T H E   U N I V E R S I T Y   O F   M I C H I G A N

COLLEGE OF ENGINEERING
Department of Aeronautical and Astronautical Engineering

Technical Report

SINGULAR PROBLEMS IN THE DETECTION OF SIGNALS IN GAUSSIAN NOISE

William L. Root

ORA Project 02905

under contract with:

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
RESEARCH GRANT NsG-2-59
WASHINGTON, D.C.

Presented at the Symposium on Time Series,
Brown University, June 11-14, 1962

administered through:

OFFICE OF RESEARCH ADMINISTRATION   ANN ARBOR

July 1962

ACKNOWLEDGMENT

I should like to acknowledge a debt to Dr. T. S. Pitcher for the use of some of his unpublished work. I have also benefitted from discussions of the mathematical material with Dr. Pitcher and Professor J. G. Wendel.

ABSTRACT

A statistical inference problem is called singular if the correct inference can be made with probability one. It has been observed by Grenander, Slepian, and others that the mathematical models used to describe the detection of radio signals in Gaussian noise sometimes appear to lead to singular inference problems. A large class of signal detection and extraction problems are examined here in the light of recent mathematical results, with the conclusion (which is necessarily a matter of opinion) that natural constraints prevent singular cases from arising.

The mathematical background is the theory of equivalence or singularity of (Gaussian) measures on function spaces. A unified treatment of some of the work in this area is given based largely on work of Kakutani, Grenander, Baxter, Slepian, Pitcher, and Feldman. In particular, a very slightly modified version of Pitcher's unpublished proof of the fundamental result (due to Feldman and Hájek) on pairs of Gaussian measures, and a drastically modified version of Feldman's theorem for the special case of rational spectral densities are given in detail.

I. INTRODUCTION

In the statistical theory of signal detection one is concerned with problems occurring in electrical communication engineering involving statistical inference from stochastic processes. Most of the work in this area has been concerned with the theory of detecting or characterizing information-bearing signals immersed in noise with Gaussian statistics. The discussion here is concerned with (1) singular cases arising in this special class of problems, and (2) implications that these singular cases carry concerning the suitability of the formulation.

We start from the model

$$y(t) = s(t) + n(t) \tag{1}$$

where t is a real variable, n(t) is a sample function from a real-valued Gaussian stochastic process $\{n_t\}$ which represents the noise, s(t) is a real-valued function representing the signal, and y(t) represents the observed waveform. We assume that y(t) is known to the observer, that s(t) is not precisely known, and that n(t) is not known but has certain known statistical properties. We want to make specified inferences about s(t) from the observation y(t).

The signal s(t) may be of the form $f(t; \alpha_1, \ldots, \alpha_n)$ where the function f is known to the observer, but the parameters $\alpha_1, \ldots, \alpha_n$ are not. For example, in the simplest detection problem, $s(t) = \alpha f(t)$ where $\alpha = 0$ or 1; the problem is then one of testing between two simple hypotheses concerning the mean of a Gaussian process. If the parameters $\alpha_1,\ldots,\alpha_n$ are real valued, the problem may be one of point or interval estimation. All such problems in which f is known and the parameters are unknown we say are of the <u>sure-signal-in-noise</u> type.

On the other hand s(t) may itself be a sample function from a stochastic process, of which only certain statistics are known to the observer. If this is so we say the problem is of the <u>noise-in-noise</u> type. It is worth noting that there is also a sort of in-between case which occurs when $s(t) = f(t; \alpha_1,\ldots,\alpha_n)$ where f is known and the α_i are random variables with known joint distribution. Properly, then, the signal is a sample function from a stochastic process $\{s_t\}$; however since the structure of $\{s_t\}$ is much better known than that of a process specified in the usual way through its family of joint distributions, it may be more appropriate to think of the resulting problem as sure-signal-in-noise than as noise-in-noise.

As in any analysis of a physical problem, the choice of an appropriate mathematical model is somewhat arbitrary, and in particular there are situations described usefully by either a sure-signal or noise-in-noise model. Usually, in fact, such is true if the mechanism whereby the channel distorts the signal is very complicated (see, for example, the article by R. Price listed in the Bibliography).

In any event, whatever inferences are to be made from the observed waveform must be made after a finite time. If we except sequential testing procedures, we can usually fix a basic time interval say of duration T, during which all the data are collected on which one decision or set of inferences is made. This interval of duration T is called the observation interval; we shall be concerned here with problems for which there is a fixed observation interval, so that y(t) in Eq. (1) will be qualified by the statement $0 \leqq t \leqq T$(or $a \leqq t \leqq a + T$). Note that s(t) or n(t) may be defined for other values of T and we may want to see what happens when T is varied.

In any electrical system whatever there is a background of thermally generated noise (Johnson noise, shot noise, etc.) which is generally assumed to be representable by a stationary Gaussian stochastic process, both because it is a macroscopic manifestation of a great many tiny unrelated motions, and because of experimental evidence. It is this background noise which is represented by n(t) in Eq. (1). This noise is always present, although it may not be the chief source of uncertainty about the received waveform. Usually one assumes the autocorrelation of the process $\{n_t\}$ to be known (although it seems almost impossible that it could be known precisely), and the mean to be zero (which in the model of Eq. (1) is equivalent to assuming it known). Thus the entire family of finite-dimensional distributions for the $\{n_t\}$ process is taken to be available.

For convenience we shall call the class of detection theory problems characterized somewhat loosely above, the Gaussian model. This term is to include both sure-signal-in-noise and noise-in-noise cases and is to imply that $\{n_t\}$, $-\infty < t < \infty$, is a stationary Gaussian process with known autocorrelation and that the observation interval is finite.

Various results obtained in the past few years show that there are classes of decision problems involving a model of the kind described for which a correct decision, or correct inference, can be made with probability one. Such problems will here be called singular. Slepian pointed out in 1958 that the problem of testing between the two simple hypotheses: that a waveform observed for a finite time be a sample function from a Gaussian process $\{x_t\}$ or from a different Gaussian process $\{\tilde{x}_t\}$, both of which are stationary and have known rational spectral density, is always singular except in a special case. From this he raised the question whether much of the noise-in-noise detection theory being developed was based on an adequate model; for it seems contrary to common sense that perfect detection of signals can be accomplished in a real-life situation. In 1950 Grenander had shown that a test between two possible mean-value functions of a Gaussian process with known statistics could be singular, even when the mean-value functions have finite "energy" (are of integrable square) and the observation period is finite. He also showed that the estimation of the "power level" of a Gaussian process with autocorrelation known except for scale is singular, again even with a finite observation interval. These results, which are quite simple, seem not to have been known or at least appreciated by engineers working on noise-theory problems for some time after 1950. In an application of Grenander's work, Davis, however, in 1955 gave a rationalization for excluding the singular cases in the problem of testing for the mean (a sure-signal-in-noise problem), and in 1958 Davenport and Root gave a different one (see Problem 14.6 in their book). Since Slepians's paper of 1958 there has been considerable interest in the appropriateness of the Gaussian model as it has been used in detection problems (see in particular the paper by Good).

I agree with the point of view that a well-posed detection theory problem should not yield a singular answer. With this as a sort of working principle, the aptness of the kind of model described above will be discussed in Section IV, where an argument is given that the Gaussian model is usually acceptable. The detection problems deal with probability measures on infinite product spaces or on function spaces. They are singular, as the term is defined here, when the measures are relatively singular. Thus one is led to the subject of relatively singular measures on function spaces, and in particular to singular Gaussian measures. In Section II a few basic results in this area are collected, and in Section III some more specialized results applicable to detection theory are given. Proofs are given for some of the propositions. It is likely that singular measures on function spaces are of interest to some who have no interest in detection theory; for them the following material will perhaps be useful as an introductory survey.

II. EQUIVALENT AND SINGULAR GAUSSIAN MEASURES

Since the eventual interest here is in continuous-parameter random processes, while many of the techniques involved use representations of these processes in terms of denumerably many random variables, one sometimes needs to carry relationships between pairs of measures on a Borel field to their induced measures on a Borel subfield, and vice versa. What is required usually turns out to be trivial, or nearly so, but it seems worthwhile to establish a procedure once and for all. For this purpose two simple lemmas are stated first.

Let Ω be a set, $\mathcal{B}$ a Borel field of subsets of Ω, and μ and ν probability measures on $\mathcal{B}$. The probability measures μ and ν are <u>mutually singular</u> (or simply <u>singular</u>) if and only if there is a set $A \in \mathcal{B}$ for which $\mu(A) = 0$, $\nu(A^c) = 0$. The condition μ, ν singular is denoted by $\mu \perp \nu$.

