AFOSR-TR-80-1300

FINAL SCIENTIFIC REPORT
Grant AFOSR-76-3062

November 26, 1980

Presented to: Dr. I.N. Shimi
Directorate of Mathematical and Information Sciences

Presented by: G.L. Wise
University of Texas at Austin

Approved for public release; distribution unlimited.
This is the Final Scientific Report of the Grant AFOSR-76-3062. In it is given a brief survey of results achieved in the areas of nonlinearities with random inputs, regression functions, detection in Laplace noise, relative efficiency of detectors, signal detection in dependent noise, estimation of probability density functions from noisy measurements, polynomial expansions, median filtering, spherically invariant random processes, support estimation, and quantization theory.
This is the Final Scientific Report of the Grant AFOSR-76-3062. In it is given a brief survey of results achieved in the areas of nonlinearities with random inputs, regression functions, detection in Laplace noise, relative efficiency of detectors, signal detection in dependent noise, estimation of probability density functions from noisy measurements, polynomial expansions, median filtering, spherically invariant random processes, support estimation, and quantization theory.
INTRODUCTION

In recent years advances in many aspects of communication theory have proven to be limited by a lack of sufficient developments in the areas of applied probability and mathematical statistics. Our investigations attempted to overcome this deficiency by contributing both to the underlying theoretical basis of the area as well as to communication engineering. Among other areas, we have obtained fundamental results relating to nonlinear transformations of random processes, nonparametric estimation of regression functions, and signal detection theory.

This report is a survey of the technical activities ensuing from the Grant AFOSR-76-3062. In the next section we list the publications which were supported by this grant. Then we name the additional personnel who contributed to the research effort. We conclude with a brief survey of the research results.
Journal Articles


Submitted Papers - Titles and Journals Tentative


Conference Proceedings


Other


ADDITIONAL PERSONNEL ASSOCIATED WITH THE RESEARCH EFFORT

1. Efren F. Abaya - graduate student  
   presently a Ph.D. student

2. Nader Bagherzadeh - graduate student  
   M.S. thesis - August 1979 - "Quantized and Linear Detection"  
   presently with Bell Laboratories, Holmdel, NJ

3. James A. Bucklew - post-doctoral research associate  
   presently Assistant Professor, University of Wisconsin

4. Luc Devroye - post-doctoral research associate  
   presently Assistant Professor, McGill University

5. Don R. Halverson - graduate student  
   Ph.D. Dissertation - August 1979 - "Discrete Time Detection of Signals in Dependent Non-Gaussian Noise"  
   presently Assistant Professor, Texas A&M University

6. Nahid Khazenie - graduate student  
   presently a Ph.D. student

7. Federico Kuhlmann - graduate student  
   presently a Ph.D. student

8. Fu-Sheng Lu - graduate student  
   presently a Ph.D. student
9. Douglas L. Michalsky - graduate student
   M.S. thesis - December 1979 - "A Relative Efficiency Study of Some Popular Detectors"
   presently with Motorola, Phoenix, AZ

10. Dariush Minoo-Hamedani - graduate student
    M.S. thesis - August 1978 - "Mean Square Prediction Using a Memoryless Nonlinearity Followed by a Linear Filter"
    presently with Bell Laboratories, Holmdel, NJ

11. H. Vincent Poor - post-doctoral research assistant
    presently Assistant Professor, University of Illinois at Urbana - Champaign

12. Terry J. Wagner - co-principal investigator (10/1/79 - 9/30/80)
    presently Professor, University of Texas at Austin
SUMMARY OF RESEARCH RESULTS

In this section we briefly survey the principal results of our research.

Nonlinearities with Random Inputs

Mean square continuity of a random process is of considerable theoretical and practical importance. In many general treatments of random processes, mean square continuity is taken as a standing assumption (see, for example, [1] and [2]). We have investigated the mean square continuity of a random process after it has undergone a (zero memory) nonlinear transformation. Such nonlinearities are frequently encountered in many signal processing schemes; for example, quantizers, limiters, rectifiers, etc. Also, one of the most common models of non-Gaussian noise is a nonlinearly distorted Gaussian process. Before the initiation of this research, the most general result of this nature, obtained by this investigator, was for the case of first order stationary random processes [3]. We have now extended this previous result to consider nonstationary random processes [4]. We have established conditions on both the nonlinearity and on the random processes. For example, it follows that if \(X(t)\) is a mean square continuous Gaussian process whose variance is not identically zero, and if \(\mathcal{G}\) is the class of all Borel measurable functions \(g\) such that \(g[X(t)]\) is a second order random process, then \(g[X(t)]\) is mean square continuous, for any \(g \in \mathcal{G}\), if and only if the variance of \(X(t)\) is never zero.

