

## Fuzzy prediction and filtering in impulsive noise

Hyun Mun Kim, Bart Kosko\*

*Department of Electrical Engineering-Systems, Signal and Image Processing Institute, University of Southern California,  
Los Angeles, CA 90089-2564, USA*

---

### Abstract

Additive fuzzy systems can filter impulsive noise from signals. Alpha-stable statistics model the impulsiveness as a parametrized family of probability density functions or unit-area bell curves. The bell-curve parameter  $\alpha$  ranges through the interval  $(0,2]$  and gives the Gaussian bell curve when  $\alpha = 2$  and gives the Cauchy bell curve when  $\alpha = 1$ . The impulsiveness grows as  $\alpha$  falls and the bell curves have thicker tails. Only the Gaussian statistics have finite variances or finite higher moments. An additive fuzzy system can learn ellipsoidal fuzzy rule patches from a new pseudo-covariation matrix or measure of alpha-stable covariation. Mahalanobis distance gives a joint set function for the learned if-part fuzzy sets of the if-then rules. The joint set function preserves input correlations that factored set functions ignore. Competitive learning tunes the local means and pseudo-covariations of the alpha-stable statistics and thus tunes the fuzzy rules. Then the covariation rules can both predict nonlinear signals in impulsive noise and filter the impulsive noise in time-series data. The fuzzy system filtered such noise better than did a benchmark radial basis neural network.

---

### 1. Filtering impulsive noise

Impulsive noise is not Gaussian. The bell-curve density of impulsive noise has thicker tails than a Gaussian bell curve has. The thicker tails give rise to more frequent bursts of noise. The Cauchy density  $f(x) = 1/\pi(1+x^2)$  has this property and so do all *alpha-stable* probability densities [17] for index parameter  $\alpha$  in  $0 < \alpha < 2$ . The Cauchy density is the special case when  $\alpha = 1$ . The thicker polynomial tails give an infinite variance and give infinite higher-order moments. The *lower-order* fractional moments are still finite.

The Gaussian density is the special case of an alpha-stable density when  $\alpha = 2$ . It is unique in this family because it has exponential tails and has finite variance and higher-order moments. Fig. 1 shows how alpha-stable noise grows more impulsive as the index  $\alpha$  falls from 2 to 1.

The Gaussian bell curve is by far the most common bell curve in science and engineering. It fits some data well and depends on just the first and second moments of a process. This gives simple closed-form solutions to many problems. Limiting sums of independent random variables converge to a standard Gaussian variable if the variables have finite variance. The key term  $e^{-x^2}$  is in effect invariant under linear transforms

---

\* Corresponding author.

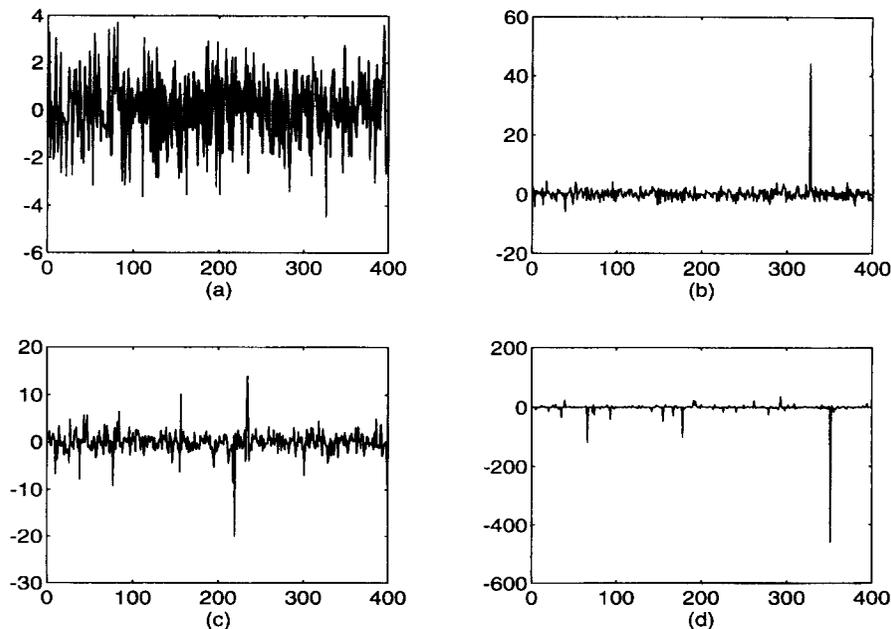


Fig. 1. Impulsive noise modeled as independent samples from alpha-stable bell-curve densities. The noise magnitudes are not to scale. A noise density with a thicker tail has more frequent bursts of high energy. (a) Classical Gaussian case when  $\alpha = 2$ . Then the noise has finite variance. (b) Impulsive noise when  $\alpha = 1.9$  and the noise has infinite variance. (c) More impulsiveness when  $\alpha = 1.5$ . (d) Still more impulsiveness in the Cauchy case when  $\alpha = 1$ .

and appears in the structure of Brownian motion. The Gaussian bell curve is the maximum-entropy density when we know just the first two (finite) moments of a process [16]. And it minimizes the “uncertainty principle” of physics and signal processing [7].

But locally a Gaussian curve looks like a Cauchy curve or other alpha-stable bell curve. Both have infinitely long tails and so are not “realistic”. The tails have nonzero mass and so both can emit or explain a noise impulse of any magnitude. And both involve a central limit theorem [3]: Limiting sums of independent alpha-stable variables converge in distribution to an alpha-stable random variable. Indeed sums of independent identically distributed random variables can converge *only* to an alpha-stable random variable. The popular Gaussian central limit theorem is the special case of this alpha-stable theorem when  $\alpha = 2$ . The real difference is that researchers first found and applied the Gaussian curve and that alpha-stable curves with  $\alpha < 2$  better model impulsive noise as found in telephone lines [19], underwater acoustics, low-frequency atmospheric signals [13] and lightning flashes, and fluctuations in gravitational fields [20] and stock prices [12] and in many other processes. Alpha-stable models seem to work well [17] when the signal data contains outliers. We seek the best alpha-stable fit to a measured noise pattern. In general this best-fit value  $\alpha^*$  will not pick out the Gaussian case and so  $\alpha^* < 2$  will hold [18].

The main problem with alpha-stable models is their math complexity. Alpha-stable statistics do not have covariance matrices. Joint alpha-stable variables do have a measure of *covariation* that acts much like a covariance. But the covariation is not symmetric.

Below we define a pseudo-covariation measure that is symmetric and then use competitive learning to update it. This new measure gives a positive definite matrix and that gives a fuzzy ellipsoidal rule [5, 6]. We also use a new way to project the rule ellipsoids onto joint multivariable fuzzy sets. This helps preserve correlations in the input data. We apply this method to predicting and filtering signals awash in impulsive noise.

## 2. Alpha-stable noise: covariation versus covariance

This section reviews the family of alpha-stable probability densities and the measure of covariation that relates alpha-stable statistics. We then define a symmetric pseudo-covariation measure and show that it acts in the limit much as the asymmetric covariation measure acts in the limit.

### 2.1. Stable distribution and its properties

An alpha-stable probability density  $f$  has the characteristic function [1, 2, 17]  $\varphi$ :

$$\varphi(t) = \exp \left[ iat - \gamma |t|^\alpha \left( 1 + i\beta \operatorname{sign}(t) \tan \frac{\alpha\pi}{2} \right) \right] \quad \text{for } \alpha \neq 1 \quad (1)$$

and

$$\varphi(t) = \exp[iat - \gamma |t|(1 + 2i\beta \ln |t| \operatorname{sign}(t)/\pi)] \quad \text{for } \alpha = 1, \quad (2)$$

where

$$\operatorname{sign}(t) = \begin{cases} 1 & \text{if } t > 0 \\ 0 & \text{if } t = 0 \\ -1 & \text{if } t < 0 \end{cases} \quad (3)$$

and  $i = \sqrt{-1}$ ,  $0 < \alpha \leq 2$ ,  $-1 \leq \beta \leq 1$ , and  $\gamma > 0$ .  $\alpha$  is the characteristic exponent parameter. An alpha-stable density with  $\alpha < 2$  has finite moments only of order less than  $\alpha$  [17]. The variance of an alpha-stable density distribution with  $\alpha < 2$  does not exist.  $a$  is a location parameter. The location parameter  $a$  is the mean of the density when  $\alpha > 1$ .  $\beta$  is a skewness parameter. The density is symmetric about  $a$  when  $\beta = 0$ . The dispersion parameter  $\gamma$  acts like a variance.  $\gamma$  is half of the Gaussian variance when  $\alpha = 2$ .

