# The Use of Noise Properties in Set Theoretic Estimation

Patrick L. Combettes, *Member, IEEE*, and H. Joel Trussell, *Senior Member, IEEE*

***Abstract*—In most digital signal processing problems, the goal is to estimate an object from noise corrupted observations of a physical system. In this paper, we describe how a wide range of probabilistic information pertaining to the noise process can be used in a general set theoretic estimation framework. The basic principle is to constrain the sample statistics of the estimation residual to be consistent with those probabilistic properties of the noise which are available and to construct sets accordingly in the solution space. Adding these sets to the collection of sets describing the solution will yield a smaller feasibility set and, hence, more reliable estimates. Pieces of information relative to quantities such as range, moments, absolute moments, second and higher order probabilistic attributes are considered and properties of the corresponding sets are established. Simulations are provided to illustrate the theoretical developments.**

## I. Introduction

At the very core of most digital signal processing concepts lies an estimation problem. Indeed, a typical digital signal processing problem can be abstracted into estimating an object $h$ (e.g., a single parameter, a collection of parameters, a function) from the data provided by observing some discrete stochastic process $(X_n)_{n \in Z}$. These problems are usually approached via conventional estimation techniques, i.e., techniques that generate a solution which is optimal with respect to some predefined criterion. Such techniques are open to criticism in that, too often, they rely on questionable critèria of optimality and involve subjective statistical hypotheses. As a result, they may produce solutions which violate known information about the problem. Set theoretic estimation is a technique which does not provide an optimal solution but, rather, a set of solutions defined as the class of objects consistent with all information arising from *a priori* knowledge and the observed data. If $(\Psi_j)_{1 \leq j \leq m}$ is the collection of propositions representing such information and $\Xi$ the solution space, a collection of so-called property sets $\Gamma = (S_j)_{1 \leq j \leq m}$ is constructed in a propositional manner, namely,

$$S_j = \{a \in \Xi \mid \Psi_j \text{ holds for } a\}. \tag{1}$$

A set theoretic estimate is any object consistent with all available information, i.e., any point in the set

$$S = \bigcap_{j=1}^{m} S_j = \{a \in \Xi \mid (\forall j \in \{1, \cdots, m\}) \Psi_j \text{ holds for } a\}. \tag{2}$$

The set $S$ is interchangeably called the solution or the feasibility set. The set theoretic estimation problem is then to find a point in $S$. This can be achieved via the method of successive projections [6] or the method of random search [7]. It is noted that the set theoretic estimation problem is posed as a feasibility problem rather than an optimality problem.

In digital signal processing, there are numerous problems which have been formulated within the framework of set theoretic estimation. We can mention in particular filter design [1], processing of electron microscopy data [4], signal restoration [5], [35], [37], speech prediction [9], linear system identification [14], [15], signal reconstruction [21], [36], image coding [29], and tomography [31].[1]

A central issue in set theoretic estimation is the construction of a collection $\Gamma$ of property sets from available information. Such sets can usually be formed from constraints pertaining to the object to be estimated or to the physical system that generated the data. In this paper, it is shown how to construct property sets in a quite general estimation problem from various pieces of information relative to the noise process that corrupts the data. The basic principle is, for each known property $\Psi_j$ of the noise, to construct the set $S_j$ of estimates which yield an estimation residual whose sample statistics are consistent with $\Psi_j$. Through the addition of these sets, one incorporates the *a priori* knowledge pertaining to the noise, thereby refining the feasibility set and improving the quality of the estimates. The idea of imposing noise-based constraints on the estimation residual was first formulated in a version of the constrained least squares problem in [25] and then applied to least squares image restoration in [18]. There, the sample second moment of the residual was forced to match that of the noise. This particular constraint has also been employed in other restoration techniques [34], [35], and in linear system identification [14].

[1] A tutorial account of some aspects of set theoretic estimation can be found in [10].

Manuscript received December 1, 1989; revised July 27, 1990.
P. L. Combettes is with the Department of Electrical Engineering, the City College of the City University of New York, New York, NY 10031.
H. J. Trussell is with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695.
IEEE Log Number 9144735.

1053-587X/91/0700-1630$01.00 © 1991 IEEE

Constraints arising from bounds on the amplitude of the noise were first employed in control [30] and then in several signal processing problems [9], [15], [17]. In the set theoretic restoration problem posed in [35], new constraints were introduced by considering other pieces of noise information (mean, spectral density) under the assumption that the noise was white and Gaussian.

The purpose of this paper is to generalize the approach taken in [35] in three respects. First, the analysis is extended to arbitrary set theoretic estimation problems with noise corrupted data. Second, one departs from the assumption that the noise is white and Gaussian. Third, it is shown how probabilistic attributes such as range, moment, and absolute moments of arbitrary order, second- and higher order properties can be used to create property sets in the solution space. The paper is organized as follows. The methodology is described in Section II and signal processing examples are given in Section III. Sets based on various pieces of noise information are constructed and analyzed in Sections IV through IX. The question of redundant noise information is addressed in Section X. Simulation results are provided in Section XI. Our conclusions appear in Section XII. All the propositions are proven in the Appendix.

## II. Methodology

Throughout this paper, $\Xi$ has the structure of a topological vector space. All the random variables (r.v.'s) are defined on a probability space $(\Omega, \Sigma, \mathsf{P})$. Lowercase letters are used to denote the value of a r.v. at a given elementary event $\omega$ in $\Omega$, representing a particular realization of the underlying stochastic process. The set of integers is denoted by $\mathbb{Z}$, the set of positive real numbers by $\mathbb{R}_+^*$, and the chi-square distribution with $n$ degrees of freedom and mean $n$ by $\chi_n^2$. For all $p$ in $\mathbb{R}_+^*$, $L^p(\mathsf{P})$ denotes the vector space of r.v.'s with finite $p$th absolute moment, a.s. stands for P-almost surely, and i.i.d. for independent and identically distributed.

Following Loève [23], the probability theoretic properties of a family of r.v.'s are defined as those properties that can be expressed in terms of the joint distribution functions (d.f.'s) of its finite subfamilies. The following basic result is central to the subsequent developments.

*Proposition 1:* If two stochastic processes $(Y_n)_{n\in Z}$ and $(U_n)_{n\in Z}$ are equivalent, i.e.,

$$(\forall n \in \mathbb{Z}) \quad Y_n = U_n \quad \text{a.s.} \tag{3}$$

then they possess the same probability theoretic properties.

In digital signal processing, a general probabilistic model for the generation of the data process $(X_n)_{n\in Z}$ is the discrete stochastic equation

$$(\forall n \in \mathbb{Z}) \quad X_n = T_n(h) + U_n \quad \text{a.s.} \tag{4}$$

In that model, $T_n$ is the signal formation operator and $(U_n)_{n\in Z}$ is the noise process. Given an estimate $a$ for $h$, the estimation residual is defined as

$$(\forall n \in \mathbb{Z}) \quad Y_n = X_n - T_n(a). \tag{5}$$

In an ideal situation where the true object would be estimated with no error, i.e., $a = h$, one would get $T_n(a) = T_n(h)$, for every integer $n$. Then, it follows easily from (4) and (5) that the residual and noise processes are equivalent. Hence, by Proposition 1, if $\Psi_j$ is a known probability theoretic property of the noise process, the estimate $a$ should lie in the subset $S_j$ of the solution space $\Xi$ defined by

$$S_j = \{a \in \Xi \mid (Y_n)_{n\in Z} \text{ satisfies } \Psi_j\}. \tag{6}$$

In practice, however, only a finite number of samples of the data process are observed, yielding a finite sample path $(x_i = X_i(\omega))_i$. Consequently, the residual process is traceable only through some sample path $(y_i = x_i - T_i(a))_{1\le i\le n}$ and the set which will actually be employed is

$$S_j = \{a \in \Xi \mid (y_1, \cdots, y_n) \text{ is consistent with } \Psi_j\}. \tag{7}$$

In the next sections, such sets will be constructed from various pieces of noise probabilistic information. Set properties such as closedness and convexity will be established since, as discussed in [6], they are of great importance in connection with the synthesis of set theoretic estimates by successive projection methods.

Henceforth, $n$ will be the length of the residual path. The $n$-tuples with $i$th component $x_i$ and $y_i$ are denoted by $x$ and $y$, respectively. Moreover, the operator $T$ is defined by

$$T: \Xi \to \mathbb{R}^n$$
$$a \mapsto (T_1(a), \cdots, T_n(a)). \tag{8}$$

## III. Examples

In order to motivate the forthcoming analysis, we shall provide four examples of digital signal processing problems whose data generation model is that displayed in (4).

