ON THE SIGN OF KURTOSIS

Ciprian Doru Giurcăneanu and Ioan Tăbuş

Signal Processing Laboratory, Tampere University of Technology
P.O. Box 553, FIN-33101 Tampere, Finland
e-mail: {cipriand,tabus}@cs.tut.fi

ABSTRACT

Starting from the attention payed in the last period to the sign of kurtosis as criterion for stability of ICA/BSS algorithms [3][12], the paper investigates how the sign is determined by the shape of the probability density function (pdf). Since the approach based on the number of intersection points between the pdf and the gaussian pdf with the same mean and variance [12] does not generalize, another approach [4][5][13], based on the number of crossing points of the pdf's is considered. Some results from literature are revisited and original proofs are presented, linking the above approaches. A non-trivial example of pdf which illustrates the general approach is presented.

1. INTRODUCTION

1.1. Kurtosis and the stability of ICA/BSS algorithms

In the linear version of ICA/BSS problem the realizations of a random vector $\underline{x} = \mathrm{A}\underline{s}$ are observed; $\underline{s}$ is an $n \times 1$ vector of independent components and A is an $n \times n$ invertible matrix with unknown entries. The goal is to find an $n \times n$ "separation" matrix B such that $\underline{y} = \mathrm{B}\underline{x}$ retrieves the original components of $\underline{s}$ (modulo a scaling and a permutation operation). A general class of on-line algorithms adapt the matrix B following the rule [3]:

$$\begin{aligned} \underline{y}(t) &= \mathrm{B}(t)\underline{x}(t) \\ \mathrm{B}(t+1) &= \left[\mathrm{I} - \mu(t) G(\underline{y}(t))\right] \mathrm{B}(t) \end{aligned} \quad (1)$$

where $\{\mu(t)\}$ is a sequence of positive learning steps, I is the $n \times n$ identity matrix and the function $G : \Re^n \to \Re^{n \times n}$ is an appropriate vector-to-matrix mapping. In [3] the stability conditions are stated for the cases

$$G(\underline{y}) = G_r(\underline{y}) = \phi(\underline{y})\underline{y}^T - \mathrm{I} \quad (2)$$

and

$$G(\underline{y}) = G_o(\underline{y}) = \underline{y} \times \underline{y}^T - \mathrm{I} + \phi(\underline{y})\underline{y}^T - \underline{y}\phi(\underline{y})^T \quad (3)$$

where

$$\begin{aligned} \phi(\underline{y}) &= \left[\phi_1(y_1), \phi_2(y_2), \ldots, \phi_n(y_n)\right]^T \\ \phi_i(s_i) &= -\frac{r_i'(s_i)}{r_i(s_i)}, \quad i = \overline{1,n} \end{aligned} \quad (4)$$

$r_i(s_i)$ being the hypothesized marginal density of the source s_i. Let's observe the marginal densities are supposed to be differentiable and the function $\phi(\underline{y})$ is component-wise nonlinear. For simplicity $E\left[\underline{s}\right] = \underline{0}$.

In [3] the stability conditions for algorithm (1) are found for the regular case (2) and the orthogonal case (3). In the regular case, the scale stability condition for the i-th source is $1 + E\left[\phi_i'(y_i)y_i^2\right] > 0$. In the orthogonal case the local stability with respect to scale is ensured by forcing the components of $\underline{y}$ to have unitary variance. To express the pair-wise stability conditions in [3] are defined the moments:

$$k_i = E\left[\phi_i'(y_i)\right] E\left[y_i^2\right] - E\left[\phi_i(y_i)y_i\right]. \quad (5)$$

With respect to these notations, the stability conditions are:

- regular case:

$$\begin{aligned} (1+k_i)(1+k_j) > 1 \quad & 1 \le i < j \le n \\ 1 + k_i > 0 \quad & 1 \le i \le n \end{aligned} \quad (6)$$

- orthogonal case:

$$k_i + k_j > 0 \quad 1 \le i < j \le n \quad (7)$$

In both (6) and (7) $k_i > 0, i = \overline{1,n}$ is a sufficient condition for stability. For $\phi_i(y_i) = ay_i + y_i^3$ (with any scalar a), we have $k_i = 3\left(E\left[y_i^2\right]\right)^2 - E\left[y_i^4\right]$ which is the fourth order cumulant of y_i with reversed sign.