Consider a collection of Borel fields, each with base space Ω, and measures on these fields related to each other as follows. $\mathcal{B}$ is a Borel field on which there are two probability measures μ, ν. The completion of μ, we denote by $\bar{\mu}$, the completion of ν, by $\bar{\nu}$, and the Borel fields of sets measurable with respect to $\bar{\mu}$ and $\bar{\nu}$, we denote by $\overline{\mathcal{B}}_\mu, \overline{\mathcal{B}}_\nu$, respectively. Let $\mathcal{B}_o$ be a Borel field contained in both $\overline{\mathcal{B}}_\mu$ and $\overline{\mathcal{B}}_\nu$, and μ_o and ν_o be the measures induced on $\mathcal{B}_o$ by $\bar{\mu}$ and $\bar{\nu}$, respectively. It follows directly from the definitions that:

1. If $\mu_o \perp \nu_o$ then $\mu \perp \nu$.

Let $\mathcal{B}$, $\mu, \nu, \overline{\mathcal{B}}_\mu$, $\bar{\mu}, \overline{\mathcal{B}}_\nu$, $\bar{\nu}, \mathcal{B}_o$, μ_o and ν_o be defined as above. Suppose now, however, that μ_0 is equivalent to $\nu_0(\mu_0 \sim \nu_0)$. Let $\bar{\mu}_0$, $\bar{\nu}_0$ be the completions of μ_0, ν_0, respectively, and denote the Borel field of sets measurable with respect to either $\bar{\mu}_0$ or $\bar{\nu}_0$ by $\overline{\mathcal{B}}_0$. Suppose further that $\mathcal{B} \subset \overline{\mathcal{B}}_0$, and write μ', ν' for the measures induced on $\mathcal{B}$ by $\bar{\mu}_0$, $\bar{\nu}_0$, respectively. Then one can readily verify that:

2. Under the hypotheses of the preceding paragraph $\mu = \mu'$, $\nu = \nu'$, $\mu \sim \nu$, and $\overline{\mathcal{B}}_0 = \overline{\mathcal{B}}_\mu = \overline{\mathcal{B}}_\nu$.

The application of these lemmas is to situations such as the following. Suppose there are two real-valued random processes $\{x_t(\omega)\}$, $\{y_t(\omega)\}$, $t \in T$ (a linear parameter set), $\omega \in \Omega$ (an abstract set), such that the smallest Borel field containing all sets of the form $[\omega | x(t,\omega) \in A]$, A a Borel set, is the same as the corresponding Borel field containing all sets of the form $[\omega | y(t,\omega) \in A]$. The probability measure on $\mathcal{B}$ for the x-process is μ and for the y-process is ν.

Suppose also there is a denumerable collection of random variables $\{x_k\}$, each of which is equal almost everywhere with respect to both μ and ν to a function measurable with respect to $\mathcal{B}$, and representations for both $\{x_t\}$ and $\{y_t\}$ in terms of the x_k such that for every t x_t and y_t are equal almost everywhere, $d\mu$ and $d\nu$, respectively, to functions measurable with respect to the Borel field $\mathcal{B}_0$ generated by the x_k. Then if it can be shown that the measures μ_0 and ν_0 induced on $\mathcal{B}_0$ are equivalent, one has that the measures μ and ν are equivalent by Lemma 2. If the measures μ_0 and ν_0 are singular, then μ and ν are singular by Lemma 1.

Singularity and Equivalence of Product Measures

In the development to be sketched here we take as starting point a theorem of Kakutani on the equivalence or singularity of two probability measures each of which is an infinite direct product of probability measures, pair by pair equivalent. Suppose μ and ν are equivalent measures defined on the same Borel field of sets from Ω, then we define

$$\rho(\mu,\nu) = \int_{\Omega} \sqrt{\frac{d\nu}{d\mu}} \; d\mu$$

The function $\rho(\mu,\nu)$ thus defined has the immediately verifiable properties: $0 < \rho(\mu,\nu) \leqq 1$, $\rho(\mu,\nu) = 1$ if and only if $\mu = \nu$, $\rho(\mu,\nu) = \rho(\nu,\mu)$. Let $\mathfrak{M}(\mathfrak{B})$ be the class of all probability measures on $\mathfrak{B}$. The definition of $\rho(\mu,\mu')$ may be extended so that $\rho(\mu,\mu')$ is defined for all $\mu,\mu' \epsilon \mathfrak{M}(\mathfrak{B})$, as follows: Let $\nu \epsilon \mathfrak{M}(\mathfrak{B})$ dominate μ and μ' (i.e. $\mu << \nu$ and $\mu' << \nu$). Define

$$\psi = \sqrt{\frac{d\mu}{d\nu}} \quad , \quad \psi' = \sqrt{\frac{d\mu'}{d\nu}}$$

Then ψ and ψ' belong to the L_2 space $L_2(\nu)$, and

$$\rho(\mu,\mu') = (\psi,\psi') \tag{2}$$

where the inner product indicated is the inner product for $L_2(\nu)$. One verifies easily that for arbitrary μ and μ', (ψ,ψ') has the same value irrespective of the dominating measure ν used in its definition. Hence, Eq. (2) may be used to define $\rho(\mu,\mu')$ for all $\mu,\mu' \epsilon \mathfrak{M}(\mathfrak{B})$. With this extended definition it is clear that $\rho(\mu,\mu') = 0$ iff $\mu \perp \mu'$.

The basic theorem is then:

Theorem 1. (Kakutani)

Let $\{m_n\}$ and $\{m'_n\}$ be two sequences of probability measures, where m_n and m'_n are defined on a Borel field $\mathcal{B}_n$ of sets from a space Ω_n, and $m_n \sim m'_n$. Then the infinite direct product measures $m = \prod_{n=1}^{\infty} m_n$ and $m' = \prod_{n=1}^{\infty} m'_n$ are either equivalent, $m \sim m'$, or mutually singular, $m \perp m'$, according as the infinite product $\prod_{n=1}^{\infty} \rho(m_n, m'_n)$ is greater than zero or equal to zero. Moreover

$$\rho(m,m') = \prod_{n=1}^{\infty} \rho(m_n, m'_n)$$

The theorem is proved by imbedding $\mathfrak{M}(\mathcal{B})$ in a Hilbert space in which the ordinary strong convergence is equivalent to some kind of convergence of the products of the derivatives $\frac{dm'}{dm}$. The completeness of the Hilbert space guarantees the existence of a limit element which corresponds to the derivative of the infinite product measures, in the case of convergence. The imbedding is accomplished by defining a metric with the aid of Eq. (2) by:

$$d(\mu,\mu') = \|\psi - \psi'\| = [(\psi - \psi', \psi - \psi')]^{1/2} = [2(1 - \rho(\mu,\mu'))]^{1/2}$$

It can be then shown that $\prod_{k=1}^{k} \sqrt{\frac{dm'_k}{dm_k}}$ converges in $L_2(m)$ to $\sqrt{\frac{dm'}{dm}}$ if the product of the $\rho(m_n, m'_n)$ converges, the case of equivalence. Thus one has as a subsidiary result that a subsequence of $\left\{ \prod^{k} \frac{dm'_k}{dm_k} \right\}$ converges with probability one (dm) to $\frac{dm'}{dm}$ if the latter exists. This last statement can be improved, of course, by application of the martingale convergence theorem which shows that the original sequence of partial products converges to $\frac{dm'}{dm}$ with probability one (dm).

Gaussian Process with Shifted Mean

Let $\{x_t\}$, $t \in I$, I an interval in E_1, be a real separable (with respect to closed sets), measurable Gaussian random process, continuous in mean square, and with mean zero. We take $I = [0,1]$ for convenience; and we let $\mathcal{B}$ be the smallest Borel field containing all ω sets of the form $\{\omega | x(t,\omega) \in A\}$, $t \in I$, where A is a Borel set. Then $R(t,s) = E\, x(t)x(s)$ is a symmetric, non-negative definite, continuous function in $[0,1] \times [0,1]$; and the integral operator R on $L_2\, [0,1]$ defined by

$$Rf(t) = \int_0^1 R(t,s)\, f(s)\, ds, \qquad t \in [0,1]$$

is Hermitian, non-negative definite and Hilbert-Schmidt. We assume in addition that R is (strictly) positive definite. Then an orthonormalized sequence of eigenfunctions of R corresponding to all of its non-zero eigenvalues is a c.o.n.s. (complete orthonormal set) in $L_2\, [0,1]$. We denote eigenvalues of R by λ_n, $\lambda_n > 0$, and corresponding eigenfunctions by $\phi_n(t)$, i.e.

$$R\phi_n = \lambda_n \phi_n$$

$$(\phi_n, \phi_m) = \delta_{nm}$$

The condition that R be strictly definite is not necessary for what is to follow, but its presence simplifies the statements a little. It will be satisfied in the case that is of real interest to us, as will be pointed out in the last section.

