Stable Symmetric Distributions and Their Role in the Signal Separation Problem

Ricky Der

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

February 2003

Stable Symmetric Distributions and Their Role in the Signal Separation Problem

1 Introduction

Stable distributions have a long history in the subject of probability. They form a subset of the class of so-called "infinitely-divisible" distributions—a class of characteristic functions at the heart of general central limit theory. Roughly speaking, the stable distributions are those which are closed under the formation of linear combinations, or more generally, affine combinations. Thus linear mixtures of stable distributions result in distributions of the same type.

The subject of blind source separation is of a more recent provenance. It is the problem of determining source signals when only their mixtures are observed. There has recently been an explosion of interest in the area with the emergence of relatively successful separation algorithms—at least under artificial mixing conditions. The resulting paradigm has a distinctly probabilistic flavor; the key concept amounts to equating *separation* with *statistical independence* of the sources. So much so that the approach is often called *Independent Component Analysis*.

The usual method of solution is as follows: the mixtures are fed through a un-mixing system T, as yet unknown. A cost function (also called a contrast), J, is chosen, measuring in some sense the inter-component dependence of the outputs of T, and ideally achieving an extremum when the outputs are independent. J is then extremised with respect to the unknown parameter T. Alternatively, an off-line procedure obtains empirical measurements for the cost-function in terms of block averages; one then solves algebraically for the mixing channel.

There are nearly as many proposed cost functions as researchers in the field. The most famous contrast is the Kullback-Leibler metric for measuring the deviation between the joint distribution of the outputs and some assumed source distribution. This metric is in some sense the "ideal" cost function from the perspective of Maximum-Likelihood estimation, and is thus statistically efficient in the Fisher sense. The drawback to the criterion is that source distributions must be known *a priori*—though not necessarily perfectly. Generally speaking, the blind source separation problem falls into the class of *semiparametric statistical problems* [1]. The main difficulty lies in the fact that one must estimate not only the unmixing matrix T, but also the source distributions F_s—a parameter of infinite dimension.

Thus contrasts involving only statistics of the mixtures have been invoked. These functions are usually heuristic and are not statistically efficient; but they suffice for many applications, offering efficient implementations without requiring the accurate knowledge of source distributions. Often these functions present some property of the *marginal* output distributions, as opposed to the joint distributions. We list from this class a few of the more prominent proposals:

1. $J = \sum_{i=1}^{N} D(p(u_i) \| p_G(u_i))$
2. $J = \sum_{i=1}^{N} {k_i}^2$
3. $J = \sum_{i=1}^{N} E[\log(\cosh u_i))]$

The first contrast is the sum of marginal output negentropies, suggested in [4], the second is the sum of marginal output kurtoses, due to [3], and the third is an "approximation" to negentropy, due to [5]. The telling characteristics of all three consist in the use of (1) marginal, as opposed to joint statistics, and (2) the use of *non-Gaussianity* as a measure. Indeed, the identification of non-Gaussian structure seems to be linked with the identification of higher-order (3rd and above) structure necessary for blind estimation [6]. There is, moreover, the heuristic idea that the mixing of independent sources drives the output distributions towards the Gaussian, and hence maximizing non-Gaussianity should lead to separation.

It is the aim of this report to disabuse such a notion. In particular, we exhibit a class of distributions which remain invariant under linear combinations—precisely the symmetric stable distributions—and hence exhibit no move towards Gaussianity when mixed. More generally, we show that any contrast which is a function of the marginal output distributions must be a constant under the usual optimization spread constraint. Hence measures such as negentropy and kurtosis cannot separate stable distributions. It will be of some comfort, however, to know that these distributions are mildly peculiar in that they must either have infinite variance, or be Gaussian. Thus marginal techniques are appropriate when the signals have finite energy.

The document is organized as follows: in Section 2 we develop the idea of stable symmetric distributions and give some elementary properties. In Section 3 we discuss its application to the blind source identification problem. A conclusion rounds the report.

2 Stable Distributions

In this section we make formal the notion of mixing invariance. It will be most convenient to work with characteristic functions, rather than probability distribution functions. The reader is reminded of the following definition:

Definition 1 *Let $F(x)$ be a probability distribution function. The characteristic function of F is defined as*

$$\Phi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x) \tag{1}$$

A characteristic function is always uniformly continuous and positive semi-definite—indeed these may be taken as *the* defining properties of a characteristic function.

