

# THE ELEMENTARY GAUSSIAN PROCESSES

By J. L. DOOB

*University of Illinois*

## 1. Introduction

One of the simplest interesting classes of temporally homogeneous stochastic processes is that for which the distributions of the defining chance variables  $\{x(t)\}$  are Gaussian. It is supposed that

(A) if  $t_1 < \dots < t_r$ , the multivariate distribution of  $x(t_1), \dots, x(t_r)$  is Gaussian,<sup>1</sup> and that

(B) this distribution is unchanged by translations of the  $t$ -axis.

The process is  $N$ -dimensional if  $x(t)$  is an  $N$ -tuple  $x_1(t), \dots, x_N(t)$ . The means  $E\{x(t)\}^2$  are independent of  $t$ , and will always be supposed to vanish in the following discussion.

The correlation matrix function  $R(t): (r_{ij}(t))$  is defined by

$$(1.1.1) \quad r_{ij}(t) = E\{x_i(s)x_j(s+t)\}.$$

This expectation is independent of  $s$ , because of condition (B). The matrix function  $R(t)$  satisfies the equation

$$(1.1.2) \quad r_{ij}(t) = r_{ji}(-t), \quad i, j = 1, \dots, N.$$

It follows that when  $t = 0$  the matrix is symmetric:

$$(1.1.3) \quad r_{ij}(0) = r_{ji}(0), \quad i, j = 1, \dots, N,$$

and it is also well known that  $R(0)$  is non-negative definite. Conditions on the functions  $r_{ij}(t)$  necessary and sufficient that  $R(t)$  be the correlation matrix function of a stochastic process were found for the case  $N = 1$  by Khintchine<sup>3</sup> and for all  $N$  by Cramér.<sup>4</sup>

Hypothesis (A), that the process is Gaussian seems at first a restriction so strong that Gaussian processes are unimportant. These processes are, however, of fundamental importance, for the following reasons.

(i) If  $R(t)$  is the correlation matrix function of any temporally homogeneous stochastic process, there is, according to Khintchine and Cramér, a Gaussian process with this same correlation function. This Gaussian process is uniquely determined by the correlation function (assuming that all first order moments vanish, as usual). Because of this intimate connection between the temporally homogeneous Gaussian processes and the most general temporally homogeneous

<sup>1</sup> Singular Gaussian distributions will not be excluded. For example the  $x(t_i)$  may all vanish identically.

<sup>2</sup> The expectation of a chance variable  $x$  will be denoted by  $E\{x\}$ .

<sup>3</sup> *Mathematische Annalen*, Vol. 109 (1934), p. 608.

<sup>4</sup> *Annals of Math.*, Vol. 41 (1940), pp. 215-230.

processes, it is not surprising that very few facts are known about specifically Gaussian processes, that is facts which are true of temporally homogeneous Gaussian processes, but not of temporally homogeneous processes in general.

(ii) It follows from (i) that in any investigation of temporally homogeneous stochastic processes involving only first and second moments—for example least squares prediction by linear extrapolation—it may be assumed that the variables are Gaussian. Under this hypothesis, the investigator may be helped by the suggestive specialized interpretations possible in the Gaussian case of results which hold in the general case. For example if  $N = 1$ , the least squares best prediction in the Gaussian case for  $x(n + 1)$  in terms of a linear combination of the variables  $x(1), \dots, x(n)$  is the conditional expectation of  $x(n + 1)$  for given  $x(1), \dots, x(n)$ , which is the least squares best prediction of  $x(n + 1)$  in terms of  $x(1), \dots, x(n)$  with no restriction on the functions involved. Thus the linearity of the prediction, which must be part of the hypothesis in the general case, is automatically true in the Gaussian case. There is necessarily a linear least squares best prediction of  $x(n + 1)$  in terms of the complete past  $\dots, x(n - 1), x(n)$  since the corresponding conditional expectation is certainly defined in the Gaussian case, and is linear in that case.

(iii) In many applications, there is a real justification for hypothesis (A) that the process is Gaussian. This is so in certain physical studies, for example, because the Maxwell distribution of molecular velocities is Gaussian. Examples will be given below.

The processes discussed in the present paper are all temporally homogeneous Gaussian processes. Most of the theorems will be valid for any temporally homogeneous processes for which the second moments of the variables exist,<sup>5</sup> with the following changes: independent chance variables which are linear combinations of the  $x(s)$  will become merely uncorrelated chance variables; the convergence with probability 1 of a series of such chance variables will become merely convergence in the mean; the conditional expectation of one such variable  $y$  for given values of others,  $y_1, y_2, \dots$  will become merely the linear approximation  $\sum_j a_j y_j$  of  $y$  in terms of the  $y_j$  which minimizes

$$E\{[y - \sum_j a_j y_j]^2\},$$

that is to say the conditional expectation becomes the least squares linear approximation.

The following theorem and its corollary are fundamental in the study of linear prediction involving infinitely many variables. The results are implicit in much of the work on the subject but do not seem to have been stated explicitly before.

**THEOREM 1.2.** *Let  $\dots, x_0, x_1, \dots$  be a sequence of one-dimensional Gaussian*

---

<sup>5</sup> The processes need not even be temporally homogeneous. It is necessary only that  $E\{x(s)\}$  and  $E\{x(s)x(s+t)\}$  be independent of  $s$ .

chance variables with the property that if  $n_1 < \dots < n_r$ , the multivariate distribution of  $x_{n_1}, \dots, x_{n_r}$  is Gaussian and that

$$(1.2.1) \quad E\{\dots, x_{m-1}, x_m; x_n\} = x_m^6$$

whenever  $m < n$ . Then  $E\{x_m\} = a$  is independent of  $m$ , and

$$(1.2.2) \quad \dots \leq E\{(x_m - a)^2\} \leq E\{(x_{m+1} - a)^2\} \leq \dots$$

If the  $\{x_n\}$  are defined for all negative integers,

$$(1.2.3) \quad \lim_{m \rightarrow -\infty} x_m = x_{-\infty}$$

exists with probability 1 and

$$(1.2.4) \quad \lim_{m \rightarrow \infty} E\{(x_{-\infty} - x_m)^2\} = 0.$$

If the  $\{x_n\}$  are defined for all positive integers, and if the dispersions in (1.2.2) form a bounded sequence,

$$(1.2.3') \quad \lim_{m \rightarrow \infty} x_m = x_{\infty}$$

exists with probability 1, and

$$(1.2.4') \quad \lim_{m \rightarrow \infty} E\{(x_{\infty} - x_m)^2\} = 0.^7$$

It follows from (1.2.1) that

$$(1.2.5) \quad E\{x_n\} = E\{E\{x_m; x_n\}\} = E\{x_m\}.$$

Hence  $\dots = E\{x_0\} = E\{x_1\} = \dots$ . It will be no restriction to assume from now on that

$$\dots = E\{x_0\} = E\{x_1\} = \dots = 0.$$

It also follows from (1.2.1) that

$$(1.2.6) \quad E\{x_m x_n\} = E\{E\{x_m; x_m x_n\}\} = E\{x_m E\{x_m; x_n\}\} = E\{x_m^2\}.^8$$

Using this equation,

$$(1.2.7) \quad E\{x_n^2\} = E\{[(x_n - x_m) + x_m]^2\} = E\{(x_n - x_m)^2\} + E\{x_m^2\},$$

and the dispersions of the  $x_n$  thus form a monotone non-decreasing sequence.

<sup>6</sup> The conditional expectation of a chance variable  $y$  for given values of a chance variable  $\eta$  will be denoted by  $E\{\eta; y\}$ .

<sup>7</sup> Much of this theorem remains true if (1.2.1) is true but only the first moments of the  $x_n$  are supposed finite, no other hypothesis being made on their distributions. Cf. Doob, *Am. Math. Soc. Trans.*, Vol. 47 (1940), pp. 458-460.

<sup>8</sup> Alternatively, (1.2.1) implies that  $x_n - x_m$  is uncorrelated with  $x_m$ . Then  $E\{x_m x_n\} = E\{x_m^2 + (x_n - x_m)x_m\} = E\{x_m^2\}$ .

Finally, using (1.2.6),

$$(1.2.8) \quad E\{(x_{m+1} - x_m)(x_{n+1} - x_n)\} = 0.$$

The series

$$(1.2.9) \quad \sum_m (x_{m+1} - x_m)$$

is therefore a series of mutually independent chance variables. According to a well known theorem of Kolmogoroff, a sequence of mutually independent chance variables converges with probability 1 if the means and dispersions form a convergent series. The present theorem follows at once from Kolmogoroff's theorem.

COROLLARY. *Let  $x$  be a one-dimensional Gaussian chance variable and let*

$$(1.2.10) \quad \begin{array}{l} \dots\dots\dots \\ x_{01}, x_{02}, \dots \\ x_{11}, x_{12}, \dots \\ x_{21}, x_{22}, \dots \\ \dots\dots\dots \end{array}$$

*be sequences of one-dimensional Gaussian chance variables with the property that if  $\nu \geq 1$ , the multivariate distribution of  $x, x_{m1}, \dots, x_{m\nu}$  is Gaussian, and suppose that each variable  $x_{mn}$  is a member of every later sequence. Then*

$$(1.2.11) \quad \begin{aligned} \lim_{m \rightarrow -\infty} E\{x_{m1}, x_{m2}, \dots; x\} &= x_- \\ \lim_{m \rightarrow +\infty} E\{x_{m1}, x_{m2}, \dots; x\} &= x_+ \end{aligned}$$

*exist with probability 1, and in the mean, and*

$$(1.2.12) \quad x_+ = E\{x_{mn}, m = 0, \pm 1, \dots, n = 1, 2, \dots; x\}.$$

It will first be shown that the sequence  $\{x_m\}$ , where

$$(1.2.13) \quad x_m = E\{x_{m1}, x_{m2}, \dots; x\},$$

has the property demanded in the theorem. In fact, from the definition of conditional expectation, the difference  $x - x_n$  has expectation zero and is independent of the variables  $\{x_{mj}\}$  for  $m \leq n$ , and therefore of the variables  $\dots, x_{n-1}, x_n$ . Hence

$$(x - x_n) - (x - x_{n+1}) = x_{n+1} - x_n$$

has expectation zero and is independent of the variables  $\dots, x_{n-1}, x_n$ . Therefore the sequence  $\{x_{n+1} - x_n\}$  is a sequence of mutually independent chance variables with vanishing expectations. This implies (1.2.1) if  $m < n$  because

$$(1.2.14) \quad \begin{aligned} E\{\dots, x_m : x_n\} &= E\left\{\dots, x_m; x_m + \sum_m^{n-1} (x_{j+1} - x_j)\right\} = E\{\dots, x_m; x_m\} \\ &+ \sum_m^{n-1} E\{\dots, x_m; x_{j+1} - x_j\} = x_m. \end{aligned}$$

Let  $a$  be the common value of  $E\{x\}$ ,  $E\{x_m\}$ . Since  $x - x_m$  is independent of  $x_m$ ,

$$(1.2.15) \quad E\{x_m - a\}^2 + E\{(x - x_m)^2\} = E\{(x - a)^2\}.$$

Hence the sequence of dispersions of the  $x_m$  is bounded and according to Theorem 1.2 the limits  $x_-$  and  $x_+$  in (1.2.1) exist with probability 1. Since  $x - x_n$  has expectation zero and is independent of  $x_m$  for  $m \leq n$ ,  $x - x_+$  also has expectation zero and is independent of  $x_m$  for all  $m$ , that is (1.2.12) is true.

The simplest non-trivial special case of this theorem is the following: Let  $x_1, x_2, \dots, x$  be one-dimensional Gaussian chance variables with the property that if  $\nu \geq 1$  the multivariate distribution of  $x, x_1, \dots, x_\nu$  is Gaussian. Then

$$(1.2.16) \quad \lim_{n \rightarrow \infty} E\{x_1, \dots, x_n; x\} = E\{x_1, x_2, \dots; x\},$$

with probability 1, and this limit is also a limit in the mean.

As stated, the theorem and corollary are true without the hypothesis that the chance variables concerned are Gaussian. (The existence of second moments must be assumed if the limits are to exist as limits in the mean.) They are stated for Gaussian variables because the proof is simple in that case, and because that is sufficient for the purposes of this paper.

In discussing t.h.G. processes whose parameter  $t$  is not restricted to be integral, the usual continuity hypothesis will be made. It will be supposed that  $R(t)$  is continuous at  $t = 0$ :

$$(1.3.1) \quad \lim_{t \rightarrow 0} [R(t) - R(0)] = -\frac{1}{2} \lim_{t \rightarrow 0} E\{[x(t) - x(0)]^2\} = 0.$$

It is then easily concluded that  $R(t)$  is everywhere continuous.

In the continuous parameter case, it would be useful to have the conditions on  $R(t)$  necessary and sufficient for the continuity in  $t$  of the chance variables  $x(t)$  and for the existence of the derivative. No set of necessary and sufficient conditions for the continuity of  $x(t)$  is known, although the fact of continuity will not be difficult to prove in the special cases to be considered in §4. The conditions for the existence of  $x'(t)$  are quite simple, and will be given in Theorem 1.4.

The spectral function of a one-dimensional t.h.G. process will play an essential role in some of the theorems to be discussed below. If  $R(n)$  is the correlation function of a one-dimensional t.h.G. process,  $R(n)$  can be expressed in either of the forms

$$(1.3.2) \quad R(n) = \int_0^\pi \cos n\lambda \, dF(\lambda) \quad n = 0, \pm 1, \dots,$$

$$(1.3.2') \quad R(n) = \int_{-\pi}^\pi e^{in\lambda} \, dG(\lambda) \quad n = 0, \pm 1, \dots,$$

where,  $F(\lambda)$ , called the spectral function of the process, and  $G(\lambda)$ , called the complex spectral function of the process, are real monotone non-decreasing functions satisfying the following conditions:

$$\begin{aligned}
 F(0) &= 0 & G(-\pi) &= 0 \\
 (1.3.3) \quad F(\lambda-) &= F(\lambda), \quad 0 < \lambda < \pi, & G(\lambda-) &= G(\lambda), \quad -\pi < \lambda < \pi \\
 & & G(\lambda) - G(0+) &= G(0) - G(-\lambda+)
 \end{aligned}$$

$$\begin{aligned}
 F(\pi) &= G(\pi) \\
 (1.3.4) \quad F(\lambda) &= G(\lambda) - G(-\lambda+) = 2G(\lambda) - G(0) - G(0+), \quad 0 < \lambda < \pi \\
 F'(\lambda) &= 2G'(\lambda)^9
 \end{aligned}$$

The last equation of course holds only at points where the derivatives exist. The forms (1.3.2), (1.3.2') are derived trivially from each other. The correlation function determines the spectral functions uniquely, if the latter are supposed to satisfy (1.3.3). In fact, at the points of continuity of  $F(\lambda)$ ,  $G(\lambda)$ :

$$\begin{aligned}
 (1.3.5) \quad F(\lambda) &= \frac{\lambda R(0)}{\pi} + \frac{2}{\pi} \sum_1^{\infty} R(n) \frac{\sin n\lambda}{n} \\
 G(\lambda) &= \frac{(\lambda + \pi)R(0)}{2\pi} + \frac{1}{2\pi i} \lim_{\nu \rightarrow \infty} \sum_{\substack{\nu \\ n \neq 0}} R(n) \frac{e^{in\lambda}}{n}.
 \end{aligned}$$

Conversely if any  $F(\lambda)$  or  $G(\lambda)$  satisfying the stated conditions is used to determine an  $R(n)$  by means of (1.3.2) or (1.3.2'),  $R(n)$  is the correlation function of a t.h.G. process. The representation of  $R(n)$  in terms of  $G(\lambda)$  is frequently more convenient than that in terms of  $F(\lambda)$ , because of the simple properties of the exponential function. The following relation, which will be used below, exhibits the elegance attained by the use of  $G(\lambda)$ :

$$\begin{aligned}
 (1.3.6) \quad E\{[\sum_m c'_m x(m)] \cdot [\sum_n c''_n x(n)]\} &= \sum_{m,n} c'_m c''_n R(m - n) \\
 &= \int_{-\pi}^{\pi} (\sum_m c'_m e^{im\lambda}) (\overline{\sum_n c''_n e^{in\lambda}}) dG(\lambda).
 \end{aligned}$$

The correlation function of a one-dimensional continuous parameter t.h.G. process can be represented in either of the following forms:

$$(1.3.7) \quad R(t) = \int_0^{\infty} \cos t\lambda \, dF(\lambda)$$

$$(1.3.7') \quad R(t) = \int_{-\infty}^{\infty} e^{it\lambda} \, dG(\lambda)$$

---

<sup>9</sup> H. Wold, *A Study in the Analysis of Stationery Time Series*, Uppsala, (1938), p. 66.

where the spectral function  $F(\lambda)$  and the complex spectral function  $G(\lambda)$  are monotone non-decreasing and satisfy the conditions

$$\begin{aligned}
 (1.3.8) \quad & F(0) = 0 & G(-\infty) &= 0 \\
 & F(\lambda-) = F(\lambda), \quad 0 < \lambda < \infty & G(\lambda-) &= G(\lambda) \\
 & & G(\lambda) - G(0+) &= G(0) - G(-\lambda+)
 \end{aligned}$$

$$\begin{aligned}
 (1.3.9) \quad & F(\infty) = G(\infty) \\
 & F(\lambda) = G(\lambda) - G(-\lambda+) = 2G(\lambda) - G(0) - G(0+), \quad 0 < \lambda < \infty, \\
 & F'(\lambda) = 2G'(\lambda).
 \end{aligned}$$

The last equation of course only holds at points where the derivatives exist. The correlation function  $R(t)$  determines the spectral functions uniquely if the latter are supposed to satisfy (1.3.8). In fact, at the points of continuity of  $F(\lambda), G(\lambda)$ :

$$\begin{aligned}
 (1.3.10) \quad & F(\lambda) = \frac{2}{\pi} \int_0^\infty R(t) \frac{\sin t\lambda}{t} dt \\
 & G(\lambda) = \lim_{T \rightarrow \infty} \frac{1}{2\pi i} \int_{-T}^T R(t) \frac{e^{i t \lambda} - 1}{t} dt + G(0).
 \end{aligned}$$

**THEOREM 1.4.** *Let  $\{x(t)\}$  be the variables of a one-dimensional continuous parameter t.h.G. process with correlation function  $R(t)$  and spectral function  $F(\lambda)$ . If*

$$(1.4.1) \quad \int_0^\infty \lambda^2 dF(\lambda) < \infty$$

then

- (i)  $R'(t), R''(t)$  exist and are continuous, and  $R'(0) = 0$ ;
- (ii)  $x(t)$  is an absolutely continuous function of  $t$ , with probability 1;
- (iii) for each  $t$ ,

$$(1.4.2) \quad \lim_{h \rightarrow 0} \frac{x(t+h) - x(t)}{h} = x'(t)$$

exists, with probability 1, and this convergence is also true in the mean:

$$(1.4.3) \quad \lim_{h \rightarrow 0} E \left\{ \left[ \frac{x(t+h) - x(t)}{h} - x'(t) \right]^2 \right\} = 0;$$

(iv) the  $x'(t)$  process is a t.h.G. process, with correlation function  $-R''(t)$  and spectral function  $\int_0^\lambda \lambda^2 dF(\lambda)$ .

Conversely if

$$(1.4.4) \quad \liminf_{h \rightarrow 0} \frac{R(0) - R(h)}{h^2} = \frac{1}{2} \liminf_{h \rightarrow 0} E \left\{ \left[ \frac{x(h) - x(0)}{h} \right]^2 \right\} < \infty,$$

then (1.4.1) is true.

This theorem is due to Slutsky. The proof will be sketched here, for completeness. (The hypothesis that the process is a Gaussian process is immaterial, since only the second moments are involved in the proof.)

*Proof of (i).* If the integral (1.4.1) exists,  $R'(t)$ ,  $R''(t)$  can be obtained by differentiating under the integral sign in (1.2.2):

$$(1.4.5) \quad R'(t) = - \int_0^\infty \lambda \sin t\lambda \, dF(\lambda)$$

$$R''(t) = - \int_0^\infty \lambda^2 \cos t\lambda \, dF(\lambda).$$

Then  $R'(t)$ ,  $R''(t)$  are continuous functions, and  $R'(0) = 0$ .