We now let a(t) and b(t) be continuous functions defined for $t \in [0,1]$ and consider the random processes

$$y(t) = a(t) + x(t), \qquad 0 \leqq t \leqq 1$$
$$z(t) = b(t) + x(t), \qquad 0 \leqq t \leqq 1 \tag{3}$$

These processes are measurable, separable and have the same Borel field of measurable ω-sets as x(t). By the well-known representation of Karhunen and Loève,

$$x(t) = \sum_n x_n \phi_n(t), \qquad t \in [0,1]$$

where the convergence is in mean-square with respect to the probability measure for each t, and where the random variables x_n are given by

$$x_n = \int_0^1 x(t) \overline{\phi_n(t)} \, dt$$

and satisfy

$$E\, x_n \bar{x}_m = \lambda_n \delta_{nm}$$
$$E\, x_n = 0$$

Since x(t) is Gaussian, the x_n are jointly Gaussian random variables. If we let

$$a_n = \int_0^1 a(t) \overline{\phi_n(t)} \, dt$$

$$b_n = \int_0^1 b(t) \overline{\phi_n(t)} \, dt$$

then the random variables $y_n = x_n + a_n$ are Gaussian and independent, as are the $z_n = x_n + b_n$. The measures μ_n and ν_n induced on E_1 by y_n and z_n respectively are equivalent, so the theorem of Kakutani quoted above may be applied to yield that the product measures, which we denote by μ_o and ν_o respectively, are either equivalent or totally singular. The probability measures μ_o and ν_o are the measures induced on the Borel field $\mathcal{B}_o \subset \mathcal{B}$ generated by the x_n. Then by Lemmas 1 and 2 the processes y(t) and z(t) are either equivalent or mutually singular.

According to the theorem, μ_o and ν_o are equivalent if and only if $\prod \rho_n$ converge. One has, since y_n and z_n are Gaussian

$$\frac{d\mu_n}{d\nu_n}(\zeta) = \exp\left\{\frac{(\zeta - b_n)^2}{2\lambda_n} - \frac{(\zeta - a_n)^2}{2\lambda_n}\right\}$$

$$= \exp \frac{(b_n - a_n)}{2\lambda_n}(a_n + b_n - 2\zeta) ,$$

$$\rho_n = \frac{1}{\sqrt{2\pi\lambda_n}} \int_{-\infty}^{\infty} \sqrt{\frac{d\mu_n}{d\nu_n}(\zeta)} \quad \exp \quad -\left[\frac{(\zeta - b_n)^2}{2\lambda_n}\right] d\zeta$$

$$= \frac{1}{\sqrt{2\pi\lambda_n}} \exp\left(\frac{b_n^2 - a_n^2}{4\lambda_n}\right) \int_{-\infty}^{\infty} \exp \left\{\frac{\zeta(a_n - b_n)}{2\lambda_n} - \frac{(\zeta - b_n)^2}{2\lambda_n}\right\} d\zeta$$

$$= \exp \frac{-(a_n - b_n)^2}{8\lambda_n} ,$$

$$\prod \rho_n = \exp\left\{-\frac{1}{8}\sum_n \frac{(a_n - b_n)^2}{\lambda_n}\right\}$$

Thus one has the result due to Grenander:

Theorem 2. (Grenander)

The Gaussian random processes y(t) and z(t) defined by Eq. (3) are either equivalent or mutually singular. They are equivalent if the series $\sum_n \frac{(a_n - b_n)^2}{\lambda_n}$ converges, and singular if the series diverges to + infinity.

Two Gaussian Processes with Different Autocorrelations

It has just been noted that two Gaussian processes defined on a finite interval and identical except for different mean-value functions have the "zero-one" property of being either equivalent or singular. The same result has been demonstrated for arbitrary Gaussian processes on a finite interval independently by Hájek and Feldman (1958 and 1959), who used entirely different methods of proof and obtained different kinds of criteria for equivalence. Here we shall sketch a third proof given by T. S. Pitcher in an unpublished memorandum, which yields a criterion for equivalence which is somewhat similar to that first obtained by Feldman.

Suppose two real-valued Gaussian processes are defined on the interval $0 \leq t \leq 1$, each with mean zero, and with autocorrelation functions R(t,s) and S(t,s) continuous in the pair t,s in $[0,1] \times [0,1]$. We shall denote sample functions by x(t) and the respective probability measures on the space of sample functions for the two processes by μ_0 and μ_1.* Thus

$$E_i x(t) \equiv \int x(t) \, d\mu_i(x) = 0 \quad , \quad i = 0,1$$

*Note that same symbol is used for sample functions of both processes.

and

$$E_0 x(t)\ x(s) \equiv \int x(t)\ x(s)\ d\mu_0(x) = R(t,s)$$

$$E_1 x(t)\ x(s) \equiv \int x(t)\ x(s)\ d\mu_1(x) = S(t,s)$$

The integral operators on L_2 [0,1] with autocorrelations as kernels are written:

$$Rf(s) \equiv \int_0^1 R(s,t)\ f(t)\ dt$$

$$Sf(s) \equiv \int_0^1 S(s,t)\ f(t)\ dt$$

where $f(t)$ is any element of L_2 [0,1].

We proceed with a series of lemmas:

3. If R and S have different zero spaces, then $\mu_0 \perp \mu_1$.
If $Rf = 0$, then

$$E_i \int_0^1 x(t)\ f(t)\ dt = 0 \quad , \quad i = 0, 1$$

$$E_0 \left[\int_0^1 x(t)\ f(t)\ dt \right]^2 = (Rf,f) = 0$$

$$E_1 \left[\int_0^1 x(t)\ f(t)\ dt \right]^2 = (Sf,f)$$

Now, since S is a non-negative definite operator, either $Sf = 0$ or $(Sf,f) > 0$.

In the latter case the Gaussian random variable

$$\int_0^1 x(t)\ f(t)\ dt \equiv \theta$$

has positive variance with respect to μ_1 measure. Hence,

$$\mu_1[x|\theta(x) \neq 0] = \mu_0[x|\theta(x) = 0] = 1$$

Henceforth we assume, without any real loss of generality, that both R and S carry only the zero element in L_2 [0,1] into zero. Then R^{-1}, S^{-1}, $(R^{1/2})^{-1}$, $(S^{1/2})^{-1}$ are densely defined, symmetric, unbounded operators. In particular, if $R\phi_n = \lambda_n\phi_n$, $(\phi_n,\phi_m) = \delta_{nm}$, then for any $f\epsilon L_2[0,1]$ one has $f = \Sigma a_n\phi_n$, $\Sigma a_n^2 < \infty$. If $f_N = \sum_1^N a_n\phi_n$ then $f_N \to f$ and

$$R^{-1}\, f_N = \sum_1^N \frac{a_n}{\lambda_n}\,\phi_n$$

$$(S^{1/2})^{-1}\, f_N = \sum_1^N \frac{a_n}{\sqrt{\lambda_n}}\,\phi_n$$

Analogous formulas can be written for S in terms of its spectral decomposition. We shall write $(R^{1/2})^{-1} \equiv R^{-1/2}$, $(S^{1/2})^{-1} \equiv S^{-1/2}$

4. If $S^{1/2}R^{-1/2}$ or $R^{1/2}S^{-1/2}$ is unbounded, then $\mu_0 \perp \mu_1$.

Suppose there exists a sequence of elements f_k in the domain of $R^{-1/2}$ satisfying $||f_k|| = 1$ and $||S^{1/2}R^{-1/2}f_k|| \geq k^3$. Let

$$\theta_k(x) = \frac{1}{k}\int_0^1 x(t)\,(R^{-1/2}f_k)(t)\,dt$$

Each $\theta_k(x)$ is Gaussian with mean zero, and

$$E_0\theta_k^2 = \frac{1}{k^2}\,(R(R^{-1/2}f_k),\, R^{-1/2}f_k) = \frac{1}{k^2}$$

$$E_1\theta_k^2 = \frac{1}{k^2}\,(S(R^{-1/2}f_k),\, R^{-1/2}f_k)$$

$$= \frac{1}{k^2}\,||S^{1/2}R^{-1/2}f_k||^2 \geq k^4$$

Now, by the Tshebysheff inequality

$$\mu_0 [x \,|\, |\theta_k(x)| \geq \epsilon] \leq \frac{1}{\epsilon^2 k^2}$$

so by the Borel-Cantelli lemma

$$\mu_0 [x \,|\, |\theta_k(x)| \geq \epsilon, \text{ infinitely many } k] = 0$$

for every $\epsilon > 0$. Also, since each $\theta_k(x)$ is Gaussian,

$$\mu_1 [x \,\, |\theta_k(x)| \leq n] \leq \frac{1}{\sqrt{2\pi}} \frac{2n}{k^2} ,$$