A rather surprising result of the investigation was that the preservation of the mean square continuity after a (zero memory) nonlinear transformation depended solely upon the univariate distribution of the random process, not the bivariate distribution. This was true even though mean square continuity is a bivariate property, not a univariate property, of a random process. As a consequence, in the above situation, it is not necessary to work with the
bivariate distribution, which may not be completely known in many practical situations.

We extended the preceding idea to the following more general situation. Consider a system with a given input and the corresponding output. If a sequence of inputs converged to that particular input, it would often be of interest to know when the corresponding sequence of outputs converged to the particular output. In [5] we were concerned with this problem in a stochastic framework. We considered random variables taking values in a separable metric space, and we considered a Borel measurable mapping $g$ from the metric space to the reals. The elements of the metric space represented the possible inputs to the system and the mapping $g$ represented the system.

Let $(S,\mathcal{S})$ be a separable metric space and let $\mathcal{A}$ be the $\sigma$-algebra in $S$ generated by the closed sets. Let $(\Omega, \mathcal{F}, P)$ be a probability space. An $S$-valued random variable will be a measurable function from $(\Omega, \mathcal{F})$ to $(S,\mathcal{S})$. Let $X$ be an $S$-valued random variable, and let $\mu$ denote the measure induced on $\mathcal{A}$ by $X$, that is, for $A \in \mathcal{A}$, $\mu(A) = P(X \in A)$. Similarly, let $(X_n; n=1,2,...)$ be a sequence of $S$-valued random variables with corresponding measures $\mu_n$ induced on $\mathcal{A}$. The random variables $X_n$ are said to converge to $X$ in probability if for any $\epsilon > 0$,

$$\lim_{n \to \infty} P(\mu(X,X_n) > \epsilon) = 0.$$ 

The measures $\mu_n$ are said to converge to $\mu$ setwise if, for any element $A$ of $\mathcal{A}$,

$$\lim_{n \to \infty} \mu_n(A) = \mu(A).$$

Let $\mathcal{B}$ denote the Borel sets on $\mathbb{R}$. Consider a measurable function $k: (S,\mathcal{S}) \to (\mathbb{R},\mathcal{B})$ and an $S$-valued random variable $Y$. Then $k(Y)$ is a real-valued random variable. We say that $k(Y)$ belongs to $L_p (p \geq 1)$ if
\[
\int_{\Omega} |k[Y(\omega)]|^p \, p(d\omega) < \infty.
\]

If \( k(Y) \in L^p \), we define the \( L^p \) norm as
\[
\|k(Y)\| = \left( \int_{\Omega} |k[Y(\omega)]|^p \, p(d\omega) \right)^{1/p}.
\]

In [5] we were interested in a sequence of \( S \)-valued random variables \( X_n \) that converge to \( X \) in such a way that \( g(X_n) \) converges to \( g(X) \) in \( L^p \) where \( g \) is a measurable function. The following result was proved.

**Theorem 1:** Assume that \( X_n \to X \) in probability and that \( \mu_n \to \mu \) setwise.

Suppose \( g \) is a measurable function from \((S, \mathcal{S})\) to \((\mathbb{R}, \mathcal{B})\) such that \( g(X) \) and \( g(X_n) \) belong to \( L^p \). Then \( g(X_n) \to g(X) \) in \( L^p \) if, and only if,
\[
\|g(X_n)\| \to \|g(X)\|.
\]

We further investigated various particular consequences of this theorem.

By proper choice of the metric space, we can use these results to establish some convergence properties of general functional transformations of random processes.

From an applied point of view, one of the most important characteristics associated with a (stationary) random process is its spectrum. Many results concerning random processes are based upon spectral representations. In the context of the transmission of random signals, the spectral distribution is used to determine how much bandwidth is required for faithful transmission.

We have studied the effect of a zero memory nonlinearity on the spectrum of a random process. Consider a random process with a spectral distribution function \( F \). The second moment bandwidth of the random process is given by
\[
\left[ \frac{\int_{-\infty}^{\infty} \omega^2 dF(\omega)}{\int_{-\infty}^{\infty} dF(\omega)} \right]^{1/2}.
\]
In [6,7] we gave the following result:

**Theorem 2:** Suppose that \( X(t) \) is a zero mean, stationary Gaussian process that has a finite second moment bandwidth \( B \) and that possesses a spectral density function. If \( g \) is a Borel measurable function that is not constant (we identify functions equal a.e.) such that \( g[X(t)] \) is second order and \( E\{g[X(t)]\} = 0 \), then the second moment bandwidth of \( Y(t) = g[X(t)] \) is greater than or equal to \( B \). Equality holds if and only if \( g \) is linear.

In [6,7] and [8,9] we also presented the following result:

**Theorem 3:** Let \( X(t) \) be a stationary, mean square continuous Gaussian random process with a nonconstant autocorrelation function, and let \( g \) be Borel measurable and such that \( g[X(t)] \) is second order. Then \( g[X(t)] \) is strictly bandlimited if and only if

(a.) \( X(t) \) is strictly bandlimited, and

(b.) \( g(\cdot) \) is a polynomial.