We would like to find the set of characteristic functions which are closed in the sense that the linear combination $aX_1 + bX_2$ gives a random variable "distributed in the same way" as its components. The following makes the quoted term precise:

Definition 2 *Let $\Phi(t)$ be a characteristic function. We define the class C_Φ to be*

$$C_\Phi = \{\Phi(at) : a \in \mathbb{R}\} \tag{2}$$

The definition sets up equivalence classes on the set of characteristic functions; two characteristic functions belong to the same class if they possess the same "shape". A few examples: the class of zero-mean Gaussians, the class of Cauchy distributions, the class of Laplacian distributions.

Definition 3 *Let C be a class of characteristic functions defined as above. Then C is multiplicatively invariant, if, for any $\Phi_1(t), \Phi_2(t) \in C$, we have that $\Phi_1(t) \cdot \Phi_2(t) \in C$.*

It is easy to see that $C = C_{\Phi(t)}$ is multiplicatively invariant iff for every real a, b,

$$\Phi(at) \cdot \Phi(bt) = \Phi(ct) \tag{3}$$

for some real number c. The number c, if it exists, must be unique up to sign or the characteristic function trivial ($\Phi(t) = 1$). It is also obvious that multiplicative invariance of characteristic functions corresponds to convolutional invariance of the respective equivalence class of distributions. We now establish some elementary properties of multiplicatively invariant characteristic functions:

Theorem 1 *Let C_Φ be multiplicatively invariant. If $\Phi(t) \in C_\Phi$, then*

1. *For any set of real numbers $a_1, \ldots, a_n$,*

$$\Phi(a_1 t) \cdot \Phi(a_2 t) \cdot \cdots \cdot \Phi(a_n t) = \Phi(ct) \tag{4}$$

for some real number c.

2. *$\Phi(t)$ is real and even.*

3. *$\Phi(t)$ is infinitely divisible.*

Proof: The first property is immediate by induction. To prove the second property, note that $\Phi(t) \cdot \Phi(-t) = \Phi(ct)$ for some c. Since $\Phi(-t) = \overline{\Phi(t)}$, it follows that $|\Phi(t)|^2 = \Phi(ct)$ and hence $\Phi(t)$ is real. But then $\Phi(-t) = \overline{\Phi(t)} = \Phi(t)$. The last property follows by setting $a_i = 1$ in the first property, hence obtaining $\Phi(t) = [\Phi(t/c_n)]^n$. ∎

A corollary of property 2 is that convolutionally invariant distributions must be symmetric. We also remark here that the converse to property 3 is not true. For example, the Poisson class given by a fixed $\lambda > 0$, with $C_\Phi = \{e^{(\lambda(e^{iat}-1))}, a \in \mathbb{R}\}$, is not convolutionally-invariant. However, every function in this class is infinitely divisible since $e^{(\frac{\lambda}{n}(e^{iat}-1))}$ is the characteristic function for aX, where X is Poisson distributed with parameter $\frac{\lambda}{n}$. Thus the set of all (equivalence classes of) convolutionally invariant distributions forms a proper subset of the set of all infinitely divisible distributions.

In the probability literature, any characteristic function for which there exists constants b, c satisfying $\Phi(a_1 t) \cdot \Phi(a_2 t) = e^{ibt}\Phi(ct)$, for any constants a_1, a_2 is called a *Lévy-stable* distribution. Comparison with (3) shows that our multiplicatively-invariant characteristic functions are a special case of these stable distributions, with $b = 0$. In fact, the class of stable distributions could be obtained if we enlarged our equivalence classes to contain *affine* forms of the type $cX + b$. We restrict our attention to the special case because it is the form applicable to the noiseless source separation problem. However, in recognition of its parenthood, and in light of property 2, we term our class of multiplicatively-invariant distributions the *stable symmetric distributions*.

2.1 Stable Symmetric Distributions

Assume that $\Phi(t)$ is not trivial. Then the number c in (3) is a function of a, b, and the characterizing functional equation becomes

$$\Phi(at) \cdot \Phi(bt) = \Phi(g(a, b) \cdot t) \tag{5}$$

Since Φ is infinitely divisible, it never vanishes [2], and hence must be of the form $\Phi(t) = e^{\varphi(t)}$ for a suitable function φ. Examination of the functional equation suggests solutions of the type

$$\Phi(t) = e^{-\alpha|t|^k}. \tag{6}$$

For positive parameters α and k, each $\Phi(t)$ is Lebesgue-integrable, hence the Fourier transform exists and is given by

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}e^{-\alpha|t|^k}dt \tag{7}$$

The question of whether $\Phi(t)$ defines a proper characteristic function can now be answered by examining whether $f(x) \geq 0$ for all real x. Lévy has shown that this is true only for $0 \leq k \leq 2$. A plot of the Fourier transform of $\Phi(t)$ for varying values of k, computed via numerical integration, may be found in Fig. 1. Assuming the range $k \in [0, 2]$, the normalisation $1 = \Phi(0) = \int_{-\infty}^{\infty} dF$ verifies that $f(x)$ is a probability density. One easily checks that (5) is also satisfied with $g(a, b) = (|a|^k + |b|^k)^{1/k}$.