*Proof of (ii), (iii), (iv).* The quantity

$$(1.4.6) \quad E \left\{ \left[ \frac{x(t+h_1) - x(t)}{h_1} - \frac{x(t+h_2) - x(t)}{h_2} \right]^2 \right\}$$

can be evaluated in terms of the correlation function  $R(t)$ , and approaches 0 with  $h_1, h_2$ , if the second derivative  $R''(t)$  exists. There is therefore a chance variable  $y(t)$  to which the difference quotient converges in the mean:

$$(1.4.7) \quad \lim_{h \rightarrow 0} E \left\{ \left[ \frac{x(t+h) - x(t)}{h} - y(t) \right]^2 \right\} = 0.$$

The  $y(t)$  process is a t.h.G. process. Moreover the equation

$$(1.4.8) \quad E\{x(s)x(s+t)\} = R(t)$$

can be differentiated to give

$$(1.4.9) \quad E\{x(s)y(s+t)\} = E\{x(s-t)y(s)\} = R'(t)$$

and this in turn when differentiated becomes

$$(1.4.10) \quad E\{y(s-t)y(s)\} = E\{y(s)y(s+t)\} = -R''(t).$$

Hence the  $y(t)$  process has correlation function  $-R''(t)$ . Finally,

$$(1.4.11) \quad E \left\{ \left[ x(t) - x(0) - \int_0^t y(s) \, ds \right]^2 \right\} = E\{[x(t) - x(0)]^2\} \\ + \int_0^t \int_0^t E\{y(s)y(s')\} \, ds \, ds' - 2 \int_0^t E\{[x(t) - x(0)]y(s)\} \, ds = 0,$$

(evaluating the right side of (1.4.11) in terms of  $R(t)$ ,  $R'(t)$ ,  $R''(t)$ ). Thus  $x(t)$  is absolutely continuous, with probability 1, and  $y(t)$  is the derived function  $x'(t)$ . Hence  $x'(t)$  exists for almost all  $t$ , with probability 1.<sup>10</sup> It follows (Fubini's theorem) that the limit in (1.4.2) exists for each  $t$ , with probability 1, except possibly for a  $t$ -set of Lebesgue measure 0. Since the process is t.h., the

<sup>10</sup> For the exact meaning and measure-theoretic justification for statements of this type, see Doob, *Am. Math. Soc. Trans.*, Vol. 42 (1937), pp. 107-40.



exceptional set must be either empty or the whole  $t$ -line. The exceptional set is therefore empty.

Conversely if (1.4.4) is true, (1.4.1) follows at once from (1.3.7).

It will be convenient to use condensed notation below. If  $x: (x_1, \dots, x_N)$ ,  $y: (y_1, \dots, y_N)$  are  $N$ -dimensional vectors and if  $A: (a_{ij})$  is an  $N$ -dimensional square matrix,  $x \cdot y$  will denote the matrix  $(x_i y_j)$ ,  $Ax$  the vector with components  $\sum_j a_{ij} x_j$  and  $(x, y)$  the number  $\sum_i x_i y_i$ . The adjoint matrix  $(a_{ij}^*): a_{ij}^* = \bar{a}_{ji}$  will be denoted by  $A^*$ . Throughout this paper, the chance variables will be real-valued, but it will be convenient to use complex constant vectors. The identity matrix will be denoted by  $I$ . It will be convenient to denote the  $i, j$ th term of the matrix  $A$  by  $(A)_{ij}$ . The following equations will be used frequently:

$$Ax \cdot By = A(x \cdot y)B^*, \quad (Ax, y) = (x, A^*y).$$

If  $x$  is a chance variable, it is clear that  $E\{x \cdot x\}$  is a symmetric non-negative definite matrix.

The simplest Gaussian processes are those in which the distribution of future states is based not on the complete past, but only on the immediate present. The precise definition of this (Markoff) property is the following.

(C) If  $t_1 < \dots < t_{\nu+1}$  the conditional distribution of  $x(t_{\nu+1})$  for given values of  $x(t_1), \dots, x(t_\nu)$  depends only on the value assigned to  $x(t_\nu)$ . The conditional distribution of  $x(t_{\nu+1})$  for given values of  $x(t_1), \dots, x(t_\nu)$  will then be simply the conditional distribution of  $x(t_{\nu+1})$  for the assigned value of  $x(t_\nu)$ .

The processes to be discussed in this paper are the processes with properties (A), (B), (C): temporally homogeneous Gaussian Markoff (t.h.G.M.) processes. The properties of t.h.G.M. processes will also be used to derive properties of the most important simple types of one-dimensional t.h.G. processes—those with rational spectral density functions. Some of the results are contained implicitly in the work of previous writers, but the presentation of the results has in all cases been chosen to stress their specific probability significance, and may therefore appeal even to readers familiar with previous work.

The condition (C) on a Gaussian process is equivalent to the condition (C') that if  $t_1 < \dots < t_{\nu+1}$

$$(1.5.1) \quad E\{x(t_1), \dots, x(t_\nu); x(t_{\nu+1})\} = E\{x(t_\nu); x(t_{\nu+1})\}.$$

In fact (C) is at least as strong as (C'). Conversely if (C') is true,

$$(1.5.2) \quad \begin{aligned} x(t_{\nu+1}) &= x(t_{\nu+1}) - E\{x(t_\nu); x(t_{\nu+1})\} + E\{x(t_\nu); x(t_{\nu+1})\} \\ &= y + E\{x(t_\nu); x(t_{\nu+1})\}, \end{aligned}$$

where  $y$  is a Gaussian chance variable with mean 0 uncorrelated with and therefore independent of  $x(t_1), \dots, x(t_\nu)$ , and the last term of (1.5.2) is simply a multiple of  $x(t_\nu)$ . Then the conditional distribution of  $x(t_{\nu+1})$  for given  $x(t_1), \dots, x(t_\nu)$  is a Gaussian variable, with mean  $E\{x(t_\nu); x(t_{\nu+1})\}$  and dispersion that of  $y$ . Since this conditional distribution depends only on  $x(t_\nu)$ , property (C')

implies property (C). Hence these properties are equivalent. The condition (C') can be written in the form

$$(1.5.3) \quad E\{x(\tau), \tau \leq s; x(s+t)\} = E\{x(s); x(s+t)\}, \quad t > 0.$$

In many applications the stochastic processes either have this property already or will have it if the dimensionality of the processes is increased by the adjunction of auxiliary chance variables. In the latter case the process is called a component process of a t.h.G.M. process. Component processes are discussed in detail below. If a process is a t.h.G. process, the right side of (1.5.3) is a linear transformation (depending only on  $t$ ) of  $x(s)$ :

$$(1.5.4) \quad E\{x(s); x(s+t)\} = A(t)x(s), \quad t > 0.$$

The matrix function  $A(t)$  will be called the transition matrix function. It satisfies the equation (obtained by performing the operation  $E\{x(s) \cdot \}$  on both sides of (1.5.4))

$$(1.5.5) \quad R(t) = R(0)A(t)^*, \quad t > 0,$$

but is otherwise unrestricted since if (1.5.5) is true, the difference  $x(t) - A(t)x(s)$  is uncorrelated with and therefore independent of  $x(s)$ . In many applications the elements of  $R(t)$  will vanish identically except in square matrices down the main diagonal. If this is true,  $A(t)$  can also be assumed in this form.

If the variables  $\{x(t)\}$  determine an  $N$ -dimensional t.h.G.M. process, and if  $B$  is a non-singular  $N$ -dimensional square matrix, the variables  $\{Bx(t)\}$  also determine a t.h.G.M. process. Two processes connected in this way will be called equivalent. If two t.h.G. processes are equivalent, and if one is a Markoff process, the other must be also. If there is a change of variable

$$(1.5.6) \quad y(t) = Bx(t).$$

taking the t.h.G.M.  $x(t)$  process with transition matrix  $A(t)$  and correlation matrix  $R(t)$  into the equivalent  $y(t)$  process with transition matrix  $A_1(t)$  and correlation matrix  $R_1(t)$ , then

$$(1.5.7) \quad A_1(t) = BA(t)B^{-1}, \quad R_1(t) = BR(t)B^*.$$

If  $\{x(t)\}$ ,  $\{y(t)\}$ ,  $\{z(t)\}$  determine t.h.G. processes of dimensions  $\alpha$ ,  $\beta$  and  $\alpha + \beta$  respectively, if the process determined by

$$\{x_1(t), \dots, x_\alpha(t), \quad y_1(t), \dots, y_\beta(t)\}$$

is equivalent to the  $z(t)$  process, and if every  $x(s)$  is independent of every  $y(t)$ , the  $z(t)$  process will be called the direct product of the  $x(t)$  and  $y(t)$  processes. The extension of the definition to direct products of more than two processes is clear. If the  $x(t)$  and  $y(t)$  processes are Markoff processes, their direct product is also a Markoff process. Conversely if the  $z(t)$  process is a Markoff process, the factor processes must also be Markoff processes. The following facts about matrices will be used below. If  $A$  is any  $N$ -dimensional matrix, there is a non-

singular  $N$ -dimensional matrix  $B$  such that  $B^{-1}AB$  is in Jordan canonical form: the elements of  $B^{-1}AB$  vanish except for those in certain submatrices down the main diagonal. Each of these submatrices has the form

$$(1.5.8) \quad \begin{pmatrix} \lambda & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & \lambda & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & \lambda \end{pmatrix}$$

or simply  $(\lambda)$  if it is one-dimensional. The  $\lambda$ 's are the characteristic values of  $A$ , that is the roots of the characteristic equation  $\det. |A - \lambda I| = 0$ , and the sum of the dimensions of the submatrices with a given  $\lambda$  is the multiplicity of  $\lambda$  as a root of this equation. The matrix  $A$  is said to have simple elementary divisors corresponding to a given root  $\lambda$  of the characteristic equation if the submatrices in (1.5.7) with that  $\lambda$  are all of dimension 1. Thus orthogonal matrices, symmetric matrices, and skew symmetric matrices have only simple elementary divisors, since they can be put in diagonal form, (with  $\lambda$ 's of modulus 1, real, pure imaginary, respectively). The transformation  $B$  and the  $\lambda$ 's may not be real. If  $A$  is real, however, there is a real matrix  $B$  such that the elements of  $B^{-1}AB$  vanish except for square submatrices down the main diagonal, and the characteristic roots of different submatrices are neither equal nor conjugate imaginary.

The powers of a matrix in Jordan canonical form are easily calculated using the fact that

$$(1.5.9) \quad \begin{pmatrix} \lambda & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & \lambda & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & \lambda \end{pmatrix}^n = \begin{pmatrix} \lambda^n & 0 & \cdot & \cdot & \cdot & 0 \\ n\lambda^{n-1} & \lambda^n & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & n\lambda^{n-1} & \lambda^n \end{pmatrix}.$$

It follows that in the general case the elements  $(A^n)_{ij}$  are linear combinations of  $\lambda_j^n, n\lambda_j^{n-1}$ , etc., where  $\lambda_1, \lambda_2, \dots$  are the characteristic values of  $A$ . Hence if  $(A^n)_{ij} \rightarrow 0$  as  $n \rightarrow \infty$ , the approach must be exponential. The terms of  $A^n$  certainly go to 0 if all the characteristic values of  $A$  have modulus less than 1.

The matrix  $e^A$  is defined by the usual series formula for the exponential function. If  $A$  has the form (1.5.8),  $e^{tA}$  can be calculated using (1.5.9):

$$(1.5.10) \quad e^{tA} = \begin{pmatrix} e^{t\lambda} & 0 & \cdot & \cdot & \cdot & 0 \\ te^{t\lambda} & e^{t\lambda} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & te^{t\lambda} & e^{t\lambda} \end{pmatrix}.$$

It follows that in the general case the elements  $(e^{tA})_{ij}$  are linear combinations of  $e^{t\lambda_j}, te^{t\lambda_j}$ , etc. where  $\lambda_1, \lambda_2, \dots$  are the characteristic values of  $A$ . If  $(e^{tA})_{ij} \rightarrow 0$  as  $t \rightarrow \infty$  the approach must be exponential. The terms of  $e^{tA}$  certainly go to 0 if all the characteristic values of  $A$  have negative real parts.

A t.h.G.M. process will be called deterministic if the least squares prediction of  $x(s+t)$  for given  $x(s)$ , ( $t > 0$ ), that is  $E\{x(s); x(s+t)\}$  is always correct:

$$(1.6.1) \quad x(s+t) = A(t)x(s) \quad t > 0,$$

with probability 1.

The following classification of deterministic processes will be useful later. It will be shown that any t.h.G.M. process is the direct product of processes of the following four deterministic types, and of a factor process with no deterministic factors.

$M(0)$ . Let  $\{x(t)\}$  be the variables of a one-dimensional t.h.G.M. process, with  $x(t) = 0$  with probability 1. (The chance variable which is 0 with probability 1 is considered as a Gaussian variable with mean 0 and dispersion 0.) The correlation function of the process vanishes identically.

$M(1)$ . A one-dimensional t.h.G.M. process which satisfies

$$(1.6.2) \quad x(t) = x(0), \quad E\{x(t)\} = 0, \quad E\{x(t)^2\} > 0,$$

will be called a process of type  $M(1)$ . The correlation function  $R(t)$  is positive and independent of  $t$ .

$M(-1)$ . A one-dimensional t.h.G.M. process with an integral-valued parameter  $n$ , satisfying

$$(1.6.3) \quad \dots = x(-1) = -x(0) = x(1) = \dots \quad E\{x(n)\} = 0, \quad E\{x(n)^2\} > 0$$

will be called a process of type  $M(-1)$ . The correlation function is alternately positive and negative:  $R(n) = (-1)^n R(0)$ .

$M(e^{i\theta})$ . A two-dimensional t.h.G.M. process with

$$(1.6.4) \quad \begin{aligned} E\{x_j(0)\} &= 0, & E\{x_j(0)^2\} &= \sigma^2 > 0, & E\{x_1(0)x_2(0)\} &= 0, \\ x_1(t) &= x_1(0) \cos t\theta - x_2(0) \sin t\theta \\ x_2(t) &= x_1(0) \sin t\theta + x_2(0) \cos t\theta, \end{aligned}$$

will be called a process of type  $M(e^{i\theta})$ . The correlation function is given by

$$(1.6.5) \quad R(t) = \begin{pmatrix} \sigma^2 \cos t\theta & \sigma^2 \sin t\theta \\ -\sigma^2 \sin t\theta & \sigma^2 \cos t\theta \end{pmatrix}.$$

A process with variables  $\{x(t)\}$  will be called degenerate if there are constants  $c_1, \dots, c_N$  not all 0, such that

$$(1.7.1) \quad \sum_j c_j x_j(t) = 0$$

with probability 1, for all  $t$ . Equation (1.7.1) is true if and only if

$$(1.7.2) \quad E\left\{\left[\sum_j c_j x_j(t)\right]^2\right\} = \sum_{j,k} (R(0))_{jk} c_j c_k = 0$$

that is if and only if the correlation matrix  $R(0)$  is singular. If a non-degenerate process is a direct product of factor processes, the latter are also non-degenerate. The only degenerate one-dimensional process is that of type  $M(0)$ .

**2. The structure of degenerate and deterministic processes**

**THEOREM 2.1.** *Every degenerate t.h.G.M. process is the direct product of processes of type  $M(0)$  and (in some cases) of a non-degenerate t.h.G.M. process.*

In proving this theorem, it can be supposed that the original process has been replaced by an equivalent process, if necessary, so that the symmetric non-negative definite matrix  $R(0)$  is in diagonal form, with only 0's and 1's down the main diagonal, say 0 to the  $\nu$ th place and 1 thereafter. Then  $x_j(t) = 0$ , when  $j \leq \nu$  and the process is obviously the direct product of  $\nu$  processes of type  $M(0)$  and an  $(N - \nu)$ -dimensional non-degenerate process.

**THEOREM 2.2.** *Let  $\{x(t)\}$  be the variables of a deterministic t.h.G.M. process, with correlation function  $R(t)$ .*

(i) *The process is the direct product of factor processes of types  $M(0)$ ,  $M(\pm 1)$ ,  $M(e^{i\theta})$ .*

(ii) *If the parameter  $t$  of the process is restricted to the integers, there is a non-singular matrix  $A$  such that*

$$(2.2.1) \quad x(n) = A^n x(0),$$

$$(2.2.2) \quad R(n) = R(0)A^{*n}, \quad n = 0, \pm 1, \pm 2, \dots$$

$$(2.2.3) \quad R(0) = AR(0)A^*.$$

*The transition matrix  $A$  is the transform  $BOB^{-1}$  of an orthogonal matrix  $O$ . If the process is non-degenerate,  $A$  is uniquely determined.*

(ii') *If the parameter of the process runs through all real numbers, there is a matrix  $Q$  such that*

$$(2.2.1') \quad x(t) = e^{tQ}x(0), \quad -\infty < t < \infty,$$

$$(2.2.2') \quad R(t) = R(0)e^{tQ^*},$$

$$(2.2.3') \quad QR(0) + R(0)Q^* = 0.$$

*The matrix  $Q$  is the transform  $BKB^{-1}$  of a skew symmetric matrix  $K$ . If the process is non-degenerate,  $Q$  is uniquely determined.*

(iii) *Conversely if  $R(0)$  is any symmetric non-negative definite matrix, and if  $A(Q)$  is any matrix satisfying (2.2.3) ((2.2.3')), where  $A$  is non-singular, there is a deterministic t.h.G.M. process with correlation function given by (2.2.2) ((2.2.2')) and satisfying (2.2.1) ((2.2.1')).*

In proving (i) (ii) and (ii') it will be permissible to go to processes equivalent to the original one, if convenient. Moreover if the given process can be expressed as a direct product, it will be sufficient to prove (i) (ii) and (ii') for each factor. Since (i) (ii) and (ii') are certainly true for processes of type  $M(0)$  (with  $A$  in (ii) the identity, and  $Q$  in (ii') the null matrix), and since according to Theorem 2.1, processes of type  $M(0)$  can be factored out of the given process to leave a non-degenerate remaining factor, if any, it will be sufficient to prove (i) (ii) and (ii') for non-degenerate processes.

*Proof (t integral) of (i) and (ii) for non-degenerate processes.* If the process determined by  $\{x(n)\}$  is deterministic, (1.6.1) is true. Hence

$$(2.2.4) \quad x(\nu + 1) = Ax(\nu).$$

Then (2.2.1) is true for  $n \geq 0$ , and will be established for all  $n$  as soon as it is shown that  $A$  is non-singular. Using (2.2.1),

$$(2.2.5) \quad R(n) = E\{x(0) \cdot x(n)\} = E\{x(0) \cdot A^n x(0)\} = R(0)A^{*n}, \quad n \geq 0,$$

and

$$(2.2.6) \quad R(0) = E\{x(1) \cdot x(1)\} = E\{Ax(0) \cdot Ax(0)\} = AR(0)A^*.$$

Under the present hypotheses,  $R(0)$  is non-singular. Then  $A$  is determined uniquely by (2.2.5) with  $n = 1$ , and  $A$  cannot be singular because of (2.2.6). There is an equivalent process in which  $R(0)$  is the identity. Considering this process, (2.2.6) becomes  $I = AA^*$ , so that  $A$  is orthogonal. Finally there is an equivalent process (obtained by an orthogonal change of variables) in which  $R(0)$  is still the identity and the matrix  $A$  is in the (real) normal form of orthogonal matrices: all the elements of  $A$  are 0 except for two-dimensional rotation matrices or 1's or  $-1$ 's down the main diagonal. It is now obvious that the process is the direct product of processes of types  $M(\pm 1)$ ,  $M(e^{i\theta})$ .