This final observation stresses the role played by kurtosis in the context of studying the stability for ICA/BSS algorithms. The kurtosis is the fourth order normalized cumulant and for symmetric, zero mean sources (the case we address in this paper) it has the expression $\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3$ where μ_4 is the fourth and μ_2 is the second moment about the mean for the considered distribution. Is is immediate to observe that the kurtosis is null for the gaussian density with zero mean and unitary variance, denoted in the sequel by $\alpha(x)$.It is well known [11] that $\gamma_2 \ge -2$ for any random variable. We can note for the random variable which takes only the values ± 1 with probability 0.5 each, the kurtosis attains the minimum value -2.

1.2. Definitions and notations

In [12] it is investigated how the kurtosis sign describes the asymptotic behavior of the distribution and the next definition is introduced:

Definition 1[12]: A symmetric, zero mean pdf $f(x)$ with unitary variance is said to be *over-gaussian* (respectively *sub-gaussian*) if exists $x_0 \in \Re^+$ such that $\forall x \geq x_0, f(x) > \alpha(x)$ (respectively $f(x) < \alpha(x)$).

A tantalizing question: a pdf is over-gaussian (sub-gaussian) if and only if it has positive (negative) kurtosis ? This is the main question in [12]. *Theorem 1* from [12] proves that any symmetric, zero mean, unitary variance density $p(x)$ for which the equation $p(x) = \alpha(x)$ has two roots, is *over-gaussian* (*sub-gaussian*) if and only if the kurtosis fulfills the condition $\gamma_2 > 0$ ($\gamma_2 < 0$).

The community of statisticians has also investigated this problem; probably the first reference is the seminal paper of Pearson [15]. In this paper the author claims that not all the experimental frequency curves are well-fitted to gaussian distributions. In his work, Pearson uses the degree of flat-toppedness of the curve to establish the gaussianity feature and also for classifying the non-gaussian curves:

"Given two frequency distributions which have the same variability as measured by the standard deviation, they may be relatively more or less flat-topped than the normal curve. If more flat-topped I term them *platykurtic*, if less flat-topped *leptokurtic*, and if equally flat-topped *mesokurtic*".

In the same paper [15], the author observes for some pdf's $f(x)$ that $\gamma_2 > 0$ for *leptokurtic* sources (those with $f(0) > \frac{1}{\sqrt{2\pi}}$), $\gamma_2 < 0$ for *platikurtic* sources (those with $f(0) < \frac{1}{\sqrt{2\pi}}$) and $\gamma_2 = 0$ for *mesokurtic* sources and empirically correlates the kurtosis sign with the shape of the distribution curve in the neighborhood of the mean.

We are using in the following the symbols $f(x)$ and $g(x)$ to denote two symmetric, zero mean pdf's. The notation $\mathcal{C}(f,g)$ is for how many times the sign of $f(x) - g(x)$ is changing for $x > 0$. Due to symmetry, the total number of sign changes for $f(x) - g(x)$ when $x \in \Re$ is $2 \times \mathcal{C}(f,g)$. The notation μ_k^f is used for the k-th order moment about the mean of distribution $f(x)$; γ_2^f is the kurtosis of the distribution. For any non-negative integer s, $v_s^f = \int_{-\infty}^{+\infty} |x|^s f(x)dx$. From definition results $v_0^f = 1$ and for any integer $p > 0$, $v_{2p}^f = \mu_{2p}^f$.

In the sequel we will analyze the relation between γ_2^f and γ_2^g with respect to different values for $\mathcal{C}(f,g)$.

2. SIGN OF KURTOSIS AND SHAPE OF THE PROBABILITY DENSITY FUNCTION

The idea that the kurtosis sign can be inferred from the shape of the distribution function has been popular since the Pearson's paper, and more results have been formulated, which we review in the following. We will illustrate them with several examples and present also original results.

