ARKIV FÖR MATEMATIK Band 2 nr 8

Communicated 10 October 1951 by HARALD CRAMÉR and HANNES ALFVÉN

Some statistical problems in the theory of servomechanisms

By MAURITZ SUNDSTRÖM

With 12 figures in the text

Contents

Introduction

The statistical problems dealt with here are more or less discrete, but in one respect there is a main thread between them, as they are all concerned in a wide sense with questions about servomechanisms. Servo technics are to a great extent a matter of measuring, and these measurements are, of course, subjected to a great many random errors. It will be our task to study these errors for some important servotechnic problems. At the same time we will pay some attention to errors committed through the use of approximation formulas. Further, in servo circuits there exists a great deal of noise which will have some influence on the signals going through the circuits. Such questions will also be dealt with here. Finally a couple of problems in the information theory will be touched on.

By a servomechanism according to I. A. GETTING (Theory of Servomechanisms; Radiation Laboratory Series 25) is meant "a combination of elements for the control of a source of power in which the output of the system, or some function of the output is fed back for comparison with the input, and the difference between these quantities is used in controlling the power". This definition will be adopted here. Suppose we have to study the behaviour of an aircraft. This aircraft may be driven from its right course by a gust of wind the moment of which may be $M(t)$ where t denotes the time. This moment (the input) will cause an angular deviation (the output), say $\alpha(t)$, of the airplane. The principle task of the servomechanism is now to reduce $\alpha(t)$ to zero. As long as $\alpha(t)$ is not equal to zero, the actual value of $\alpha(t)$ will affect the motion of the rudders in order to bring down the effect of $M(t)$.

The reader is assumed to have at least an elementary knowledge of servo technics, and thus only a few definitions will be mentioned here.

Suppose that we have to study a servo cycle with the input $f_i(t)$ and the output $f_0(t)$ where t is an independent variable, generally the time. The relation between these two quantities and eventually other auxiliary quantities may often be described mathematically by a system of differential or integro-differential equations, e.g.

$$g_\nu(f_i, f_i', \ldots, f_0, f_0', \ldots; t) = 0 \qquad \nu = 1, 2, \ldots \tag{1}$$

the number of which must be equal to the number of variables minus two (t and f_i).

When the system (1) is linear in the dependent variables and their derivatives and has constant coefficients, this being a very common case, the equations are generally solved by the introduction of Laplace transforms. We define the Laplace transform of $f(t)$ as

$$F(s) = \int_0^\infty e^{-st} f(t)\, dt, \tag{2}$$

s being a complex quantity. We then obtain

$$F_0(s) = Y(s) \cdot F_i(s) \tag{3}$$

where $Y(s)$ (the transfer function) is assumed to be a rational function of s with the degree of the numerator not higher than that of the denominator. In the case of a stable system all poles of $Y(s)$ must lie in the left halfplane. On the imaginary axis we will sometimes admit of a simple pole in the origin.

For $s = j\omega$ $(j = \sqrt{-1})$ we write

$$Y(j\omega) = \mathrm{Re}\{Y(j\omega)\} + j\,\mathrm{Im}\{Y(j\omega)\} = \varrho\, e^{j\varphi} \tag{4}$$

where Re { } means the real and Im { } the imaginary part. When nothing special is said about it, we always assume that $s = j\omega$. As usual we call

$$\varrho = \varrho(\omega) = |Y(j\omega)| \tag{5}$$

the amplitude and

$$\varphi = \varphi(\omega) = \mathrm{arctg}\frac{\mathrm{Im}\{Y\}}{\mathrm{Re}\{Y\}} \tag{6}$$

the phase of the transfer function.

From the Laplace transform $F(s)$ we can obtain the corresponding time function by using the formula

$$f(t) = \frac{1}{2\pi j}\int_{b-j\infty}^{b+j\infty} e^{st} F(s)\, ds \tag{7}$$

where we have to choose b in such a way that the integral converges for all t for which $f(t)$ exists.

If the equations (1) are non-linear[1], the methods generally used in linear cases[1] will fail, and the treatment becomes much more complicated. In the sequel we will be concerned also with non-linear cases.

Before we attack our principal problems, we are going to make some simple general investigations about random errors (chapter I). The notations given there will be used throughout the whole paper.

[1] When talking about "linear equations" in servo technics one generally means linear equations with constant coefficients.

I. A few notes on errors

a) Various errors

As in all investigations based on an empiric material we have to consider two principal kinds of errors: 1) *Errors due to observations* and 2) *Errors due to approximations.*

The error of observation of a certain quantity y will be measured from the mean value (the probable value) of y, denoted by My, and will be represented by the standard deviation

$$Dy = \sqrt{M(y - My)^2}.$$

The error of approximation will be denoted by Ey.

b) Reduction of errors by smoothing

Suppose we have to study the relation between two variables x and y, x *being exactly given while* y *is assumed to be subjected to random errors of observation* (on the contrary, systematical errors are not considered here). According to the fundamental theorem of the calculus of errors the observations generally can be supposed to be normally distributed. Let the result of observation be plotted as points P_ν with the coordinates x_ν / y_ν in a coordinate system (fig. 1). *We presume that the different observations of* y *are independent of each other.* The footpoints $x_\nu / 0$ may be called F_ν. We now draw a smoothing curve

$$S = \overset{\frown}{Q_1 Q_2 \ldots Q_n}$$

with the most simple algebraical equation (a polynomial), using n points of observation $P_1, P_2, \ldots P_n$ $(x_1 < x_2 < \cdots < x_n)$ the footpoints of which do not occupy too large an interval, so that the standard deviations Dy_ν of the different observations may be regarded as approximately equal. Furthermore we require the sum of the squares of deviations

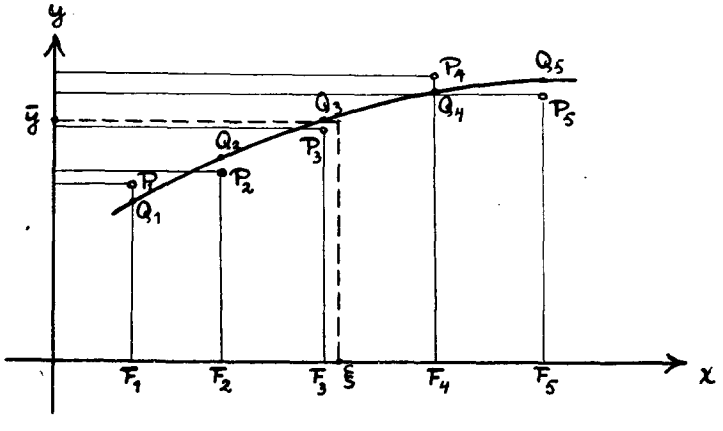

Fig. 1.

$$\sum \overline{P_\nu Q_\nu}^2$$

to be a minimum. We now ask for the remaining random errors of y after the smoothing.[1]

Under the above conditions the smoothing function, as we shall see, will be linear in $y_1, y_2, \ldots y_n$. The coefficients of this linear expression are polynomials of x of the same degree as that of the smoothing curve and with coefficients which are rational functions of $x_1, x_2, \ldots x_n$.

If the smoothing curve is

$$y = a_0 + a_1 x + \cdots + a_N x^N \quad (N < n),$$

the sum of the squares of deviations becomes

$$\sum_{\nu=1}^{n} (a_0 + a_1 x_\nu + \cdots + a_N x_\nu^N - y_\nu)^2.$$

The derivation with respect to a_μ gives

$$\sum_{\nu=1}^{n} x_\nu^\mu (a_0 + a_1 x_\nu + \cdots + a_N x_\nu^N - y_\nu) = 0 \quad (\mu = 0, 1, \ldots N).$$

Let us put

$$a_\mu = \sum_{\nu=1}^{n} x_\nu^\mu$$

$$A = \begin{vmatrix} a_0 & a_1 & \ldots & a_N \\ a_1 & a_2 & \ldots & a_{N+1} \\ \ldots & \ldots & \ldots & \ldots \\ a_N & a_{N+1} & \ldots & a_{2N} \end{vmatrix}$$

and let A_{ij} be the minor of the i-th row and the j-th column of this determinant. It follows

$$A\, a_\lambda = \sum_{\mu=0}^{N} A_{\mu+1,\lambda+1} \sum_{\nu=1}^{n} x_\nu^\mu y_\nu \quad (\lambda = 0, 1, \ldots N).$$

Hence we have the development

$$y(x) = \sum_{p=1}^{n} k_p(x) \cdot y_p \tag{8}$$

with the coefficients

$$k_p(x) = \sum_{\lambda=0}^{N} x^\lambda \sum_{\mu=0}^{N} \frac{A_{\mu+1,\lambda+1}}{A} x_p^\mu. \tag{9}$$

From this we obtain

$$\sum_{p=1}^{n} k_p(x) = \sum_{\lambda=0}^{N} x^\lambda \sum_{\mu=0}^{N} \frac{A_{\mu+1,\lambda+1}}{A} a_\mu.$$

[1] More precisely, we ask for the random errors under the assumption that the function used by the smoothing is the right one.

But

$$\sum_{\mu=0}^{N} \frac{A_{\mu+1,\lambda+1}}{A} \alpha_\mu = \begin{cases} 1 \text{ for } \lambda = 0, \\ 0 \text{ for } \lambda \neq 0. \end{cases}$$

Hence

$$\sum_{p=1}^{n} k_p(x) = 1. \tag{10}$$

Then $y(x)$ (formula 8) is a sort of modified weighted mean value[1] of the y_p:s with the weights k_p.

I will only give here the expression of $k_p(x)$ in the case of a linear smoothing curve. We then obtain

$$k_p(x) = \frac{\left(n x_p - \sum_{\nu=1}^{n} x_\nu\right) x + \sum_{\nu=1}^{n} x_\nu^2 - x_p \sum_{\nu=1}^{n} x_\nu}{n \sum_{\nu=1}^{n} x_\nu^2 - \left(\sum_{\nu=1}^{n} x_\nu\right)^2}. \tag{11}$$

This expression will sometimes be used in the following for numerical computations of correlation coefficients.

Let us represent the quantities $k_p(x)$ in a space of n dimensions (in fig. 2 $n = 3$). Hereby x is to be considered as a parameter. The points

$$\Pi_\nu \{k_p(x_\nu) \ (p = 1, 2, \ldots n)\}$$

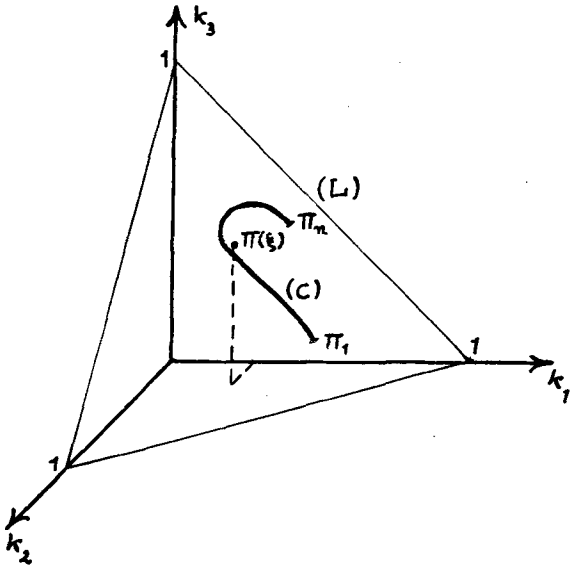

Fig. 2.

[1] Observe that $k_p(x)$ is not necessarily positive.

are lying in a plane (L) defined by (10). When x varies from x_1 to x_n the point Π_p moves along a curve (C) from Π_1 to Π_n in the plane L, while the point Q (fig. 1) follows the smoothing curve S from Q_1 to Q_n.

Of particular interest, as later will be shown, is the point $\Pi(\xi)$ defined by

$$k_1(\xi) = k_2(\xi) = \cdots = k_n(\xi) = \frac{1}{n} \tag{12}$$

which represents the average value $\bar{y} = \frac{1}{n}\sum y_p$.

