USCIPI REPORT #200

Signal Processing with Fractional Lower Order Moments: Stable Processes and Their Applications

by

Min Shao and Chrysostomos L. Nikias

March 1992

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.

Abstract

Non-Gaussian signal processing is becoming increasingly important as more and more phenomena in signal processing are found to deviate from the ideal Gaussian model. Stable distributions are among the most important non-Gaussian models. They share important characteristics with the Gaussian distribution, such as the stability property and central limit theorems, and have found applications in such diverse fields as physics, economics as well as electrical engineering. To help engineers better understand stable models and develop methodologies for their applications in signal processing, this paper presents a tutorial review of the basic characteristics of stable laws and stable signal processing. The emphasis will be on the differences and similarities between stable signal processing methods based on fractional lower order moments and Gaussian signal processing methods based on second-order moments.

1 Introduction

One of the basic objectives of signal processing is to extract desired information from observed data (signals). Since in most cases signals are either non-deterministic or contaminated by random noise, mathematical statistics plays an important role in signal processing [69, 75, 52]. A general formulation of statistical theory as methods of signal processing assumes the knowledge of a probability model which describes, either partially or completely, the way in which the observed signals and noise are generated. The probability model for the underlying signals and noise is usually a function of the desired information which is often parameterized by a set of parameters. This set of parameters are then determined from the data by using the probability model and certain optimality criteria. Obviously, the accuracy of the information obtained in this way depends heavily on the probability model and optimality principles used in the inference.

1.1 Gaussian Signal Processing with Second-Order Moments

The signal processing literature has traditionally been dominated by a Gaussian assumption and a linear least-squares criterion of optimality. Even when the Gaussian assumption is not made explicitly, signal processing problems and their solutions are often stated in terms of the second-order statistical properties of underlying signals and noise. Since Gaussian processes are completely described by the second-order structures, this second-order approach is, in most cases, equivalent to explicitly making the Gaussian assumption. The second-order formulation together with the linear least-squares criterion form the foundation of stochastic signal modeling and processing and have, over the last fifty years, provided many important concepts and structures in signal processing, such as spectral representation, adaptive filtering and prediction theory.

In many instances the Gaussian assumption is a reasonable one and can usually be justified by the Central Limit Theorem. The linear least-squares criterion, on the other hand, is intimately related with the Gaussian assumption. For example, under the Gaussian assumption, it can easily be shown that the linear least-squares estimator is the same as the maximum likelihood estimator and hence, is asymptotically efficient [52, 5]. Another main reason for the dominance of the Gaussian assumption and linear least-squares criterion is that they often lead to analytically tractable solutions for signal processing problems. For example, the additive white Gaussian noise assumption in communication theory greatly

simplifies the design and analysis of receiver structures. Any non-Gaussian assumptions will usually introduce nonlinearity to the design of receivers. This explains why the ideal Gaussian model is often used even when it is found to be not very accurate. This was especially true in the early days when numerical computations were expensive.

Unfortunately, many signals and noise sources encountered in practice are decidedly non-Gaussian [85, 38, 84]. For example, underwater acoustic signals, low-frequency atmospheric noise and most of the man-made noises are found to be non-Gaussian [8, 53, 41, 76, 45, 54]. Non-Gaussianity often results in significant performance degradation for systems optimized under the Gaussian assumption. A well-known example is the matched filter for coherent reception of deterministic signals in Gaussian white noise. If the noise statistics deviate from the Gaussian model, serious degradation in performance occurs, such as increased false alarm rate or error probability [73, 72, 35, 34]. On the other hand, a modest degree of nonlinear signal processing based on the approximately true noise statistics can lead to a much better receiver than the matched filter [55, 41, 62, 50]. Thus, there is a trade-off between model complexity and accuracy. Generally speaking, the more realistic (and complicated) signal models are, the more complex signal processing algorithms become.

The role of the least-squares criterion (i.e., the L_2 norm) in signal processing has also been under re-examination. Although it is adequate under the Gaussian assumption and usually leads to analytically tractable results, the least-squares criterion is no longer appropriate in a non-Gaussian environment, largely due to its non-robustness against a small number of big errors (outliers) in the data set [7, 26]. When the least-squares criterion is used, little attention is paid to relatively minor errors in order to make very large errors as small as possible. In many situations in signal processing, however, it is more important to make as many errors small as possible, even if it is necessary to tolerate occasional large errors. When the error distribution is Gaussian, it does not matter which criterion is used because the most probable error is small. However, it can easily be demonstrated that the least-squares estimates change dramatically when only a small proportion of extreme observations is present in the data [81]. In such cases more robust optimality criteria and procedures are needed [17, 15, 82, 36, 68].

1.2 Non-Gaussian Signal Processing with Higher Order Moments

For the past few years, non-Gaussian signal processing has received more and more attention in the literature for several reasons. One main reason is recent advances in computer software and hardware. The tremendous growth in computational power, backed by cheap VLSI hardware, makes it possible to implement very sophisticated signal processing algorithms. Computational complexity is no longer a major concern in designing signal processing systems. More realistic stochastic models can now be used to describe signals and noise. Another reason for the recent active research in non-Gaussian signal processing is the growing demand for products that serve real-world. The loss of resolution due to the Gaussian assumption in a non-Gaussian environment is no longer tolerable. Nowadays, when facing the trade-off between computational complexity of signal processing systems and realistic modeling of signals and noise, more often than not, researchers in signal processing will choose the latter. These two main factors have led to a significant amount of research activities in re-examining structures and inference methods in the context of filtering, estimation, detection and signal extraction under non-Gaussian environments [85, 84, 38].

For non-Gaussian signal processing second-order descriptions are no longer adequate. They do not provide enough information for understanding the non-Gaussianity involved. In this case, it has been suggested in the literature [58] that we should look beyond the second-order statistics and extract information about deviations from Gaussianity from higher order statistics, if they exist. A well-known example is that second-order descriptions or power spectrum suppress phase information but polyspectra preserve both magnitude and phase information of non-Gaussian signals [43]. Thus, in the cases where accurate non-minimum phase information is desired, one has to utilize higher-order statistics. This is especially important in deconvolution problems that arise in geophysics, telecommunications, etc., in which the non-minimum phase of signals must be preserved. Other attractive features of higher order spectra include their abilities to suppress Gaussian noise and detect and characterize nonlinearities in a time series [58].

Although higher order spectral analysis is a powerful tool for non-Gaussian signal processing, it has its own limitations. First of all, higher order spectral analysis is applicable only if the higher order statistics exist. Although this may not seem to be a major prob-

lem, we will soon see that there are important applications where the underlying signals do not even have second-order moments. Secondly, there is still lack of robust procedures for estimating higher order statistics from noisy observations. Finally, there is still no unifying theory for understanding higher order spectra as a whole in the context of filtering, estimation, detection and signal extraction. In particular, unlike the Gaussian case, signal processing using higher order statistics rarely has adequate optimality criteria. This makes it difficult to analyze and compare analytically signal processing algorithms based on higher order statistics.

1.3 Non-Gaussian Stable Signal Processing with Lower Order Fractional Moments

A broad and increasingly important class of statistical models for non-Gaussian phenomena is that of so-called heavy-tailed distributions, whose density functions decay in the tails less rapidly than the Gaussian density functions. These heavy-tailed distributions tend to produce large-amplitude excursions from the average value more frequently than the Gaussian distribution. As a result, they usually have very large or infinite variances, in which case the second-order moment theory and higher order spectral analysis methods no longer apply. Optimal signal processing algorithms have to be developed based on the complete signal statistics instead of a few moments.

Among all the heavy-tailed distributions, the family of *stable distributions*[1] has been found to provide useful models for phenomena observed in many diverse fields, such as economics, physics and electrical engineering [90]. It includes the Gaussian and Cauchy distributions as special cases and is capable of modeling a wide variety of non-Gaussian phenomena, from those similar to the Gaussian to those similar to the Cauchy. It also enjoys similar properties that have made the Gaussian distribution popular. For example, a basic feature of stable distributions is the so-called *stability property*: if X, Y are two independent stable random variables with the same characteristic exponent, then $aX + bY$ is also stable. In addition, stable distributions form a natural class of limit distributions, just like the Gaussian distributions. Recall that the Central Limit Theorem says that a physical phenomenon is Gaussian if there are infinitely many independent identically

[1] For a detailed summary of stable distributions and the Generalized Central Limit Theorem, see the Appendix.

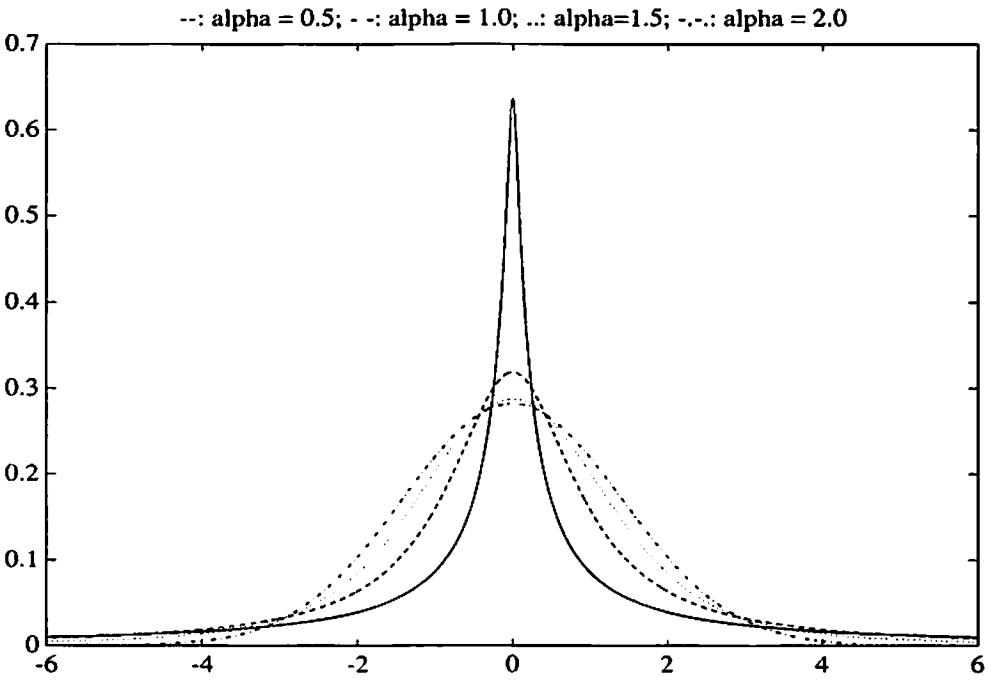

Figure 1: Density functions of $S\alpha S$ distributions

distributed (i.i.d) contributing factors with finite variances. There is, however, a more powerful theorem, called the Generalized Central Limit Theorem, for the sum of infinitely many i.i.d random variables. This theorem states that if the sum of any i.i.d random variables with or without variances converges to a distribution by increasing the number of variables, the limit distribution must be one of the stable laws [23, 9]. Thus, non-Gaussian stable distributions arise as sums of random variables in the same way as the Gaussian distributions. This suggests that we may justify the use of stable laws to describe non-Gaussian phenomena as we justify the use of Gaussian laws.

Although skewed stable distributions are important in certain applications [46], we will focus our attention on symmetric α-stable ($S\alpha S$) distributions in this paper. For simplicity, all stable distributions are assumed to be non-Gaussian $S\alpha S$, unless specified otherwise. The standard $S\alpha S$ density functions for some values of the characteristic exponent α are

shown in Figure 1, with $\alpha = 2$ corresponding to the zero-mean Gaussian distribution with variance equal to 2 and $\alpha = 1$ corresponding to the Cauchy distribution. Observe that $S\alpha S$ distributions have many features of the Gaussian. They are smooth, unimodal, symmetric with respect to the median and bell-shaped. A detailed comparison between the standard normal and $S\alpha S$ distributions shows that non-Gaussian stable distributions depart from the corresponding Gaussian distributions in the following ways. For small $|x|$, the $S\alpha S$ distributions are more peaked (have higher densities) than the normal. For some intermediate range of $|x|$, the $S\alpha S$ distributions have lower densities than the normal. Most importantly, unlike the Gaussian distributions which have exponential tails, the stable distributions have algebraic tails [25]. Thus $S\alpha S$ distributions have heavier (longer) tails than the Gaussian (higher kurtosis) and have no second moments. The smaller α is, the longer the tail. This is a desirable feature for many applications in signal processing since a lot of non-Gaussian phenomena are similar to the Gaussian, only have longer tails [53, 41, 45, 67].

Because of the importance of stable models for non-Gaussian phenomena appearing in a variety of different fields, it is very desirable to develop a theory of linear estimations of stable processes, including predictions, filtering and identifications, similar to that of Gaussian processes (second-order processes, in general). Such a theory will be especially useful for the optimal processing of signals that are nearly Gaussian and have large variances.

From the signal processing point of view, the adoption of a stable model for signals or noise has important consequences. In this case, the traditional *minimum mean square error* (MMSE) criterion is meaningless because of infinite variance. Instead, the *minimum dispersion* (MD) criterion is used as a measure of optimality. The dispersion (i.e., the scale parameter) of a stable random variable plays a analogous role of the variance. For example, the larger the dispersion, the more spread the stable random variable around the median. Thus, by minimizing the error dispersion, we minimize the average magnitude of estimation error. Furthermore, it has been shown that minimizing the dispersion is also equivalent to minimizing the probability of large estimation errors [16]. The MD criterion is thus well justified in the stable case. It is a direct generalization of the MMSE criterion (MD and MMSE criteria are the same in the Gaussian case) and reasonably simple to calculate.

It will be shown that the MD criterion is also equivalent to minimizing the fractional lower order moments (FLOM's) of estimation errors. These FLOM's measure the L_p

distance between an estimate and its true value, for $p < \alpha$. This result is not surprising since the L_p norms for $p < 2$ are well known for being robust against outliers such as those that may be described by stable laws. It will also be shown that all of the FLOM's are equivalent for stable random variables. A common choice is the L_1 moment, which is sometimes very convenient.

Stable signal processing based on FLOM's will inevitably introduce nonlinearity to even linear problems. The basic reason for the nonlinearity is that we have to solve linear estimation problems in Banach or metric spaces instead of Hilbert spaces. It is well-known that, while the linear space generated by a Gaussian process is a Hilbert space, the linear space of a stable process is a Banach space when $1 \leq \alpha < 2$ and only a metric space when $0 < \alpha < 1$ [13]. Banach or metric spaces do not have as nice properties and structures as Hilbert spaces for linear estimation problems. For example, while any finite number of Gaussian random variables can be expressed as linear combinations of *independent* Gaussian random variables, it is shown in [70] that representation of even two stable variables of the same characteristic exponent as linear combinations of finitely many independent stable variables is impossible.

Despite the difficulties, significant progresses have been made in developing the linear estimation theory for stable processes over the past thirty years. In this paper, we will give an overview of some of the results in which researchers in signal processing community might be interested.

1.4 Applications of Non-Gaussian Stable Signal Processing

The concept of stable distributions was first introduced by Lévy in the study of generalized central limit theorems [42]. They are direct generalizations of Gaussian distributions and share a lot of useful properties of Gaussian laws. Yet, despite all the attractive properties, the stable laws have attracted little attention from researchers in signal processing [80, 79, 13]. There are basically two reasons for this. First, stable laws do not have explicit expressions for their densities or distributions except for the Gaussian ($\alpha = 2$), Cauchy ($\alpha = 1$) and Pearson ($\alpha = \frac{1}{2}$) distributions. But power series expansions do exist for the probability densities functions. With today's computational power, numerical integrations are inexpensive to carry out. Moreover, much of the work that ordinarily uses probability density functions can be carried out in the transformed domain of characteristic functions. The second main reason for the obscurity of stable laws among engineers is apparently due

to the fact that the p^{th} moments of a $S\alpha S$ random variable exist only for $p < \alpha$ [23]. Thus for all non-Gaussian stable distributions there are no finite second-order moments and for $\alpha \leq 1$ even finite first-order moments do not exist. Since the second-order moment or the variance is often associated with the concept of power, it seems to be widely felt that infinite variance is inappropriate in almost any signal processing context.