*Proof (t continuously varying) of (i) and (ii') for non-degenerate processes.* If the t.h.G.M. process determined by  $\{x(t)\}$  is deterministic, (1.6.1) is true. Hence

$$(2.2.5') \quad R(t) = E\{x(s) \cdot x(s + t)\} = R(0)A(t)^*$$

$$(2.2.6') \quad R(0) = E\{x(s + t) \cdot x(s + t)\} = A(t)R(0)A(t)^*.$$

The matrix  $A(t)$  is uniquely determined by (2.2.5') since  $R(0)$  is non-singular. It then follows from (1.6.1) that

$$(2.2.7) \quad A(s + t) = A(s)A(t).$$

The continuity hypothesis (1.3.1) becomes

$$(2.2.8) \quad \lim_{t \rightarrow 0} R(0)A(t)^* = R(0),$$

which implies that

$$(2.2.9) \quad \lim_{t \rightarrow 0} A(t) = I.$$

It is well known that any solution of (2.2.7) and (2.2.9) can be written in the form  $A(t) = e^{tQ}$ . If now the right side of (2.2.6') is expanded in powers of  $t$  and the coefficient of  $t$  is set equal to 0, the resulting equation is (2.2.3'). It can be supposed, going to an equivalent process if necessary, that  $R(0)$  is the identity. Then  $A(t)A(t)^* = I$ ,  $Q + Q^* = 0$ . An equivalent process can be chosen so

that  $R(0)$  is still the identity, and so that  $Q$  is in the real canonical form of skew symmetric matrices: its elements vanish except for possible two rowed matrices

$$\begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix}$$

down the main diagonal. It is now clear that the non-degenerate process is a direct product of factors of type  $M(e^{i\theta})$  corresponding to two rowed matrices just described, and factors of type  $M(1)$ .

*Proof of (iii).* If  $R(0)$  and  $A(Q)$  satisfy the conditions of Theorem 1.2 (iii), choose  $x(0)$  as any Gaussian variable with correlation matrix  $R(0)$ . Then define  $x(n)$  by (2.2.1) ((2.2.1')). The resulting stochastic process is temporally homogeneous if and only if  $E\{x(s) \cdot x(s + t)\}$  depends only on  $t$ . The details of the calculation will be carried out for only for  $t$  integral.

In the first place

$$(2.2.10) \quad E\{x(n) \cdot x(n + \nu)\} = E\{A^n x(0) \cdot A^{n+\nu} x(0)\} = A^n R(0) A^{*n+\nu}.$$

Now (2.2.3) can be developed further:

$$(2.2.11) \quad R(0) = AR(0)A^* = A^2R(0)A^{*2} = \dots$$

so that (2.2.10) reduces to

$$(2.2.12) \quad E\{x(n) \cdot x(n + \nu)\} = R(0)A^{*\nu}.$$

The process is thus temporarily homogeneous, and obviously satisfies the other parts of the definition of a deterministic t.h.G.M. process. Theorem 2.2 is now completely proved.

The restriction imposed on  $R(0)$ ,  $A(Q)$  by (2.2.3) ((2.2.3')) is quite loose. Given  $R(0)$ , there is always an  $A(Q)$  satisfying (2.2.3) ((2.2.3')) for example the identity (null matrix). Given an  $A$  which is the transform of an orthogonal matrix (a  $Q$  which is the transform of a skew symmetric matrix) there is always a corresponding  $R(0)$ : In fact  $A(Q)$  can be assumed to be orthogonal (skew symmetric) and the  $R(0)$  can be taken as the identity.

### 3. T.H.G.M. processes with an integral valued parameter

In this section, the parameter  $t$  will range through all the integers. The condition (1.5.3) that a t.h.G. process be a t.h.G.M. process can be simplified in the integral parameter case. In fact it will be shown that it is sufficient if

$$(3.1.1) \quad E\{\dots, x(n - 1), x(n); x(n + 1)\} = E\{x(n); x(n + 1)\},$$

$n = 0, \pm 1, \dots$

with probability 1. If (3.1.1) is true,

$$(3.1.2) \quad x(\nu + 1) = Ax(\nu) + \eta(\nu)$$

where  $A$  is the transition matrix of the process and  $\eta(\nu)$  has mean 0 and is independent of  $\dots, x(\nu - 1), x(\nu)$ . It follows that

$$(3.1.3) \quad x(n) = A^{n-m}x(m) + A^{n-m-1}\eta(m) + A^{n-m-2}\eta(m+1) \\ + \dots + \eta(n-1).$$

The terms involving the  $\eta(j)$  are all independent of  $\dots, x(m-1), x(m)$ . This equation therefore implies that

$$(3.1.4) \quad E\{\dots, x(m-1), x(m); x(n)\} = A^{n-m}x(m) = E\{x(m); x(n)\},$$

and (3.1.4) is precisely the condition that the process has the Markoff property.

The following lemma will be useful.

**LEMMA 3.2.** *Let  $x: (x_1, \dots, x_N)$  be any Gaussian chance variable, with  $E\{x\} = 0$ . Then there is a uniquely determined symmetric non-negative definite matrix  $S$ , and a Gaussian chance variable  $y$ , such that*

$$(3.2.1) \quad E\{y \cdot y\} = I$$

and

$$(3.2.2) \quad x = Sy, \quad S^2 = E\{x \cdot x\}.$$

If  $x = Sy$ , and if  $S$  is symmetric, then the second equation in (3.2.2) is certainly true. It is easily seen, by examination of the characteristic values and vectors of the matrix  $E\{x \cdot x\}$  that this matrix has a unique symmetric non-negative definite square root  $S$ . Hence if there is an  $S$  satisfying (3.2.2), there can be only one. The chance variables  $x_1, \dots, x_N$  can be written as linear combinations of  $N$  uncorrelated Gaussian chance variables  $\xi_1, \dots, \xi_N$  satisfying  $E\{\xi \cdot \xi\} = I$ :

$$(3.2.3) \quad x = A\xi.$$

If  $A$  is written in the polar form  $A = SU$ , where  $S$  is symmetric and non-negative definite and  $U$  is orthogonal, (3.2.3) becomes

$$(3.2.4) \quad x = SU\xi = Sy$$

where  $y = U\xi$  satisfies (3.2.1).

It will be shown below that every t.h.G.M. process can be represented as the direct product of factors of certain types. The deterministic types have already been catalogued:  $M(0)$ ,  $M(\pm 1)$ ,  $M(e^{i\theta})$ . The non-deterministic factor type (integral valued parameter) will now be described.

*M.* Let  $\{\eta(n)\}$  be a sequence of mutually independent  $N$ -dimensional Gaussian chance variables with 0 means and a common distribution function. Let  $A$  be any  $N$ -dimensional square matrix. Define  $x(n)$  by

$$(3.3.1) \quad x(n) = \sum_{m=0}^{\infty} A^m \eta(n-m)$$



where it is supposed that<sup>11</sup>  $A$  is so chosen that the series converges with probability 1. This will be true, for example, if all the characteristic values of  $A$  have modulus less than 1, so that the terms of  $A^m$  go exponentially to 0.<sup>12</sup> It will be shown below that it is no restriction to assume that  $A$  has this character. The variables  $\{x(n)\}$  determine a t.h.G.M. process. Since  $x(n)-Ax(n-1)$  is independent of  $\dots, x(n-2), x(n-1)$ , the  $x(n)$  process is a Markoff process with transition matrix  $A$ :

$$(3.3.2) \quad E\{\dots, x(n-1); x(n)\} = Ax(n-1).$$

A process defined in this way will be called a process of type  $M$ . A non-singular change of variables  $y(n) = Bx(n)$  leads to a process of the same type:

$$(3.3.3) \quad y(n) = \sum_{m=0}^{\infty} (BAB^{-1})^m B\eta(n-m).$$

It will sometimes be convenient to write a process of type  $M$  in a form slightly different from (3.3.1). Using Lemma 3.2 it is evident that there are Gaussian variables  $\{\xi(n)\}$  satisfying

$$(3.3.4) \quad E\{\xi(n)\} = 0, \quad E\{\xi(m) \cdot \xi(n)\} = \delta_{m,n}I, \quad m, n = 0, \pm 1, \dots,$$

and a symmetric non-negative definite matrix  $S$  such that  $\eta(n) = S\xi(n)$ . Then  $S^2 = E\{\xi(n) \cdot \xi(n)\}$  and

$$(3.3.5) \quad x(n) = \sum_{m=0}^{\infty} A^m S\xi(n-m).$$

Under the change of variable  $y(n) = Bx(n)$ ,  $A$  becomes  $BAB^{-1}$  and  $S^2$  becomes  $BS^2B^*$ .

The only condition on  $A$  required for convergence in (3.3.5) is that  $A^m S \rightarrow 0$ . It will now be shown that  $A$  can always be assumed to have only characteristic values of modulus less than 1, in the sense that there is an  $A$  with this property, and satisfying the equations

$$(3.3.6) \quad A^m S = \tilde{A}^m S, \quad m = 1, 2, \dots$$

It is no restriction, going to an equivalent process if necessary, to assume that the elements of  $A$  vanish except for those in two square submatrices down the main diagonal, where one submatrix  $A_1$  has only characteristic values of modulus less than 1 and the other,  $A_2$ , of modulus greater than or equal to 1. If the matrix  $S$  is written in terms of the corresponding submatrices:

$$(3.3.7) \quad A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \quad S = \begin{pmatrix} S_1 & S_3 \\ S_4 & S_2 \end{pmatrix}$$

<sup>11</sup> Throughout this paper, if  $A$  is any matrix,  $A^0$  is defined as the identity matrix  $I$ .

<sup>12</sup> We shall use repeatedly Kolmogoroff's theorem that an infinite series of mutually independent chance variables with zero means converges with probability 1 if the series of their dispersion is convergent. (Kolmogoroff only states the theorem in one dimension, but the extension to  $n$  dimension is trivial.) If a series of mutually independent Gaussian variables converges, the series of dispersions converges to the dispersion of the sum.

the condition on  $A$  implies that  $A_2^m S_2 \rightarrow 0$ . If it is shown that  $S_2 = 0$ , it will follow that  $S_3 = S_4 = 0$ , because  $S$  is symmetric and non-negative definite. The matrix  $\tilde{A}$  will then be defined by

$$(3.3.8) \quad \tilde{A} = \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix}$$

and  $A$  will satisfy (3.3.6) and will have only characteristic values of modulus less than 1. The problem has thus been reduced to the case where  $A_1$  is absent:  $A$  only has characteristic values of modulus at least 1, and it must be proved that  $A^n S \rightarrow 0$  implies that  $S = 0$ . The proof of this is immediate when  $A$  is put into its Jordan canonical form.

The symmetric non-negative definite matrix  $S$  satisfies the equations

$$(3.3.9) \quad R(0) = \sum_0^{\infty} A^m S^2 A^{*m} = S^2 + AR(0)A^*.$$

**THEOREM 3.4.** *A direct product of processes of type  $M$  is also of type  $M$ . Conversely any factor process of a process of type  $M$  is itself of type  $M$ .*

The direct part of the theorem is obvious. To prove the converse, suppose an  $N$ -dimensional process of type  $M$  has an  $l$ -dimensional factor, corresponding to the variables  $x_1(n), \dots, x_l(n)$ . It can be supposed, that all factors of type  $M(0)$  are separated out, so that there are indices  $j, k: 1 \leq j \leq l \leq k < N$  such that the  $\{x_1(n), \dots, x_j(n)\}$  and  $\{x_{k+1}(n), \dots, x_N(n)\}$  processes are non-degenerate and that the variables  $x_{j+1}(n), \dots, x_k(n)$  vanish identically. Making a change of variables, if necessary, it can be supposed that  $R(0)$  has the form

$$(3.4.1) \quad R(0): \begin{pmatrix} j & k-j & N-k \\ I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I \end{pmatrix}$$

and that  $R(n)$  has the blocks of zeros indicated in (3.4.1). Since  $R(1) = R(0)A^*$ ,  $A$  must have the form

$$(3.4.2) \quad A: \begin{pmatrix} j & k-j & N-k \\ - & - & 0 \\ 0 & - & 0 \\ 0 & - & - \end{pmatrix}.$$

Then  $A^n$  will have this same form. Finally, because of (3.3.9),  $S^2$ , and therefore  $S$  must have the form

$$(3.4.3) \quad S^2: \begin{pmatrix} - & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & - \end{pmatrix}.$$

Let  $A_0$  be the matrix whose elements are the same as those of  $A$  except that the  $(j+1)$ th to  $k$ th columns of  $A_0$  vanish. Since

$$(3.4.4) \quad A_0^m S = A^m S, \quad m = 0, 1, \dots$$

it follows that

$$(3.4.5) \quad x(n) = \sum_{m=0}^{\infty} A_0^m S \xi(n - m).$$

It is now obvious that the  $\{x_1(n), \dots, x_i(n)\}$  process is of type  $M$ .

**THEOREM 3.5.** *A non-degenerate process of type  $M$  has no deterministic factors.*

Any factor process is non-degenerate and of type  $M$ . To prove the theorem, it will therefore be sufficient to prove that the process itself cannot be deterministic. If it were, we should have

$$x(n) - Ax(n - 1) = S \xi(n - 1) = 0.$$

Then  $S = 0$ . But then the process is certainly degenerate, contrary to hypothesis.

**THEOREM 3.6.** (i) *Every t.h.G.M. process (discrete parameter) is the direct product of processes of type  $M(0)$ ,  $M(\pm 1)$ ,  $M(e^{i\theta})$ ,  $M$ .*

(ii) *Let  $A$  be a transition matrix of a t.h.G.M. process, with variables  $\{x(n)\}$ . There are mutually independent Gaussian variables  $\dots, \xi(0), \xi(1), \dots, \xi$  satisfying*

$$(3.6.1) \quad \begin{aligned} E\{\xi(n)\} &= E\{\xi\} = 0 \\ E\{\xi(n) \cdot \xi(n)\} &= E\{\xi \cdot \xi\} = I, \end{aligned}$$

and symmetric non-negative definite matrices  $S, T$  such that

$$(3.6.2) \quad x(n) = \sum_{m=0}^{\infty} A^m S \xi(n - m) + A^n T \xi, \quad n = 0, 1, \dots,$$

$$(3.6.3) \quad T^2 = AT^2A^*,$$

$$(3.6.4) \quad R(0) = \sum_{m=0}^{\infty} A^m S^2 A^{*m} + T^2 = AR(0)A^* + S^2,$$

where the series in (3.6.2) converges with probability 1. If  $A$  is non-singular, (3.6.2) holds for all  $n$ . The sum and last term in (3.6.2) are linear transformations of  $x(n)$ : (3.6.2) exhibits in part the decomposition into factor processes described in (i). The correlation function is given by

$$(3.6.5) \quad \begin{aligned} R(n) &= R(0)A^{*n} \\ R(-n) &= A^n R(0). \end{aligned} \quad n = 0, 1, \dots$$

(iii) *The transition matrix  $A$  is uniquely determined if and only if the process is non-degenerate. In any case, there is a transition matrix whose characteristic values are all of modulus less than or equal to 1, and whose characteristic values of modulus 1 correspond to simple elementary divisors. The transition matrix  $A$  furnishes the solution of the prediction problem of the process:*

$$(3.6.6) \quad E\{\dots, x(m - 1), x(m); x(m + n)\} = A^n x(m), \quad n = 0, 1, \dots$$

The matrix  $S^2$ , which is uniquely determined, measures the dispersion of  $x(n + 1)$  about its predicted value in terms of  $x(n)$ :

$$(3.6.7) \quad E\{[x(n + 1) - Ax(n)]^2\} = S^2.$$

(iv) Conversely if  $A$  is a matrix with at least one characteristic value of modulus less than 1 or of modulus 1 and corresponding to a simple elementary divisor,  $A$  is the transition matrix of a t.h.G.M. process, with  $R(0)$  not the null matrix. If all the characteristic values of  $A$  are as just described,  $A$  is the transition matrix of a non-degenerate t.h.G.M. process. If  $R(0)$ ,  $S$ ,  $A$  are matrices satisfying (3.6.4) with  $R(0)$ ,  $S$ , symmetric and non-negative definite, there is a t.h.G.M. process whose variables can be written in the form (3.6.2) with the given  $R(0)$ ,  $S$ ,  $A$ .

This decomposition of a t.h.G.M. process into deterministic factors can be considered as a special case of the general decomposition theorem of Wold, which is applicable to all t.h.G. process.<sup>13</sup> (Wold only considered the one-dimensional case.) The proof in the present special case is simpler, however, and illuminates the general case.

*Proof of (i) and (ii).* Equations (1.5.3) and (1.5.4), in the present case, lead to

$$(3.6.8) \quad E\{\dots, x(n - 1), x(n); x(n + 1)\} = E\{x(n); x(n + 1)\} = Ax(n).$$

The first two terms are equal because the process has the Markoff property. The last term is linear in  $x(n)$  because the process is Gaussian. The matrix  $A$  can be taken independent of  $n$  because the process is temporally homogeneous. Thus (3.6.8) involves the three fundamental properties of the  $x(n)$  process. From the definition of conditional expectation, it follows that  $x(n + 1) - Ax(n)$  is independent of the chance variables  $\dots, x(n - 1), x(n)$ . Hence the variables

$$\dots, [x(n) - Ax(n - 1)], [x(n + 1) - Ax(n)], \dots$$

are mutually independent. According to Lemma 3.2, there are mutually independent chance variables  $\{\xi(n)\}$  satisfying

$$(3.6.9) \quad x(n) - Ax(n - 1) = S\xi(n), \quad E\{\xi(n) \cdot \xi(n)\} = I, \quad E\{\xi(n)\} = 0,$$

where  $S$  is symmetric, non-negative definite, and satisfies (3.6.7). The matrix  $S^2$  thus measures the dispersion of  $x(n)$  about its predicted value  $Ax(n - 1)$ .

The representation (3.6.2) can be obtained very simply. In fact

$$(3.6.10) \quad \begin{aligned} x(n) &= [x(n) - Ax(n - 1)] + A[x(n - 1) - Ax(n - 2)] + \dots \\ &\quad + A^{n-\nu-1}[x(\nu + 1) - Ax(\nu)] + A^{n-\nu}x(\nu) \\ &= \sum_{j=0}^{n-\nu-1} A^j S\xi(n - j) + A^{n-\nu}x(\nu), \end{aligned}$$

<sup>13</sup> *A Study in the Analysis of Stationery Time Series*, Uppsala (1938), p. 89. See also Kolmogoroff, *Bull. Acad. Sci. URSS Ser. Math.*, Vol. 5 (1941), pp. 3-14 and *Bolletín Moskovskogo Gosudarstvennogo, Matematika*, Vol. 2 (1941), pp. 1-40, in whose papers the decomposition theorem is brought out in its full significance.

and it will be shown that when  $\nu \rightarrow -\infty$  (3.6.10) leads to (3.6.2). Before going to the limit, however, we note that in (3.6.10) the sum is independent of the variables  $\dots, x(\nu - 1), x(\nu)$ , so that

$$(3.6.11) \quad E\{\dots, x(\nu - 1), x(\nu); x(n)\} = A^{n-\nu}x(\nu)$$

which is another way of writing (3.6.6). Moreover, using (3.6.11),

$$(3.6.12) \quad R(n - \nu) = E\{x(\nu) \cdot x(n)\} = R(0)A^{*n-\nu},$$

which is another way of writing (3.6.5). (The value of  $R(n)$  for  $n < 0$  is obtained using the fact that  $R(-n) = R(n)^*$ .) The last term in (3.6.10) is the conditional expectation of  $x(n)$  for preassigned  $\dots, x(\nu - 1), x(\nu)$ . It follows from the corollary to Theorem 1.2 that this conditional expectation converges with probability 1 when  $\nu \rightarrow -\infty$ , but this convergence will be proved directly in the present particular case.

From (3.6.10),

$$(3.6.13) \quad E\{x(n) \cdot x(n)\} = R(0) = \sum_{j=0}^{n-\nu-1} A^j S^2 A^{*j} + A^{n-\nu} R(0) A^{*n-\nu}.$$

The terms of the sum and the last term are all symmetric and non-negative definite matrices. It follows that there is convergence in (3.6.13) when  $\nu \rightarrow -\infty$ :

$$(3.6.14) \quad R(0) = \sum_{j=0}^{\infty} A^j S^2 A^{*j} + \lim_{m \rightarrow \infty} A^m R(0) A^{*m}.$$

The convergence of the series of dispersions in (3.6.14) implies that the series of chance variables in (3.6.2) converges, with probability 1. Then when  $\nu \rightarrow -\infty$  (3.6.10) becomes

$$(3.6.15) \quad x(n) = \sum_{j=0}^{\infty} A^j S \xi(n - j) + z(n),$$

where

$$(3.6.16) \quad z(n) = \lim_{\nu \rightarrow -\infty} A^{n-\nu} x(\nu).$$

Since  $x(n)$  is independent of  $\xi(n + 1), \xi(n + 2), \dots, z(n)$  is independent of every  $\xi(m)$ . Moreover, writing  $z(0) = T\xi$ , where  $\xi$  satisfies (3.6.1) and  $T$  is symmetric and non-negative definite,

$$(3.6.17) \quad z(n) = A^n z(0) = A^n T \xi, \quad n \geq 0.$$

Thus (3.6.3) and (3.6.4) are satisfied. If  $A$  is non-singular, (3.6.17) will be correct for negative  $n$  also.

