SYMMETRIC ALPHA-STABLE FILTER THEORY

John S. Bodenschatz
University of Southern California, Los Angeles, CA, USA
bodensch@sipi.usc.edu

ABSTRACT

Symmetric α-Stable (SαS) processes are used to model impulsive noise. Wiener filter theory is generally not meaningful in SαS environments because the expectations may be unbounded. To develop a filter theory for linear finite impulse response systems with independent identically distributed SαS inputs, we propose median orthogonality as a linear filter criterion, present a generalized Wiener-Hopf solution equation, and show a necessary condition for a filter to achieve the criterion. For non-Gaussian SαS densities, zero-forcing least-mean-square is the only well-known filter that satisfies the criterion, but others can easily be designed. We present a second algorithm and simulations showing that both converge to the generalized Wiener-Hopf solution.

1. INTRODUCTION

Conventional Wiener filter theory describes the behavior of adaptive filters based upon the least squares criterion. The Wiener-Hopf equation predicts the final values of the filter taps based upon the data statistics. The theory has many applications such as system identification, inverse modeling, prediction, and interference cancelation. After convergence, the variance of the filter error signal is minimized and the error signal is orthogonal to the input signal [1]. These concepts are based upon L_2 measurements and are therefore only of extremely limited use when dealing with signals that have infinite variance.

Symmetric α-stable (SαS) processes do have infinite variance and are very useful for modeling impulsive environments. The SαS distributions arise from varying the exponent in the Gaussian characteristic function; the SαS characteristic function is

$$\varphi(\omega) = e^{-\gamma|\omega|^{\alpha}}, \tag{1}$$

where $0 < \alpha \leq 2$. With $\alpha = 2$, a Gaussian distribution results, and, with $\alpha = 1$, the distribution is Cauchy. For other values of α, there is no closed form representation of the density function. The dispersion, γ, is a scale parameter. The densities are closed under addition and scalar multiplication. The justification for modeling with stable distributions is based upon the Generalized Central Limit Theorem which states that if a limit exists for a sum of Independent Identically Distributed (i.i.d.) variables, then this limit must be a stable distribution. Many problems have symmetry and, for these, a symmetrical distribution is appropriate [2].

A large range of phenomena can be modeled by α-stable theory. The first use was by Holtsmark, a Danish astronomer, who found that gravitational fields can fluctuate with an α of 1.5. A number of economic variables including stock prices have been shown to be α-stable. Many types of noise are α-stable such as, underwater acoustic, low-frequency atmospheric, phone line, and several man-made noises [2].

In this work, we create a foundation of filter theory for linear Finite Impulse Response (FIR) systems of SαS processes. We start by defining SαS generalized linear variables and processes, projection vectors, and median orthogonality. To provide background, several properties of projection vectors are described. We summarize the relationship between median orthogonality and projection vectors in (8). Median orthogonality is proposed as a criterion for a linear filter and two filter algorithms are presented. The generalized Wiener-Hopf (16) arises from this criterion and the properties of projection vectors. Simulations illustrate that the two filter algorithms do converge to the solution given by the generalized Wiener-Hopf equation.

2. DEFINITIONS

2.1. Median Orthogonality

We will use *MO* as an abbreviation for both *median orthogonality* and *median orthogonal*. Let u_1 and u_2 be two random variables (or processes) and let $M\{\}$ denote the median operator, which is similar to the expectation operator $E\{\}$. If the median of the product is zero, notated as

$$M\{u_1 u_2\} = 0 \tag{2}$$

then u_1 and u_2 are said to be MO,

$$u_1 \perp_M u_2. \tag{3}$$

In the Gaussian case, MO reduces to the conventional definition of orthogonality and is synonymous with independence. For the non-Gaussian symmetric case, independence is sufficient for MO; MO is necessary (but not sufficient) for independence.

2.2. SαS Generalized Linear Variables (SαSGLVs) & Projection Vectors

The established definition of a linear stable process [2] is not broad enough to represent the statistics of a linear

FIR system with SαS inputs, so we define SαS Generalized Linear Variables (SαSGLVs) and processes. With $\{x_i, i = 0, \pm1, \pm2, ...\}$ as set of i.i.d. SαS random variables, a set SαSGLVs $\mathbf{u} = \{u_m\}$ is defined using

$$u_m = \sum_{i=-\infty}^{\infty} \lambda_m(i) x_i. \tag{4}$$

Thus, the joint density of the two SαSGLVs u_m and u_n is completely determined by the corresponding vectors λ_m and λ_n, which will be termed *projection vectors* since they determine the projection onto the space of independent random variables.