and again by the Borel-Cantelli lemma,

$$\mu_1 [x \,|\, |\theta_k(x)| \leq n, \text{ infinitely many } k] = 0$$

for every $n > 0$. That is,

$$\mu_0 [x \,|\, \lim |\theta_k(x)| = 0] = 0$$

$$\mu_1 [x \,|\, \lim |\theta_k(x)| = \infty] = 1$$

5. Let $\{\theta_j(x)\}$ be any sequence of real-valued $\mathcal{B}$-measurable functions on the space of sample functions which are independent Gaussian random variables with respect to both μ_0 and μ_1, and which satisfy

$$E_0 \theta_j = E_1 \theta_j = 0$$

$$E_0 \theta_j^2 = \alpha_j > 0$$

$$E_1 \theta_j^2 = \beta_j > 0, \quad j = 1,2,\ldots$$

α_j and β_j arbitrary positive numbers. Then the measures μ_0' and μ_1' induced by μ_0 and μ_1 on the Borel field generated by the $\{\Theta_j\}$ are either mutually singular or equivalent. They are equivalent if and only if

$$\Sigma(1 - \frac{\alpha_j}{\beta_j})^2 < \infty.$$

Both statements follow from Kakutani's theorem. The first is immediate. For the second we need to calculate the product of the ρ_j defined in that theorem. Let ℓ_j be the likelihood ratio for Θ_j with respect to μ_0 and μ_1:

$$\ell_j = \exp \frac{1}{2} \left[\Theta_j^2 \left(\frac{1}{\beta_j} - \frac{1}{\alpha_j} \right) + \log \beta_j/\alpha_j \right].$$

Then

$$\rho_j = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sqrt{\ell_j} \exp - \frac{1}{2} \left[\frac{\Theta_j^2}{\beta_j} + \log \beta_j \right] d\Theta_j$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left[- \frac{1}{4} \Theta_j^2 \left(\frac{1}{\beta_j} + \frac{1}{\alpha_j} \right) + \frac{1}{4} \log \frac{1}{\alpha_j \beta_j} \right] d\Theta_j$$

$$= \sqrt{2} \frac{(\alpha_j \beta_j)^{1/4}}{(\alpha_j + \beta_j)^{1/2}}$$

Now, the convergence of the product

$$\prod_j \frac{\sqrt{2} (\alpha_j \beta_j)^{1/4}}{(\alpha_j + \beta_j)^{1/2}}$$

is equivalent to the convergence of the series

$$\sum_j \left(1 - \frac{4\alpha_j\beta_j}{(\alpha_j + \beta_j)^2}\right) = \sum_j \frac{(1 - \frac{\alpha_j}{\beta_j})^2}{(1 + \frac{\alpha_j}{\beta_j})^2}$$

The convergence of this series is equivalent to the convergence of

$$\sum (1 - \frac{\alpha_j}{\beta_j})^2$$

It follows immediately from this lemma that either $\mu_0 \perp \mu_1$ or $\sum (1 - \frac{\alpha_j}{\beta_j})^2 < \infty$, since $\mu_0' \perp \mu_1$ implies $\mu_0 \perp \mu_1$.

6. If $\sum_j (1 - \frac{\alpha_j}{\beta_j})^2 < \infty$, the Radon-Nikodym derivative of μ_0 with respect to μ_1 on the Borel field generated by the $\theta_j(x)$ is

$$\frac{d\mu_0'}{d\mu_1'} = \exp \frac{1}{2} \sum_j \left[\theta_j^2 (x) \left(\frac{1}{\beta_j} - \frac{1}{\alpha_j}\right) + \log \frac{\beta_j}{\alpha_j} \right]$$

This formula follows from Kakutani's theorem and the expression ℓ_j above.

We know that $S^{1/2}R^{-1/2}$ is densely defined. If it is also bounded, let X be its bounded extension to all of L_2 [0,1]. $R^{1/2}S^{-1/2}$ is also densely defined; if it is bounded, its extension is X^{-1}.

7. If $f_1, f_2, \ldots \epsilon L_2$ [0,1], there exist random variables $\theta_i(x)$, Gaussian with respect to both μ_0 and μ_1, satisfying

$$E_0\theta_i\theta_j = (f_i, f_j)$$

$$E_1\theta_i\theta_j = (X^*Xf_i, f_j)$$

Since $R^{-1/2}$ is densely defined for each i, i = 1,2,... there exists a sequence $\{f_{ij}\}_j$ such that $\lim_j f_{ij} = f_i$ and such that $h_{ij} = R^{-1/2} f_{ij}$ is defined. Let

$$\phi_{ij}(x) = \int_0^1 h_{ij}(t)\, x(t)\, dt$$

Then

$$\lim_{k,j \to \infty} E_0 \phi_{ij}\phi_{ik} = \lim_{k,j \to \infty} (Rh_{ij}, h_{ik}) = ||f_i||^2$$

and

$$\lim_{k,j \to \infty} E_1 \phi_{ij}\phi_{ik} = \lim_{k,j \to \infty} (Sh_{ij}, h_{ik}) = ||Xf_i||^2.$$

The existence of these limits implies that the sequences $\{\phi_{ij}\}_j$ have mean-square limits Θ_{i0} and Θ_{i1} with respect to both μ_0 and μ_1, and that Θ_{i0} and Θ_{i1} are measurable $\overline{\mathcal{B}}_0$ and $\overline{\mathcal{B}}_1$, respectively. It also follows that the $\{\phi_{ij}\}_j$ converge in mean-square with respect to $\mu_0 + \mu_1$ to elements Θ_i in $L_2(\mu_0 + \mu_1)$ and that $\Theta_{i0} = \Theta_i[\mu_0]$, $\Theta_{i1} = \Theta_i[\mu_1]$. Since Θ_{i0} and Θ_{i1} satisfy the second moment requirements, the Θ_i do also. The Θ_i are measurable with respect to $\overline{\mathcal{B}}_0$ and $\overline{\mathcal{B}}_1$.

We now state the main result:

Theorem 3. (Modified version of Feldman's theorem)

Either $\mu_0 \sim \mu_1$ or $\mu_0 \perp \mu_1$. A necessary and sufficient condition that $\mu_0 \sim \mu_1$ is that $X^*X = \sum \lambda_i P_i$, where each P_i is the projection on the one-dimensional subspace of $L_2[0,1]$ spanned by some f_i from an orthonormal sequence $\{f_i\}$, and $\sum(1 - \lambda_i)^2 < \infty$.

If $\mu_0 \sim \mu_1$ and random variables θ_i are formed from the f_i as in Lemma 7, then

$$x(t) = \Sigma(R^{1/2} f_i)(t)\ \theta_i(x)$$

almost everywhere dt $d\mu_0$ and dt $d\mu_1$, and

$$\frac{d\mu_0}{d\mu_1}(x) = \exp \frac{1}{2} \Sigma \theta_j^2(x)\left(\frac{1}{\lambda_j} - 1\right) + \log \lambda_j$$

We show first that if μ_0 and μ_1 are not totally singular then $X^*X = \Sigma\lambda_i P_i$, P_i one-dimensional, and $\Sigma(1 - \lambda_i)^2 < \infty$. For by Lemma 4 X is bounded, so X^*X has a spectral decomposition, $\int \lambda dP_\lambda$. Let I be the identity operator, and suppose that for some $\epsilon > 0$ $I - P_{1+\epsilon}$ is infinite dimensional. Then there exists an infinite sequence $\{\lambda_j\}$, $1 + \epsilon \equiv \lambda_1 < \lambda_2 < \ldots$, and normalized f_j's in $L_2[0,1]$ such that $(P_{\lambda_{k+1}} - P_{\lambda_k})\, f_k = f_k$. Hence by Lemma 7 there exist Gaussian random variables θ_k satisfying

$$E_0 \theta_j(x) \theta_k(x) = \delta_{jk}$$

and

$$E_1 \theta_j(x)\theta_k(x) = (X^*Xf_j,\ f_k) = \delta_{jk} \int_{\lambda_k}^{\lambda_{k+1}} \lambda d(P_\lambda f_k,\ f_k)$$

$$\geq (1 + \epsilon)\ \delta_{jk}$$

But then by Lemma 5, μ_0 and μ_1 would have to be totally singular on the Borel field generated by the θ_i's, which is a contradiction. Hence $I - P_{1+\epsilon}$ must be finite-dimensional for every $\epsilon > 0$. A similar argument shows that $P_{1-\epsilon}$ must be finite-dimensional for every $\epsilon > 0$. Hence X^*X has discrete spectrum and $X^*X = \sum \lambda_i P_i$, where the P_i are projections on the one-dimensional subspaces spanned by the f_i. If $\{\Theta_j(x)\}$ is a sequence of Gaussian random variables corresponding to $\{f_j\}$ as in Lemma 7, then by Lemma 5, μ_0 and μ_1 are equivalent when restricted to the Borel field $\mathfrak{F}(\Theta_i)$ generated by the Θ_j's, and $\sum(1 - \lambda_j)^2 < \infty$. Eq. (5) holds for the restriction of μ_0 and μ_1 to $\mathfrak{F}(\Theta_i)$ by Lemma 6.