We believe that this kind of reasoning is superficial and inappropriate. This is like saying that we should not use irrational numbers at all because we can never have irrationals from any physical measurements. Besides, as it is pointed out in [79], the Gaussian distribution would not be a physically appropriate assumption because it is unbounded. The purpose of mathematical modeling is not to account for every single detail of how a physical process is generated but to explain important and relevant features of the process in order to optimally and efficiently extract desired information from the observed data. For example, linear models in system theory are applicable only in a limited range. Beyond this range, nonlinear models must be used. By the same token, Gaussian models may be adequate for modeling a limited range of observed data. For a larger range, an infinite-variance stable model may have to be used [79]. This is especially true when outliers or heavy tails appear in the observed data. There are of course certain situations when neither Gaussian nor non-Gaussian stable distributions may be appropriate.

The most important features of non-Gaussian stable distributions are probabilistic stability and long inverse power tails, in addition to the basic fact that they arise naturally as limit distributions. If a physical phenomenon has both the stability property and long tails, stable distributions could provide useful models. Stability is a natural assumption. A lot of physical processes such as natural noise sources have this property. On the other hand, in recent years more and more physical phenomena are found to have heavier tails than the Gaussian [41, 24, 71, 67, 53, 57, 84]. Thus it can be expected that stable distributions and stable processes will provide useful models for many phenomena observed in diverse fields. Indeed, stable laws have seen applications in physics, economics, hydrology, biology as well as electrical engineering.

The earliest application of stable laws was discovered in physics by the Danish astronomer Holtsmark [31]. He found that random fluctuations of gravitational fields of stars in space under certain natural assumptions obey the stable law with characteristic exponent $\alpha = 1.5$. For other applications of stable laws in physics as well as other fields, see [86, 90, 88].

The recent surge of interest in stable laws is largely due to the work of Mandelbrot and his followers. Because of the failure of the Gaussian assumption and least-squares criterion in economic time series, he proposed a revolutionary approach based on stable distributions to the problem of price movement [46]. Many economical variables, such as common stock price changes, changes in speculative prices and interest rate changes have already been shown to have properties that conform closely to those of non-Gaussian stable laws [20, 64, 83]. Mandelbrot and van Ness also used Gaussian and stable fractional stochastic processes to describe long-range dependence arising in engineering, economics and hydrology [47].

An important applications of stable laws signal processing is in modeling impulsive noise. It is widely acknowledged that, in many communication problems, the conventional additive Gaussian noise assumption is inadequate. This is often due to the occurrence of noise with low probability but large amplitudes. This impulsive component of noise has been found to be significant in many problems, including atmospheric noise, underwater problems such as sonar and submarine communication, where the ambient acoustic noise may include impulses due to ice cracking in antic regions [85, 84, 38, 53, 67, 78, 24, 8, 76]. These types of impulsive noise are often observed to be close to Gaussian distributions but have heavier tails. For example, it has been reported in [45] that under-ice and shallow water noise has a distribution which is similar to that of Gaussian noise (symmetric, smooth, unimodal, etc.), but has heavier tails. The atmospheric noise may be considered as the result of a large number of independent sources in space so that central limit theorems apply. But they have been shown [41, 53, 67] to have algebraic tails $x^{-\alpha}$, for $1 < \alpha < 3$, a characteristic associated with stable laws. In [57], it is proposed to find a useful class of noise distributions with algebraic rather than exponential decay of the density of impulsive non-Gaussian noise that would approach the Gaussian distribution as some parameter approaches some limit. All these evidences suggest the use of stable laws as appropriate models.

In fact, the Cauchy distribution itself has been used in several papers such as [67] to represent severe impulsive noise. Stuck and Kleiner [80] empirically found that noise over certain telephone lines can be best described by stable laws with characteristic exponent α close to 2. They suggested that the design of receivers should take into account this noise characteristic. [3]. More recently, it has been shown theoretically that, under some conventional assumptions, a broad class of impulsive noise is indeed stable [74]. The

problem of signal detections under stable impulsive noise is also discussed in [74].

Stable laws are flexible in modeling impulsive noise in the sense that the characteristic exponent α may be used to control the degree of impulsiveness. A small value of α makes the noise highly impulsive, while a value of α close to 2 indicates a more Gaussian type of behavior. We do not think that lack of finite variance is a disadvantage. It just reflects the fact that the amount of outliers (noise with large amplitudes) is severe or simply unknown.

1.5 Objective

In this tutorial, we summarize some of the methodologies of processing non-Gaussian stable signals using FLOM's. The emphasis will be on demonstrating the differences and similarities between signal processing procedures based on stable distributions and FLOM's and those based on Gaussian distributions and second-order moments.

The objectives of this paper are:

1. To present the definitions, properties and applications of stable processes and FLOM's to the signal processing community and to aid engineers in utilizing stable laws to model and analyze non-Gaussian phenomena.

2. To demonstrate some of the gains in signal processing that can be obtained by adequately modeling non-Gaussian environments with stable processes and FLOM's.

3. To motivate further research in this area for developing efficient and more robust signal processing algorithms.

It should be pointed out that although stable laws and fractional lower order moment theory are in many ways similar to the Gaussian distributions and second-order moment theory, they are also substantially different from each other. Caution must be used when attempts are made to generalize the results for the Gaussian distributions to stable laws.

The organization of the paper is as follows. Section 2 presents some of the most frequently used properties of stable distributions and stable processes. Section 3 describes parameter estimation methods for $S\alpha S$ distributions. Procedures for estimating covariations are presented in Section 4. Section 5 discusses parameter estimation methods for AR, MA and $ARMA$ models with $S\alpha S$ processes as inputs, using FLOM's. Some of the results on filtering, predictions and identifications are presented in Section 6. Finally, Section 7 is devoted to concluding remarks.

2 Symmetric Stable Random Variables and Processes

In this paper, only $S\alpha S$ random variables will be considered. This is not a very restrictive assumption since most of the physical phenomena we deal with in practice are symmetric in nature. We also restrict our attentions to real random variables. For general discussions on complex $S\alpha S$ random variables, see [11].

Let us first briefly recall that a real random variable (r.v.) X is $S\alpha S$, if its characteristic function is of the form:

$$\varphi(t) = \exp\{i\delta t - \gamma|t|^\alpha\} \tag{1}$$

where α is the characteristic exponent, γ the dispersion and δ the location parameter. When $\alpha = 2$, X is Gaussian and when $\alpha = 1$, X is Cauchy. The real r.v.'s $X_1, \ldots, X_n$ are jointly $S\alpha S$, or the real random vector $\mathbf{X} = (X_1, \ldots, X_n)^T$ is $S\alpha S$, if and only if all the linear combinations $a_1X_1 + a_2X_2 + \cdots + a_nX_n$ are $S\alpha S$; or equivalently if their jointly characteristic function is of the form

$$\varphi(t) = \exp\{i\mathbf{t}^T\mathbf{a} - \int_S |\mathbf{t}^T\mathbf{s}|^\alpha \Gamma(d\mathbf{s})\} \tag{2}$$

where Γ is the spectral measure which is symmetric, i.e., $\Gamma(A) = \Gamma(-A)$ where A is any measurable set of the unit sphere S. A class of random variables $\{X(t), t \in T\}$ where T is an arbitrary index set is said to be a $S\alpha S$ stochastic process if for any $n \geq 1$, and distinct indices $t_1, \ldots, x_n \in T$, the random variables $X(t_1), \ldots, X(t_n)$ are jointly $S\alpha S$ with the same characteristic exponent α.

Just like the second-order case where we assume all the random variables have zero mean, it will be assumed that all $S\alpha S$ random variables have zero location parameters. For convenience, we consider only discrete-time $S\alpha S$ random processes in this paper.

2.1 Basic Types of Stable Processes

One of the difficulties in dealing with stable processes is due to the richness of the family of stable processes. There exist many types of stable processes with mutually exclusive properties. In the following, we discuss three types of stable processes that are commonly seen in practice.

2.1.1 Sub-Gaussian Processes

A stable process $\{X(t), t \in T\}$ is said to be an α-sub Gaussian process, or briefly α-SG(R), if for all $n \geq 1$ and all distinct indices $t_1, \ldots, t_n$, $(X(t_1), \ldots, X(t_n))$ has characteristic function given by

$$\varphi(\mathrm{u}) = \exp(-[\frac{1}{2}\sum_{m,n=1}^{n} u_m u_n R(t_m, t_n)]^{\alpha/2}) \tag{3}$$

where $R(t,s)$ is a positive definite function and α is restricted to $(1,2)$. When $\alpha = 2$, $X(t)$ is of course a Gaussian process with zero mean and covariance function $R(t,s)$. Obviously, the sub-Gaussian process is stationary if and only if $R(t,s) = R(t-s) = R(s-t)$.

It is well-known that sub-Gaussian distributions are variance mixtures of Gaussian distributions [13]. Specifically, if $X(t)$ is α-SG(R), then

$$X(t) = S^{1/2}Y(t) \tag{4}$$

where S is a positive $S\frac{\alpha}{2}S$ random variable with characteristic function $\exp(-|t|^{\alpha/2})$, $Y(t)$ a Gaussian process with zero mean and covariance function $R(t,s)$. In addition, S, $Y(t)$ are independent. Because of this, sub-Gaussian processes are among the simplest stable processes to deal with. They share many common features with the Gaussian processes. Yet they are also quite distinct from the Gaussian processes. For example, one of the striking properties about sub-Gaussian random variables is that they can not be independent [14].

2.1.2 Linear Stable Processes

Let $\{U(n), n = 0, \pm 1, \pm 2, \cdots\}$ be a family of i.i.d $S\alpha S$ random variables. Then

$$X(n) = \sum_{i=-\infty}^{\infty} a_i U(n-i)$$

defines a stationary $S\alpha S$ random process if $\sum_{i=-\infty}^{\infty} |a_i|^{\alpha-\delta} < \infty$ for some $0 < \delta < \alpha$ when $0 < \alpha < 1$, or if $\sum_{i=-\infty}^{\infty} |a_i| < \infty$ when $\alpha \geq 1$ [32]. Those processes are called *linear* stable processes or stable processes with moving-average representations.

Examples of linear stable processes include finite-order autoregressive (AR), moving-average (MA) and autoregressive moving-average ($ARMA$) processes. Specifically, let $a_1, \ldots, a_p$ be real numbers such that the polynomial $1 - \sum_{k=1}^{p} a_k z^{-k}$ has all zeros inside

unit circle in the complex plane. Then, the following equation

$$X(n) = a_1 X(n-1) + \cdots + a_p X(n-p) + U(n) \tag{5}$$

where $U(n)$'s are i.i.d $S\alpha S$ random variables with $\alpha > 1$, has a unique stationary $S\alpha S$ solution given by

$$X(n) = \sum_{k=0}^{\infty} h_k U(n-k) \tag{6}$$

where h_k is absolutely summable and

$$\frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \sum_{k=0}^{\infty} h_k z^{-k} \tag{7}$$

on the unit circle. For a generalization of this result, see [89]. The process $X(n)$ thus defined is called a p^{th} order AR stable process. Similarly, an $ARMA$ stable process of order (p, q) is defined as the unique stationary solution of the following equation

$$X(n) = a_1 X(n-1) + \cdots + a_p X(n-p) + b_0 U(n) + \cdots + b_q U(n-q) \tag{8}$$

where $U(n)$'s are i.i.d $S\alpha S$ random variables with $\alpha > 1$ and $a_1, \ldots, a_p$ are real numbers such that the polynomial $1 - \sum_{k=1}^{p} a_k z^{-k}$ has all zeros inside unit circle in the complex plane. If $a_1, \ldots, a_p$ are all zero, this process is called a q^{th} order MA stable process.

2.1.3 Harmonizable Stable Processes

It is well known that every wide sense stationary second-order random sequence $X(n)$ has a spectral representation of the form:

$$X(n) = \int_{-\pi}^{\pi} e^{in\omega} dZ(\omega) \tag{9}$$

where $Z(\omega)$ is a second-order random process with orthogonal increments, defined on $(-\pi, \pi]$ [59]. Although not every stationary stable random process has a spectral representation, Equation (9) does define an important type of stable processes called the *harmonizable* stable processes, under some appropriate interpretation of the integral and some conditions on $Z(\omega)$. For the details, see [13].

It is well known that the classes of linear and harmonizable Gaussian processes coincide.

But in the stable case they are mutually exclusive. Namely, harmonizable stable processes have no moving-average representations and stable processes with moving-average representations are not harmonizable. In addition, sub-Gaussian processes are neither linear nor harmonizable [14].

2.2 Linear Space of $S\alpha S$ Random Processes

In the second-order moment theory, it is shown that the set of random variables with finite variances forms a Hilbert space. The norm of a random variable is simply defined to be its second-order moment. In this section we show that the linear space of a $S\alpha S$ process is a Banach space if $1 \leq \alpha < 2$ and a metric space if $0 < \alpha < 1$.

Let X be a $S\alpha S$ random variable with dispersion $\gamma > 0$ and zero location parameter. Define the norm of X as

$$\|X\|_\alpha = \begin{cases} \gamma^{\frac{1}{\alpha}} & 1 \leq \alpha \leq 2 \\ \gamma & 0 < \alpha < 1 \end{cases} \tag{10}$$

Thus the norm $\|X\|_\alpha$ determines the distribution of X via the characteristic function

$$\varphi(t) = \begin{cases} \exp\{-\|X\|_\alpha^\alpha |t|^\alpha\} & 1 \leq \alpha \leq 2 \\ \exp\{-\|X\| |t|^\alpha\} & 0 < \alpha < 1 \end{cases}$$

The following Proposition is often found to be very useful [70]:

Proposition 1 *If (X, Y) are $S\alpha S$ and if X is independent of Y, then $\|X + Y\|_\alpha = \|X\|_\alpha + \|Y\|_\alpha$ if $0 < \alpha < 1$, and $\|X + Y\|_\alpha^\alpha = \|X\|_\alpha^\alpha + \|Y\|_\alpha^\alpha$ if $1 \leq \alpha \leq 2$.*

If X, Y are jointly $S\alpha S$, the distance between X and Y is defined as

$$d(X, Y) = \|X - Y\|_\alpha \tag{11}$$

To see how this distance between two $S\alpha S$ random variables measures the dissimilarity between them, we notice that all $S\alpha S$ random variables belong to $L_p(\Omega)$ (the collection of random variables with finite p^{th} order moments) for $0 < p < \alpha$. The following Proposition provides the link between the norm of a $S\alpha S$ random variable and its FLOM's [13, 90]:

Proposition 2 *Let X be $S\alpha S$ with zero location parameter and $0 < p < \alpha$. Then*

$$(\mathbf{E}\{|X|^p\})^{1/p} = C(p, \alpha)\|X\|_\alpha \tag{12}$$

where

$$C(p,\alpha) = [2\Gamma(1-p/\alpha)\Gamma(p)\sin\frac{\pi}{2}p]^{\frac{1}{p}}$$

depending only on α and p. A sequence of $S\alpha S$ random variables X_n converges to a $S\alpha S$ Y in $\|\cdot\|_\alpha$ if and only if X_n converges to Y in probability.

Thus, the distance between two $S\alpha S$ random variables measures the pth order moment of the difference of these two random variables. In the case of $\alpha = 2$ this distance measures the variance of the difference of the two random variables, which is consistent with the second-order moment theory. In addition, the lower order moments of a $S\alpha S$ random variable are equivalent in the sense that the p^{th} and q^{th} moments differ by a constant factor independent of the $S\alpha S$ random variable for all $p, q < \alpha$.

Let $\{X(t), t \in T\}$ be a $S\alpha S$ process. Then all finite linear combinations of elements in $\{X(t), t \in T\}$ form a linear space $l(X(t), t \in T)$. In this space all the random variables are jointly $S\alpha S$ with the same characteristic exponent [13]. The following theorem provides a metric on the space $l(X(t), t \in T)$ [70]:

Theorem 1 *For all $0 < \alpha \leq 2$, the distance defined in (11) is a true metric on $l(X(t), t \in T)$. In addition, for $1 \leq \alpha \leq 2$, $\|\cdot\|_\alpha$ is a norm on $l(X(t), t \in T)$.*

Let $L(X(t), t \in T)$ be the completion of $l(X(t), t \in T)$ with respect to this metric. It will be called the *linear space* of $\{X(t), t \in T\}$. It follows from the multivariate $S\alpha S$ characteristic functions that $L(X(t), t \in T)$ is a class of jointly $S\alpha S$ random variables. In particular, the random variables in the linear space of an α-sub-Gaussian process are α-sub-Gaussian [13].

As we will see later, the fundamental difficulty in stable signal processing with FLOM's is that the tools of Hilbert Space Theory are no longer applicable: although the linear space of a Gaussian process is a Hilbert Space, the linear space of a stable process is a Banach space for $1 \leq \alpha < 2$ and only a metric space for $0 < \alpha < 1$. For both mathematical and practical reasons we restrict α to $(1, 2]$ for the rest of the paper. Under this assumption, all $S\alpha S$ random variables considered in the following discussion have finite means which are assumed, without loss of generality, to be zero.