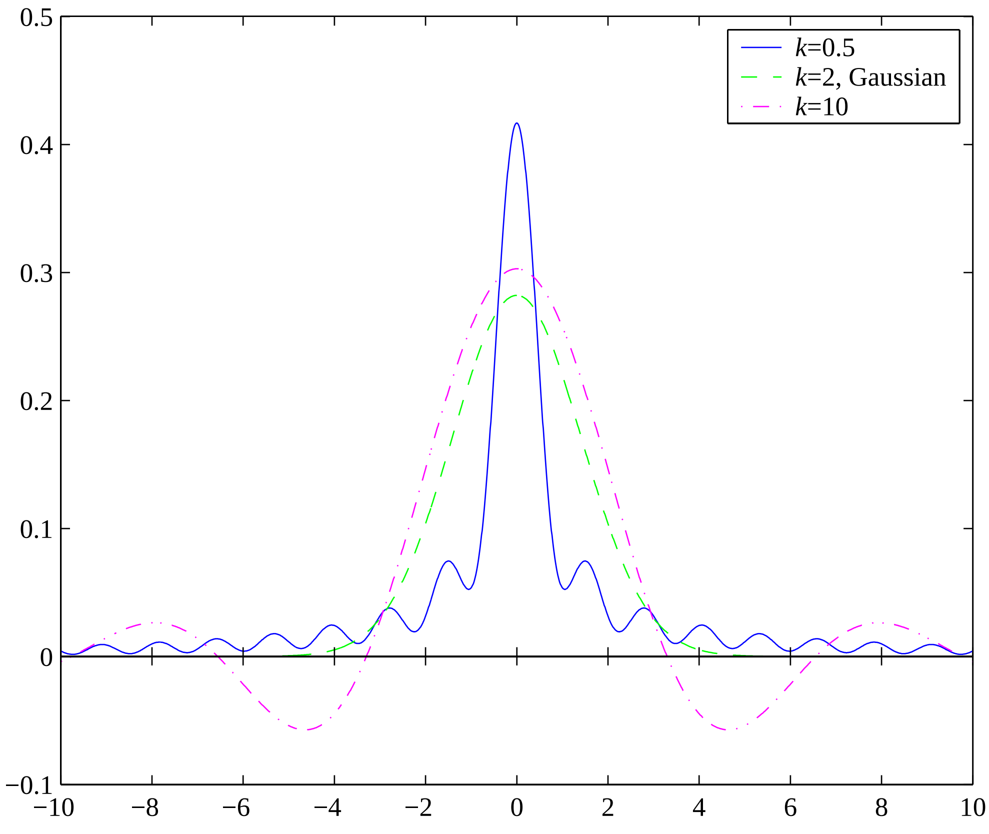

Fig. 1 Fourier transform of $\Phi(t) = e^{-|t|^k}$

The archetype of (6) is a generalisation of the Gaussian characteristic function. The next theorem shows that nearly all of them have infinite energy.

Theorem 2 *Let* $\Phi(t) = e^{-\alpha|t|^k}$ *have finite variance. Then* $\Phi(t)$ *is of Gaussian class.*

Proof: Assume $k < 2$ in (6) and suppose, for a contradiction, that $\Phi(t)$ has finite variance. Then it has finite second moment and mean, and by a well-known theorem in probability [2], $\Phi(t)$ is twice differentiable with $\Phi''(0) = -E[X^2]$, where X is a random variable with characteristic function Φ. A simple computation gives

$$\Phi''(t) = -\alpha k e^{-\alpha|t|^k} \cdot \left[(k-1)|t|^{k-2} + \text{sgn}(t) \cdot |t|^{2(k-1)}\right] \tag{8}$$

But since $k < 2$, $\Phi''(0)$ is infinite and hence the variance is infinite, obtaining the contradiction. ■

Thus the only stable symmetric distribution where the variance exists is the Gaussian. This property may seem somewhat pathological to the engineer; such distributions, however, have proven very useful for the modelling of noise [7].

