PROCEEDINGS *of the* FIFTH BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY

Held at the Statistical Laboratory
University of California
June 21–July 18, 1965
and
December 27, 1965–January 7, 1966

with the support of
University of California
National Science Foundation
National Institutes of Health
Air Force Office of Scientific Research
Army Research Office
Office of Naval Research

VOLUME II

PART 1

CONTRIBUTIONS TO PROBABILITY THEORY

EDITED BY LUCIEN M. LE CAM
AND JERZY NEYMAN

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY AND LOS ANGELES
1967

OUTLINE OF SOME TOPICS IN LINEAR EXTRAPOLATION OF STATIONARY RANDOM PROCESSES

A. M. YAGLOM
ACADEMY OF SCIENCES OF THE USSR

1. Introduction

We shall consider the frequently met problem of linear extrapolation of the stationary random processes $x(s)$, $-\infty < s < \infty$, with $Ex(s) = 0$. The problem consists of finding that linear functional $\hat{x}(t;\tau)$, $\tau > 0$, of the values $x(s)$ for $s \leq t$ (extrapolation according to the entire past of the process) or for $t - T \leq s \leq t$ (extrapolation of a process given on a finite interval) which would give the best approximation to the random variable $x(t + \tau)$. "Best" here is intended in the sense of least-squares; that is, it is required of the functional $\hat{x}(t;\tau)$ that the mean-square prediction error

$$\sigma^2(\tau) = E|x(t + \tau) - \hat{x}(t;\tau)|^2 \tag{1}$$

takes on its minimum value.

A. N. Kolmogorov [1], [2] initiated the theory of linear least-squares extrapolation of stationary processes. This theory was developed further by M. G. Kreĭn [3], N. Wiener [4], K. Karhunen [5], and others. At present, it has achieved a significant degree of completion (see, for example, Doob [6], chapter XII, or Rozanov [7]). We may formulate the general solution of this problem in the following way.

Let us start from the spectral representation of the stationary stochastic process in the form

$$x(s) = \int_{-\infty}^{\infty} e^{is\lambda}\, dZ(\lambda) \tag{2}$$

where $Z(\lambda)$ is the stochastic measure on the $-\infty < \lambda < \infty$ axis. This measure is connected to the spectral function $F(\lambda)$ of the process $x(s)$ by the relationship

$$E\left\{\int_S dZ(\lambda) \cdot \int_{S_1} \overline{dZ(\lambda)}\right\} = \int_{S \cap S_1} dF(\lambda), \tag{3}$$

where the bar above the symbol signifies the complex conjugate. If $F'(\lambda)$ is zero on a set of nonzero Lebesgue measure, or if $F'(\lambda)$ is not zero almost everywhere but

$$\int_{-\infty}^{\infty} \frac{|\log F'(\lambda)|}{1 + \lambda^2}\, d\lambda = \infty, \tag{4}$$

then the best linear extrapolator $\hat{x}(t;\tau)$ agrees almost surely with $x(t+\tau)$; in other words, in this case $\sigma(\tau) = 0$ for all τ. If the integral in the left-hand side of (4) converges, then

$$\hat{x}(t;\tau) = \int_{-\infty}^{\infty} e^{it\lambda}\Phi_\tau(\lambda)\, dZ(\lambda) \tag{5}$$

where $\Phi_\tau(\lambda) = 0$ for $\lambda \in S$, and

$$\Phi_\tau(\lambda) = \frac{1}{2\pi\varphi(\lambda)} \int_0^{\infty} e^{-ip\lambda}\, dp \int_{-\infty}^{\infty} e^{i(p+\tau)u}\varphi(u)\, du \qquad \text{for} \quad \lambda \notin S. \tag{6}$$

Here S is a set of zero Lebesgue measure consisting of the discontinuities of $F(\lambda)$ and of the growth points of the singular component of $F(\lambda)$, and $\varphi(\lambda)$ is defined by the condition $\varphi(\lambda) = \lim_{\mu \downarrow 0} \varphi(\lambda - i\mu)$ for almost all λ, where

$$\varphi(w) = \exp\left\{\frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{1+\lambda w}{\lambda - w}\frac{\log F'(\lambda)}{1+\lambda^2}\, d\lambda\right\}. \tag{7}$$

The function $\varphi(w)$ is analytic and has no zeroes in the lower half-plane of complex variable w, and its boundary value on the real axis $\varphi(\lambda)$ satisfies the condition $|\varphi(\lambda)|^2 = F'(\lambda)$ almost everywhere.

The function $\Phi_\tau(\lambda)$ is called the spectral characteristic for linear extrapolation. When the analytic expression for this function is known, it is also usually possible to give an explicit expression for the best extrapolator $\hat{x}(t;\tau)$. In fact, from (5) and (2) we have

$$\hat{x}(t;\tau) = \int_0^{\infty} x(t-p)w(p)\, dp, \tag{8}$$

where $w(p)$ is the generalized function (a Schwartz distribution) which is the Fourier transformation of the function $\Phi_\tau(\lambda)$:

$$w(p) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ip\lambda}\Phi_\tau(\lambda)\, d\lambda, \qquad \Phi_\tau(\lambda) = \int_0^{\infty} e^{-ip\lambda}w(p)\, dp. \tag{9}$$

The mean-square extrapolation error is expressed in terms of the spectral characteristic of the extrapolation by using the formula

$$\sigma^2(\tau) = \int_{-\infty}^{\infty} |e^{i\tau\lambda} - \Phi_\tau(\lambda)|^2\, dF(\lambda) = E|x(t)|^2 - \int_{-\infty}^{\infty} |\Phi_\tau(\lambda)|^2\, dF(\lambda). \tag{10}$$

In a number of cases the function $\Phi_\tau(\lambda)$ may also be found without using the complicated formulas (6) and (7). Thus, for example, in the case of an absolutely continuous spectral function $F(\lambda)$, it is easy to show that *if there exists a function ψ of the real variable λ such that ψ* (a) *belongs to the space $L^2(dF)$ (has an integrable square modulus in the measure $F'(\lambda)\, d\lambda$)*, (b) *may be continued analytically in the lower half-plane so that there it will not have an order of growth higher than a power of $|\lambda|$, and* (c) *satisfies the condition that $[e^{i\tau\lambda} - \psi(\lambda)]F'(\lambda)$ may be continued analytically in the upper half-plane so that it will fall off not slower than a power of $|\lambda|$ at infinity, then ψ will indeed be the spectral characteristic* $\Phi_\tau(\lambda)$ (see [8]).

A general solution for the problem of the best least-squares linear extrapola-

tion of a stationary process $x(s)$ (such as given by formulas (5)–(7)) cannot be obtained by means of values on the finite interval $t - T \leq s \leq t$. However, some sufficient conditions similar to the conditions (a), (b), and (c) presented above, which permit the direct selection of the spectral extrapolation characteristic $\Phi_\tau(\lambda)$ in several special cases, may be formulated for this case too. For example, *let us suppose that the nondecreasing bounded function $F(\lambda)$ is absolutely continuous and that there exists a function ψ such that $\psi(a_T)$ belongs to $L^2(dF)$, (b_T) it is an entire function of complex variable λ of the form $\psi(\lambda) = \sum_{k=1}^{r} e^{-is_k\lambda}\psi_k(\lambda)$ where r is an integer, $0 \leq s_k \leq T$ for all k and all $\psi_k(\lambda)$ are rational functions, and (c_T) satisfies the condition that $[e^{i\tau\lambda} - \psi(\lambda)]F'(\lambda)$ may be represented in the form $\varphi_1(\lambda) + e^{-iT\lambda}\varphi_2(\lambda)$ where $\varphi_1(\lambda)$ may be continued analytically in the upper half-plane and $\varphi_2(\lambda)$ may be continued analytically in the lower half-plane so that both functions will fall off in the corresponding half-planes not slower than a power of $|\lambda|$. Then it is possible to show that $\psi(\lambda)$ will indeed be the spectral characteristic for the linear extrapolation of the stationary process $x(s)$ with the spectrum $F(\lambda)$ in terms of the values $x(s)$ on the interval $t - T \leq s \leq t$* (see [9]).