2.3 Covariations

The concept of covariance between two random variables plays an essential role in the second-order moment theory. The theory of signal prediction, filtering and smoothing, in fact, the whole theory of statistical signal processing is built on covariances. Obviously, covariances do not exist on the space of $S\alpha S$ random variables, due to the lack of finite variances. Instead, a quantity called *covariation* [56, 13], which under certain circumstances plays an analogous role for $S\alpha S$ random variables to covariance for random variables with finite second-order moments, has been proposed and is defined as follows. Let (X, Y) be a $S\alpha S$ vector with spectral measure Γ, the covariation $[X, Y]_\alpha$ of X with Y is defined as

$$[X, Y]_\alpha = \int_S xy^{<\alpha-1>}\Gamma(ds) \tag{13}$$

where $S = \{(x, y) \in R^2 : x^2 + y^2 = 1\}$ is the unit circle and for any real number z and $a > 0$ we use the convention

$$z^{<a>} = |z|^{a-1}z$$

In polar coordinate system, the covariation can be written as

$$[X, Y]_\alpha = \int_0^{2\pi} \cos\phi(\sin\phi)^{<\alpha-1>}\Gamma(d\phi) \tag{14}$$

Some of the useful properties of covariations [56, 13, 86] are:

1. The covariation $[X, Y]_\alpha$ is linear in X: if X_1, X_2, Y are jointly $S\alpha S$ then

$$[aX_1 + bX_2, Y]_\alpha = a[X_1, Y]_\alpha + b[X_2, Y]_\alpha \tag{15}$$

   for any a and b.

2. When $\alpha = 2$, i.e., when X, Y are jointly Gaussian with zero mean, the covariation of X with Y reduces to the covariance of X and Y:

$$[X, Y]_\alpha = \mathbf{E}(XY)$$

3. $[X, Y]_\alpha$ is not, in general, linear with respect to the second variable Y. But it does possess the following pseudo-linearity property with respect to Y: if Y_1, Y_2 are

independent and X, Y_1, Y_2 are jointly $S\alpha S$, then

$$[X, aY_1 + bY_2]_\alpha = a^{<\alpha-1>}[X, Y_1]_\alpha + b^{<\alpha-1>}[X, Y_2]_\alpha \tag{16}$$

4. If X, Y are independent and jointly $S\alpha S$, then

$$[X, Y]_\alpha = 0$$

while the converse is not true in general.

5. The dispersion of a $S\alpha S$ r.v X is equal to the covariation of X with itself, i.e.,

$$\|X\|_\alpha^\alpha = [X, X]_\alpha$$

6. For any jointly $S\alpha S$ random variables X, Y:

$$|[X, Y]_\alpha| \leq \|X\|_\alpha \|Y\|_\alpha^{<\alpha-1>}$$

2.4 Conditional Expectations and Linear Regressions

As an application of the concept of covariation, let us look at the problem of regressions with $S\alpha S$ random variables. Let $X_0, X_1, \ldots, X_n$ be jointly $S\alpha S$ random variables with $1 < \alpha \leq 2$ and spectral measure Γ. The regression of X_0 in terms of $X_1, \ldots, X_n$ is the conditional expectation $\mathbf{E}(X_0 \mid X_1, \ldots, X_n)$. It is well known that in the case where $X_0, X_1, \ldots, X_n$ are jointly Gaussian, $\mathbf{E}(X_0 \mid X_1, \ldots, X_n)$ is a linear function of $X_1, \ldots, X_n$, thus a Gaussian random variable itself, and is the minimum variance estimate of X_0. In the $S\alpha S$ case, the regression estimate $\mathbf{E}(X_0 \mid X_1, \ldots, X_n)$ is not linear in general, thus not even $S\alpha S$.

The following theorem states a necessary and sufficient condition for the regression estimate to be linear [56]:

Theorem 2 *If $X_0, X_1, \ldots, X_n$ are jointly $S\alpha S$ random variables with $1 < \alpha \leq 2$ and spectral measure Γ on the unit sphere S in R^{n+1}, then*

$$\mathbf{E}\{X_0 \mid X_1, \ldots, X_n\} = a_1 X_1 + \ldots + a_n X_n$$

if and only if for all $r_1, \ldots, r_n$.

$$\int_S (x_0 - a_1 x_1 - \ldots - a_n x_n)(r_1 x_1 + \ldots + r_n x_n)^{<\alpha-1>} \Gamma(d\mathbf{x}) = 0 \tag{17}$$

Remark: If the regression is linear, then the coefficients $a_1, \ldots, a_n$ are uniquely determined by Γ if and only if $X_1, \ldots, X_n$ are linearly independent elements in the space of integrable random variables. For each choice of $r_1, \ldots, r_n$ the condition of the theorem provides a linear equation involving the a_j's, but it is not known in general what choices of $r_1, \ldots, r_n$ will provide n linearly independent equations which can be solved for a_j's. The case $n = 2$, however, is easily solved by the following Corollary [56]:

Corollary 1 *If* X_0, X_1, X_2 *are jointly* $S\alpha S$ *and if*

$$\mathbf{E}(X_0 \mid X_1, X_2) = a_1 X_1 + a_2 X_2$$

then a_1, a_2 *satisfy*

$$\begin{aligned} a_1[X_1, X_1]_\alpha + a_2[X_2, X_1]_\alpha &= [X_0, X_1]_\alpha \\ a_1[X_1, X_2]_\alpha + a_2[X_2, X_2]_\alpha &= [X_0, X_2]_\alpha \end{aligned} \tag{18}$$

Moreover, the above equations uniquely determine a_1, a_2 *iff neither* X_1 *nor* X_2 *is a multiple of the other.*

There are few known cases where the regressions are indeed linear. The next theorem provides one example [56]:

Theorem 3 *If* $X_0, X_1, \ldots, X_n$ *are jointly* $S\alpha S$ *random variables and if* $X_1, \ldots, X_n$ *are independent and nondegenerate, then*

$$\mathbf{E}(X_0 \mid X_1, \ldots, X_n) = a_1 X_1 + \ldots + a_n X_n$$

and the coefficients a_k *are given by*

$$a_k = [X_0, X_k]_\alpha / [X_k, X_k]_\alpha$$

In particular:

Corollary 2 *For any two jointly* $S\alpha S$ *random variables* X, Y.

$$\mathbf{E}(X|Y) = \lambda_{XY} Y \tag{19}$$

where

$$\lambda_{X,Y} = \frac{[X,Y]_\alpha}{[Y,Y]_\alpha} \tag{20}$$

is the covariation coefficient *of X with Y.*

Regressions of sub-Gaussian random variables are also linear. In fact, something stronger is true. Let us define that a $S\alpha S$ random process $\{X(t), t \in T\}$ has the *linear regression property* if $\mathbf{E}(X_0|X_1,\ldots,X_n)$ is a linear function of $X_1,\ldots,X_n$ whenever the X_i's are elements of the linear span of $\{X(t), t \in T\}$. Then, the following remarkable result holds [29]:

Theorem 4 *A $S\alpha S$ process has the linear regression property if and only if it is sub-Gaussian.*

3 Parameter Estimates for Symmetric Stable Distributions

Recall that a symmetric stable distribution is determined by three parameters: the characteristic exponent α with $0 < \alpha \leq 2$, the dispersion γ with $\gamma > 0$ and the location parameter δ with $-\infty < \delta < \infty$. A practical problem is to estimate these three parameters from the realizations of a symmetric stable random variable. For convenience, we shall replace γ by a new parameter c, defined by

$$c = \gamma^{1/\alpha} \tag{21}$$

when we discuss parameter estimations.

For $\alpha > 1$, the sample mean will provide a consistent estimate for the location parameter δ. But the problem of estimating the parameters of a stable distribution is, in general, severely hampered by the lack of known closed-form density functions for all but a few members of the stable family. Most of the conventional methods in mathematical statistics can not be used in this case, since these methods are usually based on the availability of an explicit form for the density. However, there are some numerical methods that offer promises. Here we briefly survey a few estimators that are popular in the literature.

3.1 Method of Maximum Likelihood

Maximum likelihood estimates of α and γ (assuming $\delta = 0$) were obtained by DuMouchel [18]. A multinomial approximation to the likelihood function is used to find approximate maximum likelihood estimators of the parameters. The estimates have the usual desirable properties of maximum likelihood estimation. However, the computational effort involved seems considerable.

A direct method can be formulated as follows. Following Zolotarev [90, 10], the standard symmetric stable density function is:

$$f_\alpha(x) = \frac{\alpha}{|1-\alpha|\pi} x^{1/(\alpha-1)} \int_0^{\pi/2} v(\theta) e^{-x^{\alpha/(\alpha-1)} v(\theta)} d\theta \quad \text{for} \quad \alpha \neq 1, x > 0 \tag{22}$$

where

$$v(\theta) = \frac{1}{(\sin \alpha\theta)^{\alpha/(\alpha-1)}} \cos[(\alpha-1)\theta](\cos\theta)^{1/(\alpha-1)} \tag{23}$$

Furthermore

$$\begin{aligned} f_1(x) &= \frac{1}{\pi(1+x^2)} \\ f_0(0) &= \frac{1}{\pi}\Gamma((\alpha+1)/\alpha) \\ f_2(x) &= \frac{1}{2\sqrt{\pi}} e^{-x^2/4} \end{aligned} \tag{24}$$

Therefore, the parameters α, δ, c can be estimated from the observations $x_1, x_2, \ldots, x_N$ by maximizing the log likelihood function:

$$\begin{aligned} \sum_{i=1}^N \log[f_\alpha(z_i)] &= n\log\alpha - n\log(\alpha-1)\pi + \sum_{i=1}^N (\log z_i)/(\alpha-1) \\ &\quad + \sum_{i=1}^N \log \int_0^{\pi/2} v(\theta) e^{-z_i^{\alpha/(\alpha-1)}} v(\theta) d\theta \end{aligned} \tag{25}$$

where

$$z_i = |x_i - \delta|/c$$

To avoid the discontinuity and nondifferentiability at $\alpha = 1$, α is restricted to be greater than 1. Caution must be used when evaluating the integral in (22) and (25), since the integrand is singular at $\theta = 0$.

Based on (25), Brorsen and Yang in [10] performed Monte Carlo simulations with fairly good results. An obvious disadvantage of this method is that it is a highly nonlinear optimization problem and no initialization and convergence analysis are available.

3.2 Method of Sample Fractiles

The most frequently used method to estimate the parameters of symmetric stable laws with $1 \leq \alpha \leq 2$ was suggested by Fama and Roll in [22], based on order statistics. They suggested estimating c by

$$\hat{c} = \frac{1}{1.654}[\hat{x}_{0.72} - \hat{x}_{0.28}] \tag{26}$$

where $\hat{x}_f$ ($f = 0.72, 0.28$) is the estimated f fractile of the $S\alpha S$ distribution. A consistent estimate of the f fractile, $\hat{x}_f$, is usually the $f(N+1)$st order statistic where N is the size of observation[2]. It is shown that the estimate of c given by (26) has an asymptotic bias of less than 0.4% and is asymptotically normal with variance

$$\sigma^2(\hat{c}) \approx \frac{0.09c^2}{N[p(\alpha, 0.72)]} \tag{27}$$

where $p(\alpha, f)$ is the density of the distribution of X at the f fractile of the standard stable distribution of characteristic exponent α [22].

The characteristic exponent α, on the other hand, can be estimated from the tail behavior of the distribution. Specifically, for some large f ($f = 0.95$, for example), first calculate

$$\hat{z}_f = \frac{\hat{x}_f - \hat{x}_{1-f}}{2\hat{c}} = 0.827\frac{\hat{x}_f - \hat{x}_{1-f}}{\hat{x}_{0.72} - \hat{x}_{0.28}} \tag{28}$$

from the sample. Given that X is $S\alpha S$ with characteristic exponent α and dispersion $\gamma = c^2$, $\hat{z}_f$ is an estimator of the f fractile of the standard $S\alpha S$ distribution. Thus an estimate, $\hat{\alpha}_f$ can be obtained by searching a table of standard $S\alpha S$ distribution functions, such as those in [22, 30]. Monte Carlo simulations suggest that $\hat{\alpha}_{0.95}$ and $\hat{\alpha}_{0.97}$ are fairly robust. If the true value of α is close to 2 ($\alpha > 1.9$) then the best estimator is $\hat{\alpha}_{0.99}$, in terms of both low bias and standard deviation.

For $1 < \alpha \leq 2$, the $S\alpha S$ has finite mean. Thus the sample mean is a consistent estimate of the location parameter δ. A more robust estimate is the truncated mean. A p percent truncated sample mean is the arithmetic mean of the middle p percent of the ranked observations. It has been shown by Monte Carlo simulations that the truncated mean estimate is very efficient and asymptotically unbiased [21, 19]. It is also found that

[2] As pointed out in [51], to avoid spurious skewness in x_f, a correction must be made. Specifically, if the x_i's are arranged in increasing order, the correction must be performed by identifying x_i with $\hat{x}_{q(i)}$ where $q(i) = (2i-1)/2N$, and then interpolating linearly to f from the two adjacent $q(i)$ values.

the 50% truncated mean works best when the range of α is unknown.

Fama-Roll's method is simple but suffers from a small asymptotic bias and is not asymptotically efficient. Also, α is restricted to $1 \leq \alpha \leq 2$. McCulloch [51] generalized Fama-Roll's method to provide consistent estimates for α and c. He also eliminated the asymptotic bias in the Fama-Roll estimators of α and c. Specifically, it is found that for symmetric stable laws,

$$\hat{v}_\alpha = \frac{\hat{x}_{0.95} - \hat{x}_{0.05}}{\hat{x}_{0.75} - \hat{x}_{0.25}} \tag{29}$$

is independent of both c and δ. Thus a consistent estimate $\hat{\alpha}$ can be found by searching tables, such as those in [51] with matched value of v_α. For fixed α, the following quantity

$$v_c = \frac{\hat{x}_{0.975} - \hat{x}_{0.25}}{c} \tag{30}$$

as function of α, is independent of δ. Since $\hat{\alpha}, \hat{x}_{0.75}, \hat{x}_{0.25}$ are all consistent estimators, the following is a consistent estimator of c:

$$\hat{c} = \frac{\hat{x}_{0.975} - \hat{x}_{0.25}}{v_c(\hat{\alpha})} \tag{31}$$

McCulloch's method is actually more general than what is presented here. It provides consistent estimators for all four parameters, with $-1 \leq \beta \leq 1$ and $\alpha \geq 0.6$ while retaining the computational simplicity of Fama-Roll's method.

3.3 Method of Sample Characteristic Functions

The sample characteristic function is defined as

$$\hat{\varphi}(t) = \frac{1}{N} \sum_{j=1}^{N} \exp(itx_j) \tag{32}$$

where N is the sample size, and $x_1, \ldots, x_N$ are the observations. It is a consistent estimator of the true characteristic function that uniquely determines the density function. Note that the sample characteristic function, $\{\hat{\varphi}(t), -\infty < t < \infty\}$, is a stochastic process (non-stationary) with the useful property that $0 < |\hat{\varphi}(t)| \leq 1$. So all of the moments of $\hat{\varphi}(t)$ are finite.

A few estimation methods have been proposed based on the sample characteristic func-

tion. Among these methods are the method of moments of Press [65], the method of Paulson, Holcomb, and Leitch [61], and the regression-type method of Koutrouvelis [40, 39]. It has been shown through simulations that Koutrouvelis' regression-type method is better than the other two in terms of consistency, bias as well as efficiency [1]. In the following we will outline Koutrouvelis' regression-type method for the $S\alpha S$ distributions. For the estimation of β, see [40, 39].

The Koutrouvelis' regression method is based on the following relations between the characteristic function of a $S\alpha S$ distribution and its parameters:

$$\log(-\log|\varphi(t)|^2) = \log(2c^\alpha) + \alpha \log|t| \tag{33}$$

and

$$\frac{\mathrm{Re}\varphi(t)}{\mathrm{Im}\varphi(t)} = \tan \delta t \tag{34}$$

From (33), the parameters α and c can be estimated from the linear regression

$$y_k = \mu + \alpha w_k + \varepsilon_k, \quad k = 1, 2, \ldots, K \tag{35}$$

where

$$y_k = \log(-\log|\hat{\varphi}(t_k)|^2), \mu = \log(2c^\alpha), w_k = \log|t_k|.$$

ε_k denotes an error term which is assumed to be i.i.d with mean zero. $t_1, \ldots, t_K$ is an appropriate set of real numbers.