### A. Digital Signal Restoration

A common model assumes that the degraded signal is an observation of a data process $(X_n)_n$ obtained by convolving the original signal $h = (h_1, \cdots, h_q)$ with some blurring kernel $(t_{-l}, \cdots, t_l)$ and by addition of noise. The $n$th sample of the degraded signal is given by

$$X_n = \sum_{k=-l}^{l} t_k h_{n-k} + U_n. \tag{9}$$

The goal is to restore the original signal, i.e., to estimate $h$.

### B. Autoregressive Estimation

The $n$th data sample of an autoregressive information signal $(X_n)_n$ of order $q$ is given by

$$X_n = \sum_{k=1}^{q} h_k X_{n-k} + U_n \tag{10}$$

where $(U_n)_n$ is a random excitation sequence. The problem is then to identify the regression parameters, i.e., to estimate $h = (h_1, \cdots, h_q)$.

### C. Processing of Radar Signals

A typical sample of a returned radar signal can be written as

$$X_n = A \cos\left(\frac{2\pi(\nu + h_1)}{\tau}(n - h_2) + \phi\right) + U_n \quad (11)$$

where $A$ is the amplitude of the received signal, $\tau$ the sampling period of the receiver, $\nu$ the frequency of the transmitted signal, $h_1$ the Doppler shift induced by the motion of the target, $h_2$ some delay which is proportional to the distance to the target, $\phi$ some phase reference, and $(U_n)_n$ the noise process. Here, one seeks to estimate the velocity and the range of the target, i.e., $h = (h_1, h_2)$.

### D. Harmonic Retrieval

A model for the $n$th sample of a harmonic data process is

$$X_n = \sum_{k=1}^{q} b_k \sin(2\pi h_k n + \phi_k) + U_n \quad (12)$$

where $q$ is the number of sinusoids, $b_k$ their amplitude, $\phi_k$ their phase, and $(U_n)_n$ the noise process. The harmonic retrieval problem is to estimate the unknown frequencies, i.e., $h = (h_1, \cdots, h_q)$.

The general data formation model (4) is also encountered in problems such as system identification, parametric spectral estimation, signal reconstruction, and array processing.

## IV. Sets Based on Range Information

It is assumed that all the r.v.'s in the noise process $(U_n)_{n\in Z}$ are i.i.d., all distributed as a nondegenerate r.v. $U$ with known d.f. $F$. Let us fix a confidence coefficient $1 - \epsilon$ in $]0, 1[$. Then it is always possible to find two real numbers $\kappa$ and $\lambda$ such that

$$(1 - \epsilon)^{1/n} = \mathsf{P}\{\omega \in \Omega | \kappa \le U(\omega) \le \lambda\}. \quad (13)$$

Since $(Y_n)_{n\in Z}$ and $(U_n)_{n\in Z}$ are equivalent, with probability $1 - \epsilon$, all the points in the residual path should lie in the confidence interval $[\kappa, \lambda]$. The set of estimates which satisfy this constraint is

$$S_r = \bigcap_{i=1}^{n} \{a \in \Xi | \kappa \le x_i - T_i(a) \le \lambda\}. \quad (14)$$

*Proposition 2:* $S_r$ is closed in $\Xi$ if $T$ is continuous and $S_r$ is convex in $\Xi$ if $T$ is linear.

In cases where $F$ is not available but where, for some $p$ in $\mathbb{R}_+^*$, the $p$th absolute moment of $U$ is known, a confidence interval can be found by invoking Markov's inequality [23]

$$(\forall \lambda \in \mathbb{R}_+^*) \qquad \mathsf{P}\{\omega \in \Omega | |U(\omega)| > \lambda\} \le \lambda^{-p}\mathsf{E}|U|^p. \quad (15)$$

It is seen that, for a given confidence coefficient $1 - \epsilon$, all the residual samples should lie in the interval $[-\lambda, \lambda]$, where $\lambda = (\mathsf{E}|U|^p/(1 - (1 - \epsilon)^{1/n}))^{1/p}$. Finally, note that $S_r$ can be used with a 100% confidence coefficient provided the $U_n$'s are a.s. uniformly bounded, even if they are not i.i.d.

## V. Sets Based on Absolute Moment Information

It is assumed that the noise sequence $(U_n)_{n\in Z}$ consists of i.i.d.r.v.'s, all distributed as a nondegenerate r.v. $U$. Moreover, it is supposed that, for a fixed positive real number $p$, $U$ belongs to $L^{2p}(\mathsf{P})$, and the $p$th and $2p$th absolute moments of $U$ are known. The functional $N_p$ is defined as follows:

$$(\forall x \in \mathbb{R}^n) \qquad N_p(x) = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}. \quad (16)$$

Since the residual and the noise processes are equivalent, the $Y_n$'s are i.i.d.r.v.'s with $p$th absolute moment $\mathsf{E}|U|^p$. The $p$th sample absolute moment of the residual is defined as

$$M_p = \frac{1}{n}\sum_{i=1}^{n} |Y_i|^p. \quad (17)$$

Under the above hypotheses, as the sample size $n$ tends to infinity, $M_p$ is asymptotically normal with mean and standard deviation, respectively, given by [13]

$$\mathsf{E}M_p = \mathsf{E}|U|^p \quad \text{and} \quad \sigma_p = \sqrt{\frac{\mathsf{E}|U|^{2p} - \mathsf{E}^2|U|^p}{n}}. \quad (18)$$

Therefore, by invoking the limiting distribution, given a confidence coefficient, one can compute a confidence interval $[-\alpha, \alpha]$ for $(M_p - \mathsf{E}|U|^p)/\sigma_p$ from the tables of the normal distribution. The value of the sample absolute moment at the elementary event $\omega$ can be written as

$$M_p(\omega) = \frac{1}{n}\sum_{i=1}^{n} |y_i|^p = \frac{1}{n}N_p^p(y) = \frac{1}{n}N_p^p(x - T(a)). \quad (19)$$

Hence, after some algebra, the subset of $\Xi$ of estimates which yield an observed residual sample absolute moment within the desired confidence interval is found to be

$$S_p = \{a \in \Xi | \eta_p \le N_p(x - T(a)) \le \zeta_p\} \quad (20)$$

where

$$\eta_p = \begin{cases} n^{1/p}(\mathsf{E}|U|^p - \alpha\sigma_p)^{1/p} & \text{if } \mathsf{E}|U|^p > \alpha\sigma_p \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

and

$$\zeta_p = n^{1/p}(\mathsf{E}|U|^p + \alpha\sigma_p)^{1/p}. \quad (22)$$

Let $S_p^-$ denote the deficiency of $S_p$ and $S_p^+$ its hull, i.e.,

$$\begin{cases} S_p^- = \{a \in \Xi | N_p(x - T(a)) < \eta_p\} \\ S_p^+ = \{a \in \Xi | N_p(x - T(a)) \le \zeta_p\}. \end{cases} \quad (23)$$

We can write $S_p = S_p^+ - S_p^-$.

*Proposition 3:* $S_p$ and $S_p^+$ are closed in $\Xi$ if $T$ is continuous and $S_p^+$ is convex in $\Xi$ if $T$ is linear and $p \geq 1$.

When $n$ is small, the set $S_p^-$ is usually empty for realistic values of the confidence coefficient (say at least 90%) and we need not distinguish the absolute moment set $S_p$ from its hull $S_p^+$. Finally, let us note that in the particular case when $p = 2$ and $U$ is zero mean Gaussian, the exact distribution of $nM_p/\mathsf{E}|U|^2$ is a $\chi_n^2$ [13]. Thus, from the tables of the $\chi_n^2$, one can obtain a value of $\zeta_2$ which is more accurate than that resulting from the normal approximation, especially if $n$ is small.

## VI. Sets Based on Moment Information

It is assumed that the noise sequence $(U_n)_{n\in Z}$ consists of i.i.d.r.v.'s, all distributed as a nondegenerate r.v. $U$. It is also supposed that, for a fixed positive integer $k$, $U$ belongs to $L^{2k}(\mathsf{P})$, and that the $k$th and $2k$th moments of $U$ are known. In many instances, moments and absolute moments coincide and, therefore, the results of Section V apply. Such is the case when $U$ is nonnegative or when $k$ is even.

The $k$th sample moment of the residual process is defined as

$$M_k = \frac{1}{n}\sum_{i=1}^{n} Y_i^k. \tag{24}$$

Under our assumptions, $M_k$ is asymptotically normal with mean and standard deviation, respectively, given by [13]

$$\mathsf{E}M_k = \mathsf{E}U^k \quad \text{and} \quad \sigma_k = \sqrt{\frac{\mathsf{E}U^{2k} - \mathsf{E}^2U^k}{n}}. \tag{25}$$

Hence, given a confidence coefficient, a confidence interval $[-\alpha, \alpha]$ for $(M_k - \mathsf{E}U^k)/\sigma_k$ is determined by making the normal approximation. The set of estimates which yield an observed residual sample moment within this confidence interval is

$$S_k = \left\{a \in \Xi \,|\, \gamma_k \leq \sum_{i=1}^{n} (x_i - T_i(a))^k \leq \delta_k\right\} \tag{26}$$

where

$$\gamma_k = n(\mathsf{E}U^k - \alpha\sigma_k) \quad \text{and} \quad \delta_k = n(\mathsf{E}U^k + \alpha\sigma_k). \tag{27}$$

*Proposition 4:* $S_k$ is closed in $\Xi$ if $T$ is continuous and $S_1$ is convex in $\Xi$ if $T$ is linear.