2.3. SαS Generalized Linear Processes (SαSGLPs) & Projection Vectors

The concept of generalized linear variables can readily be applied to random processes. Let $\{X_{i,j}, i, j = 0, \pm1, \pm2, ...\}$ be a two-dimensional infinite set of i.i.d. SαS random variables, we define a set of SαS generalized linear processes $\{u_m\}$ with

$$u_m(i) = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \lambda_m(k, j) X_{k,i-j}, \tag{5}$$

where λ_m is the projection vector for u_m. The set projection vectors will represent the complete statistics of a linear FIR system with i.i.d. SαS inputs.

3. PROPERTIES OF PROJECTION VECTORS

3.1. Addition of SαS variables

When a set of independent SαS variables with the same characteristic exponent α are added, the distribution of the sum is also SαS with the same exponent α. The dispersion of the sum is obtained by summation of the individual dispersions. If x is an SαS random variable with $\gamma_x = 1$, then the random variable ax has a dispersion given by $\gamma_{ax} = |a|^\alpha$. Let $\{x_0, x_1, x_2, ...x_N\}$ be set of i.i.d. SαS random variables, the density of ax_0 will be the same as the density of the sum $\sum_{i=1}^{N} b_i x_i$ iff

$$|a| = \left(\sum_{i=1}^{N} |b_i|^\alpha \right)^{1/\alpha}. \tag{6}$$

3.2. Dimensionality

The dimensionality of the projection vectors should be viewed as the number of non-zero elements of the projection vector; these correspond to the dimensions of the i.i.d. probability space. The fact that the notation $\lambda_m(k, j)$ in (5) implies a two-dimensional vector is not meaningful for our work. To avoid this potential ambiguity and simplify notation, we will work with SαSGLVs, and the results will hold for SαSGLPs as well.

Apart from α, the probability density of a vector $\mathbf{u}$ composed of SαSGLVs will be determined by the set of projection vectors. If $\mathbf{u}$ has only one element, the density is merely specified by the width statistic. If $\mathbf{u}$ has more than one element, the joint density will be specified by a set of projection vectors, one for each element of $\mathbf{u}$; each projection vector may require an infinite number of non-zero elements. However, in this work, we focus on linear FIR systems which can be modeled with projection vectors of finite length.

3.3. Matrix Notation & Nonuniqueness

Equation (4) can be rewritten as

$$\mathbf{u} = \mathbf{\Lambda}^T \mathbf{x}, \tag{7}$$

where the ordering of the rows of the projection matrix $\mathbf{\Lambda}$ is completely arbitrary. Using the addition properties, a variety of other manipulations may also be applied without changing the statistics of $\mathbf{u}$. In the Gaussian case, we can always reduce the width so that $\mathbf{\Lambda}$ is square. Also, for finite length projection vectors, equation (5) can easily be reshaped and represented using (7).

4. MO & PROJECTION VECTORS

If two SαSGLVs (or SαSGLPs) are MO, their product must also have an even density. By using the properties of projection vectors and other manipulations, we can show that a simple condition upon the projection vectors corresponds to MO. The relations are summarized by

$$\begin{aligned} u \perp_M v,\ uv = z \quad &\Leftrightarrow M\{z\} = 0 \Leftrightarrow p_z(z) = p_z(-z) \\ &\Leftrightarrow 0 = \textstyle\sum_{i=-\infty}^{\infty} \left[\lambda_u(i)\, \lambda_v(i)\right]^{<\alpha/2>}, \end{aligned} \tag{8}$$

where the scalar signed exponential operator is given by

$$a^{<b>} = |a|^b * \mathrm{sign}(a). \tag{9}$$

Later, this operator is applied to a matrix, where it operates on each element individually, without changing the dimensions.

