Is the General Form of Renyi's Entropy a Contrast for Source Separation?

Frédéric Vrins[1], Dinh-Tuan Pham[2], and Michel Verleysen[1]

[1] Machine Learning Group
Université catholique de Louvain
Louvain-la-Neuve, Belgium
{vrins,verleysen}@dice.ucl.ac.be
[2] Laboratoire de Modélisation et Calcul
Centre National de la Recherche Scientifique
Grenoble, France
dinh-tuan.pham@imag.fr

Abstract. Renyi's entropy-based criterion has been proposed as an objective function for independent component analysis because of its relationship with Shannon's entropy and its computational advantages in specific cases. These criteria were suggested based on "convincing" experiments. However, there is no theoretical proof that globally maximizing those functions would lead to separate the sources; actually, this was implicitly conjectured. In this paper, the problem is tackled in a theoretical way; it is shown that globally maximizing the Renyi's entropy-based criterion, in its general form, does not necessarily provide the expected independent signals. The contrast function property of the corresponding criteria simultaneously depend on the value of the Renyi parameter, and on the (unknown) source densities.

1 Introduction

Blind source separation (BSS) aims at recovering underlying source signals from mixture of them. Under mild assumptions, including the mutual independence between those sources, it is known from Comon [1] that finding the linear transformation that minimizes a dependence measure between outputs can solve the problem, up to acceptable indeterminacies on the sources. This procedure is known as Independent Component Analysis (ICA).

This problem can be mathematically expressed in a very simple way. Consider the square, noiseless BSS mixture model: a K-dimensional vector of independent unknown sources $\mathbf{S} = [S_1, \ldots, S_K]^{\mathrm{T}}$ is observed via an instantaneous linear mixture of them $\mathbf{X} = \mathbf{AS}$, $\mathbf{X} = [X_1, \ldots, X_K]^{\mathrm{T}}$, where $\mathbf{A}$ is the full-rank square mixing matrix. Many separation methods are based on the maximization (or minimization) of a criterion. A specific class of separation criteria is called "contrast functions" [1]. The contrast property ensures that a given criterion is suitable to achieve BSS. Such a function i) is scale invariant, ii) only depends on the demixing matrix B and of the mixture densities iii) reaches its global maximum if and only if the transfer matrix $\mathbf{W} = \mathbf{BA}$ is non-mixing [1]. A matrix $\mathbf{W}$

M.E. Davies et al. (Eds.): ICA 2007, LNCS 4666, pp. 129–136, 2007.

is said non-mixing if it belongs to the subgroup $\mathcal{W}$ of the general linear group $\mathcal{GL}(K)$ of degree K, and is defined as:

$$\mathcal{W} \doteq \{\mathbf{W} \in \mathcal{GL}(K) : \exists \mathbf{P} \in \mathcal{P}^{\mathbf{K}},\ \boldsymbol{\Lambda} \in \mathcal{D}^{\mathbf{K}}, \mathbf{W} = \mathbf{P}\boldsymbol{\Lambda}\} \tag{1}$$

In the above definition $\mathcal{P}^K$ and $\mathcal{D}^K$ respectively denote the groups of permutation matrices and of regular diagonal matrices of degree K.

Many contrast functions have been proposed in the literature. One of the most known contrast function is the opposite of mutual information $I(\mathbf{Y})$ [2] where $\mathbf{Y} = \mathbf{BX}$, which can be equivalently written as a sum of differential entropies $h(.)$:

$$I(\mathbf{Y}) \doteq \sum_{i=1}^{K} h(Y_i) - h(\mathbf{Y}) = \sum_{i=1}^{K} h(Y_i) - \log|\det \mathbf{B}| - h(\mathbf{X}). \tag{2}$$

