Correspondence

On the Strong Information Singularity of Certain Stationary Processes

BRUCE E. HAJEK

Abstract—**In an exploratory paper, T. Berger studied discrete random processes which generate information slower than linearly with time. One of his objectives was to provide a physically meaningful definition of a deterministic process, and to this end he introduced the notion of strong information singularity. His work is supplemented by demonstrating that a large class of convariance stationary processes are strongly information singular with respect to a class of stationary Gaussian processes. One important consequence is that for a large class of covariance stationary processes the information rate equals that of the process associated with the Brownian motion component of the spectral representation.**

I. Introduction

Information singular random processes, as defined by Berger [1], are those which are negligible or deterministic in some information-theoretic sense. Berger demonstrated that for the class of strictly stationary square integrable random processes, information-singular processes are simply those having zero Kolmogorov entropy (or equivalently, those processes which are completely determined by their infinite past). Hence the study of information singularity promises a generalization of the notion of zero entropy to wide-sense stationary and even more general processes. One of Berger's main results was to classify a certain collection of wide-sense stationary random processes as information singular by considering the behavior of the processes appearing in their spectral representations. Our main results involve similar spectral considerations.

A second aspect of Berger's work involves determining the information conveyed by a random process in the presence of measurement inaccuracies or noise. This leads to considering the information content of a random process in the presence of other random processes and opens the possibility of decomposing a random process into singular and "regular" components. Berger defined a process to be strongly information singular with respect to a second process if it is deterministic in a specific physically meaningful sense, even when "corrupted" by the second process.

We will prove that the class of processes which Berger proved were information singular are actually strongly information singular with respect to Gaussian processes having bounded spectral density. This solves open problem number 3 listed by Berger. The class of information-singular processes considered here consists of those complex-valued wide-sense stationary processes which have a continuous-in-probability independent increment jump process in their spectral representation. The proof, given in the next section, involves the construction of a nonlinear estimator which is interesting in its own right. First we state the definitions of information singularity presented by Berger and then consider two revealing examples.

Manuscript received December 15, 1976; revised February 6, 1979. This work was supported by the National Science Foundation under Grant ENG75-20864. This paper was presented at the Johns Hopkins Conference, Baltimore, MD, April 1977.

The author is with the Coordinated Science Laboratory, and the Department of Electrical Engineering, University of Illinois at Urbana, Urbana, IL 61801.

Let $\{Y_k, k=0, \pm 1, \cdots\}$ and $\{W_k, k=0, \pm 1, \cdots\}$ be two independent random processes with components taking values in $\mathbb{C}$, the set of complex numbers. For any $m=2n+1$, let $\boldsymbol{Y}=(Y_{-n}, \cdots, Y_n)$ and $\boldsymbol{W}=(W_{-n}, \cdots, W_n)$. Let $\{0,1\}^*$ be the set of all finite binary strings, and let $l(s)$ denote the length of $s \in \{0,1\}^*$. $\{Y_k\}$ is defined to be *strongly variable-length information singular relative to* $\{W_k\}$ if there exists a sequence of encoder-estimators and decoders $\{e_m: \mathbb{C}^m \to \{0,1\}^*\}$ and $\{d_m: \{0,1\}^* \to \mathbb{C}^m\}$ such that

$$\lim_{m\to\infty} \frac{1}{m} El(e_m(\boldsymbol{Y}+\boldsymbol{W})) = \lim_{m\to\infty} \frac{1}{m} E(\|\boldsymbol{Y}-\hat{\boldsymbol{Y}}\|^2) = 0$$

where $\hat{\boldsymbol{Y}} = d_m(e_m(\boldsymbol{Y}+\boldsymbol{W}))$ and $\|\cdot\|$ denotes the Euclidean norm in complex m-space. We could alternatively require that there exist good block codes with length arbitrarily small compared with m, yielding the definition of strongly "block" information singular. It is unknown whether or not the two definitions coincide. We will henceforth implicitly consider only variable-length information singularity. The process $\{Y_k\}$ is simply *information singular* if it is strongly information singular with respect to the identically zero process.

One implication of the strong information singularity of $\{Y_k\}$ with respect to $\{W_k\}$ is that for the purpose of transmitting blocks from the "source" process $\{Y_k+W_k\}$ with small average mean square distortion, the component $\{Y_k\}$ is negligible. From another point of view, we could be given a "signal" process $\{Y_k\}$ and a "noise" process $\{W_k\}$. The information singularity of $\{Y_k\}$ with respect to $\{W_k\}$ implies that the process $\{Y_k\}$ is deterministic in the following sense. Even when a noisy version of it is observed, namely $\{Y_k+W_k\}$, the process $\{Y_k\}$ can be estimated with arbitrarily small mean square error and then can be conveyed over any channel with positive capacity.

We should point out that the notion of strong information singularity seems to involve two separate concepts—estimation and information singularity. We conjecture that an information singular random process which may be accurately (in the mean square sense) estimated from a noisy version of itself is strongly information singular with respect to the noise. The converse is obviously true and adds further significance to the notion of strong information singularity.