*Proposition 5:* Let $F$ be the d.f. of $U$ and suppose that

$$(\forall c \in \mathbb{R}) \quad F(c) + F(-c) = 1 - \mathsf{P}\{\omega \in \Omega \,|\, U(\omega) = c\}. \tag{28}$$

Then, if $k$ is odd, $S_k$ can be written as

$$S_k = \left\{a \in \Xi \,\Big|\, \left|\sum_{i=1}^{n} (x_i - T_i(a))^k\right| \leq \alpha\sqrt{n\mathsf{E}U^{2k}}\right\}. \tag{29}$$

The conditions of Proposition 5 are satisfied, in particular, when the r.v. $U$ is absolutely continuous with an even density (e.g., zero mean uniform, Laplacian, or Gaussian).

## VII. Sets Based on Second-Order Information

In this section, second-order properties of the noise, i.e., properties which can be defined by means of its mixed second-order moments, are investigated (see Doob [12] for related definitions and results). The processes being real-valued, the spectral distributions are defined on $[0, 1/2]$. It is assumed that $n$ is even (if not, $n/2$ should be replaced by $(n-1)/2$ thereafter).

### A. The Case of Gaussian White Noise

Let us first recall a basic result of spectral analysis.

*Theorem 1 [26]:* Let $(Y_n)_{n\in Z}$ be a zero mean Gaussian discrete white noise process with power $\sigma^2$. Define

$$(\forall k \in \{0, \cdots, n/2\}) \quad I_k = \frac{2}{n}\left|\sum_{i=1}^{n} Y_i \exp\left(-j\frac{2\pi}{n}ki\right)\right|^2. \tag{30}$$

Then

i) $I_0, \cdots, I_{n/2}$ are independent r.v.'s.
ii) $I_0/2\sigma^2$ and $I_{n/2}/2\sigma^2$ have a $\chi_1^2$ distribution.
iii) $I_1/\sigma^2, \cdots, I_{n/2-1}/\sigma^2$ have a $\chi_2^2$ distribution.

If $(U_n)_{n\in Z}$ satisfies the assumptions of Theorem 1, so does $(Y_n)_{n\in Z}$. Hence, for a confidence coefficient $(1-\epsilon)^{2/(2+n)}$, from Theorem 1 and the tables of the $\chi_1^2$ and $\chi_2^2$ distributions, one can determine confidence intervals $[0, \beta_1]$ and $[0, \beta_2]$ for the r.v.'s in ii) and iii), respectively. The observed values of the periodogram are given by

$$(\forall k \in \{0, \cdots, n/2\}) \quad I_k(\omega) = \frac{2}{n}\left|\sum_{i=1}^{n} (x_i - T_i(a)) \exp\left(-j\frac{2\pi}{n}ki\right)\right|^2. \tag{31}$$

Consequently, the set of estimates which produce a residual path consistent, to within a $1-\epsilon$ confidence coefficient, with the whiteness and normality of the noise process is

$$S_d = \bigcap_{k=0}^{n/2} \left\{a \in \Xi \,\Big|\, \left|\sum_{i=1}^{n} (x_i - T_i(a)) \cdot \exp\left(-j\frac{2\pi}{n}ki\right)\right|^2 \leq \xi_k\right\} \tag{32}$$

where

$$(\forall k \in \{0, \cdots, n/2\}) \quad \xi_k = \begin{cases} n\sigma^2\beta_1 & \text{if } k = 0 \text{ or } n/2 \\ \frac{n}{2}\sigma^2\beta_2 & \text{if } 0 < k < n/2. \end{cases} \tag{33}$$

*Proposition 6:* $S_d$ is closed in $\Xi$ if $T$ is continuous and $S_d$ is convex in $\Xi$ if $T$ is linear.

### *B. The Case of Non-Gaussian White Noise*

Suppose that $(U_n)_{n\in Z}$ is a discrete white noise process consisting of i.i.d.r.v.'s all distributed as a zero mean r.v. $U$ in $L^4(\mathrm{P})$, with variance $\sigma^2$. Then the r.v.'s in ii) and iii) of Theorem 1 are symptotically distributed as a $\chi_1^2$ and a $\chi_2^2$, respectively [19]. Thus, under relatively mild conditions, the conclusions of Theorem 1 hold in an asymptotic sense. Consequently, provided that $n$ is large, the set $S_d$ of Section VII-A can be used.

### *C. The General Case of Correlated Noise*

In this section, we further generalize the analysis by dropping the whiteness assumption. We shall base the construction of a spectral set in this case on the following theorem.[2]

*Theorem 2 [28]:* Let $(Y_n)_{n\in Z}$ be a zero mean strictly stationary strongly mixing process with summable second- and fourth-order cumulant functions and spectral density $g$. Let $0 = \nu_0 < \nu_1 < \cdots < \nu_m = \frac{1}{2}$ and

$$(\forall k \in \{0, \cdots, m\}) \qquad I_k = \frac{2}{n}\left|\sum_{i=1}^{n} Y_i \exp(-j2\pi\nu_k i)\right|^2. \tag{34}$$

Then

i) $I_0, \cdots, I_m$ are asymptotically independent r.v.'s.
ii) $I_0/g(0)$ and $I_m/g(\frac{1}{2})$ are asymptotically distributed as a $\chi_1^2$.
iii) $2I_1/g(\nu_1), \cdots, 2I_{m-1}/g(\nu_{m-1})$ are asymptotically distributed as a $\chi_2^2$.

Loosely speaking, Theorem 2 states that if the span of dependence of the process is small enough, the results of Theorem 1 can be generalized for large $n$. Now suppose that $(U_n)_{n\in Z}$ satisfies the hypotheses of Theorem 2 and that its spectral density $g$ is known at points $0 \le \nu_0 < \nu_1 < \cdots < \nu_m \le \frac{1}{2}$. Then, $(U_n)_{n\in Z}$ and $(Y_n)_{n\in Z}$ being equivalent, given a confidence coefficient, one can compute the confidence intervals $[0, \beta_1]$ and $[0, \beta_2]$ for the r.v.'s in ii) and iii), respectively, by invoking their asymptotic properties. This leads to the set

$$S_d = \bigcap_{k=0}^{m}\left\{a \in \Xi \,\middle|\, \left|\sum_{i=1}^{n}(x_i - T_i(a)) \cdot \exp(-j2\pi\nu_k i)\right|^2 \le \xi_k\right\} \tag{35}$$

with

$$\xi_0 = \begin{cases} \frac{n}{2}\, g(0)\beta_1 & \text{if } \nu_0 = 0 \\ \frac{n}{4}\, g(\nu_0)\beta_2 & \text{if } \nu_0 > 0 \end{cases} \tag{36.a}$$

$$(\forall k \in \{1, \cdots, m-1\}) \qquad \xi_k = \frac{n}{4}\, g(\nu_k)\beta_2 \tag{36.b}$$

$$\xi_m = \begin{cases} \frac{n}{4}\, g(\nu_m)\beta_2 & \text{if } \nu_m < \frac{1}{2} \\ \frac{n}{2}\, g(\frac{1}{2})\beta_1 & \text{if } \nu_m = \frac{1}{2}. \end{cases} \tag{36.c}$$

It is noted that Proposition 6 still holds.