The differential (Shannon) entropy of $X \sim p_X$ is defined by

$$h(X) \doteq -\mathrm{E}[\log p_X]. \tag{3}$$

Since $I(\mathbf{Y})$ has to be minimized with respect to $\mathbf{B}$, its minimization is equivalent to the following optimization problem under a prewhitening step:

$$\max_{\mathbf{B}\in\ \mathcal{SO}(K)} C(\mathbf{B}),\quad C(\mathbf{B}) \doteq -\sum_{i=1}^{K} h(\mathbf{b}_i\mathbf{X}), \qquad \text{problem 1} \tag{4}$$

where $\mathbf{b}_i$ denotes the i-th row of $\mathbf{B}$ and $\mathcal{SO}(K)$ is the special orthogonal group

$$\mathcal{SO}(K) \doteq \{\mathbf{W} \in \mathcal{GL}(K) : \mathbf{W}\mathbf{W}^T = \mathbf{I}_K,\ \det \mathbf{W} = +1\}$$

with $\mathbf{I}_K$ the identity matrix of degree K. The $\mathbf{B} \in \mathcal{SO}(K)$ restriction, yielding $\log|\det \mathbf{B}| = 0$, results from the fact that, without loss of generality, the source can be assumed to be centered and unit-variance ($\mathrm{E}[\mathbf{S}\mathbf{S}^{\mathrm{T}}] = \mathbf{I}_K$ and $\mathbf{A} \in \mathcal{SO}(K)$ if the mixtures are whitened [8]). Clearly, if $\mathbf{W} = \mathbf{BA}$, problem 1 is equivalent to problem 2:

$$\max_{\mathbf{W}\in\ \mathcal{SO}(K)} \widetilde{C}(\mathbf{W}),\quad \widetilde{C}(\mathbf{W}) \doteq -\sum_{i=1}^{K} h(\mathbf{w}_i\mathbf{S}). \qquad \text{problem 2}$$

Few years ago, it has been suggested to replace Shannon's entropy by Renyi's entropy [4,5]. More recent works still focus on that topic (see e.g. [7]). Renyi's entropy is a generalization of Shannon's one in the sense that they share the same key properties of information measures [10]. The Renyi entropy of index $r \geq 0$ is defined as:

$$h_r(X) \doteq \frac{1}{1-r} \log \int_{\Omega(X)} p_X^r(x)dx, \tag{5}$$

where $r \geq 0$ and $\lim_{r\to 1} h_r(X) = h_1(X) = h(X)$ and $\Omega(X) \doteq \{x : p_X(x) > 0\}$. Based on simulation results, some researchers have proposed to modify the above BSS contrast $C(\mathbf{B})$ defined in problem 1 by the following modified criterion

$$C_r(\mathbf{B}) \doteq -\sum_{i=1}^{K} h_r(Y_i), \tag{6}$$

assuming implicitly that the contrast property of $C_r(\mathbf{B})$ is preserved even for $r \neq 1$. This is clearly the case for the specific values $r = 1$ (because obviously $C_1(\mathbf{B}) = C(\mathbf{B})$) and $r = 0$ (under mild conditions); this can be easily shown using the Entropy Power and the Brunn-Minkowski inequalities [3], respectively. However, there is no formal proof that the contrast property of $C_r(\mathbf{B})$ still holds for other values of r.

In order to check if this property may be lost in some cases, we restrict ourselves to see if a necessary condition ensuring that $C_r(\mathbf{B})$ is a contrast function is met. More specifically, the criterion $\widetilde{C}_r(\mathbf{W})$ should admit a local maximum when $\mathbf{W} \in \mathcal{W}$. To see if this condition is fulfilled, a second order Taylor development of $\widetilde{C}_r(\mathbf{W})$ is provided around a non-mixing point $\mathbf{W}^\star \in \mathcal{W}$ in the next section. For the sake of simplicity, we further assume $K = 2$ and that a prewhitening is performed so that we shall constraint $\mathbf{W} \in \mathcal{SO}(2)$ since it is sufficient for our purposes, as shown in the example of Section 3 (the extension to $K \geq 3$ is easy).