We will now consider two simple examples. Let $U_0, U_1, \cdots$ be mutually independent with $P(U_i=0)=P(U_i=1)=0.5$. Define Y_k for positive integers k by

$$Y_k = \sum_{j=0}^{\infty} i_j U_j \pmod 2, \qquad \text{if } k = \sum_{j=0}^{\infty} i_j 2^j (i_j \in \{0,1\}).$$

Construct $(Y_0, Y_{-1}, \cdots)$ so that $(Y_0, Y_{-1}, \cdots)$ and $(Y_1, Y_2, \cdots)$ are independent and so that $(Y_0, Y_{-1}, \cdots)$ has the same distribution as $(Y_1, Y_2, \cdots)$. The variables of the process $\{Y_k\}$ are pairwise independent so that $\{Y_k\}$ is covariance stationary. For $j>0$, $(Y_1, \cdots, Y_{2^j-1})$ is completely determined by $(U_0, \cdots, U_{j-1})$, or j bits of information. We conclude that for any $m=2n+1$, $\boldsymbol{Y}=(Y_{-n}, \cdots, Y_n)$ is completely determined by at most $2+\log_2 m$ bits of information, so that $\{Y_k\}$ is obviously information singular. The sequence $Y_1, Y_2, \cdots$ continues to generate information, for if $j \geq 0$, Y_{2^j} is independent of the collection of all variables with lesser index. (This provides a counterexample to Berger's conjecture that covariance-stationary information-singular processes are "predictable" or subordinate to their past.)

Let $\{W_k\}$ be a process consisting of independent identically distributed Gaussian random variables with mean zero and

0018-9448/79/0900-0605$00.75 ©1979 IEEE

variance σ^2. We show that $\{Y_k\}$ is strongly information singular with respect to $\{W_k\}$. Let $m=2n+1$ where $n=2^r-1$ for some $r>0$. Consider the problem of estimating $(U_0,\cdots,U_{r-1})$ from $(Y_1+W_1,\cdots,Y_n+W_n)$. For any j with $0\leqslant j\leqslant r-1$, subtract the sum of those Y_k+W_k, $1\leqslant k\leqslant n$ such that U_j is not a term in the sum defining Y_k, from the sum of those Y_k+W_k such that U_j is a term in the sum defining Y_k. Upon division by $(n+1)/2$ we obtain an estimate of U_j which, given U_j, is normally distributed with mean U_j and variance $4n(n+1)^{-2}$. We can in fact estimate $(U_0,\cdots,U_{r-1})$ from $(Y_1+W_1,\cdots,Y_n+W_n)$ with error probability converging to zero with n faster than $\exp(-n/2\sigma^2)$. The number of bits needed to specify an estimate of $(U_0,\cdots,U_{r-1})$ is $r=\log_2 n$. An estimate of $(Y_1,\cdots Y_n)$ which is exactly equal to $(Y_1,\cdots,Y_n)$ with probability $1-o(\exp(-n/2\sigma^2))$ can be constructed from the estimate of $(U_0,\cdots,U_{r-1})$. This estimation-encoding procedure may be easily modified to prove the strong information singularity of $\{Y_k\}$ with respect to $\{W_k\}$.

The next example we consider was studied by Berger. If Λ is distributed on $(-\pi,\pi)$ and if A is a complex square integrable random amplitude, then $\{Z_k\}$ given by $Z_k=A\exp(ik\Lambda)$ is information singular. This is intuitively pleasing, since a good estimate of $(Z_{-n},\cdots,Z_n)$ may be specified by giving a good estimate of only two random variables, Λ and A. The information singularity of $\{Z_k\}$ does not imply the strong information singularity of $\{Z_k\}$ with respect to an arbitrary process $\{N_k\}$. For example, if A and Λ are nondegenerate and $\{N_k\}$ has the same distribution as $\{Z_k\}$, then $\{Z_k\}$ is not strongly information singular with respect to $\{N_k\}$. A fundamental idea espoused by Berger is that strong information singularity naturally arises when an information singular process is spectrally "orthogonal" to a second process. Our main results, presented in the next section, support this idea.

II. Strong Information Singularity of a Class of Processes

Suppose $\{X_k\}$ is a complex-valued wide-sense stationary random process. Then, according to the spectral representation theorem [2], there exists a complex orthogonal increment process $\{\beta(\lambda),\frac{1}{2}\leqslant\lambda\leqslant\frac{1}{2}\}$ such that, using quadratic mean integrals,

$$X_k=\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda}\,d\beta(\lambda)\qquad\text{almost everywhere (a.e.).}$$

The power spectral measure of $\{X_k\}$ is given by $S(d\lambda)=E(|\beta(d\lambda)|^2)$, which is also the infinitesimal covariance of the process $\{\beta(\lambda)\}$. The variance of X_k is $\sigma_X^2=S(\{\lambda:-\frac{1}{2}\leqslant\lambda\leqslant\frac{1}{2}\})$. Linear prediction theory uses only the second order properties of $\{X_k\}$. Equivalently, since the autocorrelation function of $\{X_k\}$ is determined by $S(d\lambda)$ (formally by Fourier transform), only the power spectral measure $S(d\lambda)$ is needed to apply linear prediction theory. However, as we shall see, the information rate of $\{X_k\}$ is not determined by $S(d\lambda)$, but depends on the detailed behavior of the sample paths of $\{\beta(\lambda)\}$.

