# Zero-Order Statistics: A Mathematical Framework for the Processing and Characterization of Very Impulsive Signals

Juan G. Gonzalez, José L. Paredes, *Member, IEEE*, and Gonzalo R. Arce, *Fellow, IEEE*

***Abstract*—Impulsive or heavy-tailed processes with infinite variance appear naturally in a variety of practical problems that include wireless communications, teletraffic, hydrology, geology, and economics. Most signal processing and statistical methods available in the literature have been designed under the assumption that the processes possess finite variance, and they usually break down in the presence of infinite variance. Although methods based on fractional lower-order statistics (FLOS) have proven successful in dealing with infinite variance processes, they fail in general when the noise distribution has *very heavy* algebraic tails. In this paper, we introduce a new approach to statistical moment characterization which is well defined over all processes with algebraic or lighter tails. Unlike FLOS, these *zero-order statistics (ZOS)*, as we will call them, provide a common ground for the analysis of basically *any* distribution of practical use known today. Three new parameters, namely the geometric power, the zero-order location and the zero-order dispersion, constitute the foundation of ZOS. They play roles similar to those played by the power, the expected value and the standard deviation, in the theory of second-order processes. We analyze the properties of the new parameters, and derive a ZOS framework for location estimation that gives rise to a novel mode-type estimator with important optimality properties under very impulsive noise. Several simulations are presented to illustrate the potential of ZOS methods.**

***Index Terms*—Algebraic tails, alpha-stable distributions, fractional lower-order statistics, geometric power, heavy tails, logarithmic order processes, robust signal processing, very impulsive processes, zero-order statistics (ZOS).**

## I. INTRODUCTION

SECOND-ORDER processes have been historically the main subject of study in statistical signal processing. Second-order-based estimation techniques are commonly recognized as the natural tools to be used in the presence of Gaussian noise. Research efforts on higher-order statistics (HOS) have led to the development of improved estimation algorithms for non-Gaussian environments. This work has been based on the assumptions that second- and higher-order statistics of the processes exist and are finite [1], [2]. Important non-Gaussian impulsive processes are found in a variety of practical problems that include telecommunications, teletraffic, hydrology, geology, and economics. These processes can be efficiently modeled by heavy-tailed distributions with infinite variance, for which neither the classical second-order theory nor the theory of HOS are well defined [3], [4].

It has been shown repeatedly in the literature that infinite-variance processes that appear in practice are well modeled by probability distributions with algebraic tails, i.e., random variables for which,[1]

$$\Pr(|X| > x) \sim cx^{-\alpha} \tag{1}$$

for some fixed $c, \alpha > 0$. Examples of such noise processes include those modeled by $\alpha$-stable distributions [4], [5], Hall's generalized t-model [6], the generalized Cauchy model [7], [8], and the Pareto distribution [9], [10]. The tail-heaviness of these distributions is mainly determined by the tail constant $\alpha$, with increased impulsiveness corresponding to smaller values of $\alpha$.

Algebraic-tailed random variables exhibit finite absolute moments for orders less than $\alpha$; i.e., $\mathbf{E}|X|^p < \infty$ if $p < \alpha$. Conversely, if $p \geq \alpha$, the absolute moments become infinite, and thus unsuitable for statistical characterization. When $\alpha < 2$, the processes have infinite variance, and the standard second– or higher-order statistics cannot be effectively used. Alternative attempts to characterize the behavior of impulsive signals in this scenario have relied on *fractional lower-order statistics* (FLOS) in the context of non-Gaussian $\alpha$-stable distributions $(\alpha < 2)$. Here, given a fixed $\alpha$, appropriate choices of $p$ in the interval $(0, \alpha)$ can lead to useful characterizations of the process structure [4], [11], [12].

While FLOS are useful in the characterization and processing of impulsive signals, they do not provide a universal framework for the characterization of algebraic-tailed processes: for a given $p > 0$, there is a class of processes (those with $\alpha \leq p$), for which the associated FLOS do not exist. Also, restricting the values of $p$ to the valid interval $(0, \alpha)$ requires either *a priori* knowledge of $\alpha$ or a numerical procedure to estimate it. *A priori* knowledge is unrealistic in many practical applications, and numerical estimation may be inexact and/or computationally expensive.

Manuscript received May 16, 2005; revised October 6, 2005. This work was supported in part by the New Technologies Division of Intellep, LLC, Weston, FL 33326 USA, the National Science Foundation under the grant ANI-0312851, through collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-1-2-0011 and by the University of Los Andes Council of Scientific, Humanistic and Technological Development (CDCHT-ULA) by means of the project I-768-04-02-B. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Vitor Heloiz Nascimento.

J. G. Gonzalez is with the Intellep, LLC, Weston, FL 33326 USA (e-mail: research@intellep.com).

J. L. Paredes is wih the Electrical Engineering Department, University of Los Andes Mérida 5101, Venezuela (e-mail: paredesj@ula.ve).

G. R. Arce is with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716 USA (e-mail: arce@ee.udel.edu).

Digital Object Identifier 10.1109/TSP.2006.880306

[1]The symbol $\sim$ denotes asymptotic similarity. Formally, $X$ has algebraic tails if there exist $c, \alpha > 0$ such that $\lim_{x\to\infty} x^{\alpha} \Pr(|X| > x) = c$.

In this paper, we introduce zero-order statistics (ZOS) as an approach to statistical moment characterization which is well defined over all distributions with algebraic or lighter tails. In the same way as $\mathrm{p}^{\mathrm{th}}$-order moments constitute the basis of FLOS and HOS techniques, the theory of ZOS is based on logarithmic "moments" of the form $\mathbf{E}\log|X|$. We introduce the fundamental ZOS for signal power, dispersion and location, and develop a theoretical ground which enables the sound characterization of signals in very impulsive environments. Depending on the nature of the underlying impulsiveness, ZOS methods enable the treatment of signal processing problems that are either unaddressable or inefficiently managed by the classical second-order based methods.

In the search for efficient estimation methods in very impulsive environments, we discovered a singular zero-order estimator of location with important optimality properties under heavy impulsiveness conditions. In the same way as the sample mean and related linear estimation techniques play a fundamental role in classical statistics and signal processing, we believe that the zero-order location estimator represents an important statistic to build upon towards the effective characterization and processing of very impulsive processes.

Early success stories of the methods introduced in this paper include applications in channel coding and equalization [13]–[17], multiuser communications [18]–[23], digital video processing [24], image processing [25], and robust data fitting [26].

## II. Logarithmic-Order Processes

The following Theorem provides a natural motivation for characterizing the class of processes of interest in this paper. The proof is known and can be found for example in [27].

*Theorem 1:* Let $X$ be a random variable with algebraic or lighter tails. Then, $\mathbf{E}\log|X| < \infty$.

Since our main goal is to develop a signal processing framework under which all random processes with algebraic tails can be characterized, Theorem 1 allows us to restrict our attention to processes with finite logarithmic moments. We will refer to such processes as being of "logarithmic order," in analogy with the term "second order," used to denote processes with finite variance. Although it is possible to find extremely heavy-tailed probability distributions which are not of logarithmic order [27], no distributions with infinite logarithmic moments have been claimed, to the best of our knowledge, as useful models for practical applications.

Of particular interest in this paper is the class of logarithmic-order processes with infinite variance, which includes all algebraic-tailed processes with tail constant $\alpha < 2$. The evident limitations of classical methods in the context of logarithmic-order processes make it necessary to develop new basic statistics under which an efficient theory of estimation can be built. In the following section, we introduce a new indicator of process strength, namely the *geometric power*, which overcomes many of the limitations of second-order theory in the framework of logarithmic-order processes. Next, in Section IV, we develop a location indicator intimately related to the geometric power. These two parameters constitute the foundation of the ZOS framework. Much as the (second-order) power and the expected value play a central role in the theory of second-order processes, the geometric power and its related location parameter are of fundamental importance in the development of a theory of estimation for logarithmic-order processes.

## III. The Geometric Power

The *power* of a second-order process, $\mathbf{E}X^2$, has been widely accepted in signal processing as a standard measure of signal strength. Although this (second-order) power is often associated with the physical concepts of power and energy, its meaning is not universal and may be troublesome when the processes exhibit heavy tails. In the especial case of heavy algebraic tails ($\alpha < 2$), the second-order power is always infinite and does not give useful information about process strength. In order to develop signal processing tools for the class of logarithmic-order processes, it is necessary to define alternative strength measures.

Although the definition of a general strength indicator involves an inexorable dose of arbitrariness, we are interested in an indicator that 1) gives useful characterizations along the class of logarithmic-order processes; 2) has a rich set of properties that can be effectively used; and 3) is mathematically and conceptually simple. In the following definition, we introduce what we believe is the simplest parameter satisfying these conditions.

*Definition 1 (Geometric Power):* Let $X$ be a logarithmic-order random variable. We define the *geometric power* of $X$ as

$$S_0 = S_0(X) = e^{\mathbf{E}\log|X|}. \tag{2}$$

As discussed in Section III-E, $S_0$ is simply the geometric mean of $|X|$. We coined the name *geometric power* as a result of this fact.

### A. Properties of the Geometric Power

It can be easily shown that the geometric power is a scale parameter, and as such, it can be effectively used as an indicator of process strength or "power" in situations where second-order methods are inadequate. In the following we describe some of the most important properties of this parameter. The proofs are not difficult, and are omitted for brevity of presentation. The interested reader is referred to [27] for detailed proofs.

*Property 1 ($S_0$ is a Scale Parameter):* For any logarithmic-order process $X$, and any constant $c$

(i) $S_0(X) \geq 0$.
(ii) $S_0(cX) = |c|S_0(X)$.