The decomposition of the process into factor processes of the types described in the theorem will be obtained by a detailed analysis of the significance of (3.6.2). Under the change of variable  $y(n) = Bx(n)$ ,  $T^2$  becomes  $BT^2B^*$ , and

$A$  becomes  $BAB^{-1}$ . Making a suitable change of variables, if necessary, it can be supposed that  $A$  has the form

$$(3.6.18) \quad A: \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$$

where the characteristic values of  $A_1$  have moduli unequal to 1 and those of  $A_2$  have modulus 1. The matrix  $T^2$  can be written in terms of submatrices of the same dimensions in a corresponding way:

$$(3.6.19) \quad T^2: \begin{pmatrix} T_1^2 & - \\ - & T_2^2 \end{pmatrix}$$

where  $T_1, T_2$  are symmetric and non-negative definite. A further change of variables may be made, if necessary (transforming only the last  $n$  variables) preserving the forms (3.6.18) and (3.6.19) and transforming  $T_2$  into the identity. Then using (3.6.3)

$$(3.6.20) \quad A_1 T_1^2 A_1^* = T_1^2, \quad A_2 A_2^* = I.$$

Hence  $A_2$  is orthogonal. Developing (3.6.20) further,  $A_1^m T_1^2 A_1^{*m} = T_1^2$ , for all  $m \geq 0$ . When  $m \rightarrow \infty$  in this equation, the terms in the matrix product on the left involve the  $m$ th power of the characteristic values of  $A_1$  (all of modulus different from 1, by hypothesis). Then those characteristic values which actually appear can only be those of modulus less than 1, and the matrix on the left must go to 0 as  $m \rightarrow \infty$ :  $T_1 = 0$ . Since  $T$  is non-negative definite,  $T$  must have the simple form

$$(3.6.21) \quad T: \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}.$$

The matrix  $S$  can also be divided into corresponding submatrices:

$$(3.6.22) \quad S: \begin{pmatrix} S_1 & - \\ - & S_2 \end{pmatrix}.$$

The convergence of the series in (3.6.4) implies that

$$\lim_{m \rightarrow \infty} A_2^m S_2^2 A_2^{*m} = 0.$$

Since  $A_2$  is orthogonal, this means that  $S_2 = 0$ , and since  $S$  is symmetric and non-negative definite,  $S$  has the form

$$(3.6.23) \quad S: \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix}.$$

It is now clear from (3.6.2) that the  $x(n)$  process is the direct product of a process of type  $M$  and a deterministic process corresponding to the division of the variables determining the above submatrices. The deterministic factor process is the direct product of the elementary types already discussed. The variable  $z(n)$  and the sum in (3.6.2) are linear transformations of  $x(n)$ .

*Proof of (iii).* If the process is non-degenerate,  $R(0)$  is non-singular, and the transition matrix is determined uniquely by (3.6.5) with  $n = 1$ . If the process is degenerate, there will be one or more factor processes of type  $M(0)$ , and their transition matrices are quite unrestricted. In the non-degenerate case the characteristic values will be of modulus less than 1 (corresponding to a factor of type  $M$ , if one is present), or equal to 1 (corresponding to the factors of type  $M(\pm 1)$ ,  $M(e^{i\theta})$  making up the deterministic factor, if one is present), and in the latter case the elementary divisors are simple. If the process is degenerate, and if the part of  $A$  corresponding to the factors of type  $M(0)$  is taken to be the identity, there will be simple elementary divisors corresponding to the characteristic value 1 for each such factor. The remaining statements of (iii) have already been proved.

*Proof of (iv).* Let  $A$  be a matrix with at least one characteristic value of modulus less than 1 or equal to 1 and corresponding to a simple elementary divisor. Then some transform  $BAB^{-1}$  has the form

$$(3.6.24) \quad \begin{pmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{pmatrix}$$

where  $A_1$  (if present) has only characteristic values of modulus less than 1,  $A_2$  (if present) is orthogonal, and both  $A_1, A_2$  are not absent. For the purposes of the present proof it can be supposed that  $A$  is already in this form. Define  $S, T$  by

$$(3.6.25) \quad S: \begin{pmatrix} S_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad T: \begin{pmatrix} 0 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where the indicated submatrices of  $S$  and  $T$  are in the same positions as those of  $A$ , and where  $S_1$  is any symmetric positive definite (and therefore non-singular) matrix of the proper dimension. The series in (3.6.4) converges and the first equation in (3.6.4) defines a matrix  $R(0)$  which obviously satisfies the continued equality. If all the characteristic values of  $A$  are as described in the beginning of this paragraph,  $A_3$  can be supposed absent. In this case

$$R(0) = S^2 + T^2 + \dots$$

is non-singular. The proof of the first two parts of (iv) has now been reduced to that of the last part. Suppose then that  $R(0), A, S$  satisfy the hypotheses of the last part of (iv). Then

$$(3.6.26) \quad \begin{aligned} R(0) &= AR(0)A^* + S^2 \\ AR(0)A^* &= A^2R(0)A^{*2} + AS^2A^* \\ &\dots\dots\dots \\ A^{n-1}R(0)A^{*(n-1)} &= A^nR(0)A^{*n} + A^{n-1}S^2A^{*(n-1)}. \end{aligned}$$

Adding these equations

$$(3.6.27) \quad R(0) = \sum_{m=0}^{n-1} A^m S^2 A^{*m} + A^n R(0) A^{*n}.$$

This equation leads to (3.6.14), and  $T^2$ , defined as the limit in (3.6.14), satisfies (3.6.3). Let  $\dots, \xi(-1), \xi(0), \dots, \xi$  be mutually independent Gaussian variables satisfying (3.6.1). Then the  $x(n)$  defined by (3.6.2) determine the variables of a Gaussian process with non-negative values of  $n$ , but a slight modification is needed to obtain an expression defined for all  $n$ . To obtain this, it can be supposed that  $A, T, S$  are in the forms (3.6.24), (3.6.25). Define  $\tilde{A}$  by

$$(3.6.28) \quad \tilde{A}: \begin{pmatrix} I & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & I \end{pmatrix}.$$

Then  $A$  is orthogonal and  $\tilde{A}T = AT$ . If now (3.6.2) is used to define  $x(n)$  for all  $n$  with  $A^n T$  replaced by  $\tilde{A}^n T$ , the  $x(n)$  process is a t.h.G.M. process with the desired properties.

The properties of the process reversed in time are of some interest. It is easy to see that if  $n$  is replaced by  $-n$ , a t.h.G.M. process remains a t.h.G.M. process. If the original process is non-degenerate, the new transition matrix is  $R(0)A^*R(0)^{-1}$ . If the transition matrix remains unchanged when  $n$  is replaced by  $-n$ ,  $R(0)A^*R(0)^{-1} = A$ . This is equivalent to the equation  $R(n) = R(-n)$ .

The simplest generalization of a t.h.G.M. process is the following. Let the chance variables  $\{y(n)\}$  determine a t.h.G. process with the property that for some  $N > 0$ ,

$$(3.7.1) \quad E\{\dots, y(n-1); y(n)\} = E\{y(n-N), \dots, y(n-1); y(n)\},$$

with probability 1. If  $N = 1$ , the process is a t.h.G.M. process. This type process will be called a t.h.G.M. $_N$  process. To avoid notational complications only the one-dimensional case will be considered. The right hand side of (3.7.1) is a linear combination of the variables  $y(n-N), \dots, y(n-1)$ . The variables thus satisfy a difference equation of the form

$$(3.7.2) \quad y(n) - a_1 y(n-1) - \dots - a_N y(n-N) = \eta(n)$$

generalizing (3.6.9), where  $\eta(n)$  is independent of the chance variables  $\dots, y(n-2), y(n-1)$ . The  $\{\eta(n)\}$  are mutually independent chance variables with zero means and dispersions independent of  $n$ . Equation (3.7.2) leads to

$$(3.7.3) \quad y(n) - a_1^{(n-m)} y(m-1) - \dots - a_N^{(n-m)} y(m-N) = \eta^{(n-m)}(n) \tag{m \le n}$$

where  $\eta^{(n-m)}(n)$  has zero mean and is independent of the chance variables  $\dots, y(m-2), y(m-1)$ . Hence

$$(3.7.4) \quad E\{\dots, y(m-1); y(n)\} = E\{y(m-N), \dots, y(m-1); y(n)\}, \tag{m \le n}$$



The difference equation (3.7.2) has been studied in some detail in the past.<sup>14</sup> We shall use an approach which adds insight into the structure of the solution and which clarifies the place of the solution in the general theory of t.h.G. processes. This approach is in terms of  $N$ -dimensional t.h.G.M. processes. Define the variables  $\{x(n)\}$  of an  $N$ -dimensional process by

$$(3.7.5) \quad x_j(n) = y(n + j), \quad n = 0, \pm 1, \dots, \quad j = 1, \dots, N.$$

The  $x(n)$  process is evidently a t.h.G.M. process. If the index  $N$  of the  $y(n)$  process is the minimum for which (3.7.1) is true, the corresponding  $x(n)$  process will be non-degenerate. Then the transition matrix  $A$  is uniquely determined, and is evidently

$$(3.7.6) \quad A: \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \\ a_N & \cdot & \cdot & \cdot & \cdot & \cdot & a_1 \end{pmatrix} \quad a_1 \neq 0.$$

The matrix  $S$ , measuring the dispersion of the prediction  $Ax(n - 1)$  of  $x(n)$ , has the form

$$(3.7.7) \quad S: \begin{pmatrix} 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & s \end{pmatrix}.$$

The characteristic equation of  $A$  is simply

$$(3.7.8) \quad \alpha^N - a_1\alpha^{N-1} - \dots - a_N = 0.$$

The matrix  $A$  has only a single characteristic vector corresponding to each characteristic value  $\lambda$ , the vector  $(1, \lambda, \dots, \lambda^{N-1})$ . Hence if  $\lambda$  is a multiple root of (3.7.8), it does not correspond to a simple elementary divisor. Therefore, according to Theorem 3.6, all roots of (3.7.8) of modulus 1 must be simple roots. It will be proved below that either no roots have modulus 1 or all roots have modulus 1.

If an  $N$ -dimensional non-degenerate t.h.G.M. process is given whose transition matrix  $A$  and dispersion matrix  $S$  have the forms (3.7.6) and (3.7.7) respectively,

$$x_j(n) - x_{j+1}(n - 1) = 0$$

with probability 1, for  $j < N$ . Then a  $y(n)$  process can be defined unambiguously by (3.7.5). Since for fixed  $j$ ,  $x_j(n)$  determines a one-dimensional t.h.G. process, the  $y(n)$  process is a t.h.G. process, and (3.7.1) is obviously true, with  $N$  minimal if  $A$  is non-singular.

*Case 1.  $S = 0$  (deterministic case).* In this case the  $x(n)$  process is deterministic, and the  $y(n)$  process satisfies the equation

$$(3.7.2') \quad y(n) = a_1y(n - 1) + \dots + a_Ny(n - N).$$

<sup>14</sup> Cf. for example H. Wold, *A Study in the Analysis Of Stationery Time Series*, Uppsala, 1938.

All the roots of (3.7.8) are simple roots, of modulus 1. Since  $S = 0$ ,

$$(3.7.9) \quad x(n) = A^n T \xi \quad n = 0, \pm 1, \dots$$

and therefore

$$(3.7.10) \quad y(n) = x_1(n - 1) = \sum_{j=1}^N (A^{n-1} T)_{1j} \xi_j \quad n = 0, \pm 1 \dots$$

Using either the well known form of the solution of the  $N$ th order difference equation (3.7.2') or of the powers of an orthogonal matrix, it follows that

$$(3.7.11) \quad y(n) = \sum_{j=1}^N (\eta_j \cos n\theta_j + \zeta_j \sin n\theta_j)$$

where the  $\eta_j$  and  $\zeta_j$  are (one-dimensional) Gaussian variables, and

$$\{\cos \theta_j + i \sin \theta_j\}$$

are the  $N$  distinct characteristic values of  $A$ , that is the roots of (3.7.8).

*Case 2.  $S \neq 0$  (non-deterministic case).* In this case it will now be shown that the  $x(n)$  process can have no deterministic factors: that is that the roots of (3.7.8) all have modulus less than 1. In fact let  $\beta$  be a root of (3.7.8), corresponding to the characteristic vector  $z$  of  $A^*$ :

$$(3.7.12) \quad z = (a_N \beta^{N-1}, a_N \beta^{N-2} + a_{N-1} \beta^{N-1}, \dots, a_N + a_{N-1} \beta + \dots + a_1 \beta^{N-1}) \\ = (a_N \beta^{N-1}, \dots, \beta^N).$$

Then using (3.6.4),

$$(3.7.13) \quad (R(0)z, z) = (AR(0)A^*z, z) + (S^2z, z) \\ = (R(0)A^*z, A^*z) + (Sz, Sz) \\ = |\beta|^2 (R(0)z, z) + s^2 |\beta|^{2N}.$$

Hence  $|\beta|$  cannot be 1, and the  $x(n)$  process can have no deterministic factors. Equation (3.6.2) becomes

$$(3.7.14) \quad x(n) = \sum_{m=0}^{\infty} A^m S \xi(n - m)$$

which leads to

$$(3.7.15) \quad y(n) = \sum_{j=1}^N \sum_{m=0}^{\infty} (A^m S)_{1j} \xi_j(n - m - 1) = s \sum_{m=0}^{\infty} (A^m)_{1N} \xi_N(n - m - 1).$$

According to Theorem 3.6 the only restriction on the coefficients  $a_1, \dots, a_N$  in the two cases  $S = 0, S \neq 0$ , are respectively that equation (3.7.8) has  $N$  distinct roots of modulus 1 and all roots of modulus less than 1. Hence (3.7.10) and (3.7.15) furnish (with the stated restrictions on  $A$ ) the most general t.h.G.M. <sub>$N$</sub>  processes.

It was shown in Theorem 3.6 that if  $R(n)$  is the correlation function of a t.h.G.M. process,  $R(n)$  can be expressed in the form (3.6.5), where  $A$  is some suitably chosen matrix. Conversely if the correlation function of a t.h.G. process has the form (3.6.5), the process is a t.h.G.M. process since  $x(n + 1) - Ax(n)$  is then orthogonal to (and therefore independent of) the variables  $\dots, x(n - 1), x(n)$ . (This fact implies the truth of (3.1.1)). The characterization of t.h.G.M. processes in terms of their correlation functions is thus easily solved. The following theorems characterize one-dimensional t.h.G.M.<sub>N</sub> processes from various points of view. It will be convenient, and also intrinsically interesting to treat at the same time a slightly larger class of processes: the class of *component processes of t.h.G.M. processes*. A one-dimensional t.h.G. process with variables  $\{x_1(n)\}$  will be called a component process of an  $N$ -dimensional t.h.G.M. process if there are  $N - 1$  t.h.G. processes with variables  $\{x_2(n)\}, \dots, \{x_N(n)\}$  such that the  $N$ -dimensional process with variables  $\{x_1(n), \dots, x_N(n)\}$  is a t.h.G.M. process. If the variables  $\{x(n)\}$  determine an  $N$ -dimensional t.h.G.M. process, the t.h.G. processes determined by  $\{x_1(n)\}, \dots, \{x_N(n)\}$  will be called its  $N$  component processes. If an  $x(n)$  process is not of type  $M(0)$  and is a component process of an  $N$ -dimensional t.h.G.M. process, it is a component process of a non-degenerate  $N_1$ -dimensional t.h.G.M. process, for some  $N_1 \leq N$ . It has already been seen that one-dimensional t.h.G.M.<sub>N</sub> processes are component processes of  $N$ -dimensional t.h.G.M. processes.

**THEOREM 3.8.** *Let  $\dots, x(0), x(1), \dots$  be the variables of a one-dimensional t.h.G. process. The process is a component process of an  $N$ -dimensional t.h.G.M. process if and only if the chance variables*

$$(3.8.1) \quad x(0), E\{\dots, x(-1), x(0); x(n)\}, \quad n = 1, 2, \dots$$

*are linearly dependent on the first  $N$ .*

Suppose that the  $x(n)$  process is a component process of an  $N$ -dimensional  $y(n)$  process:  $x(n) = y_1(n)$ , with correlation function  $R_y(n)$  and transition matrix  $A$ . Since  $A$  satisfies its characteristic equation

$$(3.8.2) \quad \det |\alpha I - A| = \alpha^N - a_1\alpha^{N-1} - \dots - a_N = 0,$$

it follows from (3.6.2) that if  $\eta(n + N)$  is defined by

$$(3.8.3) \quad y(n + N) - a_1y(n + N - 1) - \dots - a_Ny(n) = \eta(n + N)$$

then  $\eta(n + N)$  is independent of  $\dots, y(n - 1), y(n)$ . Then

$$(3.8.4) \quad x(n + N) - a_1x(n + N - 1) - \dots - a_Nx(n) = \eta_1(n + N)$$

where  $\eta_1(n + N)$  is independent of the chance variables  $\dots, x(n - 1), x(n)$ . Equation (3.8.4) leads to

$$(3.8.5) \quad x(n + N + \nu) - a_1^{(\nu)}x(n + N - 1) - \dots - a_N^{(\nu)}x(n) = \eta_1^{(\nu)}(n + N + \nu)$$

where  $\eta_1^{(\nu)}(n + N + \nu)$  has zero mean and is independent of the chance variables  $\dots, x(n - 1), x(n)$ . If the operator

$$E\{\dots, x(n - 1), x(n); \cdot\}$$

is applied to this equation, the result is

$$(3.8.6) \quad E\{\dots, x(n - 1), x(n); x(n + N + \nu)\} - \sum_{m=1}^N a_m^{(\nu)} E\{\dots, x(n - 1), x(n); x(n + N - m)\} = 0.$$

The last term in the sum is  $x(n)$  and (3.8.6) is thus the desired linear relation.

Conversely suppose that the  $(N + 1)$ th chance variable in (3.8.1) is linearly dependent on the first  $N$ :

$$(3.8.7) \quad E\{\dots, x(n - 1), x(n); x(n + N)\} = \sum_{m=1}^N a_m E\{\dots, x(n - 1)x(n); x(n + N - m)\}, \quad n = 0, \pm 1, \dots$$

Define the variables  $y_1(n), \dots, y_N(n)$  of an  $N$ -dimensional t.h.G. process by

$$(3.8.8) \quad \begin{aligned} y_1(n) &= x(n) \\ y_2(n) &= E\{\dots, x(n - 1), x(n); x(n + 1)\} \\ &\dots\dots\dots \\ y_N(n) &= E\{\dots, x(n - 1), x(n); x(n + N - 1)\}. \end{aligned}$$

Then

$$(3.8.9) \quad \begin{aligned} E\{\dots, y(n - 1), y(n); y_1(n + 1)\} &= y_2(n) \\ &\dots\dots\dots \\ E\{\dots, y(n - 1), y(n); y_{N-1}(n + 1)\} &= y_N(n) \\ E\{\dots, y(n - 1), y(n); y_N(n + 1)\} & \\ &= E\{\dots, x(n - 1), x(n); x(n + N)\} = \sum_{m=1}^N a_m y_{N+1-m}(n). \end{aligned}$$

The  $y(n)$  process is therefore a t.h.G.M. process, with transition matrix (3.7.6), and the  $x(n)$  process is a component process.