Notice that many common zero memory nonlinearities are not polynomials. In particular, it follows that if \( X(t) \), given in Theorem 3, is passed through any type of limiter, then the output cannot be strictly bandlimited.

In actual practice, the validity of the Gaussian assumption is often questionable, and the preceding results were known to be valid for certain specific non-Gaussian processes. Recently we extended our analysis to some very wide (nonparametric) classes of non-Gaussian processes.

Let \( X(t) \) and \( N(t) \) be independent random processes that are second order, mean square continuous, and second order stationary. Assume that \( X(t) \) is a Gaussian process and that the autocorrelation function of \( X(t) \) is not a constant function. In [9] we obtained the following result.
Theorem 4: Let $Y(t) = X(t) + N(t)$, and let $g(\cdot)$ be any Borel measurable function such that $g[Y(t)]$ is a second order random process. We regard as identical two Borel measurable functions $g_1(\cdot)$ and $g_2(\cdot)$ such that $g_1[Y(t)]$ and $g_2[Y(t)]$ are equivalent random processes.

A. If $g(\cdot)$ is not a polynomial, then $g[Y(t)]$ cannot be bandlimited for any mean square continuous second order stationary random process $N(t)$.

B. If $X(t)$ is not bandlimited, then $g[Y(t)]$ cannot be bandlimited for any nonconstant Borel measurable function $g(\cdot)$ such that $E\left(\left[g[Y(t)]\right]^2\right) < \infty$.

In Theorem 4 $Y(t)$ can be regarded as a contaminated Gaussian process where $N(t)$ is the contamination component. Other than the very mild restrictions mentioned above, $N(t)$ is totally arbitrary.

In [10] we presented the following theorem which concerns the effect of a ZNL on the spectrum of randomly modulated Gaussian noise. In this theorem $X(t)$ and $N(t)$ are as above.

Theorem 5: Let $Y(t) = N(t) X(t)$ and let $g(\cdot)$ be a Borel measurable function such that $g[Y(t)]$ is a second order random process. We regard as identical two Borel measurable functions $g_1(\cdot)$ and $g_2(\cdot)$ such that $g_1[Y(t)]$ and $g_2[Y(t)]$ are equivalent random processes. Then statements A and B of Theorem 3 hold.

In [11] we presented results concerning equivalent classes of zero memory nonlinearities; that is, different nonlinearities which produce the same spectral transformations upon a stationary random process.

There exist a great many results based upon the second moment characterization of random processes. Almost all of linear filtering theory and linear estimation is based upon second moment theory. Many classes of random processes
are defined in terms of their second moment properties, for example, purely nondeterministic random processes, wide sense Markov processes, bandlimited processes, etc. Except for the case where a class of random processes is defined in terms of its second moment properties, there are few results concerning the restrictions placed upon the second moment properties of a random process by virtue of the random process belonging to a certain class. For a Gaussian random process, there are no restrictions placed upon the second moment properties, other than those restrictions which are common to all second moment properties. However, this is not true for non-Gaussian processes. We have established some results of this nature. Results such as these have application in modeling the second moment statistics of random signals and noise. Notice that since much filter design is based upon second moment theory, results of this nature will also be important from the viewpoint of system design.

In a related context, an investigation of a discrete time nonlinear Wiener filter was initiated. The filter was constrained to be composed of a memoryless nonlinearity followed by a linear filter. The study was concerned with determining how to specify the memoryless nonlinearity. Once the nonlinearity is known, the linear filter can be determined with standard techniques. The results of this effort are given in [12] and [13], where several methods are investigated for determining the nonlinear systems. It is shown that in many cases a nonlinear system of this form can significantly outperform the optimal linear system.

Regression Functions

In this area we investigated two different aspects of the regression function
\[ m(x) = E(Y|X=x), \]

where \( Y \) is an integrable random variable and \( X \) is a random variable or a random vector.

In [14, 15] we were concerned with determining the regression function \( m(x) \) from only a partial characterization of the joint distribution of \( X \) and \( Y \). We showed the following:

**Theorem 6:** Let \( Y \) be an integrable random variable, let \( X \) be an arbitrary random variable, and let \( g(\cdot) \) be an invertible Borel measurable function mapping the reals into a bounded set. Then the regression function \( m \) is determined up to probability one equivalence by the quantities

\[ E([g(x)]^k), \quad k = 1,2,... \]

and

\[ E(Y[g(x)]^k), \quad k = 0,1,2,... . \]

Thus from this theorem we see that statistical information consisting of various moments and joint moments is sufficient to characterize a regression function. In [14, 15] the extension to the case where \( X \) is a random vector taking values in \( \mathbb{R}^n \) or a random process, e.g. \( \{X(t), t \in \mathbb{T}\} \), is given.