It remains to prove the expansion of Eq. (4), for then by Lemmas 1 and 2 the equivalence of the restrictions of μ_0 and μ_1 to $\mathfrak{F}(\Theta_i)$ will imply the equivalence of μ_0 and μ_1. For the dt $d\mu_1$ case it is sufficient to show that

$$E_1 \int_0^1 dt \left| x(t) - \sum (R^{1/2} f_i)(t) \Theta_i(x) \right|^2 \tag{6}$$

converges to zero as $N \to \infty$. Now

$$\begin{aligned} E_1 x(t) \Theta_i(x) &= \lim_{j \to \infty} E_1 x(t) \phi_{ij}(x) \\ &= \lim_{j \to \infty} E_1 x(t) \int_0^1 h_{ij}(u)\, x(u)\, du = \lim_{j \to \infty} S h_{ij}(t) \\ &= \lim_{j \to \infty} S R^{-1/2} f_{ij}(t) \end{aligned}$$

Hence,

$$E_1 \int_0^1 x(t) (R^{1/2} f_i)(t) \Theta_i(x)\, dt$$

$$= \lim_{j \to \infty} (R^{1/2}f_i, SR^{-1/2}f_{ij}) = \lim_{j \to \infty} (XRf_i, Xf_{ij})$$

$$= (Rf_i, X^*Xf_i) = \lambda_i ||R^{1/2}f_i||^2$$

A similar verification shows that

$$E_1 \int_0^1 (R^{1/2}f_i)(t)\theta_i(x) \cdot (R^{1/2}f_j)(t)\theta_j(x)\, dt$$

$$= -\delta_{ij}\lambda_i ||R^{1/2}f_i||^2$$

Therefore Expression (6) above can be written

$$\int_0^1 R_1(t,t)\, dt - \sum_1^N \lambda_i ||R_0^{1/2}f_i||^2$$

We now show that this expression converges to zero. In fact, since $S = R^{1/2}X^*XR^{1/2}$,

$$\int_0^1 S(t,t)\, dt = \Sigma\, (Sf_i, f_i)$$

$$= \sum_i (X^*XR^{1/2}f_i, R^{1/2}f_i)$$

$$= \sum_i (\sum_j \lambda_j (R^{1/2}f_i, f_j)f_j, R^{1/2}f_i)$$

$$= \sum_i \sum_j \lambda_j (R^{1/2}f_i, f_j)^2 = \sum_j \lambda_j \sum_i (R^{1/2}f_i, f_j)^2$$

$$= \sum_j \lambda_j ||R^{1/2}f_i||^2$$

An analogous calculation shows that Eq. (4) holds almost everywhere dt $d\mu_0$, which completes the proof of the theorem.

One will observe that the proof just given is based on an infinite-dimensional analog of the simultaneous diagonalization of two covariance matrices.

The representation that results, and in terms of which the derivative is written, is perhaps interesting, but it is of limited usefulness because the Θ_i are not given explicitly. The restriction to processes with mean zero is not essential; neither Feldman nor Hájek required it, and it can be removed in the above.

The proof given here is somewhat similar to Feldman's. Hájek's proof is different, and is in fact essentially information-theoretic. Let $x_1, \ldots, x_N$ be measurable functions on Ω which are Gaussian random variables with respect to two different measures; and suppose they have probability densities $p(x_1, \ldots, x_N)$, $q(x_1, \ldots, x_N)$. The J-divergence (see Kullback and Leibler) of these two densities is defined as

$$J = E_p \log \frac{p}{q} - E_q \log \frac{p}{q} \tag{7}$$

where E_p, E_q denote expectation with respect to p and q measures. The first term of Eq. (7) can be interpreted as the information in p relative to q; hence, J can be interpreted as the sum of the information in p relative to q and the information in q relative to p. Now if $\{x_t,\ t \in T\}$ is a real-valued Gaussian process with respect to two different probability measures on Ω, the J-divergence of the processes is

$$J_T = \sup_{t_1, \ldots, t_n \in T} J_{t_1, \ldots, t_n}$$

Hájek's theorem states that the processes are singular iff J_T is infinite, intuitively a highly satisfying conclusion.

In addition to those already mentioned, there are papers by Middleton and Rozanov containing results similar or related to Theorem 3.*

*Other interesting results, not used here, on the differentiability and derivatives of measures corresponding to random processes are contained in Prokhorov (Appendix 2), Skorokhod, Pitcher. It should be noted that some of the material discussed can be regarded as a development of earlier work of Cameron and Martin, (not included in the Bibliography). Also it would appear to be closely related to parts of extensive work on functional integration by, e.g., Segal, Friedrichs, Gelfand (not included in the Bibliography).

III. SPECIAL RESULTS

An interesting consequence of Theorem 3 is:

Theorem 4. (Feldman)

If A_j and B_j are polynomials, with degrees respectively a_j and b_j, $j = 1,2$, and $b_j > a_j$, then the Gaussian processes (restricted to a finite parameter interval) whose spectral densities are $|A_j(x)/B_j(x)|^2$ have equivalent measures on path space if and only if

(a) $b_1 - a_1 = b_2 - a_2$

(b) the ratio of the leading coefficients of A_1 and B_1 has the same absolute value as the ratio of the leading coefficients of A_2 and B_2.

The necessity of these conditions was first shown by Slepian, using a theorem of Baxter. Baxter's theorem applied to stationary processes states that if x(t) is Gaussian, real-valued, with continuous covariance function possessing a bounded second derivative except at the origin and with mean-value function possessing a bounded derivative in [0,1] then

$$\sum_{n=1}^{2^n} \left[x\left(\frac{k}{2^n}\right) - x\left(\frac{k-1}{2^n}\right) \right]^2$$

converges with probability one to the difference between the right-hand and left-hand derivatives of the covariance function at the origin. Suppose two processes have rational densities which violate condition (a) of Theorem 4. Then if both processes are differentiated k times, $k = \min_{j=1,2} (b - a) - 1$,

the sum of squared differences will converge to zero for samples drawn from one differentiated process and to a number different from zero for the other, with probability one. If condition (b) is violated and (a) is satisfied, the sums will converge to different numbers not equal to zero. Slepian showed further that by using higher order differences an equivalent test for singularity can be obtained directly without first differentiating the processes.

The sufficiency (and a different proof of necessity) of the conditions of Theorem 4 was demonstrated by Feldman (1960). Feldman stated Theorem 4 as a corollary to a somewhat more general theorem in which only one of the processes involved need have a rational spectral density. This result was made to follow from his basic theorem referred to earlier, by techniques depending largely on certain properties of entire functions. Here we shall give a proof of the sufficiency of the conditions of Theorem 4 using Pitcher's conditions as stated in Theorem 3. The proof is an adaptation of Feldman's, modified to fit the different equivalence condition we are using. In particular we shall use Feldman's lemmas on entire functions without proof.

We assume to start with that both processes have mean value zero. The autocorrelation functions R(t,s) and S(t,s) are stationary and (with a slight abuse of notation) we write them as R(t-s) and S(t-s). They are defined for all real s,t, are integrable and of integrable square, and have rational Fourier transforms. The operators R and S on $L_2[-1,1]$ are defined as before. We also need now, however, to define operators R_0 and S_0 on $L_2(-\infty, \infty)$ by

$$(R_0 f)(t) = \int_{-\infty}^{\infty} R(t - s)f(s)\, ds, \qquad -\infty < t < \infty$$

$$(S_o f)(t) = \int_{-\infty}^{\infty} S(t - s) f(s)\, ds, \qquad -\infty < t < \infty$$

Inner products and norms on $L_2[-1,1]$ will be denoted by $(.,.)$, $|| \cdot ||$ and on $L_2(-\infty,\infty)$ (which will be written just L_2) by $(.,.)_o$, $|| \cdot ||_o$, respectively. The Fourier transform $\mathfrak{F}(f)$ (in whatever sense it may be defined) of a function f will be denoted by $\hat{f}$. We now proceed with a series of lemmas.

1. If f, g $\in L_2$ and are supported on $[-1,1]$ then

$$(Rf,g) = (R_o f, g)_o$$
$$(Sf,g) = (S_o f, g)_o$$

2. If f, g $\in L_2$,

$$(R_o f, g)_o = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R(t - s)\, f(s)\, \overline{g(t)}\, ds\, dt$$
$$= \int_{-\infty}^{\infty} \hat{R}(\mu)\, \hat{f}(\mu)\, \overline{\hat{g}(\mu)}\, d\mu$$

and analogous formulas hold for $(S_o f, g)_o$.