If m_p and σ_p denote the theoretical mean value and standard deviation of any particular observation y_p, we have from (8) for the smoothed values $y(x)$ the mean value

$$My(x) = \sum_{p=1}^{n} k_p(x) \cdot m_p \tag{13}$$

and the standard deviation

$$Dy(x) = \sqrt{\sum_{p=1}^{n} [k_p(x)]^2 \sigma_p^2} \tag{14}$$

according to the assumption *that the different values are independent.*

In practical investigations we are often able *to consider* σ_p $(p = 1, 2, \ldots n)$ *as a constant*, and thus obtain

$$Dy(x) = \sigma \sqrt{\sum_{p=1}^{n} [k_p(x)]^2}. \tag{15}$$

The exact calculation of $\sum k_p^2(x)$ is, of course, a tedious procedure. In order to diminish the work of calculation, we can consider the value $x = \hat{x}$ *where the greatest error can be expected.* We then have

$$Dy(x) \leq \sigma \sqrt{\sum_{p=1}^{n} [k_p(\hat{x})]^2}. \tag{16}$$

A useful piece of information about the average deviation can be obtained by studying the projections on the y-axis of the results of observation (fig. 1). In the most common case, i.e. when the observations are normally distributed with the same standard deviation σ, the arithmetic mean $\bar{y}$ of the projections is also normally distributed with the mean value $M\bar{y} = \frac{1}{n}\sum m_p$ and the standard deviation

$$D\bar{y} = \frac{\sigma}{\sqrt{n}}. \tag{17}$$

Often we can be content with the last expression, especially if it is not possible to obtain a good estimation of σ.

c) Correlation between smoothed values

Between two neighbouring values $y(\xi_1)$ and $y(\xi_2)$ defined by formula (1) there is, in consequense of the smoothing, a certain correlation that may be characterized by the coefficient of correlation

$$r_{12} = \frac{M\{[y(\xi_1) - My(\xi_1)][y(\xi_2) - My(\xi_2)]\}}{Dy(\xi_1) \cdot Dy(\xi_2)}. \tag{18}$$

We see at once that $r \to 1$ when $\xi_1 \rightleftarrows \xi_2$.

If we assume that the standard deviations of all y_p are equal, we have

$$r_{12} = \frac{\sum_{p=1}^{n} k_p(\xi_1) \cdot k_p(\xi_2)}{\sqrt{\Sigma[k_p(\xi_1)]^2 \cdot \Sigma[k_p(\xi_2)]^2}}. \tag{19}$$

Finally, if we can choose ξ_2 in such a way that for this ξ-value (12) is valid with satisfactory approximation, (19) changes to

$$r_{12} \approx \frac{1}{\sqrt{n\,\Sigma[k_p(\xi_1)]^2}}.$$

In many problems of correlation it is enough to draw a linear smoothing curve using a few observations in the neighbourhood of each other.

d) Errors of derivatives

An important problem is to determine the error of the derivative $y' = \frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$ at an arbitrary point $P(x/y)$ of the smoothing curve $y = y(x)$. We take two points $P_1(\xi_1/\eta_1)$ and $P_2(\xi_2/\eta_2)$ in the vicinity of P, one of each side of P. Then

$$\frac{dy}{dx} = \lim_{\xi_2 \to \xi_1 \to x} \frac{\eta_2 - \eta_1}{\xi_2 - \xi_1}.$$

It follows

$$D^2\frac{dy}{dx} = \lim \frac{1}{(\xi_2 - \xi_1)^2}[D^2\eta_1 + D^2\eta_2 - 2\,r_{12}\,D\eta_1 \cdot D\eta_2] \tag{20}$$

where r_{12} denotes the coefficient of correlation between η_1 and η_2. The nearer the points P_1 and P_2 are lying to each other, the more r_{12} approaches to $+1$. The expression for r_{12} was given in the foregoing section (formula 18). If $D\eta_1 = D\eta_2 = \sigma$ we have

$$D\frac{dy}{dx} = \lim \frac{\sigma\sqrt{2}}{\xi_2 - \xi_1}\sqrt{1 - r_{12}}. \tag{21}$$

According to (8) the deviation of a smoothed value y from its real mean value can be written

$$\varepsilon = \sum_{p=1}^{n} k_p(x) \cdot \varepsilon_p,$$

ε_p being the deviations of the different observations. As the ε_p-values are independent of each other, we have, $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ being the deviations of η_1 and η_2,

$$M(\varepsilon^{(1)} \cdot \varepsilon^{(2)}) = r_{12} D\eta_1 D\eta_2 = M\left(\sum_{\substack{p=1\\q=1}}^{n} k_p(\xi_1) k_q(\xi_2) \varepsilon_p \varepsilon_q\right) = \sum_{p=1}^{n} k_p(\xi_1) \cdot k_p(\xi_2) D^2 y_p.$$

Further

$$D^2\eta_\nu = \sum_{p=1}^{n} [k_p[\xi_\nu)]^2 D^2 y_p \quad (\nu = 1, 2).$$

From (20) and the last two relations follows

$$D^2\frac{dy}{dx} = \lim_{\xi_2 \to \xi_1 \to x} \sum_{p=1}^{n} \left[\frac{k_p(\xi_2) - k_p(\xi_1)}{\xi_2 - \xi_1}\right]^2 D^2 y_p \tag{22}$$

i.e.

$$D^2\frac{dy}{dx} = \sum_{p=1}^{n} \left[\frac{d\,k_p(x)}{dx}\right]^2 D^2 y_p. \tag{23}$$

Of course this limit can also be found directly by taking the derivative of (8).

With the formula (23) the problem is theoretically solved, as soon as we know the smoothing function (8), since it is easy to put up an expression for the derivative of $k_p(x)$. But if we use a graphical method for determining the derivative $\frac{dy}{dx}$, we cannot go to the limit $\xi_2 - \xi_1 = 0$. Then there must be made a systematical error

$$E\frac{dy}{dx} = \frac{1}{2}(\xi_2 - \xi_1)\left(\frac{d^2y}{dx^2}\right)_{x=\xi_1+\vartheta(\xi_2-\xi_1)} \tag{24}$$

where ϑ is some number in the interval $0 < \vartheta < 1$. For the second derivative it is satisfactory to use a crude estimation.

To make it easier to determine the derivative from a given curve it is convenient to prolonge the corde between P_1 and P_2. This does not enlarge the probable error.

e) Errors of integrals

Now we are going to calculate random errors of integrals. Suppose that $y(x)$ is a random variable depending on a parameter x. Then the integral

$$I = \int_a^b y(x)\,dx = \lim_{\Delta x \to 0} \sum_{x_\nu = a}^{b-\Delta x} y(x_\nu)\,\Delta x$$

is a random variable too. For the sake of simplicity we write

$$I = \eta_1 \Delta x + \eta_2 \Delta x + \cdots + \eta_m \Delta x$$

where $\eta_1, \ldots \eta_m$ are a sort of average values in the corresponding intervals and $\Delta x = (b-a):m$.

If σ_μ denotes the standard deviation of η_μ and $r_{\mu\nu}$ the coefficient of correlation between η_μ and η_ν, we obtain

$$D^2 I = \sum_{\mu=1}^{m} \Delta x^2 \sigma_\mu^2 + \sum_{\mu \neq \nu} \Delta x^2 \sigma_\mu \sigma_\nu r_{\mu\nu}. \tag{25}$$

If we can assume *that the standard deviations of all* η_μ *are equal* $(= \sigma)$, there exists an average coefficient of correlation (r), so that

$$D^2 I = \frac{b-a}{\Delta x} \sigma^2 \Delta x^2 + \frac{b-a}{\Delta x}\left(\frac{b-a}{\Delta x} - 1\right) r \sigma^2 \Delta x^2$$

i.e.

$$D^2 I = \sigma^2 [(b-a)\Delta x + (b-a)(b-a-\Delta x) r].$$

If $\Delta x = \mathrm{o}\left(\frac{b-a}{1-r} r\right)$ we have approximately[1]

$$DI \approx \sigma (b-a) \sqrt{r}. \tag{26}$$

(A more exact method will be given in V:e).

When $y(x)$ is a smoothing function calculated with the method of least squares from n values $y_1, \ldots y_n$, we have according to (8)

$$I = \sum_{p=1}^{n} y_p \int_a^b k_p(x)\, dx$$

and

$$D^2 I = \sum_{p=1}^{n} D^2 y_p \left[\int_a^b k_p(x)\, dx\right]^2. \tag{27}$$

II. Experimental determination of transfer functions

a) General reasoning

In many questions of servo technics it is a difficult problem to determine the differential equations of the studied system, but then we often have the possibility to make an experimental investigation of the transfer function $Y(s)$ (formula 3), here supposed to be *at least approximately rational.*[2] But as it is

[1] Of course, in this case r must be > 0. The symbol o( ) means (according to Landau) "small in relation to".

[2] With this expression I intend to say that $Y(s)$ with a high degree of accuracy can be approximated by a rational function where the numerator and denominator have fairly low degrees.

much more simple to read off the amplitude ϱ than the phase φ, we wish to avoid experimental determination of the phase. As a matter of fact, when *all the roots of* $Y(s) = 0$ *have real parts* < 0, this being assumed here, there exists a simple relation between the phase and the amplitude.

Suppose that we have observed the amplitude ϱ from $\omega = 0$ to $\omega = \omega_m$ and wish to determine $\varphi = \varphi(\omega_c)$ in that interval. For the sake of brevity we write ϱ_c instead of $\varrho(\omega_c)$ and φ_c instead of $\varphi(\omega_c)$. We then have[1]

$$\varphi_c = \frac{2\,\omega_c}{\pi} \int_0^\infty \frac{\ln \varrho - \ln \varrho_c}{\omega^2 - \omega_c^2}\, d\omega \tag{28}$$

where ln denotes the natural logarithm.

Let us write

$$\varphi_c = \frac{2\,\omega_c}{\pi} \int_0^{\omega_m} \frac{\ln \varrho - \ln \varrho_c}{\omega^2 - \omega_c^2}\, d\omega + R_c$$

where, after the substitution $z = \frac{\omega_m}{\omega}$,[2]

$$R_c = -\frac{2\,z_c}{\pi} \int_0^1 \frac{\ln \varrho - \ln \varrho_c}{z^2 - z_c^2}\, dz. \tag{29}$$

I do not intend to say much about the first part of φ_c, its determination being very simple. On the other hand it is a rather sensitive matter to calculate the integral of R_c, as it often has a comparatively large numerical value which depends on the goodness of the observations in the vicinity of $\omega = \omega_m$.

Let us assume that the error of a simple observation of ϱ at $\omega = \omega_c$ is ε_c. Generally ε_c is normally distributed with the mean value 0. We call the standard deviation σ_c. Further we presume that for $z < 1$ $(\omega > \omega_m)$ ϱ can be written in the form

$$\varrho = |Y(j\omega)| = \varrho(z) = \varrho_m \cdot z^{\mu(z)} \tag{30}$$

where

$$\mu(z) = \mu_m + (z-1)\,\mu_m' + \frac{1}{2!}(z-1)^2\,\mu_m'' + \cdots + \frac{1}{(h-1)!}(z-1)^{h-1}\,\mu_m^{(h-1)} + (\delta\mu)_h. \tag{31}$$

It is easy to see that $\mu(z)$ gives the slope of the curve-vector $\overrightarrow{\ln \varrho - \ln \varrho_m}$ representing the relation between ϱ and z in a logarithmic scale. In most

[1] See H. W. Bode: "Network Analysis and Feedback Amplifier Design".

[2] The reason that I have put $z = \frac{\omega_m}{\omega}$ instead of $\frac{\omega}{\omega_m}$ is that I prefer to have the upper limit of the integral of R_c equal to 1 instead of equal to ∞.

practical problems the slope for decreasing z very quickly approaches to a constant, μ_∞, which in the case of a linear system must be a positive integer.

In order to investigate the convergence of the series (31) we put

$$\zeta = \frac{j\,\omega_m}{s} = \frac{j\,\omega_m}{\lambda + j\,\omega}.$$

The singular points of $\mu(\zeta)$ are those points ζ_ν for which $Y(s_\nu) = 0$ or $Y(s_\nu) = \infty$. These points must all be situated outside a circle with the centre $\zeta = 1$ and with a radius equal to 1. Thus

$$|\zeta_\nu - 1| > 1,$$

$$\left|\frac{-\lambda_\nu - j\,(\omega_\nu - \omega_m)}{\lambda_\nu + j\,\omega_\nu}\right| > 1.$$

$$\because\ \omega_m^2 - 2\,\omega_\nu\,\omega_m > 0,$$

$$\omega_m > 2\,\omega_\nu.$$

If the greatest value of ω_ν is denoted by $\widehat{\omega}$, we must have

$$\omega_m > 2\,\widehat{\omega}.$$

If $\omega_m > 2\,\widehat{\omega}$ the only critical point of $\mu(\zeta)$ could be $\zeta = 0$, but for this value μ is assumed to be finite ($Y(s)$ approximately rational).

Concerning the determination of the roots and zeros of $Y(s)$, I shall not deal with this problem here. I refer to the ordinary textbooks. I will however remark, that if $\ln \varrho$ as a function of $\ln z$ shows a tendency to tend to a straight line at $\omega = \omega_m$, we have strong reasons to believe that $\omega_m > 2\,\widehat{\omega}$.

It is evident that we can construct the amplitude curve for $\omega > \omega_m$ as soon as we know ϱ_m, μ_m and the variation of the slope for $\omega > \omega_m\,(z < 1)$ *presuming that* $\omega_m > 2\,\widehat{\omega}$. But this variation of the slope is uniquely determined by the values of μ_m', μ_m'', ... A necessary condition is of course that the series (31) converges in the actual region of z, but this must, as already mentioned, always be the case under the conditions that we have put up here.

