What should we know about the KURTOSIS ?

A. Mansour[(1)], C. Jutten[(2)]
[(1)] BMC Research Center (RIKEN), Moriyama-ku, Nagoya 463 (JAPAN)
[(2)] INPG - LIS, 46 avenue Félix Viallet, 38031 Grenoble, France
email: mansour@nagoya.riken.go.jp, chris@tirf.inpg.fr
http://www.bmc.riken.go.jp/sensor/Mansour/mansour.html

IEEE Signal Processing Letters, Vol 6 N 12, December 1999

Abstract

In various studies on blind separation of sources, one assumes that sources have the same sign of kurtosis. In fact, this assumption seems very strong and in this paper we studied relation between signal distribution and the sign of the kurtosis. A theoretical result has been found in a simple case. However, for more complex distributions, the kurtosis sign cannot be predicted and may change with parameters. The results give theoretical explanation to tricks, like non-permanent adaptation, used in non stationary situations.

Keywords: kurtosis, high order statistics, blind identification and separation, probability density function.

1 Introduction

In various works [9, 5, 4, 3, 10, 2] concerning the problem of blind separation of sources, authors propose algorithms whose efficacy demands conditions on the source kurtosis, and sometimes that all the sources have the same sign of kurtosis. In fact, this assumption seems very strong and in this paper we studied relation between signal distribution and the sign of its kurtosis.

2 Definition and Properties

Let us denote by $x(t)$ a zero-mean random process and by $p(x)$ its probability density function (pdf).

Definition 1: The kurtosis $K[p(x)]$ is the normalized fourth-order cumulant of the process [1, 8]:

$$K[p(x)] = \frac{Cum_4(x)}{E(x^2)^2} = \frac{E(x^4) - 3E(x^2)^2}{E(x^2)^2}, \tag{1}$$

where $E()$ denotes the average.
If the process is not zero-mean, the equation (1) becomes [8]:

$$K[p(x)] = \frac{Cum_4(x)}{E(x^2)^2} = \frac{E(x^4) - 3E(x^2)^2 + 12E(x)^2E(x^2) - 4E(x)E(x^3) - 6E(x)^4}{E(x^2)^2}. \tag{2}$$

Clearly, the kurtosis has the same sign than the fourth-order cumulant, then we will only study the sign of the fourth-order cumulant.

Let $ks(x)$ denote the kurtosis sign. Some properties can be easily derived :

1. The kurtosis sign, $ks(x)$, is invariant by any linear transformation. From (2), we deduce:
$$Cum_4(ax + b) = a^4 Cum_4(x), \tag{3}$$
then $ks(ax + b) = ks(x)$.
2. If we express the pdf $p(x)$ as a sum of two functions: $p(x) = p_e(x) + p_o(x)$, where $p_e(x)$ is even and $p_o(x)$ is odd, then:
    - $ks(x)$ only depends on the even function $p_e(x)$, because the fourth-order cumulant (1) depends only on the fourth and second-order moments (so it only depends on the even moments).
    - The even function $p_e(x)$ has the properties of a pdf:
$$\begin{aligned} & p_e(x) \geq 0, \quad \forall x \\ \text{and} \quad & \int_{\mathbb{R}} p(x)dx = \int_{\mathbb{R}} p_e(x)dx = 1. \end{aligned}$$

Therefore, in the following, the study may be restricted to a zero-mean process $x(t)$ whose the pdf $p(x)$ is even and has a variance $\sigma_x^2 = 1$. Clearly from (1), it is clear that the kurtosis of a Gaussian distribution is equal to zero. Moreover, for generalized exponential distributions $p(x) = K_1 \exp(-K_2|x|^\alpha)$ (K_i are normalization parameters), it is easy to show that for $\alpha > 2$, the kurtosis is negative and for $\alpha < 2$, the kurtosis is positive. For other distributions, the kurtosis may be positive or negative (see table 1). But intuitively, the sign of the kurtosis is related to the comparaison between $p(x)$ and Gaussian distribution. As examples, in the table 1, we computed $ks(x)$ for four well known distributions.