3. The operator R_o is Hermitian, positive-definite, and has a positive-definite square root $R_o^{1/2}$ which satisfies

$$(R_o^{1/2} f, g)_o = \int_{-\infty}^{\infty} (\hat{R}(\mu))^{1/2} \hat{f}(\mu)\, \overline{\hat{g}(\mu)}\, d\mu$$

We now further specialize the autocorrelation function R(t). In particular, let

$$\hat{R}(x) = \frac{1}{(1 + x^2)^u}, \qquad u \text{ an integer} \geq 1$$

Let $p(x) = (i + x)^u$, then

$$\hat{R}(x) = \frac{1}{|p(x)|^2}$$

and

$$(R_0 f, g)_0 = \int_{-\infty}^{\infty} \hat{f}(\mu)\ \hat{g}(\mu)\ \frac{d\mu}{|p(\mu)|^2}$$

The operator R_0 has an inverse R_0^{-1} which is unbounded but densely defined on L_2. Where defined,

$$R_0^{-1} f = \mathfrak{F}^{-1}\{|p(\mu)|^2 \hat{f}(\mu)\}$$

Let us now define operators $R_0^{-1/2}$, Q by

$$R_0^{-1/2} f = \mathfrak{F}^{-1}\{|p(\mu)|\hat{f}(\mu)\}$$

$$Qf = \ ^{-1}\{p(\mu)f(\mu)\}$$

for all f for which the expressions in brackets belong to L_2. Here, $\mathfrak{F}^{-1}$ is the inverse Fourier transform in the sense of Plancherel theory. One notes immediately that $(Qf, Qg)_0 = (R_0^{-1} f, g)_0$ when either side exists.

By the conditions on S, we can write

$$\hat{S}(x) = \left|\frac{A(x)}{B(x)}\right|^2$$

where A(x), B(x) are polynomials, deg (B) - deg (A) $\geqq$ 1 and there are no poles on the real axis.

4. Let deg (B) - deg (A) = u. Then $|p(x)|^2\ [\hat{R}(x) - \hat{S}(x)]$ has a $\mathfrak{F}^{-1}$ - transforms $\psi(t)$ in L_2, and

$$\int_{-1}^{1}\int_{-1}^{1} \psi|(t - s)|^2 \, dt \, ds = a^2 < \infty$$

Proof: The inverse transform exists in the Plancherel sense, since

$$\frac{1}{|p(x)|^2} - \left|\frac{A(x)}{B(x)}\right|^2 = \frac{1}{|p(x)|^2} \cdot \frac{P(x)}{|B(x)|^2}$$

where $\frac{P(x)}{|B(x)|^2} \in L_2$. The second assertion is a trivial consequence.

Now let $\mathcal{D}$ denote the class of functions belonging to C_∞ for which the closure of their supports is contained in (-1,1).

5. Let $f \in \mathcal{D}$. Then $p(\frac{d}{dx})\, f \in \mathcal{D}$, and

$$\mathfrak{F}\left\{p(\frac{d}{dx})f\right\} = p(u)\,\hat{f}(u)$$

Furthermore, $p(u)\,\hat{f}(u) \in L_2$ and is of exponential type.

6. Let $\{f_n\}$ be a complete orthonormal sequence (c.o.n.s.) for $L_2[-1,1]$, $f_n \in \mathcal{D}$. Let $\hat{f}_n = \mathfrak{F}(f_n)$, $\hat{g}_n = p\hat{f}_n$. Then

$$\sum_{n,m=1}^{\infty} \left|(R_o\hat{g}_n , \hat{g}_m)_o - (S_o\hat{g}_n, \hat{g}_m)_o\right|^2 = a^2$$

Proof:

$$(R_o\hat{g}_n, \hat{g}_m)_o - (S_o\hat{g}_n, \hat{g}_m)_o = \int_{-\infty}^{\infty} \hat{f}_n(x)\overline{\hat{f}_m(x)} \left[1 - |p(x)|^2\left|\frac{A(x)}{B(x)}\right|^2\right] dx$$

$$= \int_{-\infty}^{\infty} f_n(t) \int_{-\infty}^{\infty} \overline{f_m(s)}\; \overline{\psi(t - s)} \, ds \, dt$$

$$= \int_{-1}^{1}\int_{-1}^{1} f_n(t)\; \overline{f_m(s)}\; \overline{\psi(t - s)} \, dt \, ds$$

But $f_n(t)\,\overline{f_n(s)}$ is a c.o.n.s. in $L_2([-1,1]x[-1,1])$, hence

$$\sum_{n,m=1}^{\infty} \left|(R_o\hat{g}_n, \hat{g}_m)_o - (S_o\hat{g}_n, \hat{g}_m)_o\right|^2 = \int_{-1}^{1}\int_{-1}^{1} |\psi(t - s)|^2 \, dt \, ds = a^2$$

7. Let $A = S_o^{1/2}Q$. Then

$$\sum_{n,m=1}^{\infty} \left|\Big((I - A^*A)f_n, f_m\Big)_o\right|^2 = a^2$$

Proof: This follows from Lemma 6 since

$$\begin{aligned}\Big((I - A^*A)f_n, f_m\Big)_o &= (f_n, f_m) - (Af_n, Af_m)_o \\ &= (R_o\hat{g}_n, \hat{g}_m)_o - (S_o\hat{g}_n, \hat{g}_m)_o\end{aligned}$$

8. The sequence $\{z_n\}$, $z_n = R^{1/2}Qf_n$ is an o.n.s. in $L_2[-1,1]$.

Proof: Qf_n is defined and has its support contained in $(-1,1)$. Hence $R^{1/2}Qf_n$ is defined. Then

$$\begin{aligned}(z_n, z_m) &= (R^{1/2}Qf_n, R^{1/2}Qf_m) = (RQf_n, Qf_m) \\ &= (R_oQf_n, Qf_m)_o = (f_n, f_m)\end{aligned}$$

by Lemmas 1 and 3.

9. If E is the closed subspace of $L_2[-1,1]$ spanned by the z_n, then $L_2[-1,1] \ominus E$ is finite dimensional.

Proof: Let $Y = L_2[-1,1] \ominus E$. Then $y \in Y$ if and only if $(z_n, y) = (R^{1/2}Qf_n, y) = (Qf_n, R^{1/2}y) = 0$, $n = 1,2,\ldots$

We know that the orthogonal complement of the closed subspace spanned by $\{Qf_n\}$ is finite dimensional, say of dimension N—by Feldman (1960)—Lemma 5. So now suppose that Y is of dimension greater than N. Then there are

$y_k \in Y$, $k = 1,2,\ldots, N + 1$, such that for any choice of numbers α_k not all zero $\sum_1^{N+1} \alpha_k y_k \neq 0$. Hence

$$R^{1/2} \left[\sum_1^{N+1} \alpha_k y_k \right] = \sum_1^{N+1} \alpha_k (R^{1/2} y_k) \neq 0$$

by the strict definiteness of R and hence of $R^{1/2}$. Since $R^{1/2}y_k \neq 0$, this contradicts the fact just stated that the orthogonal complement of the subspace spanned by $\{Qf_n\}$ has dimension N. Hence Y is of dimension N.

10. The operator $S^{1/2}R^{-1/2}$ is defined and bounded on a dense subset of $L_2[-1,1]$ and hence has a bounded extension X with $\mathcal{D}(X) = L_2[-1,1]$. The bounded self-adjoint operator $I - X^*X$ is Hilbert-Schmidt on $L_2[-1,1]$.

Proof: From Lemma 7 it follows routinely that A is bounded. Since

$$(S^{1/2}R^{-1/2}z_i, S^{1/2}R^{-1/2}z_j) = (S^{1/2}Qf_i, S^{1/2}Qf_j)$$

$$= (SQf_i, Qf_j) = (S_oQf_i, Qf_j)_o = (S_o^{1/2}Qf_i, S_o^{1/2}Qf_j)_o$$

$$= (Af_i, Af_j)_o$$

one has $||Xz_n|| = ||Af_n||_o \leq B$. Hence X is densely defined and bounded on the closed linear manifold E spanned by the z_n, and can be extended to a bounded operator on E. Furthermore $S^{1/2}R^{-1/2}$ is densely defined on the finite-dimensional subspace $L_2[-1,1] \ominus E$. Hence $S^{1/2}R^{-1/2}$ has a unique bounded extension X with domain $L_2[-1,1]$.

In order to prove the second assertion we augment the o.n. sequence $\{z_n\}$, $n = 1,2,\ldots$, with elements $z_{-N+1}, z_{-N+2},\ldots,z_o$ so that $\{z_n\}$, $n = -N, -N+1,\ldots$ is a c.o.n.s. for $L_2[-1,1]$.