2. Explicit expressions for the best extrapolator

The general case of an arbitrary stationary process was considered in the Kolmogorov [1], [2] and Kreĭn [3] works on the theory of extrapolation. However, since it is impossible to give any uniquely defined "most natural" representation for the functional $\hat{x}(t;\tau)$ in the general case, the problem of finding this extrapolator was not even posed in the works mentioned above, and all attention was turned to finding an expression for the mean-square extrapolation error $\sigma^2(\tau)$ and, especially, to the clarification of the conditions under which $\sigma^2(\tau) = 0$ or, conversely, $\sigma^2(\tau) \neq 0$. Wiener's great contribution was that he was the first to direct attention to the possibility of obtaining very simple and convenient explicit expression for the best extrapolator $\hat{x}(t;\tau)$ in some particular cases. Namely, in his book [4] Wiener examined the case of stationary processes with an absolutely continuous spectral function $F(\lambda)$, and an everywhere positive rational spectral density $F'(\lambda)$. The general form of such a spectral density is given by

$$F'(\lambda) = B\frac{\left|\prod_{k=1}^{M}(\lambda - \beta_k)\right|^2}{\left|\prod_{j=1}^{N}(\lambda - \alpha_j)\right|^2}, \qquad -\infty < \lambda < \infty, \tag{11}$$

where $B > 0$, $N > M$ and the imaginary parts of all roots α_j and β_k are positive. It is easy to verify that in the case of (11) the function $\varphi(w)$ in formula (7) acquires the extremely simple form:

$$\varphi(w) = \sqrt{B}\frac{\prod_{k=1}^{M}(w - \beta_k)}{\prod_{j=1}^{N}(w - \alpha_j)}. \tag{12}$$

Substituting this expression into (6) (for the $S = 0$ case), Wiener obtained an explicit expression for $\Phi_\tau(\lambda)$ in the form

$$\Phi_\tau(\lambda) = \frac{\gamma_\tau(\lambda)}{\prod_{k=1}^{M}(\lambda - \beta_k)}, \qquad \gamma_\tau(\lambda) = \sum_{j=0}^{N-1} C_j \lambda^{N-j-1}, \tag{13}$$

where the coefficients $C_j = C_j(\tau)$ (dependent on τ) are determined from a simple algebraic system of N linear equations. The same result is obtained even more simply by starting from the sufficient conditions (a), (b), and (c) defining $\Phi_\tau(\lambda)$ mentioned at the end of section 1. The extrapolator $\hat{x}(t;\tau)$ of the form

$$\hat{x}(t;\tau) = \sum_{j=0}^{N-M-1} B_j x^{(j)}(t) + \sum_{k=1}^{M} B_{N-M-1+k} \int_0^\infty e^{i\beta_k p} x(t-p)\, dp, \tag{14}$$

where the coefficients $B_0, \cdots, B_{N-1}$ are linear combinations of the $C_0, \cdots, C_{N-1}$, corresponds to the spectral characteristic (13) under the condition that the β_k are different. When multiple roots $\beta_k = \beta_{k+1} = \cdots = \beta_{k+z}$ exist among the roots $\beta_1, \cdots, \beta_M$, the weight functions $e^{i\beta_k p}, \cdots, e^{i\beta_{k+z} p}$ in (14) must be replaced by $e^{i\beta_k p}, pe^{i\beta_k p}, \cdots, p^{z-1}e^{i\beta_k z}$. All these results are widely known at present and may be found in several advanced mathematical and engineering texts.

It is far less known, however, that there are many examples of processes with irrational spectral density $F'(\lambda)$ for which the explicit formula for the best linear extrapolator is no more complicated than in the case of a rational spectral density. Apparently, the author gave one of the first of such examples around ten years ago in [10]. The question of the least-square extrapolation of stochastic processes with spectral density of the form $F'(\lambda) = A\lambda^{-\alpha}$, $-\infty < \lambda < \infty$, was considered there. Clearly, such a function $F'(\lambda)$ may not be spectral density of a stationary stochastic process $x(s)$ since it is nonintegrable. Nevertheless, the function $F'(\lambda) = A\lambda^{-\alpha}$ for $A > 0$ and $\alpha > 1$ is the spectral density of a stochastic process $x(s)$ with stationary increments of some order, and the whole theory of the linear extrapolation of stationary stochastic processes is extended without difficulty to such processes. In particular, formula (6) is only slightly changed when applied to processes with stationary increments. It is furthermore easy to show that for $F'(\lambda) = A\lambda^{-\alpha}$ the function $\varphi(w)$ in (7) will be given by $\varphi(w) = \sqrt{A}w^{-\alpha/2}$ where $w = |w|e^{i\theta}$, $0 \geq \theta \geq -\pi$.

Substituting this value of $\varphi(w)$ in the appropriately modified formula (6), we find an analytic expression first for the spectral extrapolation characteristic $\Phi_\tau(\lambda)$, and then for the best extrapolator $\hat{x}(t;\tau)$. For example, if $1 < \alpha \leq 2$, it can thereby be shown that the best extrapolator here has the form

$$x(t;\tau) = \frac{\sin\frac{\pi\alpha}{2}}{\pi}\, \tau^{\alpha/2} \int_0^\infty \frac{x(t-p)}{p^{\alpha/2}(p+\tau)}\, dp; \tag{15}$$

and if $2 \leq \alpha < 3$, then

$$\hat{x}(t;\tau) = x(t) + \frac{\sin\frac{\pi\alpha}{2}}{\pi}\tau^{\alpha/2}\int_0^\infty \frac{x(t-p)-x(t)}{p^{\alpha/2}(p+\tau)}\,dp. \tag{15'}$$

In both these cases, the process $x(s)$ is a process with stationary first increments which has the structure function $E|x(t+s)-x(s)|^2$ of the form

$$D(t) = E|x(t+s)-x(s)|^2 = D|t|^{\alpha-1}, \qquad D = -4\Gamma(1-\alpha)\sin\frac{\pi\alpha}{2}. \tag{16}$$

Kolmogorov [11] first considered such stochastic processes; it later turned out that they play an essential part in the statistical theory of turbulent flows (see, for example, [12]).

The method applied in [10] may even be used to solve problems on the extrapolation of some stationary stochastic processes with irrational spectral density. For example, let the spectral density of a stationary random process $x(s)$ be

$$F'(\lambda) = A(\lambda^2+a^2)^{-\alpha/2}, \tag{17}$$

where $\alpha > 1$, $A > 0$, $a > 0$. In this case, the covariance function $B(t)$ of the process $x(s)$ is

$$\begin{aligned} B(t) &= Ex(t+s)x(s) = D|t|^{(\alpha-1)/2}K_{(\alpha-1)/2}(a|t|), \\ D &= \sqrt{\pi}2^{-(\alpha-3)/2}a^{-(\alpha-1)/2}[\Gamma(2\alpha)]^{-1}, \end{aligned} \tag{18}$$

where K_ν is the so-called Basset's function (the modified Bessel function of the second kind). The function $\varphi(w)$ in (7) has the form $\varphi(w) = \sqrt{A}(w^2+a^2)^{-\alpha/2}$ in this case, where the argument θ of the complex number $w^2+a^2 = (\lambda-i\mu)^2+a^2$ is assumed to satisfy the inequality $0 \geq \theta > -2\pi$. Moreover, repeating the reasoning in [10] which results in formulas (15) and (15'), we find that

$$\hat{x}(t;\tau) = \begin{cases} \dfrac{\sin\frac{\pi\alpha}{2}}{\pi}e^{-a\tau}\tau^{\alpha/2}\displaystyle\int_0^\infty \frac{e^{-ap}}{p^{\alpha/2}(p+\tau)}x(t-p)\,dp & \text{for } 1<\alpha<2, \\ \dfrac{\sin\frac{\pi\alpha}{2}}{\pi}e^{-a\tau}\tau^{\alpha/2}\displaystyle\int_0^\infty \left[\frac{e^{-ap}}{p+\tau}x(t-p)-\frac{1}{\tau}x(t)\right]\frac{dp}{p^{\alpha/2}} & \text{for } 2<\alpha<4 \end{cases} \tag{19}$$

(analogous formulas for $\hat{x}(t;\tau)$ may also be obtained for $\alpha > 4$). More complex results of the same kind, referring to the problem of the extrapolation of homogeneous and isotropic stochastic fields $x(t_1, t_2, \cdots, t_n)$ with a spectral density of the form (17), in terms of their values in the $t_n < 0$ half-space may be found in the Fortus work [13].