*Property 2 ($S_0$ is an Indicator of Process Strength):* For any logarithmic-order process $X$, and constants $c$, $c_1$ and $c_2$,

(i) $S_0(c) = |c|$.
(ii) $0 \leq c_1 < |X| < c_2$ implies $c_1 < S_0(X) < c_2$.
(iii) $S_0(X) = 0$ if and only if $\Pr(|X| < \epsilon) > 0$ for all $\epsilon > 0$, which implies that zero power is only attained when there is a "pile up" of probability mass around zero.

*Property 3 (Multiplicativity):* For any pair of logarithmic-order random variables $X, Y$, and any real constant $c$

(i) $S_0(XY) = S_0(X)S_0(Y)$;
(ii) $S_0(X/Y) = S_0(X)/S_0(Y)$;
(iii) $S_0(X^c) = S_0(X)^c$.

Note that the multiplicativity property is a direct consequence of the properties of the logarithm, and it is always valid, independently of the correlation structure between $X$ and $Y$.

*Property 4 (Absolute Value Inequality):* For any pair of logarithmic-order random variables $X$ and $Y$

$$S_0\left(|X|+|Y|\right) \geq S_0(X) + S_0(Y). \tag{3}$$

*Proof:* Assume that $\Pr(|X|+|Y|=0)=0$ (the case $\Pr(|X|+|Y|=0)>0$ is straightforward from Property 2(iii)). Then,

$$1 = \mathbf{E}\left(\frac{|X|}{|X|+|Y|}\right) + \mathbf{E}\left(\frac{|Y|}{|X|+|Y|}\right). \tag{4}$$

Let $Z$ be a variable with finite expectation. It is easy to show[2] that $\mathbf{E}|Z| \geq S_0(Z)$. Applying this inequality to both terms in (4) we get

$$1 \geq S_0\left(\frac{|X|}{|X|+|Y|}\right) + S_0\left(\frac{|Y|}{|X|+|Y|}\right) \tag{5}$$

and using Property 3(ii),

$$1 \geq \frac{S_0\left(|X|\right) + S_0\left(|Y|\right)}{S_0\left(|X|+|Y|\right)} \tag{6}$$

which leads to the desired result. ■

## B. *The Geometric Power of $\alpha$-Stable Processes*

Alpha-stable processes constitute one of the most important infinite-variance families in the logarithmic-order class. Since they are the only processes that satisfy a form of generalized Central Limit Theorem [28], they can appear in practice as a result of natural stochastic phenomena. Symmetric $\alpha$-stable processes are being the subject of increased attention as a suitable framework for efficient signal processing in impulsive environments [4], [27], [29]–[37].

A "zero-centered" symmetric $\alpha$-stable distribution is commonly described through its characteristic function

$$\phi(\omega) = e^{-\gamma|\omega|^{\alpha}}. \tag{7}$$

The parameter $\alpha$ is usually called the *characteristic exponent* or *index*. It can be proven that, in order for (7) to define a characteristic function, the values of $\alpha$ must be restricted to the interval (0, 2]. When $\alpha < 2$, the distribution is algebraic-tailed with tail constant $\alpha$, implying infinite variance. When $\alpha = 2$, the distribution is Gaussian, implying lighter-than-algebraic tails.

The parameter $\gamma$, usually called the *dispersion*, is a positive constant related to the scale of the distribution. For a fixed $\alpha$, larger values of $\gamma$ correspond to larger strengths of the process. It is easy to see that $\gamma^{1/\alpha}$ is, in fact, a scale parameter of the distribution.

The following proposition gives us a closed-form expression for the geometric power of symmetric $\alpha$-stable random variables:

*Proposition 1:* The geometric power of a symmetric $\alpha$-stable variable as defined in (7), is given by

$$S_0 = \frac{(C_g\gamma)^{1/\alpha}}{C_g} \tag{8}$$

where $C_g = e^{C_e} \approx 1.78$, is the exponential of the Euler constant.

*Proof:* From [38, p. 215], the logarithmic moment of a zero-centered symmetric $\alpha$-stable random variable with unitary dispersion is given by

$$\mathbf{E}\log|X| = \left(\frac{1}{\alpha}-1\right)C_e \tag{9}$$

where $C_e = 0.5772\ldots$ is the Euler constant. This gives

$$S_0(X)\big|_{\gamma=1} = e^{\mathbf{E}\log|X|} = \left(e^{C_e}\right)^{\frac{1}{\alpha}-1} = \frac{C_g^{1/\alpha}}{C_g} \tag{10}$$

where $C_g = e^{C_e}$. If $X$ has a nonunitary dispersion $\gamma$, it is easy to see that

$$S_0(X) = \gamma^{1/\alpha} S_0(X)\Big|_{\gamma=1} = \frac{(C_g\gamma)^{1/\alpha}}{C_g}. \tag{11}$$

■

## C. *On the Calculation of the Geometric Power*

In some cases, the computation of $S_0$ may be greatly simplified from the use of the following result.

*Proposition 2:* Define the "moment function," $L_X(p) = \mathbf{E}|X|^p$. Then

$$S_0(X) = e^{L'_X(0)} \tag{12}$$

where $L'_X(p)$ denotes the derivative of $L_X(p)$.

*Proof:* By definition

$$L'_X(p) = \frac{d}{dp}\int_{-\infty}^{\infty} |x|^p f(x)dx. \tag{13}$$

Since $X$ has algebraic or lighter tails, we can enter the derivative into the integral

$$L'_X(p) = \int_{-\infty}^{\infty} \frac{d}{dp}|x|^p f(x)dx = \int_{-\infty}^{\infty} |x|^p \log|x| f(x)dx. \tag{14}$$

Making $p = 0$ in (14) and taking exponentials, we get the desired result. ■

In some cases, computing the derivative of the "moment function," $L_X(p)$, may lead to an easy way to obtain the Geometric Power rather than using the direct definition in (2). Moreover, (12) provides an alternative representation of the Geometric Power.

Table I shows the values of the geometric and second-order powers for several common distributions. Proposition 2 was used for the calculation of some of them [27]. With the exception of the uniform and Gaussian distributions, all the other distributions in the table have algebraic tails, parameterized by the tail constant $\alpha$ and a scale parameter $\sigma$. Note that for values

[2]This is a particular case of a broader result introduced in Section III-F.

TABLE I
GEOMETRIC AND SECOND-ORDER POWERS OF SEVERAL COMMON DISTRIBUTIONS. THE SECOND-ORDER POWER BECOMES INFINITE IN MANY CASES FOR VALUES OF $\alpha < 2$. THE GEOMETRIC POWER IS ALWAYS WELL DEFINED

| NAME | PARAMETERIZATION | GEOMETRIC POWER | 2ND-ORDER POWER |
|---|---|---|---|
| $Uniform$ | $f(x) = 1/\sigma, \quad \lvert x\rvert < \sigma/2$ | $\sigma/2e$ | $\sigma^2/12$ |
| $Gaussian$ | $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$ | $\sigma/\sqrt{2C_g}$ | $\sigma^2$ |
| $Cauchy$ | $f(x) = \frac{\sigma}{\pi}\frac{1}{\sigma^2+x^2}$ | $\sigma$ | $\infty$ |
| $Symmetric$ $\alpha - stable$ | $\phi(w) = e^{-\lvert\sigma w\rvert^\alpha}$ | $\sigma C_g^{\frac{1}{\alpha}-1}$ | $\begin{cases} 2\sigma^2 & \alpha = 2 \\ \infty & \alpha < 2 \end{cases}$ |
| $Generalized - t$ | $f(x) = a_t(\sigma^2 + x^2)^{-\frac{1+\alpha}{2}}$ | $\frac{\sigma}{2}[C_g e^{\psi(\alpha/2)}]^{-1/2}$ | $\begin{cases} \frac{\sigma^2}{\alpha-2} & \alpha > 2 \\ \infty & \alpha \leq 2 \end{cases}$ |
| $Generalized$ $Cauchy$ | $f(x) = a_c(\sigma^p + \lvert x\rvert^p)^{-\frac{1+\alpha}{p}}$ | $\sigma[e^{\psi(1/p)-\psi(\alpha/p)}]^{1/p}$ | $\begin{cases} \sigma^2\frac{\Gamma(\frac{3}{p})\Gamma(\frac{\alpha-2}{p})}{\Gamma(\frac{1}{p})\Gamma(\frac{\alpha}{p})} & \alpha > 0 \\ \infty & \alpha \leq 2 \end{cases}$ |
| $Pareto$ | $\alpha\sigma^\alpha x^{-\alpha}, \quad x > \sigma$ | $\sigma e^{1/\alpha}$ | $\begin{cases} \sigma^2\frac{\alpha}{2-\alpha} & \alpha > 2 \\ \infty & \alpha \leq 2 \end{cases}$ |

$C_g = e^{C_e} \approx 1.78$ is the exponential of the Euler constant.

$\Gamma(\cdot)$ is the gamma function.

$a_t = \frac{\Gamma(\frac{1+\alpha}{2})}{\sqrt{\pi}\Gamma(\frac{\alpha}{2})}\sigma^\alpha$.

$a_c = \frac{p\Gamma(\frac{1+\alpha}{2})}{2\Gamma(\frac{1}{p})\Gamma(\frac{\alpha}{p})}\sigma^\alpha$.

$\psi(z) = \frac{d}{dz}\log\Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)}$ is the psi function.

of $\alpha$ lower than 2, the second-order power becomes infinite and consequently useless as a measure of process strength. The geometric power, on the other hand, is well defined for any value of $\alpha > 0$. Furthermore, being a scale parameter, it is always a multiple of $\sigma$ and, more interesting, it decreases as a function of $\alpha$.