In instances when no knowledge of the spectral distribution of the noise is available, second-order properties can be enforced if the value of the correlation function, $r(\cdot)$, of the noise is known at some lag $m$. Let $R_m$ be the sample correlation coefficient of the residual for lag $m$. Then the residual path should yield an observed value of $R_m$ within some confidence interval around $r(m)$ determined by the d.f. of $R_m$. Unfortunately, the distribution theory for the sample correlations is extremely complicated. Under relatively involved assumptions, it can be shown that the r.v.'s defined by

$$(\forall m \in \{0, \cdots, n-1\}) \qquad R_m = \frac{1}{n}\sum_{i=1}^{n-m} Y_i Y_{i+m} \tag{37}$$

are asymptotically jointly Gaussian [28]. Exact results have also been established for low order correlation lags of special processes [26]. For an approximate general result based on simpler assumptions, we now follow [26]. Suppose $(U_n)_{n\in Z}$ is a zero mean wide-sense stationary process. Let us define the normalized correlation function of the equivalent process $(Y_n)_{n\in Z}$ as $\bar{r}(\cdot) = r(\cdot)/r(0)$ and its normalized sample correlation function as

$$(\forall m \in \{-n+1, \cdots, n-1\}) \qquad \bar{R}_m = \frac{\sum_{i=1}^{n-|m|} Y_i Y_{i+|m|}}{\sum_{i=1}^{n} |Y_i|^2}. \tag{38}$$

Now suppose that $|\bar{r}(i)|$ goes to zero as $|i|$ goes to infinity. Then a crude estimate for the asymptotic distribution of $\bar{R}_m$ is a normal distribution with mean $\bar{r}(m)$ and variance

$$\sigma_r^2 = \frac{1}{n}\sum_{m=1-n}^{n-1}\left|\frac{\sum_{i=1}^{n-|m|} y_i y_{i+|m|}}{\sum_{i=1}^{n}|y_i|^2}\right|^2. \tag{39}$$

It follows that the set of estimates which produce a resid-

[2]A strictly stationary process $(Y_n)_{n\in Z}$ is strongly mixing if $\lim_{n\to+\infty} \sup \{|\mathrm{P}A \cap B - \mathrm{P}A\mathrm{P}B| \,|\, A \in \Sigma_0^-, B \in \Sigma_n^+\} = 0$, where $\Sigma_n^-$ and $\Sigma_n^+$ are the sub $\sigma$-algebras of $\Sigma$ generated by $(Y_i)_{i\le n}$ and $(Y_i)_{i\ge n}$, respectively.

ual consistent with $\bar{r}(m)$ is

$$S_c = \left\{ a \in \Xi \left| \left| \frac{\sum_{i=1}^{n-|m|} (x_i - T_i(a))(x_{i+|m|} - T_{i+|m|}(a))}{\sum_{i=1}^{n} |x_i - T_i(a)|^2} - \bar{r}(m) \right| \leq \alpha\sigma_r \right. \right\}. \tag{40}$$

Again, $\alpha$ is determined from the tables of the normal distribution in accordance with some confidence coefficient. Finally, let us mention that in the particular case when the $U_n$'s are i.i.d.r.v.'s in $L^2(\mathsf{P})$, if $m \neq 0$, $\bar{r}(m) = 0$ and (39) may be reduced to $\sigma_r^2 = 1/n$ [26].

## VIII. Sets Based on Higher Order Information

If a process $(U_n)_{n \in Z}$ is Gaussian, all of the information about $(U_n)_{n \in Z}$ is contained in the means and the second-order mixed moments. First- and second-order probabilistic attributes, however, do not provide a complete description of non-Gaussian (in particular nonlinear) processes. In this section, we indicate how to construct sets based on information available via mixed moments of order greater than two. Thereafter, the term polyspectral estimate refers to the class of asymptotically normal smoothed higher order periodograms discussed in [28] (for an account of the Fourier analysis of higher order cumulants, see [28]).

### A. The Normality Set

If $(U_n)_{n \in Z}$ is a Gaussian process, all of its cumulants of order greater than two are zero and it follows that all of its spectral densities of higher order than the second vanish. Thus, for every integer $k$ greater than two, the observed value of the $k$th order polyspectral estimate based on the residual path $y$ should be within some interval around zero. Given a confidence coefficient, the bounds of this interval can be computed by invoking the asymptotic normal distribution of the polyspectral estimate. The corresponding set is that of all $a$'s which yield a residual whose observed polyspectral estimate falls in the confidence interval.

### B. The Linearity Set

Here, we use the linearity test of [33]. Suppose that $(U_n)_{n \in Z}$ is a linear process, i.e.,

$$(\forall n \in \ ) \quad U_n = \sum_{i=-\infty}^{+\infty} b_i V_{n-i} \quad \text{with} \sum_{i=-\infty}^{+\infty} b_i^2 < +\infty \tag{41}$$

where the $V_n$'s are zero mean i.i.d.r.v.'s in $L^3(\mathsf{P})$ with $\mathsf{E}V_n^2 = \mu_2$ and $\mathsf{E}V_n^3 = \mu_3$. Then, if $g(\cdot)$ and $g(\cdot, \cdot)$ denote respectively the spectral and the bispectral density of $(U_n)_{n \in Z}$, it can be shown that for every frequency $\nu_1$ and $\nu_2$

$$z(\nu_1, \nu_2) = \frac{|g(\nu_1, \nu_2)|^2}{g(\nu_1)g(\nu_2)g(\nu_1 + \nu_2)} = \frac{\mu_3^2}{\mu_2}. \tag{42}$$

Let $\hat{z}(\nu_1, \nu_2)$ be the estimate of $z(\nu_1, \nu_2)$ computed from $y$. Then, the constancy of $\hat{z}(\nu_1, \nu_2)$ over some grid $G$ in the frequency plane can be used to test the linearity of the residual process. The corresponding set is that of all $a$'s which produce a residual such that $\hat{z}(\nu_1, \nu_2)$ is within some confidence interval around the constant $\mu_3^2/\mu_2$ for all $(\nu_1, \nu_2)$ in $G$.

### C. The Independence Set

Suppose that $(U_n)_{n \in Z}$ consists of i.i.d.r.v.'s distributed as a r.v. $U$ whose cumulants at all order exist and are finite, the $k$th being denoted by $c_k$. The cumulant function of this process is given by

$$(\forall (n_1, \cdots, n_k) \in {}^{k}) \quad c(n_1, \cdots, n_k) = \begin{cases} c_k & \text{if } n_1 = \cdots = n_k = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{43}$$

Hence, the real part of the $k$th order spectral density is the constant $c_k$ and the imaginary part is zero [3]. It is noted that the flatness of the spectral density merely translates the uncorrelatedness of the $U_n$'s. On the other hand, the independence of the $U_n$'s shows up in the flatness of the higher order spectral densities. For a given $k$, the corresponding set is that of all $a$'s which produce a residual whose $k$th order polyspectral estimate falls within some confidence interval around the expected constant value. Again, given a confidence coefficient, the bounds of this interval can be computed by invoking the asymptotic normal distribution of the polyspectral estimate.

### D. The Reversibility Set

Suppose that $(U_n)_{n \in Z}$ is time reversible, meaning that the finite dimensional d.f.'s of $(U_{-n})_{n \in Z}$ are the same as that of $(U_n)_{n \in Z}$. Then its cumulant function satisfies [3]

$$(\forall (n_1, \cdots, n_k) \in {}^{k}) \quad c(n_1, \cdots, n_k) = c(-n_1, \cdots, -n_k) \tag{44}$$

and the imaginary part of the $k$th order spectral density is therefore identically zero. For a given $k$, the corresponding set is that of all $a$'s which produce a residual whose observed polyspectral estimate has an imaginary part in some confidence interval (determined as above) around zero.

## IX. Sets Based on Other Information

In statistics, procedures exist to test various hypotheses concerning a discrete stochastic process. Therefore, if $\Psi_j$ is a piece of *a priori* knowledge pertaining to the noise process, we can construct the property set $S_j$ of all $a$'s which yield a residual path $y$ that passes the test associated with $\Psi_j$ within some significance level. For instance, if it is known that $(U_n)_{n \in Z}$ is stationary, the procedure described in [27] can be used to test the residual path for stationarity; if it is known that the $U_n$'s are i.i.d., the tests of randomness discussed in [20] can be applied to the residual path. In other instances, the $U_n$'s may all be known to be distributed as a r.v. $U$ with d.f. $F$, some attributes of which are known. Thus, the residual path can be tested for departure from various functional forms of $F$ (e.g., uniformity [22], normality [24], or exponentially [32]) or from some of its properties (e.g., symmetry [8]).

## X. Redundancy of Noise Information

In some problems, noise information is scarce and few property sets can be constructed. Nonetheless, there are many situations in which several pieces of *a priori* information about the noise process are available and, naturally, the question of redundancy arises. To illustrate this point, consider a problem where it is known that the noise sequence consists of zero mean Gaussian i.i.d.r.v.'s. From the knowledge of some absolute moment of order $p$, one can infer the value of any other absolute moment and, therefore, infinitely many absolute moment sets can be created. One has then to decide which of these sets should be used.

In our framework, a piece of information $\Psi_j$ is redundant in the presence of a piece of information $\Psi_i$ if the property sets $S_i$ and $S_j$ formed in accordance with (1) are such that $S_i \subset S_j$. In other words, in the presence of $S_i$, $S_j$ does not contribute to a smaller feasibility set in (2). In general, the question of identifying analytically redundant information is a delicate one. A partial answer is brought by the following propositions where $S_r$ denotes the range set of Section IV, $S_p$ the $p$th absolute moment set of Section V, $S_k$ the $k$th moment set of Section VI, and $S_d$ the spectral set of Sections VII-A and B.

*Proposition 7:* If $k$ is odd, $S_r \subset S_k$ if $n\kappa^k \geq \gamma_k$ and $\delta_k \geq n\lambda^k$. $S_r \subset S_p^+$ if $\zeta_p \geq n^{1/p} \max\{|\kappa|, |\lambda|\}$.

