate the contrast nature of the proposed criteria with a very simple example of extraction. Let $\text{rect}(t)$ denote the square pulse, and let $\text{tri}(t) = \text{rect}(t) * \text{rect}(t)$ denote a triangular pulse, both of unit area and centered at the origin. We consider to two independent sources whose histograms are shown in figure 2. The samples of first

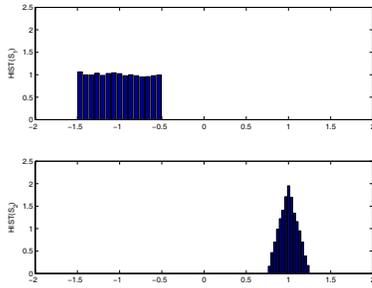


Fig. 2. Histograms of the considered sources.

source were drawn from a uniform random variable of density $f_{S_1}(s_1) = \text{rect}(s_1 + 3/2)$, whereas the samples of the second source were drawn from a triangular random variable of density $f_{S_2}(s_2) = \text{tri}(4s_2 - 4)$. The observations were a mixture of the sources by the matrix $A = [1, -1; 1, 1]$. The extraction vector $\mathbf{u} = [u_1, u_2]^T$ was parametrized in terms of its angle with the axis of abscissas. For the previous mixture, an angle of $\pi/4$ extracts the first source, whereas an angle of $-\pi/4$ extracts the second source.

Figure 3 presents different normalized criteria for the zero-order entropy. From the figure, one can observe that, in all cases, the minimum of the proposed criteria are obtained at the solutions that extract one of the sources. The example also shows how the choice of $F(\mathbf{u}, \mathbf{X})$ gives a certain freedom that can be used to favor the extraction of sources with certain stochastic properties. In the example, the criteria with cumulants of odd order favor the extraction of the second source, whereas the criteria with cumulants of even order favor the extraction of the first source. This can be explained from the fact that the sample cumulants of odd order have a similar value for the two sources, so they do not influence the order of preference in the extraction. For cumulants of even order the situation is different. The sample cumulants of even order of the first source are much more important than those of the second source, thus, influencing the preferences of extraction.

6. CONCLUSIONS

In this paper we have presented criteria for the blind extraction of the independent components with minimum normalized support or with minimum normalized entropy. These normalizations can be used to alter the preferred order of extraction and to improve the robustness of the criteria to the presence of certain kinds of additive noise. This robustness was not analyzed in this paper and it will be presented elsewhere.

7. REFERENCES

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

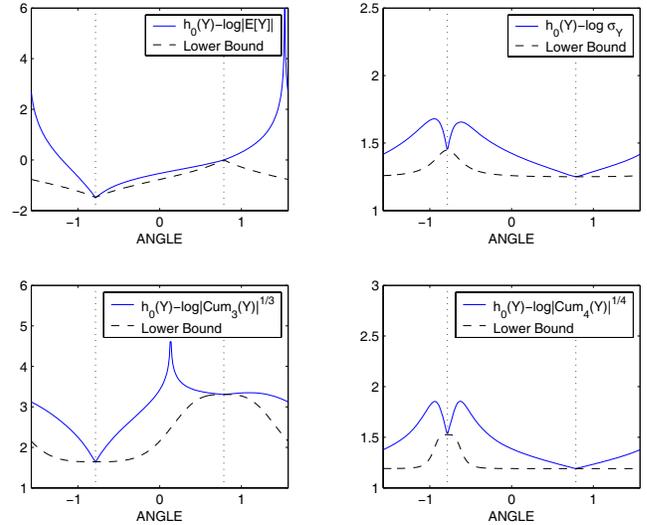


Fig. 3. Zero-order entropy criterion for different normalizations and its lower bound of eq. (23). The positions of the two solutions, the angles $-\pi/4$ and $\pi/4$ of the vector \mathbf{u} that extract one of the sources, are indicated with dotted lines.

[2] D. Donoho, *On minimum entropy deconvolution*, Applied Time Series Analysis II, D. F. Findley Editor, Academic Press, New York, pp. 565–608, 1981.

[3] D. T. Pham, "Blind separation of instantaneous mixture of sources based on order statistics", *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 363–375, 2000.

[4] S. Cruces, I. Durán, "The minimum support criterion for blind signal extraction," *proc. of the int. conf. on Independent Component Analysis and Blind Signal Separation*, Granada, Spain, pp. 57–64, 2004.

[5] F. Vrins, M. Verleysen, C. Jutten, "SWM : A class of convex contrasts for source separation" *proc. of the Int. Conf. on Acoustics, Speech and Signal Processing*, Philadelphia (USA), vol. V, pp. 161–164, 2005.

[6] J.-F. Bercher, C. Vignat, "A Renyi entropy convolution inequality with application", in *Proc. of EUSIPCO*, Toulouse, France, 2002.

[7] D. Erdogmus, J. C. Principe, L. Vielva, "Blind deconvolution with minimum Renyi's entropy", in *Proc. of EUSIPCO*, vol. 2, pp. 71–74, Toulouse, France, 2002.

[8] R. J. Gardner, "The Brunn-Minkowski Inequality", *Bulletin of the American Mathematical Society*, vol. 39(3), pp. 355–405, 2002.

[9] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, Wiley series in telecommunications. John Wiley, 1991.

[10] S. Cruces, A. Cichocki, and S-i. Amari, "From blind signal extraction to blind instantaneous signal separation: criteria, algorithms and stability", *IEEE Trans. on Neural Networks*, vol 15, no. 4, pp. 859–873, July 2004.