2.1. Equal variance continuous pdf's $f(x)$ and $g(x)$ cannot have two crossing points

Proposition 1: There are no continuous and distinct pdf's $f(x)$ and $g(x)$ with the same variance ($\mu_2^f = \mu_2^g$) for which $\mathcal{C}(f,g) = 1$.

Proof: Suppose there are two distinct, continuous, symmetric, zero mean pdf's $f(x)$ and $g(x)$ with $\mathcal{C}(f,g) = 1$. Let $a_1 (0 < a_1 < \infty)$ be the sign changing point, i.e. $f(x) \geq g(x)$ if $|x| < a_1$ and $f(x) \leq g(x)$ if $|x| > a_1$. We define $h : \Re \to \Re$, $h(x) = (x^2 - a_1^2)(g(x) - f(x))$. It is direct to observe $h(x) \geq 0, \forall x \in \Re$ and $I = \int_{-\infty}^{+\infty} h(x)dx \geq 0$. But $I = (\mu_2^g - \mu_2^f) - a_1^2(v_0^g - v_0^f) = 0$ and from the continuity condition it results $\forall x \in \Re, h(x) = 0 \Leftrightarrow f(x) = g(x)$, contradiction.

□

Corollary 1.1: There are no continuous pdf's $f(x)$ and $g(x)$, with $\mu_2^f = \mu_2^g$ such that the equation $f(x) = g(x), x \in \Re$ has exactly two different solutions.

Proof: Let's consider ρ and $-\rho$ are the real-valued solutions of the equation $f(x) = g(x)$. The condition $\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{+\infty} g(x)dx = 1$ implies there is at least one point $a_1, 0 < a_1 < \infty$ which is a crossing point of the distributions f and g. From Bolzano theorem [2] it results a_1 is a root for the equation $f(x) = g(x)$. Due to the uniqueness of the positive root of the equation $f(x) = g(x)$, it results $a_1 = \rho$. It is immediate to observe that any other positive value which is a crossing point should be a root of the equation $f(x) = g(x)$, therefore equals ρ. So ρ and $-\rho$ are the only values for which the sign of the function $f(x) - g(x)$ is changing and $\mathcal{C}(f,g) = 1$.*Proposition 1* applies directly.

□

Corollary 1.2: There is no continuous pdf $f(x)$, $\mu_2^f = \mu_2^\alpha = 1$ such that the equation $f(x) = \alpha(x), x \in \Re$ has exactly two different solutions.

Proof: From *Corollary 1.1* by choosing $g(x) = \alpha(x)$.

□

Observation: Corollary 1.2 shows that the conditions of *Theorem 1* from [12] are accomplished only if at least one of the pdf's is not continuous. As one pdf in the theorem is $\alpha(x)$ (continuous),remains that the distribution for which we decide if is over-gaussian (sub-gaussian) has necessarily to have discontinuous pdf.

In Section 1.1 we shown that the kurtosis based comparison of two distributions is interesting when both distributions have the same variance. As a result of *Corollary 1.2* the case $f(x) = \alpha(x)$ has exactly two roots (treated in *Theorem 1* [12]) is possible only for discontinuous pdf $f(x)$.

2.2. Densities with four crossing points

We can observe that comparing kurtosis of distributions in the hypothesis of equal variances reduces to the comparison of fourth order moments.

Proposition 2: If the pdf's $f(x)$ and $g(x)$ are such that $\mu_2^f = \mu_2^g$ and two numbers a_1, a_2 exist such that $0 < a_1 < a_2 < \infty$ and

$$\begin{array}{lll} f(x) > g(x) & \text{if} & |x| < a_1 \text{ or } |x| > a_2 \\ f(x) < g(x) & \text{if} & a_1 < |x| < a_2 \end{array}$$

then

a) $\gamma_2^f \geq \gamma_2^g$ which is equivalent with $\mu_4^f \geq \mu_4^g$ [4][5].

b) $v_1(f) \leq v_1(g)$ [1][13].