Then

$$\sum_{\substack{i=-N+1 \\ j=-N+1}}^{\infty} \left|((I - X^*X)z_i, z_j)\right|^2 = \sum_{i,j=1}^{\infty} + \sum_{\substack{i=N+1,\ldots,0 \\ j=-N+1,\ldots,\infty}} + \sum_{\substack{j=N+1,\ldots,0 \\ i=-N+1,\ldots,\infty}} + \sum_{\substack{i=-N+1,\ldots,0 \\ j=-N+1,\ldots,0}} \left|((I - X^*X)z_i, z_j)\right|^2$$

By the preceding calculation, the first sum on the right is equal to

$\sum_{i,j=1}^{\infty} \left|((I - A^*A)f_i, f_j)\right|^2 = a^2$. The second and third sums are finite since $\sum_j \left|((I - X^*X)z_k, z_j)\right|^2 = ||(I - X^*X)z_k||^2 = ||I - X^*X||$, and the fourth sum is obviously finite. Thus $I - X^*X$ is Hilbert-Schmidt.

The sufficiency part of Theorem 4 now follows directly from Theorem 3.

Although there are various criteria for the equivalence of Gaussian measures, Theorem 4 is particularly apt for noise-in-noise detection theory problems because it states a criterion for equivalence that is fairly general and is explicitly in terms of properties of the autocorrelation functions. Results of this kind for wider classes of processes would be useful.

For discussing singularity and equivalence in sure-signal in noise problems, the following theorem can be used in connection with Theorem 2.

<u>Theorem 5. (Kelly, Reed, and Root)</u>

Let $R(t)$ be a stationary, continuous autocorrelation function with the properties:

(1) $\int_{-\infty}^{\infty} |R(t)| \, dt < \infty$

(2) The integral operator defined by

$$R_T f(t) = \int_{-T}^{T} R(t - u)\, f(u)\, du$$

is strictly positive definite for every T.

Let $\{\phi_{n,T}\}$, $\{\lambda_n(T)\}$ be respectively a c.o.n.s. of eigenfunctions and the set of associated eigenvalues of R_T. Then if $s(t) \in L_2$, $s_n(T) = (s, \phi_{n,T})$, $\hat{s}(\mu)$ is an L_2 - Fourier transform of s(t), and $\hat{R}(\mu)$ is the Fourier transform of R(t),

$$\sum_{n=1}^{\infty} \frac{|s_n(T)|^2}{\lambda_n(T)} \uparrow \int_{-\infty}^{\infty} \frac{|\hat{s}(\mu)|^2}{\hat{R}(\mu)}\, d\mu, \text{ as } T \to \infty$$

in the sense that the left-hand side converges monotonically if the right-hand side exists and diverges monotonically to $+\infty$ otherwise.

One can show by example that the sum on the left side above may be finite for fixed T while the integral on the right diverges, even with the support of s(t) contained in (-T,T).

A recurring hypothesis in the preceding discussion has been that if $\{x_t\}$ is a stationary random process with autocorrelation function R(t), the integral operator R_T as defined above is strictly definite, or what is equivalent, $R_T f = 0$ implies $f = 0$. For a large class of processes this is true; an essentially well-known sufficient condition, useful for our purposes is the following theorem.

Theorem 6.

Let the random process $\{x_t, -\infty < t < \infty\}$ be defined by the stochastic integral

$$x(t) = \int_{-\infty}^{\infty} h(t - u)\, d\zeta(u)$$

where $\{\zeta_t\}$ is a Brownian motion, and h is a real-valued function in L_2. Then if $R(t) = E\, x_u x_{u+t}$, the operator R_T, $T > 0$, is strictly positive definite.

The proof follows easily from inspection of $(R_T f, f)$ written in terms of the Fourier transforms of $R(t)$ and $f(t)$.

IV. SUITABILITY OF THE STATIONARY GAUSSIAN MODEL

As remarked earlier, it seems unreasonable to expect that arbitrarily small error probabilities can be achieved in a radio communication or radio measurement system, which is what Theorems 2 and 4 might appear to show if the Gaussian model is to be believed. The two most commonly offered explanations of why these results do not really violate intuition are first, that the measurements are always inaccurate, and second, that the a priori data are always imperfect—in particular, autocorrelation functions and spectra are not completely or precisely known. Both explanations are obviously true statements, but I feel they do not meet the objection raised. Neither shows the existence of an absolute lower bound on error probabilities. With enough care and elaboration in obtaining a priori data and in making and processing the measurements, it would seem that arbitrarily good performance could still be achieved in some instances. So, although these points are important, I shall try to explain away the paradox of the singular cases in a different way, in fact in the simplest way possible, by showing the existence of constraints that prevent their occurrence. The essence of the explanation is that in all cases we know about, singularity occurs only if the spectral densities of the two signal plus noise processes differ at infinity, but a reasonable model of the problem indicates that the spectral densities at infinity are always determined by the residual noise, and hence are the same for both.*

*This idea appears in Davenport and Root, in Middleton, and is developed at some length in Wainstein and Zubakov, Appendix III.

To fix the domain of the argument, consider the class of systems that may be represented as in Fig. 1. A signal $s'''(t)$ is generated, processed at the transmitter, sent through the channel, received, and processed at the receiver. Gaussian thermal noise is added everywhere,

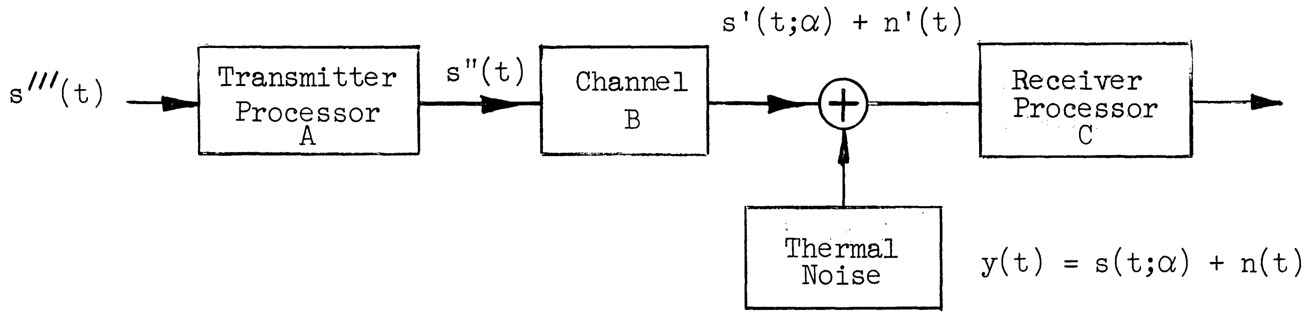

Fig. 1

but presumably the most important increment of noise is added at the point where the signal power level is lowest, at the input to the receiver, as indicated in the figure. The generated signal, $s'''(t)$, has finite energy, that is $\int |s(t)|^2\, dt < \infty$, and begins and ends in a finite time interval. It is arbitrary, but once chosen is fixed, even though we may let the observation interval, T, change. The processing at the transmitter and at the receiver must preserve the finite energy constraint and must be realizable in the usual sense that the present does not depend on the future. The channel must meet these same conditions; it may, however, perturb the signal into any one of a parametrized family of functions. The output of the receiver processor is the observed waveform, which is available for decision making. In different contexts the receiver processor might be taken to be a whole radio receiver in the usual sense; it might be only the antenna system at the receiver, or anything between these two extremes. In fact, in a particular instance there can be a good deal of arbitrariness about the breakdown into transmitter, channel, and receiver. Always, however, the noise has one property: there is at least a part, generated by thermal mechanisms, which can be thought of as <u>entering</u> the system as white noise, or as white up to frequencies at which quantum effects become important.

Let us look first at sure signals in noise. For one of the simplest situations the observed waveform is

$$y(t) = \alpha s(t) + n(t), \qquad 0 \leq t \leq T$$

where $n(t)$ is stationary, Gaussian, of mean zero and with a known continuous autocorrelation function $R(t)$, as prescribed for the Gaussian model; where $s(t)$ is known and of integrable square on $[0,T]$, and α is unknown but either zero or one. A statistical decision is to be made as to whether α is zero or one. As Grenander observed in 1950 this problem, with no further constraints imposed, can be singular in two ways. First, the integral operator R_T with noise autocorrelation as kernel may have a non-zero null space while $s(t)$ has a non-zero projection in this null space. Then there is an element $\psi \in L_2[0,T]$ such that $(\psi,\phi_n) = 0$, $n = 1,2,\ldots$, $\{\phi_n\}$ a complete set of eigenfunctions for R, but $(\psi,s) \neq 0$. Obviously, then, the statistic (ψ,y) will distinguish between the two hypotheses with probability one. Second, the series

$$\sum \frac{|s_n|^2}{\lambda_n}$$

may diverge, so that again, from Theorem 2, there is a test to distinguish between the two hypotheses with probability one. Suppose now, however, that the receiver processor C is linear as well as realizable and in fact can be represented by an integral operator with L_2 kernel $h(t)$. Then from Theorem 6 R has a zero null space, and the first kind of singularity mentioned above cannot happen. Let $\hat{h}(\mu)$ be the Fourier transform of $h(t)$ (i.e. $\hat{h}(\mu)$ is the so-called transfer function of C), then

$$\int_{-\infty}^{\infty} \frac{|\hat{s}(\mu)|^2}{\hat{R}^2(\mu)}\, d\mu = \int_{-\infty}^{\infty} \frac{|s'(\mu)|^2 |\hat{h}(\mu)|^2}{|\hat{h}(\mu)|^2}\, d\mu < \infty \tag{8}$$

so by Theorem 5 the second kind of singularity mentioned cannot happen either. Indeed, for any observation interval T,

$$\sum_n \frac{|s_n|^2}{\lambda_n} \leq \int_{-\infty}^{\infty} |\hat{s}'(\mu)|^2\, d\mu \tag{9}$$

and for a maximum-likelihood test (non-zero) error probabilities may be calculated depending only on the quantity on the left side of the inequality, which plays the role of a signal-to-noise ratio.