Suppose now that $\{\beta(\lambda)\}$ has independent, not merely orthogonal, increments. In addition, assume that $\{\beta(\lambda)\}$ is continuous in probability and (without loss of generality) right continuous with finite limits from the left. Then $\{\beta(\lambda)\}$ may be decomposed into the sum of two independent processes $\{\xi(\lambda)\}$ and $\{\eta(\lambda)\}$ such that $\{\xi(\lambda)\}$ is a pure jump process with independent increments and $\{\eta(\lambda)\}$ is a complex Wiener process (a continuous independent increment process). The process $\{\xi(\lambda)\}=\{\beta(\lambda)-\eta(\lambda)\}$ is continuous in probability, right continuous, and has finite left limits. This decomposition induces an additive decomposition of the process $\{X_k\}$, for if we define

$$Z_k=\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda}\,d\xi(\lambda)\qquad\text{a.e.}\tag{1}$$

and

$$N_k=\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ik\lambda}\,d\eta(\lambda)\qquad\text{a.e.}\tag{2}$$

then, with probability one, $X_k=Z_k+N_k$ for all k. It was shown by Berger that the process $\{Z_k\}$ is information singular. To simplify matters we will require that the process $\{N_k\}$ have bounded spectral density. That is, if $S_N(d\lambda)$ is the power spectral measure of $\{N_k\}$, we require that $S_N([-\frac{1}{2},\lambda])$ have a bounded derivative for $-\frac{1}{2}<\lambda<\frac{1}{2}$. This requirement is satisfied if, for example, $S_N(d\lambda)=\sigma_N^2 d\lambda$ so that $\{N_k\}$ is Gaussian white noise. Our main result is the following theorem.

Theorem: $\{Z_k\}$ is strongly information singular with respect to $\{N_k\}$.

Remark: When $N_k\equiv 0$ we obtain Berger's Theorem 2 that $\{Z_k\}$ is information singular.

Proof: Choose $m=2n+1$ and set $\boldsymbol{Z}=(Z_{-n},\cdots,Z_n)$ and $\boldsymbol{N}=(N_{-n}\cdots N_n)$. Define the discrete Fourier transforms

$$\xi_j=\frac{1}{\sqrt{m}}\sum_{k=-n}^{n} e^{-i2\pi k(j/m)}Z_k\tag{3}$$

and

$$\eta_j=\frac{1}{\sqrt{m}}\sum_{k=-n}^{n} e^{-i2\pi k(j/m)}N_k.\tag{4}$$

Let $\boldsymbol{\xi}=(\xi_{-n},\cdots,\xi_n)$ and $\boldsymbol{\eta}=(\eta_{-n},\cdots,\eta_n)$. The reader should keep in mind that ξ_j, $\boldsymbol{\xi}$, etc., depend on $m=2n+1$. Note that the mappings $\boldsymbol{Z}\to\boldsymbol{\xi}$ and $\boldsymbol{N}\to\boldsymbol{\eta}$ are linear Euclidean norm preserving mappings of $\mathbb{C}^m$ onto $\mathbb{C}^m$ with inverses given by

$$Z_k=\frac{1}{\sqrt{m}}\sum_{j=-n}^{n} e^{i2\pi k(j/m)}\xi_j\tag{5}$$

and

$$N_k=\frac{1}{\sqrt{m}}\sum_{j=-n}^{n} e^{i2\pi k(j/m)}\eta_j.\tag{6}$$

It follows that if we have an estimate of $\boldsymbol{\xi}$ we can, by inverse transforming the estimate, estimate $\boldsymbol{Z}$ with the same average squared error. We will show that estimators $\hat{\boldsymbol{\xi}}$ based on $\boldsymbol{\xi}+\boldsymbol{\eta}$ (or equivalently, on $\boldsymbol{Z}+\boldsymbol{N}$) exist such that

$$\lim_{m\to\infty}\frac{1}{m}\|\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}\|^2=0\tag{7}$$

and

$$\lim_{m\to\infty}\frac{1}{m}H(\hat{\boldsymbol{\xi}})=0.\tag{8}$$

By the basic theorem for variable-length source coding ([3, p. 50]), $\hat{\boldsymbol{\xi}}$ can be noiselessly encoded with a variable-length code of average length less than $H(\hat{\boldsymbol{\xi}})+1$. The construction of $\hat{\boldsymbol{\xi}}$ satisfying (7) and (8) will therefore complete the proof of the theorem.

Intuitively, the reason for taking Fourier transforms, and thereby shifting to the frequency domain, is that the singularity of $\{Z_k\}$ with respect to $\{N_k\}$ is best observed in the frequency domain. Formally, the power frequency spectrum of a typical realization of $\{Z_k\}$ consists of at most a countable infinity of impulses, whereas a typical realization of $\{N_k\}$ has a continuous power spectrum.

The construction of $\hat{\boldsymbol{\xi}}$ will proceed via a lemma, proved in the next section, which provides another estimator $\tilde{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}$. The estimator $\hat{\boldsymbol{\xi}}$ will then simply be taken to be a quantized version of $\tilde{\boldsymbol{\xi}}$.