### D. *Rationale for the Use of the Geometric Power*

Fig. 1 illustrates the usefulness of the geometric power for characterizing process strength in the $\alpha$-stable framework. The scatter plot on the left side is an independent, identically distributed (i.i.d.) realization, stemming from an $\alpha$-stable distribution with $\alpha = 1.99$ and geometric power $S_0 = 1$. On the right-hand side, the scatter plot stems from a Gaussian distribution with the same geometric power. An intuitive inspection of Fig. 1 suggests that both of the generating processes possess the same strength, in accordance with the values of the geometric power. In contrast, the values of the second-order power lead to the misleading conclusion that the process on the left is much stronger than the one on the right.

It is easy to find examples like the above that also disqualify FLOS-based indicators of strength in the class of logarithmic-order processes. In fact, fractional moments of order $p$ present the same type of "discontinuities" like the one illustrated in Fig. 1 for processes with tail constants close to $\alpha = p$. The geometric power, on the other side, is consistently continuous along the entire range of values of $\alpha$. This characteristic of the geometric power provides a useful framework for comparing the strengths of any pair of logarithmic-order signals, in the same way as the (second-order) power is used in the classical framework.

The use of the geometric power as a universally well-defined indicator of signal strength, include early success stories recently reported in the literature for impulsive multiuser communications [18]–[23], channel coding [13], [16], channel equalization [14], [15], [17], digital video processing [24] and robust data fitting [26].

*Remark:* In conducting comparisons between "zero-centered" $\alpha$-stable random variables with different impulsiveness levels, one could resort for example to the scale parameter $\sigma = \gamma^{1/\alpha}$ as an indicator of signal strength [39]. Although this approach can give insights into the strength characterization of $\alpha$-stable processes, it is not sufficiently general and does not allow to make comparisons between logarithmic-order processes outside the $\alpha$-stable class.

### E. *Estimation of the Geometric Power*

Let $x_1, x_2, \ldots,$ be a sequence of independent samples originated from a distribution with geometric power $S_0$. By virtue of the law of large numbers and taking into account that $S_0$ exists and is finite [28], we can replace the expected value operator by the sample average in the definition in (2), so that the resulting statistic

$$\hat{S}_0 = \exp\left(\frac{1}{N}\sum_{i=1}^{N}\log|x_i|\right) \tag{15}$$

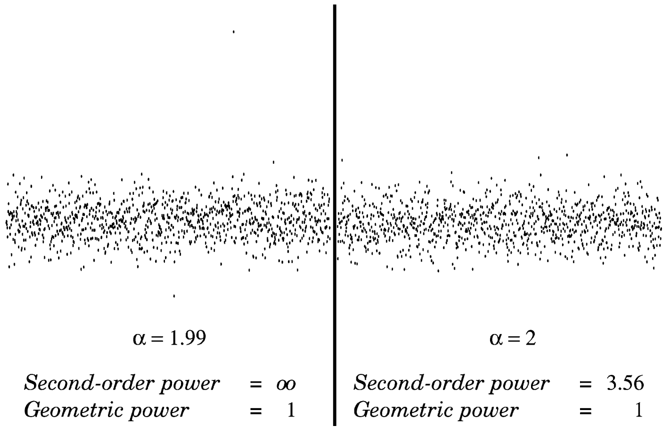



Fig. 1. Comparison of second-order power versus geometric power. Left: scatter plot stemming from an $\alpha$-stable distribution ($\alpha = 1.99$). Right: scatter plot stemming from a Gaussian distribution. The geometric power gives an intuitive idea of the relative strengths of the signals. The second-order power provides misleading results.

is always a consistent estimator of $S_0$. Furthermore, it has been shown in [40] that if $x_i$ obeys a stable distribution, the logarithmic averages are asymptotically normal emphasizing, thus, the consistency of the estimator.

A closer look at (15) indicates that $\hat{S}_0$ is, in fact, the geometric mean of the absolute values of the data

$$\hat{S}_0 = \left( \prod_{i=1}^{N} |x_i| \right)^{1/N}. \tag{16}$$

We coined the name *geometric power* as a consequence of this result. In addition to its consistency, it is known that $\hat{S}_0$ is scale invariant, and it is equivalent to the maximum likelihood estimator of $S_0$ when the underlying distribution is Pareto [41].

### *F. Relation to FLOS*

The geometric power is intimately linked to FLOS parameters as indicated in the following.

*Theorem 2:* Let $S_p = (\mathbf{E}|X|^p)^{1/p}$ denote the scale parameter derived from the $p^{th}$-order moment of $X$. If $S_p$ exists for sufficiently small values of $p$, then

$$S_0 = \lim_{p \to 0} S_p. \tag{17}$$

Furthermore, $S_0 \le S_p$, for any $p > 0$.

*Proof:* It is enough to prove that

$$\lim_{p \to 0} \frac{\log \mathbf{E}|X|^p}{p} = \mathbf{E} \log |X|. \tag{18}$$

Applying L'Hospital rule

$$\lim_{p \to 0} \frac{\log \mathbf{E}|X|^p}{p} = \lim_{p \to 0} \frac{d}{dp} \log \mathbf{E}|X|^p \tag{19}$$

$$= \lim_{p \to 0} \frac{\frac{d}{dp}\mathbf{E}|X|^p}{\mathbf{E}|X|^p} \tag{20}$$

$$= \lim_{p \to 0} \frac{\mathbf{E}\left(|X|^p \log |X|\right)}{\mathbf{E}|X|^p} \tag{21}$$

$$= \mathbf{E} \log |X|. \tag{22}$$

To prove that $S_0 \le S_p$, Jensen's inequality [28] guarantees that for a convex function $\phi$ and a random variable $Z$, $\phi(\mathbf{E}Z) \le \mathbf{E}\phi(Z)$. Making $\phi(x) = e^x$ and $Z = \log |X|^p$ we get

$$(S_0)^p = e^{(\mathbf{E} \log |X|^p)} \le \mathbf{E} e^{\log |X|^p} = \mathbf{E}|X|^p = (S_p)^p \tag{23}$$

which leads to the desired result. ∎

Theorem 2 indicates that techniques derived from the geometric power are the limiting *zero-order* relatives of FLOS. We coined the name ZOS as a consequence of this property.

### *G. The Geometric Signal-to-Noise Ratio (G-SNR)*

A natural application of the ZOS framework arises in digital communications, where it is necessary to quantify signal quality to assess and control system performance. Signal quality is usually defined as the ratio between the channel information and noise powers, typically in the second-order sense. This important statistic is known as the "signal-to-noise ratio" (SNR) of the communication system. Large values of the SNR indicate good system quality, whereas low values of the SNR indicate poor and "noisy" performance.

In channels corrupted by infinite-variance impulsive noise, the standard SNR is always zero, and, thus, becomes meaningless as an indicator of signal quality. This has been a common inconvenience in the robust signal processing literature, where authors have been forced to use *ad hoc* reformulations of the SNR that "make sense" in the particular environments at hand. The geometric power allows us to define a universal indicator of signal quality that is both meaningful and model-independent.

*Definition 2 [Geometric Signal-to-Noise Ratio]:* Let $A$ be the amplitude of a modulated signal in an additive-noise channel with noise geometric power $S_0$. Then, we define the G-SNR as

$$\text{G-SNR} = \frac{1}{2C_g} \left( \frac{A}{S_0} \right)^2 \tag{24}$$

where $C_g = e^{C_e} \approx 1.78$ is the exponential of the Euler constant.

The normalization constant $2C_g$ is used to ensure that the definition of the G-SNR corresponds to that of the standard SNR if the channel noise is Gaussian.

The G-SNR, originally introduced by Gonzalez in [27],[3] has since then been successfully exploited by several researchers in the field of signal processing and communications [13]–[24], [26].

## IV. Zero-Order Location and Dispersion

In the framework of second-order processes, the mean of a random variable can be conveniently described as the parameter $\mu_2$ that minimizes the power of the shifted variable $X - \mu$ over all possible shifts $\mu$ [36], [42]. This is,[4]

$$\mathbf{E}X = \mu_2 = \arg\min_{\mu} S_2(X-\mu) = \arg\min_{\mu} \mathbf{E}(X-\mu)^2. \tag{25}$$

[3]In Gonzalez's original work, it should be noticed that there is a typo in Equation (3.117), p. 68.

[4]We use the subscript 2 to refer to the fact that the statistics are of "second order." Likewise, we will use the subscript 0 to denote the ZOS to be introduced later.

Equation (25) states the optimality of the mean as an indicator of location which minimizes the mean squared error (mse). Historically, (25) has been a strong theoretical argument favoring the use of the mean and related linear techniques in signal processing problems involving second-order processes.

Following the same reasoning, we can use the geometric power to derive the zero-order measures of location and dispersion that will serve as the basis for a framework of estimation in the class of logarithmic-order processes. We begin with the definition of the zero-order location parameter.

*Definition 3 (Zero-Order Location):* Let $X$ be a logarithmic-order variable. We define the zero-order indicator of location, $\mu_0$, as the value of $\mu$ that minimizes the geometric power of the shifted variable $X - \mu$. This is

$$\mu_0 = \mu_0(X) = \arg\min_{\mu} S_0(X - \mu). \tag{26}$$

The following properties of $\mu_0$ can be easily proven for any logarithmic-order process (see, e.g., [27]).

*Property 5 ($\mu_0$ is a Location Parameter):* Let $X$ be symmetric and unimodal[5] with symmetry center $c$. Then, $\mu_0(X) = c$.

*Property 6 (Shift and Scale Invariance):* For any constants $a$ and $b$

$$\mu_0(aX + b) = a\mu_0(X) + b. \tag{27}$$

In contrast to the mean, the zero-order location exists for every logarithmic-order process.