The following particular type of t.h.G. process will be involved in the proof of Theorem 3.9. If the chance variables  $\{\eta(n)\}$  determine a t.h.G. process whose correlation function  $R_\eta(n)$  vanishes when  $n \geq N$ , then according to (1.3.5) the complex spectral function  $G_\eta(\lambda)$  of the  $\eta(n)$  process is continuous, with derivative  $G'_\eta(\lambda)$  given by

$$(3.9.1) \quad G'_\eta(\lambda) = \frac{1}{2\pi} \sum_{n=-(N-1)}^{N-1} R_\eta(n).$$

It is easily verified (using the fact that  $R_\eta(n) = R_\eta(-n) = \overline{R_\eta(n)}$ ) that if  $\alpha$  is a root of the equation

$$(3.9.2) \quad \sum_{n=-(N-1)}^{N-1} R_\eta(n) z^n = 0,$$

then  $\bar{\alpha}$ ,  $1/\alpha$ ,  $1/\bar{\alpha}$  are also roots, of the same multiplicity. Moreover if  $|\alpha| = 1$ ,  $\alpha$  is a root of even multiplicity, since the sum in (3.9.2) is real and non-negative when  $|z| = 1$ . When  $|z| = 1$ ,

$$(3.9.3) \quad |\alpha| |(z - \alpha)(z - 1/\bar{\alpha})| = |z - \alpha|^2.$$

Hence  $G'_\eta(\lambda)$  can be written in the following simple form:

$$(3.9.4) \quad G'_\eta(\lambda) = |b_0 e^{i(N-1)\lambda} + b_1 e^{i(N-2)\lambda} + \dots + b_{N-1}|^2$$

where the roots of the indicated polynomial have modulus at most 1, and the coefficients are real.

**THEOREM 3.9.** *Let  $\dots, x(0), x(1), \dots$  be the variables of a one-dimensional t.h.G. process. The process is a component process of a finite-dimensional t.h.G.M. process if and only if the complex spectral function is the sum of the integral of the square of the absolute value of a rational function of  $e^{i\lambda}$  with real coefficients, and of a monotone non-decreasing function increasing only in a finite number of jumps.<sup>15</sup> Specifically:*

(i) *The process is a component process of an  $N$ -dimensional t.h.G.M. process if and only if the complex spectral function has the form*

$$(3.9.5) \quad G(\lambda) = \int_{-\pi}^{\lambda} \frac{|\beta_0 e^{i(N-1)\lambda} + \dots + \beta_{N-1}|^2}{|\alpha_0 e^{iN\lambda} + \dots + \alpha_N|^2} d\lambda + \hat{G}(\lambda)$$

where

- (a)  $\hat{G}(\lambda)$  is a monotone non-decreasing function satisfying (1.3.3), increasing only by jumps, at no more than  $N$  points;
- (b) the denominator of the integrand vanishes at every discontinuity of  $G(\lambda)$ , and the numerator vanishes at every zero of the denominator, to at least the same order;
- (c) the coefficients  $\alpha_0, \dots, \alpha_N, \beta_0, \dots, \beta_{N-1}$  are real,  $\alpha_0 \neq 0, \beta_0 \neq 0$  unless the integrand vanishes identically, and the roots of the polynomials in the integrand have modulus less than or equal to 1.

*The integral vanishes identically if and only if the  $x(n)$  process is a component process of an  $N$ -dimensional deterministic process, and  $\hat{G}(\lambda)$  vanishes identically if and only if the variables  $x(n)$  vanish identically or the  $x(n)$  process is a component process of an  $N$ -dimensional t.h.G.M. process with no deterministic factor.*

(ii) *The process is a t.h.G.M.<sub>N</sub> process (deterministic case) if and only if the*

<sup>15</sup> It is easily seen that the first term of the two can also be described simply as the integral of a rational function of  $e^{i\lambda}$ , which is non-negative for  $\lambda$  real and is an even function of  $\lambda$  like all complex spectral density functions,

complex spectral function  $G(\lambda) = \hat{G}(\lambda)$  is a monotone non-decreasing function satisfying (1.3.3) increasing only in jumps, at no more than  $N$  points; (non-deterministic case) if and only if the complex spectral function has the form

$$(3.9.6) \quad G(\lambda) = \int_{-\pi}^{\lambda} \frac{d\lambda}{|\alpha_0 e^{iN\lambda} + \dots + \alpha_N|^2}$$

where  $\alpha_0, \dots, \alpha_N$  are real and  $\alpha_0 \neq 0$ .

*Proof of (i).* If the  $x(n)$  process is a one-dimensional component of an  $N$ -dimensional t.h.G.M. process, it has already been seen that for properly chosen real numbers  $a_1, \dots, a_N$ , (3.8.4) is true, where  $\eta_1(n + N)$  is independent of the chance variables  $\dots, x(n - 1), x(n)$ . Equation (3.8.2) can be assumed to have all its roots of modulus less than or equal to 1. It follows from (3.8.4) that  $\eta_1(n)$  is independent of  $\eta_1(m)$  if  $|n - m| \geq N$ . The complex spectral function of the  $\eta_1(n)$  process is therefore continuous, with derivative given by (3.9.4). It will be no restriction to assume that  $b_0 \neq 0$  unless the derivative vanishes identically. According to (1.3.6), if  $G(\lambda)$  is the complex spectral function of the  $x(n)$  process,

$$(3.9.7) \quad \begin{aligned} E\{\eta_1(0)\eta_1(n)\} &= \int_{-\pi}^{\pi} e^{i\lambda n} |b_0 e^{i(N-1)\lambda} + \dots + b_{N-1}|^2 d\lambda \\ &= \int_{-\pi}^{\pi} e^{i n \lambda} |e^{iN\lambda} - a_1 e^{i(N-1)\lambda} - \dots - a_N|^2 dG(\lambda). \end{aligned}$$

Hence if  $\hat{G}(\lambda)$  is the jump function of  $G(\lambda)$  ( $\hat{G}(-\pi) = 0$ , and  $\hat{G}(\lambda)$  is constant except for jumps at the same points as those of  $G(\lambda)$ , and of the same magnitude),

$$(3.9.8) \quad \begin{aligned} \int_{-\pi}^{\lambda} |b_0 e^{i(N-1)\lambda} + \dots + b_{N-1}|^2 d\lambda \\ = \int_{-\pi}^{\lambda} |e^{iN\lambda} - \dots - a_N|^2 d[G(\lambda) - \hat{G}(\lambda)] \\ + \int_{-\pi}^{\lambda} |e^{iN\lambda} - \dots - a_N|^2 d\hat{G}(\lambda). \end{aligned}$$

Since the first two integrals are continuous in  $\lambda$ , the last must be continuous also. Hence the last integrand must vanish at every discontinuity of  $\hat{G}(\lambda)$ , that is at every discontinuity of  $G(\lambda)$ , and the last integral vanishes identically. It follows that

$$(3.9.9) \quad G(\lambda) - \hat{G}(\lambda) = \int_{-\pi}^{\lambda} \frac{|b_0 e^{i(N-1)\lambda} + \dots + b_{N-1}|^2}{|e^{iN\lambda} - \dots - a_N|^2} d\lambda$$

where the numerator vanishes at each zero of the denominator, with the same or greater multiplicity. Since the denominator vanishes at each discontinuity of  $G(\lambda)$ , there can be at most  $N$  discontinuities. If the  $N$ -dimensional process is a deterministic process, it can be assumed that all the roots of equation (3.8.2)

have modulus 1, that is that the denominator and hence also the numerator in (3.9.9) have  $N$  roots. This can be true only if the numerator vanishes identically:  $G(\lambda) \equiv \hat{G}(\lambda)$ . If the  $N$ -dimensional process has no deterministic factor, it can be assumed that all the roots of equation (3.8.2) have modulus less than 1. Then  $G(\lambda)$  can have no discontinuities:  $\hat{G}(\lambda) \equiv 0$ .

Conversely if  $G(\lambda)$  has the form described in Theorem 3.9 (i),  $G(\lambda)$  can be assumed in the form (3.9.9) with real coefficients in numerator and denominator and the stated relations between the jumps of  $G(\lambda)$  and the zeros of the numerator and denominator in the integrand. (If the integrand vanishes identically and if  $G(\lambda)$  has  $N$  discontinuities,  $a_1, \dots, a_N$  can be chosen as those numbers making the polynomial

$$e^{iN\lambda} - a_1 e^{i(N-1)\lambda} - \dots - a_N$$

vanish at the discontinuities of  $G(\lambda)$ .) Then

$$\begin{aligned} (3.9.10) \quad & R(n + N) - a_1 R(n + N - 1) - \dots - a_N R(n) \\ &= \int_{-\pi}^{\pi} e^{i\lambda(n+1)} \frac{[b_0 + \dots + b_{N-1} e^{i(N-1)\lambda}][b_0 e^{i(N-1)\lambda} + \dots + b_{N-1}]}{1 - a_1 e^{i\lambda} - \dots - a_N e^{iN\lambda}} d\lambda \\ &\quad + \int_{-\pi}^{\pi} e^{i\lambda n} [e^{iN\lambda} - \dots - a_N] d\hat{G}(\lambda). \end{aligned}$$

The last integral vanishes since the bracket vanishes at every jump of  $\hat{G}(\lambda)$ . The denominator in the first integrand is the value on  $|z| = 1$  of a polynomial all of whose roots are outside  $|z| = 1$ , or on  $|z| = 1$ . Any zero on  $|z| = 1$  corresponds to one of the numerator at the same point. The integral therefore vanishes if  $n \geq 0$  (Cauchy Integral Theorem):

$$(3.9.11) \quad R(n + N) - a_1 R(n + N - 1) - \dots - a_N R(n) = 0, \quad n \geq 0.$$

This equation implies that

$$(3.9.12) \quad x(n + N) - a_1 x(n + N - 1) - \dots - a_N x(n)$$

is independent of the chance variables  $\dots, x(n - 1), x(n)$ , that is that (3.8.4) is true, where  $\eta_1(n + N)$  has the stated properties. It has already been seen in the proof of Theorem 3.8 that this implies (3.8.6) and that this in turn implies that the process is a component process of an  $N$ -dimensional t.h.G.M. process whose transition matrix  $A$  has characteristic equation (3.8.2). In particular if  $G(\lambda) \equiv \hat{G}(\lambda)$ , the roots of the characteristic equation are of modulus 1, so that the  $N$ -dimensional process must be deterministic. If  $\hat{G}(\lambda) \equiv 0$ , the  $x(n)$  process is a component process of an  $N$ -dimensional process whose transition matrix  $A$  has only characteristic values of modulus less than 1. This  $N$ -dimensional process can have no deterministic factors other than one or more of type  $M(0)$ . If these exist, (and if the  $x(n)$  process is not of type  $M(0)$ ) they can be replaced by non-degenerate factors of type  $M$ , to obtain an  $N$ -dimensional process with no deterministic factor, having the  $x(n)$  process as a component process.

*Proof of (ii).* If the  $x(n)$  process is a t.h.G.M<sub>N</sub>. process, (3.8.4) is true with  $\eta_1(m)$  independent of  $\eta_1(n)$  if  $m \neq n$ . The discussion in (i) is therefore simplified by the fact that the numerator in (3.9.9) is constant. If this constant is 0, the spectral function is a function of jumps:  $G(\lambda) \equiv \hat{G}(\lambda)$ . If this constant is not 0, the denominator in (3.9.9) does not vanish, and  $\hat{G}(\lambda)$  therefore vanishes identically. The converse is proved as in (i).

**THEOREM 3.10.** (i) *If  $a_1, \dots, a_N$  are real numbers, there is a one-dimensional t.h.G. process not of type  $M(0)$  with correlation function  $R(n)$  satisfying*

$$(3.10.1) \quad R(n + N) - a_1R(n + N - 1) - \dots - a_NR(n) = 0$$

for  $n \geq 0$  if and only if the equation

$$(3.10.2) \quad \alpha^N - a_1\alpha^{N-1} - \dots - a_N = 0$$

has at least one root of modulus less than or equal to 1.

Let  $\dots, x(0), x(1), \dots$  be the variables of a one-dimensional t.h.G. process not of type  $M(0)$ .

(ii) *This process is a component process of an  $N$ -dimensional t.h.G.M. process if and only if the correlation function  $R(n)$  satisfies an  $N$ th order linear difference equation (3.10.1) for  $n \geq 0$ .*

(iii) *The process is a t.h.G.M<sub>N</sub>. process if and only if the difference equation (3.10.1) is true for  $n \geq -(N - 1)$ . In this case the vectors  $\{x(n), \dots, x(n + N - 1)\}$  determine an  $N$ -dimensional t.h.G.M. process.*

(iv) *Equation (3.10.1) is satisfied for  $n \geq -N$  if and only if*

$$(3.10.3) \quad x(n + N) - a_1x(n + N - 1) - \dots - a_Nx(n) = 0, \quad n = 0, \pm 1, \dots$$

*Proof of (ii), (iii), (iv).* Let the  $x(n)$  process be a component process of an  $N$ -dimensional t.h.G.M.  $y(n)$  process with correlation function  $R_y(n)$ :  $x(n) = y_1(n)$ , and transition matrix  $A$ . Since  $A$  satisfies its characteristic equation (3.8.2), it follows from (3.6.5) that

$$(3.10.4) \quad R_y(n + N) - a_1R_y(n + N - 1) - \dots - a_NR_y(n) = 0, \quad n \geq 0.$$

Then  $R(n) = (R_y(n))_{11}$  satisfies this same difference equation. Conversely if (3.10.1) is true for  $n \geq 0$ , it has already been proved in the course of the proof of Theorem 3.9 that the  $x(n)$  process is a component process of an  $N$ -dimensional t.h.G.M. process. This finishes the proof of (ii). Parts (iii) and (iv) are proved similarly.

*Proof of (i).* According to (ii), if there is a one-dimensional t.h.G. process whose correlation function  $R(n)$  satisfies (3.10.1) for  $n \geq 0$ , the process is a component process of an  $N$ -dimensional t.h.G.M. process whose transition matrix  $A$  has (3.10.2) as characteristic equation. Since  $A$  has at least one characteristic value of modulus less than or equal to 1, (unless the  $x(n)$  process is of type  $M(0)$ ), (3.10.2) must have at least one root of modulus less than or equal to 1. Conversely if (3.10.2) has at least one such root, there is a real  $N$ -dimensional matrix  $A$  whose characteristic equation is (3.10.2), and which has simple ele-



mentary divisors. According to Theorem 3.6 (ii),  $A$  is then the transition matrix of some t.h.G.M. process. The correlation function of this process and hence that of each component process satisfies (3.10.1) for  $n \geq 0$ .

**THEOREM 3.11.** (i) *If  $a_1, \dots, a_N$  are real numbers, there is a one-dimensional t.h.G. process not of type  $M(0)$  satisfying*

$$(3.11.1) \quad x(n + N) - a_1x(n + N - 1) - \dots - a_Nx(n) = \eta(n + N),$$

$$n = 0, \pm 1, \dots$$

with  $\eta(m), \eta(n)$  independent for  $|m - n| \geq N$  if and only if (3.10.2) has at least one root of modulus less than or equal to 1.

Let  $\dots, x(0), x(1), \dots$  be the variables of a one-dimensional t.h.G. process.

(ii) *This process is a component process of an  $N$ -dimensional t.h.G.M. process if and only if (3.11.1) is true with  $\eta(m), \eta(n)$  independent for  $|m - n| \geq N$ . In this case  $\eta(n + N)$  will be independent of the chance variables  $\dots, x(n - 1), x(n)$ .*

(iii) *The process is a t.h.G.M. process if and only if in addition to the condition in (ii),  $\eta(n)$  is independent of the chance variables  $\dots, x(n - 2), x(n - 1)$ : deterministic case if  $\eta(n) = 0$  with probability 1, nondeterministic case otherwise.*

Since this theorem follows readily from the preceding theorems, the proof will be omitted.

The problem of predicting  $x(n)$  in terms of  $\dots, x(n - 2), x(n - 1)$  is trivial (theoretically at least) for t.h.G.M. <sub>$N$</sub>  processes. In fact these were defined as those processes for which the solution of the prediction problem is simply a linear combination  $\sum_{j=1}^N a_jx(n - j)$  of the preceding  $N$  variables. The solution will now be given for the more general class of component processes of  $N$ -dimensional t.h.G.M. processes, processes which have been described from several points of view in the preceding theorems.

The prediction problem for component processes of  $N$ -dimensional t.h.G.M. processes will be put into a more general setting. If the one-dimensional chance variables  $\{x(n)\}$  determine a t.h.G. process, with correlation function  $R(n)$ , the problem of finding  $E\{\dots, x(n - 2), x(n - 1); x(n)\}$  is that of finding a series  $\sum_{m=1}^{\infty} \gamma_m x(n - m)$ <sup>16</sup> such that

$$(3.12.1) \quad x(n) - \sum_{m=1}^{\infty} \gamma_m x(n - m)$$

is uncorrelated with every  $x(n - \nu)$  ( $\nu > 0$ ):

$$(3.12.2) \quad R(\nu) - \sum_{m=1}^{\infty} \gamma_m R(\nu - m) = 0, \quad \nu > 0.$$

---

<sup>16</sup> We are neglecting all convergence difficulties. They become trivial for the applications to be made below.

If the complex spectral function is  $G(\lambda)$ , (3.12.2) becomes

$$(3.12.3) \quad \int_{-\pi}^{\pi} e^{i\nu\lambda} \left\{ 1 - \sum_{m=1}^{\infty} \gamma_m e^{-im\lambda} \right\} dG(\lambda) = 0, \quad \nu = 0.$$

Let  $G(\lambda)$  be the integral of its derivative  $G'(\lambda)$ , that is let  $G(\lambda)$  be absolutely continuous. According to (3.12.3) the problem reduces to that of finding a function

$$(3.12.4) \quad f(z) = 1 - \frac{\gamma_1}{z} - \frac{\gamma_2}{z^2} - \dots \quad z = e^{i\lambda}$$

such that  $f(z)G'$  is of power series type, a function corresponding to an expansion in non-negative powers of  $z$ . The dispersion of the error of the prediction is

$$(3.12.5) \quad E \left\{ \left[ x(n) - \sum_{m=1}^{\infty} \gamma_m x(n-m) \right]^2 \right\} = \int_{-\pi}^{\pi} \left| e^{in\lambda} - \sum_{m=1}^{\infty} \gamma_m e^{i(n-m)\lambda} \right|^2 dG(\lambda) \\ = \int_{-\pi}^{\pi} |f|^2 dG(\lambda).$$

In particular if the  $x(n)$  process is a component process of an  $N$ -dimensional t.h.G.M. process,  $G(\lambda)$  is given by (3.9.5). It will be supposed throughout the following that  $\hat{G} \equiv 0$ . Then

$$(3.12.6) \quad G'(\lambda) = \frac{z(\beta_0 z^{N-1} + \dots + \beta_{N-1})(\beta_0 + \dots + \beta_{N-1} z^{N-1})}{(\alpha_1 z^N + \dots + \alpha_N)(\alpha_0 + \dots + \alpha_N z^N)}, \quad z = e^{i\lambda}.$$

In this case  $f \equiv 1$  if  $G' \equiv 0$ , and otherwise  $f$  is given by

$$(3.12.7) \quad f(z) = \frac{\beta_0(\alpha_0 z^N + \dots + \alpha_N)}{\alpha_0 z(\beta_0 z^{N-1} + \dots + \beta_{N-1})}$$

so that

$$(3.12.8) \quad |f(z)|^2 = \frac{\beta_0^2}{\alpha_0^2 G'}.$$

The dispersion of the prediction error is  $R(0)$  if  $G' \equiv 0$  and otherwise is  $2\pi\beta_0^2/\alpha_0^2$ . The prediction formula for  $x(n)$  in terms of the variables  $\dots, x(n - \nu - 1), x(n - \nu)$  has now been derived for  $\nu = 1$ , for the chance variables under discussion in this section. The solution for general  $\nu$  is easily obtained.

As  $\nu \rightarrow \infty$ , the prediction converges with probability 1, according to the corollary to Theorem 1.2. If the process is a component process of an  $N$ -dimensional t.h.G.M. process, and if  $\hat{G} \equiv 0$  in (3.9.5), the limit is 0. That is, in this case, the best predicted value of  $x(n)$  in terms of the distant past is near  $E\{x(n)\} = 0$ , the same predicted value which would be used with no knowledge of the past.