The alpha-stable characteristic function in (1) reduces to

$$\varphi(t) = e^{iat - \gamma |t|^2} \quad (4)$$

if  $\alpha = 2$  and  $\beta = 0$ . Eq. (4) is the characteristic function of the Gaussian density with mean  $a$  and variance  $2\gamma$ . So (4) gives the standard normal  $N(0,1)$  if  $a = 0$  and  $\gamma = \frac{1}{2}$ . The characteristic function in (2) reduces to

$$\varphi(t) = e^{-\gamma |t|} \quad (5)$$

if  $a = 0$ ,  $\alpha = 1$ , and  $\beta = 0$ . The exponential (5) is the characteristic function of the Cauchy density function  $f(x) = \gamma/\pi(\gamma^2 + x^2)$ . Appendix A proves the following “stability” theorem [17] that says alpha-stable statistics are closed or “stable” under linear combination. The theorem holds for Cauchy variables when  $\alpha = 1$  but requires the symmetry condition  $\beta = 0$ .

**Theorem 1.** *Suppose  $X_1, X_2, \dots, X_n$  are independent random variables and have an alpha-stable distribution with the same parameters  $(\alpha, \beta)$ . Then all linear combinations of the form  $\sum_{i=1}^n a_i X_i$  are alpha-stable with the parameter  $(\alpha, \sum_i |a_i|^\alpha \operatorname{sign}(a_i) \beta / \sum_i |a_i|^\alpha)$  if  $\alpha \neq 1$ .*

### 2.2. Covariation and pseudo-covariation

Local means and covariances define ellipsoidal fuzzy rules in [5, 6]. But alpha-stable random variables do not have a covariance if  $\alpha < 2$ . They do have a measure of covariation.

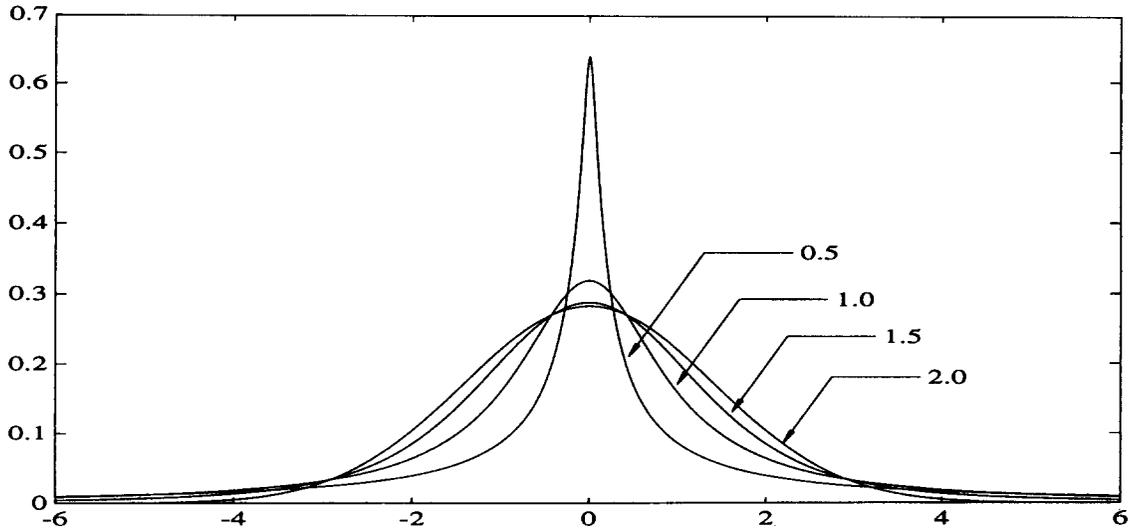


Fig. 2. Alpha-stable probability densities. Only the Gaussian bell curve ( $\alpha = 2$ ) has exponential tails and finite variation.

Let  $X$  and  $Y$  be symmetric alpha-stable random variables with  $1 < \alpha \leq 2$ , zero location parameters  $a_x$  and  $a_y$ , and dispersions  $\gamma_x$  and  $\gamma_y$ . The *covariation* of  $X$  with  $Y$  is [17]

$$(X, Y)_\alpha = \frac{E(X \text{ sign}(Y))}{E(|Y|)} \gamma_y. \tag{6}$$

Then

$$(X, X)_\alpha = \gamma_x \quad \text{and} \quad (Y, Y)_\alpha = \gamma_y. \tag{7}$$

So the covariations are not symmetric. The dispersion  $\gamma$  is also hard to estimate. Shao and Nikias [17] use three methods to estimate  $\gamma$ . These methods are computationally heavy and do not favor competitive learning. We can replace the dispersion  $\gamma$  with lower-order fractional moments [17] from Theorem 2. The proof [21, 14] of Theorem 2 lies beyond the scope of this paper.

**Theorem 2.** Let  $X$  be a symmetric alpha-stable random variable with dispersion  $\gamma$  and location parameter  $a = 0$ . Then

$$E(|X|^p) = C(p, \alpha) \gamma^{p/\alpha} \quad \text{for } 0 < p < \alpha, \tag{8}$$

where  $C(p, \alpha)$  is a constant that depends only on  $\alpha$  and  $p$ :

$$C(p, \alpha) = \frac{2^{p+1} \Gamma((p+1)/2) \Gamma(-p/\alpha)}{\alpha \sqrt{\pi} \Gamma(-p/2)} \tag{9}$$

for the gamma function  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ . Then

$$\gamma = C_1 E(|X|^p) \quad \text{as } p \rightarrow \alpha \tag{10}$$

if  $C_1 = 1/C(p, \alpha)$ .

Now define a *pseudo-covariation* that is symmetric. The pseudo-covariation of  $X$  with  $Y$  is

$$(X, Y)_{\alpha, \text{pseudo}} = E(\text{sign}(XY)|X|^{p/2}|Y|^{p/2}) \quad \text{for } 0 < p < \alpha. \quad (11)$$

Then (10) implies

$$(X, X)_{\alpha} = C_1(X, X)_{\alpha, \text{pseudo}} \quad \text{as } p \rightarrow \alpha. \quad (12)$$

We use lower-order fractional moments instead of dispersion for competitive learning. We use the same power for each random variable to make the pseudo-covariation symmetric. Note that the pseudo-covariation equals the covariance when  $p = 2$ . Below we use  $p = 1$  to learn fuzzy rules that can filter Cauchy-level impulsive noise.

We can define a pseudo-covariation matrix  $C_X$  for  $X = (X_1, \dots, X_n)^T$  that acts as a symmetric and positive-definite covariance matrix:

$$C_X = \begin{bmatrix} (X_1, X_1)_{\alpha, \text{pseudo}} & \cdots & (X_1, X_n)_{\alpha, \text{pseudo}} \\ (X_2, X_1)_{\alpha, \text{pseudo}} & \cdots & (X_2, X_n)_{\alpha, \text{pseudo}} \\ \cdots & \cdots & \cdots \\ (X_n, X_1)_{\alpha, \text{pseudo}} & \cdots & (X_n, X_n)_{\alpha, \text{pseudo}} \end{bmatrix}. \quad (13)$$

Appendix C presents the competitive learning system [8] to estimate the local covariation matrix. Then the local mean and covariation matrix define hyperellipsoids in the input–output space. We use Mahalanobis distance to produce joint if-part fuzzy sets from the ellipsoidal covariation rules. Section 6 tests how well these covariation rules filter alpha-stable impulsive noise.

### 3. Additive fuzzy systems

This section reviews the standard additive model (SAM) of the fuzzy system  $F : R^n \rightarrow R$  that we use to predict and filter nonlinear signals in alpha-stable noise. The pseudo-covariation measure (11)–(12) acts as standard covariance and yields ellipsoidal fuzzy rules. We use a new scheme based on Mahalanobis distance to project the ellipsoidal rule onto joint if-part fuzzy sets that maintain the correlation structure among the inputs. Other projection schemes factor the input sets and ignore the correlation.

#### 3.1. Additive fuzzy systems

A fuzzy system  $F : R^n \rightarrow R$  stores  $m$  rules of the word from “If  $X = A_j$ , then  $Y = B_j$ ” or the patch form  $A_j \times B_j \subset X \times Y = R^n \times R$ . The if-part fuzzy sets  $A_j \subset R^n$  and then-part fuzzy sets  $B_j \subset R$  have set functions  $a_j : R^n \rightarrow [0, 1]$  and  $b_j : R \rightarrow [0, 1]$ . The system can use the joint set function  $a_j$  or some factored form such as  $a_j(x) = a_j^1(x_1) \cdots a_j^n(x_n)$  or  $a_j(x) = \min(a_j^1(x_1), \dots, a_j^n(x_n))$  or any other conjunctive form for input vector  $x = (x_1, \dots, x_n) \in R^n$ .