Still another class of stationary stochastic processes with irrational spectral density for which an explicit formula may be written for the best extrapolator

is the class of processes $x(s)$ with spectral density expressed in terms of polynomials in λ and trigonometric functions in λ as follows:

$$F'(\lambda) = B \frac{\left| \prod_{k=1}^{M} (\lambda - \beta_k) \prod_{m=1}^{K} (1 + b_m e^{-i\delta_m \lambda}) \right|^2}{\left| \prod_{j=1}^{N} (\lambda - \alpha_j) \prod_{n=1}^{L} (1 + a_n e^{-i\gamma_n \lambda}) \right|^2}, \tag{20}$$

where $B > 0$, $N > M$, the imaginary parts of all α_j and β_k are positive and a_n, b_m, γ_n, and δ_m are real numbers such that $\gamma_n > 0$, $\delta_m > 0$, $|a_n| < 1$, $|b_m| < 1$ for all n and m. It is easy to verify that in this case

$$\varphi(w) = \sqrt{B} \frac{\prod_{k=1}^{M} (w - \beta_k) \prod_{m=1}^{K} (1 + b_m e^{-i\delta_m w})}{\prod_{j=1}^{N} (w - \alpha_j) \prod_{n=1}^{L} (1 + a_n e^{-i\gamma_n w})}. \tag{21}$$

Substituting this expression for $\varphi(w)$ in (6) (with $S = 0$), we may obtain after some analytical manipulations an explicit formula for the spectral extrapolation characteristic $\Phi_\tau(\lambda)$ and then for the extrapolator $\hat{x}(t; \tau)$ also. The same result may be obtained more simply by direct selection of the function $\Phi_\tau(\lambda)$ satisfying the conditions (a), (b), and (c) mentioned at the end of section 1; that is, by using the method developed in my book [8] to solve problems on linear extrapolation for the case of a rational spectral density $F'(\lambda)$. Finally, it is also possible to use here the fact that under condition (20), the difference equation

$$\prod_{n=1}^{L} [x(s) + a_n x(s - \gamma_n)] = \prod_{m=1}^{K} [y(s) + b_m y(s - \delta_m)] \tag{22}$$

will have the solution $y(s)$, which is a stationary stochastic process with rational spectral density of the form (11), such that $H_x^-(t) = H_y^-(t)$ (here $H_x^-(t)$ denotes the linear span of the set of stochastic variables $x(s)$, $s \leq t$, which is closed relative to mean-square convergence) and

$$\prod_{n=1}^{L} [\hat{x}(t; \tau) + a_n \hat{x}(t; \tau - \gamma_n)] = \prod_{m=1}^{K} [\hat{y}(t; \tau) + b_m \hat{y}(t; \tau - \delta_m)] \tag{22'}$$

(the last approach has been recently developed in his candidate dissertation by S. Grigoryev at Kazan University for the cases $K = 1$, $L = 0$ and $K = 0$, $L = 1$). In the particular case where

$$F'(\lambda) = B \frac{|1 + be^{-i\delta\lambda}|^2}{|\lambda - i\alpha|^2} = B \frac{(1 + b^2) + 2b \cos \delta\lambda}{\lambda^2 + \alpha^2} \tag{23}$$

where B, b, α, and δ are real parameters, $B > 0$, $\alpha > 0$, $\delta > 0$, $|b| < 1$, each of the three methods we have described leads to the formula

$$\Phi_\tau(\lambda) = \begin{cases} \dfrac{e^{-\alpha\tau} + be^{-i(\delta - \tau)\lambda}}{1 + be^{-i\delta\lambda}} & \text{for } \tau < \delta, \\ \dfrac{e^{-\alpha\tau}(1 + be^{\alpha\delta})}{1 + be^{-i\delta\lambda}} & \text{for } \tau \geq \delta. \end{cases} \tag{24}$$

It follows that

$$\hat{x}(t;\tau) = \begin{cases} e^{-\alpha\tau}\sum_{k=0}^{\infty}(-1)^k b^k x(t-k\delta) - \sum_{k=1}^{\infty}(-1)^k b^k x(t+\tau-k\delta) & \text{for } \tau < \delta, \\ e^{-\alpha\tau}(1+be^{\alpha\delta})\sum_{k=0}^{\infty}(-1)^k b^k x(t-k\delta) & \text{for } \tau \geq \delta \end{cases} \tag{25}$$

(the result (25) for the case $b = -e^{-\alpha\delta}$ has been published by Grigoryev [14], who used a more artificial method in his paper). If, for example,

$$F'(\lambda) = \frac{B}{|\lambda - i\alpha|^2|1+ae^{-i\gamma\lambda}|^2} = \frac{B}{(\lambda^2+\alpha^2)(1+a^2+2a\cos\gamma\lambda)}, \tag{26}$$

where $B > 0$, $\alpha > 0$, $\gamma > 0$, a is real and $|a| < 1$, then

$$\Phi_\tau(\lambda) = \frac{e^{-\alpha\tau}[1-(-1)^r a^r e^{r\alpha\gamma}]}{1+ae^{\alpha\gamma}}(1+ae^{-i\gamma\lambda}) + (-1)^r a^r e^{-i(r\gamma-\tau)\lambda} \tag{27}$$

for $(r-1)\gamma \leq \tau < r\gamma$, which means that for such τ

$$\hat{x}(t;\tau) = \frac{e^{-\alpha\tau}[1-(-1)^r a^r e^{r\alpha\gamma}]}{1+ae^{\alpha\gamma}}[x(t)+ax(t-\gamma)] + (-1)^r a^r x(t+\tau-r\gamma). \tag{28}$$

As is seen from these examples for specific spectral densities of the form (20), explicit formulas for the best extrapolator turn out to be no more complex than for rational spectral densities of the form (11), which contain the same number of factors in the numerator and denominator. However, the form of the extrapolators in these cases differs considerably from the forms of the extrapolators for rational spectral densities.

Generally, it is considerably more difficult to find an explicit expression for the best extrapolator $\hat{x}(t;\tau)$ for the best least-square linear extrapolation in terms of the values $x(s)$ on the finite interval $t - T \leq s \leq t$ than for extrapolation in terms of its values on the half-axis $s \leq t$. However, in the particular case of a rational spectral density of the form (11) (where the numerator may even vanish; that is, the imaginary parts of some β's may be equal to zero), this expression may also be effectively determined (for example, by using direct selection of characteristics $\Phi_\tau(\lambda)$ satisfying the conditions (a_T), (b_T), and (c_T) of section 1 or by some other similar method; see, for example, [15], [7], [9]). It turns out that in this case the extrapolator $\hat{x}(t;\tau)$ has the form

$$\begin{aligned}\hat{x}(t;\tau) = {} & \sum_{k=0}^{W-M-1} B_k x^{(k)}(t) + \sum_{k=0}^{N-M-1} B_{N-M-k} x^{(k)}(t-T) \\ & + \sum_{k=1}^{M} B_{2N-2M-1+k}\int_0^T e^{i\beta_k p} x(t-p)\,dp \\ & + \sum_{k=1}^{M} B_{2N-M-1+k}\int_0^T e^{i\bar{\beta}_k p} x(t-p)\,dp,\end{aligned} \tag{29}$$

where $B_0, B_1, \cdots, B_{2N-1}$ are τ-dependent coefficients determined from some system of $2N$ linear equations (for simplicity, we consider all the roots $\beta_1, \cdots, \beta_M$ to be different). It is, however, essential that N of these linear equations be homogeneous equations not containing the parameter τ; hence, only N of the $2N$ coefficients $B_0, \cdots, B_{2N-1}$ are independent. Therefore, for any τ the best extrapolator $\hat{x}(t; \tau)$ may be represented as the sum of N definite linear combinations of values and derivatives of the process $x(s)$ at the points $s = t$ and $s = t - T$ and of integrals of $x(t - p)$, $0 < p < T$, with the weight functions $e^{i\beta_k p}$ and $e^{i\bar{\beta}_k p}$, where every combination is multiplied by some τ-dependent coefficients.

The conditions (a_T), (b_T), and (c_T) may also be applied to finding the explicit expression for the best extrapolator $\hat{x}(t; \tau)$ in terms of the values $x(s)$ for $t - T \leq s \leq t$ in the case of more general spectral densities of the form (20) (where the imaginary part of some β's may even be zero and some b's may be equal to $+1$ or -1). For special cases where either $K = 1$, $L = 0$, or $K = 0$, $L = 1$, the expression for $\hat{x}(t; \tau)$ has recently been obtained by this method by Grigoryev in his dissertation (the results for the spectral density (23) where $b = -e^{-\alpha\delta}$ were published in [14]). In the general case $K = 1$, $L = 0$ the best extrapolator for $\tau > \delta_1$ consists of the integral term and the linear combination of the values and derivatives of the process $x(s)$ in the points of the form $t - j\delta_1$ and $t - T + j\delta_1$, $j = 0, 1, \cdots$, belonging to the interval $[t - T, T]$; the extrapolator for $\tau < \delta_1$ contains additionally the values and the derivatives of the process in the points of the form $t + \tau - j\delta_1$, $j = 1, 2, \cdots$. In the case where $K = 0$, $L = 1$, the best extrapolator $\hat{x}(t; \tau)$ contains the integral term, the values, and the derivatives of the process at the points t, $t - T$, $t - \gamma_1$, $t - T + \gamma_1$, belonging to the interval $[t - T, t]$, and the value of the process at the point $t + \tau - r\gamma_1$, where $(r - 1)\gamma_1 \leq \tau < r\gamma_1$.