5. MO & ADAPTIVE FILTERS

5.1. Adaptive Transversal Filters

Since transversal filters (or tapped delay lines) are the most common and easiest filters to describe, we will focus on them. The linear filter equation is

$$e_i = d_i - \mathbf{w}_i^T \mathbf{u}_i \tag{10}$$

where e_i is the error signal. d_i is the desired response which is provided to the filter. $\mathbf{u}_i$ is the input vector. $\mathbf{w}_i$ is the tap vector, and $\mathbf{w}_i^T \mathbf{u}_i$ is the filter output.

5.2. MO Filter Criterion

The MO criterion is that the error should be median orthogonal to all elements of the input vector,

$$e \perp_{MO} \mathbf{u}. \tag{11}$$

This extends the criterion of conventional orthogonality without restricting the densities of e or $\mathbf{u}$. It does not necessarily define a unique solution, nor does the least squares criterion in an underdetermined system.

5.3. Tap Updates for MO

5.3.1. General Condition

To satisfy the criterion at the stability point, we can use any odd function $\mathbf{h}(\cdot)$, where $\mathbf{h}(\cdot)$ has the same sign as its argument and $E\{\mathbf{h}(e\mathbf{u})\}$ is finite. When $\mathbf{h}(\cdot)$ is applied to a vector, it must operate on the elements individually, without changing the dimensions. A tap vector update of the form

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \mathbf{h}(e\mathbf{u}), \tag{12}$$

will have a stability point where $e \perp_M \mathbf{u}$, because the integral of the even density function and the odd $\mathbf{h}(\cdot)$ is sufficient for $E\{\mathbf{h}(e\mathbf{u})\} = 0$. Unfortunately, this does not show which functions have the best convergence properties.

5.3.2. Cost Functions

The novelty is that we are not starting with a cost function and then finding a gradient. Instead, we define a solution criterion and find a type of tap update that has a stability point at the solution. To view the tap update as a stochastic gradient algorithm, we define a cost function by the integral of the expectation of the stepping in the limit as the step size goes to zero. Different updates with the same stability points can often be created. In this case, the corresponding cost functions will differ in shape but will have the same point (or points) of minimization.

There is a very good reason we take this novel approach for SαS variables. A traditional cost function $\mathbf{J}$ would be a function of e. Differentiating $\mathbf{J}$ with respect to $\mathbf{w}$ gives rise to $\frac{\partial e}{\partial \mathbf{w}}$. Since $\frac{\partial e}{\partial \mathbf{w}} = \mathbf{u}$ which has an undefined mean value for $\alpha \le 1$, there does not appear to be a way to create a good algorithm from the expectation of the derivative with respect to $\mathbf{w}$.

5.3.3. Zero-Forcing least-mean-square (ZFLMS)

Zero-forcing least-mean-square, also known as the sign-sign algorithm [2], has an update that can be written as

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mu \operatorname{sign}(\mathbf{u}_i e_i). \tag{13}$$

This fulfills the MO criterion.

5.3.4. Symmetric Least Mean P-norm (SLMP)

We introduce the Symmetric Least Mean P-norm (SLMP) algorithm which also satisfies the MO criterion. The filter update is

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mu \left(\mathbf{u}_i e_i\right)^{<p/2>}. \tag{14}$$

When $\alpha < 2$, we must have $p < \alpha$ for a bounded update. When $p = 2$, this is the well-known least-mean-square algorithm, and, with $p = 0$, we have ZFLMS.

6. GENERALIZED WIENER-HOPF EQUATION

For the general case, we use the SαSGLV matrix notation (7) to represent the statistics of the filter variables

$$\begin{bmatrix} d \\ \mathbf{u}_{Lx1} \end{bmatrix} = \begin{bmatrix} (\mathbf{r}_{Mx1})^T \\ (\mathbf{Q}_{MxL})^T \end{bmatrix} [\mathbf{x}_{Mx1}], \tag{15}$$

where

- $\mathbf{x}$ is the vector of i.i.d. SαS variables,
- $\mathbf{r}$ is the projection vector for the desired signal d, and
- $\mathbf{Q}$ is the matrix of projection vectors for the filter input $\mathbf{u}$.