2 Taylor Development of Renyi's Entropy

Setting $K = 2$, we shall study the variation of the criterion $\widetilde{C}_r(\mathbf{W})$ due to a slight deviation of $\mathbf{W}$ from any $\mathbf{W}^\star \in \mathcal{W} \cap \mathcal{SO}(2)$ of the form $\mathbf{W} \leftarrow \mathcal{E}\mathbf{W}^\star$ where

$$\mathcal{E} \doteq \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{7}$$

and $\theta \simeq 0$ is a small angle. This kind of updates covers the neighborhood of $\mathbf{W}^\star \in \mathcal{SO}(K)$: if $\mathbf{W}, \mathbf{W}^\star \in \mathcal{SO}(2)$, there always exists $\boldsymbol{\Phi} \in \mathcal{SO}(2)$ such that $\mathbf{W} = \boldsymbol{\Phi}\mathbf{W}^\star$; $\boldsymbol{\Phi}$ can be written as $\mathcal{E}$ and if $\mathbf{W}$ is further restricted to be in the neighborhood of $\mathbf{W}^\star$, θ must be small enough. In order to achieve that aim, let us first focus on a first order expansion of the criterion, to analyse if non-mixing matrices are stationary points of the criterion. This is a obviously a necessary condition for $C_r(\mathbf{B})$ to be a contrast function.

2.1 First Order Expansion: Stationarity of Non-mixing Points

Let Z be a random variable independent from Y. From the definition of Renyi's entropy given in eq. (5), it comes that Renyi's entropy of $Y + \epsilon Z$ is

$$h_r(Y + \epsilon Z) = \frac{1}{1-r} \log \int p^r_{Y+\epsilon Z}(x)dx, \tag{8}$$

where the density $p_{Y+\epsilon Z}$ reduces to, up to first order in ϵ [9]:

$$p_{Y+\epsilon Z}(y) = p_Y(y) - \epsilon \left. \frac{\partial \mathrm{E}[(Z|Y=x)p_Y(x)]}{\partial x} \right|_{x=y} + o(\epsilon). \tag{9}$$

Therefore, we have:

$$p^r_{Y+\epsilon Z}(y) = p^r_Y(y) - r\epsilon p^{r-1}_Y(y) \left.\frac{\partial[\mathrm{E}(Z|Y=x)p_Y(x)]}{\partial x}\right|_{x=y} + \phi(\epsilon, y), \tag{10}$$

where $\phi(\epsilon, y)$ is $o(\epsilon)$. Hence, noting that $\log(1+a) = a + o(a)$ as $a \to 0$, equations (8) and (9) yield[1]

$$h_r(Y+\epsilon Z) = h_r(Y) - \epsilon \frac{r}{1-r} \frac{\int p^{r-1}_Y(y)[\mathrm{E}(Z|Y)p_Y]'(y)dy}{\int p^r_Y(y)dy} + o(\epsilon). \tag{11}$$

But, by integration by parts, one gets

$$\frac{1}{r-1}\int p^{r-1}_Y(y)[\mathrm{E}(Z|Y)p_Y]'(y)dy = -\int p^{r-1}_Y(y)\mathrm{E}(Z|Y=y)p'_Y(y)dy, \tag{12}$$

yielding

$$-\epsilon \frac{r}{1-r} \frac{\int p^{r-1}_Y(y)[\mathrm{E}(Z|Y)p_Y]'(y)dy}{\int p^r_Y(y)dy} = -\epsilon r \frac{\int p^{r-1}_Y(y)\mathrm{E}(Z|Y=y)p'_Y(y)dy}{\int p^r_Y(y)dy}. \tag{13}$$

From the general iterated expectation lemma (p. 208 of [6]), the right-hand side of the above equality equals

$$-\epsilon r \frac{\mathrm{E}[p^{r-2}_Y(Y)p'_Y(Y)Z]}{\int p^r_Y(y)dy} = \epsilon \mathrm{E}[\psi_r(Y)Z], \tag{14}$$

if we define the r-score function $\psi_r(Y)$ of Y as

$$\psi_r(Y) \doteq -\frac{rp^{r-2}_Y(Y)p'_Y(Y)}{\int p^r_Y(y)dy} = -\frac{1}{p_Y(Y)}\frac{(p^r_Y)'(Y)}{\int p^r_Y(y)dy}. \tag{15}$$

Observe that the 1-score reduces to the score function of Y, defined as $-(\log p_Y)'$.