The location parameter δ can be estimated in a similar way, through the following linear regression:

$$z_k = \delta u_k + \varepsilon_k, \quad l = 1, 2, \ldots, L \tag{36}$$

where

$$z_k = \mathrm{Arctan}(\mathrm{Im}(\hat{\varphi}(u_k))/\mathrm{Re}(\hat{\varphi}(u_k)))$$

and $u_1, \ldots, u_L$ is an appropriate set of real numbers The error terms ε_k are again assumed to be i.i.d with mean zero.

The whole procedure may be performed iteratively until some prespecified convergence criterion is satisfied. The initial estimates may be provided by Fama and Roll's method or McCulloch's method. For the details about implementation of the regression estimators, see [40, 39].

The regression estimators $\hat{\alpha}, \hat{c}$ and $\hat{\delta}$ described above are consistent, asymptotically unbiased. According to the simulation results in [40], the regression-type method is better than the maximum likelihood method and Fama and Roll's fractile method. This method involves minimal computational effort and is easy to implement.

4 Estimation of Covariations

The covariation of two $S\alpha S$ random variables X and Y is, in general, difficult to calculate analytically. An important exception is when X, Y are both linear combinations of independent $S\alpha S$ random variables. Specifically, we have

Proposition 3 *Let U_i's be independent $S\alpha S$ random variables with dispersions γ_i, $i = 1, \ldots n$. For any numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$, form*

$$\begin{aligned} X &= a_1 U_1 + \cdots + a_n U_n, \\ Y &= b_1 U_1 + \cdots + b_n U_n \end{aligned}$$

Then

$$\begin{aligned} [X, X]_\alpha &= \gamma_1 |a_1|^\alpha + \cdots + \gamma_n |a_n|^\alpha, \\ [Y, Y]_\alpha &= \gamma_1 |b_1|^\alpha + \cdots + \gamma_n |b_n|^\alpha, \\ [X, Y]_\alpha &= \gamma_1 a_1 b_1^{<\alpha-1>} + \cdots + \gamma_n a_n b_n^{<\alpha-1>}, \\ \lambda_{XY} &= \frac{\gamma_1 a_1 b_1^{<\alpha-1>} + \cdots + \gamma_n a_n b_n^{<\alpha-1>}}{\gamma_1 |b_1|^\alpha + \cdots + \gamma_n |b_n|^\alpha} \end{aligned} \tag{37}$$

As we will see in the next section, covariations (covariation coefficients) among symmetric stable random variables play the roles, in certain estimation problems, of correlations (correlation coefficients) for second-order random variables. It is thus important to have good (unbiased, efficient) estimators for covariations.

In this section, we will discuss various methods for estimating the covariation coefficient λ_{XY} of two $S\alpha S$ random variables X, Y. The reason we focus our attention on estimating the covariation coefficient λ_{XY} instead of the covariation $[X, Y]_\alpha$ is twofold. Firstly, since $[Y, Y]_\alpha$ is the scale parameter in the characteristic function of Y, it can be estimated by the methods in the previous section. So if we know the covariation coefficient λ_{XY}, we can find $[X, Y]_\alpha$ by multiplying λ_{XY} and $\|Y\|_\alpha^\alpha$. Secondly, as we will see soon, most of the time we need only to know the covariation coefficients. The knowledge of covariation itself is unnecessary.

4.1 Fracional Lower Order Moment Estimator

An important property of the covariation coefficient is stated in the following theorem [48, 12]:

Theorem 5 *Suppose X, Y are jointly $S\alpha S$ random variables with $1 < \alpha \leq 2$. Then their covariation coefficient is given by*

$$\lambda_{XY} = \frac{\mathbf{E}(XY^{<p-1>})}{\mathbf{E}(|Y|^p)} \tag{38}$$

for any $0 < p < \alpha$.

This theorem immediately suggests a consistent method for estimating the covariation coefficient λ_{XY}, which will be called fractional lower order moment (FLOM) estimator. Specifically, for the independent observations $(X_1, Y_1), \ldots, (X_n, Y_n)$, we let

$$\hat{\lambda}_{FLOM} = \frac{\sum_{i=1}^{N} X_i |Y_i|^{p-1} \text{sign}(Y_i)}{\sum_{i=1}^{N} |Y_i|^p} \tag{39}$$

for some $0 < p < \alpha$. A computationally efficient choice is $p = 1$. In this case

$$\hat{\lambda}_{FLOM} = \frac{\sum_{i=1}^{N} X_i \text{sign}(Y_i)}{\sum_{i=1}^{N} |Y_i|} \tag{40}$$

4.2 Screened Ratio Estimator

Kanter and Steiger [37] proposed an unbiased and consistent estimator for covariation coefficients, based upon the following theorem:

Theorem 6 *Let X and Y be random variables that satisfy $\mathbf{E}(|X|) < \infty$ and*

$$\mathbf{E}(X \mid Y) = \lambda Y \quad a.s. \tag{41}$$

for some λ. *For* $0 < c_1 < c_2 \leq \infty$ *write* B *for the event* $\{|Y| \in (c_1, c_2)\}$ *and*

$$I_B = \begin{cases} 1 & \text{if } |Y| \in (c_1, c_2) \\ 0 & \text{otherwise} \end{cases}$$

is the indicator of the event B. *Then*

$$\mathbf{E}(XY^{-1}I_B)/P(B) = \lambda$$

i.e,

$$\hat{\lambda} = (XY^{-1}I_B)/P(B) \tag{42}$$

is an unbiased estimate of λ.

As with any unbiased estimate, the strong law of large numbers implies that the above estimate is strongly consistent. Specifically,

$$\hat{\lambda}_{SCR} = \sum_{i=1}^{N}(X_iY_i^{-1}I_{B_i})/\sum_{i=1}^{N} I(B_i) \tag{43}$$

converges to λ almost surely as $n \to \infty$, where X_i, Y_i are independent copies of X and Y, respectively, and $B_i = \{|Y_i| \in (c_1, c_2)\}$. $\hat{\lambda}$ in (43) is called the *screen ratio estimate* (SCR) of λ. Constants c_1, c_2 are arbitrary. A common choice is to let $c_2 = \infty$.

4.3 Least-Squares Estimators

Another method, which works well for estimating correlation coefficients of Gaussian random variables, is the least-squares (LS) method. Specifically, we may estimate λ by minimizing the error

$$\sum_{i=1}^{N}(X_i - \lambda Y_i)^2$$

The solution is of course the usual least-squares solution:

$$\hat{\lambda}_{LS} = \sum_{i=1}^{N} X_iY_i/\sum_{i=1}^{N} Y_i^2 \tag{44}$$

As pointed out in [37], the least square estimate $\hat{\lambda}_{LS}$ is not consistent in the infinite

variance case.

4.4 Sampling Results and Some Comparisons

In this section we compares the performances of the three estimates by Monte-Carlo simulations. Two $S\alpha S$ random variables, X and Y, are generated by

$$X = a_1U_1 + a_2U_2,$$
$$Y = b_1U_1 + b_2U_2$$

where U_1, U_2 are independent, symmetrically distributed stable random variables with characteristic function $\exp(-|t|^\alpha)$. So the true covariation coefficient λ of X with Y is

$$\lambda = \frac{a_1b_1^{\alpha-1} + a_2b_2^{\alpha-1}}{|b_1|^\alpha + |b_2|^\alpha}$$

5000 independent copies of U_1, U_2 are generated and the three estimators are computed as follows:

FLOM estimator:

$$\hat{\lambda}_p(N) = \frac{\sum_{i=1}^{N} X_i \text{sign}(Y_i)}{\sum_{i=1}^{N} |Y_i|} \tag{45}$$

Screened ratio estimator:

$$\hat{\lambda}_{SCR}(N) = \sum_{i=1}^{N} (X_iY_i^{-1}I_{B_i}) / \sum_{i=1}^{N} I(B_i) \tag{46}$$

where we chose $c_1 = 0.1, c_2 = \infty$.

Least-squares estimator:

$$\hat{\lambda}_{LS}(N) = \sum_{i=1}^{N} X_iY_i / \sum_{i=1}^{N} Y_i^2 \tag{47}$$

Finally this process is repeated independently 50 times, and the means and standard deviations of the three estimators are computed. The results appear in Table 1 with standard deviations in parentheses.

Model (a_1, a_2, b_1, b_2)	α	LS	Screened Ratios	FLOM $(p = 1)$	True λ
(-0.75, 0.25, 0.18, 0.78)	1.1	0.3340 (5.0539)	-0.4327 (0.3919)	-0.4707 (3.1266)	-0.4252
	1.3	0.3752 (5.1640)	-0.2591 (0.4991)	-0.2599 (1.0352)	-0.2602
	1.5	0.4059 (4.5142)	-0.1112 (0.3337)	-0.1222 (0.9202)	-0.1273 (0.9202)
	1.9	0.1069 (0.8870)	0.0654 (0.3175)	0.06104 (0.4598)	0.0599
	2.0	0.0976 (0.0102)	0.08544 (0.2790)	0.0954 (0.1252)	0.0936

Table 1: Sampling properties of covariation estimates

Several features stand out in the simulation results in Table 1. The least-squares estimate works only in the Gaussian case. It completely fails for $\alpha < 2$. The screened ratio estimate performs well for $\alpha < 2$. The smaller α is, the better its performance. These are consistent with the theoretic result that the screened ration estimates are unbiased and consistent. On the other hand, the FLOM method with $p = 1$ is very robust against changes of α, although it becomes more volatile when α approaches 1. Thus the FLOM method seems to be a very promising method for estimating covariations, especially when α is close to 2. In addition it is very simple and inexpensive to compute.

In closing, let us point out the following special case of Theorem 5:

Corollary 3 *If X, Y are jointly Gaussian with zero means, then the correlation coefficient of X, Y can be written as*

$$\rho_{XY} = \frac{\mathbf{E}(XY)}{\sqrt{\mathbf{E}(X^2)}\sqrt{\mathbf{E}(Y^2)}} = \frac{\mathbf{E}(X\,sign(Y))}{\mathbf{E}(|Y|)} \tag{48}$$

The significance of the above result is that it provides a computationally efficient way of estimating the correlation of any Gaussian random variables. This is very useful, for example, in conventional power spectral estimation or the Yule-Walker method of AR spectrum estimation.

5 Parametric Models of Stable Random Processes

Unlike the Gaussian random processes which are completely determined by the autocorrelation sequences, stable processes are difficult to characterize. In this section, we consider a class of linear $S\alpha S$ processes which are generated by systems with rational transfer functions driven by i.i.d $S\alpha S$ inputs. This class of models includes AR, MA, and $ARMA$ models. The output process can be completely described in terms of model parameters and the dispersion of the input. It will be shown how to estimate the model parameters and input dispersion from the covariation sequence of the output. The development here is quite similar to that of second-order processes. The covariation sequence plays the role of covariance sequence in the second-order case.

5.1 Parameter Estimation of AR Stable Processes

Let us consider processes generated by the following AR system

$$X(n) = a_1 X(n-1) + \cdots + a_p X(n-p) + U(n) \tag{49}$$

where $\{U(n)\}$ is a sequence of i.i.d $S\alpha S$ random variables of characteristic exponent α and dispersion γ_u, i.e., the random variables $U(n)$ have the same characteristic function $\exp(-\gamma_u |t|^\alpha)$. As in the second-order case, we are only interested in stationary solutions of (49) which is unique if the AR system in (49) is stable, i.e., all the poles are inside the unit circle. In addition, the stationary process $X(n)$ is $S\alpha S$ and $X(n)$ and $U(n+j)$ are independent for any $j > 0$.

We are interested in the problem of identifying the AR coefficients $a_1, \ldots, a_p$ from the observation of output $X(n)$. There are several of methods that have been proposed in the literature. In the following we review some of these approaches and compare the performances through Monte Carlo simulations.

5.1.1 Generalized Yule-Walker Equation

Taking the conditional expectation of both sides of (49), one has that for $n-p \le m \le n-1$:

$$\mathbf{E}(X(n)|X(m)) = a_1 \mathbf{E}(X(n-1)|X(m)) + \cdots + a_p \mathbf{E}(X(n-p)|X(m)) \tag{50}$$

where we used the fact that $\mathbf{E}(U(n)|X(m)) = 0$ if $U(n)$ and $X(m)$ are independent. Since $\{X(n)\}$ is stable and stationary, we have

$$\mathbf{E}(X(n+l)|X(n)) = \lambda(l)X(n)$$

where $\lambda(l)$ is the covariation coefficient of $X(n+l)$ with $X(n)$ with $\lambda(0) = 1$. Let us also define

$$\mathbf{p} = \begin{bmatrix} \lambda(1) \\ \lambda(2) \\ \vdots \\ \lambda(p) \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

and

$$\mathbf{C} = \begin{bmatrix} \lambda(0) & \lambda(-1) & \cdots & \lambda(1-p) \\ \lambda(1) & \lambda(0) & \cdots & \lambda(2-p) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda(p-1) & \lambda(p-2) & \cdots & \lambda(0) \end{bmatrix} \tag{51}$$

Then the coefficients of the AR system can be found by solving the following system of linear equations:

$$\mathbf{Ca} = \mathbf{p} \tag{52}$$

Equation (52) is a direct generalization of the Yule-Walker equation for the case $\alpha = 2$. The covariation matrix C is Toeplitz. Thus, if it is nonsingular, we can solve (52) fairly efficiently. On the other hand, unlike the Gaussian case where C is symmetric and positive definite, the covariation matrix C is not symmetric. It may even be singular.

In practice, we can only estimate the covariation matrix $\mathbf{C}$ and vector $\mathbf{p}$. Thus the estimates of the coefficients $\mathbf{a}$ in (52) are based on those for $\lambda(l)$. For example, we can use the screened ratio estimates or the FLOM estimates defined in the previous section. Then forming the covariation matrix $\hat{\mathbf{C}} = (\hat{C}(i,j)) = (\hat{\lambda}(i-j))$ and the vector $\hat{\mathbf{p}} = (\hat{\lambda}(i))$, we estimate $\mathbf{a}$ by the relation (assuming $\hat{\mathbf{C}}$ is invertible)

$$\hat{\mathbf{a}} = \hat{\mathbf{C}}^{-1}\hat{\mathbf{p}} \tag{53}$$

Since the screened ratio and FLOM estimates are consistent, we know that $\hat{\mathbf{a}}$ given by (53) is also consistent.

5.1.2 Least-Squares Method

Another approach to estimate the AR coefficients is treating the AR model as if it were driven by a second-order process and use least-squares estimates. The least-squares estimates of $a_1, \ldots, a_p$ are found by minimizing

$$\sum_{n=N1}^{N2} (X(n) - a_1 X(n-1) - \cdots - a_p X(n-p))^2$$

This is in fact equivalent to using the least-squares estimates of the covariation coefficient $\lambda(l)$

$$\hat{\lambda}_{LS}(l) = \sum_{n=N_1}^{N_2} X(n-l)X(n) / \sum_{n=N_1}^{N_2} X^2(n) \tag{54}$$

Forming the covariation matrix $\hat{\mathbf{C}}_{LS} = (\hat{C}(i,j)) = (\hat{\lambda}_{LS}(i-j))$ and the vector $\hat{\mathbf{p}}_{LS} = (\hat{\lambda}_{LS}(i))$, the least square estimates of the AR coefficients are given by:

$$\hat{\mathbf{a}}_{LS} = \hat{\mathbf{C}}_{LS}^{-1} \hat{\mathbf{p}}_{LS}$$

There is a considerable amount of literature dealing with the properties of the least squares estimates for AR processes with infinite variances [37, 89, 28]. We summarize the results in the following theorem.

Theorem 7 *Let N be the number of observations used in computing $\hat{\lambda}_{LS}(l)$ as shown in (54). Then for any $\delta > \alpha$, $\lim_{N\to\infty} N^{1/\delta}(\hat{a}_j - a_j) = 0$ a.s. for $j = 1, 2, \cdots, p$.*

Thus, least square estimate is consistent and convergence of the least-squares estimates to the true coefficients is very rapid, on the order of $O(N^{-\frac{1}{\delta}})$ for any $\delta < \alpha$. The smaller α is, the faster the convergence.

5.1.3 Least Absolute Deviation Estimates

It is well known that the least-squares method is closely related to the minimum mean square error estimation method. It is surprising that the least-squares method has such nice properties for estimating AR processes with infinite variance. After all, in this case, the minimum mean square error estimation does not even apply.