For the process with stationary increments having the structure function (16) and the spectral density $F'(\lambda) = A\lambda^{-\alpha}$, $1 < \alpha \leq 3$, it is also possible to obtain the explicit expression for the best extrapolator in terms of the values of the process on the interval $t - T \leq s \leq t$ (Kreĭn [16], [17], Grigoryev). According to Grigoryev, the extrapolator $\hat{x}(t; \tau)$ for $1 < \alpha < 2$ has the form

$$\hat{x}(t; \tau) = \frac{\sin \frac{\pi\alpha}{2}}{\pi} [\tau(T + \tau)]^{\alpha/2} \int_0^T \frac{x(t - p)}{[p(T - p)]^{\alpha/2}} \left[\frac{1}{p + \tau} - c\right] dp, \tag{30}$$

where $c = (2(\alpha - 1)/\alpha T)F(1, \alpha; 1 + \alpha/2; -\tau/T)$ and $F(a, b; c; z)$ is the usual symbol for the hypergeometric series.

3. Simplified linear extrapolators. The use of the decomposition of the random process into the principal components

The problem of finding the explicit expressions for the best linear extrapolator is an interesting, purely analytical problem. However, the solutions of the problem are rarely used in practice, as they are usually not simple enough.

Besides, even when the explicit expressions for the best extrapolator are used, they are often not the best in reality. The derivation of the expression for $\hat{x}(t;\tau)$ requires the knowledge of the precise form of the spectral (or covariance) function of the process, and the spectrum or covariance which is used in many cases is only an approximation to the precisely unknown or too complicated true function $F(\lambda)$ or $B(t)$.

From the theoretical point of view, the use of the best extrapolators corresponding to the approximate expression for the covariance function or the spectrum seems not to be justifiable. There are some special examples where the best extrapolation becomes meaningless (for example, as it contains nonexisting derivatives) or very far from being optimal after very small changes of the functions $B(t)$ and $F(\lambda)$.

(In the special case where it is known that the approximation $F_1(\lambda)$ to the true spectrum $F(\lambda)$ which is used has the property that the difference $F_1(\lambda) - F(\lambda)$ is itself the spectral function, the situation is simpler. In this case the best linear extrapolator $\hat{x}_1(t;\tau)$ which corresponds to the spectrum $F_1(\lambda)$ can obviously be applied to the process $x(s)$ with the spectrum $F(\lambda)$. It is also easy to show that if $F_1(\lambda) - F(\lambda)$ is a spectral function and $\max_\lambda [F_1(\lambda) - F(\lambda)]$ is small enough, the error of the extrapolator $\hat{x}_1(t;\tau)$ will be quite close to the error of the best linear extrapolator $\hat{x}(t;\tau)$ for all τ (see Rozanov [18]).)

However, in almost all practical applications the use of the best extrapolators corresponding to the rather rough approximations to covariance or spectrum functions as a rule leads only to a very small excess of root-mean square error of extrapolation over the root-mean square error of the true best linear extrapolator. But the excess of root-mean square error over its minimum value $\sigma(\tau)$ will also be usually very small for many linear extrapolators of different forms. Therefore, in many cases it is possible to fix beforehand a form of the extrapolator containing a few undetermined parameters and to select only the values of the parameters from the condition of minimization of mean square error. From this point of view the most interesting result of the theory of linear extrapolation is the evaluation of the minimum value of mean square error. The knowledge of this irremovable mean square error of extrapolation permits us to make sure that the selected simplified extrapolator cannot be significantly improved.

One of the simplest possible extrapolators is evidently the following:

$$\hat{x}(t;\tau) = a(\tau)x(t). \tag{31}$$

Root-mean square error of the extrapolator (31) will have the minimum value $\sigma_1(\tau) = \{B(0)[1 - B^2(\tau)/B^2(0)]\}^{1/2}$ when $a(\tau) = B(\tau)/B(0)$. In the case of the convex covariance function $B(t)$, the error $\sigma_1(\tau)$ can be compared with the root-mean square error $\sigma(\tau)$ of the best linear extrapolator with the help of Hájek's result [19]. According to this result, if $B(t)$ is a convex function, then $\sigma(\tau) \geq \{B(0)[1 - B(\tau)/B(0)]\}^{1/2}$. It follows that for the convex function $B(t)$ the error $\sigma_1(\tau)$ exceeds $\sigma(\tau)$ by no more than the factor $[1 + B(\tau)/B(0)]^{1/2}$ (that is, by no more than 50%). Hájek's estimation for $\sigma(\tau)$ is sharp (it is

attained exactly when $B(t) = \max\{1 - |t|, 0\}$); however, for many individual covariances it appears to be rather rough. As to the nonconvex covariances $B(t)$, there is no general estimation of the ratio $\sigma_1(\tau)/\sigma(\tau)$ (since it is possible that $\sigma(\tau) = 0$ and $\sigma_1(\tau) > 0$). Nevertheless, even for the nonconvex covariances met in applications, the value $\sigma_1(\tau)$ is often surprisingly near $\sigma(\tau)$. For example, if $B(t) = e^{-\alpha|t|}\cos \alpha t$, then $\max_\tau \sigma_1(\tau)/\sigma(\tau) \approx 1.01$ (that is, $\sigma_1(\tau)$ exceeds $\sigma(\tau)$ by no more than 1% for all τ). The ratio $\sigma_1(\tau)/\sigma(\tau)$ takes somewhat larger values in cases where the function $B(t)$ is twice differentiable, and the best extrapolator $\hat{x}(t;\tau)$ contains values of the derivatives of $x(s)$ at the point $s = t$. However, even in these cases the replacement of the best linear extrapolator by the best extrapolator of the form (31) has in many practical cases sufficient accuracy.

A still better approximation to the minimum value of the root-mean square error of extrapolation can be attained using two-term extrapolators of the form

$$\hat{x}(t;\tau) = a(\tau)x(t) + a_1(\tau)x(t - t_1). \tag{32}$$

When t_1 is fixed, the optimal values of the coefficients $a(\tau)$ and $a_1(\tau)$ are determined from the simple system of two linear regression equations. Determination of the optimal value of t_1 in equation (32) is a complicated mathematical problem having, in some cases, no solution. (For example, if $B(t) = Ce^{-\alpha|t|}(1 + \alpha|t|)$, then the root-mean square error of the extrapolator (32) will decrease with the decreasing of t_1 tending to the root-mean square error of the best linear extrapolator as $t_1 \to 0$.) However, by means of two or three tests, in almost all cases it is easy to select a value t_1 such that the root-mean square of the extrapolator (32) will exceed the root-mean square of the best linear extrapolator no more than by a few percent. If still greater accuracy is required, it is possible to use an extrapolator $\hat{x}(t;\tau)$ having the form of a linear combination of three values $x(s)$ at the points $s \leq t$.

Note that in the case of extrapolation of a multidimensional stationary random process (that is, of an homogeneous random field) the number of terms in the right-hand part of equation (32) necessary to attain accuracy of extrapolation close to the one of the best linear extrapolator appears to be markedly greater than in the one-dimensional case. For some special cases of extrapolation of a two-dimensional process $x(t_1, t_2)$ in terms of its values in the half-plane $t_2 \leq 0$, it was shown by Fortus [20] that a good approximation to the root-mean square error of the best linear extrapolator can be attained only by means of the linear combination of several known values of the process containing no less than ten terms.

One can also find in the scientific literature a great number of functionals different from linear combinations of some values $x(s)$ at $s \leq$ t used as extrapolators $\hat{x}(t;\tau)$. For example, Yudin suggested in [21] to extrapolate the process $x(s)$ with stationary increments and with structure function (16) by the mean arithmetical moving average of the form

$$\hat{x}_a(t;\tau) = \frac{1}{a}\int_0^a x(t - p)\,dp, \tag{33}$$

where the best value $a = a_{\text{opt}}$ is determined by the condition of minimization of mean square error $\sigma^2(\tau; a) = E|x(t + \tau) - \hat{x}_a(t; \tau)|^2$. He found that $a_{\text{opt}} = 0$ if $\alpha \geq 2$, so that the best extrapolator of the form (33) is the "inertial extrapolator" $\hat{x}(t; \tau) = x(t)$ if $\alpha \geq 2$. However, if $1 < \alpha < 2$, then the ratio a_{opt}/τ takes a finite value different from zero, and in this case, $\sigma(\tau; a_{\text{opt}})$ exceeds the root-mean square error of the best linear extrapolator (15) (found after the publication of Yudin's paper) by no more than 10%. The fact that extrapolator (33) is of no use when $\alpha > 2$ is the consequence of the negativeness of the correlation coefficient between $x(t + \tau) - x(t)$, $\tau > 0$, and $x(t - p) - x(t)$, $p > 0$, in these cases. It is clear from equation (15′) that when $2 < \alpha < 3$, it is much more reasonable to select an approximate extrapolator of the form

$$\hat{x}_a(t; \tau) = x(t) + \frac{1}{a}\int_0^a [x(t) - x(t - p)]\, dp \tag{34}$$

$$= 2x(t) - \frac{1}{a}\int_0^a x(t - p)\, dp.$$

If the mean square error of the best extrapolator (34) is again denoted as $\sigma^2(\tau; a_{\text{opt}})$, then $\sigma(\tau; a_{\text{opt}})$ will also be very close to the root-mean square error of the best linear extrapolator for $2 < \alpha < 3$.