If we put

$$I_{c\nu} = \frac{z_c}{\nu!}\int_0^1 (z-1)^\nu \frac{\ln z}{z^2 - z_c^2}\,dz \tag{32}$$

and

$$J_c = z_c \ln z_c \int_0^1 \frac{dz}{z^2 - z_c^2}, \tag{33}$$

we obtain

$$R_c = -\frac{2}{\pi}\left\{\sum_{\nu=0}^{\infty} I_{c\nu}\,\mu_m^{(\nu)} - J_c\,\mu_c\right\}. \tag{34}$$

Here μ_c is defined by

$$\mu_c = \frac{\ln \varrho_c - \ln \varrho_m}{\ln z_c}.$$

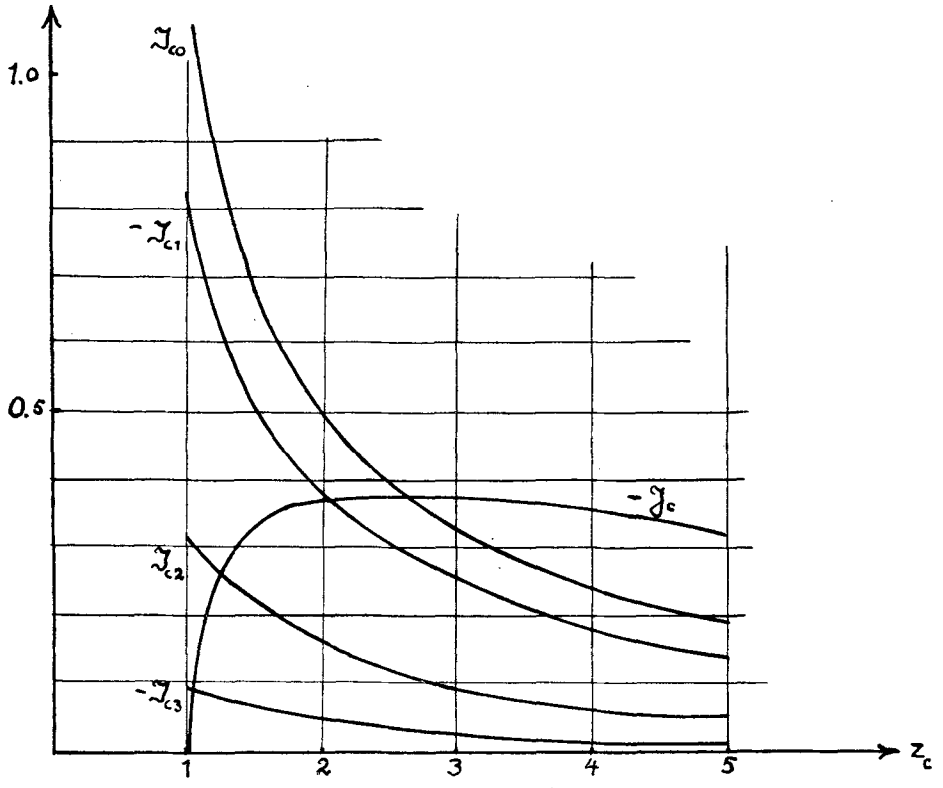

Fig. 3. I- and J-curves. For $z_c > 5$ use the approximation formulas.

In fig. 3 the curves representing $I_{c\nu}$ ($\nu = 0, 1, 2, 3$) and J_c are drawn. In order to obtain the same sign for all values I have there considered $(-1)^\nu I_{c\nu}$ instead of $I_{c\nu}$. The I-curves are very rapidly approaching to the z_c-axis when ν is increasing.

When z_c is not too small (> 5), we have the good approximation formulas:

$$I_{c0} \approx \frac{1}{z_c},$$

$$I_{c1} \approx -\frac{1}{z_c}\int_0^1 (z-1)\ln z\, dz = -\frac{1}{z_c}\sum_{\nu=1}^{\infty}\frac{1}{\nu(\nu+2)} = -\frac{3}{4z_c},$$

$$I_{c2} \approx -\frac{1}{2z_c}\int_0^1 (z-1)^2\ln z\, dz = +\frac{1}{2z_c}\sum_{\nu=1}^{\infty}\frac{1}{\nu(\nu+3)} = \frac{11}{36z_c},$$

.....

and

$$J_c \approx -\frac{\ln z_c}{z_c}.$$

The quantities μ_m, μ'_m, μ''_m, ... and μ_c are random variables, while I_{cv} and J_c are constants for given c and v. Hence R_c is also a random variable.

We now have to calculate the mean value and the error of R_c. We find directly

$$M R_c = -\frac{2}{\pi}\left\{\sum_{v=0}^{\infty} I_{cv} M \mu_m^{(v)} - J_c M \mu_c\right\}. \tag{35}$$

If ω_m is not too small, the series of formula (31) generally converges very quickly. Thus, in practical investigations, we generally need consider only a few terms. Very often two terms will be enough. If ${}_h\mu$ denotes the sum of the first h terms of (31), there exists for every arbitrarily small number η another number h, such that for every $h' > h$

$$|\mu - {}_{h'}\mu| < \eta.$$

According to (29) and (30) we have

$$R_c = -\frac{2 z_c}{\pi}\int_0^1 \frac{\mu(z) \ln z}{z^2 - z_c^2} dz + \frac{2 z_c}{\pi}\int_0^1 \frac{\mu_c \ln z_c}{z^2 - z_c^2} dz.$$

Then the corresponding error of R_c becomes, if μ_c and I_{c0} are exact,

$$|R_c - {}_h R_c| = E_\mu R_c < \frac{2}{\pi} I_{c0}\, \eta. \tag{36}$$

The total error of computation concerning R_c can be written

$$E R_c = E_\mu R_c + E_{\mu_c} R_c \tag{37}$$

where the last term arises from errors committed when estimating μ_c, the quantities I and J being considered as exact. The last error is often of so little importance that it can be omitted.

An exact computation of $D\,{}_h R_c$ is rather tedious, because the variables $\mu_m^{(v)}$ are not mutually independent. But from a practical point of view we could be satisfied with an approximation formula that gives a tolerably close upper limit.

Let $x = x_1 + x_2 + \cdots + x_h$ be a sum of mutually dependent variables. Then we always have, according to the inequality of Schwarz,

$$(x - Mx)^2 = \left[\sum_{v=1}^{h} (x_v - M x_v)\right]^2 \leq h \sum_{v=1}^{h} (x_v - M x_v)^2$$

and

$$D^2 x \leq h \sum_{v=1}^{h} D^2 x_v.$$

Thus we can conclude that[1]

[1] We have here assumed that there is no dependance between μ_c and $\{\mu_m^{(v)}\}$.

$$D^2{}_hR_c \leq \frac{4}{\pi^2}\left\{h\sum_{\nu=0}^{h-1} I_{c\nu}^2 \cdot D^2\mu_m^{(\nu)} + J_c^2 \cdot D^2\mu_c\right\}. \tag{38}$$

We now require a method to determine $D^2\mu_m^{(\nu)}$ ($\nu = 0, 1, \ldots h-1$). *Suppose that we can consider all observations of* ϱ *in the vicinity of* ϱ_m *as being normally distributed with the same standard deviation,* σ. Let $\bar{\varrho}$ be the mean value of ϱ and put

$$\varrho = \bar{\varrho} + \varepsilon.$$

(Of course, $\bar{\varrho}$ and ε are functions of z.) Then

$$\ln\varrho = \ln\bar{\varrho} + \ln\left(1 + \frac{\varepsilon}{\bar{\varrho}}\right) = \ln\bar{\varrho} + \frac{\varepsilon}{\bar{\varrho}} - \frac{1}{2}\left(\frac{\varepsilon}{\bar{\varrho}}\right)^2 + \frac{1}{3}\left(\frac{\varepsilon}{\bar{\varrho}}\right)^3 - \cdots,$$

$$M\ln\varrho = \ln\bar{\varrho} - \frac{1}{2}\left(\frac{\sigma}{\bar{\varrho}}\right)^2 - \frac{3}{4}\left(\frac{\sigma}{\bar{\varrho}}\right)^4 - \cdots,$$

$$(\ln\varrho - M\ln\varrho)^2 = \left(\frac{\varepsilon}{\bar{\varrho}}\right)^2 - \left(\frac{\varepsilon}{\bar{\varrho}}\right)^3 + \frac{11}{12}\left(\frac{\varepsilon}{\bar{\varrho}}\right)^4 - \cdots + \frac{1}{4}\left(\frac{\sigma}{\bar{\varrho}}\right)^4 + \frac{\varepsilon}{\bar{\varrho}}\left(\frac{\sigma}{\bar{\varrho}}\right)^2 - \frac{1}{2}\left(\frac{\varepsilon}{\bar{\varrho}}\right)^2\left(\frac{\sigma}{\bar{\varrho}}\right)^2 \cdots,$$

$$D^2\ln\varrho = M(\ln\varrho - M\ln\varrho)^2 = \left(\frac{\sigma}{\bar{\varrho}}\right)^2 + \frac{5}{2}\left(\frac{\sigma}{\bar{\varrho}}\right)^4 \cdots < K\left(\frac{\sigma}{\bar{\varrho}}\right)^2$$

where $K \approx$ and > 1, *provided that* $\frac{\sigma}{\bar{\varrho}}$ *is a small number* (a few per cents).

The last formula is applicable to a single observation of ϱ. But we have assumed that the ϱ-values are smoothed and that for the observation of ϱ we have used n different values. Thus we can write according to the approximation formula (17)

$$D\ln\varrho < \frac{K}{\sqrt{n}} \cdot \frac{\sigma}{\bar{\varrho}} \quad (K \approx \text{and} > 1). \tag{39}$$

If ϱ is a value in the neighbourhood of ϱ_m, there is, as a consequence of the smoothing, a strong correlation between $\ln\varrho$ and $\ln\varrho_m$. Let the coefficient of correlation be $r_{(0)}$ and put $z - z_m = \Delta z$. Then, according to (21) and (39),

$$D\mu_m = D\lim_{z\to 1}\frac{\ln\varrho - \ln\varrho_m}{\ln z} = \lim_{\Delta z\to 0} D\frac{\ln\varrho - \ln\varrho_m}{\Delta z} = \lim\frac{K_0}{\sqrt{n}} \cdot \frac{\sqrt{2}}{\Delta z}\sqrt{1 - r_{(0)}} \cdot \frac{\sigma}{\bar{\varrho}}. \tag{40}$$

As we cannot go to the limit, we have to add an E-error, which can be calculated by means of a formula similar to (24).

Further, if $r_{(1)}$ denotes the coefficient of correlation for the first derivatives of μ, we have

$$D\mu_m' = \lim_{\Delta z\to 0}\frac{K_1}{\sqrt{n}}\left(\frac{\sqrt{2}}{\Delta z}\right)^2\sqrt{(1 - r_{(0)})(1 - r_{(1)})} \cdot \frac{\sigma}{\bar{\varrho}}.$$

Generally we can write

$$D\mu_m^{(\nu)} = \lim_{\Delta z \to 0} \frac{K_\nu}{\sqrt{n}} \left(\frac{\sqrt{2}}{\Delta z}\right)^{\nu+1} \sqrt{(1-r_{(0)})\,(1-r_{(1)}) \ldots (1-r_{(\nu)})} \cdot \frac{\sigma}{\varrho}. \tag{41}$$

After having computed the standard deviations of μ_m and its derivatives we have to put the values into (38).

The total error of R_c,

$$DR_c + ER_c,$$

can be determined from the formulas (36–41).

To obtain the total error of φ_c we have to add the error committed by the calculation of

$$\frac{2\,\omega_c}{\pi} \int_0^{\omega_m} \frac{\ln \varrho - \ln \varrho_c}{\omega^2 - \omega_c^2}\, d\omega.$$

b) Numerical examples

1) I think the best way to check the method is to first study an easy example where the mathematical expression of the transfer function is known. We choose the transfer function

$$Y(s) = \frac{1}{(s+1)\,(s+2)}$$

and then have

Tab. 1.

ω	ϱ	$\ln \varrho$	φ
0	0.500	−0.693	−0
0.1	0.497	−0.699	−8°.6
0.2	0.488	−0.717	−17°.0
0.3	0.474	−0.747	−25°.2
0.4	0.455	−0.787	−33°.1
0.5	0.434	−0.835	−40°.6
0.6	0.411	−0.889	−47°.7
0.8	0.362	−1.016	−60°.5
1	0.316	−1.152	−71°.5
1.5	0.222	−1.505	−86°.8
2	0.158	−1.845	−108°.5
3	0.088	−2.430	−127°.9
4	0.054	−2.919	−139°.4
5	0.036	−3.324	−146°.9
$6 = \omega_m$	0.026	−3.650	−152°.1

$$\varrho(\omega) = \frac{1}{\sqrt{\omega^4 + 5\omega^2 + 4}},$$

$$\varphi(\omega) = \operatorname{arc\,tg} \frac{3\omega}{\omega^2 - 2}.$$

According to the reasoning on p. 150 we have in this case $\widehat{\omega} = 0$ and therefore there is no restriction concerning the choice of ω_m. Some values of ϱ and φ are given in tab. 1 where ω_m is put $= 6$.