Proof:

a) We summarize the proof from [4][5]. Let $h : \Re \to \Re, h(x) = (x^2 - a_1^2)(x^2 - a_2^2)(f(x) - g(x))$. From $h(x) \geq 0, \forall x \in \Re \Rightarrow (\mu_4^f - \mu_4^g) - (a_1^2 + a_2^2)(\mu_2^f - \mu_2^g) + a_1^2 a_2^2 (v_0^f - v_0^g) \geq 0$. By reducing the null terms, the inequality becomes $\mu_4^f \geq \mu_4^g \Leftrightarrow \gamma_2^f \geq \gamma_2^g$.

b) We give a simpler proof than in [1][13]. Define $h : \Re^+ \to \Re$, $h(x) = (x-a_1)(x-a_2)(f(x)-g(x))$. It is easy to observe $h(x) \geq 0, \forall x \in \Re^+$ and $v_1^f = \int_{-\infty}^{+\infty} |x| f(x)dx = 2\int_0^{+\infty} x f(x)dx$. The fact that function $h(x)$ has only non-negative values implies $\int_0^{+\infty} h(x)dx \geq 0 \Rightarrow (\mu_2^f - \mu_2^g) - (a_1+a_2)(v_1^f - v_1^g) + a_1 a_2 (v_0^f - v_0^g) \geq 0$. From condition of equal variance for $f(x)$ and $g(x)$, it results that $-(a_1 + a_2)(v_1^f - v_1^g) \geq 0$ and with $a_1, a_2 > 0$ we get $v_1^f \leq v_1^g$.
□

Observation: From the hypotheses of *Proposition 2* results that the function $h(x)$ used to prove part a) is null at most for $x \in \{-a_2, -a_1, a_1, a_2\}$. In proof of part b), the function $h(x)$ is zero at most for $x \in \{a_1, a_2\}$. In both situations the number of roots of the equation $h(x) = 0$ is finite and therefore the inequalities are strict in the parts a) and b) of *Proposition 2*.

An example of such a pair of pdf's is $f(x) = \alpha(x)$ and

$$g(x) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{if } |x| \leq \sqrt{3} \\ 0 & \text{if } |x| > \sqrt{3} \end{cases}$$

Elementary calculations prove the conditions from *Proposition 2* are accomplished: $\mu_2^f = \mu_2^g = 1, \mathcal{C}(f,g) = 2$, $a_1 = \left(2\ln\left(\frac{2\sqrt{3}}{\sqrt{2\pi}}\right)\right)^{1/2} \approx 0.8$, $a_2 = \sqrt{3} \approx 1.73$. The rectangular pdf $g(x)$ is *platykurtic* [15] and also is *sub-gaussian* [12] (for x_0 can be chosen any real value greater than $\sqrt{3}$). The negative value of kurtosis $\gamma_2^g = -1.2$ is like in the part a) of the *Proposition 2* ($\gamma_2^f = \gamma_2^\alpha = 0$) and $v_1^f = \frac{2}{\sqrt{2\pi}} \approx 0.79 < v_1^g = \frac{\sqrt{3}}{2} \approx 0.86$. The negative kurtosis for $f(x)$ agrees with the intuition we have about the pdf that is *platykurtic* or *sub-gaussian*. It is easy to check that $f(x)$ fulfills the conditions of *Theorem 1* from [12], and we note that $f(x)$ is discontinuous (*Corollary 1.2*) with two discontinuity points.

2.3. Arbitrary number of crossings

Proposition 3: Let f and g be pdf's with $v_s^f = v_s^g$ for $s_k > s_{k-1} > \cdots > s_1 = 0$, $k > 2$. Let the positive axis be divided into $k+1$ successive intervals (a_0, a_1), (a_1, a_2), ..., (a_k, a_{k+1}), where $a_0 = 0$, $a_{k+1} = +\infty$, such that $f(x) > g(x)$ on the first, third, etc., intervals, and $f(x) < g(x)$ on the second, fourth,etc.,intervals. Then

$$v_s^f > v_s^g \quad \text{for} \quad s_2 < s < s_3, s_4 < s < s_5, \ldots,$$

for s such that $v_s^g < \infty$, while

$$v_s^f < v_s^g \quad \text{for} \quad s_1 < s < s_2, s_3 < s < s_4, \ldots,$$

for s such that $v_s^f < \infty$ [13].