Sometimes the exponentially weighted moving averages of the form

$$\hat{x}_a(t; \tau) = a\int_0^\infty e^{-ap} x(t - p)\, dp \tag{35}$$

are also used for extrapolation (see, for instance, Cox [22], where the time series with discrete time are studied). The extrapolator (35) is closely related to (33); in the case where the value $a = a_{\text{opt}}$ is determined from the root-mean square criterion, its root-mean square error will in many cases only be slightly in excess of the minimum value of such an error. For the cases when the extrapolator (35) is not good enough, Cox [22] suggested the use of an extrapolator of the form

$$\hat{x}_{a,b}(t; \tau) = bx(t) + a(1 - b)\int_0^\infty e^{-ap} x(t - p)\, dp. \tag{36}$$

The last extrapolator contains two parameters, a and b, the values of which can be determined by minimization of the mean square error.

All extrapolators (31)–(36) are linear combinations with variable coefficients of some fixed simple linear functionals of the past of the process. It is also possible to use a linear combination of functionals, selected not because of its simplicity but for particular theoretical reasons. For example, it seems reasonable to select the functionals involved in the extrapolator by a method based on the general analysis into principal components. The analysis was introduced by Hotelling [23] at the beginning of the 1930's for finite families of random variables and is, at present, the widely used method of multivariate statistical analysis (see, for example, Anderson [24], chapter 11). Its generalization to the case of the continuous family of random variables (to the part of the continuous

random process) was later obtained independently by several scientists (see [25] to [29]). The analysis begins by extracting the first principal component. The component is the normalized linear combination or the normalized linear functional of the given random variables having maximum variability (that is, maximum variance). The word normalized means that the sum of squares of the coefficients or the integral of the square of weight function is one. Then the second normalized linear combination or normalized linear functional is sought. It is uncorrelated with the first one and has maximum variance among all those which are uncorrelated with the first principal component, and so on. Dealing with the statistical problem concerned with the given family of random variables, it is natural to find the approximate solution which depends only on a few first principal components (supposing that the other components with small variability cannot change the solution significantly). During the last years this approach was often suggested for practical statistical extrapolation (see, for instance, Pugachev [30] and Lorenz [31]).

The principal components of the part of the stationary random process $x(s)$, $t - T \leq s \leq t$, with covariance function $B(t)$, are the Fourier coefficients of the process corresponding to the orthogonal set of eigenfunctions of the integral equation

$$\lambda \int_{t-T}^{t} B(s - s_1)\varphi(s_1)\, ds_1 = \varphi(s), \qquad t - T \leq s \leq t. \tag{37}$$

The variance of the component

$$W_k = \int_0^T x(t - p)\varphi_k(t - p)\, dp, \qquad \int_{t-T}^{t} |\varphi_k(s)|^2\, ds = 1, \tag{38}$$

is equal to λ_k^{-1}, where λ_k is the corresponding eigenvalue of the equation (37). The contribution of the principal component W_k to the best linear extrapolator $\hat{x}(t; \tau)$ is equal to

$$\hat{x}_k(t; \tau) = E[x(t + \tau)W_k] \cdot W_k = \int_{t-T}^{t} B(t + \tau - s_1)\varphi_k(s_1)\, ds_1 \cdot W_k. \tag{39}$$

The sum of all contributions $\hat{x}_k(t; \tau)$, $k = 1, 2, \cdots$, is evidently equal to the best linear extrapolator in terms of the values $x(s)$ for $t - T \leq s \leq t$ (cf. Grenander [32], p. 269). So it is natural to expect that the sum of a few first terms $\hat{x}_k(t; \tau)$, with smallest indices k corresponding to the smallest eigenvalues λ_k, will form a good approximation to the best extrapolator $\hat{x}(t; \tau)$.

However, the true situation does not coincide with the expected one. Let us consider the typical case of rational spectral density (11). It is possible to show that in this case the integral equation (37) is equivalent to the eigenvalue problem for the differential equation

$$\prod_{j=1}^{N} \left(-\frac{d^2}{ds^2} + \alpha_j^2 \right) \varphi(s) = 2\pi\lambda B \prod_{k=1}^{M} \left(-\frac{d^2}{ds^2} + \beta_k^2 \right) \varphi(s) \tag{40}$$

with the special boundary conditions at the points $s = t$ and $s = t - T$ (see [33]). This statement leads to the conclusion that the eigenvalues λ_k in the rational

spectral density case are the roots of some transcendental equation accessible to numerical analysis. The corresponding eigenfunctions $\varphi_k(s)$ have simple analytical expressions which involve the parameter λ_k. In the simplest case of the Ornstein-Uhlenbeck process with covariance function $B(t) = Ce^{-\alpha|t|}$, the transcendental equation for λ_k has the form

$$\frac{C\lambda - \alpha}{\sqrt{2C\alpha\lambda - \alpha^2}} \tan\sqrt{2C\alpha\lambda - \alpha^2}T = 1, \tag{41}$$

and the functions $\varphi_k(s)$ are proportional either to $\cos\sqrt{2C\alpha\lambda_k - \alpha^2}\,(s - t + T/2)$ or to $\sin\sqrt{2C\alpha\lambda_k - \alpha^2}\,(s - t + T/2)$. These results allow one to compute easily the root-mean square error $\sigma_1(\tau)$ of the extrapolator $\hat{x}_1(t;\tau)$ for the Ornstein-Uhlenbeck process $x(s)$, $t - T \leq s \leq t$. If, for example, $\alpha\tau = 0.1$, it appears that $\sigma_1(\tau) \approx 1.3\sigma(\tau)$ for $\alpha T = \frac{1}{4}$, $\sigma_1(\tau) \approx 1.75\sigma(\tau)$ for $\alpha T = 1$, and $\sigma_1(\tau) \approx 2.2\sigma(\tau)$ for $\alpha T = 3$, where $\sigma(\tau)$ is the root-mean square error of the best linear extrapolator $\hat{x}(t;\tau)$. Similarly, if $\alpha\tau = 0.2$, then $\sigma_1(\tau) \approx 1.15\sigma(\tau)$ for $\alpha T = \frac{1}{4}$, $\sigma_1(\tau) \approx 1.4\sigma(\tau)$ for $\alpha T = 1$, and $\sigma_1(\tau) \approx 1.6\sigma(\tau)$ for $\alpha T = 3$. Therefore, the extrapolator $\hat{x}_1(t;\tau)$ involving only the first principal component is, in this case, satisfactory for a short interval T (and a not too small τ) but very inaccurate for a long interval T. The next approximations $\sum_1^n \hat{x}_k(t;\tau)$, $n = 2, 3, \cdots$, behave the same way, and consequently, in order to obtain a good approximation to $\hat{x}(t;\tau)$ for a long enough T, it is necessary to use a large number of principal components W_k.

This phenomenon may be explained by the fact that the values $x(s)$ in the beginning and in the end of the interval $t - T \leq s \leq t$ contribute equally to the principal components, whereas the last known values of the process are much more important for the extrapolation than the earliest ones. It is also clear that the Ornstein-Uhlenbeck process is the least suitable for the extrapolation by means of principal components because all the information about the future of such a process is contained in its last known value $x(t)$. However, in all other cases the best extrapolator will also be dependent mainly on the values $x(s)$ in the neighborhood of the point $s = t$. Therefore, for the extrapolator determinated by a fixed number of the first principal components W_k, the accuracy of extrapolation must decrease when the length of the interval of known values $x(s)$, and consequently, the known information, is increasing. This proves that the application of the method of principal components to extrapolation problems with long intervals T is not advisable.

4. Theory of canonical correlations for stationary random processes

The decomposition into the principal components is not convenient for the extrapolation because it is based on the selection of the functionals containing maximum total information (that is, maximum variability), whereas only the specific information about the future values of the process is of interest for extrapolation. The method of statistical analysis being most suitable for the

study of interdependencies and interrelations between two families of random variables is the method of canonical correlations. Therefore, it is interesting to investigate the application of this method to statistical extrapolation. The theory of canonical correlations was developed in the middle of the 1930's independently by Hotelling [34] and by Oboukhov [35], [36] (see also Anderson [24], chapter 12).