*Proposition 8:* Let $p \leq q$ be two positive real numbers. Then $S_q^+ \subset S_p^+$ if $\zeta_q \leq n^{1/q - 1/p}\zeta_p$.

*Proposition 9:* For $k = 1$, $S_d \subset S_k$ if $\gamma_1 \leq -\sqrt{\xi_0}$ and $\delta_1 \geq \sqrt{\xi_0}$.

It is worth noting that Proposition 9 also holds for the spectral set of Section VII-C if $\nu_0 = 0$. It is also pointed out that Propositions 7 through 9 involve merely sufficient conditions. Consequently, they may fail to detect redundancy in some cases. In such instances, the selection of noise property sets should be guided by experience with similar problems.

## XI. Simulation Results

The goal of these simulations is not to demonstrate the usefulness of the sets constructed above, since particular cases of these sets have already been successfully employed in previous studies. For instance, sets similar to $S_p$ (for $p = 2$), $S_k$ (for $k = 1$), and $S_d$ have been used in set theoretic restoration [5], [35]. Rather, the intent here is to illustrate some of the general statements made earlier via direct pictorial representation of some of the sets. This will also provide insight into the actual value of various pieces of noise information.

In order to give a pictorial representation of the sets in the Cartesian plane, estimation problems with two real unknown parameters $h_1$ and $h_2$ are considered: autoregressive estimation (problem A) and harmonic retrieval (problem B). In each problem, points $(x_i)_i$ of a data process of the form (4) are generated. To represent a given noise property set $S_j$ in the $(h_1, h_2)$ solution plane, the following Monte Carlo experiment is performed. A large number of estimates of $h = (h_1, h_2)$, $a = (a_1, a_2)$, are drawn at random from a uniform distribution over some compact region $G$, called search region. For each $a$, the residual path is computed according to $y_i = x_i - T_i(a)$. A simple acceptance/rejection procedure then takes place. Only those $a$'s which produce a residual path consistent with the noise property in question are retained. The scatter plot of these points in the solution plane represents approximately the set $S_j$. In the simulations the noise sequence consists of zero mean i.i.d.r.v.'s distributed as a r.v. $U$ with known density type (uniform in problem A and Gaussian in problem B) and variance. This information is sufficient to compute the parameters of all the sets constructed above. The number of residual samples is $n = 16$. By centering the confidence interval at zero, the range set of Section IV can be written as

$$S_r = \bigcap_{i=1}^{n} \{a \in \mathbb{R}^2 | \, |x_i - T_i(a)| \leq \lambda\}. \quad (45)$$

From Section V, the absolute moment set of order $p$ reads

$$S_p = \{a \in \mathbb{R}^2 | \eta_p \leq N_p(x - T(a)) \leq \zeta_p\}. \quad (46)$$

By Proposition 5, the expression of the moment set of odd order $k$ is

$$S_k = \left\{a \in \mathbb{R}^2 \left| \, \left|\sum_{i=1}^{n} (x_i - T_i(a))^k\right| \leq \alpha\sqrt{n\mathsf{E}U^{2k}}\right.\right\}. \quad (47)$$

Finally, the spectral set of Sections VII-A and B is given by

$$S_d = \bigcap_{k=0}^{n/2} \left\{a \in \mathbb{R}^2 \left| \, \left|\sum_{i=1}^{n} (x_i - T_i(a)) \cdot \exp\left(-j\frac{2\pi}{n}ki\right)\right|^2 \leq \xi_k\right.\right\}. \quad (48)$$

For both problems, we shall represent $S_r$, $S_p$ for $p = \frac{1}{2}$, 1, 2, and 4, $S_k$ for $k = 1$ and $k = 3$, and $S_d$. The confidence

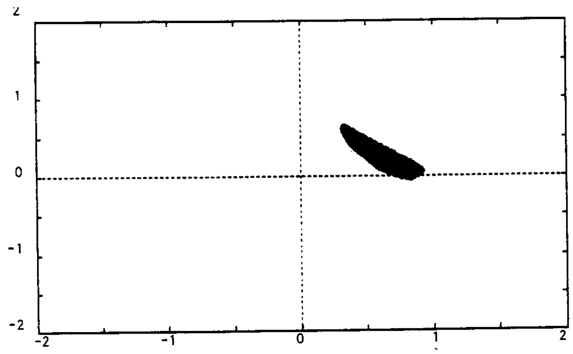



Fig. 1. AR estimation problem: Range set.

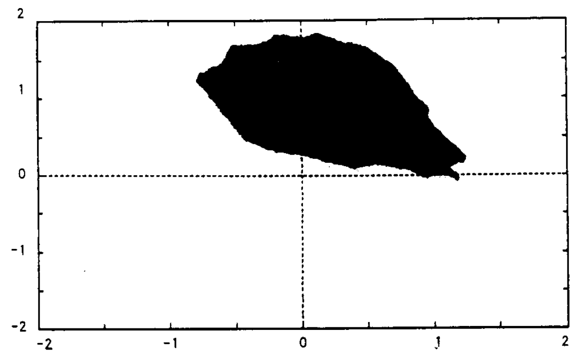



Fig. 2. AR estimation problem: Absolute moment set—$p = \frac{1}{2}$.

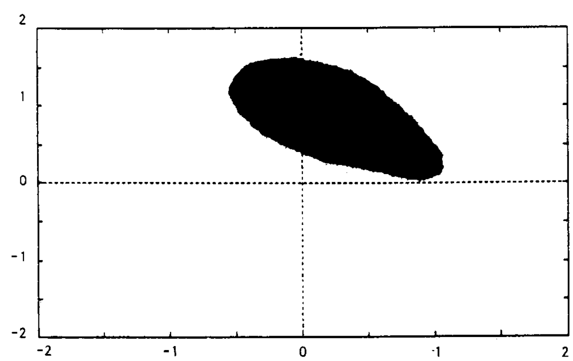



Fig. 3. AR estimation problem: Absolute moment set—$p = 1$.

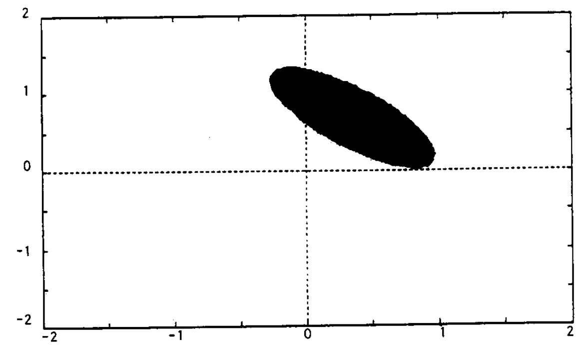



Fig. 4. AR estimation problem: Absolute moment set—$p = 2$.

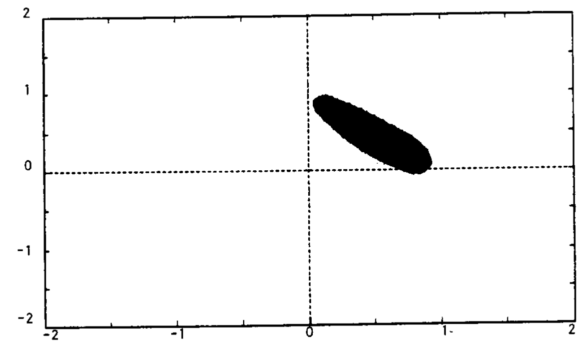



Fig. 5. AR estimation problem: Absolute moment set—$p = 4$.

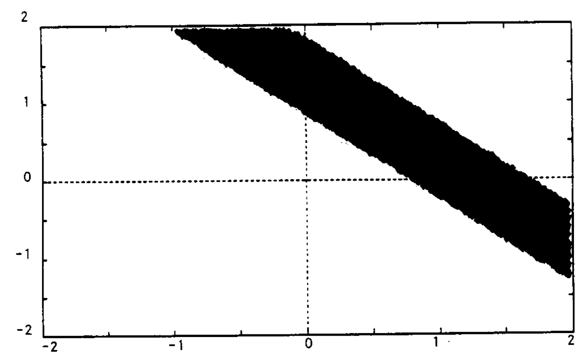



Fig. 6. AR estimation problem: Moment set—$k = 1$.

coefficient is fixed to 95% and, since $n$ is small, $S_p = S_p^+$.

### A. Autoregressive Estimation

We consider a second-order AR process with two unknown coefficients $h_1$ and $h_2$, driven by a noise sequence $(U_n)_{n \in Z}$. The signal formation operator is given by

$$T_i(h) = h_1 x_{i-1} + h_2 x_{i-2}. \tag{49}$$

The true regression coefficients are $h_1 = 0.45$ and $h_2 = 0.50$; $U$ is uniformly distributed with variance $\sigma^2 = 0.05$. The search region is $G = [-2, 2] \times [-2, 2]$. The range set is displayed in Fig. 1, the absolute moment sets in Figs. 2-5, the moment sets in Figs. 6, 7, and the spectral set in Fig. 8.