Then, using eq. (11), noting $\mathbf{Y} = \mathcal{E}\mathbf{W}^\star\mathbf{S}$, $\cos\theta = 1 + o(\theta)$ and $\sin\theta = \theta + o(\theta)$, the criterion $\widetilde{C}_r(\mathcal{E}\mathbf{W}^\star)$ becomes up to first order in θ:

$$\begin{aligned}\widetilde{C}_r(\mathcal{E}\mathbf{W}^\star) &= -h_r(Y_1) - h_r(Y_2) \\ &\approx \widetilde{C}_r(\mathbf{W}^\star) \pm \theta\Big\{\mathrm{E}[\psi_r(S_1)S_2] - \mathrm{E}[\psi_r(S_2)S_1]\Big\}.\end{aligned} \tag{16}$$

The sign of θ in the last equation depends on matrix $\mathbf{W}^\star$; for example, if $\mathbf{W}^\star = \mathbf{I}_2$, it is negative, and if the rows of $\mathbf{I}_2$ are permuted in the last definition of $\mathbf{W}^\star$, it is positive.

Remind that the criterion is not sensitive to a left multiplication of its argument by a scale and/or permutation matrix. For instance, $\widetilde{C}_r(\mathbf{W}^\star) = \widetilde{C}_r(\mathbf{I}_2) = -h_r(S_1) - h_r(S_2)$. It results that since independence implies non-linear decorrelation, both expectations vanish in eq. (16) and $\widetilde{C}_r(\mathbf{W})$ admits a stationary point whatever is $\mathbf{W}^\star \in \mathcal{W}$.

[1] Provided that there exist $\epsilon^\star > 0$ and an integrable function $\Phi(y) > 0$ such that for all $y \in \mathbb{R}$ and all $|\epsilon| < \epsilon^\star$, $\phi(\epsilon, y)/\epsilon < \Phi(y)$. It can be shown that this is indeed the case under mild regularity assumptions.

2.2 Second Order Expansion: Characterization of Non-mixing Points

Let us now characterize these stationary points. To this end, consider the second order expansion of $p_{Y+\epsilon Z}$ provided in [9] (Z is assumed to be zero-mean to simplify the algebra):

$$p_{Y+\epsilon Z} = p_Y + \frac{1}{2}\epsilon^2 \mathrm{E}(Z^2) p_Y'' + o(\epsilon^2). \tag{17}$$

Therefore, since Renyi's entropy is not sensitive to translation we have, for $r > 0$:

$$h_r(Y + \epsilon Z) = h_r(Y) + \frac{\epsilon^2}{2} \underbrace{\frac{r}{1-r} \frac{\int p_Y^{r-1}(y) p_Y''(y) dy}{\int p_Y^r(y) dy}}_{\doteq J_r(Y)} \mathrm{var(Z)} + \mathrm{o}(\epsilon^2), \tag{18}$$

where $J_r(Y)$ is called the r-th order information of Y. Observe that the first order information reduces to $J_1(Y) = \mathrm{E}[\psi_{Y,r}^2]$, i.e. to Fisher's information [2].

In order to study the "nature" of the stationary point reached at $\mathbf{W}^\star$ (minimum, maximum, saddle), we shall check the variation of $\widetilde{C}_r$ resulting from the update $\mathbf{W} \leftarrow \mathcal{E}\mathbf{W}^\star$ up to second order in θ. Clearly, $\cos\theta = 1 - \theta^2/2 + o(\theta^2)$ and $\tan\theta = \theta + o(\theta^2)$, the criterion then becomes:

$$\begin{aligned}
\widetilde{C}_r(\mathcal{E}\mathbf{W}^\star) &= -h_r(Y_1) - h_r(Y_2) \\
&= -h_r(S_1 + \tan\theta S_2) - h_r(S_2 - \tan\theta S_1) - 2\log|\cos\theta| \\
&= \widetilde{C}_r(\mathbf{W}^\star) - \frac{\theta^2}{2}\left[J_r(S_1)\mathrm{var(S_2)} + \mathrm{J_r(S_2)var(S_1)}\right] - 2\log|1 - \frac{\theta^2}{2}| + o(\theta^2) \\
&= \widetilde{C}_r(\mathbf{W}^\star) - \frac{\theta^2}{2}\left[J_r(S_1)\mathrm{var(S_2)} + \mathrm{J_r(S_2)var(S_1)} - 2\right] + o(\theta^2)
\end{aligned} \tag{19}$$

where we have used $H_r(\alpha Y) = H_r(Y) + \log|\alpha|$, for any real number $\alpha > 0$. This clearly shows that if the sources share a same density with variance var(S) and r-th order information $J_r(S)$, the sign of $\widetilde{C}_r(\mathcal{E}\mathbf{W}^\star) - C_r(\mathbf{W}^\star)$ equals, up to second order in θ to $\mathrm{sign}(1 - J_r(S)\mathrm{var(S)})$. In other words, the criterion reaches a local minimum at any $\mathbf{W} \in \mathcal{W}$ if $J_r(S_i)\mathrm{var(S_i)} < 1$, instead of an expected global maximum. In this specific case, maximizing the criterion does not yield the seeked sources.

3 Example

A necessary and sufficient condition for a scale invariant criterion $f(\mathbf{W})$, $\mathbf{W} \in \mathcal{SO}(K)$ to be an orthogonal contrast function is that the set of its global maximum points matches the set of the orthogonal non-mixing matrices, i.e. $\mathrm{argmax}_{\mathbf{W} \in \mathcal{SO}(K)} f(\mathbf{W}) = \mathcal{W}$. Hence, in the specific case where the two sources share the same density p_S, it is necessary that the criterion admits (at least) a local maximum at non-mixing matrices. Consequently, according to the results derived in the previous section, the $J_r(S)\mathrm{var(S)} < 1$ inequality implies that the sources cannot be recovered through the maximization of $C_r(\mathbf{B})$.

3.1 Theoretical Characterization of Non-mixing Stationary Points

The above analysis would be useless if the $J_r(S)\mathrm{var}(S) < 1$ inequality is never satisfied for non-Gaussian sources. Actually, it can be shown that simple and common non-Gaussian densities satisfies this inequality. This is e.g. the case of the triangular density. We assume that both sources S_1, S_2 share the same triangular density p_T[2]:

$$p_T(s) \doteq \begin{cases} \frac{1-|s/\sqrt{6}|}{\sqrt{6}} & \text{if } |s| \leq \sqrt{6} \\ 0 & \text{otherwise .} \end{cases} \tag{20}$$

Observe that $\mathrm{E}[\mathrm{S_i}] = 0$, $\mathrm{var}(\mathrm{S_i}) = 1$, $i \in \{1, 2\}$, and note that using integration by parts, the r-th order information can be rewritten as

$$J_r(Y) = r\frac{\int p_Y^{r-2}(y)[p'_Y(y)]^2 dy}{\int p_Y^r(y) dy}.$$

Then, for $S \in \{S_1, S_2\}$ and noting $u \doteq 1 - s/\sqrt{6}$:

$$J_r(S) = \frac{r}{6}\frac{\int_0^{\sqrt{6}}(1-s/\sqrt{6})^{r-2}ds}{\int_0^{\sqrt{6}}(1-s/\sqrt{6})^r ds} = \frac{r}{6}\frac{\int_0^1 u^{r-2}du}{\int_0^1 u^r du} = \begin{cases} r(r+1)/[6(r-1)] & \text{if } r > 1 \\ \infty & \text{if } r \leq 1 \end{cases}$$

Thus $J_r(S)\mathrm{var}(S) < 1$ if and only if $r(r+1)/[6(r-1)] < 1$. But for $r \geq 1$, the last inequality is equivalent to $0 > r(r+1) - 6(r-1) = (r-2)(r-3)$. Therefore $J_r(S)\mathrm{var}(S) < 1$ if and only if $2 < r < 3$, as shown in Figure 1(a). We conclude that for a pair of triangular sources, the criterion $C_r(\mathbf{B})$ is not a contrast for $2 < r < 3$.