*Remark:* When locating a logarithmic-order distribution we will refer to $\mu_0$ as the *center* of the distribution, and we will say that a process $X$ is *zero-centered* when $\mu_0(X) = 0$. Note that any zero-mean symmetric unimodal process is also zero-centered, but not all zero-centered symmetric unimodal processes can be cataloged as zero-mean.

Going back to expression (25), the standard deviation of a second-order process can be seen as the minimum value of the second-order scale $S_2(X - \mu)$, over all possible shifts $\mu$. Analogously, we can derive a zero-order parameter for quantifying dispersion directly from Definition 3.

*Definition 4 (Zero-Order Dispersion):* Let $X$ be a logarithmic-order random variable. We define $\sigma_0$, the zero-order indicator of dispersion, as

$$\sigma_0 = \min_{\mu} S_0(X - \mu) = S_0(X - \mu_0). \tag{28}$$

The relation between $\sigma_0$ and $S_0$ is similar to the relation between variance and power for second-order processes. In principle, if a logarithmic-order process is *zero-centered*, then $\sigma_0 = S_0$. Also, $\sigma_0$ will present similar properties to those presented for the geometric power in Section III. The reader is referred to [27] for a discussion of these properties.

[5]A random variable is said to be unimodal with mode $c$, if its density function is continuous and monotonic increasing on $(-\infty, c)$, and continuous and monotonic decreasing on $(c, \infty)$ [43].

### A. Zero-Order Estimation of Location

The definition of the ZOS location parameter in (26) leads to the discovery of a new estimator with strong optimality properties as a locator of very impulsive processes. Substituting (2) in (26) and using the fact that $\log(\cdot)$ is a monotonic nondecreasing function, we can reformulate the definition of $\mu_0$ as

$$\mu_0 = \arg\min_{\mu} \mathbf{E} \log |X - \mu|. \tag{29}$$

In order to propose an estimator $\hat{\mu}_0$, an intuitive inspection of (29) would suggest to replace the expected value operator by the sample average, so to obtain

$$\hat{\mu}_0 = \arg\min_{\mu} \sum_{i=1}^{N} \log |x_i - \mu|. \tag{30}$$

Note that the argument of expression (30) yields $-\infty$ at $\mu = x_i$, for $i = 1, \dots, N$, resulting in a multiple tie among all the sample values $x_i$ as solutions to the minimization. The above tie is an ill effect generated by the unbounded discontinuities of the minimization argument in (30). In order to formally break the tie, we constrain the minimization domain from the set $\Re$ of real numbers to a compact set that excludes all those points $\mu$ for which the sum of logarithms is discontinuous. This can be done by defining

$$\hat{\mu}_\delta = \arg\min_{\mu \in \Delta_\delta} \sum_{i=1}^{N} \log |x_i - \mu| \tag{31}$$

where

$$\Delta_\delta = \Re - \bigcup_{i=1}^{N} (x_i - \delta, x_i + \delta) \tag{32}$$

$\Re$ is the set of the real numbers, and $\delta$ is a small positive constant. The tie is easily solved by letting

$$\hat{\mu}_0 = \lim_{\delta \to 0} \hat{\mu}_\delta. \tag{33}$$

The rigorous reader can consider (33) as our formal definition of $\hat{\mu}_0$. The following Theorem proves the existence of the limit in (33), and gives a surprisingly simple formula for calculating $\hat{\mu}_0$ without resorting to the limit expression.[6]

*Theorem 3:* Given a sample of values, $x_1, \dots, x_N$, the zero-order estimator of location as defined in (33) can be calculated as

$$\hat{\mu}_0 = \arg\min_{x_j \in \mathcal{M}} \prod_{i=1, x_i \neq x_j}^{N} |x_i - x_j| \tag{34}$$

where $\mathcal{M} = \{x_{m_1}, \dots, x_{m_n}\}$ is the set of "modes" or most repeated values in the sample.

*Proof:* See Appendix I.

[6]Fast, numerically efficient implementations of $\hat{\mu}_0$ based on Theorem 3 are currently commercially available, e.g., in the core of the STABLE software package under the framework of estimators for highly impulsive stable distributions. The STABLE software package is developed and distributed by Robust Analysis, Inc [44].

According to (34), $\hat{\mu}_0$ will always be one of the most repeated values in the sample, resembling the behavior of a sample mode. This *mode property*, as we call it, insinuates the high effectiveness of the estimator in locating heavy impulsive processes. In the following sections we report both theoretical and experimental evidence of this effectiveness.

Also, being a sample mode, $\hat{\mu}_0$ is evidently a "selection-type" estimator, in the sense that it is always equal, by definition, to one of the sample values. Thanks to this "selection" property $\hat{\mu}_0$ has been successfully applied to statistical and deterministic image processing problems [25].

### B. Properties of the ZOS Location Estimator

In addition to the "selection" and mode properties reported above, $\hat{\mu}_0$ presents a rich set of interesting properties. We discuss some of them here.

*Property 7 (Shift and Scale Invariance):* Let $z_i = ax_i + b$, for $i = 1, \ldots, N$. Then

$$\hat{\mu}_0(z_1, \ldots, z_N) = a\hat{\mu}_0(x_1, \ldots, x_N) + b. \tag{35}$$

The proof is trivial.

*Property 8 (No Overshoot/Undershoot):* $\hat{\mu}_0$ is always bounded by

$$x_{(2)} \leq \hat{\mu}_0 \leq x_{(N-1)} \tag{36}$$

where $x_{(i)}$ denotes the $i^{th}$-order statistic of the sample.

*Proof:* We will prove that $x_{(2)} \leq \hat{\mu}_0$. The proof for $\hat{\mu}_0 \leq x_{(N-1)}$ follows analogously.

If either $x_{(1)}$ or $x_{(2)}$ are repeated more than once in the sample set, then $x_{(2)} \leq \hat{\mu}_0$ holds trivially. Hence, assume that the values of $x_{(1)}$ and $x_{(2)}$ are repeated only once in the sample set, and then, according to (34), $x_{(2)} \leq \hat{\mu}_0$ if and only if

$$\prod_{i=1, x_i \neq x_{(1)}}^{N} \left|x_i - x_{(1)}\right| \geq \prod_{i=1, x_i \neq x_{(2)}}^{N} \left|x_i - x_{(2)}\right|. \tag{37}$$

To prove that (37) is true, note that for $i = 3, \ldots, N$, $x_{(i)} - x_{(1)} \geq x_{(i)} - x_{(2)}$, and then

$$\prod_{i=1, x_i \neq x_{(1)}}^{N} \left|x_i - x_{(1)}\right| = \left|x_{(2)} - x_{(1)}\right| \prod_{i=3}^{N} \left|x_{(i)} - x_{(1)}\right| \tag{38}$$

$$\geq \left|x_{(2)} - x_{(1)}\right| \prod_{i=3}^{N} \left|x_{(i)} - x_{(2)}\right| \tag{39}$$

$$= \prod_{i=1, x_i \neq x_{(2)}}^{N} \left|x_i - x_{(2)}\right|. \tag{40}$$

■

The following is a direct consequence of the No Overshoot/Undershoot property:

*Property 9:* If $N = 3$, $\hat{\mu}_0$ is equivalent to the sample median.

*Property 10 (Unbiasedness):* Let $X_1, X_2, \ldots, X_N$ be all independent and symmetrically distributed around the symmetry center $c$. Then, $\hat{\mu}_0 = \hat{\mu}_0(X_1, X_2, \ldots, X_N)$ is also symmetrically distributed around $c$. In particular, if $\mathbf{E}\hat{\mu}_0$ exists, then $\mathbf{E}\hat{\mu}_0 = c$.

*Proof:* If $X_i$ is symmetric around $c$, then $2c - X_i$ has the same distribution as $X_i$. Thus $\hat{\mu}_0(X_1, \ldots, X_N)$ has the same distribution as $\hat{\mu}_0(2c - X_1, 2c - X_2, \ldots, 2c - X_N)$, which, thanks to Property 7, is equal to $2c - \hat{\mu}_0(X_1, X_2, \ldots, X_N)$. It follows that $\hat{\mu}_0$ is symmetric around $c$. ■

*Remark on the Consistency of $\hat{\mu}_0$:* Determining the family of probability distributions for which $\hat{\mu}_0$ is consistent, is still an open research problem. For instance, it is easy to show that there exist symmetric bimodal distributions for which $\hat{\mu}_0$ is an inconsistent estimator of location. In a very extensive simulation study, we found that the variance of $\hat{\mu}_0$ always decreased as a function of the sample size for symmetric $\alpha$-stable distributions, even with very small values of $\alpha$ ($\alpha = 0.001$). We conjecture that $\hat{\mu}_0$ is consistent and asymptotically Gaussian for all continuous and unimodal distributions in the logarithmic-order class.

*Remark on the Complexity of $\hat{\mu}_0$:* Theorem 3, and in particular expression (33), enables the computation of $\hat{\mu}_0$ through an algorithm with quadratic complexity, i.e., $O(N^2)$, where $N$ is the sample size.[7] This is the same complexity exhibited by all FLOS-based estimators of location with $p < 1$, and represents a significant complexity reduction when compared against general maximum likelihood and M-estimation algorithms, which, for non-convex problems typical of tails $\alpha < 1$ possess NP-hard complexity. In the following section we prove that $\hat{\mu}_0$ is in fact the convergence point of the maximum likelihood estimator when the impulsiveness is very high, which makes it a very attractive alternative as a location estimator for small values of $\alpha$.