### B. Harmonic Retrieval

We now consider a harmonic data process with two phaseless sinusoids of unit amplitude and unknown frequencies $h_1$ and $h_2$ corrupted by a noise process $(U_n)_{n \in Z}$. For such a problem, the signal formation operator is non-

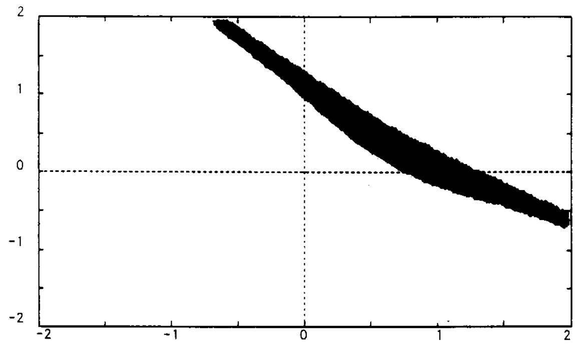



Fig. 7. AR estimation problem: Moment set—$k = 3$.

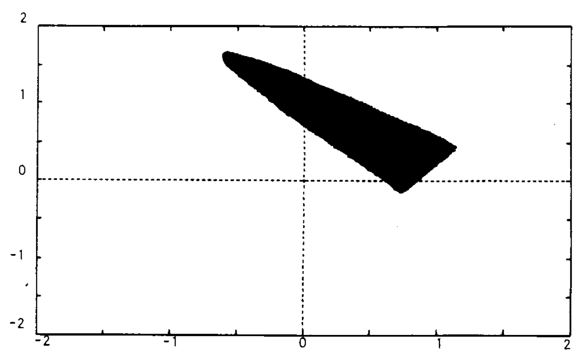



Fig. 8. AR estimation problem: Spectral set.

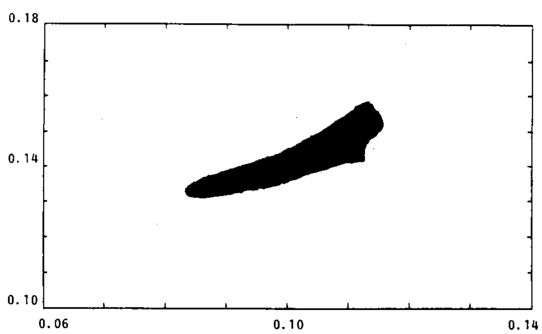



Fig. 9. Harmonic retrieval problem: Range set.

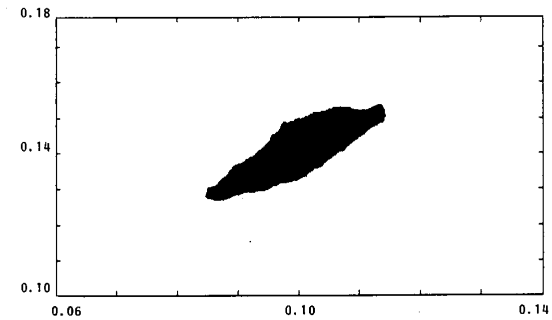



Fig. 10. Harmonic retrieval problem: Absolute moment set—$p = \frac{1}{2}$.

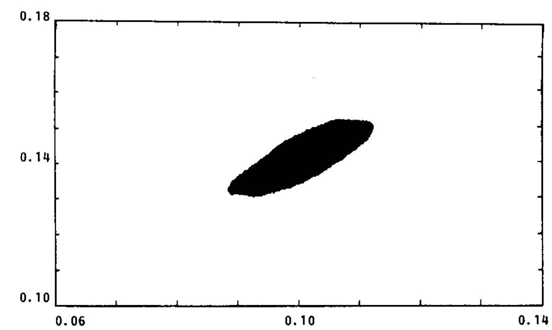



Fig. 11. Harmonic retrieval problem: Absolute moment set—$p = 1$.

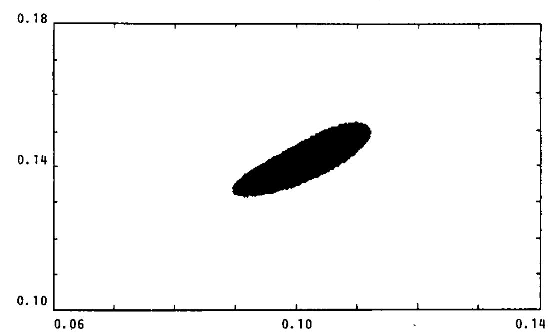



Fig. 12. Harmonic retrieval problem: Absolute moment set—$p = 2$.

linear

$$T_i(h) = \sin(2\pi h_1 i) + \sin(2\pi h_2 i). \tag{50}$$

The true frequencies are $h_1 = 0.10$ and $h_2 = 0.14$; $U$ is normally distributed with variance $\sigma^2 = 0.05$. The search region is $G = [0, \frac{1}{2}] \times [0, \frac{1}{2}]$. The range set is displayed in Fig. 9, the absolute moment sets in Figs. 10–13, the moment sets in Figs. 14 and 15, and the spectral set in Fig. 16 (since the problem is symmetric in $a_1$ and $a_2$, only the points with $a_2 > a_1$ need be represented).

### C. Discussion

In problem A, the signal formation operator is linear. From Figs. 1, 3 through 6, and 8, it is seen that the range set, the absolute moment sets (for $p \geq 1$), the moment set of order 1, and the spectral set are convex, which agrees with Propositions 2–4, and 6, respectively. In problem B the signal formation operator is nonlinear. This gives rise to sets which are geometrically more complex such as the moment sets in Figs. 14 and 15. Note also that other sets are nonconvex such as the absolute moment sets of order $\frac{1}{2}$ in Figs. 2 and 10, the third moment set of Fig. 7, the range set in Fig. 9. As was indicated in Section X, Propositions 7 to 9 may fail to detect some redundancies. For instance, the range set of Fig. 1 is contained in the absolute moment set of order 4 of Fig. 5 while $\zeta_4 < n^{1/4}\lambda$ in problem A. It should also be noted that moment sets of

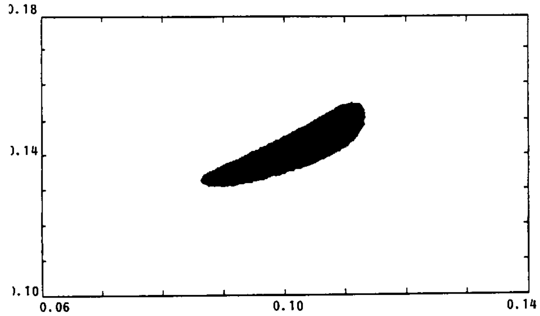



Fig. 13. Harmonic retrieval problem: Absolute moment set—$p = 4$.

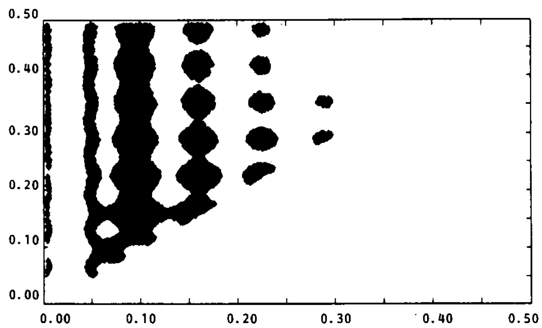



Fig. 14. Harmonic retrieval problem: Moment set—$k = 1$.

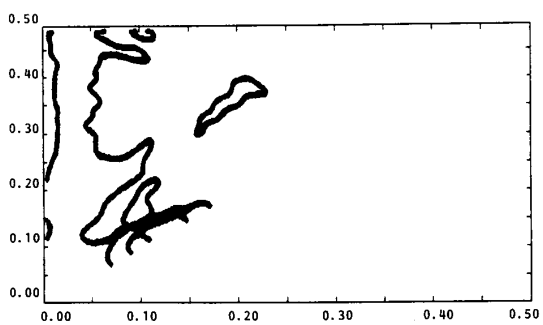



Fig. 15. Harmonic retrieval problem: Moment set—$k = 3$.

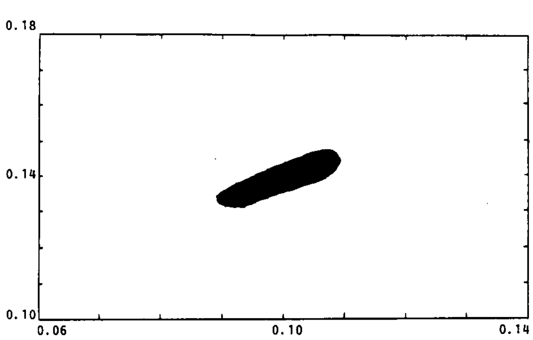



Fig. 16. Harmonic retrieval problem: Spectral set.

different order may carry nonredundant information. The same remark also applies to absolute moment sets. As a general conclusion, these simulations show that one type of information (range, moment, absolute moment, spectral density) is not necessarily redundant with another type. Therefore, leaving aside any computational aspect, one should seek to use as much of the available noise information as possible to reduce the feasibility set.