Lemma: Given ϵ, $D>0$, there exists for all large $m=2n+1$ an estimator $\tilde{\boldsymbol{\xi}}=(\tilde{\xi}_{-n},\cdots,\tilde{\xi}_n)$ of $\boldsymbol{\xi}$ based on $\boldsymbol{\xi}+\boldsymbol{\eta}$ such that

$$\frac{1}{m}\sum_{k=-n}^{n} P(\tilde{\xi}_k\neq 0)\leqslant\epsilon\tag{9}$$

and

$$\frac{1}{m}E\|\boldsymbol{\xi}-\tilde{\boldsymbol{\xi}}\|^2\leqslant D.\tag{10}$$

Starting with $\tilde{\xi}$ satisfying (9) and (10), we will now define $\hat{\xi}$. Let

$$\hat{\xi}_j = 2k\sqrt{D} \text{ if } (2k-1)\sqrt{D} \leqslant \tilde{\xi}_j \leqslant (2k+1)\sqrt{D}\ ; \quad k=0,\pm 1,\cdots.$$

It is clear that $(1/m)\|\hat{\xi}-\tilde{\xi}\|^2 \leqslant D$. Using (10) and the inequality $\|a+b\|^2 \leqslant 2\|a\|^2+2\|b\|^2$ we obtain

$$\frac{1}{m}E\|\hat{\xi}-\xi\|^2 \leqslant \frac{2}{m}E(\|\hat{\xi}-\tilde{\xi}\|^2+\|\tilde{\xi}-\xi\|^2) \leqslant 4D. \tag{11}$$

We will next show that $H(\hat{\xi})$ can be made small. Our proof is very similar to Berger's proof of Lemma 3. Note that by (11),

$$\frac{1}{m}E\|\hat{\xi}\|^2 \leqslant \frac{2}{m}E(\|\hat{\xi}-\xi\|^2+\|\xi\|^2) \leqslant 8D+2\sigma_Z^2. \tag{12}$$

In terms of $P_{j,k}(m)=P(\hat{\xi}_j=2k\sqrt{D}\,)$, (9) and (12), respectively, imply that

$$\frac{1}{m}\sum_{j=-n}^{n}\sum_{k\neq 0}P_{j,k}(m) \leqslant \epsilon \tag{13}$$

and

$$\frac{1}{m}\sum_{j=-n}^{n}\sum_{k=-\infty}^{\infty}4k^2P_{j,k}(m) \leqslant \gamma \tag{14}$$

where $\gamma=(8D+2\sigma_z^2)/D$ is a constant, independent of m. The entropy of $\hat{\xi}$ is

$$\frac{1}{m}H(\hat{\xi}) = -\frac{1}{m}\sum_{j=-n}^{n}P_{j,0}(m)\log P_{j,0}(m) - \frac{1}{m}\sum_{j=-n}^{n}\sum_{k\neq 0}P_{j,k}(m)\log P_{j,k}(m). \tag{15}$$

Using the fact that (13) implies that $(1/m)\sum_{j=-n}^{n}P_{j,0}(m) \geqslant 1-\epsilon$, the concavity and monotonicity of $-\alpha\log\alpha$ for α near 1, and Jensen's inequality, we obtain

$$\begin{aligned}-\frac{1}{m}\sum_{j=-n}^{n}P_{j,0}(m)\log P_{j,0}(m) &\leqslant -\left(\frac{1}{m}\sum_{j=-n}^{n}P_{j,0}(m)\right)\log\left(\frac{1}{m}\sum_{j=-n}^{n}P_{j,0}(m)\right)\\ &\leqslant -(1-\epsilon)\log(1-\epsilon).\end{aligned} \tag{16}$$

This tends to zero as ϵ tends to zero. Using Lagrange's method one easily finds that the second term in (15) is maximized subject to (13) and (14) when $P_{j,k}(m)=ce^{-\alpha k^2}$ (no dependence on j), where c and α are m-dependent and chosen so that (13) and (14) are satisfied with equality. Given this, the second half of Berger's proof of Lemma 3 goes through without change to yield that the second term in (15) converges to zero as $m\to\infty$ and $\epsilon\to 0$. We obtain $(1/m)H(\hat{\xi})\to 0$ as $m\to\infty$ and $\epsilon\to 0$ for any fixed $D>0$. In view of (11), the existence of a sequence of estimators $\hat{\xi}$ satisfying (7) and (8) is guaranteed.

III. An Estimation Result–Proof of Lemma

In this section we prove the lemma stated in Section II. We will begin by characterizing the distribution of ξ and η. From (1) and (3) we obtain

$$\begin{aligned}\xi_j &= \int_{-\frac{1}{2}}^{\frac{1}{2}}\left(\frac{1}{\sqrt{m}}\sum_{k=-n}^{n}e^{i2\pi kj/m}e^{i2\pi k\lambda}\right)d\xi(\lambda) \qquad \text{a.e.}\\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}}g_m\left(\lambda-\frac{j}{m}\right)d\xi(\lambda) \qquad \text{a.e.}\end{aligned} \tag{17}$$

where

$$g_m(u) \triangleq \frac{1}{\sqrt{m}}\sum_{k=-n}^{n}e^{i2\pi ku} = \frac{\sin(m\pi u)}{\sqrt{m}\sin(\pi u)}. \tag{18}$$

The function $g_m(u)$ has period one, is continuous, and $g_m(u)^2$ restricted to the interval $[-\frac{1}{2},\frac{1}{2}]$ converges to a unit impulse at the origin as m increases without limit. We conclude that ξ_j depends primarily on the behavior of the process $\{\xi(\lambda)\}$ near $\lambda=j/m$. Similarly,

$$\eta_j = \int_{-\frac{1}{2}}^{\frac{1}{2}}g_m\left(\lambda-\frac{j}{m}\right)d\eta(\lambda). \tag{19}$$

Recall that $(1/m)E\|\xi\|^2=\sigma_z^2$ and $(1/m)E\|\eta\|^2=\sigma_N^2$ for all m, so that in some average sense, the variance of the variables ξ_j and η_j is of the same order of magnitude for all j and m. However, for large m, we see from (17) that ξ_j will be very large when a jump in $\{\xi(\lambda)\}$ occurs near $\lambda=j/m$, but that with high probability ξ_j will be quite small. In fact, we shall prove below that in some average sense the ξ_j converge in distribution to zero as $m\to\infty$.