An additive fuzzy system [8, 9] sums the “fired” then part sets  $B_j'$ :

$$B = \sum_{j=1}^m B_j' = \sum_{j=1}^m a_j(x) B_j. \quad (14)$$

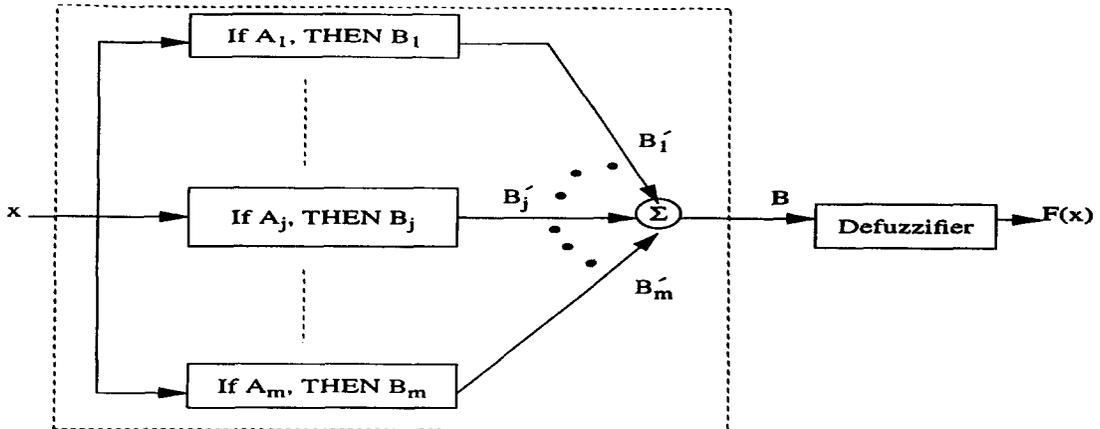


Fig. 3. Architecture of an additive fuzzy system  $F : R^n \rightarrow R^p$  with  $m$  rules. Each input  $x_0 \in R^n$  enters the system  $F$  as a numerical vector. At the set level  $x_0$  acts as a delta pulse  $\delta(x - x_0)$  that combs the if-part fuzzy sets  $A_j$  and gives the  $m$  set values  $a_j(x_0)$ :  $a_j(x_0) = \int_{R^n} \delta(x - x_0) a_j(x) dx$ . The set values “fire” the then-part fuzzy sets  $B_j$  to give  $B'_j$ . A standard additive model (SAM) scales each  $B_j$  with  $a_j(x)$ . The system then sums the  $B'_j$  sets to give  $B$ . The system output  $F(x_0)$  is the centroid of  $B$ .

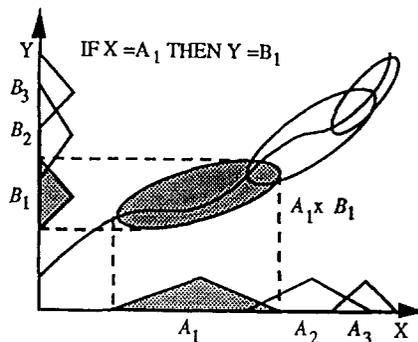


Fig. 4. Each fuzzy rule defines a patch or fuzzy subset of the input–output state space. The fuzzy system approximates a function as it covers its graph with rule patches. Lone optimal rule patches cover extrema.

Fig. 3 shows the parallel fire-and-sum structure of the SAM system. These systems can uniformly approximate any continuous (or bounded measurable) function  $f$  on a compact domain [9].

Fig. 4 shows how three rule patches can cover part of the graph of a scalar function  $f : R \rightarrow R$ . The patch cover shows that all fuzzy systems  $F : R^n \rightarrow R^p$  suffer from *rule explosion* in high dimensions. A fuzzy system  $F$  needs of the order of  $k^{n+p-1}$  rules to cover the graph and thus to approximate a vector function  $f : R^n \rightarrow R^p$ . Optimal rules can help deal with the exponential rule explosion. Lone or local mean-squared optimal rule patches cover the extrema of the approximand  $f$  [10]. They “patch the bumps”. Better learning schemes move rule patches to or near extrema and then fill in between extrema with extra rule patches if the rule budget allows.

The scaling choice  $B'_j = a_j(x)B_j$  gives a *standard additive model* or SAM. Appendix B shows that taking the centroid of  $B$  in (14) gives [8–11] the SAM ratio

$$F(x) = \frac{\sum_{j=1}^m a_j(x)V_j c_j}{\sum_{j=1}^m a_j(x)V_j}. \tag{15}$$

$V_j$  is the nonzero volume or area of then-part set  $B_j$ .  $c_j$  is the centroid of  $B_j$  or its center of mass.

The SAM theorem (15) implies that the fuzzy structure of the then-part sets  $B_j$  does not matter. The ratio depends on just the volume  $V_j$  and location  $c_j$  of the then-part sets  $B_j$ . Our SAM has then-part sets of the same area and so the volume terms  $V_j$  cancel from (15). We need to pick only the scalar centers  $c_j$ . The projection method in the next section picks the multivariable if-part sets  $a_j$ .

### 3.2. Covariation rules and Mahalanobis sets: product space clustering and projection

Fuzzy rule patches  $A_j \times B_j$  can take the form of ellipsoids [5, 6]. Input–output data  $(x_i, y_i)$  drives the competitive learning process in Appendix C. We form the concatenated vector  $z_i^T = [x_i^T, y_i]$  in the product space  $R^n \times R$ . Then we lose  $p$  quantization vectors  $m_j$  on the same product space. These vectors learn or adapt as they code for the local sample statistics and give rise to an *adaptive vector quantization* (AVQ). The AVQ vectors  $m_j$  act as the synaptic fan-in columns of a connection matrix in a neural network with  $p$  competing neurons. The points  $m_j$  learn if and only if they move in the product space. The points  $m_j$  track the distribution of the incoming data points  $z_i$  and tend to be sparse or dense where the  $z_i$  are sparse or dense.

Fuzzy rules form through AVQ *product space clustering* [8]. Each AVQ vector  $m_j$  converges to the local centroid of the sample data  $z_i$ . So the AVQ vectors  $m_j$  estimate the local first-order moments of some unknown probability density  $p(z)$  that generates the random sample data  $z_i$ . The AVQ algorithm is a form of nearest-neighbor pattern matching or  $K$ -means clustering. The  $m_j$  are random estimators since the data  $z_i$  are random. The AVQ point  $m_j$  hops about more in the presence of noisy or sparse data and thus has a larger “error” ball. The covariance matrix  $K_j$  measures this error ball. The same competitive AVQ algorithm in Appendix C updates the positive-definite matrix  $K_j$ . The inverse matrix  $K_j^{-1}$  defines an ellipsoid  $E$  in the product space  $R^n \times R$  as the locus of all points  $z$  that obey

$$e^2 = (z - m_j)^T K_j^{-1} (z - m_j) \tag{16}$$

for centering constant  $e > 0$ . Then the  $j$ th ellipsoid defines the  $j$ th rule in the additive fuzzy system  $F : R^n \rightarrow R$ . We use the pseudo-covariation matrix in (11) with  $p = 1$  to grow ellipsoidal *covariation rules*.

Fuzzy sets arise when the ellipsoids project down onto the axes. The projection is not unique because an ellipsoid  $E \subset R^n \times R$  does not factor into the Cartesian product of  $n + 1$  sets  $E_1 \times \dots \times E_n \times E_{n+1}$ . The SAM theorem in (15) lets us replace the then-part set  $E_{n+1}$  (or  $B_j$ ) with just an area or volume  $V_j$  and a centroid  $c_j$ . This presents a problem since noisy rules will have larger  $V_j$  values than smaller and more certain rule patches have. Then the SAM ratio (15) would give more weight to the least certain rules. Dickerson and Kosko [5, 6] deal with this by weighting each rule in (14) and (15) with a weight  $w_j$  that varies with the inverse of  $V_j$  or with its square. We deal with it by normalizing all then-part sets to have the same area or volume. Then  $V_1 = \dots = V_m > 0$ . This ignores some of the statistical structure of  $E$  but simplifies both the SAM ratio (15) for computing  $F(x)$ . It also finds each then-part set  $B_j$  as just the centroid  $c_j$  of the projection of the  $j$ th ellipsoid onto the output axis  $R$ .

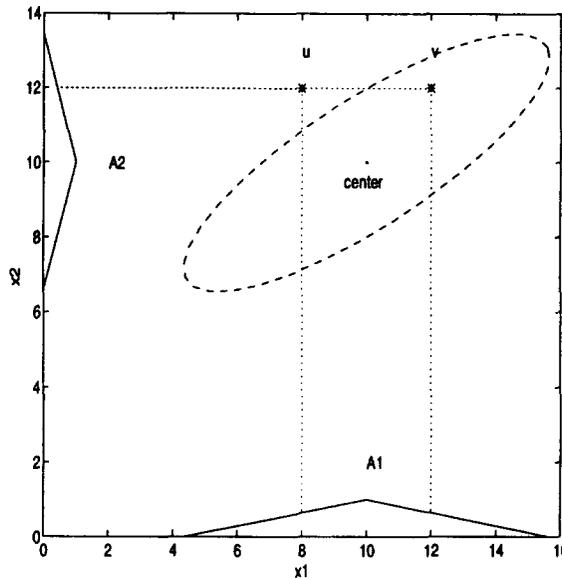


Fig. 5. Triangular fuzzy sets from ellipsoidal projection onto input axes. The vectors  $u$  and  $v$  have components that belong to the same degree to the scalar fuzzy sets  $A^1$  and  $A^2$ . Such projection ignores correlation among the input components.