## XII. Conclusions

In the literature, the use of noise properties in estimation procedures has been limited to few probabilistic attributes and to particular problems such as system identification and signal restoration. In this paper, it has been shown how a wide range of information relative to the noise can be incorporated in a general set theoretic estimation problem. The incorporation of noise information, which is seldom possible with conventional estimation techniques, is very simple in set theoretic estimation as it amounts to adding the corresponding property sets to the collection of sets describing the solution. The addition of these sets will translate into a smaller feasibility set and, thereby, more reliable estimates.

Among the pieces of information considered to construct property sets were the range of the r.v.'s, the moments and absolute moments of arbitrary order, the correlation function at some lags, and the spectral density at some points. It was also indicated how to create sets based on properties available through higher order spectral densities and, via the use of statistical tests, on other probabilistic information. Properties of these sets such as closedness and convexity were also discussed and the question of redundant information was addressed. Because of the generality of the data formation model used in the analysis, the results presented here can be applied to a vast body of estimation problems in digital signal processing and systems theory, in particular to those mentioned in Section III.

## Appendix

The basic elements of mathematical analysis needed in the following proofs can be found in Berge [2] and Dieudonné [11].

*Proof of Proposition 1:* It simply needs to be checked that $(Y_n)_{n \in Z}$ and $(U_n)_{n \in Z}$ have the same finite dimensional d.f.'s. This is easily done by letting $k$ be an arbitrary but fixed positive integer and then fixing arbitrary $k$-tuples $(n_1, \cdots, n_k)$ in $\mathbb{Z}^k$ and $(\kappa_1, \cdots, \kappa_k)$ in $\mathbb{R}^k$. Consider the events

$$\begin{cases} A = \bigcap_{i=1}^{k} A_i \quad \text{where } A_i = \{\omega \in \Omega \mid Y_{n_i}(\omega) = U_{n_i}(\omega)\} \\ B = \bigcap_{i=1}^{k} \{\omega \in \Omega \mid Y_{n_i}(\omega) < \kappa_i\} \\ C = \bigcap_{i=1}^{k} \{\omega \in \Omega \mid U_{n_i}(\omega) < \kappa_i\}. \end{cases} \tag{A.1}$$

Note that, since $(Y_n)_{n\in Z}$ and $(U_n)_{n\in Z}$ are equivalent, $\mathsf{P}A^c = \mathsf{P} \bigcup_{i=1}^{k} A_i^c \le \Sigma_{i=1}^{k} \mathsf{P}A_i^c = 0$. Consequently, $\mathsf{P}A^c \cap B = \mathsf{P}A^c \cap C = 0$. Moreover, $A \cap B = A \cap C$. Hence,

$$\begin{aligned} \mathsf{P}B &= \mathsf{P}A \cap B + \mathsf{P}A^c \cap B \\ &= \mathsf{P}A \cap C + \mathsf{P}A^c \cap C = \mathsf{P}C \end{aligned} \tag{A.2}$$

and the proof is complete. □

*Proof of Proposition 2:* Let $i$ be an arbitrary integer in $\{1, \cdots, n\}$ and let $f_i$: $a \mapsto x_i - T_i(a)$. Since $T_i$ is continuous, so is $f_i$. Hence, each $C_i = f_i^{-1}([\kappa, \lambda])$ is closed in $\Xi$ and so is their intersection $S_r$. To prove the second assertion, let $a$ and $b$ be two arbitrary vectors in $S_r$ and let $\alpha$ be an arbitrary real number in $]0, 1[$. Clearly, $a$ and $b$ both belong to $C_i$ and thus

$$\begin{cases} \alpha\kappa \le \alpha(x_i - T_i(a)) \le \alpha\lambda \\ (1-\alpha)\kappa \le (1-\alpha)(x_i - T_i(b)) \le (1-\alpha)\lambda \end{cases} \tag{A.3}$$

Since $T_i$ is linear, (A.3) yields $\kappa \le x_i - T_i(\alpha a + (1 - \alpha)b) \le \lambda$. Therefore, each $C_i$ is convex and so is their intersection $S_r$. □

*Proof of Proposition 3:* Let $f_p$: $a \mapsto N_p(x - T(a))$. If $T$ is continuous, so is $f_p$. Therefore $S_p = f_p^{-1}([\eta_p, \zeta_p])$ and $S_p^+ = f_p^{-1}(]-\infty, \zeta_p])$ are closed. To establish that $S_p^+$ is convex, it is enough to prove that $f_p$ is a convex function. Let $\alpha$ be an arbitrary real number in $]0, 1[$ and let $a$ and $b$ be two arbitrary vectors in $\Xi$. By linearity of $T$

$$\begin{aligned} &f_p(\alpha a + (1-\alpha)b) \\ &\quad = N_p(\alpha(x - T(a)) + (1-\alpha)(x - T(b))). \end{aligned} \tag{A.4}$$

But, for $p \ge 1$, $N_p$ is a norm on $\mathbb{R}^n$ and it is therefore convex. Whence

$$\begin{aligned} &N_p(\alpha(x - T(a)) + (1-\alpha)(x - T(b))) \\ &\quad \le \alpha N_p(x - T(a)) + (1-\alpha)N_p(x - T(b)). \end{aligned} \tag{A.5}$$

Consequently,

$$f_p(\alpha a + (1-\alpha)b) \le \alpha f_p(a) + (1-\alpha)f_p(b) \tag{A.6}$$

which is the desired result. □

*Proof of Proposition 4:* Similar to that of Proposition 2. □

*Proof of Proposition 5:* The odd moments of a r.v. $U$ satisfying (28) are zero [13]. Therefore (26) reduces to (29). □

*Proof of Proposition 6:* Let $f_k$: $a \mapsto |\Sigma_{i=1}^{n}(x_i - T_i(a)) \exp(-j(2\pi/n)ki)|^2$, for an arbitrary $k$ in $\{0, \cdots, n/2\}$. If $T$ is continuous so is $f_k$. Therefore, each $D_k = f_k^{-1}(]-\infty, \xi_k])$ is closed in $\Xi$ and so is their intersection $S_d$. To prove the second assertion, it is enough to show that $f_k$ is convex (if it is, then $D_k = f_k^{-1}(]-\infty, \xi_k])$ will be convex and so will the intersection $S_d$). Let $\alpha$ be an arbitrary real number in $]0, 1[$ and let $a$ and $b$ be two arbitrary vectors in $\Xi$. For every $i$ in $\{1, \cdots, n\}$ let $w_{ik} = \exp(-j(2\pi/n)ki)$. Then, by linearity of $T_i$

$$\begin{aligned} &f_k(\alpha a + (1-\alpha)b) \\ &\quad = \left| \alpha \sum_{i=1}^{n} (x_i - T_i(a))w_{ik} \right. \\ &\qquad \left. + (1-\alpha) \sum_{i=1}^{n} (x_i - T_i(b))w_{ik} \right|^2. \end{aligned} \tag{A.7}$$

Since $z \mapsto |z|^2$ is a convex function on it follows that

$$f_k(\alpha a + (1-\alpha)b) \le \alpha f_k(a) + (1-\alpha)f_k(b). \tag{A.8}$$

Thus, $f_k$ is convex. □

*Proof of Proposition 7:* Let $a$ be an arbitrary point in $S_r$ and $i$ an arbitrary point in $\{1, \cdots, n\}$. Then, since $k$ is odd, $n\kappa^k \le \Sigma_{i=1}^{n} (x_i - T_i(a))^k \le n\lambda^k$. Thus, $n\kappa^k \ge \gamma_k$ and $\delta_k \ge n\lambda^k$ yield $a \in S_k$. On the other hand, $|x_i - T_i(a)|^p \le \max\{|\kappa|^p, |\lambda|^p\}$. Summing over all $i$'s and taking the $p$th root gives $N_p(x - T(a)) \le n^{1/p} \max\{|\kappa|, |\lambda|\}$. Hence $n^{1/p} \max\{|\kappa|, |\lambda|\} \le \zeta_p$ yields $a \in S_p^+$. □

*Proof of Proposition 8:* Let $a$ be an arbitrary point in $S_q$ and let $y = x - T(a)$. Then $N_q(y) \le \zeta_q$. But $N_p(y) \le n^{1/p - 1/q}N_q(y)$ [16]. Hence, $N_p(y) \le n^{1/p - 1/q}\zeta_q$. By hypothesis, $n^{1/p - 1/q}\zeta_q \le \zeta_p$. Hence $N_p(y) \le \zeta_p$. Therefore $a \in S_p^+$. □

*Proof of Proposition 9:* Let $a$ be an arbitrary point in $S_d$. Then certainly $-\sqrt{\xi_0} \le \Sigma_{i=1}^{n} x_i - T_i(a) \le \sqrt{\xi_0}$. Thus, the hypotheses $\gamma_1 \le -\sqrt{\xi_0}$ and $\delta_1 \ge \sqrt{\xi_0}$ give $a \in S_k$ for $k = 1$. □

## References


[1] A. Abo-Taleb and M. M. Fahmy, "Design of FIR two-dimensional digital filters by successive projections," *IEEE Trans. Circuits Syst.*, vol. CAS-31, no. 9, pp. 801-805, Sept. 1984.