Proof: An elegant proof is based on the theory of total positivity [13].

Definition 2[10]: A function of two real variables $Q(x,y)$ ranging over linearly ordered sets X and Y is said to be *totally positive of order r* (TP_r) if for all $1 \leq m \leq r, x_1 < x_2 < \cdots < x_m, y_1 < y_2 < \cdots < y_m$, $(x_i \in X, y_j \in Y)$, the determinants of the matrices $[T_{ij}]_{1\leq i,j\leq m}$ are non-negative, where the entries of the matrices are given by $T_{ij} = Q(x_i, y_j)$, $1 \leq i,j \leq m$.

Let's define $h, w : \Re^+ \to \Re, h(x) = f(x) - g(x), w(s) = \int_0^{+\infty} x^s h(x)dx$. From $v_s^f - v_s^g = 2w(s)$ it results that finding the sign for $v_s^f - v_s^g$ is equivalent with finding the sign for $w(s)\forall s \in \Re^+$ (in fact we pay attention only for the case s is a positive integer). The properties of generalized Vandermonde determinants [6] imply the function $Q : \Re^+ \times \Re^+ \to \Re, Q(s,x) = x^s$ is TP_{r+1} for any integer $r \geq 1$. Moreover for the function $Q(s,x) = x^s$ the determinants invoked by *Definition 2* are strict positive [6]. Now the demonstration is direct from the *variation diminishing property* of totally positive functions [10]: $Q(s,x)$ is TP_{r+1} and $h(x)$ changes the sign $k \leq r$ times, then $w(s)$ changes sign at most k times. Since $w(s)$ changes sign exactly k times, then $w(s) = \int_0^{+\infty} Q(s,x)h(x)dx$ must have the arrangement of sign as the function $h(x)$ as s and x traverse their domains from left to right.
□

*Observations:*a) *Proposition 2* results from *Proposition 3* for the particular case $k = 2, s_1 = 0, s_2 = 2$.

b) *Proposition 3* is proved also in [1], but for the case when $a_1, a_2, \ldots, a_k$ are positive integers.

Proposition 3 provides a general statement and in the following we give an example of densities which satisfy the constraints in the hypotheses of the proposition. Let $f : \Re \to \Re, f(x) = \alpha(x)\left[1 + \frac{a_{2n}}{\sqrt{(2n)!}} H_{2n}(x)\right], n > 1$ where $H_{2n}(x)$ is the order $2n$ Chebyshev-Hermite polynomial [11] and a_{2n} is a small positive constant. The order r Chebyshev-Hermite polynomial is recursively defined by $H_0(x) = 1$ and for any positive integer r, $H_r(x) = (-1)^r \frac{1}{\alpha(x)} \frac{\mathrm{d}^r[\alpha(x)]}{\mathrm{d}x^r} = (-1)^r e^{\frac{x^2}{2}} \frac{\mathrm{d}^r}{\mathrm{d}x^r}\left[e^{-\frac{x^2}{2}}\right]$. In our approach we are using also the order r Hermite polynomial recursively defined [9] by the equations $H_0^\star(x) = 1$, $H_r^\star(x) = (-1)^r e^{x^2} \frac{\mathrm{d}^r}{\mathrm{d}x^r}\left[e^{-x^2}\right] \forall r \geq 1$.

It follows $H_r(x) = \frac{1}{2^{r/2}} H_r^\star\left(\frac{x}{\sqrt{2}}\right) \forall x \in \Re, r \geq 1$. From $|H_{2n}^\star(x)| \leq 2^n \left[(2n)!\right]^{1/2} e^{\frac{x^2}{2}} \forall n, x$ (equality occurs for $x = 0, n = 0$) [9] results

$$\left|\frac{a_{2n}}{\sqrt{(2n)!}} H_{2n}(x)\right| \leq a_{2n} e^{\frac{x^2}{4}} \tag{8}$$