For $S\alpha S$ random variables, a suitable measure of dispersion is the norm defined by (10). Since for $p < \alpha$, the L_p norm $(\mathbf{E}(|X|^p))^{1/p}$ exists for any $S\alpha S$ random variable X

and is equivalent to $\|X\|_\alpha$, we may estimate the coefficients by minimizing

$$\mathbf{E}|X(n) - a_1X(n-1) - \cdots - a_pX(n-p)|^p$$

For finite observations, we find $a_1, \cdots, a_p$ by minimizing the function

$$\sum_{n=N1}^{N2} |X(n) - a_1X(n-1) - \cdots - a_pX(n-p)|^p$$

Since L_p error norm places less weight on extreme observations than does least squares, it seems a natural estimator to apply in cases where extreme observations occur more frequently than where the disturbances are normally distributed.

Several researchers have looked at a particular case where $p = 1$ [6, 27, 2]. The estimates $\hat{a}_1, \ldots, \hat{a}_p$ obtained by minimizing

$$\sum_{n=N1}^{N2} |X(n) - a_1X(n-1) - \cdots - a_pX(n-p)| \tag{55}$$

are called the *least absolute deviation* (LAD) estimates. In general, the LAD estimates are unique. The strong consistency of LAD estimates in the case of infinite variances is given by the following theorem [2]:

Theorem 8 *Let N be the number of observations used in computing (55). Then for any $\delta > \alpha$, the LAD estimates $\hat{a}_i$ satisfy $\lim_{N\to\infty} N^{1/\delta}(\hat{a}_j - a_j) = 0$ in probability for $j = 1, 2, \cdots, p$.*

That is, the convergence rate of LAD estimates is comparable with that of LS estimates, although simulations show that *LAD* estimates actually converge faster than the LS estimates. But LAD estimates are several times costly than LS estimates.

5.1.4 Sampling Results and Performance Comparisons

To study the large sample properties of the proposed estimators, we generated 5000 samples of the stationary $S\alpha S$ output of a second-order AR defined by

$$X(n) - a_1X(n-1) - a_2X(n-2) = U(n)$$

Coefficients	α	Least squares	LAD	Yule-Walker
a_1=0.195	1.1	0.1953 (0.0277)	0.1950 (0.0021)	0.1942 (0.1178)
	1.3	0.1946 (0.0299)	0.1949 (0.0039)	0.1949 (0.0895)
	1.5	0.1944 (0.0022)	0.1945 (0.0110)	0.1929 (0.0918)
	1.9	0.1951 (0.0031)	0.1946 (0.0362)	0.1949 (0.0563)
	2.0	0.1950 (0.0347)	0.1953 (0.0437)	0.1945 (0.0392)
a_2=-0.95	1.1	-0.9486 (0.0014)	-0.9500 (0.0017)	-0.9454 (0.0049)
	1.3	-0.9500 (0.0271)	-0.95 (0.0050)	-0.9519 (0.002)
	1.5	-0.9493 (0.0008)	-0.9498 (0.0093)	-0.9513 (0.0913)
	1.9	-0.9500 (0.0001)	-0.95 (0.0001)	-0.9513 (0.0014
	2.0	-0.95 (0.00001)	-0.9503 (0.0003)	-0.9498 (0.0002)

Table 2: Estimation of coefficients of second-order AR process

where

$$a_1 = 0.195, \quad a_2 = -0.95$$

The estimates $\hat{a}_1, \hat{a}_2$ were computed based on the generalized Yule-Walk equation, the LS method and the LAD method. Since simulations in [37] shows that the AR estimates based on the screened ratio estimates of covariation coefficients may be very inefficient despite their consistency on theoretic ground, the FLOM method was used for the covariation estimates in the generalized Yule-Walker equation.

This procedure was repeated independently, 50 times, and the means and standard deviations of the three estimates were computed. These results are shown in Table 2, standard deviations in parentheses.

It is obvious that the LS and LAD methods are very efficient and reliable. The general-

ized Yule-Walker method based on FLOM estimates with $p = 1$ performs as we as the least squares despite its simplicity. This is not surprising, knowing its superior performance in estimating covariations.

5.2 Parameter Estimations of $ARMA$ Stable Processes

Consider processes generated by the following $ARMA$ system

$$X(n) = a_1 X(n-1) + \cdots + a_p X(n-p) + b_0 U(n) + \cdots + b_q U(n-q) \tag{56}$$

where $\{U(n)\}$ is a sequence of i.i.d $S\alpha S$ random variables of characteristic exponent α and dispersion γ_u. Under the assumption that the system is stable, the output process $\{X(n)\}$ is a stationary $S\alpha S$ random sequence of characteristic exponent α. In addition, $X(n)$ and $U(n+j)$ are independent for all $j > 0$. The problem is to identify the AR coefficients $a_1, \ldots, a_p$ and the MA coefficients $b_1, \ldots, b_q$ from the observations of output

Define

$$\lambda_{xx}(m) = \lambda_{X(n)X(n-m)} \tag{57}$$

as the covariation coefficient of $X(n)$ with $X(n-m)$ and

$$\lambda_{ux}(m) = \lambda_{U(n)X(n-m)} \tag{58}$$

as the covariation coefficient of $U(n)$ with $X(n-m)$. Because of the stationarity of the random processes, these covariations do not depend on time n.

By taking the covariations of both sides of (56) with $X(n-m)$ one has:

$$\lambda_{xx}(m) = \sum_{k=1}^{p} a_k \lambda_{xx}(m-k) + \sum_{k=0}^{q} b_k \lambda_{ux}(m-k) \tag{59}$$

The covariation $\lambda_{ux}(m)$ of the input and the output can be expressed in terms of the impulse response $h(n)$ of the system by using the linearity of covariations:

$$\lambda_{ux}(m) = \begin{cases} 0 & \text{for } m > 0 \\ \gamma_u & \text{for } m = 0 \\ \gamma_u [h(-m)]^{<\alpha-1>} & \text{for } m < 0 \end{cases} \tag{60}$$

Combining (59) and (60), we obtain the relationship between the $ARMA$ parameters and the covariation of the process $X(n)$:

$$\lambda_{xx}(m) = \begin{cases} \sum_{k=1}^{p} a_k \lambda_{xx}(m-k) + \gamma_u \sum_{k=m}^{q} b_k [h(k-m)]^{<\alpha-1>} & \text{for } 0 \le m \le q \\ \sum_{k=1}^{p} a_k \lambda_{xx}(m-k) & \text{for } m > q \end{cases} \tag{61}$$

Restricting the lag index m to $q+1 \le m \le q+p$, one obtains a set of linear equations for the autoregressive parameters $a_1, \ldots, a_p$ of an $ARMA$ model in terms of the covariation sequence:

$$\begin{bmatrix} \lambda_{xx}(q) & \lambda_{xx}(q-1) & \cdots & \lambda_{xx}(q-p+1) \\ \lambda_{xx}(q+1) & \lambda_{xx}(q) & \cdots & \lambda_{xx}(q-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{xx}(q+p-1) & \lambda_{xx}(q+p-2) & \cdots & \lambda_{xx}(q) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \lambda_{xx}(q+1) \\ \lambda_{xx}(q+2) \\ \vdots \\ \lambda_{xx}(q+p) \end{bmatrix} \tag{62}$$

Thus, the the autoregressive parameters may be found separately from the moving average parameters as the solution to the simultaneous equation (62). This is analogous to the $ARMA$ Yule-Walker normal equations. Note that the coefficient matrix in (62) is again Toeplitz. Fast algorithms exist for solving (62).

Unfortunately, the moving average parameters of an $ARMA$ model can not be found simply as the solution of a set of linear equations. The MA parameters are convolved with the impulse response coefficients $h(k)$, as indicated in (62), resulting a nonlinear relationship with the covariation sequence. In the simplest case where $p = 0$ (i.e., for MA models), it is easy to show that the relationship between the model parameters and output covariation sequence is given by

$$\lambda_{xx}(m) = \begin{cases} \gamma_u (\sum_{k=0}^{p-m} b_{k+m} b_k^{<\alpha-1>}) & \text{for } 0 \le m \le q \\ \gamma_u (\sum_{k=0}^{p-m} b_k b_{k-m}^{<\alpha-1>}) & \text{for } -q \le m \le 0 \\ 0 & \text{for } |m| > q \end{cases} \tag{63}$$

This is a set of highly nonlinear equations.

6 Linear Estimations of Stable Processes

6.1 The Minimum Dispersion Criterion

One of the central problems in statistical signal processing can be stated as follows: given a set of observations $\{X(t), t \in T\}$, find the "best" estimate of an unknown random variable Y from the linear space spanned by $\{X(t), t \in T\}$. This is the so-called *linear theory* of stochastic processes, which includes linear estimation, prediction and filtering.

The linear theory of second-order processes (Gaussian processes in particular) has been fully developed. In this case, the linear space of observations $X(t)$, $L(X(t), t \in T)$, is a Hilbert space. Under the *minimum mean square error* (MMSE) criterion, the best linear estimate of the unknown Y is the orthogonal projection of Y onto $L(X(t), t \in T)$. On the other hand, the linear theory of non-Gaussian stable processes has only recently been the subject of intensive research and relatively fewer results are available. A major difficulty is that the linear span of a stable process is a Banach space when $1 \leq \alpha < 2$ and only a metric space when $0 < \alpha < 1$. These spaces do not have as nice properties and structures as Hilbert spaces for the linear estimation problem.

Let us first look at what we mean by a "best" estimate. For second-order processes, the most commonly used criterion for the best estimate is the MMSE criterion. Under this criterion, the best estimate is the one that minimizes the variance of estimation error. If the process is Gaussian it can be shown that this criterion also minimizes the probabilities of large estimation errors. For stable processes, the MMSE criterion is no longer appropriate due to the lack of finite variance. But the concept of MMSE criterion can be easily generalized to stable processes. Specifically, the *minimum dispersion* (MD) criterion is used in discussing linear theory of stable processes. Under the MD criterion, the best estimate of a $S\alpha S$ random variable in the linear space of observations is the one that minimizes the dispersion of estimation error. Recall that the dispersion (i.e., the scale parameter) of a stable random variable plays an analogous role of the variance. For example, the larger the dispersion, the more spread the stable random variable around the median. Thus, by minimizing the dispersion, we minimize the average magnitude of estimation error. Furthermore, it can be shown that minimizing the dispersion is also equivalent to minimizing the probability of large estimation errors [16]. The MD criterion is thus well justified in stable case. It is a direct generalization of the MMSE criterion (MD and MMSE criteria are the same in the Gaussian case) and reasonably simple to calculate. This

criterion was introduced in [79] in an attempt to solve Kalman filtering problems associated with stable processes. It has also been used in [6] in regression models with stable errors as well as in [16] for linear prediction of $ARMA$ processes with infinite variances.

From Equation (12), we see that the MD criterion is also equivalent to minimizing the FLOM's of estimation errors. These FLOM's measure the L_p distance between Y and its estimate $\hat{Y}$ on the linear space generated by the observation $\{X(t), t \in T\}$, for $p < \alpha$. This result is not surprising since the L_p norms for $p < 2$ are well known for being robust against outliers such as those that may be described by stable laws. Notice that in stable cases all of the FLOM's are equivalent. A common choice is the L_1 moment which is sometimes very convenient.

Under the MD criterion, the linear estimation problem of stable processes can be formulated as follows: Find an element $\hat{Y}$ in the linear space $L(X(t), t \in T)$ of the observations such that

$$\|Y - \hat{Y}\|_\alpha = \inf_{Z \in L(X(t), t \in T)} \|Y - Z\| \tag{64}$$

or equivalently

$$\mathbf{E}|Y - \hat{Y}|^p = \inf_{Z \in L(X(t), t \in T)} \mathbf{E}|Y - Z|^p \tag{65}$$

for $0 < p < \alpha$. Since $L(X(t), t \in T)$ is a Banach space, $\hat{Y}$ always exists and is unique for $1 < \alpha < 2$ [77]. It is obtained by a metric projection of Y onto the convex Banach space $L(X(t), t \in T)$. For $1 < \alpha < 2$, $\hat{Y}$ is also uniquely determined by [13]:

$$[X(t), Y - \hat{Y}]_\alpha = 0 \quad \text{for all } t \in T \tag{66}$$

This is analogous to the *orthogonality principle* used extensively in linear estimations of second-order processes [59].

When $\alpha = 2$, Equation (66) is linear and thus a closed-form solution exists for $\hat{Y}$. For $\alpha < 2$, Equation (66) is nonlinear and is hard to solve for the estimate $\hat{Y}$. For example, let X, Y be jointly $S\alpha S$ with spectral measure Γ. Let aY be the optimal approximation of X in the MD sense.Then the coefficient a is the solution of the following equation

$$\int_S y(x - ay)^{<\alpha-1>} \Gamma(ds) = 0 \tag{67}$$

Even if we know Γ, the above equation is still very hard to solve.

A related estimation problem is the regression. The regression estimate of Y given $X_1, \ldots, X_n$ is the conditional expectation $\mathbf{E}(Y|X_1, \ldots, X_n)$. When $Y, X_1, \ldots, X_n$ are jointly Gaussian, the regression estimate is linear and is equal to the linear estimate determined by Equation (66). When $Y, X_1, \ldots, X_n$ are jointly stable with $\alpha < 2$, this is no longer true, as we have seen in Theorem 2. Even if the regression estimate is linear it needs not to be the same as the linear estimate. A simple example to illustrate the point is to consider the regression of two $S\alpha S$ random variables X and Y. In this case, the regression is given by $\mathbf{E}(Y|X) = aX$, where $a = [Y, X]_\alpha / [Y, Y]_\alpha$. On the other hand, the linear estimate $\hat{Y} = aX$ is determined by (67), which has no closed-form solutions.

Although the development of linear theory for general stable processes is still in the primitive stage, there are explicit results for special types of stable processes, such as harmonizable, linear and sub-Gaussian processes. A large proportion of these results are devoted to harmonizable stable processes [63, 87, 12, 33, 49]. They are usually presented in the spectral domain, rather involved and can not be easily implemented from signal processing point of view. We choose instead to present some results on r predictions of linear processes. For a detailed analysis of continuous-time linear estimation problems, see [13].

6.2 A Suboptimal State-Space Prediction

In [79], Stuck made an attempt to extend the classic Kalman filtering theory to a more general situation where the plant and observation noises are stable. He considered only the scalar discrete-time case. Specifically, let the model of the signal process $X(n)$ and observation process $Y(n)$ be given by

$$\begin{cases} X(n+1) &= aX(n) + bU(n) \\ Y(n) &= cX(n) + W(n) \end{cases} \tag{68}$$

where the plant noise $U(n)$ and observation noise $W(n)$ are two i.i.d $S\alpha S$ sequences with dispersions γ_u and γ_w respectively. In addition, we assume that $U(n), W(m)$ are independent for all n and m. The initial value $X(0)$ is also assumed to be stable with dispersion γ_{x_0}. To predict $X(n+1)$ from the observations $Y(n), Y(n-1), \ldots, Y(1)$, Stuck suggested

using the recursive formulation of Kalman filters:

$$\begin{aligned} \hat{X}(n+1) &= a\hat{X}(n) + G_n e(n) \\ e(n) &= Y(n) - c\hat{X}(n) \end{aligned} \tag{69}$$

where the gain G_n is chosen to minimize the dispersion $\gamma_{e(n)}$ of prediction error $e(n)$. Note that $e(0) = X(0)$, thus $\gamma_e(0) = \gamma_{x_0}$. It can be shown [79] that for $1 < \alpha \leq 2$

$$\begin{cases} G_n &= \dfrac{a}{c} \dfrac{[|c|^\alpha \gamma_e(n)]^{1/(\alpha-1)}}{[|c|^\alpha \gamma_e(n)]^{1/(\alpha-1)} + \gamma_w^{1/(\alpha-1)}} \\ \gamma_e(n+1) &= |b|^\alpha \gamma_u + |\dfrac{a}{c}|^\alpha \dfrac{\gamma_w |C|^\alpha \gamma_e(n)}{\{[|c|^\alpha \gamma_e(n)]^{1/(\alpha-1)} + \gamma_w^{1/(\alpha-1)}\}^{\alpha-1}} \end{cases} \tag{70}$$

Although (69) is easy to implement, it is only suboptimal. It does not truly use the MD criterion in the prediction, as can be seen from the fixed recursive formulation. In general, the MD prediction of $X(n)$ no longer has the recursive formulation of the Kalman filters for the Gaussian case.