The variables η_j, on the other hand, are all Gaussian with variance bounded above for all $m=2n+1$ and j. In fact, we have assumed that the process $\{N_k\}$ has a bounded spectral density. Thus there is a constant M such that

$$S_N(d\lambda) = E(|d\eta(\lambda)|^2) \leqslant Md\lambda.$$

Since $\{\eta(\lambda)\}$ is a Wiener process it follows that ξ_j is Gaussian for all m and that

$$\begin{aligned}E(|\eta_j|^2) &= \int_{-\frac{1}{2}}^{\frac{1}{2}}g_m^2\left(\lambda-\frac{j}{m}\right)S_N(d\lambda)\\ &\leqslant M\int_{-\frac{1}{2}}^{\frac{1}{2}}g_m^2\left(\lambda-\frac{j}{m}\right)d\lambda = M,\end{aligned} \tag{20}$$

where the last equality is apparent from (18). We conclude that for each $d>0$ there is a constant ϵ_d such that, for all j and m,

$$P(|\eta_j| \geqslant d) \leqslant \epsilon_d \qquad E\left(|\eta_j|^2\chi_{|\eta_j|\geqslant d}\right) \leqslant \epsilon_d \tag{21}$$

where χ denotes an indicator function and ϵ_d converges to zero as d increases to infinity.

The statement that the ξ_j converge to zero in distribution in some average sense is made precise by the following two facts:

$$\lim_{m\to\infty}\frac{1}{m}\sum_{j=-n}^{n}P(|\xi_j| \geqslant \tau)=0, \qquad \text{if } \tau>0 \tag{22}$$

and

$$\lim_{m\to\infty}\frac{1}{m}\sum_{j=-n}^{n}E\left(|\xi_j|^2\chi_{\{|\xi_j|<3d\}}\right)=0, \qquad \text{if } d>0. \tag{23}$$

Equation (23) follows from (22), because for all $\tau>0$,

$$\frac{1}{m}\sum_{j=-n}^{n}E\left(|\xi_j|^2\chi_{\{|\xi_j|<3d\}}\right) \leqslant \tau^2+\frac{3d}{m}\sum_{j=-n}^{n}P\left(|\xi_j^2| \geqslant \tau\right),$$

and by (22) the right side converges to τ^2 as $m\to\infty$. Berger proved (22) for the case when the process $\{\xi(\lambda)\}$ has stationarily distributed increments. In fact, in that case, one easily sees from (17) and the periodicity of $g_m(u)$ that for fixed $m=2n+1$, $\xi_{-n},\cdots,\xi_n$ are identically distributed. Berger then showed that as $m\to\infty$, the ξ_j converge to zero in distribution, implying (22). When the process $\{\xi(\lambda)\}$ has possibly nonstationary increments, the variables ξ_j will no longer converge to zero in distribution uniformly, but (22) is strong enough for our purposes.

Proof of (22): We prove (22) for the case when $\{\xi(\lambda)\}$ and hence the ξ_j take on real values. Our proof is easily modified to cover the general complex case.

Choose any $\tau>0$ and $\epsilon>0$. Since the process $\{\xi(\lambda)\}$ is an independent increment process, continuous in probability, its characteristic function may be expressed as (see [5] for general theory)

$$E\{\exp iu(\xi(\lambda_1)-\xi(\lambda_2))\} = \exp\int_{\lambda_1}^{\lambda_2}d\lambda\left(iua(\lambda)+\int_{-\infty}^{\infty}e^{iux}-1-iux\Pi(\lambda,dx)\right).$$

The spectral measure $\Pi(\lambda, A)$ of $\{\xi(\lambda)\}$ is a Borel measure for each λ and is continuous in λ for each A. Formally $\Pi(\lambda, A)$ is the instantaneous (at λ) expected number of jumps of $\{\xi(\lambda)\}$ per unit length which have magnitude in A. It is related to the power spectral measure of $\{Z_k\}$ by

$$S_Z(d\lambda) = d\lambda \int_{-\infty}^{\infty} x^2 \Pi(\lambda, dx).$$

By our assumption that $\{\xi(\lambda)\}$ was a pure jump process, it follows that for δ small enough

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} d\lambda \int_{|x|<\delta} x^2 \Pi(\lambda, dx) \leq \frac{\tau^2 \epsilon}{40}. \tag{24}$$