In a different aspect of this area, we investigated the estimation of a regression function from empirical data. It is reasonable to expect that with a large amount of empirical data we could achieve a good estimate of a regression function. However, with a large amount of data, we may be faced with computational burdens in processing them. Therefore, a recursive method of estimation may seem attractive. In [16] we presented distribution-free consistency results for the recursive nonparametric regression function estimation problem.

Assume that \((X,Y), (X_1,Y_1), \ldots, (X_N,Y_N)\) are independent identically
distributed $\mathbb{R}^d \times \mathbb{R}$-valued random vectors with $E \{ |Y| \} < \infty$. Consider estimating the regression function
\[ m(x) = E \{ Y | X = x \} \]
from the data $(X_1, Y_1), \ldots, (X_N, Y_N)$. We proposed the following estimate.

Break the data up into disjoint blocks of length $b_1, b_2, \ldots, b_n$, and among all $X_i$ in the $j$-th block, find the one that is closest to $x$ in the $l_q$ norm $|| \cdot ||$ on $\mathbb{R}^d$ (in case of a tie, pick the $X_i$ with the lowest index $i$). Let us call the corresponding $\mathbb{R}^d \times \mathbb{R}$-valued random vector $(X^*_j, Y^*_j)$. (The dependency on $x$ is suppressed for the sake of brevity.)

If \{w_{n1}, \ldots, w_{nn}\}, $n \geq 1$ is a triangular array of positive weights, then we proposed to estimate $m(x)$ by
\[ m_n(x) = \frac{\sum_{j=1}^{n} w_{nj} Y^*_j}{\sum_{j=1}^{n} w_{nj}} \tag{1} \]
when $N = b_1 + \ldots + b_n$ observations $(X_i, Y_i)$ are available. Notice that when $w_{ni} = v_i$ for all $n,i$, then the computation in (1) can be performed recursively.

That is, there is no need to store all the observations $(X_i, Y_i)$, and if we are not satisfied with $m_n$ we can collect more observations and update our estimate. Also, (1) retains the flavor of the nearest neighbor estimates (see, for example, [17, 18]), but the processing burden arising from the ranking procedure is less.

The conditions which we put upon $b_n$ and $w_{ni}$ were weak:
\[ b_n \sup_{1 \leq i \leq n} \frac{w_{ni}}{\sum_{j=1}^{n} w_{nj}} \rightarrow 0. \]
Let
\[ I_{np} = \int |m_n(x) - m(x)|^p \, \mu(dx), \]
where \( \mu \) is the probability measure of \( X \). In [16] we showed that
\[ E \{ I_{np} \} \leq 0 \text{ whenever } E \{ |Y|^p \} < \infty \text{ (for } p \geq 1 \}, \]
and that \( I_{np} \leq 0 \) with probability one when \( Y \) is almost surely bounded.

Consider the case that \( Y \) is \( \{1, \ldots, M\} \)-valued and that \( Y \) must be estimated from \( X \) and the data (the discrimination problem), by, say, \( g_n(x) \) where \( g_n \) is a Borel measurable function
\[ g_n : \mathbb{R}^d \times \mathbb{R}^d \times \{1, \ldots, M\}^N \rightarrow \{1, \ldots, M\}. \]

In [16] we considered an application to the discrimination problem, and we presented a discrimination rule that was strongly Bayes risk consistent. This is the first distribution-free strong Bayes risk consistency result in the literature.

In [19] the \( L_1 \) convergence of kernel regression function estimators was studied, and some applications to the discrimination problem were considered.

**Detection in Laplace Noise**

Recently, there has been considerable interest in the detection of signals in non-Gaussian noise. Although the assumption of Gaussian noise is frequently justified, such as in UHF; in other cases, such as ELF (extra low frequency), the assumption is definitely unjustified. One form of frequently encountered non-Gaussian noise is that known as impulsive noise. Impulsive noise is typically characterized as noise whose distribution has an associated "heavy tail" behavior. That is, the probability density function (pdf) approaches zero more slowly than a Gaussian pdf. We considered the discrete time detection
of a known constant signal in additive white Laplace noise. Laplace noise is characterized by a double exponential pdf. This noise is typical of the class of impulsive noises. The references in [20] give a summary of some forms of impulsive noise and situations where it arises. For example, Bernstein, et al. [21] comment on the non-Gaussian nature of ELF atmospheric noise, and they give a plot of a typical experimentally determined pdf associated with such noise [21, figure 10]. This experimentally determined pdf is similar to a Laplace pdf, and on a linear graph the difference is barely distinguishable. To quote Miller and Thomas [22]: "Non-Gaussian noise does not seem to be a problem for radars operating at UHF and above, but those long range radars operating at HF frequencies must contend with the same impulsive atmospheric noise that disturbs communication systems in that spectral region."

The form of the Neyman-Pearson optimal detector for this problem is well known [22, 23] and has the structure of an amplifier-limiter followed by a summer. The accumulated sum is the test statistic which is compared to a threshold to announce the presence or absence of the signal. In order to determine the performance of the detector, it is necessary to know the distribution of the test statistic. This is pertinent, for example, to the determination of how many samples must be taken to achieve a given level of performance.