6.3 MD Predictions of Linear Stable Processes

Consider the $ARMA$ stable process $X(n)$ determined by the following equation

$$X(n) - a_1 X(n-1) - \cdots - a_p X(n-p) = U(n) + b_1 U(n-1) + \cdots + b_q U(n-q) \tag{71}$$

where $\{U(n), n = 0, \pm 1, \pm 2, \ldots\}$ are i.i.d $S\alpha S$ and the system is assumed to have minimum phase. Let

$$\frac{1 - a_1 z^{-1} - \cdots - a_p z^{-p}}{1 + b_1 z^{-1} + \cdots + b_q z^{-q}} = 1 - \sum_{k=1}^{\infty} h_k z^{-k} \quad \text{for } |z| = 1 \tag{72}$$

be the inverse of the system. Then the solution of k-step prediction based on the infinite past is given by the following theorem [16]:

Theorem 9 *For the $ARMA(p,q)$ process there exists a unique minimum dispersion linear predictors $\hat{X}(n+k)$ for $X(n+k)$, $k \geq 1$, based on the infinite past $X(n), X(n-1), \ldots$, given by the following recursive relationship:*

$$\hat{X}(n+k) = \sum_{j=1}^{k-1} h_j \hat{X}(n+k-j) + \sum_{j=k}^{\infty} h_j X(n+k-j) \tag{73}$$

In practice, one would like to predict $X(n+k)$ based on finite observations $X(n),\ldots,X(1)$. In this case, the truncation of the solution in (73) gives a nearly optimal solution for large n.

The exact MD k-step predictor of $ARMA(p,q)$ process based on $X(n),\ldots,X(1)$ is analytically involved, except for the AR process. In this case, the receipt for the predictor is exactly the same as the one for Gaussian processes. Specifically, one has [16]

Theorem 10 *For the AR process given by the following stable system*

$$X(n) = a_1X(n-1) + \cdots + a_pX(n-p)$$

there exists a unique minimum dispersion k-step predictor $\hat{X}(n+k)$ *for* $X(n+k)$ *(*$k \geq 1, n \geq p$*) in terms of* $X(1),\ldots,X(n)$*. It satisfies the following recursive relationship*

$$\hat{X}(n+k) = a_1\hat{X}(n-1) + \cdots + a_p\hat{X}(n-p) \tag{74}$$

with initial conditions $\hat{X}(j) = X(j)$ *for* $1 \leq j \leq n$.

Thus, the MD predictor $\hat{X}(n+k)$ is exactly the same as the MMSE predictor for an AR Gaussian process. This is not the case for general $ARMA$ stable processes [16].

6.4 Adaptive Wiener Filters for Stable Processes

Adaptive solutions of the linear estimation problems for stable processes are much easier to implement because they do not require closed-form solutions. The dispersion of the estimation error is usually a convex function of the parameters. So numerical methods, such as stochastic gradient methods, may be used to find the parameters by minimizing the dispersion of the error function.

Let us consider designing an FIR filter with an input consisting of a stationary $S\alpha S$ process $u(0), u(1), u(2),\ldots$. The problem is to choose the tap weight $w_0, w_1, \ldots, w_{M-1}$ such that the output of the filter is as close to the desired response $d(n)$ as possible. Here we assume $d(n)$ and $u(n)$are jointly $S\alpha S$. Specifically, we would like to find $w_0, w_1, \ldots, w_{M-1}$ such that the dispersion of the error

$$e(n) = d(n) - \sum_{k=0}^{M-1} w_k u(n-k) \tag{75}$$

is minimized. The cost function is thus given by

$$J = \|d(n) - \sum_{k=0}^{M-1} w_k u(n-k)\|_\alpha \tag{76}$$

This cost function turns out to be quite intractable in general. We will use an equivalent form. Recall the norm of a $S\alpha S$ is proportional to the usual L_p norm of random variables for any $0 < p < \alpha$. So an equivalent cost function is given by

$$J = \mathbf{E}(|d(n) - \sum_{k=0}^{M-1} w_k u(n-k)|^p) \tag{77}$$

where $0 \leq p < \alpha$. A particularly simple case is when $p = 1$. In this case, the cost function is just

$$J = \mathbf{E}(|d(n) - \sum_{k=0}^{M-1} w_k u(n-k)|) \tag{78}$$

Although there is no closed-form solution for the set of coefficients minimizing the cost function J, J is convex. So we may use stochastic gradient method to solve for the coefficients in the same way as the LMS does. Hence, we propose the following LMP (Least Mean P-norm) algorithm.

LMP Algorithm: Fix $1 \leq p < \alpha$

1. Filter output:

$$y(n) = \hat{\mathbf{w}}^T(n)\mathbf{u}(n) \tag{79}$$

2. Estimation error:

$$e(n) = d(n) - y(n) \tag{80}$$

3. Tap weight adaptation:

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mu \mathbf{u}(n)|e(n)|^{p-1}\text{sign}(e(n)) \tag{81}$$

For $p = 1$ the above algorithm will be called LMAD (Least Mean Absolute Deviation) algorithm. The LMAD is actually the familiar signed LMS algorithm, although it is derived in a different context.

To compare the performances of the LMS and LMAD algorithms, we set up the fol-

lowing experiments involving a first-order AR process. Consider an AR process $\{u(n)\}$ of order 1, described by the difference equation

$$u(n) = au(n-1) + v(n) \tag{82}$$

where $a = 0.99$ is the parameter of the process, and $\{v(n)\}$ is a $S\alpha S$ white-noise process of dispersion 1. We let the AR process in (82) reach steady-state before processing the data to ensure the stationarity. To estimate the parameter a, we implement adaptive predictors of order 1 using LMAD and LMS. The LMAD algorithm for the weight adaptation can be written as

$$\begin{aligned}
&\hat{w}_{LMAD}(n+1) = \hat{w}_{LMAD}(n) + \mu u(n-1)\text{sign}(e(n)) \\
&e(n) = u(n) - \hat{w}_{LMAD}(n)u(n-1) \\
&w_{LMAD}(0) = 0.0
\end{aligned}$$

On the other hand, the LMS can be written as

$$\begin{aligned}
&\hat{w}_{LMS}(n+1) = \hat{w}_{LMS}(n) + \mu u(n-1)e(n) \\
&e(n) = u(n) - \hat{w}_{LMS}(n)u(n-1) \\
&w_{LMS}(0) = 0.0
\end{aligned}$$

Figures 2-5 show the plots of $\hat{w}_{LMAD}(n), \hat{w}_{LMS}(n)$ versus the number of iterations where $\hat{w}_{LMAD}(n), \hat{w}_{LMS}(n)$ were obtained by averaging 30 independent trials of the experiments. For each trial, a different computer realization of the AR process $\{u(n)\}$ is used. In order to get a reasonable comparison between LMAD and LMS regarding the convergence rates and misadjustment, in each experiment we first adjust the step size of LMS to the maximum while ensuring the convergence of LMS and then choose a step size of LMAD to get misadjustment comparable with that of LMS.

A couple of important observations could be made here. First, in all cases, LMAD is simpler to implement than LMS. In each iteration, it still needs M multiplications and $M-1$ additions to compute the filter output and prediction error, but the tap weight update equation is much simpler. This reduction in computation, however, comes at the expense of performance when the driving noise is Gaussian. In this case, LMAD is slower than LMS, as shown in Figure 2 This situation changes dramatically, however, when α is less than 2. As α decreases, LMS becomes slower and slower to converge. In fact, when α is close to 1, LMS hardly converges without its step size approximately equal to zero.

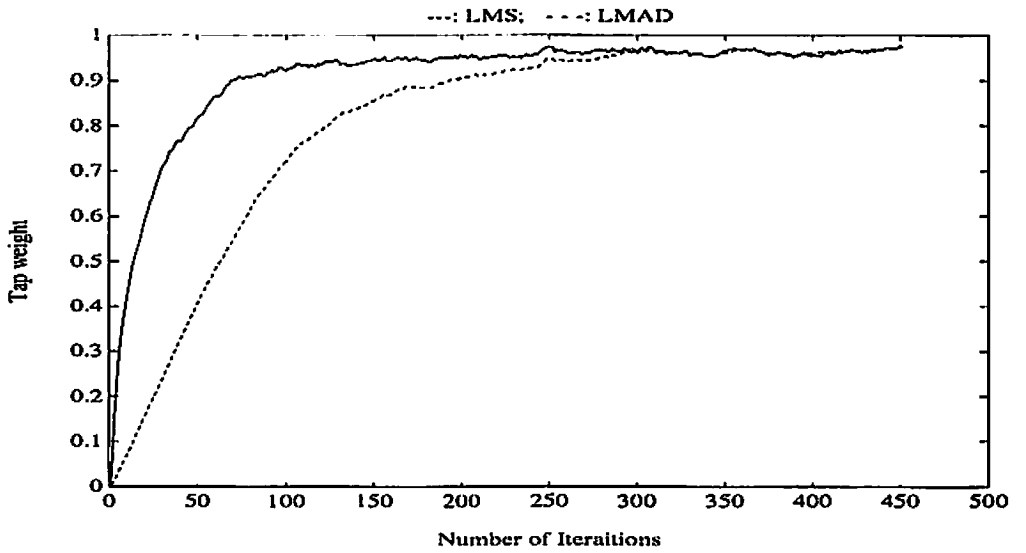

Figure 2: Transient behavior of tap weights in LMS and LMAD with $\alpha = 2.0$

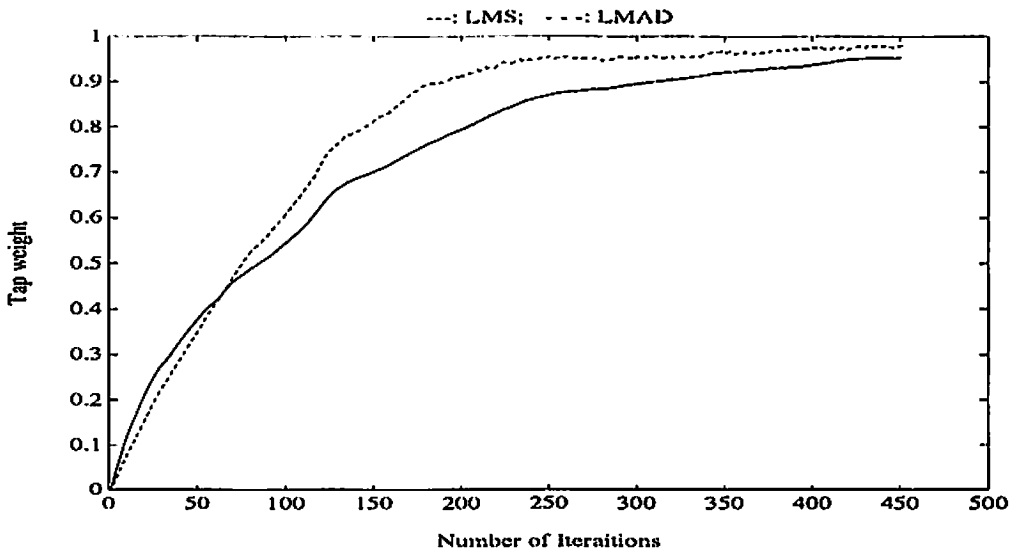

Figure 3: Transient behavior of tap weights in LMS and LMAD with $\alpha = 1.9$

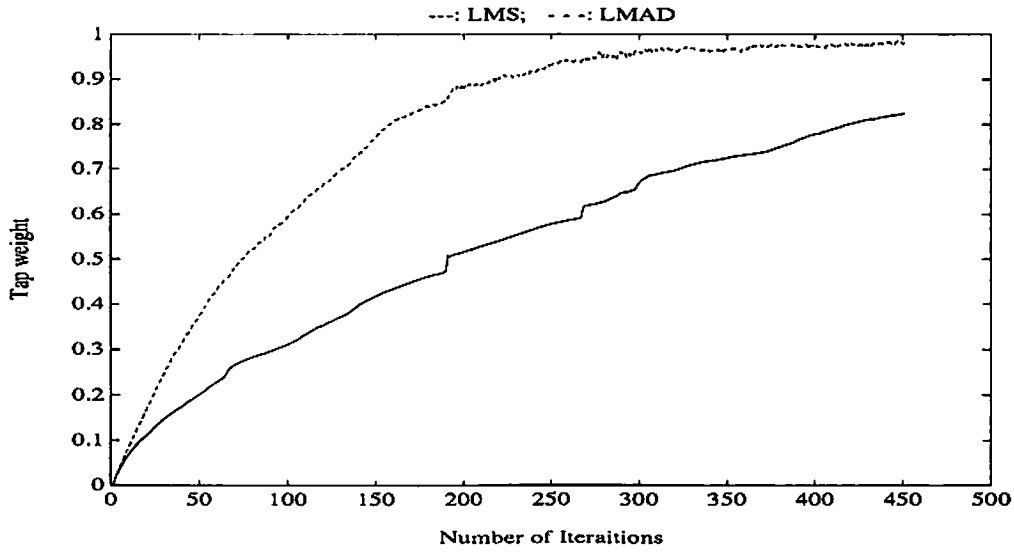

Figure 4: Transient behavior of tap weights in LMS and LMAD with $\alpha = 1.5$

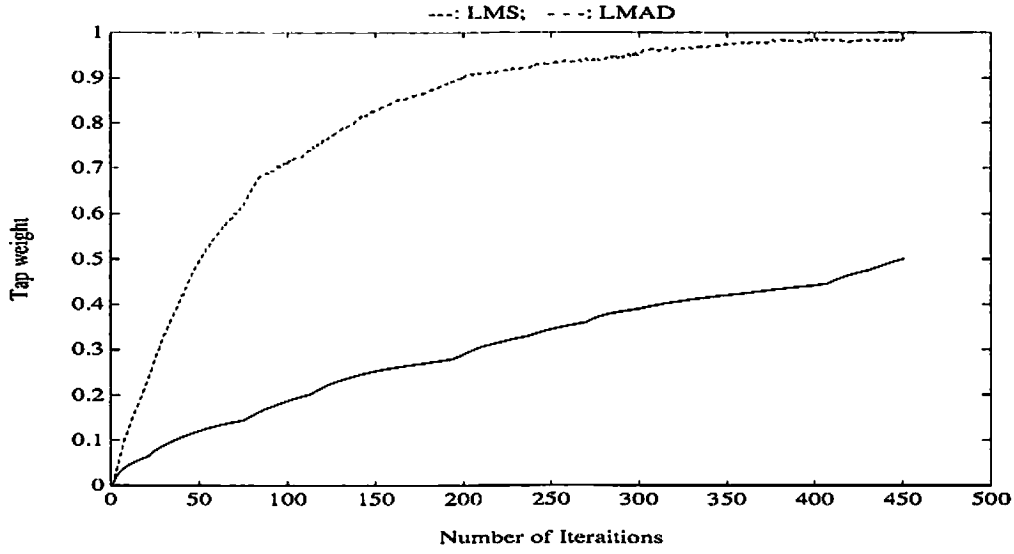

Figure 5: Transient behavior of tap weights in LMS and LMAD with $\alpha = 1.1$

LMAD, however, maintains fairly constant rate of convergence over the whole range of α. LMAD converges faster and faster relative to LMS as α becomes smaller and smaller while maintaining the same misadjustment as that of LMS.

This preliminary experiments show the distinct advantages of LMAD when the *AR* process is driven by white stable noise. It is a simple, effective method for adaptive filtering when we are dealing with stable processes.

6.5 Identification of LSI Systems

We consider a linear shift-invariant (LSI) system with a $S\alpha S$ process $X(n)$ as the input. Assume the impulse response of the system is $h(n)$. The output $Y(n)$ is then given by the following convolution

$$Y(n) = \sum_{k=0}^{\infty} h(k)X(n-k) \tag{83}$$

Observe that the output $Y(n)$ is also a $S\alpha S$ process and $Y(n)$ and $X(n)$ are jointly stationary.

In the following we show that if we know the covariation sequence $C_{xx}(n)$ and the cross covariation $C_{yx}(n)$ sequence, we can identify the system (that is, find $h(n)$). In fact by taking covariations of both sides with $X(n-m)$, one has the following relation

$$C_{yx}(m) = \sum_{k=1}^{\infty} h(k)C_{xx}(m-k) \tag{84}$$

We now take the Fourier transform of both sides to get

$$\hat{C}_{yx}(\omega) = H(\omega)\hat{C}_{xx}(\omega) \tag{85}$$

where $\hat{C}_{yx}(\omega), H(\omega), \hat{C}_{xx}(\omega)$ are the Fourier transforms of $C_{yx}(m), h(n), C_{xx}(n)$ respectively.