According to the theory, the investigation of the interrelations of the families $\mathbf{x} = (x_1, \cdots, x_n)$ and $\mathbf{y} = (y_1, \cdots, y_m)$ begins by finding out the normalized linear combinations $U_1 = \sum_1^n \alpha_{i1}x_i$ and $V_1 = \sum_1^m \beta_{j1}y_j$ having maximum correlation coefficients ρ_1. Then the second linear combinations $U_2 = \sum_1^n \alpha_{i2}x_i$ and $V_2 = \sum_1^m \beta_{j2}y_j$ are sought. They are uncorrelated with the first ones and have maximum correlation coefficient ρ_2 among all those which are uncorrelated with the first ones, and so on. As a result one manages to select coordinate systems in the spaces of variables x's and y's such that all the components of the compound vector $(U_1, U_2, \cdots, U_n, V_1, V_2, \cdots, V_m)$ (where U_i and V_j are the components of $\mathbf{x}$ and $\mathbf{y}$ in a new coordinate system) appear to be pairwise uncorrelated with the exception of the pairs (U_i, V_i), $i = 1, 2, \cdots, l$ where $l \leq \min(n, m)$.

One can show that the canonical variables $U_k = \sum \alpha_{ik}x_i = \boldsymbol{\alpha}_k'\mathbf{x}$ and $V_k = \sum \beta_{jk}y_j = \boldsymbol{\beta}_k'\mathbf{y}$ and the canonical correlations $\rho_k = \lambda$ are determined by the following algebraic eigenvalue problem:

$$-\lambda \mathfrak{B}_{xx}\boldsymbol{\alpha} + \mathfrak{B}_{xy}\boldsymbol{\beta} = 0, \qquad \mathfrak{B}_{yx}\boldsymbol{\alpha} - \lambda \mathfrak{B}_{yy}\boldsymbol{\beta} = 0, \tag{42}$$

where $\mathfrak{B}_{xx}$, $\mathfrak{B}_{xy}$, $\mathfrak{B}_{yx}$, and $\mathfrak{B}_{yy}$ are the corresponding covariance matrices.

The method of obtaining the values $U_1, \cdots, U_l, V_1, \cdots, V_l$, and $\rho_1, \cdots, \rho_l$ can also be described purely geometrically. Let us consider the multidimensional space $H_{x,y}$ of all linear combinations $w = \sum_1^n \alpha_i x_i + \sum_1^m \beta_j y_j$ with the usual scalar product $(w_1, w_2) = Ew_1w_2$. Let $\mathcal{P}_x$ be the matrix of projection in $H_{x,y}$ on the linear subspace H_x consisting of linear combinations of the form $\sum_1^n \alpha_i x_i$, and let $\mathcal{P}_y$ be the matrix of projection to the subspace H_y of combinations $\sum_1^m \beta_j y_j$. In this case the correlations $\rho_1, \cdots, \rho_l$ will coincide with the nonzero eigenvalues of the matrix $\mathfrak{B}_x = \mathcal{P}_x\mathcal{P}_y\mathcal{P}_x$ (or the matrix $\mathfrak{B}_y = \mathcal{P}_y\mathcal{P}_x\mathcal{P}_y$). The variables $U_1, \cdots, U_l$ and $V_1, \cdots, V_l$ will be eigenvectors of the matrices $\mathfrak{B}_x$ and $\mathfrak{B}_y$ corresponding to the eigenvalues $\rho_1, \cdots, \rho_l$.

It is clear that in the case where the variables $(x_1, \cdots, x_n, y_1, \cdots, y_m)$ have a multivariate Gaussian distribution, all the information about the vector $\mathbf{x}$ contained in the vector $\mathbf{y}$ is fully characterized by the values of the canonical correlations $\rho_1, \cdots, \rho_l$. Using the known Shannon's formula it is easy to calculate that in the case considered the amount of information about $\mathbf{y}$ contained in $\mathbf{x}$ is equal to $-(1/2)\sum_1^l \log(1 - \rho_k^2)$ (cf. [33]). In the course of evaluating the amount of information about a Gaussian random process contained in another random process, the theory of canonical correlations was generalized by Gelfand and Yaglom [33] to the case of two infinite families of random variables (that is to the case of two random processes $\{x(s), s \in S\}$ and $\{y(t), t \in T\}$). If S and T are two intervals of a real axis (which can coincide with each other),

the determination of canonical correlations and canonical variables for $\{x(s)\}$ and $\{y(t)\}$ can formally be reduced to the solution of the eigenvalue problem (related to (42))

$$-\lambda \int_T B_{xx}(t, t')\varphi(t')\,dt' + \int_S B_{xy}(t, s')\psi(s')\,ds' = 0, \qquad t \in T, \tag{43}$$

$$\int_T B_{yx}(s, t')\varphi(t')\,dt' - \lambda \int_S B_{yy}(s, s')\psi(s')\,ds' = 0, \qquad s \in S,$$

where B_{xx}, B_{xy}, B_{yx}, and B_{yy} are the covariance functions and the cross-covariance functions. However, the eigenfunctions of the problem, as a rule, are generalized functions (for example, they can contain the δ-function and its derivatives; cf. the similar situation in the paper [15] devoted to extrapolation and filtering). Therefore, the mathematically rigorous presentation of the theory of canonical correlations for random processes can be developed more easily by basing it on the geometrical interpretation of the theory. This interpretation can be extended to an infinite dimensional case without any changes with the exception of the fact that the matrices $\mathcal{P}_x$, $\mathcal{P}_y$ and $\mathcal{B}_x$, $\mathcal{B}_y$ turn out to be operators in the Hilbert space (see [33], [37] and related purely geometrical papers [38], [39]).

The papers [33], [40] deal with the case where S and T are the same interval of the time axis $-\infty < s < \infty$ and where $y(s) = x(s) + z(s)$, with $x(s)$ and $z(s)$ mutually uncorrelated stationary random processes with rational spectral densities. Under the additional assumption that the spectral density of $z(s)$ falls off at infinity faster than the spectral density of $x(s)$, the evaluation of the canonical correlations for this case can be reduced first to some eigenvalue problem for a linear differential operator with constant coefficients and then to the solution of some transcendental equation containing exponential and trigonometric functions. The number of nonzero canonical correlations ρ_k in this case is infinite.

For the theory of extrapolation of stationary random processes, another case is clearly more interesting. This is when $x(s)$ and $y(t)$ are the same stationary random process, but the sets S and T are different: S is the past (that is, either the semiaxis $s \leq t$ or the finite interval $t - T \leq s \leq t$), and T is the future (that is, either the semiaxis $s \geq t + \tau$ or the interval $t + \tau \leq s \leq t + \tau + T_1$, where $\tau \geq 0$). Such a theory of canonical correlations of two parts of the same stationary random process was considered in the paper [41] for the case where S is the semiaxis $s \leq t$ and T is the semiaxis $s \geq t + \tau$. Here the operator $\mathcal{P}_x$ of the projection of the future on the past is the operator which transforms $x(t + \tau_1)$, $\tau_1 \geq \tau$, in its best linear extrapolator $\hat{x}(t; \tau_1)$. The formulas (13) and (14) show that if the process $x(s)$ has rational spectral density of the form (11), the projection $\mathcal{P}_x H_y$ of the whole future into the whole past is a finite dimensional (namely N-dimensional) linear manifold. Consequently, the number of nonzero eigenvalues of the operators $\mathcal{B}_x$ and $\mathcal{B}_y$ in this case cannot be more than N. In the Gaussian case it follows from the above mentioned phenomenon that for a stationary process $x(s)$ with rational spectral density, all the information about the future contained in its past is concentrated in N special linear functionals $U_1, \cdots, U_N$ of the values $x(s)$, $s \leq t$.

The explicit evaluation of the canonical correlations ρ_k and canonical variables U_k, V_k for the rational spectral densities can be obtained with the help of a simple modification of the conditions (a), (b), and (c) mentioned in section 1 in connection with the problem of the best linear extrapolation. Let us suppose that the spectral density $F(\lambda)$ is absolutely continuous, and let us introduce the functions $\Phi_k^-(\lambda)$ and $\Phi_k^+(\lambda)$ determined by the relations

$$U_k = \int_{-\infty}^{\infty} e^{it\lambda}\Phi_k^-(\lambda)\,dF(\lambda), \qquad V_k = \int_{-\infty}^{\infty} e^{i(t+\tau)\lambda}\Phi_k^+(\lambda)\,dF(\lambda). \tag{44}$$

Then the considerations used in [8] for obtaining the sufficient conditions (a), (b), and (c) allow us to prove the following statement.