## V. Optimality of $\hat{\mu}_0$ in Very Impulsive Environments

Many popular estimators possess the valuable property of being optimal under a given class of probability distributions. The sample average for example, is known to be optimal for estimating the location of i.i.d. Gaussian processes, whereas the sample median presents optimal performance (in the maximum likelihood sense) when the underlying distribution is Laplacian. Identifying the class of distributions for which optimal (or close to optimal) performance is achieved, gives important information about the behavior of an estimator and can be a key tool at the time of designing applications. Even though distributions for which $\hat{\mu}_0$ represents the maximum likelihood estimator of location do not exist, the following characterization indicates the adequateness of $\hat{\mu}_0$ as a location estimator in very heavy-tailed environments. We begin the discussion with an important definition.

*Definition 5 (Purely Algebraic Family):* Let $\{f_\alpha\}_{\alpha>0}$ denote a family of algebraic-tailed density functions parameterized by the tail constant $\alpha$. We will refer to $\{f_\alpha\}_{\alpha>0}$ as a *purely algebraic* family if there exists $\epsilon > 0$, such that for all $\alpha < \epsilon$

(i) $f_\alpha$ is unimodal.

(ii) there exists a family of constants $C_\alpha$ such that, as $\alpha \to 0$

$$\frac{|x|^{1+\alpha} f_\alpha(x)}{C_\alpha} \to 1 \tag{41}$$

pointwise.

[7]Furthermore, since the computation of $\hat{\mu}_0$ is derived from the matrix of distances between the sample points, it is possible to design recursive linear complexity ($O(N)$) sliding window filtering algorithms based on $\hat{\mu}_0$, using the same techniques usually exploited for sliding window median filters [45].

Intuitively, $f_\alpha(x)$ is purely algebraic if, for small values of $\alpha$, it can be approximated by a "purely algebraic" density of the form $C_\alpha/|x|^{1+\alpha}$. Density functions of this form are extremely heavy tailed, exhibiting progressively heavier tails as $\alpha$ gets closer to zero. It is easy to prove that the zero-centered generalized-t distribution, according to the parameterization given in Table I, is purely algebraic. The following proposition identifies another important example of a purely algebraic family. The proof is included in Appendix II.

*Proposition 3:* The class of zero-centered symmetric $\alpha$-stable distributions is purely algebraic.

Now we are ready for the main result of this section.

*Theorem 4:* Let $\{f_\alpha\}_{\alpha>0}$ be a family of purely algebraic density functions, and let $T_\alpha$ denote the maximum likelihood location estimator associated with $f_\alpha$. Then

$$\lim_{\alpha\to 0} T_\alpha = \hat{\mu}_0. \tag{42}$$

*Proof:* See Appendix III.

Theorem 4 gives significant relevance to the discovery of $\hat{\mu}_0$. The result is very intuitive as well: it tells us that, in the presence of very impulsive noise (i.e., when the observations are very unreliable), the best location estimator is close to one of the most repeated values in the sample. The fact that $T_\alpha$ converges to one of the values in the sample, was discovered and demonstrated for the first time (for the family of $\alpha$-stable distributions) by DuMouchel in 1973 [46]. However, to our knowledge, no one had provided a characterization of the pseudomode behavior of this limit, nor given a precise formula for its computation. Theorems 3 and 4 do exactly that, not only for the family of $\alpha$-stable distributions, but for the complete class of *purely algebraic* distribution families.

## VI. Performance of ZOS in $\alpha$-Stable Noise

In this section, we illustrate via simulation the performance of the ZOS location estimator in i.i.d. $\alpha$-stable noise. Fig. 2 shows the estimated mean absolute errors (MAE) of the sample mean, the sample median and the ZOS estimator when used to locate the center of an i.i.d. symmetric $\alpha$-stable sample of size $N = 5$. The result comes from a Monte Carlo simulation with 200 000 repetitions. The values of the tail parameter range from $\alpha = 2$ (Gaussian case) down to $\alpha = 0.3$ (very impulsive). The value of $\sigma$ has been chosen to guarantee a unitary geometric power for each value of $\alpha$. Values of $\alpha$ slightly smaller than 2 indicate a distribution close to the Gaussian, in which case the sample mean outperforms both the median and the ZOS estimator. As $\alpha$ is decreased, the noise becomes more impulsive and the sample mean rapidly loses efficiency, being outperformed by the sample median for values of $\alpha$ less than 1.7. More interesting, as $\alpha$ approaches 1, the estimated MAE of the sample mean explodes. In fact, it is known that for $\alpha < 1$ it is more efficient to use any of the sample values than the sample mean itself as an estimator of location. This fact renders the sample mean useless for $\alpha < 1$.

As $\alpha$ continues to decrease, the sample median loses progressively more efficiency with respect to the ZOS estimator, and at $\alpha \approx 0.87$, the ZOS begins to outperform the sample me-

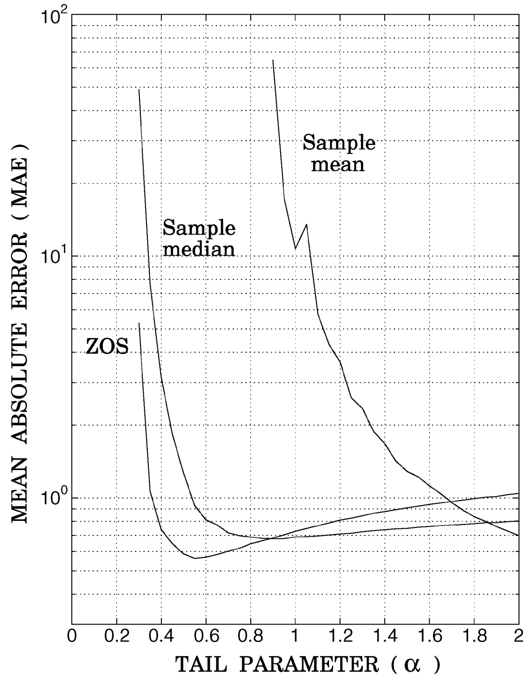



Fig. 2. Estimated mean absolute error of the sample mean, sample median and ZOS location estimator in $\alpha$-stable noise ($N = 5$).

dian. This is an expected result, given the optimality of the ZOS estimator for small values of $\alpha$. For the last value in the plot, $\alpha = 0.3$, the ZOS estimator has an estimated efficiency ten times better than the median. This increase in relative efficiency is expected to grow without bounds as $\alpha$ approaches 0 (recall that $\alpha = 0$ is the optimality point of the ZOS estimator).

The high efficiency of the ZOS location estimator in severely impulsive environments (i.e., very small values of $\alpha$) is illustrated in Fig. 3. Here, we applied mean, median, FLOS, and ZOS smoothing filters to a corrupted version of the "blocks" signal shown in Fig. 3(a). The signal corrupted with additive stable noise with $\alpha = 0.2$ is shown in Fig. 3(b), where a different scale is used to illustrate the impulsiveness of the noise. The (sample) mse between the original signal and the noisy observation is $8.3 \times 10^{88}$. The following running smoothers (location estimators) are applied, all using a window of size $N = 121$: (3c) the sample mean ($\text{MSE} = 6.9 \times 10^{86}$), (3d) the sample median ($\text{MSE} = 3.2 \times 10^4$), (3e) the FLOS with $p = 0.8$ ($\text{MSE} = 77.5$), and (3f) the ZOS location estimator ($\text{MSE} = 4.1$).

As shown in the figure, at this level of impulsiveness, the sample median and mean break down. The FLOS does not perform as well as the ZOS due to the mismatch of $p$ and $\alpha$. The performance of the FLOS estimator could certainly improve, but the parameter $p$ would have to be matched closely to the stable noise index, a task that can be difficult. The ZOS, on the other hand, performs well without the need of parameter tuning.

## VII. Conclusion

We have introduced the concept of zero-order statistics (ZOS), a statistical framework that is sound and consistent

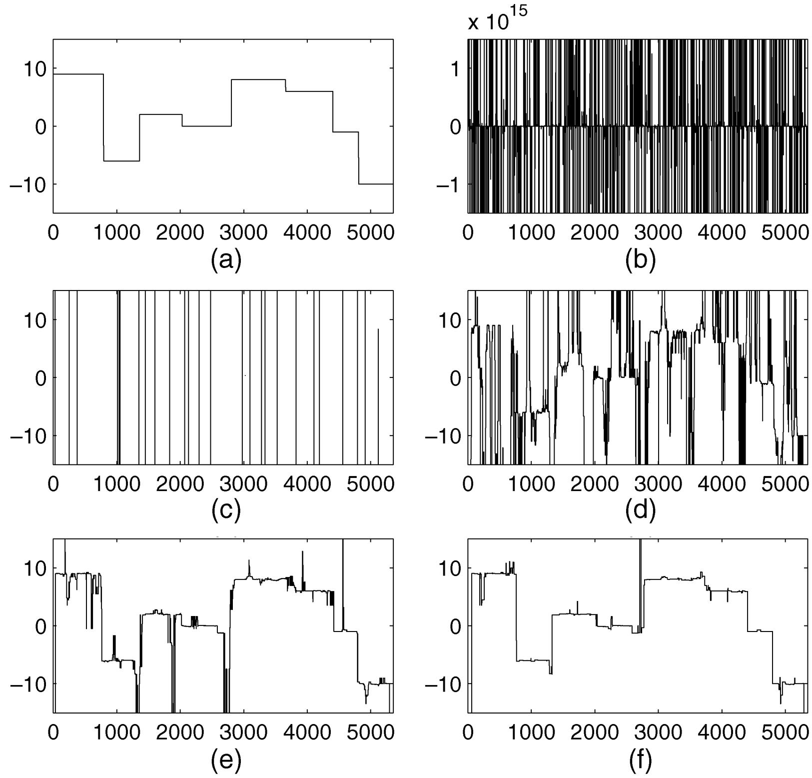



Fig. 3. Running smoothers in stable noise ($\alpha = 0.2$). All smoothers of observation window size 121. (a) Original "blocks" signal. (b) Corrupted signal with stable noise. (c) Output of the running mean. (d) Running median. (e) Running FLOS smoother. (f) Running ZOS smoother. Results were obtained using the commercial STABLE software package for stable distributions by Robust Analysis, Inc (http://www.robustanalysis.com).

for all processes with finite logarithmic moments. This "logarithmic-order" class includes impulsive-type processes modeled by algebraic-tailed distributions, and basically embraces most probability models with practical use today. Three new parameters, namely the geometric power, the zero-order location and the zero-order dispersion, constitute the basic elements of the ZOS framework. They play roles similar to the power, the expected value and the standard deviation, respectively, in the context of second-order processes. We discussed several important properties of the new parameters, and derived a ZOS framework for location estimation that led naturally to the discovery of a novel mode-type estimator with optimality properties under very impulsive noise. Given the limitations of methods based on second-order and fractional lower-order statistics, ZOS may be an attractive alternative for many characterization and processing problems in which infinite variance processes appear. Early success stories of the methods introduced in this paper include applications in channel coding and equalization [13]–[17], multiuser communications [18]–[23], digital video processing [24], image processing [25], and robust data fitting [26].