Usually, many authors only consider asymptotic properties of the distributions. It leads to the following definition.

Signal	$Cum_4(x)$	$ks(x)$	fig.
Uniform	$\frac{a^4+b^4+6a^2b^2+3.5(a^3b+b^3a)}{-30}$	$-$	1.a
Discrete	$\frac{-N(N+1)(2N^2+2N+1)}{15}$	$-$	1.b
Gamma	$\frac{26}{3}$, if $\sigma_x = 1$	$+$	1.c
Cosine	$\frac{192}{\pi^4} - 2 - 2\alpha^4$	$-$	1.d

Table 1: known distributions.

Definition 2: A pdf $p(x)$ is said over-Gaussian (respectively sub-Gaussian), if:

$$\exists \quad x_0 \in \mathbb{R}^+ \ \mid \ \forall x \geq x_0, \ p(x) > g(x) \tag{4}$$

(respectively $p(x) < g(x)$), where $g(x)$ is the normalized Gaussian pdf. From the previous examples, it seems that $ks(x)$ is positive for over-Gaussian signals and negative for sub-Gaussian signals.

3 Theoretical result

3.1 A simple Theorem

Let us consider $p(x)$ an (even) pdf and $g(x)$ a zero-mean normalized Gaussian pdf.

Theorem 1 *If $p(x) = g(x)$ have only two solutions than:*

$$Ks(x) > 0 \Longleftrightarrow p(x) \text{ is over-Gaussian}$$
$$Ks(x) < 0 \Longleftrightarrow p(x) \text{ is sub-Gaussian}$$

The demonstration is given in appendix A. This theorem shows the intuitive claim, given in the previous section, is true under the specific condition of theorem 1. So, this condition is satisfied for the generalized exponential distributions. Additionally, this result can be generalized for all unimodal distributions.

3.2 General cases

In the general case, if $p(x) = g(x)$ has more than two solutions, then there is no rule to predict $ks(x)$. More precisely, over-Gaussian as well as sub-Gaussian pdfs can lead to positive as well as negative sign of kurtosis. As an example, let us consider the pdf is a sum of two exponential functions:

$$p(x) = \frac{b}{4}(\exp(-b|x-a|) + \exp(-b|x+a|)) \tag{5}$$

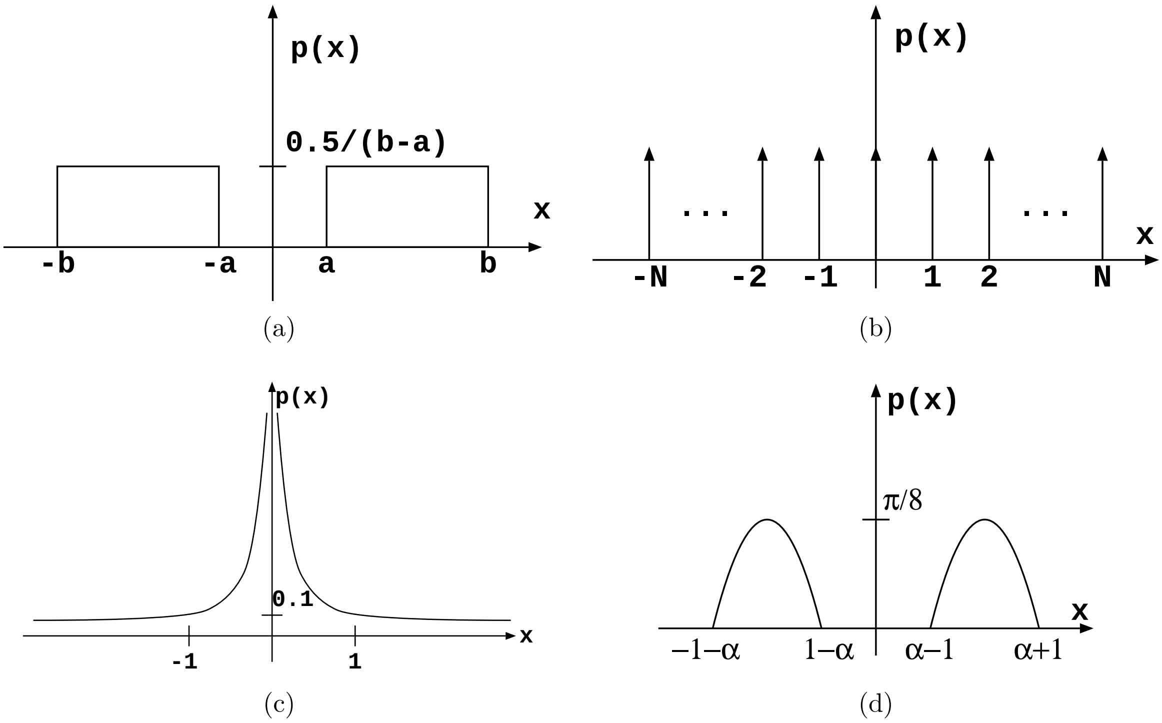


Figure 1: pdf of four random process of table 1

Figure 2 shows the general form of $p(x)$.

Figure 3 (a) and figure 3 (b) give examples (with different parameters a and b) where $ks(x) > 0$ and $ks(x) < 0$, respectively. On these figures, we show also the normalized Gaussian pdf $g(x)$: the previous theorem is not applicable, because there are morethan one solution (in $\mathbb{R}^+$) to the equation $p(x) = g(x)$.

Using the equation (5), it is easy to compute:

$$E(x^4) = a^4 + \frac{12}{b^2}a^2 + \frac{24}{b^4} \quad (6)$$

$$E(x^2) = a^2 + \frac{2}{b^2} \quad (7)$$

From equations (6) and (7), we can derive the kurtosis of (5):

$$K[p(x)] = 2\frac{6-(ab)^4}{4+4a^2b^2+a^4b^4}, \quad (8)$$

Then by choosing adequate values of the parameters a and b, it is possible to change $ks(x)$. From (8), it is clear that:

$$K[p(x)] \geq 0 \text{ if } 0 < ab \leq \sqrt[4]{6}.$$

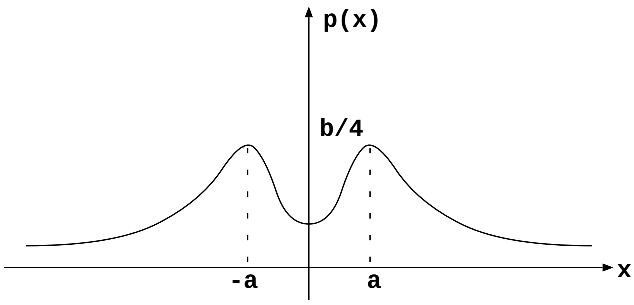


Figure 2: The exponential pdf

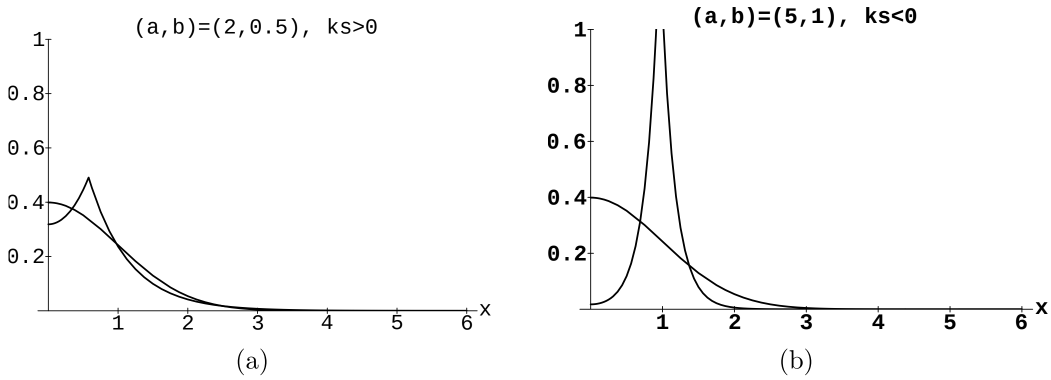


Figure 3: Comparaison of exponential pdf p(x) of (5) and the normalized Gaussian pdf $g(x)$ for various values of the parameters a and b.

$$K[p(x)] < 0 \text{ if } ab > \sqrt[4]{6}.$$

With respect to the definition (4), $p(x)$ is an even over-Gaussian pdf, and nevertheless $ks(x)$ is not always negative, but may change according to the values of the parameters a and b.

3.3 Case of bounded pdf

In practical cases, we may consider that artificial signals (for instance telecommunication signals) are bounded, and consequently their pdf are sub-Gaussian. It is after claimed that the kurtosis of such signals is negative. We show in this subsection that this claim is wrong.

Let us consider for instance quaternary sources $x(t)$ (see Fig 4), whose the fourth order cumulant is

$$Cum_4(x) = a^4 p(1-3p) - 6a^2b^2p(1-p) - b^4(1-p)(2-3p). \quad (9)$$

It is clear that the sign of $Cum_4(x)$ may change with the values of the parameters. For example let be $a = 0$, then $Cum_4(x) < 0$ if $p < 2/3$, and *vice*

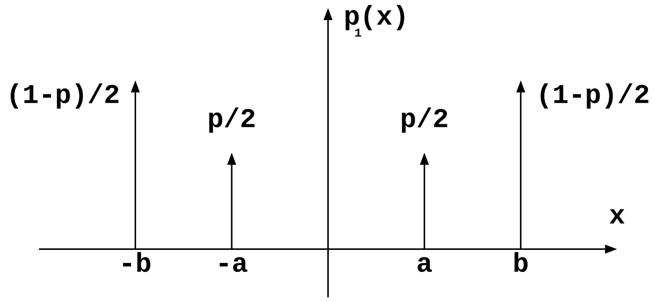
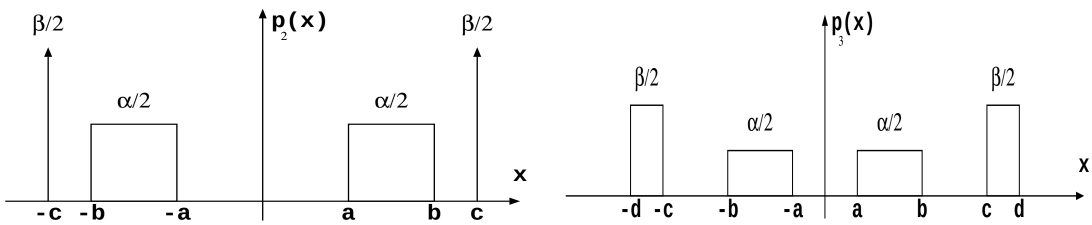


Figure 4: Pdf of quaternary sources.

versa.



(a) (b)

Figure 5: Two exemples of x-limited pdf

Finally, let us consider the exemples in figure 5. It is easy to evaluate the kurtosis of these signals (Fig 5). The kurtosis of first signal (Fig 5 (a)) can be written as:

$$K(p_2(x)) = \frac{\alpha}{5\sigma_x^4}(b^5 - a^5) + \frac{\beta}{\sigma_x^4}c^5 - \frac{\alpha^2}{3\sigma_x^4}(b^3 - a^3)^2 - \frac{3\beta^2}{\sigma_x^4}c^6 - 2\frac{\alpha\beta}{\sigma_x^4}(b^3 - a^3)c^3, \quad (10)$$