Tab. 2 contains an outline of the calculation of φ_c for some values of ω_c. The values of μ_m and its first derivative ($\mu_m = 1.8$, $\mu'_m = -0.25$) are taken

Tab. 2.

Outline of the calculation of φ_c.

		$\omega_c=0.15$, $z_c=40$ $\ln \varrho_c = -0.709$ $\mu_c=0.80$		$\omega_c=0.45$, $z_c=13.33$ $\ln \varrho_c = -0.811$ $\mu_c=1.10$		$\omega_c=0.9$, $z_c=6.67$ $\ln \varrho_c = -1.084$ $\mu_c=1.35$	
ω	ω^2	$\mu_m=1.8$ $\mu'_m=-0.25$					
		Integrand		Integrand		Integrand	
0	0	−0.711		−0.583		−0.483	
0.1	0.01	−0.800		−0.582		−0.481	
0.2	0.04	−0.457		−0.578		−0.477	
0.3	0.09	−0.563		−0.569		−0.468	
0.4	0.16	−0.567	Σ (Simpson)	−0.565	Σ (Simpson)	−0.457	Σ (Simpson)
0.5	0.25	−0.554	$\Vert$	−0.505	$\Vert$	−0.445	$\Vert$
0.6	0.36	−0.533	−10.960	−0.495	−9.988	−0.433	−8.360
Integral →			−0.365		−0.333		−0.279
0.8	0.64	−0.497		−0.469		−0.400	
1	1	−0.453	−2.974	−0.428	−2.799	−0.358	−2.391
Integral →			−0.198		−0.187		−0.159
1.5	2.25	−0.357		−0.339		−0.292	
2	4	−0.286	−2.167	−0.272	−2.056	−0.239	−1.765
Integral →			−0.361		−0.343		−0.294
3	9	−0.192		−0.184		−0.164	
4	16	−0.138		−0.133		−0.121	
5	25	−0.105		−0.101		−0.093	
6	36	−0.074	−1.824	−0.079	−1.757	−0.073	−1.582
Integral →			−0.608		−0.586		−0.527
$\frac{2\omega_c}{\pi}\int_0^{\omega_m} \frac{\ln\varrho - \ln\varrho_c}{\omega^2-\omega_c^2}\,d\omega$		− 0.146		− 0.415		− 0.721	
$I_{c0}\,\mu_m$		+ 0.046		+ 0.138		+ 0.276	
$I_{c1}\,\mu'_m$		+ 0.006		+ 0.019		+ 0.037	
$-J_c\,\mu_c$		+ 0.076		+ 0.220		+ 0.391	
R_c		− 0.081		− 0.240		− 0.448	
φ_c		−13.0°		−37.5°		−66.9°	
φ_c (right value)		−12.9°		−36.9°		−66.2°	

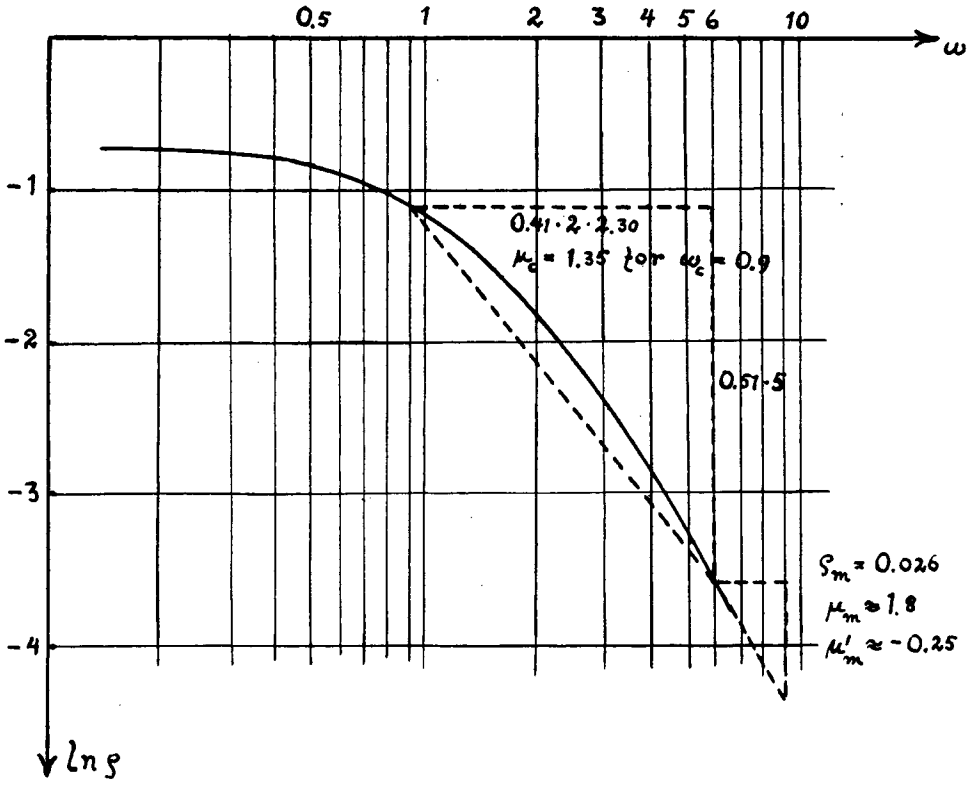

Fig. 4. Determination of μ and μ' for $\varrho = \frac{1}{\sqrt{\omega^4 + 5\omega^2 + 4}}$.

from the ϱ-curve (fig. 4). An exact calculation gives $\mu_m = 1.87$, $\mu'_m = -0.29$. The table shows a good agreement between the right φ-values and the corresponding values calculated by the method described here. The difference between these values for $\omega = \omega_c$ is what we have called $E\varphi_c$.

2) As a second example let us consider the curve of fig. 5 that gives the amplitudes of a certain regulator. The amplitudes are observed up to $\omega = 10$ ($= \omega_m$), but the observations for the highest frequencies do not seem to be reliable. This will not affect very much the phases corresponding to small ω, but for large ω the error, as will be shown, can be very important. In spite of that I have chosen this example, because it gives a good picture of the difficulties that arise in problems of this kind.

When drawing the smoothing curve we cannot avoid a certain degree of subjectivity, and especially not for large values of ω where the accuracy is small (ϱ seems here to be highly underestimated by the observator). In this way the smoothing becomes more a matter of feeling than a matter of logical reasoning. This holds true even more so concerning the calculation of coefficients of correlation (see below). If we could rely on the observations, but this is not the case here, we should determine these coefficients by means of the formulas (9) and (19).

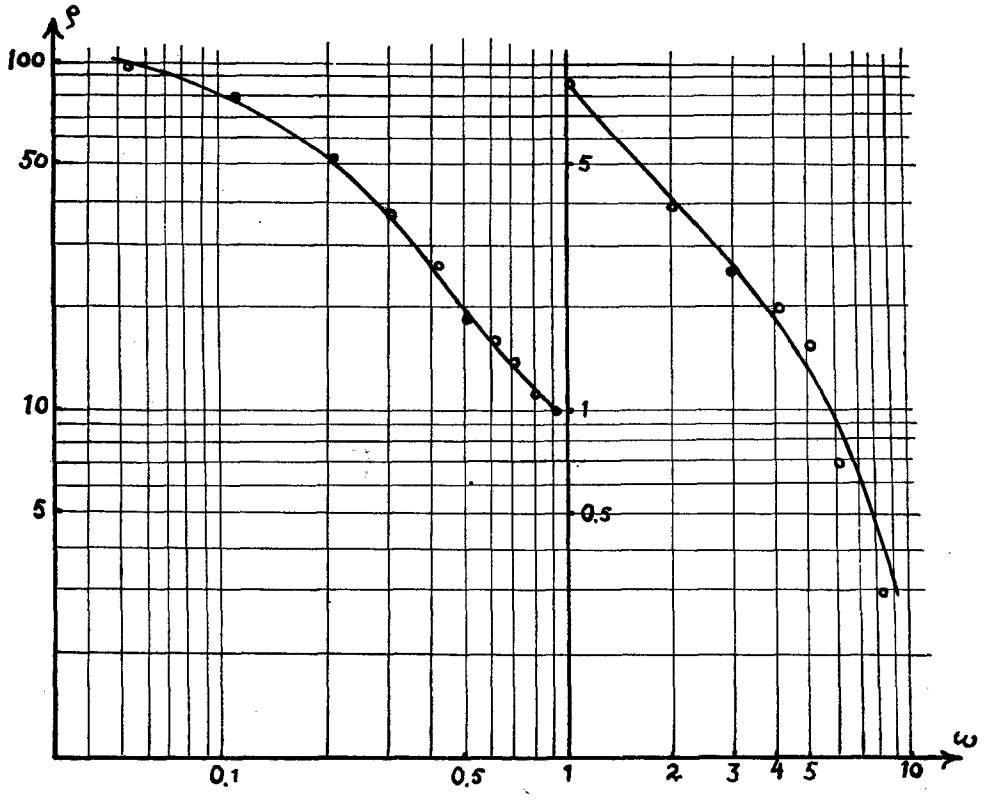

Fig. 5. Amplitudes of a certain regulator. Observed values o. Smoothing curve —.

For ϱ_m I have taken the value 0.16. Further I have put

$$\mu_m = \frac{\ln 0.27 - \ln 0.16}{\ln 10 - \ln 9} \approx 5.0 .$$

By the calculation of $D\mu_m$ according to formula (40) I have used the values:

$$n = 4,$$

$$\Delta z = \frac{10}{10} - \frac{10}{9} = -\frac{1}{9},$$

$$r = \frac{1}{2} \text{ (not computed; only a rough estimate),}$$

$$\frac{\sigma}{\bar{\varrho}} = 0.1$$

and thus obtained $D\mu_m \approx 0.5$. This value is possibly undervalued, because the quantity $\sigma/\bar{\varrho}$ may have been chosen too small. On the other hand r probably is > 0.5. As a comparison the values of $D\mu_c$ for $\omega_c = 0.15$ and $\omega_c = 3.5$ have been calculated with the following assumptions and results:

$$D^2\mu_c = D^2\left[\frac{\ln\varrho_c - \ln\varrho_m}{\ln z_c}\right] \leq \frac{1}{\ln^2 z_c}[D^2\ln\varrho_c + D^2\ln\varrho_m] \leqq$$

$$\leqq \frac{1}{n\ln^2 z_c}\left[K_c\left(\frac{\sigma_c}{\bar{\varrho}_c}\right)^2 + K_m\left(\frac{\sigma_m}{\bar{\varrho}_m}\right)^2\right] = \frac{K}{n\ln^2 z_c}\left[\left(\frac{\sigma_c}{\bar{\varrho}_c}\right)^2 + 0.01\right],$$

	$\omega_c = 0.15$	$\omega_c = 3.5$
μ_c	1.44	2.45
n	4	4
$\ln z_c$	4.29	1.06
$\sigma_c/\bar{\varrho}_c$	<0.05	0.1
$D\mu_c$	<0.02	<0.1.

Tab. 3.

Computation of φ_c for $\omega_c = 0.15$ and $\omega_c = 3.5$.

ω	ϱ	$\ln\varrho$	$\omega_c=0.15$ $z_c=66.7$ $\ln\varrho_c=4.20$ $\mu_c=1.44$			$\omega_c=3.5$ $z_c=2.86$ $\ln\varrho_c=0.74$ $\mu_c=2.45$		
			Integrand	Σ (Simpson)	= Integral	Integrand	Σ (Simpson)	= Integral
0	113	4.73	−22.7			−0.322		
0.1	82.5	4.41	−15.2			−0.296		
0.2	56	4.03	−10.9			−0.265		
0.3	38	3.64	−8.6			−0.234		
0.4	28	3.33	−6.5			−0.210		
0.5	21	3.04	−5.2			−0.188		
0.6	17	2.83	−4.1	−177.6	−5.92	−0.172	−4.316	−0.144
0.8	12	2.48	−2.80			−0.146		
1	9.5	2.25	−2.01	−17.31	−1.15	−0.130	−0.886	−0.059
1.5	5.7	1.74	−1.11			−0.095		
2	4.2	1.43	−0.70	−7.15	−1.19	−0.078	−0.588	−0.098
3	2.6	0.96	−0.36			−0.052		
4	1.9	0.64	−0.22			−0.040		
5	1.4	0.34	−0.16			−0.035		
6	1.00	0.00	−0.12			−0.033		
7	0.65	−0.43	−0.095			−0.033		
8	0.44	−0.82	−0.079			−0.031		
9	0.27	−1.31	−0.068			−0.031		
10	0.16	−1.83	−0.061	−4.33	−1.44	−0.030	−0.920	−0.307
$\frac{2\omega_c}{\pi}\int_0^{\omega_m}\frac{\ln\varrho-\ln\varrho_c}{\omega^2-\omega_c^2}d\omega$				−0.93			−1.35	
$I_{c0}\mu_m$				+0.08			+1.75	
$-I_{c1}\mu'_m$				0			0	
$-J_c\mu_c$				+0.09			+0.93	
R_c				−0.11			−1.71	
φ_c (calculated)				−60°			−175°	
φ_c (observed)				−58°			−190°	

I think there is no point in this case to study the derivatives of μ at $\omega = \omega_m$, not even the first one. On the other hand it would be valuable to know the limit of μ, when $\omega \to \infty$. But suppose we know nothing about the construction of the regulator. Then we have no possibility to estimate μ_∞ except from μ_m and its derivatives.

