Strong Sub- and Super-Gaussianity

Jason A. Palmer[1][*], Ken Kreutz-Delgado[2], and Scott Makeig[1]

[1] Swartz Center for Computational Neuroscience
[2] Department of Electrical and Computer Engineering
University of California San Diego, La Jolla, CA 92093
{jason,scott}@sccn.ucsd.edu, kreutz@ece.ucsd.edu

Abstract. We introduce the terms *strong sub- and super-Gaussianity* to refer to the previously introduced class of densities log-concave is x^2 and log-convex in x^2 respectively. We derive relationships among the various definitions of sub- and super-Gaussianity, and show that strong sub- and super-Gaussianity are related to the score function being star-shaped upward or downward with respect to the origin. We illustrate the definitions and results by extending a theorem of Benveniste, Goursat, and Ruget on uniqueness of separating local optima in ICA.

1 Introduction

In their seminal work on blind deconvolution, Benveniste, Goursat, and Ruget [3] proposed a definition of sub- and super-Gaussianity, and used it to derive conditions for blind identifiability of the unmixing deconvolutive system by minimization of the expected value of certain classes of functions. Unlike the Cardoso-Amari stability conditions [1, 7], which are local stability conditions for separating solutions based on second derivatives, the BGR conditions based on super- and sub-Gaussianity give conditions not only for stability, but also for uniqueness of separating local optima, and are based on first derivatives, not requiring finite curvature. While their work is credited with being fundamental [10, 8], the BGR definition does not seem to have been employed much in subsequent development in the ICA/BSS and latent variable communities, though the definition of super-Gaussianity turns out to coincide with that used to guarantee monotonicity in the sparse solution of underdetermined systems [15]. In this paper, we attempt to illuminate the BGR concept of sub- and super-Gaussianity, which we refer to as *strong sub-* and *super-Gaussianity*, by situating the definition in a nested hierarchy that includes the more commonly used characterizations such as kurtosis, and Gaussian density crossings. We also extend the class of functions that lead to unique globally optimal separating solutions for super-Gaussian sources to the more naturally related class of star-shaped functions.

2 Definitions and Set inclusion relationships

Qualitatively, super-Gaussianity may be thought of as implying *both* a sharper peak and heavier tail than the corresponding Gaussian, while sub-Gaussian densities have

flatter peaks, "heavier shoulders", and lighter tails. Quantitatively, we shall consider in particular the following four definitions of the classes of sub- and super-Gaussian densities.

1. *Fourth-order cumulant.* Let K_+ to be the set of symmetric densities with positive or infinite fourth-order cumulant, or excess kurtosis. Let K_- be the set of densities with negative fourth-order cumulant.
2. *Density crossing inequalities.* Let DC_+ be the set of symmetric, finite variance densities that cross a Gaussian density of equal variance exactly four times, with higher density at the origin and on the tails. Let DC_- be defined similarly, with lower density at the origin and on the tails.
3. *Strong sub-/super-Gaussianity.* Let SS_+ be the set of symmetric densities, $p(x)$, such that $\log p(\sqrt{x})$ is convex on $(0, \infty)$. or equivalently densities of the form $\exp(-f(x))$ such that $f'(x)/x$ is non-increasing on $(0, \infty)$. Let SS_- be the set of symmetric densities with $\log p(\sqrt{x})$ concave on $(0, \infty)$, or equivalently $p(x) = \exp(-f(x))$ with $f'(x)/x$ non-decreasing on $(0, \infty)$.
4. *Convexity of score function.* Let SC_+ be the set of symmetric densities $p(x) = \exp(-f(x))$ such that $f'(x)$ is concave on $(0, \infty)$. Let SC_- be the set of densities with $f'(x)$ convex on $(0, \infty)$.