The distribution of the test statistic has been extremely elusive and past attempts at obtaining a simple expression for this distribution have not been very successful. The most notable success had been achieved by Miller and Thomas [23], who gave a lengthy and complex recursion scheme for obtaining the distribution. Their results, however, were of a numerical nature and did not culminate in a closed form analytical expression for the distribution of the test statistics. In fact, for 35 samples their method required over half
an hour of time on an IBM System 360 Model 91 digital computer.

If the number of samples were sufficiently large, the Central Limit
Theorem would apply, and the distribution of the test statistic would be
approximately normal. However, the small sample performance of the detector
would still be unknown (see, for example, [23, 24]). Alternatively, one could
establish bounds on the detection and false alarm probabilities, and thus
establish a bound on detector performance; or Monte Carlo simulation may be
employed. In general, however, it would be desirable to have a convenient
expression for the probability distribution of the test statistic.

In our recent investigations [25-27] we developed a simple, convenient,
closed form analytical expression for the probability distribution function
of the test statistic for the Neyman-Pearson optimal detector. This result
enabled us to study several aspects of the detection problem. In particular,
we analyzed the small sample performance of the optimal detector. We also
considered the performance of the linear detector.

These results are pertinent to long range radars operating in spectral
regions associated with Laplace noise. They may also yield some insight
into relative efficiencies. Detectors are frequently compared on the basis
of asymptotic relative efficiency. However, as noted by Helstrom [28], when
the number of samples is not large, the detectors, or receivers, may behave
quite differently from the predictions of the asymptotic theory. Very little
work has been done in this area [23]. Our results offer the possibility of
more insight into relative efficiencies.

It should be noted that for the Neyman-Pearson discrete time detection
problem of a sure signal in non-Gaussian white noise, there are extremely few
cases where the distribution of the test statistic is known for an arbitrary
number of samples. Our result represents such a case.
As a specific comment on our work, to evaluate the distribution function of the test statistic at a given point for the above problem with 35 samples, our method requires less than one quarter of one percent of the computational time required by the previously best known method.

Relative Efficiency of Detectors

The asymptotic efficiency of a discrete time signal detection scheme is often viewed as a valid measure of its detection performance. In this case the asymptotic relative efficiency (ARE) is usually employed as a criterion for comparison of detectors. The ARE is generally held to be appropriate in the case of large sample size and small signal strength. Moreover, the employment of the ARE generally yields mathematically tractable results, due largely to the applicability of central limit theorems.

In any practical engineering situation, we can take only a finite number of samples. The number of samples available, however, may not be sufficiently large to ensure that the ARE is an appropriate indicator of detection efficiency. For example the requirement that the samples be statistically independent may set an upper bound on the sampling rate. Thus we are actually concerned with the efficiency of the detector with the number of samples available. In this case the relative efficiency between detectors is of interest. This quantity is a measure of the amount of data one detector requires, relative to a reference detector, to attain a prescribed level of performance. It is generally accepted that the ARE gives a good indication of relative efficiency for moderate sample sizes. However, the exact analysis of relative efficiency is generally hindered by mathematical difficulties, and there has been very little work done in the area of relative efficiency analysis to verify this assumption.
(see, for example [23]). In [29] we investigated the exact relative efficiencies of two pairs of widely used detection systems for some commonly assumed noise distributions, and we demonstrated that the ARE can sometimes be a poor predictor of finite-sample-size detection performance even for some very large sample sizes.

**Signal Detection in Dependent Noise**

A longstanding area of both practical and theoretical importance has been the detection of signals in corrupting noise. A situation of increasing interest and importance has been the presence of a dependent noise source. Because of modern high-speed sampling such a situation should prove to be even more important in the future. In this case Neyman-Pearson techniques have been found to be tractable only in cases where the appropriate multivariate distribution of the noise is known, e.g., if the noise process is Gaussian. There are, however, a number of cases where a non-Gaussian assumption is considered necessary (see, for example, [20, 21, 30-42]), and it would appear likely that in the future such cases will become even more numerous.

Recall that the Neyman-Pearson optimal detector for independent data consists of a memoryless nonlinearity followed by an accumulator followed by a threshold comparator [22]. The Neyman-Pearson optimal detector for dependent data consists of a more complicated structure. In some cases we may realize that there is statistical dependence in the data and not be satisfied with using the detector which is optimal for independent data, and at the same time feel that there is not enough dependence within the data to warrant a radically different structure for the detector. Also we might not have a complete enough statistical characterization of the dependent data to design the Neyman-Pearson
optimal detector. Thus we may be satisfied with the basic structure of the optimal detector for independent data but desire to choose a different (i.e. other than the one which is optimal in the independent case) non-linearity in the detector so as to account for the dependency in the data. This was the approach taken by Poor and Thomas [42] who considered the detection of a known constant signal in m-dependent noise. In our work we have significantly generalized this approach.