The calculation of φ_c has been carried out for $\omega_c = 0.15$ and $\omega_c = 3.5$. The result is given in tab. 3. We see from the table that for $\omega_c = 0.15$ there is a rather satisfactory accordance between the observed and calculated values of φ_c whereas for $\omega_c = 3.5$ the difference is considerable. But now, of course, the whole error is not to be referred to ϱ_c. The errors of the phases are at least as great as those of the amplitudes.

Let us estimate the error of φ_c for $\omega_c = 0.15$.

First we obtain (36 and 37)

$$\eta < 1,$$

$$E R_c < 0.01.$$

Further we have according to (38)

$$D_h R_c < 0.01.$$

Thus the total error of R_c is less than 0.02 (as a matter of fact this number is definitely too high). The corresponding error of φ_c becomes $\approx 1°$.

III. Determination of inverse Laplace transforms

a) General reasoning

Let a Laplace transform $F(s)$ be given (either as a mathematical function or as observations of amplitudes and phases) and suppose we have to compute the corresponding time function $f(t)$. This can be carried out by means of formula (7), i.e.

$$f(t) = \frac{1}{2\pi j} \int_{b-j\infty}^{b+j\infty} F(s)\, e^{st}\, ds.$$

We assume here that $F(s)$ *is a rational function (at least approximately) with no poles in the right half-plane or on the axis of imaginaries. Further the degree of the denominator is supposed to be higher than that of the numerator.* If there were a pole of the first order in the origin (other poles on the imaginary axis will not be considered according to the assumption of the introduction), we could separate the corresponding partial fraction $\frac{k}{s}$ where

$$k = \lim_{s \to 0} s F(s) = f(\infty)$$

from $F(s)$. Now the inverse transform of $\frac{k}{s}$ is k and consequently we have to apply a special method only to $F(s) - \frac{k}{s}$.

If $A(\omega)$ denotes the real part of $F(j\omega)$, it is easy to prove[1], $F(0)$ being $\neq \infty$, that

$$f(t) = \frac{2}{\pi}\int_0^\infty A(\omega)\cos\omega t\, d\omega. \tag{42}$$

A similar formula exists for the imaginary part $B(\omega)$ of $F(j\omega)$. In most cases $A(\omega)$ and $B(\omega)$ are random functions known by a set of observations in a limited frequency interval $(0, \omega_m)$.

A direct calculation of the integrals of (42) and especially their random errors is rather tedious, and therefore we transform the integrals before the numerical computation. Putting $\omega t = \zeta + 2\pi\nu$ $(\nu = 0, 1, \ldots)$ we obtain

$$f(t) = \frac{2}{\pi t}\sum_{\nu=0}^{\infty}\int_0^{2\pi} A\left(\frac{\zeta + 2\pi\nu}{t}\right) d\sin\zeta.$$

The substitution $\sin\zeta = \tau$ gives

$$\int_0^{2\pi} A\left(\frac{\zeta+2\pi\nu}{t}\right) d\sin\zeta = \int_0^1 A\left(\frac{\arcsin\tau+2\pi\nu}{t}\right) d\tau - \int_0^1 A\left(\frac{\pi-\arcsin\tau+2\pi\nu}{t}\right) d\tau -$$

$$- \int_0^1 A\left(\frac{\pi+\arcsin\tau+2\pi\nu}{t}\right) d\tau + \int_0^1 A\left(\frac{2\pi-\arcsin\tau+2\pi\nu}{t}\right) d\tau.$$

Using the following notations:

$$\lambda_0 = \frac{\pi}{t} \tag{43: a}$$

$$\Delta\lambda = \frac{\arcsin\tau}{t} \tag{43: b}$$

$$\left.\begin{aligned} \lambda_1 &= 2\nu\lambda_0 + \Delta\lambda \\ \lambda_2 &= (2\nu+1)\lambda_0 - \Delta\lambda \\ \lambda_3 &= (2\nu+1)\lambda_0 + \Delta\lambda \\ \lambda_4 &= (2\nu+2)\lambda_0 - \Delta\lambda \end{aligned}\right\} \tag{43: c}$$

$$\hat{A}_\nu(\tau; t) = A(\lambda_1) - A(\lambda_2) - A(\lambda_3) + A(\lambda_4) \tag{43: d}$$

$$u_\nu(t) = \int_0^1 \hat{A}_\nu(\tau; t)\, d\tau \tag{43: e}$$

[1] See "Regelungstheorie" by JOST HÄNNY.

we have

$$f(t) = \frac{2}{\pi t} \sum_{\nu=0}^{\infty} u_\nu(t). \tag{44}$$

In the case of F being a random variable we have further

$$Mf(t) = \frac{2}{\pi t} \sum_{\nu=0}^{\infty} M u_\nu(t) = \frac{2}{\pi t} \sum_{\nu=0}^{\infty} \int_0^1 M \hat{A}_\nu(\tau; t)\, d\tau. \tag{45}$$

The formula (44) is not convenient for the computation of $f(0)$ and $f(\infty)$. The value of $f(0)$ can of course be taken from (42), but often this calculation is unnecessary, as the determination can be made directly from the physical conditions of the problem. Further we have assumed that $F(s)$ has no pole for $s = 0$. Hence $f(\infty) = 0$.

In order to facilitate the calculation of $u_\nu(t)$ I have constructed a table concerning the variation of the λ-values a fragment of which is given in tab. 4. With the aid of this table and a pair of compassors it is easy to take the values of $\hat{A}$ from a curve of $A(\omega)$ and then evaluate the integrals by means of Simpsons formula. The error hereby committed is, according to the notations of section I a), $Du_\nu(t) + Eu_\nu(t)$. Of course, if we know the mathematical expression of $F(s)$, the error of observation disappears.

Before a discussion of the errors let us take into consideration two other questions.

1:o. If we do not know the value of $A(\omega)$ beyond a certain upper limit ω_m, we have a problem of the same kind as that treated in chapter II. Since $A = \mathrm{Re}\{F(s)\}$, it is evident that $\ln A$ has the same critical points as $\ln F(s)$. Let $\widehat{\omega}$ have the same meaning as in the foregoing chapter, i.e. the greatest imaginary part of any root or pole of $F(s)$. Then for $\omega \geq \omega_m > 2\widehat{\omega}$ we put

$$z = \frac{\omega_m}{\omega},$$

$$A(z) = A_m \cdot z^{\mu(z)}, \tag{46}$$

$$\mu(z) = \mu_m + (z-1)\mu_m' + \frac{1}{2!}(z-1)^2 \mu_m'' + \cdots \tag{47}$$

and let ${}_h\mu_m$ denote the first h terms of this series.

2:o. According to the assumption that $F(s)$ is a rational function with the degree of the denominator higher than that of the numerator we can find such a number ω_q that for $\omega > \omega_q$ the approximate formula

$$A(\omega) \approx A_q \left(\frac{\omega_q}{\omega}\right)^{\mu} \tag{48}$$

Tab.

Table of the func-

t	τ / ν		0	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18
0.2	0	λ_1	0	0.100	0.200	0.300	0.400	0.50	0.60	0.70	0.80	0.90
		λ_2	15.71	15.61	15.51	15.41	15.31	15.21	15.11	15.01	14.91	14.81
		λ_3	15.71	15.81	15.91	16.01	16.11	16.21	16.31	16.41	16.51	16.61
		λ_4	31.42	31.32	31.22	31.12	31.02	30.92	30.82	30.72	30.62	30.52
0.4	0	λ_1	0	0.050	0.100	0.150	0.200	0.25	0.30	0.35	0.40	0.45
		λ_2	7.85	7.80	7.75	7.70	7.65	7.60	7.55	7.50	7.45	7.40
		λ_3	7.85	7.90	7.95	8.00	8.05	8.10	8.15	8.20	8.25	8.30
		λ_4	15.71	15.66	15.61	15.56	15.51	15.46	15.41	15.36	15.31	15.26
0.6	0	λ_1	0	0.033	0.067	0.100	0.133	0.167	0.200	0.234	0.268	0.302
		λ_2	5.24	5.21	5.17	5.14	5.11	5.07	5.04	5.01	4.97	4.94
		λ_3	5.24	5.27	5.31	5.34	5.37	5.41	5.44	5.47	5.51	5.54
		λ_4	10.48	10.45	10.41	10.38	10.35	10.31	10.28	10.25	10.21	10.18
	1	λ_1	10.48	10.51	10.55	10.58	10.61	10.65	10.68	10.71	10.75	10.78
		λ_2	15.72	15.69	15.65	15.62	15.59	15.55	15.52	15.49	15.45	15.42
		λ_3	15.72	15.75	15.79	15.82	15.85	15.89	15.92	15.95	15.99	16.02
		λ_4	20.96	20.93	20.89	20.86	20.83	20.79	20.76	20.73	20.69	20.66
0.8	0	λ_1	0	0.025	0.050	0.075	0.100	0.125	0.150	0.176	0.201	0.226
		λ_2	3.93	3.91	3.88	3.86	3.83	3.81	3.78	3.75	3.73	3.70
		λ_3	3.93	3.95	3.98	4.00	4.03	4.05	4.08	4.11	4.13	4.16
		λ_4	7.86	7.84	7.81	7.79	7.76	7.74	7.71	7.68	7.66	7.63
	1	λ_1	7.86	7.88	7.91	7.93	7.96	7.98	8.01	8.04	8.06	8.09
		λ_2	11.79	11.77	11.74	11.72	11.69	11.67	11.64	11.61	11.59	11.56
		λ_3	11.79	11.81	11.84	11.86	11.89	11.91	11.94	11.97	11.99	12.02
		λ_4	15.72	15.70	15.67	15.65	15.62	15.60	15.57	15.54	15.52	15.49

where μ is an integer >1 is valid with a arbitrary degree of accuracy. Suppose that we take so many terms of the development (44) that the remaining expression will be independent of frequencies $<\omega_q$, i.e. if

$$\omega_q t < 2\pi \left(\nu = \frac{\omega_q t - \zeta}{2\pi} < \frac{\omega_q t}{2\pi} < 1 \right)$$

we use only one term, if $2\pi < \omega_q t < 4\pi$ we use two terms and so on. We put

$$f(t) = f_{\mathrm{I}}(t) + f_{\mathrm{II}}(t) \tag{49}$$

where $f_{\mathrm{I}}(t)$ is the sum of the terms of the development (44) for which some $\omega < \omega_q$ and $f_{\mathrm{II}}(t)$ the remaining expression, i.e.

$$f_{\mathrm{II}}(t) = \frac{2}{\pi t} \sum_{\nu=\varkappa}^{\infty} u_\nu(t) \approx \frac{2}{\pi} A_q \omega_q^\mu t^{\mu-1} \int_{\varkappa \cdot 2\pi}^{\infty} \frac{\cos\zeta \, d\zeta}{\zeta^\mu} \tag{50}$$

where $\varkappa = \left[\frac{\omega_q t}{2\pi} + 1 \right]$. Some values of the integral are given in the following table.

4.

tions λ (Formula 43: c).