Consider a random variable X with probability density $p(x)$, mean $\mu = E\{X\}$, and variance $\sigma^2 = E\{(X - \mu)^2\}$. We can define a measure based on the fourth moment that is invariant to changes in mean and scale, sometimes called the normalized or standardized fourth moment, or kurtosis, $\kappa = E\{(X-\mu)^4\}/\sigma^4$. Employing cumulants, we find that the first and second cumulants equal the mean and variance, and the fourth cumulant, γ, is equal to $\kappa - 3$. The fourth cumulant is sometimes called the excess kurtosis, since the kurtosis of the standard Normal density is 3. The most commonly used definition of sub- and super-Gaussianity involves the sign of the fourth cumulant, i.e. the kurtosis relative to a Gaussian of equal variance. If the kurtosis exceeds that of the Gaussian, then X, or its density $p(x)$, is said to be super-Gaussian. Likewise if γ is negative, or $\kappa < 3$, then X is said to be sub-Gaussian.

2.1 Density Crossings and Karlin's theory

Let $q(x)$ be a Gaussian density with variance equal to that of the symmetric density $p(x)$. According to [9], the proposition was known since R. A. Fisher that any symmetric density which has sharper peak and heavier tails at unit variance, defined by $p(x)$ crossing $q(x)$ four times, with $p(x) > q(x)$ near $x = 0$, and $p(x) > q(x)$ as $|x| \to \infty$, will have positive fourth cumulant. Finucan [9] proves this proposition.[3] The relationship between density crossings and moments in clarified by the work of S. Karlin, which we shall briefly describe below.

Let $p(x)$ and $q(x)$ be symmetric probability densities. We first note that any symmetric function $\psi(x)$ that is increasing on $(0, \infty)$ can be used as a normalizing condition, $\int \psi(x)p(x)dx = \int \psi(x)q(x)dx$ in the definition of a class of moment-based sub- and super-Gaussianity. We employ the two sign change form of Lemma A of [11].

---

[3] In [13] the authors prove a similar theorem, where a concept of "over-Gaussianity" is defined using the criterion of heavier tail only.

Lemma 1 *If $\int \psi(x)p(x)dx = \int \psi(x)q(x)dx$, with $\psi(x)$ symmetric and increasing on $(0, \infty)$, then $p(x)$ and $q(x)$ cross each other at least four times, i.e. $p(x) - q(x)$ has at least two sign changes on $(0, \infty)$.*

We say that $\varphi(x)$ is *convex with respect to* $\psi(x)$ on (a, b) if,

$$\begin{vmatrix} 1 & \psi(x_1) & \varphi(x_1) \\ 1 & \psi(x_2) & \varphi(x_2) \\ 1 & \psi(x_3) & \varphi(x_3) \end{vmatrix} > 0, \quad a \le x_1 < x_2 < x_3 \le b$$

Ordinary convexity is obtained for $\psi(x) = x$. Just as a convex function can be intersected by a linear function at most two times, if $\varphi(x)$ is convex with respect to $\psi(x)$, then $\varphi_{\alpha,\beta}(x) = \alpha\psi(x) + \beta$ and $\psi_{\gamma,\delta}(x) = \gamma\varphi(x) + \delta$ can intersect at most two times for all $\alpha, \beta, \gamma, \delta$. And when $\alpha, \gamma > 0$, and $\varphi_{\alpha,\beta}(x)$ and $\psi_{\gamma,\delta}(x)$ intersect two times on (a, b), then $\varphi_{\alpha,\beta}(x) > \psi_{\gamma,\delta}(x)$ in a neighborhood of b. If $\varphi(x)$ is convex with respect to $\psi(x)$, we also say that $\psi(x)$ is *concave with respect to* $\varphi(x)$. It is clear from the definition that $\varphi(x)$ is convex with respect to $\psi(x)$ on (a, b) if and only if $\varphi(\psi^{-1}(x))$ is convex on $(\psi(a), \psi(b))$.

Karlin and Novikoff's Lemma B states.