The covariance method [5, 6] projects the ellipsoid  $E$  onto the  $n$  input axes to give triangular fuzzy sets. Fig. 5 shows how a 2-D ellipsoid projects onto the  $x_1$  and  $x_2$  axes to give two triangles. The two vector points  $u$  and  $v$  lie the same distance from the center of the ellipsoid. So their components belong to the same degree to the fuzzy triangular sets  $A^1$  and  $A^2$ :  $a^1(u_1) = a^1(v_1)$  and  $a^2(u_2) = a^2(v_2)$ . We lose the correlation structure among the input components with a factored set function  $a(u) = \min(a^1(u_1), a^2(u_2)) = a(v)$ . The same holds for pairwise multiplication or other pairwise conjunctive operators.

We propose a metrical set function  $a(x)$  that preserves the correlations among the input components and that exploits the second-order statistics of the sample data  $z$ . The form of the ellipsoids in (16) suggests the squared Mahalanobis distance.

$$d(x)^2 = (x - m_j^i)^T (K_j^i)^{-1} (x - m_j^i) \tag{17}$$

as the basis for a similarity set function  $a(x)$ . The  $n$ -vector  $m_j^i$  in (17) stands for the first  $n$  components of the quantization vector  $m_j$  in the input–output product space  $R^n \times R$  with covariance submatrix  $K_j^i$ . Each  $a_j$  corresponds to just one  $m_j$  and vice versa.  $d(x)$  is the covariance-scaled distance from input vector  $x$  to local mean  $m_j^i$ .

We define the joint if-part set function  $a_j$  as the Mahalanobis similarity measure

$$a_j(x) = \begin{cases} 1 - d(x)/q & \text{if } d(x) \leq q \\ 0 & \text{if } d(x) > q \end{cases} \tag{18}$$

for some normalizing factor  $q > 0$ . We used  $q=2$ . Fig. 6 shows ellipsoidal contours of equal Mahalanobis distance. Then  $d(u) > d(v)$  and the vector components maintain their correlation structure. So  $a(u) < a(v)$  as the figure makes clear:  $u$  has less set value than  $v$  has in the concentric interior ellipsoids. Fig. 7 shows the fuzzy structure of the joint if-part Mahalanobis set function  $a$ . The cone structure shows the correlation among the components  $a^1$  and  $a^2$  of the if-part set function  $a$ .

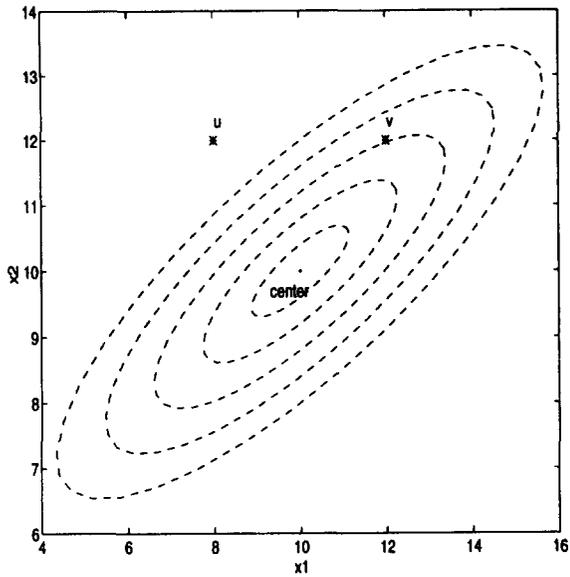


Fig. 6. Ellipsoidal contours of equal Mahalanobis distance. The point  $v$  is closer to the center in Mahalanobis distance than point  $u$  is. Both points have the same Euclidean distance to the center when the covariance matrix  $K$  is the identity matrix  $I$ .

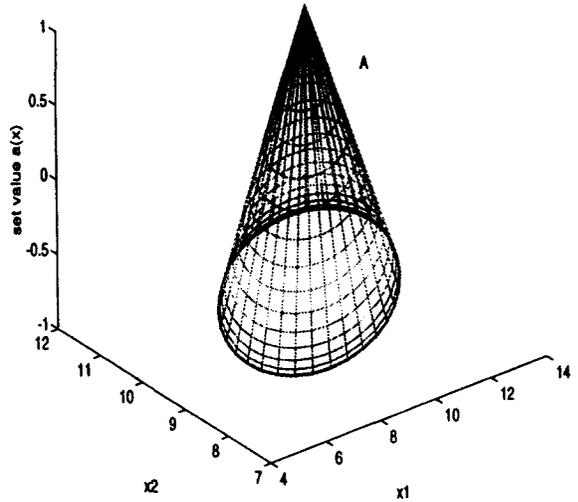


Fig. 7. Joint input fuzzy set  $A$  that arises from the ellipsoidal Mahalanobis isometric contours in Fig. 6. The conical fuzzy-set patch  $A$  does not factor into a Cartesian product of two uncorrelated fuzzy sets  $A^1$  and  $A^2$ .

#### 4. Fuzzy prediction with covariation rules

This section shows how a SAM fuzzy system  $F : R^n \rightarrow R$  with pseudo-covariation rules from Section 3 predicts a nonlinear function in the presence of impulsive alpha-stable noise. The test function was a sine wave sampled at 12 points for each period as in Fig. 8.

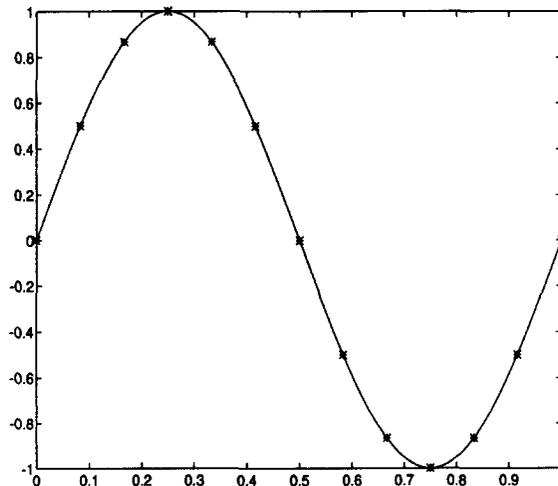


Fig. 8. Sampling points of approximand  $f(x) = \sin 2\pi x$ .

The goal was to predict the next sample value  $y(k)$  based on the past two sample differences  $y(k-1) - y(k-2)$  and  $y(k-2) - y(k-3)$ . Additive measurement noise  $n(k)$  corrupted each input to the fuzzy system. The noise  $n(k)$  came from pseudo-random numbers for the alpha-stable values  $\alpha = 2$ ,  $\alpha = 1.9$ ,  $\alpha = 1.5$ , and  $\alpha = 1$ .

The noise  $n(k)$  had finite variance for only the Gaussian case  $\alpha = 2$ . So we used the finite unbiased sample variance

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (19)$$

with sample mean  $\bar{x} = (1/n) \sum_{i=1}^n x_i$ . The fuzzy system  $F$  gave the predicted value

$$\hat{y}(k) = F(x) = F(x_1(k), x_2(k)) \quad (20)$$

for input components  $x_1(k) = y(k-2) - y(k-3) + n(k)$  and  $x_2(k) = y(k-1) - y(k-2) + n(k)$ .

We created 1000 data pairs and used the first 500 as training data and the second 500 as test data. The first 12 data pairs came from the first 12 uniform samples. The second 12 data pairs came from the second sinusoidal period and so on. We used  $p = 2$  in the pseudo-covariation measure (11) to simplify the competitive learning process. There were 12 clusters since there were 12 sample points for each period. Fig. 9 shows the resulting ellipsoidal covariation rules for the Gaussian case. Fig. 10 shows the projected Mahalanobis sets.

We compared the pseudo-covariation/Mahalanobis system with three other systems. Four types of alpha-stable noise corrupted each system. The first system was a SAM fuzzy system that projects the ellipsoidal rules onto axis triangles [5, 6]. Thus this “standard” system used  $p = 2$  in (11) to give a covariance matrix. This system ignored all input correlations and did much poorer than did the Mahalanobis system. The second system was a radial basis function (RBF) network with the Gaussian kernel [4, 15]

$$N(x) = \sum_{j=1}^m w_j e^{-\sum_{i=1}^n (x_i - c_j^i)^2 / \sigma^2} \quad (21)$$

and with mean or center vector  $c_j = (c_j^1, \dots, c_j^n)$  and variance or spread  $\sigma^2$ . We used the AVQ cluster centers to center the Gaussian balls:  $c_j = m_j$ . We chose the RBF weights  $w_j$  as the best linear weights in the sense of a least-squares data fit [4]. The weights  $w_j$  solved the normal equation  $Q^T Q \bar{w} = Q^T b$  where  $b$  is the desired output and  $Q$  is an  $m$ -by- $n$  matrix. Here  $m = 500$  is the number of training data and  $n = 12$  is the number of weights.  $Q^T Q$  is invertible if the columns of  $Q$  are linearly independent. Then  $\bar{w} = (Q^T Q)^{-1} Q^T b$ . This greatly increased the computational load of tuning the RBF net but increased its predictive accuracy.