0.20	0.25	0.30	0.35	0.40	0.5	0.6	0.7	0.8	0.9	1	λ_0
1.01	1.26	1.52	1.79	2.06	2.62	3.22	3.88	4.64	5.60	7.85	15.71
14.70	14.45	14.19	13.92	13.65	13.09	12.49	11.83	11.07	10.11	7.86	
16.72	16.97	17.23	17.50	17.77	18.33	18.93	19.59	20.35	21.31	23.56	
30.41	30.16	29.90	29.63	29.36	28.80	28.20	27.54	26.78	25.82	23.57	
0.50	0.63	0.76	0.89	1.03	1.31	1.61	1.94	2.32	2.80	3.93	7.85
7.35	7.22	7.09	6.96	6.82	6.54	6.24	5.91	5.53	5.05	3.92	
8.35	8.48	8.61	8.74	8.88	9.16	9.46	9.79	10.17	10.65	11.78	
15.21	15.08	14.95	14.82	14.68	14.40	14.10	13.77	13.39	12.91	11.78	
0.336	0.42	0.51	0.60	0.69	0.87	1.07	1.29	1.55	1.87	2.62	5.24
4.90	4.82	4.73	4.64	4.55	4.37	4.17	3.95	3.69	3.37	2.62	
5.58	5.66	5.75	5.84	5.93	6.11	6.31	6.53	6.79	7.11	7.86	
10.14	10.06	9.97	9.88	9.79	9.61	9.41	9.19	8.93	8.61	7.86	
10.82	10.90	10.99	11.08	11.17	11.35	11.55	11.77	12.03	12.35	13.10	
15.38	15.30	15.21	15.12	15.03	14.85	14.65	14.43	14.17	13.85	13.10	
16.06	16.14	16.23	16.32	16.41	16.59	16.79	17.01	17.27	17.59	18.34	
20.62	20.54	20.43	20.36	20.27	20.09	19.89	19.67	19.41	19.09	18.34	
0.252	0.316	0.381	0.447	0.51	0.65	0.80	0.97	1.16	1.40	1.96	3.93
3.68	3.61	3.55	3.48	3.42	3.28	3.13	2.96	2.77	2.53	1.97	
4.18	4.25	4.31	4.38	4.44	4.58	4.73	4.90	5.09	5.33	5.89	
7.61	7.54	7.48	7.41	7.35	7.21	7.06	6.89	6.70	6.46	5.90	
8.11	8.18	8.24	8.31	8.37	8.51	8.66	8.83	9.02	9.26	9.82	
11.54	11.47	11.41	11.34	11.28	11.14	10.99	10.82	10.63	10.39	9.83	
12.04	12.11	12.17	12.24	12.30	12.44	12.59	12.76	12.95	13.19	13.75	
15.47	15.40	15.34	15.27	15.21	15.07	14.92	14.75	14.56	14.32	13.76	

Tab. 5.

$$\int_{\varkappa \cdot 2\pi}^{\infty} \cos \zeta \cdot \zeta^{-\mu}\, d\zeta.$$

$\varkappa$ \ μ	1	2	3	4
1	0.02257	0.00651	0.00138	0.00026
2	0.00612	0.00095	0.00011	0.00001
3	0.00277	0.00029	0.00002	0.00000

Let us now return to formula (44) and try to estimate the error due to observations. Firstly, it is evident from (43: d) that[1]

$$D\hat{A} \leq 4\, DA_{\max} \cdot n^{-1/2} \tag{51}$$

where $DA_{\max}$ means the largest deviation of any A, ω varying in the interval

[1] For the sake of brevity I have often omitted the variables t and τ. I do not think this will cause the reader any trouble.

for which $\hat{A}$ is calculated, and n has the same meaning as in chapter II (number of observations used at the smoothing). If the four terms of $\hat{A}$ are uncorrelated, we have

$$D\hat{A} \leq 2\,D A_{\max} \cdot n^{-1/2}.$$

The standard deviation of u_ν can be evaluated by the aid of the formulas of section I e). I think (26) will be good enough in most cases and hence we write

$$D u_\nu(t) < 4\,D A_{\max} \cdot n^{-1/2} \sqrt{r(t)}, \tag{52}$$

$r(t)$ being a sort of average coefficient of correlation between different values of $\hat{A}$.

Summing up the errors of u_ν for different ν-values we are able to determine $D f_{\mathrm{I}}(t)$, provided that we succeed in determining the coefficients of correlation between different u_ν ($\nu = 0, 1, \ldots$). In most cases this correlation is very small.

The error of computation $E f_{\mathrm{I}}(t)$ is to be referred to the calculation of the integrals of u_ν and can easily be estimated.

We now have to consider the errors of $f_{\mathrm{II}}(t)$. The random error of $f_{\mathrm{II}}(t)$ arises by the determination of A_q and becomes from (50)

$$D f_{\mathrm{II}}(t) \approx \frac{2}{\pi} D A_q \cdot \omega_q^\mu t^{\mu-1} \int_{\varkappa \cdot 2\pi}^{\infty} \frac{\cos\zeta\, d\zeta}{\zeta^\mu}. \tag{53}$$

The error of computation arises from the fact that, when using formula (50), we put the sign $=$ instead of $\approx$. Suppose that the true value of $f_{\mathrm{II}}(t)$ corresponds to a certain average value of μ, say $\mu + \Delta\mu$. Then we may write

$$E f_{\mathrm{II}}(t) \approx \frac{2}{\pi} A_q\, \omega_q^\mu t^{\mu-1}\, \Delta\mu \int_{\varkappa \cdot 2\pi}^{\infty} \ln\left(\frac{\omega_q t}{\zeta}\right) \frac{\cos\zeta\, d\zeta}{\zeta^\mu}. \tag{54}$$

For $\Delta\mu$ it is enough to find an upper limit.

Let us now concentrate the results of this investigation.

1. Formula (44) gives the required time function and formula (45) its mean value. We have supposed that the amplitudes and phases are known for $\omega \leq \omega_m$. For $\omega > \omega_m$ we calculate the missing real parts by the aid of (46).

2. By the calculation of $f(t)$ we divide the time function into a sum of two parts according to (49), the first of which being a certain number of terms of (44). The second part is estimated by (50).

3. If the amplitudes are obtained by observation, we have for each term of (44) a certain error, the standard deviation of which being estimated by (52). Further there is an error of computation committed by the numerical integration.

4. The random error resp. the error of calculation for the expression defined by (50) is to be calculated from (53) and (54).

5. The total error is obtained by combining the errors mentioned in the points 3 and 4.

b) Numerical examples

1) To obtain a conception about the accuracy of the method let us first consider a simple example where the mathematical expression of the Laplace transform is given. Of course, the difference between such an example and the general case, when there is a schedule of observations of amplitudes and phases, is only formal.

Take the same transfer function which we have studied in the foregoing chapter

$$F(s) = \frac{1}{(s+1)(s+2)}.$$

Then

$$A(\omega) = \frac{2-\omega^2}{4+5\omega^2+\omega^4}.$$

Hence $\mu = 2$.

The curve of $A(\omega)$ has been drawn in fig. 6.

By the computations ω_q is chosen $= 15$. An outline of the calculations is given in tab. 6. Here the columns 3–5 contain the values of $\hat{A}_\nu(\tau; t)$, the difference between the τ-values being: in col. 3 0.02, in col. 4 0.05 and in col. 5 0.1. Further the table shows:

$u_\nu(t)$ according to (43: e) (col. 6)

$f_{\mathrm{I}}(t)$ the sum of a number of terms of (44) (col. 7)

$f_{\mathrm{II}}(t)$ the rest of the sum (44) according to (50) (col. 8)

$f(t)$ (col. 9)

and finally (col. 11) the difference between the values of $f(t)$ calculated by the method here described and the exact values of $f(t)$. Column 10 contains no numbers, as in our case there are no random errors.

2) If the amplitude $\varrho(\omega)$ and the phase $\varphi(\omega)$ are directly observed, they are random variables. Then $A(\omega)$ and $\hat{A}(\tau; t)$ are also to be considered as random variables. As a computation of the random errors of A or $\hat{A}$ from the distributions of ϱ and φ is rather hard work, it is often preferable to estimate the error of A or $\hat{A}$ directly from the calculated values of these quantities.

Suppose the variation of A is that of fig. 6 (the cross points). Between these "observed" values of A, i.e. the values calculated from the observations of ϱ and φ, is drawn a smoothing curve which is here assumed to be the curve given in example 1.

Looking at the figure we have the feeling that the error can nowhere in the considered interval $(0 < \omega < \omega_m = 10)$ exceed 0.01. Then according to (51)

Tab.

Outline of the cal-

1	2	3											4			
		$\hat{A}_\nu(\tau; t)$														
t	ν \ τ	0	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20	0.25	0.30	0.35	0.40
		0.														
0.2	0	508	497	473	434	387	337	282	232	187	142	106	037	−006	−033	−042
0.4	0	527	522	516	506	492	475	453	429	405	380	355	285	223	163	121
0.6	0	552	550	547	542	534	523	513	501	487	474	461	421	376	327	277
	1	−0025														
0.8	0	576	572	570	569	567	562	555	548	541	534	520	492	460	428	392
	1	−005										−004				

with $n = 5$ $D\hat{A} < 0.02$. This quantity is surely overestimated, for $DA(\lambda_4)$ for instance must here be much less than $DA(\lambda_1)$. Furthermore the correlation between $A(\lambda_1)$ and $A(\lambda_4)$ must be practically zero. With a high probability $D\bar{A} < 0.01$. If we take $r = 0.5$, formula (52) gives

$$Du_\nu(t) < 0.007.$$

For $t = 0.4$ we obtain, using only one term of $f_I(t)$ ($\omega_q = 15$ as in example 1),

$$Df_I(0.4) < \frac{2}{0.4\pi} \cdot 0.007 < 0.012$$

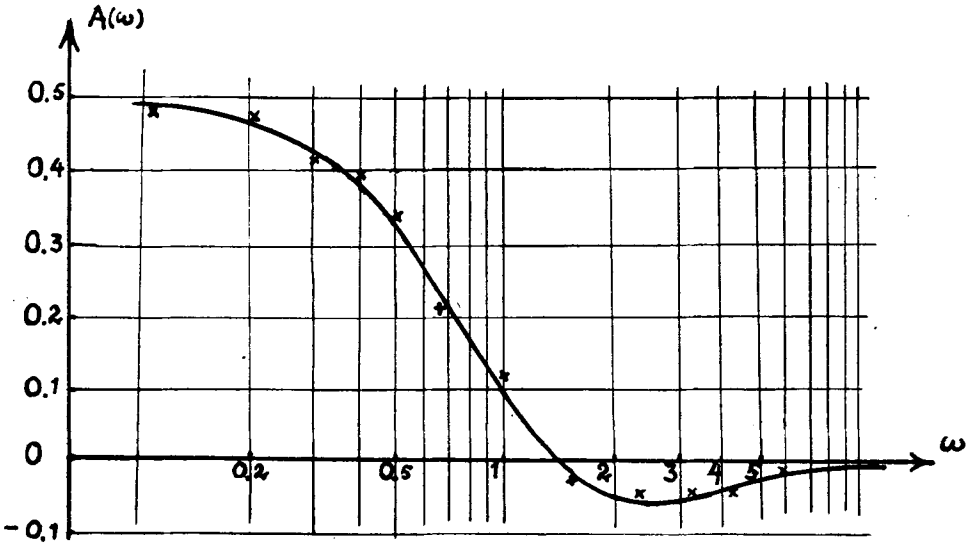

Fig. 6. $A(\omega) = \frac{2-\omega^2}{4+5\omega^2+\omega^4}$. Concerning the cross points see example 2.

ulation of $f(t)$.

5						6	7	8	9	10	11
0.5	0.6	0.7	0.8	0.9	1	$u_\nu(t)$	$f_{\mathrm{I}}(t)$	$f_{\mathrm{II}}(t)$	$f(t)$	$Df(t)$	$Ef(t)$
−048	−041	−036	−026	−016	000	0.047	0.151	−0.001	0.150	−	+0.002
043	000	−019	−025	−022	000	0.140	0.223	−0.002	0.221	−	0.000
199	131	078	033	003	000	0.236					
−0017					−0000	−0.001	0.249	−0.0004	0.249	−	+0.001
327	255	185	124	070	000	0.319					
−003					000	−0.002	0.252	−0.0005	0.251	−	+0.003

that makes about 5 % of $f_{\mathrm{I}}(0.4)$. Further we have, according to (53), putting $\mu = 2$,

$$Df_{\mathrm{II}}(0.4) < \frac{2}{\pi} 0.0005 \cdot 15^2 \cdot 0.4 \cdot 0.00651$$

which will have no effect on the total error. Thus

$$Df(0.4) < 0.012.$$

Earlier we have found

$$f(0.4) = 0.221.$$

IV. The effect on the output of omitting input frequencies in linear systems

a) Arbitrary inputs

Let the relation between the input $f_i(t)$, defined for $0 \leq t \leq T$, and the output $f_0(t)$ be given by the differential equation

$$\sum_{\nu=0}^{m} a_\nu f_0^{(\nu)}(t) = \sum_{\nu=0}^{n} b_\nu f_i^{(\nu)}(t) \tag{55}$$

where the coefficients a_ν and b_ν are real constants ($a_m = 1$, $m \geq n$). The input may be given in the form of a Fourier series

$$f_i(t) = \sum_{k=-\infty}^{\infty} A_k e^{ikt} \tag{56}$$

which is always possible, at least in a finite t-interval. In this formula k is presumed to be real but not necessarily integer. Now suppose that from this series only the terms for which $|k| < \varkappa$ will be used. Then the error committed hereby becomes

$$E f_i(t) = \sum_{|k| \geq \varkappa} A_k e^{jkt}. \tag{57}$$

We are going to determine $E f_0(t)$. In order to avoid unnecessary complications we assume that there are *no random errors*.