We are looking for $a_{2n} > 0$ such that $f(x) \geq 0, \forall x \in \Re$. The polynomial $H_{2n}(x)$ has $2n$ distinct real roots [11]. Since $H_{2n}(x)$ is an even function [11], the roots are $\pm\rho_1, \pm\rho_2, \cdots, \pm\rho_n$, $0 < \rho_1 < \rho_2 < \cdots < \rho_n$.
The roots of $H_{2n}^\star(x)$ are $\pm\frac{\rho_1}{\sqrt{2}}, \pm\frac{\rho_2}{\sqrt{2}}, \cdots, \pm\frac{\rho_n}{\sqrt{2}}$ and $\frac{\rho_n}{\sqrt{2}} < 2\sqrt{n}$ [9]. It is immediate to observe $H_{2n}(x) > 0, \forall x$ such that $|x| \geq 2\sqrt{2n}$. It results $f(x) > 0, \forall x$ such that $|x| \geq 2\sqrt{2n}$. If we choose $a_{2n} \in \left(0, e^{-2n}\right)$ it results $a_{2n} e^{\frac{x^2}{4}} < 1, \forall x$ such that $|x| < 2\sqrt{2n}$. The inequality (8) implies $\left|\frac{a_{2n}}{\sqrt{(2n)!}} H_{2n}(x)\right| < 1, \forall x$ such that $|x| < 2\sqrt{2n}$ which in turn implies that $f(x) > 0, \forall x$ such that $|x| < 2\sqrt{2n}$. This proves the existence of a_{2n} (small and positive constant) such that $f(x) \geq 0, \forall x \in \Re$.

Let's note that from $\alpha(x)$ symmetric and $H_{2n}(x)$ an even function results that $f(x)$ is also symmetric. Since the equation $H_{2n}(x)=0$ has $2n$ distinct real roots $\pm\rho_i, i=\overline{1,n}$ the borders of the intervals from the hypothesis of *Proposition 3* can be chosen $a_i=\rho_i, i=\overline{1,k}, k=n, g(x)=\alpha(x)$.

After some straight derivations we obtain for any non-negative integer p

$$v_{2p}^f=\frac{(2p)!}{2^p p!}\left[1+\frac{a_{2n}2^n}{\sqrt{(2n)!}}p(p-1)\cdots(p-n+1)\right] \qquad (9)$$

From (9), for $p=0$ it results $\int_{-\infty}^{+\infty}f(x)dx=1$ which confirms that $f(x)$ is really a pdf. Furthermore for $p\in\{1,2,\ldots,n-1\}$, $v_{2p}^f=\frac{(2p)!}{2^p p!}=v_{2p}^\alpha$. So using the notations from *Proposition 3* with $k=n$ and $g(x)=\alpha(x)$ we can write $v_s^f=v_s^g$ for $s_1<s_2<\cdots<s_k$ where $s_1=0, s_2=2,\ldots,s_k=2(k-1)$, in particular the variance of $f(x)$ is unitary.

For odd order moments the expression of $v_{2p+1}^f-v_{2p+1}^\alpha$ is given by

$$\begin{cases}\frac{a_{2n}}{\sqrt{(2n)!}}(-1)^{n-p-1}\sqrt{\frac{2}{\pi}}\frac{(2p+1)!(2n-2p-2)!}{2^{n-p-1}(n-p-1)!} & \text{if} \quad 2p+1<2n\\ \frac{a_{2n}}{\sqrt{(2n)!}}\sqrt{\frac{2}{\pi}}\frac{2^{p-n}(p-n)!(2p+1)!}{(2p-2n+1)!} & \text{if} \quad 2p+1>2n\end{cases} \qquad (10)$$

where $v_{2p+1}^\alpha=\sqrt{\frac{2}{\pi}}2^p p!$. It follows that all the inequalities between v_s^f and v_s^g are like in *Proposition 3*.

Due to $H_{2n}(0)=(-1)^n\frac{1}{2^n}\frac{(2n)!}{n!}$ [9], it follows $f(0)>\alpha(0)$ ($f(x)$ is *leptokurtic*) if n even and $f(0)<\alpha(0)$ ($f(x)$ is *platikurtic*) if n odd, but $\gamma_2^f=0$. This example shows that the condition $\gamma_2^f=0$ is not sufficient for the source $f(x)$ to be *mesokurtic*, contrary to what Pearson suggested [15].