**4. Processes whose parameter  $t$  varies continuously**

The basic process in terms of which t.h.G.M. processes without deterministic factors were expressed in section 3 was a process whose variables  $\{\xi(n)\}$  were Gaussian, with

$$(4.1.1) \quad E\{\xi(n)\} = 0, \quad E\{\xi(m) \cdot \xi(n)\} = \delta_{m,n}I.$$

The corresponding process in the continuous parameter case is not obtained by replacing the integral parameters  $m, n$  in (4.1.1) by continuous parameters. In fact the process so defined does not satisfy any useful continuity conditions. In the present discussion, sums like  $\sum_m A_m \xi(m)$  will be replaced by Stieltjes

integrals  $\int A(t) d\xi(t)$ , and  $d\xi(t)$  thus will correspond to  $\xi(n)$ . The  $\xi(t)$  process is defined as follows. For any  $t_1 < \dots < t_n$ , the chance variables

$$\xi(t_2) - \xi(t_1), \dots, \xi(t_n) - \xi(t_{n-1})$$

are mutually independent  $N$ -dimensional Gaussian chance variables, and if  $s < t$ ,

$$(4.1.2) \quad E\{\xi(t) - \xi(s)\} = 0, \quad E\{[\xi(t) - \xi(s)] \cdot [\xi(t) - \xi(s)]\} = (t - s)I.$$

This process, called simply a  $\xi$ -process below has been discussed in great detail by Bachelier, Wiener and Lévy. The function  $\xi(t)$ , considered as a function of  $t$  is known to be continuous with probability 1.<sup>17</sup> The derivative  $\xi'(t)$  does not exist, since  $E\{[\xi_i(t+h) - \xi_i(t)]^2\}$  is proportional to  $h$ , whereas this mean would be proportional to  $h^2$  if  $\xi'(t)$  existed. In fact it has been shown that  $\xi(t)$  is (with probability 1) not even of bounded variation in any finite interval. However, if  $f(t)$  is a function defined and continuous for  $a \leq t \leq b$  (where  $a$  or  $b$  or both may be infinite), the integral

$$(4.1.3) \quad \int_a^b f(t) d\xi(t)$$

can be defined as the limit in the mean of the usual Stieltjes sum. If  $f(t)$  has a continuous derivative, the integral in (4.1.3) can be evaluated by integration by parts:

$$(4.1.4) \quad \int_a^b f(t) d\xi(t) = f(b)\xi(b) - f(a)\xi(a) - \int_a^b \xi(t)f'(t) dt.$$

Integrals of the following type will be used below:

$$(4.1.5) \quad y(t) = \int_a^t f(t - \tau) d\xi(\tau) = f(0)\xi(t) - f(t - a)\xi(a) + \int_a^t \xi(\tau)f'(t - \tau) d\tau$$

---

<sup>17</sup> Paley and Wiener, "Fourier transforms in the complex domain," *Am. Math. Soc. Colloq. Pub.*, Vol. 19, p. 148.

where  $f(t)$  is continuous and has two continuous derivatives. It is then evident that  $y(t)$  is continuous, but that  $y'(t)$  exists if and only if  $f(0) = 0$ . If  $f(0) = 0$ ,  $y'(t)$  is given by

$$(4.1.6) \quad y'(t) = \int_a^t f'(t - \tau) d\xi(\tau).$$

A more general process will also come into the discussion below, and will be called a  $\zeta$ -process. The chance variables  $\{\zeta(t)\}$  of a  $\zeta$ -process are Gaussian, and have the same independence property as the variables of a  $\xi$ -process. The second equation of (4.1.2) is dropped, so that (4.1.2) is replaced by

$$(4.1.7) \quad E\{\zeta(t) - \zeta(0)\} = 0, \quad E\{[\zeta(t) - \zeta(0)] \cdot [\zeta(t) - \zeta(0)]\} = D(t),$$

where the symmetric and non-negative definite matrix  $D(t)$  will sometimes be supposed to have special properties, such as continuity in  $t$ , etc. The independence property of the  $\zeta$ -process implies that

$$(4.1.8) \quad E\{[\zeta(t) - \zeta(s)] \cdot [\zeta(t) - \zeta(s)]\} = D(t) - D(s).$$

Hence  $D'(t)$  (if this derivative exists) is symmetric and non-negative definite.

**THEOREM 4.1.** *If the dispersion matrix  $D(t)$  of a  $\zeta$ -process is continuous, the functions  $\{\zeta(t)\}$  are continuous in  $t$ , with probability 1.*

The component processes of a  $\zeta$ -process with a continuous dispersion function are also  $\zeta$ -processes with continuous dispersion functions. Hence it will be sufficient to prove the theorem in the one-dimensional case. In this case  $D(t)$  is non-negative and monotone non-decreasing, according to (4.1.7) and (4.1.8). It can be supposed that  $D(t)$  does not vanish identically. Let  $D_1(t)$  be an inverse function of  $D(t)$ :  $D[D_1(t)] = t$ . Then  $\xi(t) = \zeta[D_1(t)]$  defines a  $\xi$ -process, and the continuity of  $\xi(t)$  implies that of  $\zeta(t)$ .

The integrals of type (4.1.3) are defined for  $\zeta$ -processes as for  $\xi$ -processes, and satisfy the equations

$$(4.1.9) \quad \begin{aligned} E \left\{ \int_a^b f(t) d\zeta(t) \right\} &= 0 \\ E \left\{ \int_a^b f(t) d\zeta(t) \cdot \int_a^b g(t) d\zeta(t) \right\} &= \int_a^b f(t)g(t)D'(t) dt \\ E \left\{ \int_a^b A(t) d\zeta(t) \cdot \int_a^b B(t) d\zeta(t) \right\} &= \int_a^b A(t)D'(t)B(t)^* dt, \end{aligned}$$

where  $f, g$  are numerically valued functions and  $A, B$  are matrix functions.<sup>18</sup>

The  $\zeta$ -processes lie at the basis of t.h.G. processes. To every t.h.G. process (discrete parameter) with variables  $\{x(n)\}$  correspond two one-dimensional  $\zeta$ -processes with variables  $\{\zeta_1(t)\}, \{\zeta_2(t)\}$  such that

$$(4.1.10) \quad x(n) = \int_{-\pi}^{\pi} \cos n\lambda d\zeta_1(\lambda) + \sin n\lambda d\zeta_2(\lambda)$$

<sup>18</sup> These equations are easily proved using the fact that each integral can be approximated by the usual Riemann-Stieltjes sums.

where the two  $\zeta$ -processes are mutually independent in the sense that every  $\zeta_1(\lambda_1)$  is independent of every  $\zeta_2(\lambda_2)$  and where, if  $G(\lambda)$  is the complex spectral function of the process,

$$(4.1.11) \quad E\{\zeta_i(\lambda)^2\} = G(\lambda).$$

In the continuous parameter case (4.1.11) becomes

$$(4.1.12) \quad x(t) = \int_{-\infty}^{\infty} \cos t\lambda \, d\zeta_1(\lambda) + \sin t\lambda \, d\zeta_2(\lambda).^{19}$$

This theorem of Cramér shows that  $x(n)$ , or  $x(t)$  as the case may be, is the limit of a sum of sines and cosines, with Gaussian chance variables as coefficients. The dispersion of each coefficient, which measures the intensity of the corresponding periodic term of the sum, is determined by the spectral function of the process. In particular, if the spectral function  $F(\lambda)$  is the integral of its derivative  $F'(\lambda)$ , each integrand involving  $d\zeta_i(\lambda)$  in the above equations can be replaced by one involving  $\sqrt{F'(\lambda)}d\xi_i(\lambda)$  where  $\xi_i(\lambda)$  is the variable of a  $\xi$ -process. Thus in many important cases the processes can be written in a simple way in terms of  $\xi$ -processes.

It will be shown below that every t.h.G.M. process can be represented as the direct product of factors of certain types. The deterministic types have already been catalogued:  $M(0)$ ,  $M(1)$ ,  $M(e^{i\theta})$ . The standard non-deterministic type, as in the discrete parameter case, will be called type  $M$ .

$M$ . Let  $\{\xi(t)\}$  be the chance variables of an  $N$ -dimensional  $\xi$ -process, as described above. Let  $Q$  be an  $N$ -dimensional square matrix, and let  $S$  be an  $N$ -dimensional symmetric non-negative definite matrix. Define  $x(t)$  by

$$(4.2.1) \quad x(t) = \int_0^{\infty} e^{sQ} S \, d\xi(t-s) = \int_{-\infty}^t e^{(t-\tau)Q} \, d\xi(\tau)$$

where it is supposed that the improper integrals converge with probability 1. There will be convergence, for example, if  $Q$  has only characteristic values with negative real parts so that the elements in the matrix  $e^{sQ}$  go to 0 exponentially as  $s \rightarrow \infty$ . (Cf. section 1.) It will be shown below that it is no restriction to assume that  $Q$  has this character. The  $x(t)$  process is evidently a t.h.G. process. If  $u < t$ , the chance variable

$$(4.2.2) \quad x(t) - e^{(t-u)Q} x(u) = e^{tQ} \int_u^t e^{-sQ} S \, d\xi(s)$$

is independent of  $x(u)$  for  $v \leq u$ , since  $x(v)$  is expressed in terms of  $\xi(s)$  for  $s \leq v$ . Therefore the  $x(t)$  process is a Markoff process with transition matrix  $A(t) = e^{tQ}$ :

$$(4.2.3) \quad E\{x(v), v \leq u; x(t)\} = e^{(t-u)Q} x(u), \quad u < t.$$

<sup>19</sup> H. Cramér, *Arkiv For Matematik, Astronomi och Fysik*, Vol. 28B, No. 12, pp. 1-17. Cramér only discusses the continuous parameter case, but the other requires no change of method. He allows complex-valued  $\xi$ -processes, in terms of which (4.1.10) and (4.1.12) assume a more elegant form.

A process defined in this way will be called a process of type  $M$ . A change of variable  $y(t) = Bx(t)$  leads to a process of the same type:

$$(4.2.4) \quad y(t) = \int_0^\infty e^{sBQB^{-1}} BS d\xi(t - s).$$

The matrix  $Q$  goes into  $BQB^{-1}$  and if  $S_1O$  is the polar form of  $BS$ , where  $S$  is symmetric and non-negative definite and  $O$  is orthogonal,  $S$  goes into  $S_1$ . (We are using the fact that  $O\xi(t)$  defines a second  $\xi$ -process.) The correlation function of a process of type  $M$  is easily calculated:

$$(4.2.5) \quad R(0) = \int_0^\infty e^{sQ} S^2 e^{sQ^*} ds, \quad R(t) = R(0)e^{tQ^*}.$$

The only condition imposed on  $Q$  is that the improper integrals in (4.2.1) converge. This condition is easily seen to be equivalent to the convergence of the integral in (4.2.5). This in turn is equivalent to the condition that

$$(4.2.6) \quad \lim_{t \rightarrow \infty} e^{tQ} S = 0.$$

This condition is certainly satisfied if the characteristic values of  $Q$  all have negative real parts, and it can always be assumed that this is so. (Cf. the corresponding discussion of processes of type  $M$  in the discrete parameter case.)

The analogues in the continuous parameter case of Theorems 3.4 and 3.5 are true. The proofs are substantially the same as the proofs in the discrete parameter case, and will be omitted.

**THEOREM 4.3.** (i) *Every t.h.G.M. process (continuous parameter) is the direct product of processes of type  $M(0)$ ,  $M(1)$ ,  $M(e^{i\theta})$ ,  $M$ .*

(ii) *If  $x(t)$  are the variables of such a process, there is a matrix  $Q$  such that  $A(t) = e^{tQ}$  is a transition matrix function. There is a  $\xi$ -process, a Gaussian variable  $\xi$ , independent of the  $\xi(t)$ , satisfying*

$$(4.3.1) \quad E\{\xi\} = 0, \quad E\{\xi \cdot \xi\} = I$$

and symmetric non-negative definite matrices  $S, T$  such that

$$(4.3.2) \quad \begin{aligned} x(t) &= \int_0^\infty e^{sQ} S d\xi(t - s) + e^{tQ} T\xi \\ &= \int_{-\infty}^t e^{(t-s)Q} S d\xi(s) + e^{tQ} T\xi = \int_0^t e^{(t-s)Q} S d\xi(s) + e^{tQ} x(0), \end{aligned}$$

$$(4.3.3) \quad QT^2 + T^2 Q^* = 0,$$

$$(4.3.4) \quad R(0) = \int_0^\infty e^{sQ} S^2 e^{sQ^*} ds + T^2$$

$$(4.3.5) \quad QR(0) + R(0)Q^* = -S^2,$$

where the integrals in (4.3.2) converge with probability 1. The integral and the last term in each pair in (4.3.2) are linear transformations of  $x(t)$ : (4.3.2) exhibits

in part the decomposition into factor processes described in (i). The correlation function is given by

$$(4.3.6) \quad \begin{aligned} R(t) &= R(0)e^{tQ} \\ R(-t) &= e^{tQ}R(0) \end{aligned} \quad t \geq 0.$$

(iii) The matrix  $Q$  is uniquely determined if and only if the process is non-degenerate. In any case there is a  $Q$  whose characteristic values all have negative or zero real parts and whose characteristic values with zero real parts correspond to simple elementary divisors. The matrix  $Q$  furnishes the solution to the prediction problem of the process:

$$(4.3.7) \quad E\{x(s), s \leq t; x(t + u)\} = e^{uQ}x(t), \quad u > 0.$$

The matrix  $S$ , which is uniquely determined, measures the dispersion of  $x(t)$  about its predicted value:

$$(4.3.8) \quad E\{[x(t + u) - e^{uQ}x(t)]^2\} = R(0) - e^{uQ}R(0)e^{uQ*} \sim uS^2 \quad (u \rightarrow 0).$$

(iv) Conversely if  $Q$  is a matrix with at least one characteristic value with negative real part or with zero real part and corresponding to a simple elementary divisor,  $e^{tQ}$  is the transition matrix function of a t.h.G.M. process with  $R(0)$  not the null matrix. If all the characteristic values of  $Q$  are as just described,  $e^{tQ}$  is the transition matrix function of a non-degenerate t.h.G.M. process. If  $R(0)$ ,  $S$ ,  $Q$  are matrices satisfying (4.3.5) with  $R(0)$ ,  $S$  symmetric and non-negative definite, there is a t.h.G.M. process whose variables can be written in the form (4.3.2) with the given  $R(0)$ ,  $S$ ,  $Q$ .

The proof of Theorem 4.3 follows closely that of Theorem 3.6, and the details will not be given, except as they differ from those of the earlier proof.

*Proof of (i).* Suppose that the  $\{x(t)\}$  are the variables of a t.h.G.M. process which is non-degenerate. The transition matrix function  $A(t)$  is then uniquely determined by (1.5.5). Take the conditional expectation of both sides of (1.5.4) for given  $x(0)$ :

$$(4.3.9) \quad A(s + t)x(0) = A(t)A(s)x(0) \quad s, t > 0.$$

Since the process is non-degenerate,

$$(4.3.10) \quad A(s + t) = A(s)A(t) \quad s, t > 0.$$

According to (1.3.1) and (1.5.5)

$$(4.3.11) \quad \lim_{t \rightarrow 0} R(t) = \lim_{t \rightarrow 0} R(0)A(t)^* = R(0), \quad t > 0.$$

Hence

$$(4.3.12) \quad \lim_{t \rightarrow 0} A(t) = I.$$

It has already been noted that any solution to (4.3.10) under the continuity hypothesis (4.3.12) can be written in the form

$$(4.3.13) \quad A(t) = e^{tQ},$$

where

$$Q = \lim_{t \rightarrow 0} \frac{A(t) - I}{t}.$$

Under a change of variables  $y(t) = Bx(t)$ ,  $A(t)$  becomes  $BA(t)B^{-1}$  and  $Q$  becomes  $BQB^{-1}$ . According to Theorem 2.1, if the  $x(t)$  process is degenerate, it is the direct product of one or more factors of type  $M(0)$  and (perhaps) of a non-degenerate factor. The matrix  $Q$  of a factor of type  $M(0)$  can be taken as the null matrix. Then the form (4.3.13) is admissible for any t.h.G.M. process, although  $Q$  will only be uniquely determined if the process is non-degenerate. Define  $\zeta(t)$  by

$$(4.3.14) \quad \zeta(t) = A(t)^{-1}x(t) = e^{-tQ}x(t).$$

Then if  $s_1 < t_1 < s_2 < t_2$

$$(4.3.15) \quad E\{\zeta(t_1) - \zeta(s_1)\} = 0, \quad E\{[\zeta(t_2) - \zeta(s_2)] \cdot [\zeta(t_1) - \zeta(s_1)]\} = 0$$

and

$$(4.3.16) \quad D(t) = E[\zeta(t) - \zeta(0)] \cdot [\zeta(t) - \zeta(0)] = e^{-tQ}R(0)e^{-tQ*} - R(0).$$

Hence the  $\{\zeta(t)\}$  determine a  $\zeta$ -process, with dispersion matrix given by (4.3.16). The derivative  $D'(t)$  is easily evaluated:

$$(4.3.17) \quad D'(t) = e^{-tQ}[-R(0)Q* - QR(0)]e^{-tQ*}.$$

Since  $D'(t)$  is symmetric and non-negative definite, the bracket also has this property, and there is a non-singular matrix  $S_1$  such that

$$(4.3.18) \quad S_1[-QR(0) - R(0)Q*]S_1^* = U,$$

where  $U$  is in diagonal form, with only 0's and 1's in the main diagonal. Then the integral

$$(4.3.19) \quad \int_0^t S_1 e^{sQ} d\zeta(s)$$

defines a  $\zeta$ -process with dispersion matrix  $tU$ . There is therefore a  $\xi$ -process with variables  $\{\xi(t)\}$  such that

$$(4.3.20) \quad U\xi(t) = \int_0^t S_1 e^{sQ} d\zeta(s).$$

This equation can be solved for  $\zeta(t)$  and  $x(t)$ :

$$(4.3.21) \quad x(t) = e^{tQ}\zeta(t) = e^{tQ} \int_0^t e^{-sQ} S_2 U d\zeta(s) + e^{tQ}x(0)$$



where  $S_2 = S_1^{-1}$ . The matrix  $S_2U$  can be written in the polar form  $SO$  where  $S$  is symmetric and non-negative definite and  $O$  is orthogonal. This  $S$  is the  $S$  of (4.3.2) etc.

The remainder of the proof follows closely the proof of Theorem 3.6 and will be omitted.