The RBF net used unweighted Euclidean distance instead of Mahalanobis distance. It gave less weight to data correlations and predicted less poorly than did the Mahalanobis fuzzy system for all noise but Cauchy noise. We modified the RBF net in (21) to extend the quadratic form in the exponent from  $(x - c_j)^T (x - c_j)$  to  $(x - c_j)^T C_j^{-1} (x - c_j)$ . This made the RBF net competitive with the fuzzy Mahalanobis system.

All four models predicted poorly in Cauchy noise. We ran the test 10 times for distinct data sets and averaged the mean-squared errors (MSEs) of the prediction. All the predictors degraded as the noise impulsiveness grew. The Mahalanobis system did best when the input data were most correlated. Table 1 shows the results.

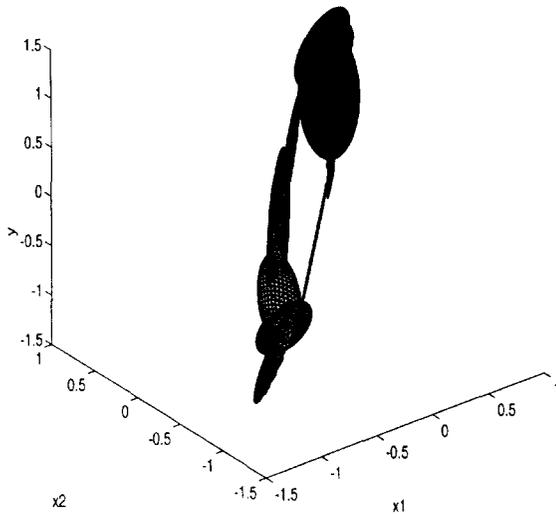


Fig. 9. Ellipsoidal pseudo-covariation rules for the noisy fuzzy predictor. The noise in this case was additive Gaussian noise.

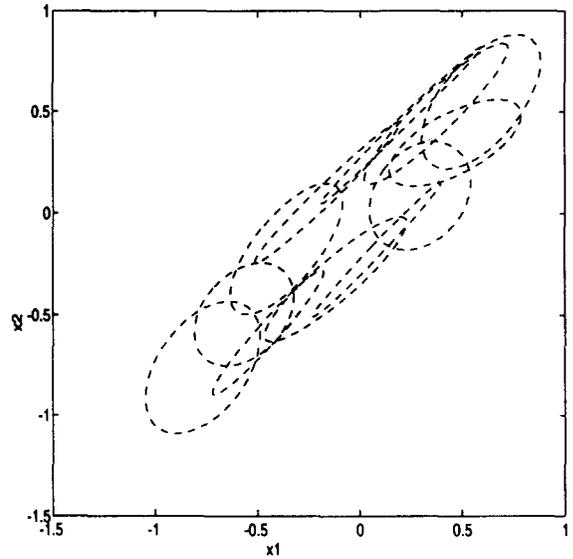


Fig. 10. Mahalanobis joint fuzzy sets of the covariation rules in Fig. 9.

Table 1

Mean-squared error of prediction for the pseudo-covariation/Mahalanobis, standard ellipsoidal, and radial basis function (RBF) predictors.

$\alpha$	$\hat{\sigma}^2$	Ellipsoidal fuzzy system	Covariation fuzzy system	RBF net	Mahalanobis RBF net
2.0	0.0197	0.0731	0.0305	0.0499	0.0308
1.9	0.0282	0.0606	0.0395	0.0504	0.0313
1.5	0.1932	0.2248	0.1503	0.1518	0.1357
1.0	2.7585	0.3967	0.3088	0.3069	0.3191

### 5. Fuzzy filters with covariation rules

This section tests the fuzzy Mahalanobis system as a filter or noise canceller. The test signal is the standard fifth-degree polynomial from [5, 6]

$$f(x) = 3x(x - 1)(x - 1.9)(x + 0.7)(x + 1.8). \tag{22}$$

The same four types of additive noise  $n(k)$  in Section 4 corrupted the polynomial signal as Fig. 11 shows. The goal was to cancel the overall noise in the sense of mean-squared error.

We sampled the noisy data pairs  $z_i$  in the planar product space. Competitive learning grew covariance ellipsoids around the  $p$  quantization vectors  $m_j$  as in Sections 3 and 4. We compared the 10 covariation rules with the 10 covariance rules [5, 6] and with a 10-radial-basis-function network. Fig. 12 shows the ellipsoidal

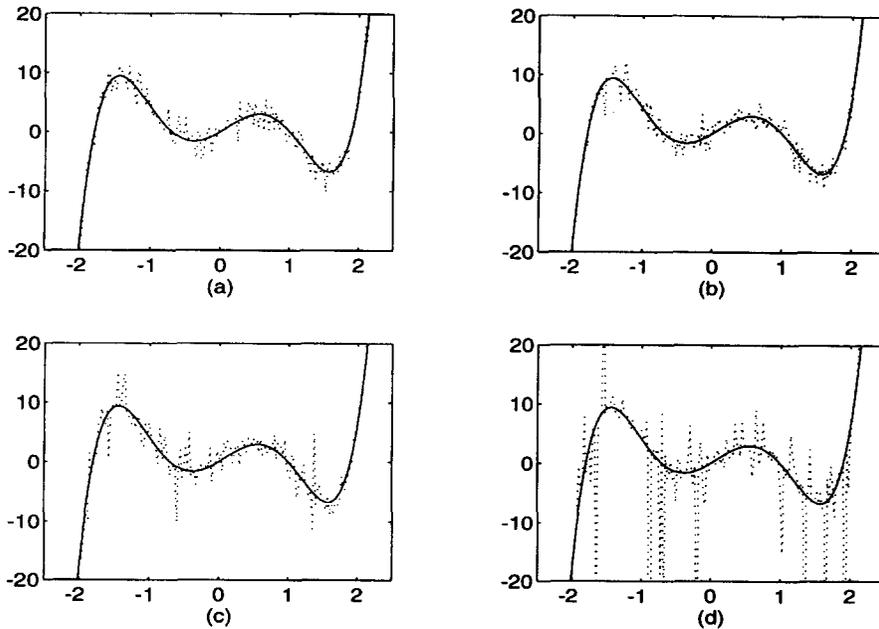


Fig. 11. Four types of alpha-stable impulsive noise corrupt a polynomial signal: (a) Gaussian noise when  $\alpha = 2$ ; (b) more impulsive noise when  $\alpha = 1.9$ ; (c) still more impulsive noise when  $\alpha = 1.5$ ; (d) Cauchy noise when  $\alpha = 1$ .

rules from the covariance matrices. These rule patches cover the signal polynomial in Fig. 13. Fig. 14 shows the ellipsoidal rules from the covariation matrices. These rule patches cover the same signal polynomial in Fig. 15. The impulsive Cauchy noise ( $\alpha = 1$ ) moves the covariation ellipsoids further apart from one another than it moves the covariance ellipsoids. The covariation rules gave a better filtering patch cover in terms of mean-squared error. The rules in Figs. 12 and 14 grew from 500 training data with Cauchy noise. We used the other 500 sample points as test data.

The covariation rule system filtered noise better than did both the covariance rule system and the RBF net in terms of mean-squared error. Fig. 16 compares the filtered noise from the covariance rule system with the noiseless polynomial signal. Fig. 17 compares the filtered noise from the covariation rule system with the noiseless signal. Fig. 18 compares the RBF filtered noise with the noiseless signal. The scalar input values did not give rise to an input covariance matrix. So there was no RBF net with Mahalanobis distance as there was for the prediction problem in Section 4.

Figs. 19–21 plot the power spectra of the three filters. The high frequency terms in Fig. 19 came from the separation among the covariation ellipsoids and thus came from the impulsive noise itself. We then used the smoothing-by-3s system

$$y(k) = \frac{1}{3}(x(k-1) + x(k) + x(k+1)) \quad (23)$$

to smooth the high-frequency components in the covariation rule system. This greatly improved how the filter canceled the four types of alpha-stable noise as in Figs. 22 and 23. Table 2 compares the four filters in terms of mean-squared error.

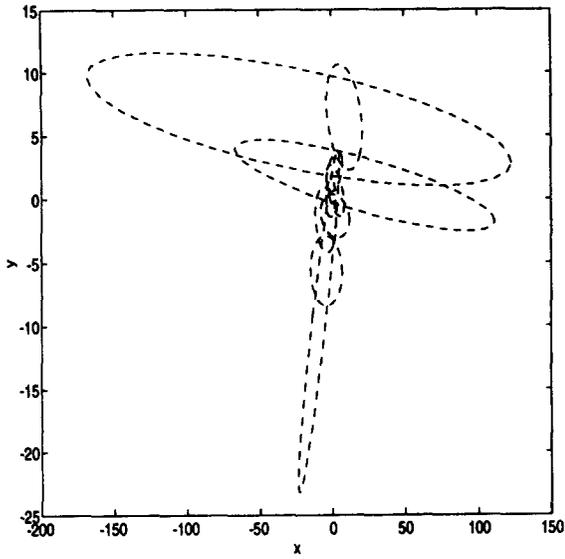


Fig. 12. Ellipsoidal rules from covariance matrices and Cauchy noise.