In [43, 44] we extended the above m-dependence assumption to the case of symmetrically \( \phi \)-mixing noise processes. Let \( \{N_i\}_{i=1}^{\infty} \) be a strictly stationary sequence of random variables. For \( a \leq b \), define \( \mathcal{N}_a^b = \sigma(N_a, N_{a+1}, \ldots, N_b) \), the \( \sigma \)-algebra generated by the indicated random variables. Then \( \{N_i\}_{i=1}^{\infty} \) is symmetrically \( \phi \)-mixing if there exists a nonnegative sequence \( \{\phi_i\}_{i=1}^{\infty} \) with \( \phi_i \to 0 \) such that for each \( k, 1 \leq k < \infty \) and for each \( i \geq 1 \), \( E_1 \in \mathcal{N}_1^k \), \( E_2 \in \mathcal{N}_1^{k+1} \) together imply

\[
|P(E_1 \cap E_2) - P(E_1)P(E_2)| \leq \phi_i \max\{P(E_1), P(E_2)\}.
\]

Thus we see that the assumption of a symmetrically \( \phi \)-mixing noise process permits a great deal of flexibility in modeling the dependency structure of the noise.

In [45, 46] we considered the same basic situation as investigated in [43, 44] (i.e. the case for symmetrically \( \phi \)-mixing noise), except we constrained the nonlinearity to be a polynomial. This polynomial constraint resulted in a great deal of simplification in determining the nonlinearity in the detector.

The class of random processes used to model the noise in the above work may be seen to be quite general; however, the assumption of a constant known
signal is in some cases overly restrictive. Instead of such an assumption, we might wish to model the signal as a random process. Also, since we allowed dependency between noise samples, it would be desirable to allow dependency between signal samples. Finally, it would seem reasonable to allow some degree of dependency between signal and noise (to encompass, for example, the signal dependent noise induced through reverberation effects).

This is the situation we considered in [47, 48] where we extended the work of [43, 44] to this area. That is, we used the same detector structure as described above for [43, 44], but we allowed the signal to be symmetrically \( \varphi \)-mixing, we allowed the noise to be symmetrically \( \varphi \)-mixing, and we allowed the noise to be dependent upon a finite window of the signal (the \( i \)-th noise sample could be dependent upon the \( (i-m) \)-th to the \( (i+m) \)-th signal samples). In [49, 50] we generalized some of the results of [43, 44] and [47, 48] by weakening the assumption of symmetrically \( \varphi \)-mixing processes to the assumption of strong mixing processes.

The above work in signal detection which we have described required some statistical knowledge of the data; in [43, 44] and [47, 48] bivariate densities were assumed to be known, and in [45, 46] bivariate moments were assumed to be known. In some practical situations, however, very little is known concerning the statistical properties of the noise. The employment of a nonparametric detector is often desirable in situations where little information about the statistics of the noise is available. If the noise sequence is independent and identically distributed, a popular choice for detection of a constant signal is the well known sign detector [51]. Because of a modern high speed sampling, however, in many situations it is unlikely that adjacent samples of the waveform could be considered to be statistically independent. What we might expect in
some situations is that samples separated sufficiently far apart in time could be considered to be independent, i.e. an assumption of m-dependence might be reasonable. In these cases the sign detector unfortunately loses its nonparametric nature. It is thus desirable, when confronted with this form of dependency in the noise, to modify standard nonparametric schemes in a way which is easily implemented and yet preserves the nonparametric nature of the detector under dependent inputs. One promising approach toward this goal was considered by Kassam and Thomas [52]. Consider the detection problem of a constant signal in m-dependent noise. Kassam and Thomas [52] considered the following scheme. Group the samples into blocks of length n with m samples skipped between the blocks. Then for each block add the samples together. Now apply the sign detector to this sequence of independent random variables. We will refer to this scheme as a modified sign detector. A question which naturally arises for the modified sign detector is what choice of block length n gives the best performance. In [52] the block length was investigated from the viewpoint of the asymptotic situation. Asymptotic performance measures are frequently used in statistics and the resulting schemes usually work well. However, in this particular scheme the block length n effectively serves to "shrink" the data (i.e. n samples are summed, thus shrinking n samples to one sample). At this point we might suspect the validity of asymptotic results, since regardless of how much the data are shrunk by the summing operation, we would still be working with an unbounded number of blocks. In a practical situation there would be a finite number of samples, and thus as n (the length of each block) becomes larger, the number of blocks will decrease. In [53, 54] we investigated how the block size for the modified sign detector may be selected for two fidelity criteria, one based on a finite number of samples and the other on the asymptotic limit. We have found by way of example that it is possible for the
two criteria to disagree radically on the optimal block size.