Lemma 2 *If $\int \psi(x)p(x)dx = \int \psi(x)q(x)dx$, with $\psi(x)$ symmetric and increasing on $(0, \infty)$, and $p(x)$ and $q(x)$ intersect exactly two times on $[0, \infty)$, then $\int \varphi(x)p(x)dx \ge \int \varphi(x)q(x)dx$ for all $\varphi(x)$ convex with respect to $\psi(x)$.*

Taking $q(x)$ to be Gaussian, $\psi(x) = x^2$, and $\varphi(x) = x^4$ yields the Finucan-Fisher theorem. Taking $q(x)$ Gaussian and $\psi(x) = |x|$ in Thm. 1 implies that $E\{\varphi(X)\} \ge E\{\varphi(Z)\}$ for Gaussian Z and strongly super-Gaussian X satisfying $E\{|X|\} = E\{|Z|\}$, allowing the moment-based definition to extend to random variables which do not have finite variance.

2.2 Strong sub- and super-Gaussianity

If we assume that $\log p(x)$ is convex with respect to $\log q(x)$, or vice-versa, then the normalized versions of $p(x)$ and $q(x)$, which intersect at least two times on $(0, \infty)$, must in fact intersect exactly two times, since otherwise $\log p(x)$ and $\log q(x)$ would also intersect more than two times contradicting the relative convexity of $\log p(x)$ and $\log q(x)$. We thus have the following theorem.

Theorem 1 *Let $p(x)$ and $q(x)$ be unimodal, symmetric probability densities, and let $\int \psi(x)p(x)dx = \int \psi(x)q(x)dx$, with $\psi(x)$ increasing on $[0, \infty)$. Let $\varphi(x)$ be convex with respect to $\psi(x)$ on $[0, \infty)$. Then if $\log p(x)$ is convex with respect to $\log q(x)$, then $\int \varphi(x)p(x)dx \ge \int \varphi(x)q(x)dx$.*

If we take $q(x)$ to be Gaussian, then the condition that $\log p(x)$ be convex (concave) with respect to $\log q(x)$ on $[0, \infty)$ is equivalent to saying that $-\log p(x)$ is convex (concave) with respect to x^2, or $-\log p(\sqrt{x})$ is convex (concave) on $(0, \infty)$. This condition forms the definition of strong sub- and super-Gaussianity proposed by [4].

Definition 1 *A symmetric probability density $p(x)$ is **strongly super-gaussian** if $p(\sqrt{x})$ is log-convex on $(0, \infty)$, and **strongly sub-gaussian** if $p(\sqrt{x})$ is log-concave on $(0, \infty)$.*

The requirement that $f(x) \triangleq -\log p(x)$ be convex or concave with respect to x^2 can be expressed differentially in the same way that ordinary convexity can. Letting x_1 tend to x_2 in the determinantal definition of relative convexity yields the first order condition that $\varphi(x)$ is convex with respect to $\psi(x)$ if $\varphi'(x)/\psi'(x)$ is non-decreasing. Thus the differential condition for $f(x)$ convex (concave) with respect to x^2 on (a, b) is that,

$$f'(x)/x \quad \text{non-decreasing (non-increasing) on } (a, b)$$

This condition on the first derivative of $f(x)$ can be expressed more intuitively using the concept of *star-shaped functions* [6]. Recall that a set S is star-shaped with respect to a point $x \in S$ if every point $z \in S$ can be "seen" by x, i.e. for every point y on the line joining x and z, $y \in S$. Classically, a function is said to be star-shaped if the epigraph of $f(x)$ (i.e. points (x, y) such that $y \geq f(x)$) is star-shaped with respect to the origin. This definition seems to be overly restrictive. We extend the classical definition of star-shaped functions to include *star-shaped upward* and *star-shaped downward* functions.