Owing to wellknown technical reasons (condition of stability) the roots p_1, $p_2, \ldots p_m$ of the characteristic equation

$$P(p) = \sum_{\nu=0}^{m} a_\nu p^\nu = 0$$

are, in most cases, all lying in the left half-plane. We will assume here that there are *no equal roots* among them. (If there were equal roots, the amendments to be made are almost selfevident and can be omitted here. Besides this case is not very common.) A simple root in the origin can exist (for $a_0 = 0$) but is easily eliminated by studying f_0' instead of f_0.

The general solution of (55) can be written

$$f_0(t) = \sum_{\nu=0}^{m} C_\nu e^{p_\nu t} + \sum_{k=-\infty}^{\infty} A_k Y(jk) e^{jkt} \quad (t > 0) \tag{58}$$

where $Y(\ )$ as before means the transfer function. From the fact that $f_0(t) \equiv 0$ when $f_i(t) \equiv 0$ follows that $C_1 = C_2 = \cdots = C_m = 0$. Thus

$$E f_0(t) = \sum_{|k| \geq \varkappa} A_k Y(jk) e^{jkt}. \tag{59}$$

If we use the notations:

$$\operatorname{Re}\{Y(s)\} = R(\omega) = \varrho(\omega) \cos \varphi(\omega),$$

$$\operatorname{Im}\{Y(s)\} = I(\omega) = \varrho(\omega) \sin \varphi(\omega),$$

we obtain

$$E f_0(t) = \sum_{|k| \geq \varkappa} A_k \varrho(k) e^{j[kt + \varphi(k)]}. \tag{60}$$

As a consequence of the assumption $m \geq n$ made in connection with equation (55) we can find a number ω_p so that for $\omega > \omega_p$

$$\varrho(\omega) \approx \varrho_p \left(\frac{\omega_p}{\omega}\right)^\mu \quad (\mu = m - n), \tag{61: a}$$

$$\varphi(\omega) = -\frac{\pi}{2}\mu + \frac{c_{-1}}{\omega} + \frac{c_{-2}}{\omega^2} + \cdots \tag{61: b}$$

with an error which can be made arbitrarily small for increasing ω_p. These expressions can be used for the highest frequencies of the expression (60).

b) Step input replaced by a Fourier series

In many practical problems the step input

$$f_i(t) = \begin{cases} c \text{ for } t > 0 \\ 0 \quad » \quad t < 0 \end{cases} \tag{A}$$

is replaced by a pulse function

$$g_i(t) = \begin{cases} c \text{ for } \ldots -2T < t < -T,\ 0 < t < T, \ldots \\ 0 \quad » \text{ other values of } t. \end{cases} \tag{B}$$

The corresponding curves are drawn in fig. 7.

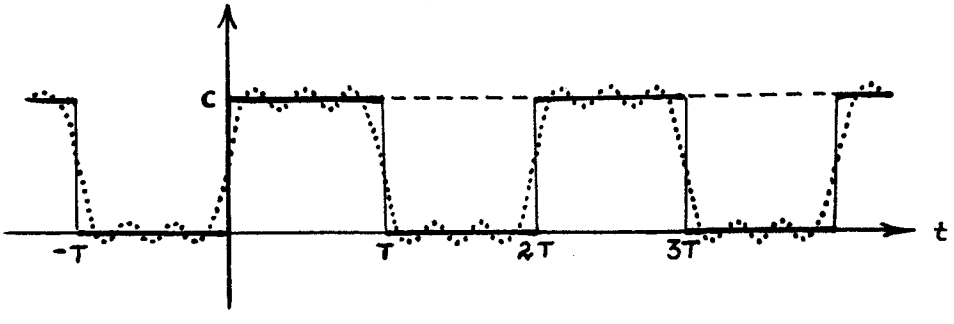

Fig. 7.

The Fourier expansion of the pulse function becomes $\left(\text{with } \omega_0 = \frac{\pi}{T}\right)$

$$g_i(t) = \frac{c}{2} + \frac{2c}{\pi} \sum_{\nu=1}^{\infty} \frac{1}{2\nu-1} \sin(2\nu-1)\,\omega_0 t. \tag{C}$$

From this series we use the terms up to and including $\frac{1}{2n-1} \sin(2n-1)\,\omega_0 t$ (the dotted line in the figure). The corresponding sum will be denoted by ${}_{2n-1}g_i(t)$. The problem here is to find the error of the output ${}_{2n-1}g_0(t)$. Note that it is only errors of computation that are dealt with here. Random errors are assumed not to occur.

We assume as before that the transfer function $Y(s)$ has the following qualities:

1. $Y(s)$ is rational.
2. The degree of the denominator is μ unites higher than that of the numerator.
3. $Y(s)$ has no poles in the right half-plane or on the axis of imaginaries.

Let us further assume n to be so large that the formulas (61: a and b) are valid for $\omega \geq (2n-1)\,\omega_0$.

If the transfer function is written in the polar form

$$Y(s) = Y(j\omega) = \varrho(\omega)\, e^{j\varphi(\omega)},$$

the output corresponding to the step input $f_i(t) = c$ can be given by the well-known formula

$$f_0(t) = \frac{c\,Y(0)}{2} + \frac{c}{\pi}\int_0^\infty \frac{\varrho(\omega)}{\omega}\sin[\omega t + \varphi(\omega)]\,d\omega. \tag{62}$$

It is important to observe here that

$$\lim_{\omega\to 0}\frac{\varrho(\omega)}{\omega}\sin[\omega t + \varphi(\omega)] = t\,\mathrm{Re}\{Y(0)\} + \lim_{\omega\to 0}\frac{\mathrm{Im}\{Y(j\omega)\}}{\omega}$$

is finite. This follows from the fact that $Y(s)$ is rational with real coefficients. Hence $Y(0)$ must be real, i.e.

$$\lim_{\omega\to 0}\mathrm{Im}\{Y(j\omega)\} = \lim_{\omega\to 0}\varrho(\omega)\sin\varphi(\omega) = 0.$$

Using the input (C) we obtain the output

$$g_0(t) = \frac{c\,Y(0)}{2} + \frac{2c\omega_0}{\pi}\sum_{\nu=1}^{\infty}\frac{\varrho[(2\nu-1)\omega_0]}{(2\nu-1)\omega_0}\sin\{(2\nu-1)\omega_0 t + \varphi[(2\nu-1)\omega_0]\}. \tag{63}$$

This formula can be derived from (62) by calculating the integral for $\omega > \omega_0$ with the trapezoidal formula by means of the ordinates at the points $\nu\omega_0$ ($\nu = 1, 3, 5, \ldots$) and then adding $\varrho(\omega_0)\sin[\omega_0 t + \varphi(\omega_0)]$. Obviously this fact could be used to determine the error of $g_0(t)$.

First there are reasons to expect that a great part of the error of $g_0(t)$ will vanish, if instead of the series (63) we use the series

$$g_0^*(t) = \frac{c\,Y(0)}{2} + \lim_{\omega\to 0}\frac{c\omega_0}{2\pi}\frac{\varrho(\omega)}{\omega}\sin[\omega t + \varphi(\omega)] + \\ + \frac{c\omega_0}{\pi}\sum_{\nu=1}^{\infty}\frac{\varrho(\nu\omega_0)}{\nu\omega_0}\sin[\nu\omega_0 t + \varphi(\nu\omega_0)], \tag{D}$$

i.e. if we use the ordinates at all points $\nu\omega_0$ ($\nu = 0, 1, 2, \ldots$). Thus we can wait to obtain a valuable piece of information, if we consider the difference

$$_{2n-1}g_0(t) - {}_{2n-1}g_0^*(t) = \frac{c}{\pi}\left\{-\frac{\omega_0}{2}\lim_{\omega\to 0}\frac{\varrho(\omega)}{\omega}\sin[\omega t + \varphi(\omega)] + \frac{\varrho(\omega_0)}{1}\sin[\omega_0 t + \varphi(\omega_0)] - \right.$$

$$- \frac{\varrho(2\omega_0)}{2}\sin[2\omega_0 t + \varphi(2\omega_0)] + \frac{\varrho[3\omega_0)}{3}\sin[3\omega_0 t + \varphi(3\omega_0)] - \cdots$$

$$\left. \cdots + \frac{\varrho[(2n-1)\omega_0]}{2n-1}\sin\{(2n-1)\omega_0 t + \varphi[(2n-1)\omega_0]\}\right\}. \tag{64}$$

I have failed in my efforts to find a general closed expression that gives a good approximation of ${}_{2n-1}g_0(t) - {}_{2n-1}g_0^*(t)$, but this has no great importance, as the series generally converges rather quickly.

The next step is to find the error of ${}_{2n-1}g_0^*(t)$. I think a good way to gain a satisfactory result is to use the following graphical method. In a coordinate system (fig. 8) we mark the points P_ν with the coordinates $\nu\omega_0$ and

$$\frac{\varrho(\nu\omega_0)}{\nu\omega_0}\sin[\nu\omega_0 t + \varphi(\nu\omega_0)]$$

for $\nu = 0, 1, \ldots (2n-1)$. If we connect the points P_ν for which ν is an odd number with straight lines, the area between the polygon and the ω-axis gives

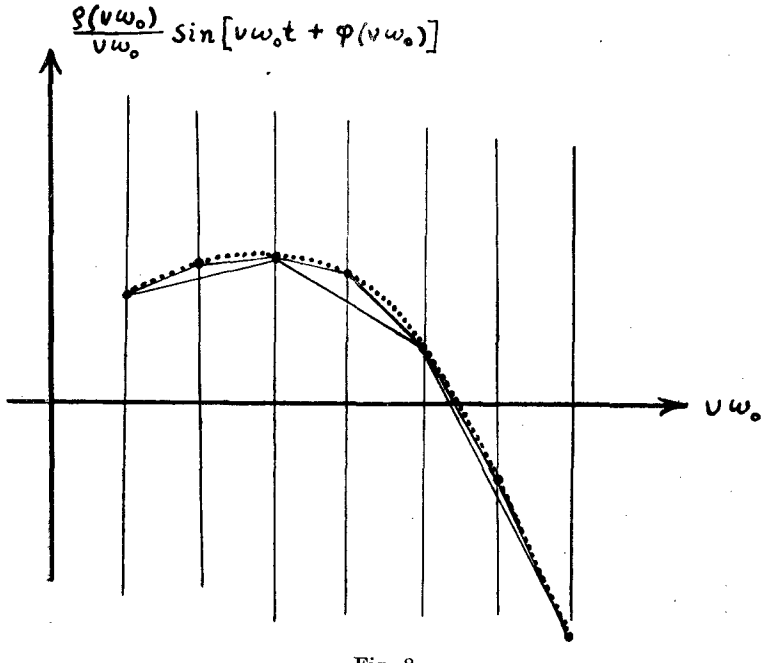

Fig. 8.

the $\sum$-expression of ${}_{2n-1}g_0(t)$. On the other hand, if the polygon is formed of all the points P_ν, we obtain the corresponding expression of ${}_{2n-1}g_0^*(t)$. The sum of the small triangels in the figure yields the difference $|{}_{2n-1}g_0(t) - {}_{2n-1}g_0^*(t)|$. The remaining error, i.e. the difference between the polygon of ${}_{2n-1}g_0^*(t)$ and the real curve, can be estimated directly from the figure, at least in such cases where we are satisfied with a rough estimation of the error.