From the proof above we observe that $x_0=2\sqrt{2n}$ is the point for which $f(x)>\alpha(x), \forall x\geq x_0$. Therefore $f(x)$ is *over-gaussian* [12], but the kurtosis $\gamma_2^f=\gamma_2^\alpha=0$, showing that *over-gaussian* does not necessarily imply strict positive kurtosis.

3. FINAL REMARKS

Proposition 2 and *3* present conditions on the shape of a pdf for which the sign of kurtosis can be predicted. In general, the converse propositions are not true. In [4] there is a counter-example for the converse of *Proposition 2*. So, in general, it is not possible to find direct links between the flat-toppedness or shape of tails of the pdf and the sign of kurtosis.

We finish by recalling the more realistic interpretation of kurtosis found in [14]. To any random variable X with mean μ and variance σ^2 we associate the standardized random variable $Z=\frac{X-\mu}{\sigma}$ having $E\left[Z^2\right]=1$. The kurtosis of X is $\gamma_2=E\left[Z^4\right]-3=\text{var}\left[Z^2\right]-2$ where $\text{var}\left[Z^2\right]$ denotes the variance of Z^2. Now it is obvious that the kurtosis of X is a measure of the dispersion of Z^2 around its mean [14]. Therefore the kurtosis of X is an inverse of the concentration of the distribution of X in the points $\mu+\sigma$ and $\mu-\sigma$. In the particular case of symmetric X with zero mean and unitary variance the kurtosis measures the dispersion of X around the values $+1$ and -1.

It remains as an open question how the ideas of quantile based measures for kurtosis [7][8][14] could be applied to ICA/BSS problems since the quantile based measures are more robust than the fourth order normalized cumulant.

4. REFERENCES

[1] M.M. Ali. Stochastic ordering and kurtosis measure. *J. of the American Statistical Association*, 69(346):543–545, Jun. 1974.

[2] T.M. Apostol. *Calculus*. Blaisdell Pub. Co., 1961.

[3] J.F. Cardoso. On the stability of some source separation algorithms. In *Proc. of the 1998 IEEE SP workshop on neural networks for signal processing (NNSP '98)*, pages 13–22, 1998.

[4] F.J. Dyson. A note on kurtosis. *J. of the Royal Statistical Soc.*, 106(4):360–361, 1943.

[5] H.M. Finucan. A note on kurtosis. *J. of the Royal Statistical Soc.,Ser.B*, 26(1):111–112, 1964.

[6] F.R. Gantmacher. *The theory of matrices*, volume 2. Chelsea Pub. Co., 1974.

[7] R.A. Groenweld. A class of quantile measures for kurtosis. *The American Statistician*, 51(4):325–329, 1998.

[8] R.A. Groenweld and G. Meeden. Measuring skewness and kurtosis. *Statistician*, 33(4):391–399, Dec. 1984.

[9] J. Indritz. An inequality for Hermite polynomials. *Proceedings of the American Mathematical Soc.*, 12(6):981–983, Dec. 1961.

[10] S. Karlin. Total positivity, absorption probabilities and applications. *Tr. of the American Mathematical Soc.*, 111(1):33–107, Apr. 1964.

[11] M. Kendall and A. Stuart. *The advanced theory of statistics*, volume 1, fourth ed. Charles Griffin and Co. Ltd., 1977.

[12] A. Mansour and C. Jutten. What should we say about the kurtosis? *IEEE Signal Processing Letters*, 6(12):321–322, Dec. 1999.

[13] G. Marsaglia, A.W. Marshall, and F. Proschan. Moment crossings as related to density crossings. *J. of the Royal Statistical Soc., Ser.B*, 27(1):91–93, 1965.

[14] J.J.A. Moors. A quantile alternative for kurtosis. *Statistician*, 37(1):25–32, 1988.

[15] K. Pearson. Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A rejoinder. *Biometrika*, 4(1-2):169–212, Jun. 1905.