So in theory one may recover the impulse response $h(n)$ and/or its Fourier transform $H(\omega)$ from the covariation sequences through

$$H(\omega) = \hat{C}_{yx}(\omega)/\hat{C}_{xx}(\omega) \tag{86}$$

Note, however, that $\hat{C}_{xx}(\omega)$ is just the Fourier transform of the covariation sequence of

$X(n)$ and has no power spectrum meaning.

For a similar development for continuous-time systems, see [13].

7 Conclusion

Stable laws are direct generalizations of the Gaussian laws. They share some of the most important characteristics of the Gaussian, such as the stability property and central limit theorems. Although they have found important applications in diverse fields such as physics, economics as well as engineering, stable laws and processes are relatively unknown among electrical engineers. To increase the accessibility of stable models to engineers, this paper presented a tutorial review of stable distributions and stable signal processing with fractional lower order moments.

There are still a lot remaining to be done to better understand stable models and to develop algorithms for their specific signal processing applications. For example, adaptive signal processing algorithms can be developed and be applied in equalization, noise cancellation, underwater acoustic signal processing and signal enhancement, under non-Gaussian stable environments. These schemes can be expected to exhibit much faster convergence and obtain more robust results than other existing methods which are generally tailored specifically for the Gaussian signals. One could also investigate signal detection, parameter estimation and classification with stable distributions and fractional lower order moments, especially for signals and/or noise with infinite or very large variances. Filtering of impulsive noise and nonlinear stable signal analysis are yet two more areas worth investigating.

Appendix

In this appendix we summarize some of the basic facts about stable distributions. Most of the following results can be found in [44, 9, 23, 25, 56, 60, 30, 90]

A Univariate Stable Distribution

A.1 Basic Definitions

Definition 1 *A random variable X is said to have a stable distribution if for all X_1, X_2*

independent with the same distribution as X and for all $a_1 > 0, a_2 > 0$, there are constants $a > 0$ and b such that

$$X_1 + X_2 \stackrel{\mathrm{d}}{=} aX + b \tag{87}$$

where the notation $X \stackrel{\mathrm{d}}{=} Y$ means that X and Y have the same distribution. Equation (87) is called the *stability* or *superposability* property of stable distributions.

The stable laws arise naturally as limiting distributions by the following *Generalized Central Limit Theorem*: [9]:

Theorem 11 (Generalized Central Limit Theorem) *X is the limit in distribution of normalized sums of the form*

$$S_n = (X_1 + \cdots + X_n)/a_n - b_n$$

where $X_1, X_2, \ldots,$ are i.i.d and $a_n \to \infty$, if and only if X has a stable distribution.

In particular, if X_i's have mean μ and variance σ^2 and $A_n = 1/\sqrt{n}$, $B_n = \sqrt{n}\mu$, then the limiting distribution is Gaussian. This is of course the result of the ordinary Central Limit Theorem. Thus, normal distributions are all stable.

In the above theorem, in order for S_n to converge in distribution, X_n must have a distribution belonging to the domain of a stable law. Specifically,

Definition 2 *A distribution $G(x)$ is said to belong to the domain of a distribution $F(x)$ if there exist $a_n \to \infty$ and b_n such that the normalized sum*

$$S_n = (X_1 + X_2 + \cdots + X_n)/a_n - b_n \tag{88}$$

converges to $F(x)$ in distribution, where X_n are i.i.d and have distribution $G(x)$.

For example, every distribution with finite variance belongs to the domain of a normal distribution. By Theorem 1, one can see that a distribution function $F(x)$ is stable iff it possesses a non-empty domain of attraction. In fact, we have the following stronger result [23]:

Proposition 4 *If $F(x)$ is stable, then it belongs to its own domain of attraction.*

A.2 Characteristic and Probability Density Functions

The characteristic functions of stable laws can be written in the following parametric form [9, 23]:

Theorem 12 *A univariate distribution function $F(x)$ is stable iff its characteristic function has the following form*

$$\varphi(t) = \exp\{iat - \gamma \mid t \mid^{\alpha} [1 + i\beta \text{sign}(t)\omega(t, \alpha)]\} \tag{89}$$

where

$$\omega(t, \alpha) = \begin{cases} \tan \frac{\alpha\pi}{2} & \text{if } \alpha \neq 1 \\ \frac{2}{\pi} \log \mid t \mid & \text{if } \alpha = 1 \end{cases} \tag{90}$$

and

$$-\infty < a < \infty, \gamma > 0, 0 < \alpha \leq 2, -1 \leq \beta \leq 1 \tag{91}$$

Here:

1. α is called the *characteristic exponent*. It is uniquely determined. The distributions and corresponding random variables are called *α-stable*. α measures the "thickness" of the tails of the distribution. Thus, if a stable random variable is observed, the larger the value of α, the less likely it is to observe values of the random variable which are far from its central location. A small value of α will mean considerable probability mass in the tails of the distribution. An $\alpha = 2$ corresponds to a normal distribution (for any β), while $\alpha = 1, \beta = 0$ correspond to a Cauchy distribution. Distributions corresponding to other values of α are unknown.

2. γ is a scale parameter, also called the *dispersion* [79]. For the normal distribution it is equal to twice the variance. For non-Gaussian stable distributions, γ behaves like the variance of a Gaussian distribution.

3. β is a symmetry parameter. $\beta = 0$ implies a distribution symmetric about a. In this case, the distribution is called *symmetric α stable*, or simply $S\alpha S$. Symmetric stable distributions represent an important subclass of stable distributions. The Gaussian and Cauchy distributions are all $S\alpha S$.

4. a is a location parameter. For $S\alpha S$ distributions, it is the mean when $1 < \alpha \leq 2$ and the median when $0 < \alpha < 1$. For the symmetric Cauchy distribution a is the

semi-interquartile range.

A stable distribution is called *standardized* if $a = 0, \gamma = 1$. It is easy to see that if X is stable with parameters α, β, γ and a, then $(X - a)/\gamma^{\frac{1}{\alpha}}$ is a standardized variable with characteristic exponent α and symmetry parameter β.

By taking the inverse Fourier transform of the characteristic function, it is easy to show that the probability density function of the standardized stable variable is given by

$$f(x;\alpha,\beta) = \frac{1}{\pi}\int_0^\infty \exp(-t^\alpha)\cos[xt + \beta t^\alpha \omega(t,\alpha)]dt \tag{92}$$

Note that $f(x;\alpha,\beta) = f(-x;\alpha,-\beta)$. It can also be shown that the probability density functions of stable distributions are bounded and have derivatives of arbitrary orders [90].

Closed-form density and distribution functions are not available for general stable laws, except for the Gaussian, Cauchy and Pearson distributions [30]. But the power series expansion of the density of a stable random variable is available. The probability density function of the standardized stable variable can be expanded into absolutely convergent series as follows [4, 44, 23, 90, 30]: for $x > 0$

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k!}\Gamma(\alpha k+1)(\frac{x}{r})^{-\alpha k}\sin[\frac{k\pi}{2}(\alpha+\zeta)] & 0<\alpha<1 \\ \frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k!}\Gamma(\frac{k}{\alpha}+1)(\frac{x}{r})^{k}\sin[\frac{k\pi}{2\alpha}(\alpha+\zeta)] & 1<\alpha\le 2 \end{cases} \tag{93}$$

where

$$\eta = \beta\tan(\pi\alpha/2), r = (1+\eta^2)^{-1/(2\alpha)}, \zeta = -(2/\pi)\arctan\eta \tag{94}$$

In particular, the pdf of a $S\alpha S$ random variable is given by

$$f_\alpha(x) = \begin{cases} \frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k!}\Gamma(\alpha k+1)x^{-\alpha k}\sin[\frac{k\alpha\pi}{2}] & 0<\alpha<1 \\ \frac{1}{\pi\alpha}\sum_{k=0}^{\infty}\frac{(-1)^{k}}{2k!}\Gamma(\frac{2k+1}{\alpha})x^{2k} & 1<\alpha\le 2 \end{cases} \tag{95}$$

For the $S\alpha S$ distributions with $\alpha > 1$, asymptotic series are available for both the density and distribution functions. Specifically, for $x > 0$,

$$f_\alpha(x) = -\frac{1}{\pi}\sum_{1}^{n}\frac{(-1)^k}{k!}\frac{\Gamma(\alpha k+1)}{x^{\alpha k+1}}\sin(\frac{k\pi\alpha}{2}) + R(x) \tag{96}$$

where the remainder

$$R(x) = O(x^{-\alpha(n+1)-1}) \tag{97}$$

Term by term integration of (97) yields the asymptotic series for the cumulative distribution function for $x > 0$:

$$F_\alpha(x) = 1 + \frac{1}{\pi}\sum_{k=1}^{n}(-1)^k \frac{\Gamma(\alpha k)}{k! x^{\alpha k}} \sin(\frac{k\pi\alpha}{2}) + H(x) \tag{98}$$

where the remainder

$$H(x) = O(\frac{x^{-\alpha(n+1)}}{\alpha(n+1)}) \tag{99}$$

For the detailed tabulation of the stable distribution functions and graphs of stable probability density functions, see [30, 90].

A.3 Important Properties of Univariate Stable Laws

1. Every stable distribution with characteristic exponent $0 < \alpha < 2$ has finite absolute moments $\mathbf{E}(|x|^p)$ of order p for $0 < p < \alpha$. All absolute moments of order $p \geq \alpha$ are infinite [23].

   Thus for $0 < \alpha \leq 1$ the stable laws have no first or higher order moments; for $1 < \alpha < 2$, the stable laws have the first moment and all the fractional moments of order p where $p < \alpha$; for $\alpha = 2$, all moments exist. In particular, except for $\alpha = 2$, all stable laws have infinite variances.

2. The tails of $S\alpha S$ distributions with zero location parameter and dispersion γ satisfy [42, 86]

   $$\lim_{t\to\infty} t^\alpha P(|X| > t) = (C(\alpha))^\alpha \gamma$$

   where

   $$C(\alpha) = [\int_0^\infty \frac{\sin u}{u^\alpha} du]^{-1/2} \tag{100}$$

   Thus stable laws have inverse power tails. In contrast, normal distributions have exponential tails. This proves that the tails of stable laws are a lot thicker than those of normal distributions .

3. If $X_1, X_2, \ldots, X_n$ are independent and follow a stable law with the same (α, β), then $\sum_{j=1}^{n} X_j$ follows the same law, except for a possible location and scale change.

4. Any $S\alpha S$ random variable X can be written as a product of two independent random variables as follows:

$$X = Y^{\frac{1}{2}} G \tag{101}$$

where G is Gaussian, Y is $S\frac{1}{2}S$ (thus positive) and independent of G.

B Multivariate Stable Distributions

B.1 Characterizations of Multivariate Stable Laws

The multivariate stable distributions are again defined by the stability property:

Definition 3 *A k-dimensional distribution function $F(x)$, $x \in R^k$ is called* stable *if, for any independent identically distributed random vectors X_1, X_2 with distribution function $F(x)$, there exist $a > 0, b \in R^k$ and a random vector X with the same distribution function $F(x)$ such that*

$$X_1 + X_2 \stackrel{\mathrm{d}}{=} aX + b \tag{102}$$

One of the main differences between univariate and multivariate stable distributions is that the family of one-dimensional stable distributions form a parametric set while the family of multivariate stable distributions form a nonparametric set, as we can see from the characteristic functions of multivariate stable distributions [42, 60]:

Theorem 13 *A k-dimensional distribution function $F(\mathbf{x}), \mathbf{x} \in R^k$ is stable iff its characteristic function has the following form*

$$\varphi(\mathbf{t}) = \begin{cases} \exp\{i\mathbf{t}^T\mathbf{a} - \mathbf{t}^T A\mathbf{t}\} & \text{if } \alpha = 2 \\ \exp\{i\mathbf{t}^T\mathbf{a} - \int_S |\mathbf{t}^T\mathbf{s}|^\alpha \Gamma(d\mathbf{s}) + i\beta_\alpha(t)\} & \text{if } 0 < \alpha < 2 \end{cases} \tag{103}$$

where

$$\beta_\alpha(\mathbf{t}) = \begin{cases} \tan\frac{\alpha\pi}{2} \int_S |\mathbf{t}^T\mathbf{s}|^\alpha \mathrm{sign}(\mathbf{t}^T\mathbf{s})\Gamma(d\mathbf{s}) & \text{if } \alpha \neq 1.0 < \alpha < 2 \\ \int_S \mathbf{t}^T\mathbf{s} \log|\mathbf{t}^T\mathbf{s}|\Gamma(d\mathbf{s}) & \text{if } \alpha = 1 \end{cases} \tag{104}$$

and where $\mathbf{a}, \mathbf{t} \in R^k, S = \{\mathbf{x} \in R^k : |\mathbf{x}|\}, \Gamma =$ a finite Borel measure on S, A a positive semidefinite symmetric matrix of order.

Remark:

1. α is the *characteristic exponent* of the stable distributions. It is uniquely determined. The distribution is called α-stable. If $\alpha = 2$, then the stable distributions are the family of normal distributions with mean $\mathbf{a}$ and covariance matrix $2A$.

2. Except for the case $\alpha = 2$, the multivariate stable distributions form a nonparametric set. They are determined by a vector $\mathbf{a} \in R^k$, a scalar $0 < \alpha < 2$ and a certain finite measure $\Gamma(d\mathbf{s})$ on the sphere S.

3. $\mathbf{a}$ is a location vector.

4. Γ is called *spectral measure*. For $1 < \alpha < 2$, Γ is uniquely determined.

5. $\beta_\alpha(t)$ is called *asymmetry function*. If $\beta_\alpha \equiv 0$ then such stable distributions are *symmetric* and called symmetric α stable ($S\alpha S$).

B.2 Properties of Multivariate Stable Distributions

Most of the following results can be found in [60, 56, 66, 13]

1. All non-degenerate ($\gamma > 0$) stable distributions are absolutely continuous and have continuously differentiable densities.

2. The moment properties of the components of a stable random vectors follow the univariate results for moments [56].

3. If $1 < \alpha \leq 2$ then a random vector $\mathbf{X}$ follows a multivariate stable (or $S\alpha S$) law with characteristic component α iff every linear combination of the components of $\mathbf{X}$ follows a univariate stable (or $S\alpha S$) law with characteristic exponent α [86].

4. The marginal distributions corresponding to a multivariate stable (or $S\alpha S$) law are all stable (or $S\alpha S$) with the same characteristic exponent.

5. The characteristic functions of the form

$$\varphi(\mathbf{t}) = \exp(-2^{-\alpha/2}(\mathbf{t}^T R\mathbf{t})) \tag{105}$$

where the matrix R is positive definite, define an important subclass of multivariate $S\alpha S$ random vectors, the so-called α-sub Gaussian random vectors [60, 13]. This subclass is often denoted by α-SG(R). It is well known [13] that if $\mathbf{x} \in \alpha - \mathrm{SG(R)}$ then

$$\mathbf{x} = \eta \mathbf{y} \tag{106}$$

where η is a positive $S\frac{\alpha}{2}S$ random variable. $\mathbf{y}$ is Gaussian with mean zero and covariance matrix R. In addition, η and $\mathbf{y}$ are independent.

6. The question of how to characterize the independence of jointly $S\alpha S$ random variables is answered by the following theorem due to Miller [56]:

Theorem 14 *Let $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ be jointly $S\alpha S$ random variables with $0 < \alpha < 2$ and spectral measure Γ on the Borel subsets of the unit sphere S in R^{m+n}. Then $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$ are independent if and only if*

$$\Gamma\{(x, y) \in S : x \in R^m, y \in R^n, |x||y| \neq 0\} = 0$$

where $|x|, |y|$ are the Euclidean norms of x, y. In addition, for a family of $S\alpha S$ random variables, independence is equivalent to pairwise independence.

In particular, the component $S\alpha S$ random vector are independent iff Γ is concentrated on the finite set of points formed by the intersection of the coordinate axes and S.

7. It is shown in most engineering probability books that any Gaussian random vector can be whitened. Specifically, if $\mathbf{x}$ is a Gaussian vector, then $\mathbf{x}$ can be written as

$$\mathbf{x} = A\mathbf{y}$$

where A is a constant matrix and y is a Gaussian random vector with independent component. But in the stable case, we have the following lemma [70].