3 The Source Separation Problem

We are now in a position to see the special role that stable distributions play in the source separation problem. Let us first give the formulation.

Let $\mathbf{s} = [s_1, s_2, \ldots, s_n]^T$ be a vector of independent source random variables. Let $\mathbf{A}$ be an $n \times n$ non-singular matrix. We define the mixture vector to be

$$\mathbf{x} = \mathbf{A}\mathbf{s}. \tag{9}$$

The goal is to find a matrix $\mathbf{W}$, such that $\mathbf{y} = \mathbf{W}\mathbf{x}$ is a vector of independent variables, and to do so only with knowledge of the mixtures $\mathbf{x}$. For sources of which at most one is Gaussian distributed, this is equivalent to determining a $\mathbf{W}$ such that $\mathbf{W}\mathbf{A} = \mathbf{P}$, where $\mathbf{P}$ is a permuted and row-scaled version of the identity matrix. As posed here, the solution is not unique; usually a constraint is placed on the spreads (e.g. variances) of the outputs, which reduces the indeterminacy to that of permutation.

Now suppose that $\mathbf{s}$ is drawn from a stable symmetric class C_Φ. Writing $x_i = \sum_{j=1}^n a_{ij} s_j$, we have the following formula for the characteristic functions:

$$\Phi_{x_i}(t) = \prod_{j=1}^{n} \Phi_{s_j}(a_{ij} t) \tag{10}$$

Since each s_j has a stable symmetric distribution, from property 1 of the first theorem there exists a function $g_i(a_1, \ldots, a_n)$ (inductively generalised from (5)), such that

$$\Phi_{x_i}(t) = \Phi(g_i(a_1, \ldots, a_n) \cdot t) \tag{11}$$

Thus the mixtures each are distributed from class C_Φ. In light of Theorem 2, variances may not exist, so some more general spread constraint must be used. The most natural measure of spread is given by the function $c = |g(a_1, \ldots, a_n)|$. If variances exist, the criterion reduces to the variance constraint $c = a_1^2 + \cdots + a_n^2$. Under this restriction, $g(\cdot)$ is unique up to sign; this is irrelevant, however, since multiplicatively invariant characteristic functions are even. The conclusion is that the marginal distributions of the mixtures must be fixed over all matrices $\mathbf{A}$ satisfying the spread constraint. This gives us our final theorem:

Theorem 3 *Let $\boldsymbol{s}$ be a vector of n independent random variables drawn from a stable symmetric class C_Φ, with associated spread function $g(a_1, \ldots, a_n)$. Let*

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} \tag{12}$$

where $\boldsymbol{A}$ is a real matrix. Define K to be the set of all $n \times n$ real matrices $[a_{i,j}]$ such that, for each i, $|g(a_{i,1}, \ldots, a_{i,n})| = \text{constant}$. If $J(x_1, x_2, \ldots, x_n)$ is a function of only the marginal distributions of $\boldsymbol{x}$, then $J(\boldsymbol{x}(\boldsymbol{A}))$ is a constant on the set K.

4 Conclusion

We have shown that any cost function extremising any property of the marginal distributions of the mixtures is a constant for stable symmetric sources, and thus cannot solve the blind source separation problem in full generality. These distributions are non-pathological, but have infinite energy. The noticeable exception is the Gaussian distribution, for which the separation problem is inherently undetermined. For finite variance signals, the use of marginal statistics is justified.

References

[1] S. Amari, J.-F. Cardoso. Blind Source Separation—Semiparametric Statistical Approach. *IEEE Transactions on Signal Processing*, 45(11): 2692–2700, 1997.

[2] C. Burrill, *Measure, Integration and Probability.* New York, USA: McGraw-Hill, 1972.

[3] P. Comon. Independent Component Analysis—a New Concept? *Signal Processing*, 36(3): 287–314, 1994.

[4] M. Girolami, C. Fyfe. Extraction of independent signal sources using a deflationary exploratory projection pursuit network with lateral inhibition. *IEEE Proceedings on Vision, Image and Signal Processing Journal*, 14(5): 299–306, 1997.

[5] A. Hyvärinen. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks*, 10(3): 626–634, 1999.

[6] T. Lee. *Independent Component Analysis.* Boston, USA: Kluwer Academic Press, 1998.

[7] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications.* New York, USA: Wiley, 1995.