Assume that there exist functions $\psi^-(\lambda)$ and $\psi^+(\lambda)$ and a nonnegative number ρ such that:

(a′) *ψ^- and ψ^+ satisfy the conditions*

$$\int_{-\infty}^{\infty} |\psi^-(\lambda)|^2 F'(\lambda)\,d\lambda = \int_{-\infty}^{\infty} |\psi^+(\lambda)|^2 F'(\lambda)\,d\lambda = 1;$$

(b′) *the function $\psi^+(\lambda)$ may be continued analytically in the upper half-plane of the complex variable λ and $\psi^-(\lambda)$ may be continued analytically in the lower half-plane so that both functions will not have an order of growth higher than a power of $|\lambda|$; and*

(c′) *the function $[e^{i\tau\lambda}\psi^+(\lambda) - \rho\psi^-(\lambda)]F'(\lambda)$ may be continued analytically in the upper half-plane of λ, and the function $[e^{-i\tau\lambda}\psi^-(\lambda) - \rho\psi^+(\lambda)]F'(\lambda)$ may be continued analytically in the lower half-plane, so that both these functions will fall off not slower than a power of $|\lambda|$ at infinity.*

Then ψ^+ and ψ^- will be the functions $\Phi_k^+(\lambda)$ and $\Phi_k^-(\lambda)$ corresponding to canonical variables U_k and V_k and to the canonical correlation $\rho_k = \rho$ (see [41]).

If the spectral density $F'(\lambda)$ is rational and has the form (11), the stated conditions may be satisfied by functions $\psi^-(\lambda)$ and $\psi^+(\lambda)$ of the form

$$\psi^-(\lambda) = \frac{\gamma^-(\lambda)}{\prod_{j=1}^{m}(\lambda - \beta_j)}, \qquad \psi^+(\lambda) = \frac{\gamma^+(\lambda)}{\prod_{j=1}^{m}(\lambda - \bar{\beta}_j)}, \tag{45}$$

where $\gamma^-(\lambda)$ and $\gamma^+(\lambda)$ are polynomials of degree $N - 1$. Then the conditions (a′) and (b′) will be evidently fulfilled (after the normalization of the coefficients of $\gamma^-(\lambda)$ and $\gamma^+(\lambda)$). In order to satisfy the condition (c′) also it is necessary to select $\gamma^-(\lambda)$ and $\gamma^+(\lambda)$ in such a way that the functions

$$\frac{e^{i\tau\lambda}\prod_{j=1}^{M}(\lambda - \beta_j)\gamma^+(\lambda) - \rho\prod_{j=1}^{M}(\lambda - \bar{\beta}_j)\gamma^-(\lambda)}{\prod_{k=1}^{N}(\lambda - \alpha_k)},$$

$$\frac{e^{-i\tau\lambda}\prod_{j=1}^{M}(\lambda - \bar{\beta}_j)\gamma^-(\lambda) - \rho\prod_{j=1}^{M}(\lambda - \beta_j)\gamma^+(\lambda)}{\prod_{k=1}^{N}(\lambda - \bar{\alpha}_k)} \tag{46}$$

should be entire functions of λ.

The last condition leads to a system of $2N$ homogeneous linear equations for the $2N$ unknown coefficients of the polynomials $\gamma^-(\lambda)$ and $\gamma^+(\lambda)$ which contain ρ as a factor in certain terms. The condition of the existence of a nonzero solution of the system gives us the algebraic determinantal equation for ρ of degree $2N$. The equation has the roots $\rho_1, \cdots, \rho_N$ and $-\rho_1, \cdots, -\rho_N$. When the canonical correlations ρ_k have been determined, the coefficients of the polynomials $\gamma_k^-(\lambda)$ and $\gamma_k^+(\lambda)$ corresponding to the functions $\Phi_k^-(\lambda)$ and $\Phi_k^+(\lambda)$ can be found from the linear system involved in (c′) and the normalizing conditions (a′).

Similarly, one may treat the more general problem about the canonical correlations and canonical variables for two finite parts $\{x(s), t - T \leq s \leq t\}$ and $\{x(s'), t + \tau \leq s' \leq t + \tau + T_1\}$ of the stationary random process $x(s)$ with the rational spectral density (11). Here the operator $\mathcal{P}_x$ transforms the variables $x(s')$, $t + \tau \leq s' \leq t + \tau + T_1$, into the best linear extrapolators in terms of the values $x(s)$ for $t - T \leq s \leq t$. Since the space of all linear functionals of $x(s)$ where $t - T \leq s \leq t$ for $T < \infty$ is a subspace of the space of the linear functionals of the whole past of the process $x(s)$, it is evident that the space $\mathcal{P}_x H_y$ cannot be more than N-dimensional for $T < \infty$ (cf. equation (29) and the statement after it). It follows that for two arbitrary disjoint finite intervals of the process $x(s)$ there cannot exist more than N nonzero canonical correlations. These correlations and the corresponding canonical variables can be found with the help of the following modification of the conditions (a′), (b′), and (c′) mentioned in section 1.

Assume that there exist functions $\psi^-(\lambda)$ *and* $\psi^+(\lambda)$ *and a nonnegative number* ρ *such that*

$$(a'') \quad \int_{-\infty}^{\infty} |\psi^-(\lambda)|^2 F'(\lambda)\, d\lambda = \int_{-\infty}^{\infty} |\psi^+(\lambda)|^2 F'(\lambda)\, d\lambda = 1;$$

(b″) *the functions* $\psi^-(\lambda)$ *and* $\psi^+(\lambda)$ *are entire functions of* λ *represented in the form* $\psi^-(\lambda) = \psi_1^-(\lambda) + e^{-iT\lambda}\psi_2^-(\lambda)$ *and* $\psi^+(\lambda) = \psi_1^+(\lambda) + e^{iT_1\lambda}\psi_2^+(\lambda)$ *where the functions* ψ_1^-, ψ_2^-, ψ_1^+, *and* ψ_2^+ *are rational; and*

(c″) *the functions* $[e^{i\tau\lambda}\psi_1^+(\lambda) + e^{i(\tau+T_1)\lambda}\psi_2^+(\lambda) - \rho\psi_1^-(\lambda)]F'(\lambda)$ *and* $\psi_2^+(\lambda)F'(\lambda)$ *may be continued analytically in the upper half-plane of* λ, *and the functions* $[e^{-i\tau\lambda}\psi_1^-(\lambda) + e^{-i(\tau+T)}\psi_2^-(\lambda) - \rho\psi_1^+(\lambda)]F'(\lambda)$ *and* $\psi_2^-(\lambda)F'(\lambda)$ *may be continued analytically in the lower half-plane so that all the functions will fall off not slower than a power of* $|\lambda|$ *at infinity in the corresponding half-planes.*

Then the functions $\psi^-(\lambda)$ *and* $\psi^+(\lambda)$ *will be the functions* $\Phi_k^-(\lambda)$ *and* $\Phi_k^+(\lambda)$ *of the equations* (44), *which determine the canonical variables* U_k *and* V_k *(corresponding to the canonical correlation* $\rho = \rho_k$*) of the parts* $\{x(s), t - T \leq s \leq t\}$ *and* $\{x(s'), t + \tau \leq s' \leq t + \tau + T_1\}$ *of the process* $x(s)$ *with the spectral density* $F'(\lambda)$.

The proof of this statement is similar to the proof of conditions (a_T), (b_T), and (c_T) mentioned at the end of section 1. For the rational spectral density (11) the conditions may be satisfied by the functions of the form

$$(47) \qquad \psi_r^+(\lambda) = \frac{\gamma_r^+(\lambda)}{\left|\prod_{j=1}^{M}(\lambda - \beta_j)\right|^2}, \qquad \psi_r^-(\lambda) = \frac{\gamma_r^-(\lambda)}{\left|\prod_{j=1}^{M}(\lambda - \beta_j)\right|^2}, \qquad r = 1, 2,$$

where $\gamma_r^+(\lambda)$ and $\gamma_r^-(\lambda)$ are the polynomials of degree $N + M - 1$. The conditions (a''), (b''), and (c'') lead to a system of linear homogeneous equations for the coefficients of the polynomials $\gamma_r^+(\lambda)$ and $\gamma_r^-(\lambda)$. After eliminating some unknowns from the system, it is possible again to obtain the determinantal equation of degree $2N$ having the roots $\rho_1, \cdots, \rho_N, -\rho_1, \cdots, -\rho_N$. When the canonical correlations ρ_k are known, the coefficients of $\gamma_r^+(\lambda)$ and $\gamma_r^-(\lambda)$ can be easily obtained for every ρ_k from the system of linear equations and the normalizing conditions (a'').