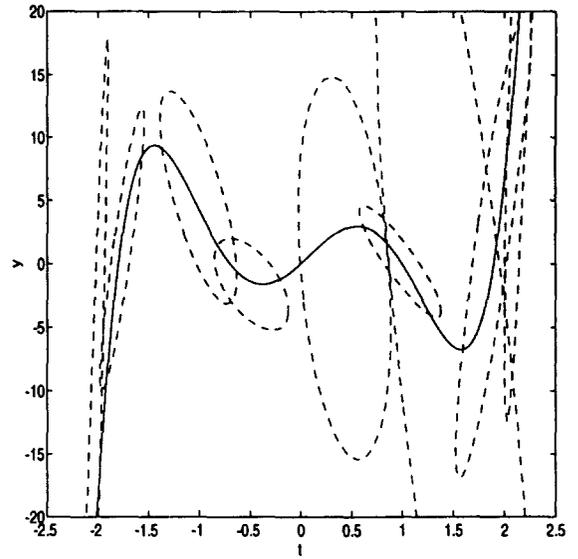


Fig. 13. Patch cover of covariance rules.

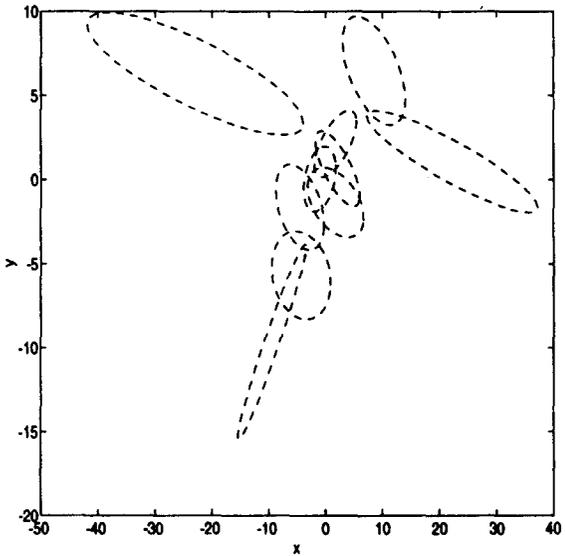


Fig. 14. Ellipsoidal rules from pseudo-covariation matrices and Cauchy noise.

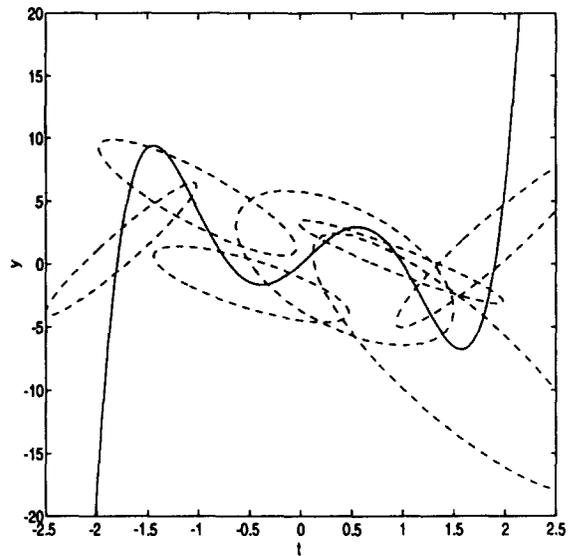


Fig. 15. Patch cover of pseudo-covariation rules.

## 6. Conclusion

Dealing with impulsive noise remains one of the great challenges of modern engineering. It is hard to model, predict, and filter—and yet it pervades the world. Fuzzy systems omit the modeling step that burdens

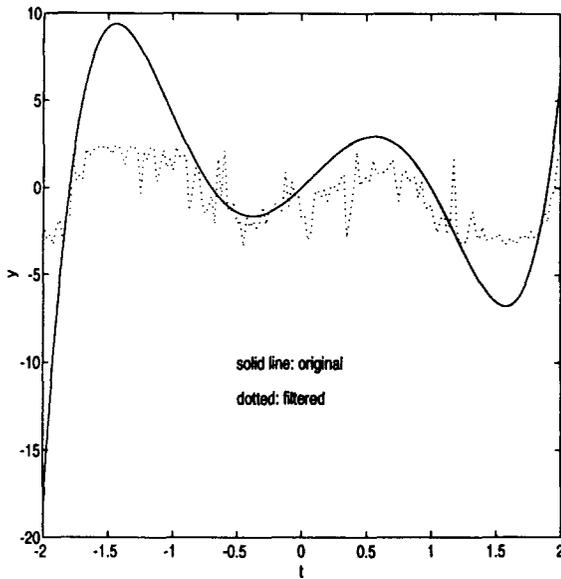


Fig. 16. Covariance fuzzy filter of Cauchy noise.

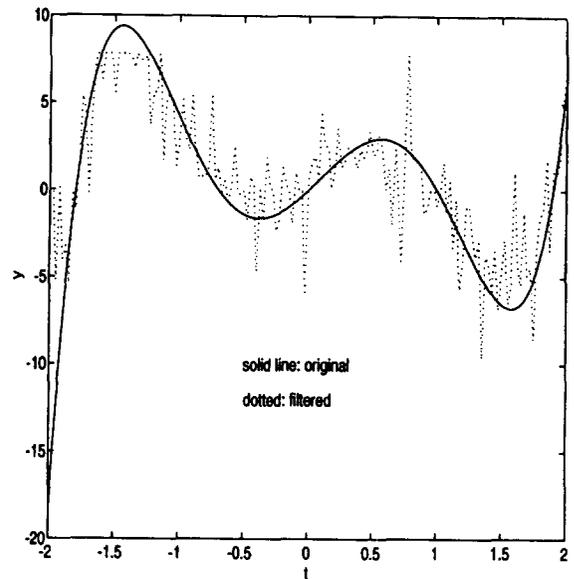


Fig. 17. Pseudo-covariance fuzzy filter of Cauchy noise.

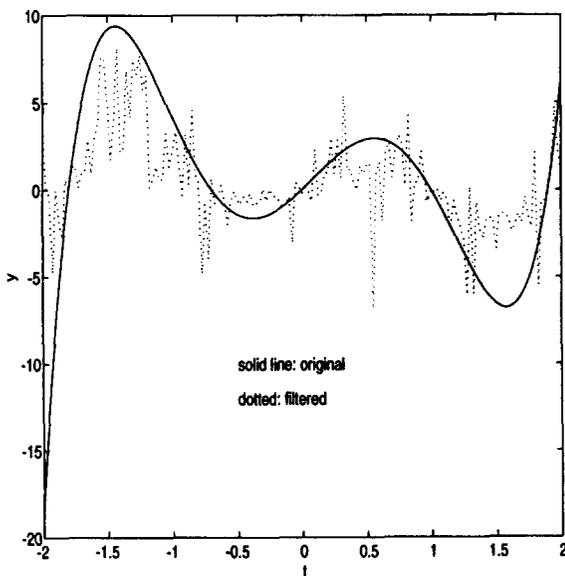


Fig. 18. Radial basis function filter of Cauchy noise.

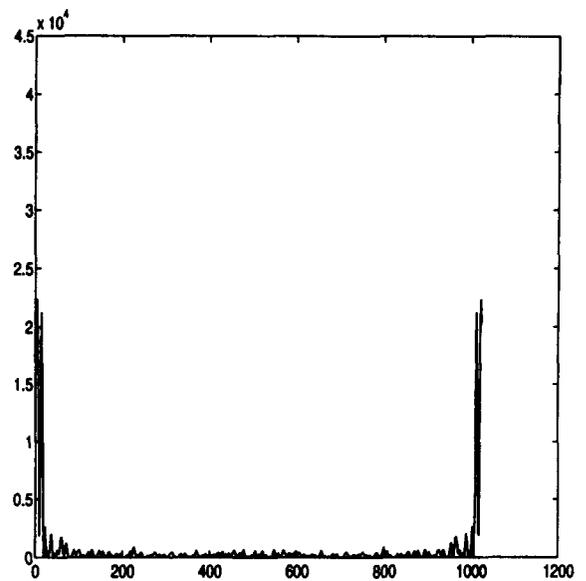


Fig. 19. Power spectrum of covariance-filtered signal.

most standard system. They also dispense with the old working fiction that real noise is Gaussian noise of finite variance. Adaptive fuzzy systems can grow and tune their rules from any source of signals or noise.