## Appendix I
## Proof of Theorem 3

Consider the "sum-of-costs" function

$$G(\mu) = \sum_{i=1}^{N} \log |x_i - \mu|. \tag{43}$$

$G(\mu)$ is a piecewise continuous function with discontinuities at $\mu = x_i$, $i = 1, \ldots, N$. At its discontinuity points, $G(\mu) = -\infty$. Under these circumstances it is easy to show that for $\delta$ sufficiently small, $\hat{\mu}_\delta$ is in the boundary of $\Delta_\delta$, i.e., there exists $k$, $1 \leq k \leq N$, such that either $\hat{\mu}_\delta = x_k + \delta$ or $\hat{\mu}_\delta = x_k - \delta$. By virtue of this result we can restrict the domain for the minimization in (31) to the finite set $\Gamma_\delta = \{x_j + \delta, x_j - \delta : j = 1, \ldots, N\}$, so that

$$\hat{\mu}_\delta = \arg\min_{\mu \in \Gamma_\delta} G(\mu) = \arg \min_{\substack{\mu = x_j \pm \delta, \\ j=1,\ldots,N}} G(\mu). \tag{44}$$

Now, let $r_j$ denote the number of times the value $x_j$ is repeated in the sample set. Then, for $j = 1, \ldots, N$

$$G(x_j \pm \delta) = \sum_{i=1}^{N} \log |x_i - x_j \mp \delta| \tag{45}$$

$$= r_j \log \delta + \sum_{i=1, x_i \neq x_j}^{N} \log |x_i - x_j \mp \delta|. \tag{46}$$

If $r_j > r_i$, then for $\delta$ sufficiently small

$$G(x_i \pm \delta) - G(x_j \pm \delta) = O\left((r_i - r_j) \log \delta\right) > 0. \tag{47}$$

Thus, $\hat{\mu}_\delta$ is of the form $x_j \pm \delta$, where $r_j$ is maximal, i.e., where $x_j$ is one of the most repeated values in the sample.

Now, let us define the "revised sum-of-costs" associated with $x_j$ as

$$G_j(\mu) = \sum_{i=1, x_i \neq x_j}^{N} \log |x_i - \mu| \tag{48}$$

and let $k$ be an index such that $r_k = \max_j r_j$. If there is only one value, say $x_k$, with repetition index $r_k$ (i.e., $x_k$ is the only sample value which is repeated $r_k$ times), $\hat{\mu}_\delta$ can be easily calculated as

$$\hat{\mu}_\delta = \begin{cases} x_k + \delta & \text{if } G_k(x_k + \delta) \leq G_k(x_k - \delta) \\ x_k - \delta & \text{if } G_k(x_k - \delta) \leq G_k(x_k + \delta) \end{cases}. \tag{49}$$

More interestingly, if the most repeated value in the sample is not unique, (i.e., if there exist different sample values $x_{k_1}$ and $x_{k_2}$ such that $r_{k_1} = r_{k_2} = r_k$), then, according to (45) and (48)

$$G(x_{k_1} \pm \delta) - G(x_{k_2} \pm \delta) = G_{k_1}(x_{k_1} \pm \delta) - G_{k_2}(x_{k_2} \pm \delta). \tag{50}$$

Expression (50) implies that the "revised sum-of-costs" can be used as the discriminant to identify the minimum sum-of-costs estimator once the most repeated values in the sample have been identified. Based on results (31) and (50), a general procedure for computing $\hat{\mu}_\delta$ in the vicinity of $\delta = 0$ can now be summarized as follows:

1) Identify $\mathcal{M} = \{x_{m_1}, \ldots, x_{m_n}\}$, the set of "modes" or most repeated values in the sample.
2) Construct $\mathcal{M}_\delta = \{x_{m_i} + \delta, x_{m_i} - \delta : x_{m_i} \in \mathcal{M}\}$.
3) Calculate the corresponding revised sum-of-costs $G_{m_i}(x_{m_i} \pm \delta)$ for each element in $\mathcal{M}_\delta$.
4) Select $\hat{\mu}_\delta$ as the element in $\mathcal{M}_\delta$ with minimum revised sum-of-costs.

Now, since $G_{m_i}(x_{m_i} \pm \delta)$ is continuous in the vicinity of $\delta = 0$, taking the limit as $\delta \to 0$ defines

$$\begin{aligned} \hat{\mu}_0 &= \lim_{\delta \to 0} \hat{\mu}_\delta \\ &= \arg \min_{x_{m_i} \in \mathcal{M}} G_{m_i}(x_{m_i}) \\ &= \arg \min_{x_j \in \mathcal{M}} \sum_{i=1, x_i \neq x_j}^{N} \log |x_i - x_j| \\ &= \arg \min_{x_j \in \mathcal{M}} \prod_{i=1, x_i \neq x_j}^{N} |x_i - x_j|. \end{aligned} \tag{51}$$

■

*Remark:* It is worth noting that the procedures described, including (49), do not necessarily describe $\hat{\mu}_0$ and $\hat{\mu}_\delta$ uniquely. In some particular situations, it could happen that several different values satisfy, for example, Step 4 above or the expression for $\hat{\mu}_0$ in (51). In order to make the definitions unique, one could define a "tie-breaker" rule that selects a particular solution among all the possible ones. The definition of such a rule could impact the behavior of $\hat{\mu}_0$ in discretely distributed processes, and it should be designed according to the specifics of the problem. If the main concern is to get unbiased results, a uniformly randomized selection should be sufficient.

Note also that for processes with continuous distributions, the probability of occurrence of the above "ties" is zero, reducing the "tie-breaking" procedure to an unimportant formality.

## Appendix II
## Proof of Proposition 3

We must prove that the class of zero-centered symmetric $\alpha$-stable distributions satisfy conditions *(i)* and *(ii)* in Definition 5. Condition *(i)* (unimodality), is a well-known property of $\alpha$-stable distributions (see for example [38] or [43]). We must then prove *(ii)*. To avoid the notational burden, let us assume that the dispersion $\gamma$ is unitary. The generalization of the result to other values of $\gamma$ is straightforward.

An integral expression for the zero-centered, unit dispersion, symmetric stable density when $\alpha < 1$ is given [38] by

$$f_\alpha(x) = \frac{\alpha}{(1-\alpha)\pi} \frac{1}{|x|^{\frac{1}{1-\alpha}}} \times \int_0^{\pi/2} \upsilon(\theta) \exp\left(-\upsilon(\theta)|x|^{-\alpha/(1-\alpha)}\right) d\theta \tag{52}$$

where

$$\upsilon(\theta) = \frac{(\sin(\alpha\theta))^{\frac{\alpha}{1-\alpha}} \cos[(1-\alpha)\theta]}{(\cos\theta)^{\frac{1}{1-\alpha}}}.$$

The value of $f_\alpha(0)$ can be calculated as the limit when $x \to 0$ in (52). A convenient rearrangement of (52) gives us

$$\frac{|x|^{1+\alpha} f_\alpha(x)}{C_\alpha} = \frac{\int_0^{\pi/2} \upsilon(\theta) \exp\left(-\upsilon(\theta)|x|^{-\alpha/(1-\alpha)}\right) d\theta}{|x|^{\frac{\alpha^2}{1-\alpha}} \int_0^{\pi/2} \upsilon(\theta) \exp(-\upsilon(\theta))\, d\theta} \tag{53}$$

where

$$C_\alpha = \frac{\alpha}{(1-\alpha)\pi} \int_0^{\pi/2} \upsilon(\theta) \exp(-\upsilon(\theta))\, d\theta. \tag{54}$$

Taking the limit as $\alpha \to 0$ in (53) leads to

$$\lim_{\alpha \to 0} \frac{|x|^{1+\alpha} f_\alpha(x)}{C_\alpha} = 1 \tag{55}$$

which is the desired result. ■

## Appendix III
## Proof of Theorem 4

*Proof:* We begin by stating the following lemma, whose proof follows this one.

*Lemma 1:* Let $\{f_\alpha\}_{\alpha>0}$ be purely algebraic, and let $C_\alpha$ be as in condition (ii) of Definition 5. Then, for any closed interval $[a, b]$, the function

$$\xi_\alpha(x) = \frac{C_\alpha}{f_\alpha(x)} - |x| \tag{56}$$

converges *uniformly* to 0 as $\alpha \to 0$.