The best linear extrapolator $\hat{x}(t;\tau)$ in terms of the values $x(s)$ for $s \leq t$ or $t - T \leq s \leq t$ can always be decomposed into the sum of contributions of different canonical variables U_k for the corresponding past values and the arbitrary part of the future which contain the point $t + \tau$ (for example, for the semiaxis $s \geq t + \tau$ or $s \geq t + \tau_0$, where $0 \leq \tau_0 \leq \tau$). Therefore,

$$\hat{x}(t;\tau) = \sum_k (Ex(t+\tau)U_k) \cdot U_k. \tag{48}$$

Usually in real situations the canonical correlations ρ_k are rapidly decreasing when the index k increases. Therefore, as a rule, the extrapolator can be approximated precisely by a few first terms in the right-hand part of (48). In the special case of the Ornstein-Uhlenbeck process, where the method of the principal component turns out to be ineffective for the purpose of extrapolation, the right-hand part of (48) contains only one term corresponding to $U_1 = x(t)$.

In the more general case of the arbitrary rational spectral density (11), the right-hand part of (48) contains a finite number (namely N) terms; however, all of them, with the exception of one or two first terms, are usually negligible. If we increase the length of the interval of the known past values of the process, the accuracy of the approximate extrapolator containing only the fixed number of right-hand terms in (48) will be increasing too. All these facts display the great advantages of the canonical variables in comparison with the principal components in studying the statistical extrapolation.

REFERENCES

[1] A. N. Kolmogorov, "Stationary sequences in Hilbert space," *Bull. Moscov. Gosud. Univ., Mat.*, Vol. 2, No. 6 (1941), 40 pp.

[2] ———, "Interpolation und Extrapolation von stationären zufälligen Folgen," *Izv. Akad. Nauk SSSR, Ser. Mat.*, Vol. 5 (1941), pp. 3–14.

[3] M. G. Kreĭn, "On a problem of extrapolation of A. N. Kolmogoroff," *Dokl. Akad. Nauk SSSR* (n.s.), Vol. 46 (1945), pp. 306–309.

[4] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, New York, Wiley, 1949.

[5] K. Karhunen, "Über die Struktur stationärer zufälliger Funktionen," *Ark. Mat.*, Vol. 1 (1950), pp. 141–160.

[6] J. L. Doob, *Stochastic Processes*, New York, Wiley, 1953.

[7] Yu. A. Rozanov, *Stationary Random Processes*, Moscow, Fizmatgiz, 1963. (English translation is in preparation by Holden-Day, San Francisco.)

[8] A. M. YAGLOM, *An Introduction to the Theory of Stationary Random Functions*, New York, Prentice-Hall, 1962.

[9] ———, "Extrapolation, interpolation and filtering of stationary random processes with rational spectral density," *Trudy Moscov. Mat. Obšč.*, Vol. 4 (1955), pp. 333–374. (English translation in the *Selected Transl. Math. Statist. and Probability*, Vol. 4 (1963), pp. 345–387.)

[10] ———, "Correlation theory of processes with stationary random increments of order n," *Mat. Sb.* (n.s.), Vol. 37 (1955), pp. 141–196. (English translation in the *Amer. Math. Soc. Transl., Ser. 2*, Vol. 8 (1958), pp. 87–141.)

[11] A. N. KOLMOGOROV, "Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum," *Dokl. Akad. Nauk SSSR* (n.s.), Vol. 26 (1940), pp. 115–118.

[12] A. M. OBOUKHOV and A. M. YAGLOM, "The microstructure of developed turbulence," *Trudy 3 Mat. Siezda SSSR*, Vol. 3 (1958), pp. 542–557. (Moscow)

[13] M. I. FORTUS, "Formulas for extrapolation of random fields," *Teor. Verojatnost. i Primenen.*, Vol. 7 (1962), pp. 104–113.

[14] S. V. GRIGORYEV, "Linear prediction for one class of stationary processes," *Izv. Vuzov. SSSR Mat.*, No. 4 (1964), pp. 56–60.

[15] L. A. ZADEH and R. RAGAZZINI, "Extension of Wiener's theory of prediction," *J. Appl. Phys.*, Vol. 21 (1950), pp. 645–655.

[16] M. G. KREĬN, "On the main approximation problem of the theory of extrapolation and filtering for stationary random processes," *Dokl, Akad. Nauk SSSR*, Vol. 94 (1954), pp. 13–16.

[17] ———, "On a method of the solution of reverse boundary problem," *Dokl. Akad. Nauk SSSR*, Vol. 94 (1954), pp. 987–990.

[18] YU. A. ROZANOV, "On the stability of solutions of linear problems for stationary processes," *Teor. Verojatnost. i Primenen.*, Vol. 9 (1964), pp. 528–530.

[19] J. HÁJEK, "Predicting a stationary process when the correlation function is convex," *Czechoslovak Math. J.*, Vol. 8 (1958), pp. 150–154.

[20] M. I. FORTUS, "On practical realization of statistical extrapolation of meteorological fields," *Izv. Akad. Nauk SSSR, Ser. Geophys.*, No. 6 (1962), pp. 795–803.

[21] M. I. YUDIN, "Some topics in the theory of meteorological fields," *Trudy Glavn. Geofiz. Obs.*, No. 19, Vol. 81 (1950), pp. 288–314.

[22] D. R. COX, "Prediction by exponentially weighted moving averages and related methods," *J. Roy. Statist. Soc. Ser. B*, Vol. 23 (1961), pp. 414–422.

[23] H. HOTELLING, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psych.*, Vol. 24 (1933), pp. 417–441; 498–520.

[24] T. W. ANDERSON, *An Introduction to Multivariate Statistical Analysis*, New York, Wiley, 1959.

[25] D. D. KOSAMBI, "Statistics in function space," *J. Indian Math. Soc.*, Vol. 7 (1943), pp. 76–88.

[26] K. KARHUNEN, "Zur Spektraltheorie stochastischer Prozesse," *Ann. Acad. Sci. Fennicae, Ser. A*, Vol. I, No. 34 (1946), pp. 3–7.

[27] V. S. PUGACHEV, "The general theory of correlation of random functions," *Izv. Akad. Nauk SSSR, Ser. Mat.*, Vol. 17 (1953), pp. 401–420.

[28] A. M. OBOUKHOV, "Statistical description of continuous fields," *Trudy Geophys. Inst. Akad. Nauk SSSR*, No. 24, Vol. 151 (1954), pp. 3–42.

[29] M. LOÈVE, *Probability Theory*, New York, Van Nostrand, 1955.

[30] V. S. PUGACHEV, *Theory of Random Functions and Their Application to Automatic Control Problems*, Moscow, Fizmatgiz, 1960. (English translation was published by Pergamon Press, London, 1965.)

[31] E. N. LORENZ, "Prospect for statistical weather forecasting," Final report statistical forecasting project, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1959.

[32] U. GRENANDER, "Stochastic processes and statistical inference," *Ark. Mat.*, Vol. 1 (1950), pp. 195–277.

[33] I. M. GELFAND and A. M. YAGLOM, "Evaluation of the amount of information about a random function which is contained in another random function," *Uspehi Mat. Nauk*, Vol. 12, No. 1 (1957), pp. 3–52. (English translation in the *Amer. Math. Soc. Transl. Ser. 2*, Vol. 12 (1959), pp. 199–246.)

[34] H. HOTELLING, "Relation between two sets of variates," *Biometrika*, Vol. 28 (1936), pp. 321–377.

[35] A. M. OBOUKHOV, "Normal correlation of vectors," *Izv. Akad. Nauk SSSR, Otd. Mat. i Estestv. Nauk*, No. 3 (1938), pp. 339–370.

[36] ———, "The theory of correlation of vectors," *Učen. Zap. Mosk. Gos. Univ., Mat.*, No. 45 (1940), pp. 73–92.

[37] E. J. HANNAN, "The general theory of canonical correlation and its relation to functional analysis," *J. Austral. Math. Soc.*, Vol. 2 (1961), pp. 229–242.

[38] M. G. KREĬN, M. A. KRASNOSELSKII, and D. P. MILMAN, "On the defect numbers of linear operators in Banach space and on some geometrical questions," *Trudy Inst. Mat. Akad. Nauk Ukraine SSR*, No. 11 (1948), pp. 97–112.

[39] C. DAVIS, "Separation of two linear subspaces," *Acta Sci. Math. (Szeged)*, Vol. 19 (1958), pp. 172–187.

[40] A. M. YAGLOM, "The explicit formulas for the extrapolation, filtering and the evaluation of the amount of information in the theory of Gaussian stochastic processes," *Trans. Second Prague Conf. Information Theory, Statistical Decision Functions, Random Processes*, Prague, 1960, pp. 251–262.

[41] ———, "Stationary Gaussian processes satisfying the strong mixing condition and best predictable functionals," *Bernoulli, Bayes, Laplace Berkeley Seminar 1963*, Heidelberg, Springer, 1965, pp. 241–252.