The ellipsoidal rule structure constrained our fuzzy systems. This need not be the case. Rules with highly irregular or impulsive shapes may better predict and filter impulsive noise. Or fuzzy systems might keep the

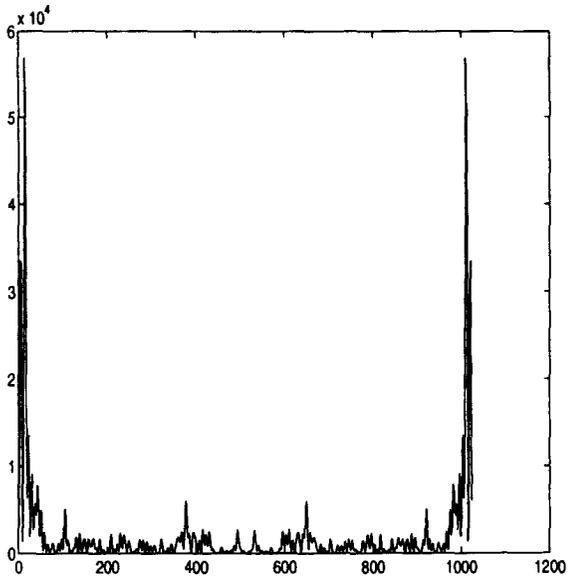


Fig. 20. Power spectrum of pseudo-covariation-filtered signal.

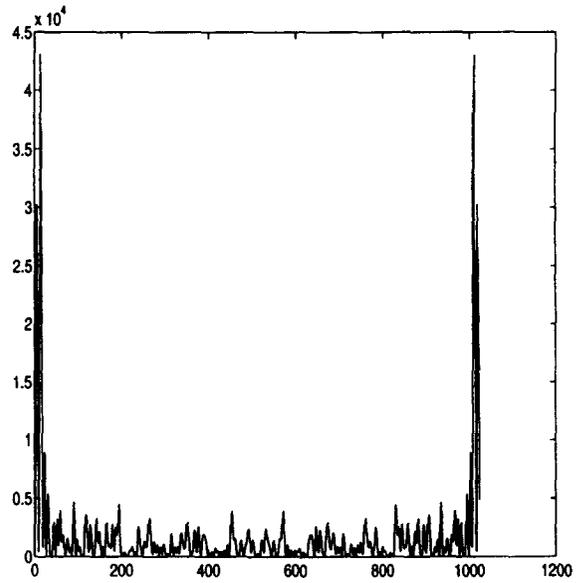


Fig. 21. Power spectrum of RBF-filtered signal.

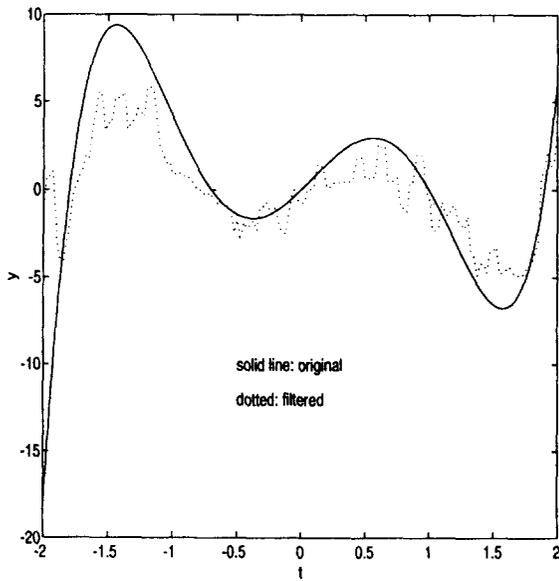


Fig. 22. Smoothed pseudo-covariation filter of Cauchy noise.

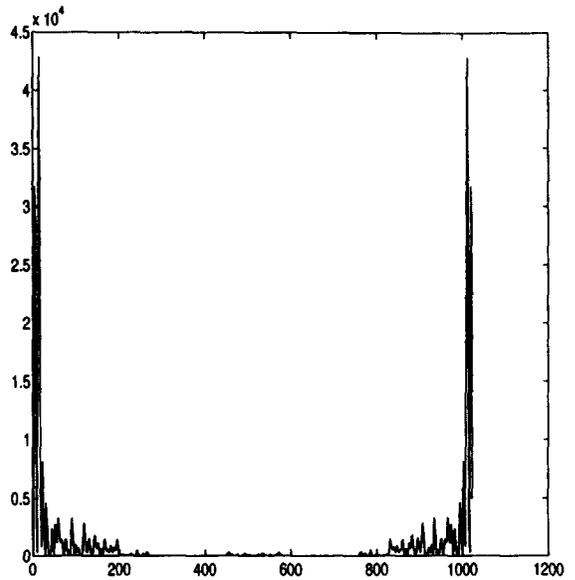


Fig. 23. Power spectrum of covariation-filtered signal with smoothing.

ellipsoidal rule structure but estimate other statistics [17] of alpha-stable random variables. The search for such rules and estimators is an open research area.

Table 2  
Mean-squared error of filtering for the pseudo-covariation/Mahalanobis, standard ellipsoidal, and radial basis function (RBF) filters

$\alpha$	$\sigma^2$	Ellipsoidal fuzzy system	Covariation fuzzy system	RBF net	Smoothed covariation fuzzy system
2.0	1.86	5.25	4.33	4.68	4.24
1.9	2.21	6.94	5.63	6.77	5.13
1.5	11.45	11.88	8.23	10.91	6.12
1.0	775.27	16.35	14.27	14.59	11.85

**Appendix A. Proof of “stability” theorem**

**Theorem 1.** Suppose  $X_1, X_2, \dots, X_n$  are independent random variables and have an alpha-stable distribution with the same parameters  $(\alpha, \beta)$ . Then all linear combinations of the form  $\sum_{j=1}^n a_j X_j$  are alpha-stable with the parameter  $(\alpha, \sum_{j=1}^n |a_j|^\alpha \text{sign}(a_j) \beta / \sum_{j=1}^n |a_j|^\alpha)$  if  $\alpha \neq 1$ .

**Proof.** Suppose a random variable  $X$  has a probability density function  $p_X(x)$  and characteristic function  $\varphi_X(t)$ . Define a new random variable  $Y = aX$ . Then it has the characteristic function  $\varphi_Y(t) = \varphi_X(at)$ . So the sum random variable  $\sum_{j=1}^n a_j X_j$  has the factored characteristic function  $\varphi_{X_1}(a_1 t) \varphi_{X_2}(a_2 t) \cdots \varphi_{X_n}(a_n t)$  since the random variables are independent. Then replacing  $t$  with  $a_j t$  in the alpha-stable characteristic function

$$\varphi_X(t) = \exp \left[ iat - \gamma |t|^\alpha \left( 1 + i\beta \text{sign}(t) \tan \frac{\alpha\pi}{2} \right) \right] \tag{A.1}$$

gives

$$\varphi_{\sum_{j=1}^n a_j X_j}(t) = \prod_{j=1}^n \varphi_{X_j}(a_j t) \tag{A.2}$$

$$= \prod_{j=1}^n \exp \left[ i a a_j t - \gamma |a_j t|^\alpha \left( 1 + i\beta \text{sign}(a_j t) \tan \frac{\alpha\pi}{2} \right) \right] \tag{A.3}$$

$$= \prod_{j=1}^n \exp \left[ i a a_j t - \gamma |a_j|^\alpha |t|^\alpha \left( 1 + i\beta \text{sign}(a_j) \text{sign}(t) \tan \frac{\alpha\pi}{2} \right) \right] \tag{A.4}$$

$$= \exp \left[ \sum_j \left( i a a_j t - \gamma |a_j|^\alpha |t|^\alpha \left( 1 + i\beta \text{sign}(a_j) \text{sign}(t) \tan \frac{\alpha\pi}{2} \right) \right) \right] \tag{A.5}$$

$$= \exp \left[ \sum_j \left( i a a_j t - \gamma |t|^\alpha \left( \sum_j \left( |a_j|^\alpha + |a_j|^\alpha i\beta \text{sign}(a_j) \text{sign}(t) \tan \frac{\alpha\pi}{2} \right) \right) \right) \right] \tag{A.6}$$

$$= \exp \left[ \sum_j \left( i a a_j t - \gamma |t|^\alpha \left( \sum_j (|a_j|^\alpha) + \sum_j \left( |a_j|^\alpha i\beta \text{sign}(a_j) \text{sign}(t) \tan \frac{\alpha\pi}{2} \right) \right) \right) \right] \tag{A.7}$$

$$= \exp \left[ i a \left( \sum_j a_j \right) t - \left( \sum_j |a_j|^\alpha \right) \gamma |t|^\alpha \left( 1 + i \frac{\sum_j |a_j|^\alpha \text{sign}(a_j)}{\sum_j |a_j|^\alpha} \beta \text{sign}(t) \tan \frac{\alpha\pi}{2} \right) \right]. \tag{A.8}$$