3.2 Simulation

Let us note $\mathbf{Y} = [Y_1, Y_2]^T$, $\mathbf{Y} = \mathbf{W}_\theta\mathbf{S}$, where $\mathbf{W}_\theta$ is a 2D rotation matrix of angle θ of the same form of $\mathcal{E}$ but where θ can take arbitrary values $[0, \pi]$. The criterion $-(h_r(Y_1)+h_r(Y_2))$ is plotted with respect to the transfer angle θ. Obviously, the set of non-mixing points reduces to $\mathcal{W} = \{\mathbf{W}_\theta : \theta \in \{k\pi/2 | k \in \mathbb{Z}\}\}$. Drawing this figure requires some approximations, and we are aware about the fact that it does not constitute a proof of the violation of the contrast property by itself; this proof is provided in the above theoretical development where it is shown that out of any problem of e.g. density estimation or exact integration approximation, Renyi's entropy is not always a contrast for BSS. The purpose of this plot is, complementary to Section 2, to show that *in practice, too*, the use of Renyi's entropy with arbitrary value of r might be dangerous.

Figure 1(b) has been drawn as follows. For each angle $\theta \in [0, \pi]$, the exact triangular probability density function p_T is used to compute the pdf of $\sin\theta S$

[2] This density is piecewise differentiable and continuous. Therefore, even if the density expansions are not valid everywhere, eq. (19) is still of use.

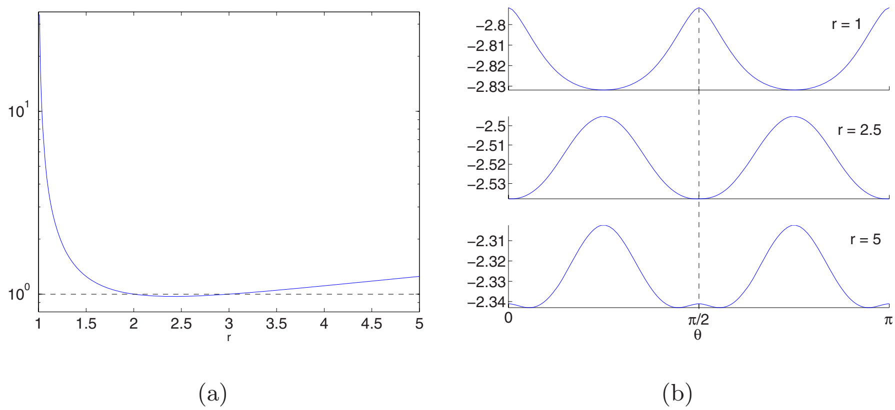

Fig. 1. Triangular sources. (a): $\log(J_r(S)\text{var(S)})$ vs r. (b): Estimated Renyi's criterion $\widetilde{C}_r(\mathcal{E})$ vs θ. The criterion is not a contrast function for $r = 2.5$ and $r = 5$.

and $\cos\theta S$, $S \sim p_T$, by using the well-known formula of the pdf of a transformation of random variables. Then, the output pdfs are obtained by convoluting the independent sources scaled by $\sin\theta$ and $\cos\theta$. Finally, Renyi's entropy is computed by replacing exact integration by Riemannian summation restricted on points were the output density is larger than $\tau = 10^{-4}$ to avoid numerical problems resulting from the log operator. At each step, it is checked that the pdfs of $\sin\theta S$, $\cos\theta S$, Y_1 and Y_2 integrate to one with an error smaller than τ and that the variance of the outputs deviates from unity with an error smaller than τ. Note that at non-mixing points, the exact density p_T is used as the output pdf to avoid numerical problems.