An important class of t.h.G.M. processes which arises frequently in physical applications is obtained in the following way. Let  $\{\xi(t)\}$  be the variables of a one-dimensional  $\xi$ -process. Consider the formal equation

$$(4.4.1) \quad \frac{d^N y(t)}{dt^N} - a_1 \frac{d^{N-1} y(t)}{dt^{N-1}} - \dots - a_N y(t) = c \xi'(t),$$

where  $a_1, \dots, a_N, c$  are constants. This equation cannot be considered precise as it stands, since  $\xi'(t)$  does not exist. The problem can however be reformulated as follows: find a  $y(t)$  process, where  $y', \dots, y^{(N-1)}$  are supposed to exist, satisfying the equation

$$(4.4.2) \quad \int_a^b f(t) dy^{(N-1)}(t) - a_1 \int_a^b f(t) dy^{(N-2)}(t) - \dots - a_N \int_a^b f(t) y(t) dt = c \int_a^b f(t) d\xi(t)$$

with probability 1, for each continuous function  $f(t)$  and each pair of numbers  $a, b$ . The formal integrals are defined as the limit in the mean of the usual sums.<sup>20</sup> The integral on the right has already been discussed. With this interpretation, equations involving  $\xi'$  can be treated in the usual way, and this will be done in the following without further comment. The formal solution of (4.4.2) is well known. Let  $\lambda_1, \dots, \lambda_N$  be the roots of the equation

$$(4.4.3) \quad \lambda^N - a_1 \lambda^{N-1} - \dots - a_N = 0$$

and suppose that these roots are distinct, and have negative real parts. Let  $\Delta_{jk}$  be the cofactor of  $\lambda_j^{k-1}$  in the determinant

$$(4.4.4) \quad \delta = \begin{vmatrix} 1 & \cdot & \cdot & \cdot & 1 \\ \lambda_1 & \lambda_2 & \cdot & \cdot & \lambda_N \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \lambda_1^{N-1} & \lambda_2^{N-1} & \cdot & \cdot & \lambda_N^{N-1} \end{vmatrix}.$$

Then the general solution of (4.4.1), that is to say of (4.4.2), is

$$(4.4.5) \quad y(t) = \frac{c}{\delta} \int_0^t \sum_{j=1}^N \Delta_{Nj} e^{\lambda_j(t-s)} d\xi(s) + \frac{1}{\delta} \sum_{j,k=1}^N \Delta_{kj} e^{\lambda_j t} y^{(k-1)}(0).$$

Since the integrand and its first  $N - 1$  derivatives vanish when  $s = t$ ,  $y', \dots, y^{(N-1)}$  as defined by (4.4.5) exist, but  $y^{(N)}(t)$  does not exist, because  $\xi'(t)$  in (4.4.1)

<sup>20</sup> For a full discussion in the case  $N = 1$  cf. Doob, *Annals of Math.*, Vol. 43 (1942), pp. 358-61.

does not exist. The  $y(t)$  process is a t.h.G. process if  $y(0), \dots, y^{(N-1)}(0)$  are chosen properly. This can be seen from the solution

$$(4.4.6) \quad y(t) = \frac{c}{\delta} \int_{-\infty}^t \sum_{j=1}^N \Delta_{Nj} e^{\lambda_j(t-s)} d\xi(s).$$

In fact this is the only solution defining a t.h.G. process. To prove this, rewrite (4.4.5) in the form

$$(4.4.5') \quad y(t) = \frac{c}{\delta} \sum_{j=1}^N \Delta_{Nj} \int_{t-\tau}^t e^{\lambda_j(t-s)} d\xi(s) + \frac{1}{\delta} \sum_{j,k=1}^N \Delta_{kj} e^{\lambda_j(t-\tau)} y^{(k-1)}(\tau).$$

If the  $y(t)$  process is a t.h.G. process, (4.4.5') becomes (4.4.6) when  $\tau \rightarrow -\infty$ . Thus there is a unique stationary solution to (4.4.1) and, by (4.4.5), every solution tends to this solution in the long run. The stationary solution (4.4.6) has the property that  $y(t)$  is written in terms of  $\xi(s)$  for  $s \leq t$ . Then in (4.4.5) the integral is independent of the terms involving the initial conditions. In other words

$$(4.4.7) \quad E\{y(s), s \leq 0; y(t)\} = \frac{1}{\delta} \sum_{j,k=1}^N \Delta_{kj} e^{\lambda_j t} y^{(k-1)}(0).$$

Hence the variables  $y(t), y'(t), \dots, y^{(N-1)}(t)$  define an  $N$ -dimensional t.h.G.M. process. The transition matrix function  $A(t)$ , and the matrices  $Q, S, T$  of Theorem 4.3 are easily calculated.

$$(4.4.8) \quad \begin{aligned} A(t) &: (\Delta_{kj} \lambda_j^{k-1} e^{\lambda_j t}), \\ Q &: \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 \\ a_N & \cdot & \cdot & \cdot & a_1 \end{pmatrix} \\ T &= 0. \end{aligned}$$

The necessary changes to be made if the  $\lambda_j$  are not distinct are well known. The case  $c = 0$  will be treated below, when the problem will be reconsidered from another point of view. In all cases the solution of (4.4.1) leads to an  $N$ -dimensional t.h.G.M. process.<sup>21</sup>

As a simple example, consider a torsion pendulum, suspended in a sealed container. The only turning forces acting on the pendulum are the molecular shocks of the surrounding gas, and the restoring torque. The equation of motion is

$$(4.4.9) \quad I \frac{d^2 y(t)}{dt^2} + \alpha_1 \frac{dy(t)}{dt} + \alpha_2 y(t) = X(t),$$

---

<sup>21</sup> According to a letter from Uhlenbeck, the differential equation (4.4.1) was solved; from a somewhat different point of view, by Miss Ming Chen Wang, in a thesis written in 1941 which is unfortunately inaccessible to me at the moment.

where  $y$  is the angular displacement measured from the equilibrium position,  $I$  is the moment of inertia,  $\alpha_2$  is the torque coefficient of the suspension, and the molecular force is resolved into a systematic Stokes term  $\alpha_1 y'$  and a remainder  $X$ . The remainder term  $X(t)$  defines a stationary process which to a first approximation is "purely random." In the present context "purely random" means that if  $t_1 < \dots < t_\nu$ ,  $X(t_1), \dots, X(t_\nu)$  are mutually independent. This is precisely the property the derived process of a  $\xi$ -process would have, if  $\xi'(t)$  existed. Unfortunately it has already been noted that  $\xi'(t)$  does not exist, since the difference quotient  $[\xi(t+h) - \xi(t)]/h$  is unbounded as  $h \rightarrow 0$ . It has already been seen, however, that (4.4.9) can be given a meaning with  $X(t)$  identified with  $c\xi'(t)$  even though  $\xi'(t)$  does not exist, and it has been seen that the solution approaches a steady state. It may still be a disappointment to some that the solution  $y(t)$  has a first derivative  $y'(t)$  but that  $y''(t)$  does not exist: there is an angular velocity but not an angular acceleration! This unhappy circumstance can either be blamed on the physical world, or on the mathematical approximation to the physical world, depending on the point of view. The corresponding electrical picture is the following. There are spontaneous currents in any electrical circuit, due to the thermal motion of the electrons. This is known as the Johnson effect. In a simple closed circuit, consisting of an inductance  $L$ , a resistance,  $R$ , and a capacitance  $C$  in series, the current equation can be written in the form

$$(4.4.10) \quad L \frac{d^2 y(t)}{dt^2} + R \frac{dy(t)}{dt} + \frac{y(t)}{C} = X(t),$$

where  $y$  is the charge on the condenser and  $X(t)$  represents a fictitious voltage set up by the motion of the electrons. The  $X(t)$  is identified with  $c\xi'(t)$  as before. In this case there is a current  $\frac{dy}{dt}$ , but the current function has no derivative. In these applications, the physical justification for the Gaussian character of the  $\xi$ -distribution lies in the Gaussian character of the Maxwell distribution of elementary particle velocities. The known mean particle kinetic energy determines the constant  $c$  in (4.4.1). The more complicated mechanical or electrical systems will lead to equations of higher order than 2, or systems of equations. For example the usual current equations of a net or resistances capacitances and inductances lead to a system of say  $\nu$  second order equations of type (4.4.10), and the corresponding pairs  $y, y'$  form a  $2\nu$ -dimensional t.h.G.M. process.<sup>22</sup>

The processes defined by linear differential equations of the type (4.4.1) are the analogues of the t.h.G.M.<sub>N</sub> processes in the discrete parameter case. Instead of defining these solutions of (4.4.1) as the t.h.G.M.<sub>N</sub> processes, however, we shall use a definition closer to the definition in the discrete parameter case. A

<sup>22</sup> Further discussion and references to papers by physicists on this subject will be found in Doob, *Annals of Math.*, Vol. 43 (1942), pp. 351-69.

one-dimensional t.h.G. process with variables  $\{y(t)\}$  will be called a t.h.G.M. $_N$  process if the derivatives  $y'(t), \dots, y^{(N-1)}(t)$  exist, and if whenever  $s < t$ ,

$$(4.5.1) \quad E\{y(\tau), \tau \leq s; y(t)\} = E\{y(s), y'(s), \dots, y^{(N-1)}(s); y(t)\}.$$

If  $N = 1$ , the process is a t.h.G.M. process. The right hand side of (4.5.1) is a linear combination of the variables  $y(s), \dots, y^{(N-1)}(s)$ . The variables  $\{y(t)\}$  thus satisfy an equation of the form

$$(4.5.2) \quad y(t) - a_1(t-s)y(s) - \dots - a_N(t-s)y^{(N-1)}(s) = \eta(s, t)$$

where  $\eta(s, t)$  is independent of the variables  $\{y(\tau)\}$  for  $\tau \leq s$ . Define the variables  $\{x(t)\}$  of an  $N$ -dimensional t.h.G. process by

$$(4.5.3) \quad \begin{aligned} x_1(t) &= y(t) \\ x_j(t) &= y^{(j)}(t), \quad j = 1, \dots, N-1. \end{aligned}$$

If this process is degenerate, there is a relation of the form

$$(4.5.4) \quad c_0 y(s) + c_1 y'(s) + \dots + c_{N-1} y^{(N-1)}(s) = 0, \quad \sum_{j=0}^{N-1} |c_j| > 0.$$

It can be assumed that  $c_{N-1} \neq 0$ , (differentiating (4.5.4) to get a term in  $y^{(N-1)}(t)$  if there is none originally). Then  $y^{(N-1)}(s)$  can be eliminated in (4.5.2), to get a relation of the same type with  $N$  replaced by  $N-1$ . Hence the process is non-degenerate if  $N$  is the minimum index for which (4.5.1) is true. It will now be proved that the  $x(t)$  process is a t.h.G.M. process. It can be assumed to be non-degenerate. Using (4.5.1),

$$(4.5.5) \quad E\{x(\tau), \tau \leq s; x_1(t)\} = E\{y(\tau), \tau \leq s; y(t)\} = E\{x(s); x_1(t)\}.$$

It must also be shown that

$$(4.5.6) \quad E\{x(\tau), \tau \leq s; x_j(t)\} = E\{x(s); x_j(t)\} \quad j = 2, \dots, N.$$

This will be shown by justifying the taking of derivatives in (4.5.5). It will be sufficient to prove (4.5.6) when  $j = 2$ . Using (4.5.1),

$$(4.5.7) \quad E\left\{x(\tau), \tau \leq s; \frac{y(t+h) - y(t)}{h}\right\} = E\left\{x(s); \frac{y(t+h) - y(t)}{h}\right\}.$$

The right hand side is a linear combination of  $x_1(s), \dots, x_N(s)$  whose coefficients are continuous in  $h$ ,  $h \geq 0$ , since the correlation function of the  $y(t)$  process is continuous. Hence the right hand side converges to

$$E\{x(s); y'(t)\} = E\{x(s); x_2(t)\}$$

when  $h \rightarrow 0$ . Since the difference

$$x(t) - E\left\{x(s); \frac{y(t+h) - y(t)}{h}\right\}$$

is uncorrelated with  $x(\tau)$  if  $\tau \leq s$ , the same is true of its limit as  $h \rightarrow 0$ . This means that (4.5.6) is true when  $j = 2$ , as was to be shown. Conversely if  $\{y(t)\}$  are the variables of a one-dimensional t.h.G. process, if  $y'(t), \dots, y^{(N-1)}(t)$  exist, and if the  $x(t)$  process defined by (4.5.3) is a t.h.G.M. process, the  $y(t)$  process is obviously a t.h.G.M.<sub>N</sub> process. The transition matrix function  $A(t)$  and the matrices  $Q, S, T$  of Theorem 4.3 are easily calculated. Suppose that the  $x(t)$  process is non-degenerate. Since  $y^{(i-1)}(t)$  is given by

$$(4.5.8) \quad \begin{aligned} y^{(i-1)}(t) &= x_i(t) = \int_{-\infty}^t \sum_{j=1}^N [e^{(t-s)Q} S]_{ij} d\xi_j(s) + \sum_{j=1}^N [e^{tQ} T]_{ij} \xi_j \\ &= \int_0^t \sum_{j=1}^N [e^{(t-s)Q} S]_{ij} d\xi_j(s) + \sum_{j=1}^N (e^{tQ})_{ij} x_j(0) \end{aligned}$$

and since  $x'_i(t)$  exists if  $i < N$ , it follows that the integrand must vanish when  $s = t$ :

$$(4.5.9) \quad (S)_{ij} = 0, \quad i = 1, \dots, N - 1, \quad j = 1, \dots, N.$$

Since  $S$  is symmetric and non-negative definite,  $S$  must have the form

$$(4.5.10) \quad S: \begin{pmatrix} 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & c \end{pmatrix}, \quad c \geq 0.$$

The fact that  $x'_i(t) = x_{i+1}(t)$  means that

$$(4.5.11) \quad \sum_{j=1}^N (e^{tQ} Q)_{ij} x_j(0) = \sum_{j=1}^N (e^{tQ})_{i+1j} x_j(0), \quad j = 1, \dots, N$$

or, since the  $x(t)$  process is non-degenerate,

$$(4.5.12) \quad (e^{tQ} Q)_{ij} = (e^{tQ})_{i+1j} \quad \begin{matrix} i = 1, \dots, N - 1 \\ j = 1, \dots, N. \end{matrix}$$

Hence  $(t \rightarrow 0) Q$  has the form

$$(4.5.13) \quad Q: \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 \\ a_N & \cdot & \cdot & \cdot & \cdot & a_1 \end{pmatrix}.$$

Conversely if there is an  $N$ -dimensional non-degenerate t.h.G.M. process with transition matrix function  $e^{tQ}$  where  $Q$  is given by (4.5.13) and dispersion matrix  $S$  given by (4.5.10),

$$x'_i(t) = x_{i+1}(t), \quad i = 1, \dots, N - 1,$$

and the  $x_1(t)$  process is a t.h.G.M.<sub>N</sub> process.

Case 1.  $S = 0$  (*deterministic case*). In this case the  $x(t)$  process is deterministic:

$$(4.5.14) \quad x(t) = e^{tQ} T \xi.$$

Since  $Q$  satisfies its characteristic equation

$$(4.5.15) \quad \alpha^N - a_1 \alpha^{N-1} - \dots - a_N = 0$$

it follows that

$$(4.5.16) \quad x^{(N)}(t) - a_1 x^{(N-1)}(t) - \dots - a_N x(t) = 0,$$

$$(4.5.17) \quad y^{(N)}(t) - a_1 y^{(N-1)}(t) - \dots - a_N y(t) = 0.$$

The roots of (4.5.15) are simple roots, and are all pure imaginary, according to Theorem 4.3. It follows that

$$(4.5.18) \quad y(t) = \sum_j (\eta_j \cos t\theta_j + \zeta_j \sin t\theta_j)$$

where the  $\eta_j$  and  $\zeta_j$  are one-dimensional Gaussian variables, and  $\{i\theta_j\}$  are the distinct roots of (4.5.15).

Case 2.  $S \neq 0$  (*non-deterministic case*). In this case it will now be shown that the  $x(t)$  process has no deterministic factor, that is that the roots of (4.5.15) all have negative real parts. In fact let  $\beta$  be a root of (4.5.15), corresponding to the characteristic vector  $z$  of  $Q^*$ :

$$(4.5.19) \quad z = (a_N \beta^{N-1}, a_N \beta^{N-2} + a_{N-1} \beta^{N-1}, \dots, a_N + a_{N-1} \beta + \dots + a_1 \beta^{N-1}) \\ = (\dots, \beta^N).$$

Then using (4.3.5)

$$(4.5.20) \quad 0 < c^2 |\beta|^{2N} = (S^2 z, z) = -(R(0)Q^*z, z) - (QR(0)z, z) \\ = -\lambda(R(0)z, z) - \bar{\lambda}(R(0)z, z) \\ = -(\lambda + \bar{\lambda})(R(0)z, z).$$

Hence  $\lambda + \bar{\lambda}$  is real and negative:  $\lambda$  has a negative real part. In this non-deterministic case, therefore, the  $x(t)$  process can have no deterministic factor. The matrix  $T$  is the null matrix, and (4.3.2) becomes

$$(4.5.21) \quad x(t) = \int_{-\infty}^t e^{(t-s)Q} S d\xi(s)$$

which leads to

$$(4.5.22) \quad y(t) = c \int_{-\infty}^t [e^{(t-s)Q}]_{1N} d\xi_N(s).$$

Moreover

$$\begin{aligned}
 (4.5.23) \quad & y'(t) = c \int_{-\infty}^t [Qe^{(t-s)Q}]_{1N} d\xi_N(s) \\
 & \dots\dots\dots \\
 & y^{(N-1)}(t) = c \int_{-\infty}^t [Q^{N-1}e^{(t-s)Q}]_{1N} d\xi_N(s).
 \end{aligned}$$

Since  $Q$  satisfies its characteristic equation (4.5.15),

$$(4.5.24) \quad \int_{-\infty}^t [Q^N e^{(t-s)Q}]_{1N} d\xi_N(s) - \sum_{j=1}^N a_j \int_{-\infty}^t [Q^{N-j} e^{(t-s)Q}]_{1N} d\xi_N(s) = 0.$$

In other words

$$(4.5.25) \quad c \int_{-\infty}^t [Q^N e^{(t-s)Q}]_{1N} d\xi_N(s) - a_1 y^{(N-1)}(t) - \dots - a_N y(t) = 0.$$

Now formally, if  $\xi'_N(t)$  existed, the last equation in (4.5.23) could be differentiated to give

$$(4.5.26) \quad y^{(N)}(t) = c[Q^{N-1}]_{1N} \xi'_N(t) + c \int_{-\infty}^t [Q^N e^{(t-s)Q}]_{1N} d\xi_N(s)$$

and (4.5.25) would become

$$(4.5.27) \quad y^{(N)}(t) - a_1 y^{(N-1)}(t) - \dots - a_N y(t) = c\xi'_N(t).$$

(We are using the fact that  $(Q^{N-1})_{1N} = 1$ .) Thus the t.h.G.M.<sub>N</sub>. processes satisfy the formal differential equation (4.5.27) already discussed above from another point of view. Equation (4.4.2) is readily justified.

**THEOREM 4.6.** (i) *Let  $\{x(t)\}$  be the variables determining a t.h.G.M. process. Then considered as functions of  $t$ , the  $x(t)$  are continuous with probability 1. Let  $\{y(t)\} = \{x_r(t)\}$  be the variables of a coordinate process.*

(ii) *If  $y'(t)$  exists, it is a linear combination of coordinate functions:*

$$y'(t) = \sum_{j=1}^N c_j x_j(t).$$

(iii) *If  $y'(t), \dots, y^{(N-1)}(t)$  exist,  $y(t)$  satisfies a generalized differential equation (4.4.1), that is the  $y(t)$  process is a t.h.G.M.<sub>N</sub>. process.*

(iv) *If  $y'(t), \dots, y^{(N)}(t)$  exist,  $y(t)$  has derivatives of all orders. The  $y(t)$  process is a t.h.G.M.<sub>N</sub>. process (deterministic case) and  $y(t)$  therefore satisfies an  $N$ th order homogeneous differential equation (4.5.17).*

(v) *If  $x'_1(t), \dots, x'_N(t)$  exist, that is if  $x'(t)$  exists, the  $x(t)$  process is deterministic and the coordinate functions have derivatives of all orders.*

*Proof of (i).* It has already been shown that the  $\{\zeta(t)\}$  determined by (4.3.14) determine a  $\zeta$ -process, and the dispersion matrix function  $D(t)$  of the  $\zeta$ -process, given by (4.3.16), is certainly continuous. Hence, by Theorem 4.1, the  $\{\zeta(t)\}$ , and therefore the  $\{x(t)\}$  are continuous in  $t$ , with probability 1.

*Proof of (ii).* If  $x'_r(t)$  exists, the  $r$ th row of  $S$  in (4.3.2) must vanish, and  $x'_r(t)$  is given by the  $r$ th coordinate of

$$(4.6.1) \quad \int_{-\infty}^t Q e^{(t-s)Q} S d\xi(s) + Q e^{tQ} x(0) = Qx(t).$$

Hence

$$(4.6.2) \quad x'_r(t) = \sum_{j=1}^N (Q)_{rj} x_j(t).$$

*Proof of (iii).* Suppose that  $x'_r(t), \dots, x_r^{(N-1)}(t)$  exist. Then ( $r$  is fixed in the following equations)

$$(4.6.3) \quad \begin{aligned} x_r(t) &= \int_0^t \sum_{j=1}^N [e^{(t-s)Q} S]_{rj} d\xi_j(s) + \sum_{j=1}^N [e^{tQ}]_{rj} x_j(0), \\ x'_r(t) &= \int_0^t \sum_{j=1}^N [e^{(t-s)Q} QS]_{rj} d\xi_j(s) + \sum_{j=1}^N [e^{tQ} Q]_{rj} x_j(0), \\ &\dots\dots\dots \\ x_r^{(N-1)}(t) &= \int_0^t \sum_{j=1}^N [e^{(t-s)Q} Q^{N-1} S]_{rj} d\xi_j(s) + \sum_{j=1}^N [e^{tQ} Q^{N-1}]_{rj} x_j(0) \end{aligned}$$

and (in order that the derivatives can exist)

$$(4.6.4) \quad \begin{aligned} S_{rj} &= 0 \\ (QS)_{rj} &= 0 \\ \dots\dots\dots & \\ (Q^{N-2} S)_{rj} &= 0, \end{aligned} \quad j = 1, \dots, N.$$

Since  $Q$  satisfies its characteristic equation, say (4.5.15),

$$(4.6.5) \quad \int_{-\infty}^t [e^{(t-s)Q} Q^N S] d\xi(s) - \sum_{j=1}^N a_j \int_{-\infty}^t [Q^{N-j} e^{(t-s)Q}] d\xi(s) = 0.$$

This vector equation can be written (using only the  $r$ th coordinate) in the form

$$(4.6.6) \quad \int_{-\infty}^t \sum_{j=1}^N [e^{(t-s)Q} Q^N S]_{rj} d\xi_j(s) - a_1 x_r^{(N-1)}(t) - \dots - a_N x_r(t) = 0.$$

If  $\xi'_j(t)$  existed, the last equation in (4.6.3) could be differentiated to give

$$(4.6.7) \quad x_r^{(N)}(t) = \int_{-\infty}^t \sum_{j=1}^N [e^{(t-s)Q} Q^N S]_{rj} d\xi_j(s) + \sum_{j=1}^N [Q^{N-1} S]_{rj} \xi'_j(t)$$

and (4.6.7) would then become

$$(4.6.8) \quad x_r^{(N)}(t) - a_1 x_r^{(N-1)}(t) - \dots - a_N x_r(t) = \sum_{j=1}^N (Q^{N-1} S)_{rj} \xi'_j(t).$$

Now the process with variables

$$(4.6.9) \quad \left\{ \frac{1}{c} \sum_{j=1}^N (Q^{N-1} S)_{rj} \xi_j(t) \right\}$$



is a  $\xi$ -process, if  $c$  is chosen properly, unless the parenthesis in (4.6.9) vanishes for all  $j$ . In either case (iii) is proved.