In [55] we analyzed the above sample and skip procedure as applied to strong mixing noise. We showed how a modified sign detector may be designed for the nonparametric detection of a constant signal in strong mixing noise.

Estimation of Probability Density Functions from Noisy Measurements

By and large, probability densities are not obtained from physical derivations, but from empirical data. Measurements are taken, and from these measurements a density function is obtained. Several methods have been proposed for the estimation of probability density functions, and numerous properties of these methods have been studied [56, 57]. However, these methods assume that the measurements from which the density is estimated are not corrupted by noise. In many practical situations, the measurements from which one constructs the estimated density are corrupted by noise. The corrupting noise might arise from background noise not associated with the random variable of interest, or it may arise from noise introduced by the measuring techniques. Although there is quite extensive literature on the estimation of probability density functions (most of it relatively new), little has been done for the case where the measurements are corrupted by noise.

As a specific example of the foregoing, we have treated the case where the measurements are independent and identically distributed and corrupted by independent additive Poisson noise. That is, each measurement is of the form

\[ Y = X + N, \]

where \( N \) is a Poisson random variable and \( X \) is the random variable whose density function we desire to estimate. We have developed a procedure [58] for estimating the density function of \( X \) from measurements corrupted by Poisson
noise. We have established the appropriate forms of convergence and we have given a practical realization of the estimator.

We also investigated various problems involving the recovery of a discrete probability density from independent observations [59, 60]. We considered estimation of the discrete density function in the presence of additive noise, and we solved the problem for the cases of Poisson, geometric, and binomial noises. We also investigated the recovery of a discrete density when some of the measurements are incorrect. Finally, we considered recovering the parameters of a mixture density from independent observations. We derived an easy-to-implement estimate of the parameters such that all of the parameter estimates are nonnegative and they sum to unity.

**Polynomial Expansions**

Two common ways of representing functions have been polynomial expansions and trigonometric expansions. In much of engineering the trigonometric expansion has useful interpretations and has dominated over the generalized Fourier series expansions in applications. However, many functions are readily expressed in terms of polynomials. We have derived [61-64] a simple linear transformation which maps the polynomial representation into a trigonometric representation. Also, we have derived the inverse transformation which maps a trigonometric expansion to a polynomial expansion.

The inverse transformation has enabled us to develop a fast algorithm for the computation of the Legendre polynomial coefficients for any $L_2[-\pi, \pi]$ function. The algorithm utilizes the Fast Fourier Transform (FFT) to compute the Fourier series coefficients and then multiplies the vector of coefficients
by a linear matrix transformation to compute the vector of polynomial coefficients. This approach can offer a considerable saving in computation time over the standard integral formula for computing these coefficients.

**Polynomial Expansions of Bivariate Densities**

The diagonal series expansion of a bivariate density function in terms of orthonormal functions yields considerable structural information about the bivariate density and, due to the previous work of this investigator [65], is readily interpretable in terms of Markov sequences. In the case where the orthonormal functions are polynomials, the bivariate density function is said to belong to the class $\Lambda$, introduced by Barrett and Lampard [66]. The class $\Lambda$ has been studied by many people and several properties of this class are known. However, the number of specific examples of bivariate densities which belong to the class $\Lambda$ is not large.

We have derived some new examples of bivariate density functions that belong to the class $\Lambda$. The examples we have derived are associated with Gegenbauer polynomials with parameter $3/2$ [67].

**Median Filtering**

In many signal processing applications the concept of a linear filter is a basic one. However, there are situations where linear filtering is inadequate. For example, if the signal displays sharp discontinuities in addition to being corrupted by high frequency noise, then a linear filter designed to eliminate the noise will also smooth out the signal. Recently a nonlinear
method called median filtering has achieved some very interesting results. Median filtering was introduced by Tukey [68-71], and it has produced promising results in picture processing [72, 73] and speech processing [74, 75]. However, most of the work in the open literature is of an empirical, a survey, or an implementation nature. The implementation of a median filter requires a very simple digital nonlinear operation. To begin, we take a sampled and quantized signal and across this signal we slide a window that spans $2N+1$ adjacent signal sample points. The filter output is set equal to the median value of these $2N+1$ signal samples. The filter output is associated with the time sample at the center of the window. To account for start up and end effects at the two endpoints of the signal, $N$ samples are appended to the beginning and end of the sequence. The appended samples are constant and equal in value to the first and last samples of the original sequence, respectively.

In [76, 77] we presented a theoretical analysis of median filters. We studied the effects of median filters, and we completely characterized the signals which are unaffected by median filters. That is, we gave a necessary and sufficient condition for a signal to be invariant to a median filter. We called a signal unaffected by a median filter a root, and we showed that by successive median filtering operations, any signal is reduced to a root. For a signal of length $L$, we showed that a maximum of $\frac{L}{2}(L-2)$ repeated filterings produces a root signal. In particular, it follows that if a signal is changed by a median filter, then this signal can never be exactly recovered by successive median filtering operations (i.e. successive operations cannot result in a cycling effect).