By the Levy decomposition theorem we can decompose $\{\xi(\lambda)\}$ into the sum of two independent processes, $\{\xi^1(\lambda)\}$ and $\{\xi^2(\lambda)\}$, such that the $\{\xi^i(\lambda)\}$ are both independent increment processes, which are both continuous in probability, right continuous, and have finite limits on the left. In addition, $\{\xi^1(\lambda)\}$ is constant except for jumps of magnitude at least d, the jumps of $\{\xi^2(\lambda)\}$ have magnitude less than d, and the spectral measures $\Pi_i^i(\lambda, dx)$ of $\{\xi_i^i(\lambda)\}$ satisfy

$$\Pi^1(\lambda, A) = \Pi(\lambda, A \cap (-\delta, \delta)^c),\ \Pi^2(\lambda, A) = \Pi(\lambda, A \cap (-\delta, \delta)).$$

Use (17) with $\{\xi(\lambda)\}$ replaced by $\{\xi^1(\lambda)\}$ or $\{\xi^2(\lambda)\}$ to define variables $\boldsymbol{\xi}^1 = (\xi^1_{-n}, \cdots, \xi^1_n)$ and $\boldsymbol{\xi}^2 = (\xi^2_{-n}, \cdots, \xi^2_n)$, respectively. Clearly $\boldsymbol{\xi} = \boldsymbol{\xi}^1 + \boldsymbol{\xi}^2$ a.e.. Hence

$$\frac{1}{m} \sum_{j=-n}^{n} P(|\xi_j| > \tau) \leq \frac{1}{m} \sum_{j=-n}^{n} P\left(|\xi_j^1| > \tau/2\right) + \frac{1}{m} \sum_{j=-n}^{n} P\left(|\xi_j^2| > \tau/2\right) \tag{25}$$

$$E\left(|\xi_j^2|^2\right) \equiv \int_{-\frac{1}{2}}^{\frac{1}{2}} g_m^2(\lambda - j/m) E[|d\xi^2(\lambda)|^2]$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} g_m^2(\lambda - j/m) \int x^2 \Pi^2(\lambda, dx)\, d\lambda.$$

Hence

$$\frac{1}{m} \sum_{j=-n}^{n} E\left[|\xi_j^2|^2\right] \leq \int_{-\frac{1}{2}}^{\frac{1}{2}} f_m(\lambda) \int_{|x|\leq\delta} x^2 \Pi(\lambda, dx)\, d\lambda \tag{26}$$

where

$$f_m(\lambda) = \frac{1}{m} \sum_{j=-n}^{n} g_m\left(\lambda - \frac{j}{m}\right)^2.$$

We claim that $f_m(\lambda) \leq 5$ for all m, λ. In fact $f_m(\lambda)$ is periodic with period $1/m$, and for $|\lambda| \leq 1/2m$

$$f_m(\lambda) = \frac{1}{m^2} \sum_{j=-n}^{n} \frac{\sin^2(m\Pi\lambda - j\Pi)}{\sin^2(\Pi(\lambda - j/m))}$$

$$\leq 1 + \frac{1}{m^2} \sum_{\substack{j=-n \\ j\neq 0}}^{n} \frac{1}{\sin^2(\Pi(\lambda - j/m))}$$

$$\leq 1 + \frac{1}{4m^2} \sum_{\substack{j=-n \\ j\neq 0}}^{j=n} \frac{1}{(\lambda - j/m)^2} \leq 1 + \frac{1}{2m^2} \sum_{j=1}^{n} \frac{1}{\left(\frac{1}{2m} - \frac{j}{m}\right)^2}$$

$$= 1 + 2 \sum_{j=1}^{n} \frac{1}{(1-2j)^2} \leq 5.$$

Therefore, by Chebyshev's inequality, (24) and (26),

$$\frac{1}{m} \sum_{j=-n}^{n} P\left(|\xi_j^2| > \frac{1}{2}\right) \leq \frac{4}{\tau^2} \frac{1}{m} \sum_{j=-n}^{n} E\left[|\xi_j^2|^2\right] \leq \frac{4}{\tau^2} \cdot 5 \frac{\tau^2 \epsilon}{40} \leq \epsilon/2. \tag{27}$$

The expected number of jumps of the process $\{\xi^1(\lambda)\}$ is $\int_{-1/2}^{1/2} \Pi(\lambda, (-\delta, \delta)^c)\, d\lambda$ which is finite. Hence there is a large constant K such that $P(A) > 1 - \epsilon/4$ where A is the event that $\{\xi^1(\lambda)\}$ has at most K jumps, and all of those jumps have magnitude less than K. When A holds, ξ_j^1 will be very near zero unless j/m is in a small neighborhood of at most K values of λ such that $\{\xi^1(\lambda)\}$ has a jump. As m increases, these neighborhoods shrink to the set of discontinuities of $\{\xi^1(\lambda)\}$ so that

$$\frac{1}{m} \sum_{j=-n}^{n} P\left(|\xi_j^1| > \tau/2\right) \leq \frac{1}{m} \sum_{j=-n}^{n} P\left(A \cap \left\{|\xi_j^1| > \tau/2\right\}\right) + \epsilon/4 \leq \epsilon/2$$

for large enough $m = 2n+1$. Combining this with (25) and (27) completes the proof of (22).