According to this lemma, the density $f_\alpha(x)$ can be expressed as

$$f_\alpha(x) = \frac{C_\alpha}{|x| + \xi_\alpha(x)} \tag{57}$$

where $\xi_\alpha(x)$ is a "remainder" which converges uniformly to 0 as $\alpha \to 0$. Based on (57), the maximum likelihood estimator of location is given by

$$T_\alpha = \arg\min_\mu \prod_i \left[|x_i - \mu| + \xi_\alpha(x_i - \mu)\right]. \tag{58}$$

Note that the constant $C_\alpha$ has been dropped out since it is irrelevant for the minimization problem. Developing the products in (58), we get

$$T_\alpha = \arg\min_\mu H_\alpha(\mu) \tag{59}$$

where

$$H_\alpha(\mu) = \prod_i |x_i - \mu| + \sum_i \left\{ \xi_\alpha(x_i - \mu) \prod_{j, j\neq i} |x_j - \mu| \right\} + \sum_{i,j} \left\{ \xi_\alpha(x_i - \mu)\xi_\alpha(x_j - \mu) \prod_{k, k\neq(i,j)} |x_k - \mu| \right\} + \cdots. \tag{60}$$

Let $x_{\min}$ and $x_{\max}$ denote, respectively, the minimum and maximum values in the sample set. The unimodality of $f_\alpha$ guarantees that $H_\alpha(\mu) > H_\alpha(x_{\min})$ for $\mu < x_{\min}$, and $H_\alpha(\mu) > H_\alpha(x_{\max})$ for $\mu > x_{\max}$. This restricts the location of $T_\alpha$ to the closed interval $[x_{\min}, x_{\max}]$, and thus, thanks to Lemma 1, all but the first term in (60) can be made arbitrarily small by letting $\alpha \to 0$. This tells us that, as $\alpha \to 0$, $T_\alpha$ tends to have a "selection" behavior, in the sense that it converges to one of the sample values.

The specific point of convergence can be identified by analyzing the behavior of the cost function $H_\alpha(\mu)$ in the vicinity of each sample value. To do so, let $r_j$ denote the number of times the value of $x_j$ is repeated in the sample. Then, for $\mu$ close to $x_j$, the first $r_j$ terms in (60) can be made arbitrarily small, independently of $\alpha$. Making $\alpha$ small, the magnitude of $H_\alpha(\mu)$ in the vicinity of $\mu = x_j$ is thus driven by the $(r_j+1)$th term in (60), which can be continuously approximated by

$$\lim_{\mu\to x_j} \sum_{i_1,\ldots,i_{r_j}} \xi_\alpha(x_{i_1} - \mu)\cdots\xi_\alpha\left(x_{i_{r_j}} - \mu\right) \times \prod_{k, k\neq\left(i_1,\ldots,i_{r_j}\right)} |x_k - \mu| = [\xi_\alpha(0)]^{r_j} \prod_{i, x_i\neq x_j} |x_i - x_j|. \tag{61}$$

Hence, the point of convergence is the value $x_j$ with the minimum associated cost (61), for small values of $\alpha$. Since $\xi_\alpha(0) \to 0$ as $\alpha \to 0$, minimizing (61) for small values of $\alpha$ requires the maximization of $r_j$, which implies the convergence of $T_\alpha$ to one of the most repeated values in the sample. Taking this fact into account, the minimization of (61) leads to the simplified result

$$\lim_{\alpha\to 0} T_\alpha = \arg\min_{x_j\in\mathcal{M}} \prod_{i, x_i\neq x_j} |x_i - x_j| \tag{62}$$

where $\mathcal{M}$ is, as before, the set of most repeated values in the sample. ■

## Appendix IV
## Proof of Lemma 1

Let us suppose that $\epsilon > 0$ is given, and let $I_\epsilon = [a_\epsilon, b_\epsilon]$ be a closed interval in $[a, b]$ with length less than $\epsilon/2$. We will show that there exists a bound $\alpha_0$ such that, for every $x \in I_\epsilon$,

$$\left| \frac{C_\alpha}{f_\alpha(x)} - |x| \right| < \epsilon \qquad \text{if} \quad 0 < \alpha < \alpha_0. \tag{63}$$

To begin, let us assume that $a_\epsilon \geq 0$. Since by definition $C_\alpha / f_\alpha(x) \to |x|$ for all $x$, there exist constants $\alpha_a$ and $\alpha_b$ such that

$$a_\epsilon - \frac{C_\alpha}{f_\alpha(a_\epsilon)} < \epsilon/2 \qquad \text{if} \quad 0 < \alpha < \alpha_a \tag{64}$$

and

$$\frac{C_\alpha}{f_\alpha(b_\epsilon)} - b_\epsilon < \epsilon/2 \qquad \text{if} \quad 0 < \alpha < \alpha_b. \tag{65}$$

Let $\alpha_0 = \min\{\alpha_a, \alpha_b\}$, and let $x$ be an arbitrary point in $I_\epsilon$. Since the construction of $I_\epsilon$ guarantees that $x - a_\epsilon < \epsilon/2$, it follows from (64) that

$$x - \epsilon < \frac{C_\alpha}{f_\alpha(a_\epsilon)} \qquad \text{if} \quad 0 < \alpha < \alpha_0. \tag{66}$$

Similarly, $b_\epsilon - x < \epsilon/2$ for all $x$ in $I_\epsilon$, and it follows from (65) that

$$\frac{C_\alpha}{f_\alpha(b_\epsilon)} < x + \epsilon, \qquad \text{if} \quad 0 < \alpha < \alpha_0. \tag{67}$$

Recalling the monotonicity of $f_\alpha$,

$$\frac{C_\alpha}{f_\alpha(a_\epsilon)} \leq \frac{C_\alpha}{f_\alpha(x)} \leq \frac{C_\alpha}{f_\alpha(b_\epsilon)} \tag{68}$$

and the subsequent application of (66) and (67) leads us to

$$x - \epsilon < \frac{C_\alpha}{f_\alpha(x)} < x + \epsilon. \tag{69}$$

This proves (63) for nonnegative intervals. The case in which $a_\epsilon \leq b_\epsilon \leq 0$, can be proven in a similar fashion. For the "mixed" case in which $a_\epsilon < 0$ and $b_\epsilon > 0$, it is sufficient to observe that

$$I_\epsilon = [a_\epsilon, 0] \cup [0, b_\epsilon] \tag{70}$$

and the result in (63) can be applied to the smaller intervals $[a_\epsilon, 0]$ and $[0, b_\epsilon]$.

Now, let $I_k$ denote a finite collection of closed intervals in $[a, b]$, such that

1) $\bigcup_k I_k = [a, b]$.
2) For every $k$, the length of $I_k$ is less than $\epsilon/2$.

The existence of such a collection is guaranteed by the compactness of $[a, b]$. Let $\alpha_k$ denote the bound value corresponding to $I_k$ according to the derivation of (63), and let $\delta = \min_k\{\alpha_k\}$. Then, for every $x \in [a, b]$

$$\left|\frac{C_\alpha}{f_\alpha(x)} - |x|\right| < \epsilon \qquad \text{if} \quad 0 < \alpha < \delta. \tag{71}$$

This completes the proof. ■

## Acknowledgment

We would like to thank Robust Analysis, Inc. for providing access to its comprehensive suite of numerical routines for stable distributions which proved invaluable for several of our simulation studies [44]. We are also thankful to J. Bacca for his assistance with several simulations reported in this paper.

## References

[1] J. M. Mendel, “Tutorial on higher order statistics (spectra) in signal processing, and system theory: Theoretical results and some applications,” *Proc. IEEE*, vol. 79, no. 3, 1991.

[2] C. L. Nikias and A. T. Petropulu*, Higher-Order Spectra Analyisis: A Nonlinear Signal Processing Framework*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[3] R. J. Adler, R. E. Feldman, and M. S. T. , Eds.*, A Practical Guide to Heavy Tails-Statistical Techniques and Applications*. Boston, MA: Birkhauser, 1998.

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

[5] G. Samorodnistsky and M. S. Taqqu*, Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance*. New York: Chapman & Hall, 1994.

[6] H. M. Hall, “A new model for ‘impulsive’ phenomena: Application to atmosheric-noise communication channels,” Stanford Electron. Lab., Stanford Univ., Stanford, CA, Tech. Rep. 3412-8 and 7050-7, Aug. 1966.

[7] J. H. Miller and J. B. Thomas, “Detectors for discrete-time signals in non-Gaussian noise,” *IEEE Trans. Inf. Theory*, vol. 18, pp. 241–250, Mar. 1972.

[8] P. R. Rider, “Generalized Cauchy distributions,” *Ann. Inst. Stat. Math.*, vol. 9, 1957.

[9] K. G. Hagstroem, “Remarks on Pareto distributions,” Skandinavisk Aktuarietidskrift pp. 59–71, 1960.

[10] V. Pareto*, Cours d’Economie Politique*. Paris, France: Rouge and Cie, 1897.

[11] D. W. Griffith, Jr., J. G. Gonzalez, and G. R. Arce, “Robust time-frequency representations for signals in alpha-stable noise using fractional lower-order statistics,” presented at the IEEE Signal Process Workshop in Higher Order Statistics, Banff, Canada, Jul. 21–23, 1997.

[12] X. Ma and C. L. Nikias, “Joint estimation of time delay and frequency delay in impulsive noise using fractional lower-order statistics,” *IEEE Trans. Signal Process.*, vol. 44, no. 11, pp. 2669–2687, Nov. 1996.

[13] T. C. Chuah, “Distance metric for soft-decision decoding in non-Gaussian channels,” *Electron. Lett.*, vol. 39, pp. 1062–1063, 2003.

[14] ——, “Enhanced sequence detector in the presence of intersymbol interference and non-Gaussian noise,” *Electron. Lett.*, vol. 40, pp. 369–371, 2004.