The two last plots of Figure 1(b) clearly indicate that the problem could be emphasized even when dealing with an approximated form of Renyi's entropy. On the top of the figure ($r = 1$), the criterion $\widetilde{C}_r(\mathbf{W}) = \widetilde{C}(\mathbf{W})$ (or more precisely, $C_r(\mathbf{B}) = C(\mathbf{B})$) is a contrast function, as expected. On the middle plot ($r = 2.5$), $\widetilde{C}_r(\mathbf{W}_\theta)$ admits a local minimum point when $\mathbf{W}_\theta \in \mathcal{W}$ (this results from $J_r(S)\text{var(S)} < 1$), and thus violates a necessary requirement of a contrast function. Finally, on the last plot (r=5), the criterion is not a contrast even though $J_r(S)\text{var(S)} > 1$ since the set of *global* maximum points of the criterion does not correspond to the set $\mathcal{W}$.

4 Conclusion

In this paper, the contrast property of a well-known Renyi's entropy based criterion for blind source separation is analyzed. It is proved that at least in one realistic case, *globally* maximizing the related criterion does not provide the expected sources, whatever is the value of Renyi's exponent; the transfer matrix $\mathbf{W}$ globally maximizing the criterion might be a mixing matrix, with possibly more than one non-zero element per row. Even worst, it is not guaranteed that the criterion reaches a *local* maximum at non-mixing solutions ! Actually, the only thing we are sure is that the criterion is stationary for non-mixing matrices. This is a mere information since if the criterion has a local maximum (resp. minimum) point at mixing matrices, then a stationary point might also exist at mixing solution, i.e. at $\mathbf{W}$ such that the components of $\mathbf{WS}$ are not proportional to distinct sources. Consequently, the value of Renyi's exponent has to be chosen with respect to the source densities in order to satisfy $\sum_{i=1}^{K} J_r(S_i)\mathrm{var}(\mathrm{S_i}) > \mathrm{K}$ (again, this is not a sufficient condition: it does not ensure that the local maximum is global). Unfortunately, the problem is that the sources are unknown. Hence, nowadays, the only way to guarantee that $C_r(\mathbf{B})$ is a contrast function is to set $r = 1$ (mutual information criterion) or $r = 0$ (log-measure of the supports criterion, this requires that the sources are bounded); it can be shown that counter-examples exist *for any other value of* r*, including* $r = 2$. To conclude, we would like to point out that contrarily to the kurtosis criterion case, it seems that it does not exist a simple mapping $\phi[.]$ (such as e.g. the absolute value or even powers) that would match the set $\mathrm{argmax}_{\mathbf{B}}\, \phi[C_r(\mathbf{B})]$ to the set $\{\mathbf{B} : \mathbf{BA} \in \mathcal{W}\}$ where $\mathcal{W}$ is the set of non-mixing matrices, because there is no information about the sign of the relevant local optima.

References

1. Comon, P.: Independent component analysis, a new concept? Signal Processing 36(3), 287–314 (1994)
2. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley Series in Telecommunications. Wiley and Sons, Inc, Chichester (1991)
3. Dembo, A., Cover, T.M., Thomas, J.A.: Information theoretic inequalities. IEEE Transactions on Information Theory 37(6), 1501–1518 (1991)
4. Erdogmus, D., Hild, K.E., Principe, J.: Blind source separation using renyi's mutual information. IEEE Signal Processing Letters 8(6), 174–176 (2001)
5. Erdogmus, D., Hild, K.E., Principe, J.: Blind source separation using renyi's α-marginal entropies. Neurocomputing 49(49), 25–38 (2002)
6. Gray, R., Davisson, L.: An Introduction to Statistical Signal Processing. Cambridge University Press, Cambridge (2004)
7. Hild, K.E., Erdogmus, D., Principe, J.: An analysis of entropy estimators for blind source separation. Signal Processing, 86, 174–176, 182–194 (2006)
8. Hyvärinen, A., Karhunen, J., Oja, E.: Independent component analysis. John Willey and Sons, Inc. New York (2001)
9. Pham, D.-T.: Entropy of a variable slightly contaminated with another. IEEE Signal Processing Letters 12(7), 536–539 (2005)
10. Renyi, A.: On measures of entropy and information. Selected papers of Alfred Renyi 2, 565–580 (1976)