Using (8), we can show that to achieve the MO criterion (11), the filter $\mathbf{w}$ must satisfy the generalized Wiener-Hopf equation,

$$\left\{ \begin{bmatrix} 1 & -(\mathbf{w}_{Lx1})^T \end{bmatrix} \begin{bmatrix} \mathbf{r}^T \\ \mathbf{Q}^T \end{bmatrix} \right\}^{<\alpha/2>} \mathbf{Q}^{<\alpha/2>} = \mathbf{0}_{1xL}, \tag{16}$$

where $\mathbf{0}$ is an all-zero vector. This is a non-linear equation; however, there are a few cases with a close-form solution.

6.1. Reduction to Wiener-Hopf with Gaussian Noise

When $\alpha = 2$, the signed exponential operators vanish and (16) reduces to the standard Wiener-Hopf equation $E\left\{\mathbf{u}\mathbf{u}^T\right\}\mathbf{w} = E\{d\mathbf{u}\}$, with the input autocorrelation $E\left\{\mathbf{u}\mathbf{u}^T\right\} = \mathbf{Q}^T\mathbf{Q}$ and $E\{d\mathbf{u}\} = \mathbf{Q}^T\mathbf{r}$.

6.2. Solution for the Square Case

When the number of nonzero columns of $\mathbf{Q}$ is equal to the number of taps plus one, the augmented matrix will be square and the solution has a closed form, because the exponential operations may be applied to the constants. The solution for a square projection matrix,

$$\mathbf{\Lambda}^T_{(L+1)x(L+1)} = \begin{bmatrix} \left(\mathbf{r}_{(L+1)x1}\right)^T \\ \left(\mathbf{Q}_{(L+1)xL}\right)^T \end{bmatrix}, \tag{17}$$

is

$$\begin{bmatrix} 1 & -\mathbf{w}^T \end{bmatrix} = \begin{bmatrix} \mathbf{r}^T \\ \mathbf{Q}^T \end{bmatrix}^{-1} \left[\beta^T\right]^{<2/\alpha>} \lambda', \tag{18}$$

where β is a vector in the left nullspace of $\mathbf{Q}^{<\alpha/2>}$. (Obviously, the inverse must exist.) λ' is a constant which is adjusted so that the leftmost element of the right-hand side of (18) is unity.

6.3. Solution for a System Identification Model

If each input u_j is independent with all other u_k (for $j \neq k$) and each u_j only overlaps d in one dimension of the i.i.d. space, we can use the model

$$\begin{bmatrix} d_{1x1} \\ \mathbf{u}_{Lx1} \end{bmatrix} = \begin{bmatrix} (\mathbf{a}_{Lx1})^T & (\mathbf{0}_{Lx1})^T & b_{1x1} \\ \mathbf{S}_{LxL} & \mathbf{N}_{LxL} & \mathbf{0}_{Lx1} \end{bmatrix} \left[\mathbf{x}_{(2L+1)x1}\right], \tag{19}$$

where $\mathbf{S}$ is a square diagonal matrix representing the signal subspace in $\mathbf{u}$ and $\mathbf{N}$ is a square diagonal matrix representing the noise subspace in $\mathbf{u}$. The solution is

$$\mathbf{w} = \left(\mathbf{S}^2 + \mathbf{N}^2\right)^{-1} \mathbf{S}\mathbf{a}, \tag{20}$$

assuming the inverse exists. In the Gaussian case, $\mathbf{S}^2$ is the autocorrelation of the signal in $\mathbf{u}$, $\mathbf{N}^2$ is the autocorrelation of the noise, and $E\{\mathbf{u}d\} = \mathbf{S}\mathbf{a}$. As an aside, equation (20) can be derived when the i.i.d. vector $\mathbf{x}$ in (19) is only symmetric; a stable density is not required.

7. SIMULATIONS

We study the convergence of ZFLMS and SLMP to a filter value which satisfies the generalized Wiener-Hopf equation. The values produced during the iterations are compared to the numerical solution of the non-linear generalized Wiener-Hopf equation.