with the normalization condition:

$$\alpha(b-a) + \beta = 1. \quad (11)$$

The kurtosis of second signal (figure 5 (b)) is equal to:

$$\begin{aligned} K(p_3(x)) = {} & \frac{\alpha}{5\sigma_x^4}(b^5 - a^5) + \frac{\beta}{5\sigma_x^4}(d^5 - c^5) \\ & - \frac{\alpha^2}{3\sigma_x^4}(b^3 - a^3)^2 - \frac{\beta^2}{3\sigma_x^4}(d^3 - c^3)^2 - 2\frac{\alpha\beta}{3\sigma_x^4}(b^3 - a^3)(d^3 - c^3), \quad (12) \end{aligned}$$

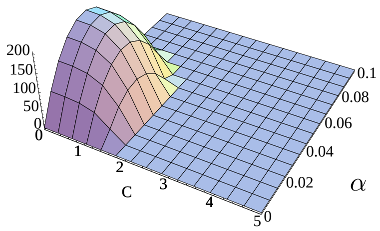
with the normalization condition:

$$\alpha(b-a) + \beta(d-c) = 1. \quad (13)$$

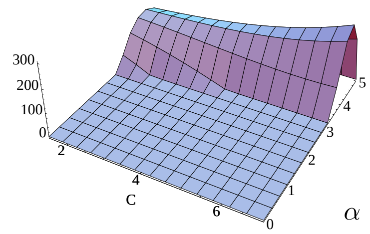
For scale reasons, we do not draw directly $K(p(x))$ but:

$$K^{\star}(p(x)) = \frac{1}{2}[K(p(x)) + \mid K(p(x)) \mid]. \tag{14}$$

Thus, if $K(p(x)) > 0$, $K^{\star}(p(x)) = K(p(x))$, otherwise, if $K(p(x)) \leq 0$, $K^{\star}(p(x)) = 0$. Then, we remark that the sign of the kurtosis may be easily controlled with adequate values of the pdf parameters (see Fig 6).



(a) $K^{\star}(p_2(x))$, with $a = 2$ and $b = 9$.

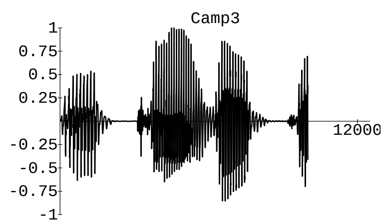


(b) $K^{\star}(p_3(x))$, with $a = 0.9$, $b = 1.1$ and $d = 9$.

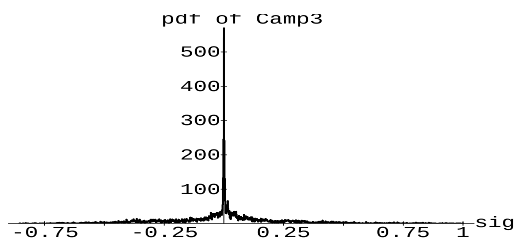
Figure 6: Representation of $K^{\star}(p(x))$ according to parameters c and α

4 Experimental results

In the case of real signals, the kurtosis estimation will be done on finite moving windows [7]. For stationary signals, the window may be very long. But for non-stationnary signals (speech signals for exemple, see Fig 7), the length of the window must be short enough (about 20-30 ms i.e. 2000 to 3000 samples at $F_s = 10$KHz).
Moreover, in case of non stationary signals, the pdf can vary a lot : for instance, silent periods in speech signals imply a peak in the pdf around $x = 0$ (see figure 7).



Signl respecting time



Estimated pdf

Figure 7: Speech signal: Camp3

According to the size of the window, and its location, we observe changes in the kurtosis sign. Figure 8 shows the kurtosis time evolution of the speech signal of figure 7. The kurtosis is estimated on 500-sample windows, every 50 samples. We remark that the kurtosis is negative during a silent period, and it becomes positive during the speech transient.