Definition 2 *A function $f(x)$ is **star-shaped upward** if the epigraph of $f(x)$ is star-shaped with respect to the origin. $f(x)$ is **star-shaped downward** if the hypograph (i.e. (x, y) such that $y \leq f(x)$) is star-shaped with respect to the origin.*

Geometrically, if a line segment is drawn joining the origin to the point $(x, f(x))$, the slope is given by $f(x)/x$. If $f(x)$ is star-shaped upward, then by definition for $0 \leq \alpha \leq 1$, we have $f(\alpha x) \leq \alpha f(x)$, or $f(\alpha x)/(\alpha x) \leq f(x)/x$. Thus an equivalent defining criterion of star-shaped upward (downward) functions is that $f(x)/x$ be non-decreasing (non-increasing). We may thus formulate a natural geometric definition of strong sub- and super-Gaussianity in terms of the (location) score function $\psi(x) \triangleq (d/dx) \log p(x)$: *A symmetric density $p(x)$ is strongly super-Gaussian (sub-Gaussian) if and only if the score function $\psi(x)$ is star-shaped upward (downward).*

We note from the functions $\exp(-|x|^p)$, which are super-Gaussian for $p < 2$ and sub-Gaussian for $p > 2$ (by all definitions) that the score function in this case is convex in the super-Gaussian case, and concave in the sub-Gaussian case. We might thus define a (somewhat strict) form of sub- and super-Gaussianity accordingly as the score function is convex or concave on $(0, \infty)$. Since functions that are non-positive and convex are also star-shaped upward, and functions that are non-positive and concave are star-shaped downward, we see that the score convex and concave classes SC_+ and SC_- are strict subsets of strong super- and sub-Gaussian classes SS_+ and SS_- respectively.

Putting these relationships together, we have the following.

Theorem 2 *We have,*

$$K_+ \supset DC_+ \supset SS_+ \supset SC_+$$

and similarly for the sub-Gaussian sets, with the unique intersection of all sub- and super-Gaussian sets being the Gaussian density, as illustrated in the Venn diagram in Figure 2a.
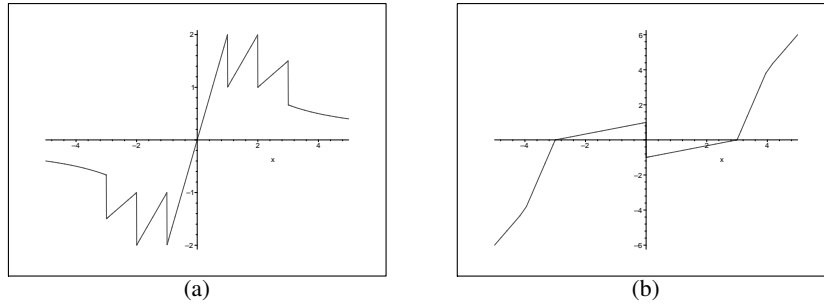
$$\text{(a)} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(b)}$$

Fig. 1. Star-shaped functions. (a) Star-shaped downward function corresponding to a strong super-Gaussian density. (b) Star-shaped upward function corresponding to a (non-unimodal) strongly sub-Gaussian density. In each function, the points on the graph are "visible" to the origin (joined by a line segment without intersecting the graph). Star-shaped downward functions must be odd and non-negative on $(0, \infty)$, so that strong super-Gaussians must be unimodal. Star-shaped downward functions may be negative near the origin, and thus strong sub-Gaussians can be bimodal. Neither require continuity, monotonicity, differentiability, or convexity or concavity.

2.3 Scale Mixture Representations

Gaussian scale mixtures constitute a large class of super-Gaussian densities, closed under convolution, and in fact are the uniformly convergent limit of a sequence functions that may be called n-times monotone [17].

Definition 3 $p(x)$ *is **n-times monotone** on* (a, b) *if* $(-1)^k f^{(k)}(x)$ *is non-negative, non-increasing and convex on* (a, b) *for* $k = 0, 1, 2, \ldots, n - 2$.

Thus n-times monotone functions have derivatives of alternating sign up to order n when sufficiently differentiable. If this holds for all n, then we have *complete monotonicity* [16]: A function $f(x)$ is *completely monotonic* on (a, b) if $(-1)^n f^{(n)}(x) \geq 0$, $n = 0, 1, \ldots$ for every $x \in (a, b)$. Bernstein's theorem [16, Thm. 12b] states: A necessary and sufficient condition that $p(x)$ should be completely monotonic on $(0, \infty)$ is that $p(x) = \int_0^\infty e^{-tx} d\alpha(t)$, where $\alpha(t)$ is non-decreasing on $(0, \infty)$. Similarly, the following theorem of Williamson [17] states the conditions for n-times monotonicity.