Proposition 5 *In general, representation of of even two stable variables of index α, $0 < \alpha < 2$ as the linear combination of a finite number of independent stable variables of the same index is impossible.*

The above lemma shows why we have to be very careful when we generalize something about Gaussian processes to stable processes.

References

[1] V. Akgiray and C. G. Lamoureux, "Estimation of Stable-Law Parameters: A Comparative Study," *J. of Business & Economic Statistics*, Vol. 7, pp. 85–93, 1989.

[2] H. An and Z. Chen, "On Convergence of LAD Estimates in Autoregression with Infinite Variance," *Journal of Multivariate Analysis*, Vol. 12, pp. 335–345, 1982.

[3] J. Berger and B. Mandelbrot, "A New Model for Error Clustering in Telephone Circuits," *IBM J. Research and Development*, Vol. 7, pp. 224–236, 1963.

[4] H. Bergstrom, "Some Expansions of Stable Distributions," *Arkvi for Mathematik*, Vol. 2, pp. 375 378, 1952.

[5] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, San Fransico: Holden-Day, Inc., 1977.

[6] R. Blattberg and T. Sargent, "Regression with non-Gaussian Stable Disturbances: Some Recent Results," *Econometrica*, Vol. 39, No. 3, pp. 501–510, 1971.

[7] P. Bloomfield and W. L. Steiger, *Least Absolute Deviations: Theory, Applications and Algorithms*, Boston: Birkhöuser, 1983.

[8] M. Bouvet and S. C. Schwartz, "Comparison of Adaptive and Robust Receivers for Signal Detection in Ambient Underwater Noise," *IEEE Transaction on Acoustic, Speech and Signal Processing*, Vol. ASSP-37, No. 5, pp. 621–626, 1989.

[9] L. Breiman, *Probability*, Addison-Wesley, 1968.

[10] B. W. Brorsen and S. R. Yang, "Maximum Likelihood estimates of Symmetric Stable Distribution Parameters," *Commun. Statist.-Simula.*, Vol. 19, pp. 1459–1464, 1990.

[11] S. Cambanis, "Complex Symmetric Stable Variables and Processes," in *Contributions to Statistics: Essays in Honour of Norman L. Johnson* (P. Sen, ed.), pp. 63–79, North-Holland Publishing Company, 1973.

[12] S. Cambanis and A. G. Miamee, "On Prediction of Harmonizable Stable Processes," *Sankhya A*, Vol. 51, pp. 269–294, 1989.

[13] S. Cambanis and G. Miller, "Linear Problems in pth Order and Stable Processes," *SIAM Journal of Applied Mathematics*, Vol. 41, No. 1, pp. 43–69, 1981.

[14] S. Cambanis and A. R. Soltani, "Prediction of Stable Processes: Spectral and Moving Average Representations," *Z. Wahrscheinlichkeitstheorie ver. Gebiete*, Vol. 66, pp. 593–612, 1984.

[15] J. F. Claerbout and F. Muir, "Robust Modeling with erratic Data," *Geophysics*, Vol. 38, No. 5, pp. 826–844, 1973.

[16] D. B. Cline and P. J. Brockwell, "Linear Prediction of ARMA Processes with Infinite Variance," *Stochastic Processes and Their Applications*, Vol. 19, pp. 281–296, 1985.

[17] E. Denoél and J. P. Solvay, "Linear Prediction of Speech with a Least Absolute Error Criterion," *IEEE Transaction on Acoustic, Speech and Signal Processing*, Vol. ASSP-33, No. 6, pp. 1397–1403, 1985.

[18] W. DuMouchel, *Stable Distributions in Statistical Inference*, Ph.D. dissertation, Department of Statistics, Yale University, 1971.

[19] W. DuMouchel, "Stable Distributions in Statistical Inference:2. Information from Stably Distributed Samples," *Journal of the American Statistic Association*, Vol. 70, No. 350, pp. 386 393, 1975.

[20] E. Fama, "The Behavior of Stock-Market Prices," *Journal of Business*, Vol. 38, pp. 34–105, 1965.

[21] E. Fama and R. Roll, "Some Properties of Symmetric Stable Distributions," *American Statistical Association Journal*, Vol. 63, pp. 817–836, 1968.

[22] E. Fama and R. Roll, "Parameter Estimates for Symmetric Stable Distributions," *American Statistical Association Journal*, Vol. 66, pp. 331-338, 1971.

[23] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2, New York: Wiley, 1966.

[24] E. C. Field, Jr. and M. Lewinstein, "Amplitude-Probability Distribution Model for VLF/ELF Atmospheric Noise," *IEEE Transaction on Communications*, Vol. COM-26, No. 1, 1978.

[25] B. Gnedenko and A. Kolmogorov, *Limit Distributions of Sums of Independent Random Variables*, Cambridge, Massachusetts: Addison-Wesley, 1954.

[26] R. Gonin and A. H. Money, *Nonlinear L_p-Norm Estimation*, New York: Marcel Dekker, Inc., 1989.

[27] S. Gross and W. L. Steiger, "Least Absolute Deviation Estimates in Autoregression with Infinite Variance," *Journal of Applied Probability*, Vol. 16, pp. 104–116, 1979.

[28] E. J. Hannan and M. Kanter, "Autoregressive Processes with Infinite Variance," *Journal of Applied Probability*, Vol. 14, pp. 411–415, 1977.

[29] C. D. Hardin, Jr., "On the linearity of regression," *Z. Wahr. und verw Gebiete*, Vol. 61, pp. 293–302, 1982.

[30] D. R. Holt and E. L. Crow, "Tables and Graphs of the Stable Probability Density Functions," *Journal of Research of the National Bureau of Standard - B. Mathematical Sciences*, Vol. 77B, No. 3 & 4, 1973.

[31] J. Holtsmark, "Über die Verbreiterung von Sektrallinien," *Ann. Physik*, Vol. 58, No. 4, pp. 577–630, 1919.

[32] Y. Hosoya, "Discrete-time Stable Processes and Their Certain Properties," *The Annals of Probability*, Vol. 6, No. 1, pp. 94–105, 1978.

[33] Y. Hosoya, "Harmonizable Stable Processes," *Z. Wahrschenlichkeitstheorie verw. Gebiete*, Vol. 60, pp. 517–533, 1982.

[34] L. Izzo, L. Panico, and L. Paura, "Error Rates for Fading NCFSK Signals in an Additive Mixture of Impulsive and Gaussian Noise," *IEEE Transactions on Communications*, Vol. COM-30, pp. 2434–2438, 1982.

[35] L. Izzo and L. Paura, "Error Probability for Fading CPSK Signals in Gaussian and Impulsive Atmospheric noise environments," *IEEE Transactions on Aerospace Electronic Systems*, Vol. AES-17, pp. 719–722, 1981.

[36] R. Y. J. Schroeder and J. Hershey, "L_p Normed Minimization with Applications to Linear Predictive Modeling for Sinusoidal Frequency Estimation," *Signal Processing*, Vol. 24, pp. 193–216, 1991.

[37] M. Kanter and W. Steiger, "Regression and Autoregression with Infinite Variance," *Adv. Appl. Prob.*, Vol. 6, pp. 768–783, 1974.

[38] S. A. Kassam, *Signal Detection in non-Gaussian Noise*, New York: Springer-Verlag, 1988.

[39] I. Koutrouvelis, "An Iterative Procedure for the Estimation of the Parameters of Stable Laws," *Communications in Statistics: Simulation and Computation*, Vol. 10, No. 1, 1981.

[40] I. A. Koutrouvelis, "Regression-Type Estimation of the Parameters of Stable Laws," *Journal of the American Statistical Association*, Vol. 75, No. 412, pp. 918–928, 1980.

[41] R. M. Lerner, "Design of Signals," in *Lectures on Communication System Theory* (E. J. Baghdady, ed.), pp. 243–277, New York: McGraw-Hill Book Company, Inc., 1961.

[42] P. Lévy, *Calcul des probabilités*, Paris: Gauthier-Villars, 1925.

[43] K. Lii and M. Rosenblatt, "Deconvolution and Estimation of Transfer Function Phase and Coefficients for NonGaussian Linear Processes," *The Annals of Statistics*, Vol. 10, No. 4, pp. 1195–1208, 1982.

[44] E. Lukacs, *Characteristic Functions*, London: Charles Griffin and Co., 1960.

[45] F. W. Machell, C. S. Penrod, and G. E. Ellis, "Statistical Characteristic of Ocean Acoustic Noise Processes," in *Topics in Non-Gaussian Signal Processing* (E. J. Wegman, S. C. Wegman, and J. B. Thomas, eds.), pp. 29–57, Springer-Verlag, 1989.

[46] B. Mandelbrot, "The Variation of Certain Speculative Prices," *Journal of Business*, Vol. 36, pp. 394–419, 1963.

[47] B. Mandelbrot and J. W. Van Ness, "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, Vol. 10, pp. 422–437, 1968.

[48] M. Marques and S. Cambanis, "Admissible and Singular Translates of Stable Processes," in *Probability Theory on Vector Spaces IV* (S. Cambanis, ed.), Lecture Notes in Mathematics, pp. 239–257, Springer-Verlag, 1987.

[49] E. Mastry and S. Cambanis, "Spectral Density Estimation for Stationary Stable Processes," *Stochastic Processes and their Applications*, Vol. 18, pp. 1–31, 1984.

[50] W. B. McCain and C. D. McGillem, "Performance Improvement of DPLL's in Non-Gaussian Noise Using Robust Estimators," *IEEE Transaction on Communications*, Vol. COM-35, No. 11, 1987.

[51] J. H. McCulloc, "Simple Consistent Estimators of Stable Distribution Parameters," *Communications in Statistics: Simulation and Computation*, Vol. 15, No. 4, 1986.

[52] J. M. Mendel, *Lessons in Digital Estimation Theory*, Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1987.

[53] P. Mertz, "Model of Impulsive Noise for Data Transmission," *IRE Transaction on Communications Systems*, June 1961.

[54] D. Middleton, "Statistical-Physical Models of Electromagnetic Interference," *IEEE Transaction on Electromagnetic Compatibility*, Vol. EMC-19, No. 3, 1977.

[55] D. Middleton, "Procedures for Determining the Parameters of the First-Order Canonical Models of Class A and Class B Electromagnetic Interference," *IEEE Transaction on Electromagnetic Compatibility*, Vol. EMC-21, No. 3, 1979.

[56] G. Miller, "Properties of Certain Symmetric Stable Distribution," *Journal of Multivariate Analysis*, Vol. 8, pp. 346–360, 1978.

[57] J. H. Miller, "Detections for Discrete-Time Signals in Non-Gaussian Noise," *IEEE Transaction on Information Theory*, Vol. IT-18, No. 2, 1972.

[58] C. Nikias and M. R. Raghuveer, "Bispectrum Estimation: A Digital Signal Processing Framework," *IEEE Proc.*, Vol. 75, No. 7, pp. 869–891, 1987.

[59] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, New York: McGraw-Hill, Inc., third ed., 1991.

[60] V. J. Paulauskas, "Some Remarks on Multivariate Stable Distributions," *Journal of Multivariate Analysis*, Vol. 6, pp. 356–368, 1976.

[61] A. S. Paulson, E. W. Holcomb, and R. Leitch, "The Estimation of the Parameters of the Stable Laws," *Biometrika*, Vol. 62, pp. 163–170, 1975.

[62] S. S. Pillai and M. Harisankar, "Simulated Performance of a DS Spread-Spectrum System in Impulsive Atmospheric Noise," *IEEE Transaction on Electromagnetic Compatibility*, Vol. EMC-29, No. 1, 1987.

[63] M. Pourahmad, "On Minimality and interpolation of Harmonizable Stable Processes," *SIAM J. of Applied Mathematics*, Vol. 44, No. 5, 1984.

[64] S. Press, "A Compound Events Model For Security Prices," *Journal of Business*, Vol. 40, pp. 317–335, 1968.

[65] S. J. Press, "Estimation in Univariate and Multivariate Stable Distributions," *Journal of the American Statistical Association*, Vol. 67, pp. 842–846, 1972.

[66] S. J. Press, "Stable Distributions: Probability, Inference, and Applications in Finance - A Survey, and A review of Recent Results," in *Statistical Distributions in Scientific Work* (G. Patil, ed.), pp. 87–102, Dordrecht, Holland: D. Reidel Publishing Company, 1975.

[67] S. S. Rappaport and L. Kurz, "An Optimal Nonlinear Detector for Digital Data Transmission Through Non-Gaussian Channels," *IEEE Transaction on Communication Technology*, Vol. COM-14, No. 3, 1966.

[68] J. R. Rice and J. S. White, "Norms for Smoothing and Estimation," *SIAM Review*, Vol. 6, No. 3, pp. 243–256, 1964.

[69] L. Scarf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*, Reading, Mass.: Addison-Wesley, 1991.

[70] M. Schilder, "Some Structure Theorems for the Symmetric Stable Laws," *The Annals of Mathematical Statistics*, Vol. 41, No. 2, pp. 412–421, 1970.

[71] M. R. Schroeder, *Fractals, Chaos, Power Laws: Minutes from an Infinite Universe*, New York: W. H. Freeman, 1991.

[72] S. C. Schwartz and J. B. Thomas, "Detection in a Non-Gaussian Environment," in *Statistical Signal Processing* (E. J. Wegman and J. G. Smith, eds.), pp. 93–105, Marcel Dekker, Inc., 1984.

[73] J. Seo, S. Cho, and K. Feher, "Impact of Non-Gaussian Impulsive Noise on the Performance of High-Level QAM," *IEEE Transaction on Electromagnetic Compatibility*, Vol. EMC-31, No. 2, 1989.

[74] M. Shao and C. Nikias, "Impulsive Noise Modeling with Stable Distributions." Working Paper, 1992.

[75] R. Shiavi, *Introduction to Applied Statistical Signal Analysis*, Homewood, IL.: Aksen Associates, 1991.

[76] M. P. Shinde, "Signal Detection in the Presence of Atmospheric Noise in Tropics," *IEEE Transaction on Communications*, Vol. COM-22, No. 8, 1974.

[77] I. Singer, *Best Approximation in Normed Linear Space by Elements of Linear Subspaces*, New York: Springer-Verlag, 1970.

[78] A. D. Spaulding, "Optimum Threshold Signal Detection in Broad-Band Impulsive Noise Employing Both Time and Spatial Sampling," *IEEE Transaction on Communications*, Vol. COM-29, No. 2, 1981.

[79] B. Stuck, "Minimum Error Dispersion Liner Filtering of Scalar Symmetric Stable Processes," *IEEE Transaction on Automatic Control*, Vol. AC-23, No. 3, pp. 507–509, 1978.

[80] B. Stuck and B. Kleiner, "A Statistical Analysis of Telephone," *The Bell System Technical Journal*, Vol. 53, No. 7, pp. 412–421, 1974.

[81] A. Tarantola, *Inverse Problem Theory*, Amsterdam: Elsevier, 1987.

[82] H. L. Taylor, S. C. Banks, and J. F. McCoy, "Deconvolution with the l_1 Norm," *Geophysics*, Vol. 44, No. 1, pp. 39–45, 1979.

[83] J. Teichmoeller, "A Note on the Distribution of Stock Price Changes," *Journal of American Statistics Association*, Vol. 66, pp. 282–284, 1971.

[84] E. J. Wegman, S. C. Schwartz, and J. B. Thomas, eds., *Topics in Non-Gaussian Signal Processing*, New York: Springer-Verlag, 1989.

[85] E. J. Wegman and J. G. Smith, eds., *Statistical Signal Processing*, New York: Marcel Dekker, Inc., 1984.

[86] A. Weron, "Stable Processes and Measures: A Survey," in *Probability Theory on Vector Spaces III* (S. Cambanis, ed.), Lecture Notes in Mathematics, pp. 306–364, Springer-Verlag, 1983.

[87] A. Weron, "Harmonizable Stable Processes on Groups: Spectral, Ergodic and Interpolation Properties," *Z. Wahr. verw. Gebiete*, 1985.

[88] A. Weron and K. Weron, "Stable Measures and Processes in Statistical Physics," in *Probability Theory on Vector Spaces IV* (S. Cambanis, ed.), Lecture Notes in Mathematics, pp. 440–452, Springer-Verlag, 1987.

[89] V. J. Yohai and R. A. Maronna, "Asymptotic Behavior of Least-Squares Estimates for Autoregressive Processes with Infinite Variances," *The Annals of Statistics*, Vol. 5, No. 3, pp. 554–563, 1977.

[90] V. Zolotarev, *One-dimensional Stable Distributions*, Providence, Rhode Island: American Mathematical Society, 1986.