Now suppose the channel perturbs the signal by delaying it, shifting its frequency spectrum, changing its amplitude, etc. As long as it does not amplify the signal to give it infinite energy, a bound of the kind in Inequality (8) still exists, and the detection problem is non-singular. The situation is a little different if a radio measurement is to be made. The signal will be known to exist and a statistical estimate is to be made of the parameter α in $s(t;\alpha)$. Let α_1, α_2 be any two possible values of α (which may be vector-valued). Then the two Gaussian processes

$$y_t = s(t;\alpha_1) + n_t, \qquad 0 \leqq t \leqq T$$

$$y_t = s(t;\alpha_2) + n_t, \qquad 0 \leqq t \leqq T$$

are mutually singular if and only if

$$\sum_n \frac{|s_n(\alpha_1) - s_n(\alpha_2)|^2}{\lambda_n} = + \infty$$

Again, by an application of the Schwarz inequality, and with the conditions on the noise imposed above, this series cannot diverge if

$$\int_{-\infty}^{\infty} |s'(t;\alpha_i)|^2 \, dt < \infty, \qquad i = 1,2$$

as we have assumed. The conclusion does not depend on whether α is considered to be an unknown or a random variable.

Two weaknesses in the above argument are the assumptions that the receiver processing is linear and that the noise enters the system as pure white noise. Let us try to patch these up. First, the point of observation at which $y(t)$ is available after the noise has been introduced (actually noise is introduced everywhere, as mentioned) is arbitrary for purposes of discussion. Thus if it is possible to observe the processed received waveform at some point past the point of noise entry where the waveform is a linear functional of $s'(t;\alpha) + n'(t)$, $y(t)$ can be taken as the waveform at that point and the above arguments apply. No further processing of the sample functions can reduce the problem to a singular one.

Second, I suggest that there is no mechanism for generating the signal $s'''(t)$ so that the square of its Fourier transform falls off faster at infinity than thermally generated noise, and that the filtering action of the transmitter and channel is such as to attenuate the Fourier transform of the signal at high frequencies by more than the reciprocal of the frequency (the effect of a simple R-C filter). If this be true, then obvious modifications of Eq. (8) will restore the argument for non-singularity.

The discussion for noise in noise is similar to the foregoing, and can therefore be shortened. Consider the simple detection problem:

$$y(t) = \beta\, s_i(t) + n(t), \qquad 0 \leq t \leq T, \qquad i = 0,1$$

where $s_0(t) \equiv 0$ and $s_1(t)$ is a section of a sample function from a stationary Gaussian process with mean zero. β is a constant. We assume $\{s_{1t}\}$ and $\{n_t\}$ are mutually independent, so that $\{y_t\}$ is again a Gaussian process under either hypothesis. The only readily applicable criterion available for the singularity of two stationary Gaussian processes is that of Theorem 4; so we require the processors and channel as shown in Fig. 1 to be linear with rational transfer functions. Then if $\{n_t\}$ is white noise and $\{s_{it}\}$, $i = 0,1$, has rational spectral density, $\{y_t\}$ has rational spectral density under either hypothesis. If the transmitter and channel have an over-all transfer function which vanishes at least as the reciprocal of the frequency at infinity, then the behavior of the spectral density of $\{y_t\}$ at infinity is determined entirely by the noise, $\{n_t\}$, <u>under either hypothesis</u>. Thus by Theorem 4 the non-singular case obtains, for any observation interval T. Obviously, operations on the transmitted signal of translation (time delay) or amplification or linear combinations of these do not affect this conclusion.*

The aim here has not been to try to "prove" the faithfulness to reality of the Gaussian model, which would be foolish, but merely to try to rescue it from one rather important apparent difficulty. This seems to me to be important if the Gaussian model is to be used with confidence as a basis for more sophisticated analyses.

*The concept of band-limited noise, which is common in engineering literature, does not appear here. Actually, band-limited noise is a special case of the class of analytic Gaussian processes, which has been completely characterized by Belyaev. It is redundant to our argument, but perhaps of interest, to note that neither received signal nor noise can be analytic with the constraints adopted here. See Belyaev, Theorems 2 and 3.

BIBLIOGRAPHY

Baxter, G., "A Strong Limit Theorem for Gaussian Processes," Proc. A. M. S., v. 7, n. 3 (1956) pp. 522-528.

Belyaev, Yu. K., "Analytic Random Processes," Theory of Prob. and its Applications, v. IV, n. 4 (1959) pp. 402-409. (Translation.)

Davenport, W. and Root, W., "Introduction to the Theory of Random Signals and Noise," McGraw-Hill (1958), New York.

Davis, R., "On the Detection of Sure Signals in Noise," J. Appl. Phys., v. 25 (1933) p. 76.

Feldman, J., "Equivalence and Perpendicularity of Gaussian Processes," Pac. J. of Math., v. 8, n. 4 (1958), pp. 699-708.

Feldman, J., "Correction to Equivalence and Perpendicularity of Gaussian Processes," Pac. J. of Math., v. 9, n. 4 (1959), pp. 1295-1296.

Feldman, J., "Some Classes of Equivalent Gaussian Processes on an Interval," Pac. J. of Math., v. 10, n. 4 (1960), pp. 1211-1220.

Good, I. J., "Effective Sampling Rates for Signal Detection: or Can the Gaussian Model be Salvaged," Information and Control, v. 3 (1960), p. 116-140.

Grenander, U., "Stochastic Processes and Statistical Inference," Ark. f. Mat., v. 1 (1950), pp. 195-277.

Hájek, J., "On a Property of Normal Distribution of any Stochastic Process," Čy. Math. J., v. 8 (1958), pp. 610-617. (Also: Selected Translations in Math. Stat. and Prob., v. 1, pp. 245-253.)

Kakutani, S., "On Equivalence of Infinite Product Measures," Ann. of Math., v. 49 (1948).

Kelly, Reed, and Root, "The Detection of Radar Echoes in Noise, I," J. Soc. Indust. Appl. Math., v. 8, n. 2 (1960), pp. 309-341.

Kullback, S., and Leibler, R. A., "On Information and Sufficiency," Ann. Math. Stat., v. 22 (1951), pp. 79-86.

Middleton, D., "On Singular and Nonsingular Optimum (Bayes) Tests for the Detection of Normal Stochastic Signals in Normal Noise," I.R.E. Trans. IT-7, n. 2 (1961), pp. 105-113.

Pitcher, T., "Likelihood Ratios for Gaussian Processes," Unpublished M.I.T. Lincoln Laboratory memorandum.

Pitcher, T., "Likelihood Ratios of Gaussian Processes," Ark. f. Mat., v. 4, n. 5 (1959), pp. 35-44.

Pitcher, T., "Likelihood Ratios for Diffusion Processes with Shifted Mean Values," Trans. A.M.S., v. 101, n. 1 (1961), pp. 168-176.

Price, R., "Optimum Detection of Random Signals in Noise, with Application to Scatter - Multipath Communication, I," I.R.E. Trans. IT-2, n. 4 (1956), pp. 125-135.

Prokhorov, Yu. U., "Convergence of Random Processes and Limit Theorems in Probability Theory," Theory of Prob. and its Appl., v. 1, n. 2 (1956), pp. 157-214. (Translations.)

Rozanov, Yu. A., "On a Density of One Gaussian Distribution with Respect to Another," Teor. Veroj. i Prim., v. VII (1962), pp. 84-89.

Skorokhod, A. V., "On the Differentiability of Measures which Correspond to Stochastic Processes, I.," Theory of Prob. and its Appl., v. II, n. 4 (1957), pp. 407-432. (Translations.)

Slepian, D., "Some Comments on the Detection of Gaussian Signals in Gaussian Noise," I.R.E. Trans. IT-4, n. 2 (1958), pp. 65-68.

Wainstein, L. A. and Zubakov, V. D., "Extraction of Signals from Noise," (Translation from the Russian), Prentice-Hall (1962). Englewood Cliffs.