Theorem 3 *A necessary and sufficient condition that* $p(x)$ *should be* n-*times monotonic on* $(0, \infty)$ *is that* $p(x) = \int_0^\infty (1 - tx)_+^{n-1} d\alpha(t)$, *where* $\alpha(t)$ *is non-decreasing and bounded below on* $(0, \infty)$.

We define the class $M^2(n)$ to be the set of functions n-times monotone in x^2:

Definition 4 *The class of functions* $M^2(n)$ *consists of all functions of the form,*

$$p(x) = \int_0^\infty (1 - tx^2)_+^{n-1} d\alpha(t)$$

where $\alpha(t)$ *is non-decreasing and bounded below on* $(0, \infty)$.
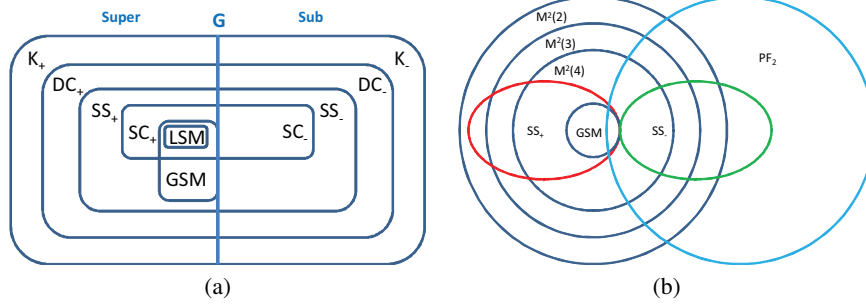
Fig. 2. (a) Venn diagram showing set inclusions among sub- and super-Gaussian desnties. Gaussian (G) is represented by the central line, and lies in the intersection of all sets except LSM. K_+ and K_- are defined by moment-based criteria, e.g. excess kurtosis, and form the outermost shell. The set of densities satisfying crossing properties with respect to a normalized Gaussian density, DC_+ and DC_- form subsets of K_+ and K_- respectively, as shown by Karlin's theory. The set of strong sub- and super-Gaussians, SS_- and SS_+ are again strict subsets of the density crossing classes, and the score concave and convex classes, SC_+ and SC_- are strict subsets of SS_+ and $SS_=$. The class of Gaussian scale mixtures, GSM, is a strict subset of SS_+, and strictly contains the set of Laplacian scale mixtures, LSM, which is itself strictly contained in the class of score convex densities, SC_+. (b) Venn diagram showing set inclusions among strong sub- and super-Gaussian classes, various scale mixtures, and the log-concave class. The class of densities n-times monotone in x^2, $M^2(n)$, tends to GSM as $n \to \infty$. The class of strong super-Gaussians, SS_+ is strictly contained in $M^2(2)$, but not in $M^2(n)$ for $n \geq 3$. The set of unimodal strong sub-Gaussians is strictly contained in the log-concave class, PF_2 (Polya frequency functions of order 2), which however also contains densities that are in SS_+.

By Williamson's theorem, this is equivalent to $(-1)^k (d/dx)^k p(\sqrt{x})$ non-negative, non-increasing, and convex on $(0, \infty)$, for $k = 1, 2, \dots, n-2$. It is obvious from this that $M^2(n) \subset M^2(m)$ for $m < n$. By the Bernstein-Widder theorem, we have the result that [12, 2], a function $p(x)$ can be represented as a Gaussian scale mixture if and only if $p(\sqrt{x})$ is completely monotonic on $(0, \infty)$. Concavity of $-\log p(\sqrt{x})$ follows from the complete monotonicity of $p(\sqrt{x})$ since sums of log-convex functions are again log-convex [5, §3.5.2]. Thus, completely monotonic functions, being scale mixtures of the log-convex on $(0, \infty)$ function, $\exp(-x)$, are also log-convex on $(0, \infty)$. We thus have the following [14], *All Gaussian scale mixtures are strongly super-Gaussian, i.e.* $GSM \subset SS_+$. The inclusion is strict.