The Estimation Procedure: We now specify our estimate $\tilde{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}$. Choose $d > 0$, and for $-n \leq j \leq n$ let

$$\tilde{\xi}_j = \xi_j + \eta_j, \quad \text{if } |\xi_j + \eta_j| > 2d$$
$$= 0, \quad \text{if } |\xi_j + \eta_j| \leq 2d.$$

It is to be demonstrated that if $m = 2n+1$ and d are chosen large enough, then (9) and (10) are satisfied. First note that

$$\frac{1}{m} \sum_{k=-n}^{n} P\left(\tilde{\xi}_k \neq 0\right) \leq \frac{1}{m} \sum_{k=-n}^{n} P(|\xi_j| > d) + \frac{1}{m} \sum_{k=-n}^{n} P(|\eta_j| > d)$$

$$\leq \frac{1}{m} \sum_{k=-n}^{n} P(|\xi_j| > d) + \epsilon_d$$

which by (22) converges to zero as $n, d \to \infty$. Thus (9) is satisfied if n and d are large. Now for $-n \leq j \leq n$,

$$E\left(|\xi_j - \tilde{\xi}_j|^2\right) = E\left(|\eta_j|^2 \chi_{|\eta_j + \xi_j| > 2d}\right) + E\left(|\xi_j|^2 \chi_{|\eta_j + \xi_j| \leq 2d}\right)$$

$$\leq E\left(|\eta_j|^2 \chi_{|\eta_j| > d}\right) + E\left(|\eta_j|^2 \chi_{|\xi_j| > d}\right)$$

$$+ E\left(|\xi_j|^2 \chi_{|\eta_j| > d}\right) + E\left(|\xi_j|^2 \chi_{|\xi_j| \leq 3d}\right)$$

$$\leq \epsilon_d + MP(|\xi_j| > d) + E\left(|\xi_j|^2\right)\epsilon_d + E\left(|\xi_j|^2 \chi_{|\xi_j| \leq 3d}\right).$$

It then follows from (22), (23), and the fact that $(1/m)\|\boldsymbol{\xi}\|^2 = \sigma_z^2$ that if $m = 2n+1$ and d are chosen large enough, then (10) is satisfied.

IV. Concluding Remarks

Many important questions remain to be solved in the theory of information singularity. For example, when does (strong) variable-length information singularity imply (strong) block information singularity? Is strong information singularity a consequence of information singularity together with a "well-estimatable" property as alluded to in Section I?

Another important problem is to find the extent to which the results of [1] and the present paper generalize to arbitrary wide-sense stationary processes. After all, the independent increment property assumed here for the spectral representation processes seems much stronger than the most general, or orthogonal increment, property. It is interesting to note, however, that whenever a wide-sense stationary Gaussian process (which is, of course, strictly stationary) has a spectral density, then its spectral representation process can be taken to be a continuous *independent* increment process, or Brownian motion. Hence the main result of this correspondence is that any process $\{Z_k\}$ with a continuous-in-probability pure jump process in its spectral representation is strongly information singular with respect to any stationary Gaussian process with bounded spectral density. In fact, we may dispense with the boundedness condition. One way to handle the unbounded case is to choose M so large that, except for a λ set of small measure, the spectral density of the Gaussian process is less than M. Our estimators would then, loosely speaking, only attempt to estimate that part of the frequency spectrum of $\{Z_k\}$ for which the average noise power per unit bandwidth was less than M.

REFERENCES

[1] T. Berger, "Information singular processes," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 5, pp. 502–511, Sept. 1975.
[2] H. Cramér and M. R. Leadbetter, *Stationary and related stochastic processes*. New York: Wiley, 1967.
[3] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[4] I. I. Gihman and A. V. Skorohod, *The Theory of Stochastic Processes I*. New York: Springer-Verlag, 1974.
[5] I. I. Gihman and A. V. Skorohod, *The Theory of Stochastic Processes II*. New York: Springer-Verlag, 1975.

Symmetries of Binary Goppa Codes

OSCAR MORENO

***Abstract*—It is known that extended Goppa codes are invariant under the group of transformations $Z \to (AZ+B)/(CZ+D)$, with $AD+BC \neq 0$. This invariance is used here to classify cubic and quartic irreducible Goppa codes and to investigate their symmetry groups. A computer has been used to determine the actual group of the codes of length 33 (for cubics and quartics). It has been said, concerning the trends in symmetry groups with respect to the Gilbert bound, that "a good family of codes can be linear or have many symmetries, but not both" [8]. The groups found here are rather small; and so the results reinforce that statement.**

I. INTRODUCTION

There are four sections in this correspondence. In Section II results from [3] and [1] are restated and used to classify cubic and quartic extended Goppa codes. In the next section the largest possible group of symmetries (to be found from the group action given in [3]) is studied; that is to say its order and main characteristics are found. In the last section a computer is used to compare the results found here and the actual classification of Goppa codes (cubic and quartic) and their symmetry groups for $n=33$. Finally, we would like to comment that we have considered only binary irreducible Goppa codes, even though some of the results can be generalized. This provides a simpler structure to the correspondence.