[15] J. Choi, M. Bouchard, and T. H. Yeap, “Recurrent neural equalization for communication channels in impulsive noise environments,” in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN 2005)*, Montreal, QC, Canada, Jul. 31–Aug. 4, 2005 .

[16] T. C. Chuah, “Robust iterative decoding of turbo codes in heavy-tailed noise,” *Inst. Elect. Eng. Proc.—Commun.*, vol. 152, pp. 29–38, 2005.

[17] ——, “Turbo equalisation in non-Gaussian impulsive noise,” *Inst. Elect. Eng. Proc. Visual, Image Signal Process.*, vol. 152, pp. 52–60, 2005.

[18] T. C. Chuah, B. S. Sharif, and O. R. Hinton, “Robust adaptive spread-spectrum receiver with neural-net preprocessing in non-Gaussian noise,” *IEEE Trans. Neural Netw.*, vol. 12, no. 3, pp. 546–558, May 2001.

[19] T. C. Chuah, B. S. Sharif, and O. R. Hinton, “Robust signal processing for cdma communications in non-Gaussian noise,” presented at the 5th Int. Conf. Mathematics in Signal Process., Warwick, U.K., 2000.

[20] ——, “Nonlinear decorrelator for multiuser detection in non-Gaussian impulsive environments,” *Electron. Lett.*, vol. 36, pp. 920–922, 2000.

[21] S. Yoon, I. Song, and S. Y. Kim, “Code acquisition for ds/ss communications in non-Gaussian impulsive channels,” *IEEE Trans. Commun.*, vol. 52, pp. 187–190, Feb. 2004.

[22] A. Hmidat, B. Sharif, and W. Woo, “Robust fuzzy-based multiuser detector for non-Gaussian CDMA channel,” *Inst. Elect. Eng. Electron. Lett.*, vol. 40, no. 15, pp. 948–949, 2004.

[23] A. Hmidat, B. Sharif, W. Woo, and S. Jimaa, “Nonlinear space-time multiuser detector for non-Gaussian channels,” presented at the 9th Int. Symp. Comput. Commun. (ISCC’2004), Alexandria, Egypt, 2004.

[24] M. Gok and A. Cetin, “Moving object detection in video by detecting non-Gaussian regions in subbands and active contours,” in *Proc. 2003 IEEE Int. Conf. Image Process. (ICIP’2003)*, Sep. 14–17, 2003, vol. 3, pp. 965–968.

[25] P. Zurbach, “Weighted myriad filters for image processing,” Master’s thesis, Dept. Elect. Eng., Univ. Delaware, Newark, DE, 1996.

[26] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, “Robust iterative fitting of multilinear models,” *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2678–2689, Aug. 2005.

[27] J. G. Gonzalez, “Robust techniques for wireless communications in non-Gaussian environments,” Ph.D. dissertation, Dept. Electr. Eng., Univ. of Delaware, Newark, DE, 1997.

[28] R. Durrett*, Probability: Theory and Examples*, 2nd ed. Belmont: Duxbury, 1996.

[29] J. Ilow and D. Hatzinakos, “Analytic alpha-stable noise modeling in a Poisson field of interferers or scatterers,” *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1601–1611, Jun. 1998.

[30] P. G. Georgiou, P. Tsakalides, and C. Kyriakakis, “Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise,” *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 291–301, Sep. 1999.

[31] S. Kalluri and G. R. Arce, “Adaptive weighted myriad filter algorithms for robust signal processing in alpha-stable noise environments,” *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 322–334, Feb. 1998.

[32] R. F. Brcich, D. R. Iskander, and A. M. Zoubir, “The stability test for symmetric alpha-stable distributions,” *IEEE Trans. Signal Process.*, vol. 53, no. 3, pp. 977–986, Mar. 2005.

[33] J. G. Gonzalez and G. R. Arce, “Optimality of the myriad filter in practical impulsive-noise environments,” *IEEE Trans. Signal Process.*, vol. 49, no. 2, pp. 438–441, Feb. 2001.

[34] S. Kogon and D. Williams*, A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Boston, MA: Birkhäuser, 1998, ch. Characteristic Function Based Estimation of Stable Distribution Parameters.

[35] S. T. Rachev, Ed.*, Handbook of Heavy Tailed Distributions in Finance*. Amsterdam, The Netherlands: Elsevier, 2003.

[36] G. Arce*, Nonlinear Signal Processing: A Statistical Approach*. New York: Wiley-Interscience, 2004.

[37] M. L. Ryznar*, Potential Analysis*. New York: Springer Science+Business Media B.V., 2002, pp. 1–23, Estimates of Green Function for Relativistic $\alpha$-Stable Process.

[38] V. M. Zolotarev*, One-Dimensional Stable Distributions*. Providence, RI: Amer. Math. Soc., 1986, translation from Russian, 1983.

[39] J. G. Gonzalez, D. W. Griffith, Jr., and G. R. Arce, "Matched myriad filters for robust communications," presented at the 1996 Conf. Inf. Sci. Syst., Princeton, NJ, Mar. 20–22, 1996.

[40] I. Berkes, L. Horvath, and D. Khoshnevisan, "Logarithmic averages of stable rv are asymptotically normal," *Stoch. Process. Appl.*, vol. 77, pp. 35–51, 1998.

[41] A. N. M. Muniruzzaman, "On measures of location and dispersion and tests of hypotheses on a Pareto population," *Bull. Calcuta Statist. Assoc.*, vol. 7, pp. 115–123, 1957.

[42] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[43] S. Dharmadhikari and K. Joag-Dev, *Unimodality, Convexity, and Applications*. New York: Academic, 1988.

[44] Robust Analysis, Inc.. Tokaoma Park, MD, User Manual for the STABLE 3.14 Library [Online]. Available: http://www.robustanalysis.com 2005

[45] J. T. Astola and T. G. Campbell, "On computation of the running median," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 4, pp. 572–574, Apr. 1989.

[46] W. H. DuMouchel, "On the asymptotic normality of the maximum likelihood estimate when sampling from a stable distribution," *Ann. Stat.*, vol. 1, no. 5, pp. 948–957, 1973.


**Juan G. Gonzalez** holds several advanced degrees in electrical engineering and applied mathematics, including the Ph.D. degree in electrical engineering (with honors) from the University of Delaware, Newark. His doctoral work introduced the concepts that led to the invention of the myriad filter, a digital filtering framework with important optimality properties in practical impulsive noise environments. Myriad filters are currently taught at major universities, and have been successfully used in fields as varied as communications, biomedical engineering, astronomy, statistics, and digital imaging.

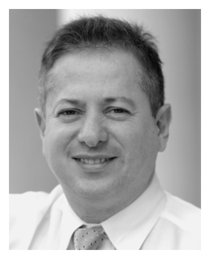
He has held research appointments with Bell Laboratories and the University of Delaware, and has been a cofounder and/or an R&D executive for several technology start-ups in the field of algorithm development. He has received numerous awards for his technical contributions, including the 1997 Allan P. Colburn Prize from the University of Delaware, the 1995 Whitaker Best Paper Award from the Rehabilitation Engineering Society of North America, and the Honor Degree Distinction in Electrical Engineering from the Universidad Pontificia Bolivariana, Medellin, Colombia.

Dr. Gonzalez is listed in *Who's Who in America*, *Who's Who in Science and Technology*, and *Who's Who in Finance and Business*.

**José L. Paredes** (M'00) was born in Mérida, Venezuela. He received the Diploma in electrical engineering from the Universidad de Los Andes, Mérida, in 1995, being the first student to graduate *summa cum laude* from that department. He received the M.S. and Ph.D. degrees in electrical engineering from the University of Delaware, Newark, in 1999 and 2001, respectively.

From 1996 to 2001, he was a Research Assistant with the Department of Electrical and Computer Engineering, University of Delaware. Since 2001, he has been a Professor with the Electrical Engineering Department, Universidad de Los Andes. He has consulted with industry in the areas of digital communication and signal processing. His research interests include robust signal and image processing, nonlinear filter theory, adaptive signal processing, and digital communications. He holds one U.S. patent.

Dr. Paredes is the recipient of several awards. In 1997, he won the Best Paper award from the U.S. Army's ATIRP Federated Laboratory Consortium. He also was awarded three scholarships from the Government of Venezuela (in 1988, 1992, and 1996).



**Gonzalo R. Arce** (F'00) received the Ph.D. degree from Purdue University, West Lafayette, IN, in 1982.

Since 1982 he has been a member of the faculty of the Department of Electrical and Computer Engineering, University of Delaware, Newark, where he is the Charles Black Evans Distinguished Professor and Department Chairman. His research interests include statistical and nonlinear signal processing, multimedia security, electronic imaging, and signal processing for communications and networks. He is a coauthor of the textbooks *Digital Halftoning* (New York: Marcel Dekker, 2001), *Nonlinear Signal Processing and Applications* (Boca Raton, FL: CRC, 2003, and *Nonlinear Signal Processing: A Statistical Approach* (New York: Wiley, 2004). He is a frequent consultant to industry in the areas of image printing and digital video. He holds 10 U.S. patents.

Dr. Arce received the NSF Research Initiation Award. He is a Fellow of the IEEE for his contributions on nonlinear signal processing and its applications. He was the co-chair of the 2001 EUSIPCO/IEEE Workshop on Nonlinear Signal and Image Processing (NSIP'01), co-chair of 1991 SPIE's Symposium on Nonlinear Electronic Imaging, and the co-chair of the 2002 and 2003 SPIE ITCOM conferences. He has served as Associate Editor for the IEEE TRANSACTIONS FOR SIGNAL PROCESSING, Senior Editor of the *Applied Signal Processing Journal*, as Guest Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, and as Guest Editor for *Optics Express*.