We also note that since strong super-Gaussianity of $p(x)$ implies that $p(\sqrt{x})$ is convex (and non-negative and non-increasing) on $(0, \infty)$, $p(x) \in M^2(2)$ by Williamson's theorem. Hence $SS_+ \subset M^2(2)$. $M^2(n)$ in fact includes sub- as well as super-Gaussian densities (the functions $(1-x^2)_+^n$ are all sub-Gaussian,) but the scale mixture representation using $M^2(2)$ will be useful when we consider the blind deconvolution theorem. It is unclear whether the class SS_+ admits an equivalent scale mixture representation, but we shall show in the next section that for $p(x) \in SS_+$, $\log p(x)$ can be represented by a scale mixture, and thus $p(x)$ can be represented by a type of product mixture.

We finally note that the classes $M^2(n)$, along with the limiting class GSM, are closed under convolution. The closure of GSM is straightforward, owing to the well-known closure of the Gaussian density under convolution. In fact, for X and Y Gaussian scale mixtures, $X = \xi_1^{1/2} Z_1$, $Y = \xi_2^{1/2} Z_2$ where ξ_1, ξ_2 are non-negative i.i.d, and Z_1 and Z_2 are i.i.d. Gaussian, we have $X + Y \overset{d}{=} (\xi_1 + \xi_2)^{1/2} Z$ [12]. The following are used in the sequel is stated without proof due to space constraints.

Theorem 4 *The classes $M^2(n)$ are closed under convolution. Furthermore, if $p(x) \in M^2(m)$ and $q(x) \in M^2(n)$, then $\int p(x-t)q(t)dt \in M^2(\min(m,n))$.*

Theorem 5 *If $\psi(x)$ is star-shaped downward, and $p(x) \in M^2(2)$, then $\int_{-\infty}^{\infty} p(x - t)\psi(t)dt$ is star-shaped downward.*

3 Uniqueness of Separating Local Optima in ICA

To illustrate the application of the definition of strong sub- and super-Gaussianity (the BGR definition) we prove the following theorem.

Theorem 6 *Let $\mathbf{x} = \mathbf{A}\mathbf{s}$ where s_i are i.i.d., zero mean, and strongly super-Gaussian. If $\varphi'(y)$ is star-shaped downward, then,*

$$L(\mathbf{w}) = E\{\varphi(\mathbf{w}^T\mathbf{x})\}$$

has a local minimum $\mathbf{w}^*$ with $\|\mathbf{w}^*\| = 1$, if and only if $\mathbf{w}^{*T}\mathbf{A} = c\,\mathbf{e}_j^T$ for some j, i.e. $\mathbf{w}^{*T}$ is a row of the inverse of $\mathbf{A}$ up to scaling and permutation.

Proof. Define $\mathbf{c}^T = \mathbf{w}^T\mathbf{A}$. Suppose $c_i \neq 0$ and $c_j \neq 0$. Define $\tilde{\varphi}(y) \triangleq \int \varphi(u)h(y - u)du$, where $h(u)$ is the symmetric probability density function of $u = \sum_{k \neq i,j} c_k s_k$. We have $h(u) \in M^2(2)$ by Theorem 4, and,

$$E\{\varphi(\mathbf{w}^T\mathbf{x})\} = E\{\tilde{\varphi}(c_i s_i + c_j s_j)\}$$

with $\tilde{\varphi}$ star-shaped downward by Theorem 5. Let $y = R\cos(\theta)s_i + R\sin(\theta)s_j$, where $R = 1 - \sum_{k \neq i,j} w_k^2$. Consider the function, $(d/d\theta)\, E\{\varphi(y)\,;\,\theta\}$. From [3] we have,