II. CLASSIFICATION THEOREMS

The binary Goppa code of length $n=2^m$ with Goppa polynomial $g(Z)$[1] consists of all vectors $X=(X(\alpha_i))=(X(\alpha_1),X(\alpha_2),\cdots,X(\alpha_n)) \in Z_2^n$ (where $\{\alpha_1 \cdots \alpha_n\}=\mathrm{GF}\ (2^m)$), satisfying

$$\sum_{i=1}^{n} \frac{X(\alpha_i)}{Z-\alpha_i} \equiv 0(g(Z)).$$

We call this code the g-code. We assume here that g is irreducible; then if u is any root of g, we may write

$$\sum_{i=1}^{n} \frac{X(\alpha_i)}{u-\alpha_i}=0, \qquad u \in \mathrm{GF}(2^{mt})$$

where t is the degree of g. Correspondingly, for $u' \in \mathrm{GF}(2^{mt})$,

$$\sum_{i=1}^{n} \frac{X(\alpha_i)}{u'-\alpha_i}=0$$

defines the code with Goppa polynomial $g_{u'}$, the minimal polynomial of u'.

Manuscript received April 3, 1978; revised October 10, 1978. This work was supported in part by the University of Puerto Rico under an O. C. E. G. I. grant and in part by the National Science Foundation under Grant RIM78-16787.

The author is with the Department of Mathematics, University of Puerto Rico, Rio Piedras, Puerto Rico 00931.

[1]The coefficients of $g(Z)$ are in $\mathrm{GF}(2^m)$.

The proofs of the following three theorems are essentialy contained in [3] or [7].

Theorem 1: $(X(\alpha_i)) \to (X(A\alpha_i+B))$ (for $A \neq 0$, $A,B \in \mathrm{GF}\ (2^m)$) permutes the g_u-code into the g_{Au+B}-code.

Theorem 2: For $A,B,C,D \in \mathrm{GF}(2^m)$, $AD+BC \neq 0$,

$$(X(\alpha_i)) \to \left(X\left(\frac{A\alpha_i+B}{C\alpha_i+D}\right)\right)$$

permutes the extended g_u-code into the extended g_v code, where $v=(Au+B)/(Cu+D)$.

Theorem 3: $(X(\alpha_i)) \to (X(\alpha_i^2))$ permutes the g_u-code into the g_{u^2}-code.

Corollary 1: $(X(\alpha_i)) \to (X(A\alpha_i^{2^j}+B))$ permutes the g_u-code into the g_v-code, where $v=Au^{2^j}+B$.

Corollary 2:

$$(X(\alpha_i)) \to \left(X\left(\left(A\alpha_i^{2^j}+B\right)/\left(C\alpha_i^{2^j}+D\right)\right)\right)$$

permutes the extended g_u-code into the extended g_v-code, where $v=(Au^{2^j}+B)/(Cu^{2^j}+D)$.

We will now use these theorems to classify Goppa codes.

Theorem 4: There is only one g-code for irreducible g of degree 2.

Proof: We will work in the extension field $\mathrm{GF}(2^m)[u]$ where u is a root of g. Since the degree of g is 2, it is known that every element of GF (2^m) $[u]$ can be written as $Au+B$ with $A,B \in \mathrm{GF}(2^m)$. Also any other irreducible polynomial of degree 2 is the minimum polynomial of some $u' \in \mathrm{GF}(2^m)$ $[u]$ and $u'=Au+B$. Therefore, by Theorem 1, the $g_{u'}$-code is a permutation of the g_u-code. □

Theorem 5: There is only one extended g-code for irreducible g of degree 3.

Proof: As above let u be a root of g, of degree 3, and let us work in GF (2^m) $[u]$. Under the action of the group $X \to (AX+B)/(CX+D)$ $(AD \neq BC)$, u is mapped into a subset of GF (2^m) $[u]$. If we prove that this subset is $\mathrm{GF}(2^m)(u) \setminus \mathrm{GF}(2^m)$, then Theorem 2 will give us the result. Two such images cannot be the same; otherwise, if

$$\frac{Au+B}{Cu+D}=\frac{A'u+B'}{C'u+D'},$$

by cross multiplying we would get an equation of degree 2 in u, contradicting the fact that the minimal polynomial of u has degree 3. Therefore, since the group $(X \to (AX+B)/(CX+D)$, $AD \neq BC)$ has n^3-n elements, where $n=2^m$, there is exactly that number of elements in the image. But since there are precisely n^3-n elements in GF (2^m) $(u) \setminus \mathrm{GF}(2^m)$, this means, by using Theorem 2, that there is only one extended cubic code. □

The following theorem gives a canonical form for quartics under the equivalence relation of extended codes given by the group $X \to (AX+B)/(CX+D)$ $(AD \neq BC)$, by using Theorem 2.

Theorem 6: The Goppa polynomial for any quartic extended code can be chosen to be $g_B=X^4+X^2+BX+C_B$ or X^4+X+1, the latter only if m is odd. Here B varies in $\mathrm{GF}(2^m)$ in such a way that $\mathrm{tr}(B^{-1}) \neq \mathrm{tr}(1)$, and C_B is any given element not in the image of the map of $\mathrm{GF}(2^m) \to \mathrm{GF}(2^m)$ given by $X \to X^4+X^2+BX$.

Proof: Given any irreducible quartic $X^4+AX^3+BX^2+CX+D$ we can simplify it by means of our group in the following way. First, if $A \neq 0$ we can eliminate A by the map $X \to X+(CA^{-1})^{1/2}$ followed by $X \to 1/X$. Therefore, by using the group we can take any irreducible quartic to X^4+AX^2+BX+C. Now

0018-9448/79/0900-0609$00.75 ©1979 IEEE