Obviously, the generalized Wiener-Hopf equation cannot be studied for all possible values. In the Gaussian case, if $M < L$, we have an underdetermined system, so we restricted the simulations so that $M \geq L$. We choose $\mathbf{r}$ and $\mathbf{Q}$ as random matrices with i.i.d. Gaussian entries. Chambers' algorithm is used to generate the SαS deviates [2]. For computational expediency, we use a recursive block implementation with an adaptive step-size parameter for ZFLMS and SLMP. This merely speeds computation; it does not affect the value of the solution. We made many runs with different matrix sizes. Convergence can be very fast or very slow. Sometimes ZFLMS is much faster than SLMP; sometimes the opposite occurs, and sometimes they step almost identically. Misadjustment, the error after the convergence of a stochastic gradient algorithm, is also varied.

For illustration, we choose $\mathbf{Q}$ to be $12x4$ to give a meaningful system size. Each iteration is an adjustment of $\mathbf{w}$ based upon the errors in a block of $1,000$ values of $\mathbf{u}$ and d. The convergence of the ZFLMS and SLMP tap vectors toward this true solution is shown in Figure 1. The error after $5,000$ such iterations is shown as a function of α in Figure 2. The variable step-size parameter permits a very fast convergence in the first 30 steps. Allowing for a small misadjustment error, these plots illustrate that the generalized Wiener-Hopf equation does correctly predict the solution.

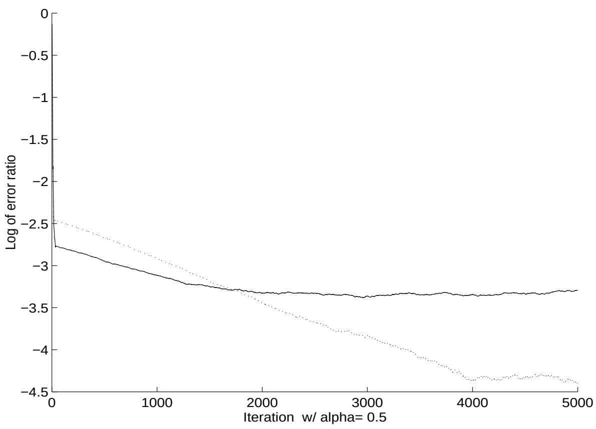

Figure 1. Convergence of the tap vector to solution of the generalized Wiener-Hopf equation at $\alpha = 0.5$, using ZFLMS (solid line) and SLMP (dotted line). The vertical scale is the natural logarithm of the ratio L_2 distance to the L_2 norm of the generalized Wiener filter. The value of this error ratio at initialization is 0.

8. CONCLUDING REMARKS

We have given a filter theory for linear FIR systems with i.i.d. SαS inputs. MO serves as a filter criterion and gives rise to a generalized version of the Wiener-Hopf equation

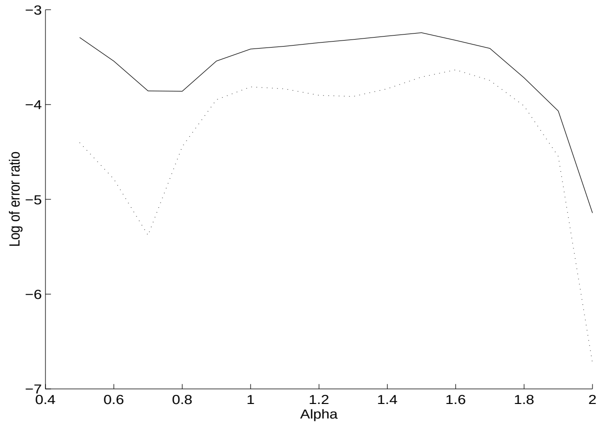

Figure 2. Natural Log of the error of the tap vector after 5,000 iterations as a function of α, using ZFLMS (solid line) and SLMP (dotted line). The vertical scale is the natural logarithm of the ratio L_2 distance to the L_2 norm of the generalized Wiener filter.

which predicts the solution of MO filters such as the ZFLMS and SLMP adaptive filters. Since we have a point of stability and filter updates, we can define corresponding cost functions and view ZFLMS and SLMP as stocastic gradient algorithms. If a system has an input that is not SαS, the theories in this work will not strictly hold. However, the non-SαS inputs, even those with finite variance, may be modeled as SαS with the appropriate width statistics. Some accuracy is obviously lost, but the results will generally be useful.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, NJ, third edition, 1996.

[2] Chrysostomos L. Nikias and Min Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, John Wiley & Sons, New York, 1995.