$$
\begin{aligned}
(d/d\theta)\, E\{\varphi(y)\,;\,\theta\} = {} & \int_0^{\infty} r^3 \int_0^{\pi/4} \exp\left(-g(r\cos(\phi)) - g(r\sin(\phi))\right) \sin(2\phi) \\
& \times \left(\frac{g'(r\cos(\phi))}{r\cos(\phi)} - \frac{g'(r\sin(\phi))}{r\sin(\phi)}\right)\left(b(\phi - \theta) - b(\phi + \theta)\right)d\phi\, dr \quad (1)
\end{aligned}
$$

where $b(\phi) \triangleq \varphi(r\cos(\phi)) + \varphi(r\sin(\phi))$. Note that the function $b(\phi)$ satisfies the symmetries, $b(\phi) = b(-\phi) = b(\pi/2 - \phi)$ by the symmetry in φ. Also, we have,

$$b'(\phi) = \tfrac{1}{2}r^2 \sin(2\phi)\left(\frac{\varphi'(r\sin(\phi))}{r\sin(\phi)} - \frac{\varphi'(r\cos(\phi))}{r\cos(\phi)}\right)$$

Since $g'(x)/x$ is non-decreasing (sources are strongly super-Gaussian), and $\varphi'(x)/x$ is non-decreasing ($\varphi'(x)$ is star-shaped downward) on $(0,\infty)$, and both are increasing on a common interval of nonzero measure, we have that the integrand of (1) is positive, and thus,

$$(d/d\theta)\, E\{\varphi(y)\,;\,\theta\} > 0, \quad \theta \in (0,\pi/4)$$

and $\mathbf{w}$ cannot be a local optimum, and one of c_i or c_j must be zero at all local optima.

4   Conclusion

We defined strong sub- and super-Gaussianity, and derived relationships among various definition of sub- and super-Gaussianity. We extended a result of Benveniste, Goursat, and Ruget [3] to include star-shaped downward non-linearities. Similar results can be derived for sub-Gaussians and star-shaped upward functions, but under more restrictive conditions due to limitations on closure under convolution.

References

1. S.-I. Amari, T.-P. Chen, and A. Cichocki. Stability analysis of learning algorithms for blind source separation. *Neural Networks*, 10(8):1345–1351, 1997.
2. D. F. Andrews and C. L. Mallows. Scale mixtures of normal distributions. *J. Roy. Statist. Soc. Ser. B*, 36:99–102, 1974.
3. A. Benveniste, M. Goursat, and G. Ruget. Robust identification of a nonminimum phase system. *IEEE Transactions on Automatic Control*, 25(3):385–399, 1980.
4. A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*. Springer-Verlag, 1990.
5. S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
6. A. M. Bruckner and E. Ostrow. Some function classes related to the class of convex functions. *Pacific J. Math.*, 12(4):1203–1215, 1962.
7. J.-F. Cardoso and B. H. Laheld. Equivariant adaptive source separation. *IEEE. Trans. Sig. Proc.*, 44(12):3017–3030, 1996.
8. P. Comon and C. Jutten, editors. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Elsevier, 2010.
9. H. M. Finucan. A note on kurtosis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(1):111–112, 1964.
10. S. Haykin. *Neural Networks: a comprehensive foundation*. Prentice Hall, 1999.
11. S. Karlin and A. Novikoff. Generalized convex inequalities. *Pacific J. Math.*, 13:1251–1279, 1963.
12. J. Keilson and F. W. Steutel. Mixtures of distributions, moment inequalities, and measures of exponentiality and Normality. *The Annals of Probability*, 2:112–130, 1974.
13. A. Mansour and C. Jutten. What should we say about the kurtosis? *IEEE Signal Processing Letters*, 6(12):321–322, 1999.
14. J. A. Palmer, K. Kreutz-Delgado, D. P. Wipf, and B. D. Rao. Variational EM algorithms for non-gaussian latent variable models. In *Advances in Neural Information Processing Systems*. MIT Press, 2006.
15. B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado. Subset selection in noise based on diversity measure minimization. *IEEE Trans. Signal Processing*, 51(3), 2003.
16. D. V. Widder. *The Laplace Transform*. Princeton University Press, 1946.
17. R. Williamson. Multiply monotone functions and their Laplace transforms. *Duke Math. J.*, 23:189–207, 1956.