So the sum  $\sum_{j=1}^n a_j X_j$  is alpha-stable and has the parameter  $(\alpha, \sum_{j=1}^n |a_j|^\alpha \text{sign}(a_j) \beta / \sum_{j=1}^n |a_j|^\alpha)$ . □

### Appendix B. SAM Theorem

**Theorem B.1.** *Suppose the fuzzy system  $F : R^n \rightarrow R^p$  is a standard additive model:  $F(x) = \text{Centroid}(B) = \text{Centroid}(\sum_{j=1}^m a_j(x) B_j)$ . Then  $F(x)$  is a convex sum of the  $m$  then-part set centroids:*

$$F(x) = \frac{\sum_{j=1}^m a_j(x) V_j c_j}{\sum_{j=1}^m a_j(x) V_j} = \sum_{j=1}^m p_j(x) c_j. \tag{B.1}$$

The convex coefficients or discrete probability weights  $p_1(x), \dots, p_m(x)$  depend on the input  $x$  through

$$P_j(x) = \frac{a_j(x) V_j}{\sum_{k=1}^m a_k(x) V_k}. \tag{B.2}$$

$V_j$  is the finite positive volume (or area if  $p = 1$ ) and  $c_j$  is the centroid of then-part set  $B_j$ :

$$V_j = \int_{R^p} b_j(y_1, \dots, y_p) dy_1 \cdots dy_p > 0, \tag{B.3}$$

$$c_j = \frac{\int_{R^p} y b_j(y_1, \dots, y_p) dy_1 \cdots dy_p}{\int_{R^p} b_j(y_1, \dots, y_p) dy_1 \cdots dy_p}. \tag{B.4}$$

**Proof.** There is no loss of generality to prove the theorem for the scalar-output case  $p = 1$  when  $F : R^n \rightarrow R^p$ . This simplifies the notation. We need but replace the scalar integrals over  $R$  with the  $p$ -multiple or volume integrals over  $R^p$  in the proof to prove the general case. The scalar case  $p = 1$  gives (B.3) and (B.4) as

$$V_j = \int_{-\infty}^{\infty} b_j(y) dy \tag{B.5}$$

$$c_j = \frac{\int_{-\infty}^{\infty} y b_j(y) dy}{\int_{-\infty}^{\infty} b_j(y) dy} \tag{B.6}$$

Then the theorem follows by expanding the centroid of  $B$  and invoking the SAM assumption  $F(x) = \text{Centroid}(B) = \text{Centroid}(\sum_{j=1}^m a_j(x) B_j)$  to rearrange terms:

$$F(x) = \text{Centroid}(B) \tag{B.7}$$

$$= \frac{\int_{-\infty}^{\infty} y b(y) dy}{\int_{-\infty}^{\infty} b(y) dy} \tag{B.8}$$

$$= \frac{\int_{-\infty}^{\infty} y \sum_{j=1}^m b'_j(y) dy}{\int_{-\infty}^{\infty} \sum_{j=1}^m b'_j(y) dy} \tag{B.9}$$

$$= \frac{\int_{-\infty}^{\infty} y \sum_{j=1}^m a_j(x) b_j(y) dy}{\int_{-\infty}^{\infty} \sum_{j=1}^m a_j(x) b_j(y) dy} \tag{B.10}$$

$$= \frac{\sum_{j=1}^m a_j(x) \int_{-\infty}^{\infty} y b_j(y) dy}{\sum_{j=1}^m a_j(x) \int_{-\infty}^{\infty} b_j(y) dy} \tag{B.11}$$

$$= \frac{\sum_{j=1}^m a_j(x) V_j \frac{\int_{-\infty}^{\infty} y b_j(y) dy}{V_j}}{\sum_{j=1}^m a_j(x) V_j} \quad (\text{B.12})$$

$$= \frac{\sum_{j=1}^m a_j(x) V_j c_j}{\sum_{j=1}^m a_j(x) V_j}. \quad \square \quad (\text{B.13})$$

### Appendix C. Competitive AVQ algorithm for local means and covariations

The following procedure describes the competitive adaptive vector quantization (AVQ) algorithm for local means and pseudo-covariations.

1. Initialize cluster centers from sample data:  $m_i(0) = x(i)$  for  $i = 1, \dots, p$ .
2. Find the “winning” or closest synaptic vector  $m_j(k)$  to sample vector  $x(k)$ :

$$\|m_j(k) - x(k)\| = \min_i \|m_i(k) - x(k)\|,$$

where  $\|\cdot\|$  denotes the Euclidean norm:  $\|x\|^2 = x_1^2 + \dots + x_n^2$ .

3. Update the winner  $m_j(k)$ ,

$$m_j(k+1) = \begin{cases} m_j(k) + c_k[x(k) - m_j(k)] & \text{if the } j\text{th neuron wins,} \\ m_j(k) & \text{if the } j\text{th neuron loses,} \end{cases} \quad (\text{C.1})$$

and its pseudo-covariation estimate  $C_j(k)$ ,

$$C_j(k+1) = \begin{cases} C_j(k) + d_k[(x(k) - m_j(k))^{p/2}((x(k) - m_j(k))^{p/2})^T - C_j(k)] & \text{if the } j\text{th neuron wins,} \\ C_j(k) & \text{if the } j\text{th neuron loses,} \end{cases} \quad (\text{C.2})$$

where  $(x(k))^{p/2} \triangleq (\text{sign}(x_1(k))|x_1(k)|^{p/2}, \dots, \text{sign}(x_n(k))|x_n(k)|^{p/2})^T$ . We used  $p = 1$  in all simulations.

### References

- [1] V. Akgiray and C.G. Lamoureux, Estimation of stable-law parameters: a comparative study, *J. Business Econ. Statist.* **7** (1989) 85–93.
- [2] H. Bergstrom, On some expansions of stable distribution functions, *Ark. Math.* **2** (1952) 375–378.
- [3] L. Breiman, *Probability* (Addison-Wesley, Reading, MA, 1968).
- [4] S. Chen, C.F.N. Cowan and P.M. Grant, Orthogonal least squares learning algorithm for radial basis function networks, *IEEE Trans. Neural Networks* **2** (1991) 302–309.
- [5] J.A. Dickerson and B. Kosko, Fuzzy function learning with covariance ellipsoids, *Proc. 2nd IEEE Internat. Conf. on Fuzzy Systems (IEEE FUZZ-93)* (1993) 1162–1167.
- [6] J.A. Dickerson and B. Kosko, Fuzzy function approximation with supervised ellipsoidal learning, *World Congress on Neural Networks (WCNN '93)*, Vol. II (1993) 9–13.
- [7] R.W. Hamming, *Digital Filters* (Prentice-Hall, Englewood Cliffs, NJ, 2nd ed., 1983).
- [8] B. Kosko, *Neural Networks and Fuzzy Systems* (Prentice-Hall, Englewood Cliffs, NJ, 1991).
- [9] B. Kosko, Fuzzy systems as universal approximators, *IEEE Trans. Comput.* **43** (1994) 1329–1333. An early version appears in *Proc. 1st IEEE Internat. Conf. on Fuzzy Systems (IEEE FUZZ-92)* (1992) 1153–1162.
- [10] B. Kosko, Optimal fuzzy rules cover extrema, *Internat. J. Intelligent Systems* **10** (1995) 249–255.
- [11] B. Kosko, Combining fuzzy systems, *Proc. IEEE Internat. Conf. on Fuzzy Systems (IEEE FUZZ-95)*, Vol. IV (1995) 1855–1863.
- [12] B. Mandelbrot, The variation of certain speculative prices, *J. Business* **36** (1963) 394–419.
- [13] B. Mandelbrot and J.W. Van Ness, Fractional Brownian motions, fractional noises and applications, *SIAM Rev.* **10** (1968) 422–437.

- [14] E. Masry and S. Cambanis, Spectral density estimation for stationary stable processes, *Stochastic Processes Applic.* **18** (1984) 1–31.
- [15] J. Moody and C.J. Darken, Fast learning in networks of locally-tuned processing units, *Neural Comput.* **1** (1989) 281–294.
- [16] A. Papoulis, *Probability, Random Variables, and Stochastic processes* (McGraw-Hill, New York, 2nd ed., 1984).
- [17] M. Shao and C.L. Nikias, Signal processing with fractional lower order moments: stable processes and their applications, *Proc. IEEE* **81** (1993) 984–1010.
- [18] J. Shen and C.L. Nikias, Estimation of multilevel digital signals in the presence of arbitrary impulsive interference, *IEEE Trans. Signal Processing* **43** (1995) 196–203.
- [19] B.W. Stuck and L. Kleiner, A statistical analysis of telephone noise, *Bell Systems Tech. J.* **53** (1974) 1263–1320.
- [20] A. Wernon and K. Wernon, Stable measures and processes in statistical physics, in: S. Cambanis, Ed., *Probability Theory on Vector Spaces IV* (Springer, Berlin, 1987) 440–452.
- [21] V.M. Zolotarev, Mellin–Stieltjes transforms in probability theory, *Theory Probab. Applic.* **2** (1957) 433–460.