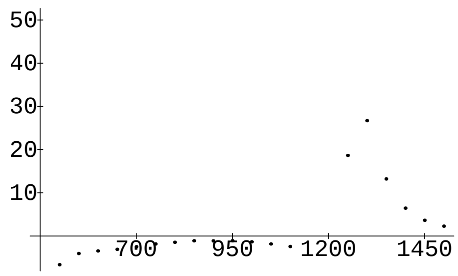
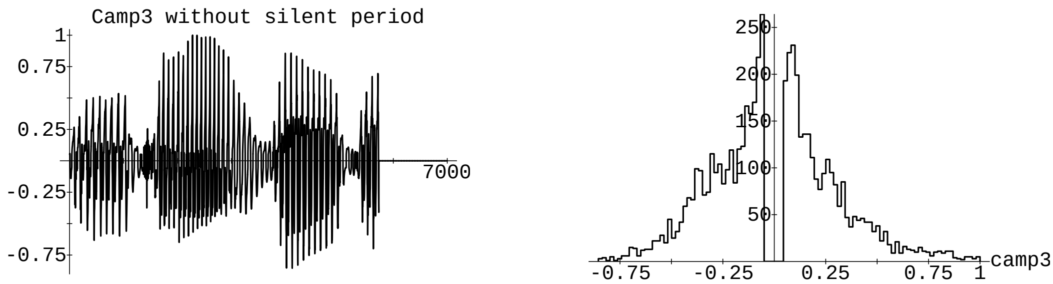


Figure 8: The estimated kurtosis of speech signal "Camp3"

Experimentally, we remark that: for speech signal, the kurtosis sign fluctuations can be eliminated by estimating the kurtosis on all the samples excepted those of silent periods (see Fig 9).
This result can explain that the non-permanent learning (freezing the parameter estimation) in speech separation algorithms [10] enforces source pdf to have a negative kurtosis sign and then allows algorithm convergence.



(a) Speech signal "Camp3" without its silent periods. (b) Estimated pdf of this signal.

Figure 9: Experimental results

5 Conclusion

In the paper, we point out some relations between pdf and kurtosis sign. First, we show the kurtosis sign is not modified by any scale or translation factors, and it only depends on the even part of the pdf.

Usually, people associates the kurtosis sign of a distribution $p(x)$ to its over-Gaussian or sub-Gaussian nature. We prove that this claim is only relevant for unimodal pdf $p(x)$ having only two intersections (in $\mathbb{R}$) with the Gaussian pdf.

In the general case, even for bounded pdf, we show by a few examples that the kurtosis sign can be positive or negative according to the pdf parameters.

From a practical point of view, kurtosis sign of non stationary signals, which must be estimated on short moving windows, can change. A previous experimental study proves that the kurtosis sign of speech signal can be affected by the silent period [6]. Additionally, this paper gives a theoretical explanation to the necessity and the efficacy of intermittent adaptation which is used for separation of non stationary sources [10].

A Proof of Theorem 1

Let us consider that for $x > 0$, there is one and only one intersection point ρ between $p(x)$ and $g(x)$.

It is known that the fourth-order cumulant of a Gaussian signal is zero. As a consequence, we can write:

$$\int_{\mathbb{R}} x^4 g(x)dx = 3\int_{\mathbb{R}} x^2 g(x)dx = 3. \tag{15}$$

Using (1) and the unit variance signal, the kurtosis can be rewritten as:

$$K[p(x)] = \int_{\mathbb{R}} x^4 (p(x) - g(x))dx. \tag{16}$$

According to result of section 2, we may only consider the even pdf. In addition, we just may study the sign of Υ:

$$\begin{aligned} \Upsilon &= \frac{1}{2}K[p(x)] = \int_0^{\infty} x^4(p(x) - g(x))dx \\ &= \int_0^{\rho} x^4(p(x) - g(x))dx + \int_{\rho}^{\infty} x^4(p(x) - g(x))dx. \end{aligned} \tag{17}$$