*Proof of (iv).* If in (iii),  $x_r^{(N)}(t)$  exists, (4.6.3) can be augmented to include

$$(4.6.3') \quad x_r^{(N)}(t) = \int_{-\infty}^t \sum_{j=1}^N [e^{(t-s)Q} Q^N S]_{rj} d\xi_j(s) + \sum_{j=1}^N [e^{tQ} Q^N]_{rj} x_j(0)$$

and (4.6.4) now includes

$$(4.6.4') \quad (Q^{N-1}S)_{rj} = 0, \quad j = 1, \dots, N.$$

In this case the last term in (4.6.7) vanishes and (4.6.8), with zero on the right hand side, is strictly true.

*Proof of (v).* If  $x_1'(t), \dots, x_N'(t)$  exist,  $S$  must vanish and (4.6.3) yields

$$(4.6.10) \quad x(t) = e^{tQ}x(0), \quad x^{(v)}(t) = Q^v x(t).$$

Thus the  $x(t)$  process is deterministic and  $x(t)$  has derivatives of all orders.

**THEOREM 4.7.** *Let  $\{x(t)\}$  be the variables of a one-dimensional t.h.G. process. The process is a component process of an  $N$ -dimensional t.h.G.M. process if and only if the chance variables*

$$(4.7.1) \quad x(0), \{E\{x(s), s \leq 0; x(t)\}\} \quad 0 < t < \infty$$

are linearly dependent on  $N$  variables.

Suppose that the  $x(t)$  process is a component process of an  $N$ -dimensional t.h.G.M.  $y(t)$  process:  $x(t) = y_r(t)$ , and let  $A(t)$  be the transition matrix function of the  $y(t)$  process. When if  $\epsilon > 0$  and if  $n$  is any integer, the difference

$$y[(n + 1)\epsilon] - A(\epsilon)y(n\epsilon)$$

is independent of every  $y(s)$  with  $s \leq n\epsilon$ , and therefore independent of every  $y(m\epsilon)$  with  $m \leq n$ . Hence the  $y(n\epsilon)$  process is a t.h.G.M. process (discrete parameter case). Equation (3.8.5) becomes, in this case, if  $n = 0$ ,

$$(4.7.2) \quad x[(N + \nu)\epsilon] - a_1^{(\nu)}x[(N - 1)\epsilon] - \dots - a_N^{(\nu)}x(0) = \eta_1^{(\nu)}(N + \nu)$$

where  $\eta_1^{(\nu)}(N + \nu)$  is not merely independent of the variables  $\dots, x(-\epsilon), x(0)$ , but is even independent of every  $x(s)$  with  $s \leq 0$ . It then follows, applying the operator  $E\{x(s), s \leq 0; \cdot\}$  to both sides of (4.7.2), that the variables in (4.7.1) are linearly dependent on  $N$  variables if  $t$  is restricted to be a multiple of  $\epsilon$ . Allowing  $\epsilon$  to run through the values

$$\left\{ \frac{1}{m!} \right\}, \quad m = 1, 2, \dots$$

it follows that the statement of the theorem is true if  $t$  is restricted to be rational. The proof will be complete when it is shown that the subject<sup>23</sup> variables for rational  $t$  are dense in the whole class in the sense that for any  $t$ , the expectation

$$(4.7.3) \quad \delta = E\{[E\{x(s), s \leq 0; x(t')\} - E\{x(s), s \leq 0; x(t)\}]^2\}$$

<sup>23</sup> Courtesy of U. S. Navy.

converges to 0 when  $t' \rightarrow t$ . In fact, using the Schwarz inequality

$$(4.7.4) \quad \begin{aligned} \delta &= E\{[E\{x(s), s \leq 0; x(t') - x(t)\}]^2\} \\ &\leq E\{E\{x(s), s \leq 0; [x(t') - x(t)]^2\}\} = E\{[x(t') - x(t)]^2\} \end{aligned}$$

and the basic continuity hypothesis (1.3.1) imposed on continuous processes is precisely that the last expectation converges to 0 when  $t' \rightarrow t$ .

Conversely suppose that the chance variables (4.7.1) are linearly dependent on  $N$  variables. It can be supposed that  $x(0)$  is one of these  $N$ . Let the others be those for which  $t = t_2, \dots, t_N$ , and define  $y_1(t), \dots, y_N(t)$  by

$$(4.7.5) \quad \begin{aligned} y_1(t) &= x(t) \\ y_j(t) &= E\{x(s), s \leq t; x(t + t_j)\} \quad j = 2, \dots, N. \end{aligned}$$

The  $y(t)$  process is obviously an  $N$ -dimensional t.h.G. process. Moreover

$$(4.7.6) \quad \begin{aligned} E\{y(s), s \leq 0; y_j(t)\} &= E\{x(s), s \leq 0; y_j(t)\} \\ &= E\{x(s), s \leq 0; x(t + t_j)\} \quad j = 1, \dots, N \end{aligned}$$

(where  $t_1$  is defined as 0). Since the right side is by hypothesis, for each  $j$ , a linear combination of  $y_1(0), \dots, y_N(0)$ , the  $y(t)$  process is a t.h.G.M. process, and the  $x(t)$  process is a component process, as was to be shown.

A detailed examination will now be made of t.h.G.M.<sub>N</sub> processes, and of the more general class of component processes of t.h.G.M. processes. The following theorem will be useful.

**THEOREM 4.8.** *Let  $\{x(t)\}$  be the variables determining a t.h.G. continuous parameter process. The process is a component process of an  $N$ -dimensional t.h.G.M. process if and only if for each  $\epsilon > 0$  the discrete parameter process with variables  $\{x(n\epsilon)\}$  is a component process of an  $N$ -dimensional t.h.G.M. process.*

If the  $x(t)$  process is a component process of an  $N$ -dimensional t.h.G.M.  $y(t)$  process, the  $x(n\epsilon)$  process is a component process of the  $N$ -dimensional t.h.G.M.  $y(n\epsilon)$  process. Conversely suppose that the  $x(n\epsilon)$  process is a component process of an  $N$ -dimensional t.h.G.M. process (which may depend on  $\epsilon$ ) for every  $\epsilon > 0$ . It follows that for each  $\epsilon > 0$  the chance variables

$$(4.8.1) \quad E\{\dots, x(-\epsilon), x(0); x(n\epsilon)\}, \quad n = 0, 1, \dots$$

are linearly dependent on  $N$  of their number. Hence the same is true of the following chance variables, if  $\nu, m$  are fixed and  $\nu > m$ :

$$(4.8.2) \quad E\{\dots, x(-1/\nu!), x(0); x(n/m)\} \quad n = 0, 1, \dots$$

According to the Corollary to Theorem 1.2, when  $\nu \rightarrow \infty$  the conditional expectations in (4.8.2) converge to

$$(4.8.3) \quad E\{x(s), s \leq 0, s \text{ rational}; x(n/m)\} \quad n = 0, 1, \dots$$

Hence the chance variables ( $t$  rational)

$$(4.8.4) \quad E\{x(s), s \leq 0, s \text{ rational}; x(t)\} = E\{x(s), s \leq 0; x(t)\}^{24}, \quad 0 < t < \infty,$$

are linearly dependent on  $N$  of their number. As in the proof of Theorem 4.7 it follows that the same is true if  $t$  runs through all positive real numbers, and according to Theorem 4.7, the  $x(t)$  process is therefore a component process of an  $N$ -dimensional t.h.G.M. process.

**THEOREM 4.9.** *Let  $\{x(t)\}$  be the variables of a one-dimensional continuous parameter t.h.G. process. The process is a component process of a finite-dimensional t.h.G.M. process if and only if the complex spectral function of the process is the sum of the integral of the square of the absolute value of a rational function of  $\lambda$  and of a mono'one non-decreasing function increasing only in a finite number of jumps.<sup>25</sup> Specifically:*

(i) *The process is a component process of an  $N$ -dimensional t.h.G.M. process if and only if the complex spectral function has the form*

$$(4.9.1) \quad G(\lambda) = \int_{-\infty}^{\lambda} \frac{|\beta_0(i\lambda)^{N-1} + \dots + \beta_{N-1}|^2}{|(i\lambda)^N + \alpha_1(i\lambda)^{N-1} + \dots + \alpha_N|^2} d\lambda + \hat{G}(\lambda)$$

where

- (a)  $\hat{G}(\lambda)$  is a monotone non-decreasing function satisfying (1.3.3) and increasing only in jumps, at no more than  $N$  points.
- (b) the denominator of the integrand vanishes at every discontinuity of  $\hat{G}(\lambda)$ , and the numerator vanishes at every zero of the denominator, to at least the same order;
- (c) the coefficients in the integrand are real, and the roots of the  $\lambda$  polynomials are all on the real axis or in the upper half plane.

*The integral vanishes identically if and only if the  $x(n)$  process is a component process of an  $N$ -dimensional deterministic process, and  $\hat{G}(\lambda)$  vanishes identically if and only if the variables  $\{x(t)\}$  vanish identically or the  $x(t)$  process is a component process of an  $N$ -dimensional t.h.G.M. process with no deterministic factor.*

(ii) *The process is a t.h.G.M.<sub>N</sub> process, in the deterministic case, if and only if the complex spectral function  $G(\lambda) = \hat{G}(\lambda)$  is a function increasing only in jumps, at no more than  $N$  points; non-deterministic case if and only if the complex spectral function has the form*

$$(4.9.2) \quad G(\lambda) = \int_{-\infty}^{\lambda} \frac{c d\lambda}{|(i\lambda)^N + \dots + \alpha_N|^2}.$$

<sup>24</sup> The equality (4.8.4) is proved as follows. Let  $t$  be fixed, and let  $x$  be the chance variable on the left. Then  $x(t) - x$  has mean 0 and is uncorrelated with every  $x(s)$  with  $s \leq 0$  and rational. It follows at once from the continuity of hypothesis (1.3.1) that then  $x(t) - x$  is uncorrelated with every  $x(s)$  with  $s \leq 0$ : it follows that (4.8.4) is true.

<sup>25</sup> It is easily seen that the first term of the two can also be described simply as the integral of a rational function of  $\lambda$ , which is non-negative for real  $\lambda$  and is integrable and an even function, like all complex spectral density functions.

Proof of (i). Suppose that the  $x(t)$  process is a one-dimensional component process of an  $N$ -dimensional t.h.G.M.  $y(t)$  process,  $x(t) = y_1(t)$ . It is no restriction to assume that the  $y(t)$  process is non-singular. Then the correlation function of the  $y(t)$  process is given by

$$\begin{aligned}
 (4.9.3) \quad R_y(t) &= R_y(0)e^{tQ^*} & t \geq 0 \\
 R_y(t) &= e^{-tQ} R_y(0) & t \leq 0,
 \end{aligned}$$

where  $Q$  is uniquely determined and

$$\begin{aligned}
 (4.9.4) \quad G(\lambda) - G(0) &= \left[ \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{it\lambda} - 1}{it} R_y(t) dt \right]_{11} \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{it\lambda} - 1}{it} [R_y(t)]_{11} dt,
 \end{aligned}$$

at the points of continuity of  $G(\lambda)$ .

The correlation function  $R_y(t)$  has derivatives of all orders for  $t > 0$ :

$$\begin{aligned}
 (4.9.5) \quad R_y^{(r)}(t) &= R_y(0)Q^{*r}e^{tQ^*} & t > 0 \\
 &= (-1)^r Q^r e^{-tQ} R_y(0) & t < 0.
 \end{aligned}$$

Suppose first that the  $y(t)$  process has no deterministic factor, in other words that it is non-degenerate and of type  $M$ . Then the characteristic values of  $Q$  have negative real parts and  $R(t) \rightarrow 0$  exponentially when  $|t| \rightarrow \infty$ . Hence  $G(\lambda)$  has a continuous derivative  $G'(\lambda)$ :

$$(4.9.6) \quad G'(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{it\lambda} [R_y(t)]_{11} dt.$$

Integrating by parts,

$$\begin{aligned}
 (4.9.7) \quad G'(\lambda) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{it\lambda}}{i\lambda} [R_y'(t)]_{11} dt \\
 &= \frac{R_y'(0+) - R_y'(0-)}{2\pi(i\lambda)^2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{it\lambda}}{(i\lambda)^2} R_y''(t) dt \\
 &= \frac{R_y'(0+) - R_y'(0-)}{2\pi(i\lambda)^2} - \frac{R_y''(0+) - R_y''(0-)}{2\pi(i\lambda)^3} - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{it\lambda}}{(i\lambda)^3} R_y'''(t), \\
 &\quad \dots \dots \dots \dots \dots \dots \dots
 \end{aligned}$$

Since  $Q$  satisfies its characteristic equation

$$(4.9.8) \quad \alpha^N - a_1\alpha^{N-1} - \dots - a_N = 0,$$

it follows that

$$\begin{aligned}
 (4.9.9) \quad R_y^{(N)}(t) - a_1 R_y^{(N-1)}(t) - \dots - a_N R_y(t) &= 0 & t > 0 \\
 R_y^{(N)}(t) + a_1 R_y^{(N-1)}(t) - \dots + (-1)^{N-1} R_y(t) &= 0 & t < 0
 \end{aligned}$$

and therefore if  $U$  is the operator  $\frac{d}{dt}$

$$(4.9.10) \quad [U^N - a_1U^{N-1} - \dots - a_NU^0][U^N + a_1U^{N-1} - \dots + (-1)^{N-1}U^0]R_y(t) = 0, \quad t \neq 0.$$

Applying (4.9.10) to (4.9.6)

$$(4.9.11) \quad \begin{aligned} &[(i\lambda)^N - a_1(i\lambda)^{N-1} - \dots - a_N][(i\lambda)^N + a_1(i\lambda)^{N-1} + \dots \\ &+ (-1)^{N-1}a_N]G'(\lambda) \\ &= |(i\lambda)^N - a_1(i\lambda)^{N-1} - \dots - a_N|^2 G'(\lambda) = P(i\lambda) \end{aligned}$$

where  $P(i\lambda)$  is a polynomial of degree  $2n - 2$ . Since  $P(i\lambda)$  is real and non-negative, when  $\lambda$  is real, the roots on the real axis are of even multiplicity and those off the axis are symmetric in the axis. Moreover  $P(i\lambda)$  is even, since the left side of (4.9.11) is even. It follows easily that  $P(i\lambda)$  can be written in the form

$$(4.9.12) \quad P(i\lambda) = \left| \sum_{j=0}^{N-1} \beta_j (i\lambda)^{N-j-1} \right|^2$$

where the roots of the  $\beta$  polynomial are all on or to the left of the imaginary axis. Finally

$$(4.9.13) \quad G'(\lambda) = \frac{|\beta_0(i\lambda)^{N-1} + \dots + \beta_{N-1}|^2}{|(i\lambda)^N - \dots - a_N|^2}.$$

The denominator polynomial in  $\lambda$  vanishes only at points where  $i\lambda$  has a negative real part, that is where  $\lambda$  has a positive imaginary part. This completes the proof in the case where the  $N$ -dimensional  $y(t)$  process has no deterministic factor. If there are such factors, it is easily verified that  $G(\lambda)$  has corresponding discontinuities and the above proof then applies to  $G(\lambda)$  less its jump function. The result can finally be summarized as in the statement of the theorem. If the  $y(t)$  process has only deterministic factors  $[R_y(t)]_{11}$  will be a sum of trigonometric functions and  $G(\lambda)$  will be a function of jumps.

Conversely suppose that the  $x(t)$  process has the complex spectral function (4.9.13). Then following the ideas of the proof of the analogous section of Theorem 3.9, it follows that  $R(t)$  satisfies the differential equation (cf. (3.9.10) and (3.9.11)).

$$(4.9.14) \quad R^{(N)}(t) - a_1R^{(N-1)}(t) - \dots - a_NR(t) = 0, \quad t > 0.$$

Any solution of (4.9.14) is a linear combination of (at most  $N$ ) functions

$$(4.9.15) \quad e^{\beta t}, te^{\beta t}, \dots$$

where  $\beta$  is a root of the equation

$$(4.9.16) \quad \alpha^N - a_1\alpha^{N-1} - \dots - a_N = 0$$

and where powers of  $t$  may appear if  $\beta$  is a multiple root. Let  $\epsilon$  be a positive number. The discrete parameter process determined by the variables  $\{x(n\epsilon)\}$

has correlation function  $R(n\epsilon)$ . This function is a linear combination of function

$$(4.9.15') \quad (e^{\beta\epsilon})^n, n(e^{\beta\epsilon})^n, \dots$$

corresponding to those of (4.9.15). There is an equation

$$(4.9.17) \quad \alpha^N - a_1(\epsilon)\alpha^{N-1} - \dots - a_N(\epsilon) = 0$$

with the  $\{e^{\beta\epsilon}\}$  as roots, of the same multiplicity as that of  $\beta$  in (4.9.16). Hence

$$(4.9.18) \quad R[(n + N)\epsilon] - a_1(\epsilon)R[(n + N - 1)\epsilon] - \dots - a_N(\epsilon)R(n\epsilon) = 0, \quad n \geq 0.$$

According to Theorem 3.10 the  $x(n\epsilon)$  discrete parameter process is therefore a component process of an  $N$ -dimensional discrete parameter t.h.G.M. process. Since this is true for all  $\epsilon$ , the  $x(t)$  process is a component process of an  $N$ -dimensional continuous parameter t.h.G.M. process.

If the integral vanishes identically, the non-deterministic factors in the  $N$ -dimensional process are irrelevant to the  $x(t)$  process and can be replaced by factors of type  $M(0)$ . If on the other hand the spectral function is continuous, the deterministic factors are irrelevant and can be replaced by factors of type  $M$ .

*Proof of (ii).* Since the t.h.G.M. $_N$  processes are characterized among the component processes of  $N$ -dimensional t.h.G.M. processes by the fact that the first  $N - 1$  derived process exist, their spectral functions (according to Theorem 1.4) are characterized by the property that

$$\int_{-\infty}^{\infty} \lambda^{2(N-1)} dG(\lambda) < \infty$$

that is the numerator in (4.9.1) must be identically constant. If this constant is not 0,  $\hat{G}(\lambda)$  can have no jumps, since each jump corresponds to a zero of numerator and denominator. Hence  $G(\lambda)$  is either identically  $\hat{G}(\lambda)$  or is in the form (4.9.2). The two possibilities obviously correspond to the deterministic and non-deterministic cases, respectively.

**COROLLARY.** *The t.h.G.M. $_N$  one dimensional process which is the solution of (4.4.1) has complex spectral function*

$$(4.9.19) \quad \int_{-\infty}^{\lambda} \frac{c^2 d\lambda}{|(i\lambda)^N - a_1(i\lambda)^{N-1} - \dots - a_1|^2}.$$

In fact the complex spectral function has the form (4.9.2), where the coefficients in the polynomial are those of the differential equation for the correlation function  $R_{\nu}(t)$  in (4.9.9), that is the coefficients of the characteristic equation of the infinitesimal transition matrix  $Q$ , (cf. (4.4.8)). The evaluation (4.9.19) is also easily proved directly.

The analogues of Theorems 3.10 and 3.11 in the continuous parameter case are easy to prove and will be omitted.