In [78, 79] we derived an expression for the bivariate distribution
function of the output of a median filter with independent identically distributed random variables for the input, and we analyzed the effect of a median filter upon the second moment properties of a sequence of independent identically distributed random variables. In the cases that we analyzed, we found that the power spectrum of the output of the median filter suggested a low sensitivity to the input distribution. Our results also suggested a low pass characteristic of the median filter.

Spherically Invariant Random Processes

Communication engineers have traditionally relied upon the Gaussian model, both because of practical considerations and important theoretical properties. Often, extensions of the Gaussian process have been investigated, which are frequently more general models but retain many useful properties of this process. One particularly attractive property of a Gaussian process has been the linearity of all minimum mean squared error estimation problems. One such generalization of the Gaussian case has been the spherically invariant random process (SIRP).

SIRP's were introduced by Vershik [80] when he was investigating a class of random processes which shared some properties characteristic of Gaussian processes. In particular, SIRP's are the most general class for which minimum mean squared error estimates admit linear solutions, and this class of processes is closed under linear operations. In an interesting paper, Blake and Thomas [81] explored some important properties of SIRP's. Then in a recent paper [82] Yao presented some very significant results concerning SIRP's. In particular, he presented a representation theorem for the family of finite dimensional distribution of SIRP's. The references in [82] provide a summary
of other work done in this area.

We have established [83, 84] the following representation theorem for SIRP's.

**Theorem 7:** A random process is a (centered) spherically invariant random process if and only if it is equivalent to a random process of the form \( AY(t) \), where \( A \) is an arbitrary random variable and \( Y(t) \) is a zero mean Gaussian process independent of \( A \).

This theorem explicitly illustrates the relation between a SIRP and a Gaussian process, and most properties of SIRP's follow in an elementary fashion from the theorem. This result will find applications in any situation where a SIRP is used to model random phenomena.

**Support Estimation**

A problem of increasing significance to engineers concerns the detection of abnormal or faulty behavior of a system, plant, or machine. Assume that we have observed the system in normal operation and that we have taken measurements of the normal behavior. A measurement is assumed to be an \( \mathbb{R}^d \)-valued random vector. The randomness may be due to measurement noise, parasitic effects, or random inputs. Thus the measurements are given by \( X_1, X_2, \ldots, X_n \), a sequence of \( \mathbb{R}^d \)-valued random vectors which we assume are independent with a common unknown probability measure \( \mu \).

Classically, the assumption is made that one has access at the present time to \( m \) independent observations \( X'_1, X'_2, \ldots, X'_m \) with common probability measure \( \nu \), and the system is said to behave differently, or abnormally, if \( \nu \neq \mu \). To detect such a change in distribution, several tests have been proposed (for example, [85-91]).
In [92] we treated the problem concerned with taking only one new observation. For economic reasons, lack of time, or practical limitations, only one new observation $X$ can be made and there is no hope to recover or approximate $\nu$ as with the large sample $X'_1, X'_2, \ldots, X'_m$. Regardless of $\nu$, we say that the system behaves abnormally if $X$ does not belong to $S$, the support of $\mu$. In several practical applications, the complement $S^c$ of $S$ can be considered as a danger area because under normal behavior (with probability measure $\mu$) the probability that some of the $X_i$ take values in $S^c$ is zero. Thus the problem is reduced to one of estimating the support $S$ from $X_1, X_2, \ldots, X_n$. This problem is treated in [92].

Another problem that we considered was concerned with taking $n$ new measurements which are independent with common unknown probability measure $\nu$. We assumed that the system might have changed, but we were concerned with whether or not the system might exhibit abnormal behavior. We assumed that the system still functions normally if the support of $\nu$ is contained within $S$. This problem was also treated in [92].

**Topics in Quantization Theory**

The quantization of continuous amplitude, discrete time signals combined with the transmission of the quantized samples over noisy channels is a problem that was considered in [93]. We investigated the total mean squared distortion suffered by a companded, continuous amplitude memoryless source which is uniformly quantized and transmitted over a noisy channel with a known capacity. We were interested in a small distortion analysis, i.e. quantizers with very large numbers of quantization levels and channels whose capacities are large
enough to carry the data rates coming out of the quantizer. The twin tools of asymptotic quantization theory and rate distortion theory were used to find an expression for the approximate total mean squared distortion. In [93] the approximate total mean squared distortion was minimized over a class of parameterized compressor characteristics for input processes whose univariate probability density functions were members of the generalized Gaussian family.

In [94] we investigated the asymptotic theory of k dimensional quantization for r-th power distortion measures. Subject only to a moment condition, it was shown [94] that the infimum over all N level quantizers of the quantity $N^{r/k}$ times the r-th power distortion measure converged to a finite constant as $N \to \infty$. This work was more general than any of the previous efforts for this distortion measure.
References