Let us consider that the pdf $p(x)$ is an over-Gaussian signal ($p(x) > g(x)$, when $x \to \infty$). Then, the sign of $p(x) - g(x)$ remains constant on each interval [0,ρ], and [ρ, ∞]. Using the second mean value theorem, Υ can be rewritten as:

$$\begin{aligned} \Upsilon &= \xi^4 \int_0^{\rho} (p(x) - g(x))dx + \lambda^4 \int_{\rho}^{\infty} (p(x) - g(x))dx \\ &= \lambda^4 \int_{\rho}^{\infty} (p(x) - g(x))dx - \xi^4 \int_0^{\rho} (g(x) - p(x))dx \end{aligned} \tag{18}$$

where:

$$0 < \xi < \rho < \lambda. \tag{19}$$

In fact, $p(x)$ and $g(x)$ are both pdf, so we have:

$$\int_0^\infty (p(x) - g(x))dx = \int_0^\rho (p(x) - g(x))dx + \int_\rho^\infty (p(x) - g(x))dx = 0 \tag{20}$$

From (20), and taking into account that $p(x)$ is over-Gaussian, we deduce:

$$\int_\rho^\infty (p(x) - g(x))dx = \int_0^\rho (g(x) - p(x))dx > 0. \tag{21}$$

Using (18), (19) and (21), we remark that:

$$\Upsilon = (\lambda^4 - \xi^4) \int_\rho^\infty (p(x) - g(x))dx > 0. \tag{22}$$

Finally, if $p(x)$ is an over-Gaussian pdf (with the assumption of an unique intersection positive point between $p(x)$ and $g(x)$), then its kurtosis is positive. Using the same reasoning, we can claim that a sub-Gaussian pdf has a negative kurtosis.

References

[1] A. Benvéniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations.* Spring-Verlag, 1990.

[2] A. Cichocki, S. Douglas, S. I. Amari, and P. Mierzejewski. Independent component analysis for noisy data. In *Proc. of International Workshop on Independence & Artificial Neural Networks*, pages 52–58, Tenerife, SPAIN, February 9-10 1994.

[3] N. Delfosse and P. Loubaton. Adaptive blind separation of independent sources: A deflation approach. *Signal Processing*, 45(1):59–83, July 1995.

[4] B. Laheld and J. F. Cardoso. Adaptive source separation with uniform performance. In M.J.J. Holt, C.F.N. Cowan, P.M. Grant, and W.A. Sandham, editors, *Signal Processing VII, Theories and Applications (EUSIPCO'94)*, pages 1–4, Edinburgh, Scotland, September 1994. Elsevier.

[5] A. Mansour and C. Jutten. Fourth order criteria for blind separation of sources. *IEEE Trans. on Signal Processing*, 43(8):2022–2025, August 1995.

[6] A. Mansour, C. Jutten, and N. Ohnishi. Kurtosis: Definition and properties. In H. R. Arabnia and D. D. Zhu, editors, *International Conference on Multisource-Multisensor Information Fusion*, pages 40–46, Las Vegas, USA, 6-9 July 1998.

[7] A. Mansour, A. Kardec Barros, and N. Ohnishi. Comparison among three estimators for high order statistics. In S. Usui and T. Omori, editors, *Fifth International Conference on Neural Information Processing (ICONIP'98)*, pages 899–902, Kitakyushu, Japan, 21-23 October 1998.

[8] P. McCullagh. *Tensor methods in statistics.* Chapman and Hall, 1987.

[9] E. Moreau and O. Macchi. New self-adaptive algorithms for source separation based on contrast functions. In *IEEE Signal Processing Workshop on Higher-Order Statistics*, pages 215–219, South Lac Tahoe, USA (CA), June 1993.

[10] L. Nguyen Thi and C. Jutten. Blind sources separation for convolutive mixtures. *Signal Processing*, 45(2):209–229, 1995.