[2] C. Berge, *Espaces Topologiques, Fonctions Multivoques*, second ed. Paris: Dunod, 1966.

[3] D. R. Brillinger and M. Rosenblatt, "Computation and interpretation of $k$th order spectra," in *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, pp. 189-232, 1967.

[4] J. M. Carazo and J. L. Carrascosa, "Information recovery in missing angular data cases: An approach by the convex projections method in three dimensions," *J. Microscopy*, vol. 145, pt. 1, pp. 23-43, Jan. 1987.

[5] P. L. Combettes and H. J. Trussell, "Methods for digital restoration of signals degraded by a stochastic impulse response," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 3, pp. 393-401, Mar. 1989.

[6] P. L. Combettes and H. J. Trussell, "Method of successive projections for finding a common point of sets in metric spaces," *J. Optimiz. Theory Appl.*, vol. 67, no. 3, pp. 487-507, Dec. 1990.

[7] P. L. Combettes and H. J. Trussell, "Set theoretic estimation by random search," *IEEE Trans. Signal Processing*, this issue, pp. 1669-1671.

[8] S. Csörgö and C. R. Heathcote, "Testing for symmetry," *Biometrika*, vol. 74, no. 1, pp. 177-184, Mar. 1987.

[9] J. R. Deller, Jr., and T. C. Luk, "Linear prediction analysis of speech based on set-membership theory," *Comput. Speech, Language*, vol. 3, no. 4, pp. 301-327, Oct. 1989.

[10] J. R. Deller, Jr., "Set membership identification in digital signal processing," *IEEE ASSP Mag.*, vol. 6, no. 4, pp. 4-20, Oct. 1989.

[11] Jean Dieudonné, *Foundations of Modern Analysis*, second ed. New York: Academic, 1969.

[12] Joe Doob, *Stochastic Processes*. New York: Wiley, 1953.

[13] M. Fisz, *Probability Theory and Mathematical Statistics*, third ed. New York: Wiley, 1963.

[14] E. Fogel, "System identification via membership set constraints with energy constrained noise," *IEEE Trans. Automat. Contr.*, vol. AC-24, no. 5, pp. 752-758, Oct. 1979.

[15] E. Fogel and Y. F. Huang, "On the value of information in system identification—bounded noise case," *Automatica*, vol. 18, no. 2, pp. 229-238, Mar. 1982.

[16] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, second ed. Cambridge, England: Cambridge University Press, 1952.

[17] Y. F. Huang, "A recursive estimation algorithm using selective updating for spectral analysis and adaptive signal processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 5, pp. 1331-1334, Oct. 1986.

[18] B. R. Hunt, "The application of constrained least squares estimation to image restoration by digital computer," *IEEE Trans. Comput.*, vol. C-22, no. 9, pp. 805-812, Sept. 1973.

[19] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*. San Francisco, CA: Holden-Day, 1968.

[20] M. Kendall, A. Stuart, and J. K. Ord, *The Advanced Theory of Statistics*, vol. 3, fourth ed. New York: Macmillan, 1983.

[21] A. Levi and H. Stark, "Signal reconstruction from phase by projection onto convex sets," *J. Opt. Soc. Amer.*, vol. 73, no. 6, pp. 810-822, June 1983.

[22] C. Locke and J. D. Spurrier, "On tests of uniformity," *Commun. Stat. A*, vol. 7, no. 3, pp. 241-258, 1978.

[23] Michel Loève, *Probability Theory*, fourth ed. New York: Springer, 1977 (vol. 1), 1978 (vol. 2).

[24] E. S. Pearson, R. B. D'Agostino, and K. O. Bowman, "Tests for departure from normality: Comparison of powers," *Biometrika*, vol. 64, no. 2, pp. 231-246, Aug. 1977.

[25] D. L. Phillips, "A technique for the numerical solution of certain integral equations of the first kind," *J. Ass. Comput. Mach.*, vol. 9, no. 1, pp. 84-97, Jan. 1962.

[26] M. B. Priestley, *Spectral Analysis and Time Series*, vol. 1. New York: Academic, 1981.

[27] M. B. Priestley, *Nonlinear and Nonstationary Time Series Analysis*. New York: Academic, 1988.

[28] M. Rosenblatt, *Stationary Sequences and Random Fields*. Boston, MA: Birkhäuser, 1985.

[29] P. Santago and S. A. Rajala, "Using convex set techniques for combined pixel and frequency domain coding of time-varying images," *IEEE J. Selected Areas Commun.*, vol. SAC-5, no. 7, pp. 1127-1139, Aug. 1987.

[30] F. C. Schweppe, "Recursive state estimation: Unknown but bounded errors and system inputs," *IEEE Trans. Automat. Contr.*, vol. AC-13, no. 1, pp. 22-28, Feb. 1968.

[31] M. I. Sezan and H. Stark, "Tomographic image reconstruction from incomplete view data by convex projections and direct Fourier inversion," *IEEE Trans. Med. Imaging*, vol. MI-3, no. 2, pp. 91-98, June 1984.

[32] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for the exponential distribution (complete samples)," *Technometrics*, vol. 14, no. 2, pp. 355-370, May 1972.

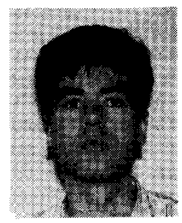
[33] T. Subba Rao and M. M. Gabr, "A test for linearity of stationary time series," *J. Time Series Analysis*, vol. 1, no. 2, pp. 145-158, 1980.

[34] H. J. Trussell, "Convergence criteria for iterative restoration methods," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, no. 1, pp. 129-136, Feb. 1983.

[35] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, no. 2, pp. 201-212, Apr. 1984.

[36] S. J. Yeh and H. Stark, "Iterative and one-step reconstruction from nonuniform samples by convex projections," *J. Opt. Soc. Amer. A*, vol. 7, no. 3, pp. 491-499, Mar. 1990.

[37] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1—Theory," *IEEE Trans. Med. Imaging*, vol. MI-1, no. 2, pp. 81-94, Oct. 1982.



**Patrick L. Combettes** (S'84–M'90) was born in Salon de Provence, France, on May 24, 1962. He received the Diplôme d'Ingénieur from l'Institut National des Sciences Appliquées de Lyon, France, the M.S. and Ph.D. degrees from North Carolina State University, Raleigh, all in electrical engineering, in 1985, 1987, and 1989, respectively.

During the academic year 1989–1990, he was a Visiting Assistant Professor in the Department of Electrical and Computer Engineering at North Carolina State University, Raleigh. In August 1990, he joined the Department of Electrical Engineering at the City College of the City University of New York, New York, as an Assistant Professor. His primary interests include digital signal processing and probability theory.

Dr. Combettes is a member of Sigma Xi, Phi Kappa Phi, Pi Mu Epsilon, the IEEE Signal Processing Society, the IEEE Information Theory Society, and the American Mathematical Society.

**H. Joel Trussell** (S'75–M'76–SM'91) was born in Atlanta, GA, on February 3, 1945. He received the B.S. degree from the Georgia Institute of Technology in 1967, the M.S. degree from Florida State University in 1968, and the Ph.D. degree from the University of New Mexico in 1976.

He joined the Los Alamos Scientific Laboratory, Los Alamos, NM, in 1969 where he began working in image and signal processing in 1971. During 1978–1979, he was a Visiting Professor at Heriot-Watt University, Edinburgh, Scotland. In 1980, he joined the Electrical and Computer Engineering Department at North Carolina State University, Raleigh, NC. During 1988–1989, he was a Visiting Scientist at the Eastman Kodak Company in Rochester, NY. His current interests include signal restoration/reconstruction, color reproduction, and mathematical methods.

Dr. Trussell is a member of Sigma Xi, Tau Beta Pi, Phi Kappa Phi, and the IEEE Signal Processing Society. He has won the IEEE ASSP Society Senior Paper Award (with M. R. Civanlar). He has been an Associate Editor for the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING and a member of the Multidimensional Digital Signal Processing Committee of that society.