The total error can be written

$${}_{2n-1}g_0(t) - f_0(t) = [{}_{2n-1}g_0(t) - {}_{2n-1}g_0^*(t)] + [{}_{2n-1}g_0^*(t) - {}_{2n-1}f_0(t)] + [{}_{2n-1}f_0(t) - f_0(t)].$$

From (62) follows that

$$|f_0(t) - {}_{2n-1}f_0(t)| < \frac{c\varrho_p}{\pi\mu}\left(\frac{\omega_p}{(2n-1)\omega_0}\right)^{\mu}. \tag{65}$$

Let us also consider $g_0(t) - {}_{2n-1}g_0(t)$. According to (61) we have

$$g_0(t) - {}_{2n-1}g_0(t) \approx \approx \frac{2c\varrho_p}{\pi}\left(\frac{\omega_p}{\omega_0}\right)^{\mu}\sum_{\nu=n+1}^{\infty}\frac{1}{(2\nu-1)^{\mu+1}}\sin\left(\omega_\nu t - \frac{\pi}{2}\mu + \frac{c_{-1}}{\omega_\nu} + \cdots\right) \tag{66}$$

where ω_ν stands for $(2\nu-1)\omega_0$. The majorant series

$$\sum_{\nu=n+1}^{\infty}\frac{1}{(2\nu-1)^{\mu+1}}$$

has its sum less then

$$\frac{1}{2}\int_{2n-1}^{\infty}\frac{dx}{x^{\mu+1}} = \frac{1}{2\mu(2n-1)^{\mu}},$$

and thus we have the approximation formula

$$|g_0(t) - {}_{2n-1}g_0(t)| < \frac{c\varrho_p}{\pi\mu}\left(\frac{\omega_p}{(2n-1)\omega_0)}\right)^{\mu}. \tag{67}$$

c) Numerical example

We are now going to study a numerical example of the formulas given in section b). For a certain servo system we have found the transfer function

$$Y(s) = \frac{1.313\,s^4 + 5.735\,s^3 + 477.49\,s^2 + 110.02\,s + 2.948}{3.176\,s^6 + 17.855\,s^5 + 229.09\,s^4 + 313.48\,s^3 + 489.54\,s^2 + 110.04\,s + 2.948}.$$

Thus $\mu = 2$.

Putting $\omega_0 = 0.2$, i.e. $T = 5\pi$, and using 14 terms of the series (C), $n = 13$, we obtain the following table.

ν	$2\nu-1 = k$	$k\omega_0$ (rad)	$\varrho(k\omega_0)$	$\varphi(k\omega_0)$
1	1	0.2	1.040	− 5°
2	3	0.6	1.260	− 21°
3	5	1.0	1.570	− 48°
4	7	1.4	1.550	− 88°
5	9	1.8	1.030	−130°
6	11	2.2	0.700	−143°
7	13	2.6	0.500	−153°
8	15	3.0	0.420	−166°
9	17	3.4	0.280	−180°
10	19	3.8	0.230	−195°
11	21	4.2	0.185	−203°
12	23	4.6	0.150	−216°
13	25	5.0	0.125	−228°

For $\omega > 5.0$ $\varrho(\omega)$ and $\varphi(\omega)$ can be calculated with satisfactory accuracy with the aid of the formulas (61: a and b).

For the time function ${}_{25}g_0(t)$ there have been obtained the following values ($c = 1$).

t	0	0.25	0.5	0.75	1.0	1.5	2.0
${}_{25}g_0(t)$	−0.029	+0.012	0.136	0.351	0.610	1.091	1.322
t	2.5	3.0	3.5	4.0	4.5	5.0	5.5
${}_{25}g_0(t)$	1.367	1.274	1.132	1.006	0.931	0.937	0.956
t	6.0	6.5	7.0				
${}_{25}g_0(t)$	0.998	1.023	1.029				

Let us compute the error of g_0 for $t = 3$. The terms of the series (D) are given in the last column of the following table. Fig. 9 contains the same values in a logarithmic scale.

ν	$\nu\omega_0$ (rad)	$\varrho(\nu\omega_0)$	$3\nu\omega_0$	$\varphi(\nu\omega_0)$	$3\nu\omega_0 + \varphi(\nu\omega_0)$	$\frac{\varrho(\nu\omega_0)}{\nu} \cdot \sin[3\nu\omega_0 + \varphi(\nu\omega_0)]$
1	0.2	1.040	34°	− 5°	29°	+0.504
2	0.4	1.130	69°	− 12°	57°	+0.474
3	0.6	1.260	103°	− 21°	82°	+0.416
4	0.8	1.450	138°	− 35°	103°	+0.353
5	1.0	1.570	172°	− 48°	124°	+0.260
6	1.2	1.600	206°	− 68°	138°	+0.178
7	1.4	1.550	241°	− 88°	153°	+0.101
8	1.6	1.300	275°	−110°	165°	+0.042
9	1.8	1.030	309°	−130°	179°	+0.002
10	2.0	0.830	344°	−139°	205°	−0.035
11	2.2	0.700	378°	−143°	235°	−0.052
12	2.4	0.580	413°	−150°	263°	−0.048
13	2.6	0.500	447°	−153°	294°	−0.035
14	2.8	0.420	481°	−163°	318°	−0.020
15	3.0	0.400	516°	−166°	350°	−0.005
16	3.2	0.360	550°	−170°	380°	+0.008
17	3.4	0.280	584°	−180°	404°	+0,011
18	3.6	0.250	619°	−185°	434°	+0.013
19	3.8	0.230	653°	−195°	458°	+0.012
20	4.0	0.200	688°	−199°	489°	+0.008
21	4.2	0.185	722°	−203°	519°	+0.003
22	4.4	0.175	756°	−211°	545°	−0.001
23	4.6	0.150	791°	−216°	575°	−0.004
24	4.8	0.140	825°	−224°	601°	−0.005
25	5.0	0.125	859°	−228°	631°	−0.005

From the figure we can see at once that the error of ${}_{25}g_0^*(3)$ is so small that it can be omitted. For ${}_{25}g_0(3) - {}_{25}g_0^*(3)$ we obtain the value −0.018.

We now ask for the error due to the abbreviation of the input series. This error is for all t less than

$$\frac{\varrho_p}{\pi\mu} = \frac{0.125}{2\pi} = 0.020.$$

Thus the total error must be less than 0.040. In another way the real error for $t = 3$ has been found $= 0.021$.

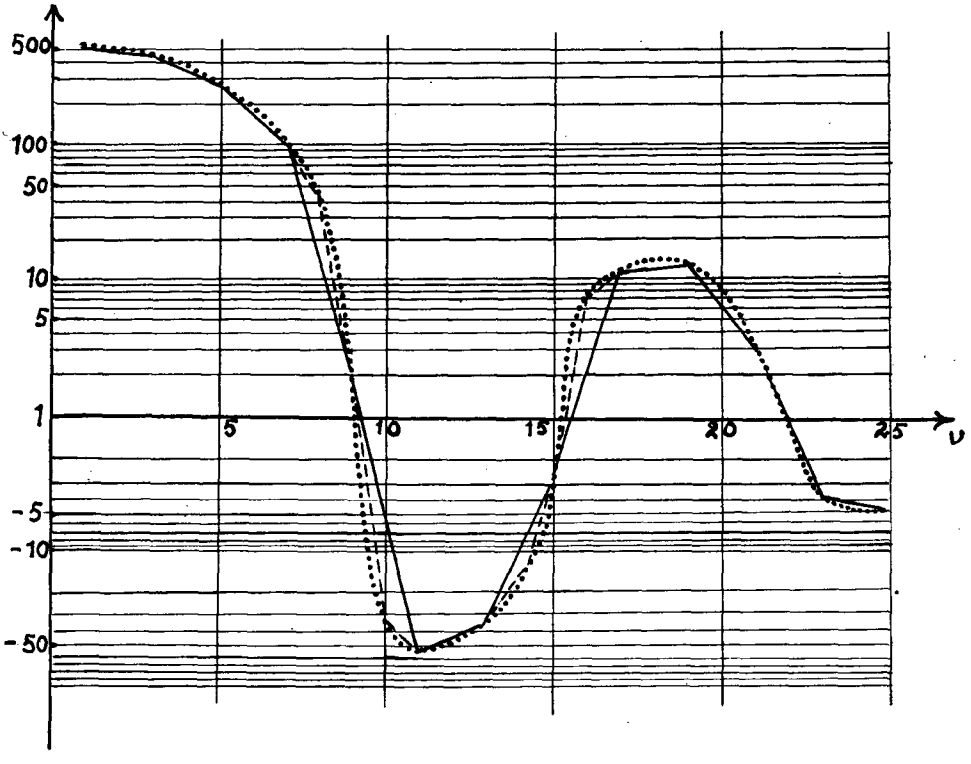

Fig. 9. $^{10}\log \left| 1000 \frac{\varrho(\nu\omega_0)}{\nu} \sin[3\nu\omega_0 + \varphi(\nu\omega_0)] \right| \cdot \operatorname{sign} \left(\frac{\varrho(\nu\omega_0)}{\nu} \sin[3\nu\omega_0 + \varphi(\nu\omega_0)] \right)$.

V. Some fundamental investigations of the probability distributions of the input and the output

a) General considerations

If we are concerned with the problem to decide which of two available servomechanisms is the best one for a certain purpose, it is not sufficient to send a certain input signal through the servomechanisms and with the aid of the responses obtained try to find an answer. It may happen that one of the servomechanisms in a special case shows an obvious superiority above the other while the latter is to be preferred in many other and perhaps practically more important cases. Therefore and also for other reasons it is necessary to try to get some experience about the probability distributions of the inputs. Such problems are treated, among others, by R. S. Phillips in a book earlier referred to: Theory of Servomechanisms; Radiation Laboratory Series 25. His treatment is based on the concepts "autocorrelation function" and "spectral density". These concepts are of great theoretical interest concerning stationary processes. However, in this section will be considered not only stationary processes and therefore I prefer to use ordinary time functions.

Any input signal is a random variable with a certain probability distribution which can vary with the time. To make the reasoning easier we divide the time axis in intervals each of which may have the length Δt which is assumed to be able to become arbitrarily small. Then an arbitrary time can be represented by an interval of the length $n\,\Delta t$. During the ν-th time interval, $(\nu-1)\,\Delta t < t < \nu\,\Delta t$, the input $x = x(t)$ is supposed to have the change Δx_ν (higher differences will not be considered here) with the distribution function

$$G_\nu(u; x) = P\{\Delta x_\nu \leq u \text{ when } x_{\nu-1} = x\}$$

where $P\{\ \}$ means "probability of" the facts given within the brackets. By acting in this way we lose the finest details of the structure, but often we are not interested in those details. In any case it is always possible to take with "the granulation" to as high a degree as we want.

For the derivative of x we have the distribution function

$$\lim_{\Delta t \to 0} G_\nu(u\,\Delta t; x) = P\{x' \leq u \text{ when } x_\nu = x\}$$

under the assumption, of course, that this limit exists.

The distribution function of the combined variable $(\Delta x_1, \Delta x_2, \ldots \Delta x_n)$ will be denoted here by $G(u_1, u_2, \ldots u_n; x)$;

$$G(u_1, u_2, \ldots u_n; x) = P\{\Delta x_1 \leq u_1,\ \Delta x_2 \leq u_2, \ldots \Delta x_n \leq u_n \text{ when } x_0 = x\}.$$

The value of the input at the end of the n-th interval

$$x_n = x_0 + \sum_{\nu=1}^{n} \Delta x_\nu = x_{n-1} + \Delta x_n$$

will become a random variable the distribution function of which we denote $F_n(u)$. The functions $F_n(u)$ can be determined from the functions $G_\nu(u; x)$ by the aid of the recursion formula

$$F_n(u) = \int_{-\infty}^{\infty} G_n(u - x; x)\, d F_{n-1}(x). \tag{68}$$

On the other hand, this formula can uniquely give G_n from F_{n-1} and F_n only when all quantities Δx_ν are independent of each other. In this case we have the well-known composition formulas

$$F_n(u) = \int_{-\infty}^{\infty} G_n(u - x)\, d F_{n-1}(x) = F_{n-1}(u) * G_n(u)$$

and

$$F_n(u) = F_0(u) * G_1(u) * G_2(u) * \cdots * G_n(u).$$

The output signal after the time $n\,\Delta t$ is denoted by y_n (in the continuous case by $y(t)$) and the increases by Δy_ν;

$$y_n = y_0 + \sum_{\nu=1}^{n} \Delta y_\nu;$$

y_n and Δy_ν are functions of $\{\Delta x_\nu\}$ and thus random variables.

We go back now for a moment to the question put at the beginning of this chapter. Suppose that y_n, always being bounded, with increasing n has to tend to a function $f_n = f(n\,\Delta t)$ in such a way that the mean deviation from this function will be as small as possible. In order to study this question we form the mathematical expectation of $(y_n - f_n)^2$;

$$M(y_n - f_n)^2 = \int (y_n - f_n)^2\, dG,$$

the integral taken over all values of $u_1, u_2, \ldots u_n$. The quantity $\mathcal{M}_T$ defined by

$$\mathcal{M}_T^2 = \frac{\Delta t}{T} \sum_{\nu=1}^{T/\Delta t} M(y_\nu - f_\nu)^2 \tag{69}$$

is commonly called the rms error (root mean square error) and is of great importance in investigations of disturbances. In the actual case $\mathcal{M}_T$ is a finite function of T. Supposing T to be large, we are tempted to say that the best servo system is the one which minimizes $\mathcal{M}_T$. We come back to this question many times in the sequel.

Let us put for the sake of simplicity $f_n \equiv 0$. For stationary processes $M y_n^2$ is independent of n. Then $\mathcal{M}_T$ is independent of T and we use the notation $\mathcal{M}$ instead of $\mathcal{M}_T$;

$$\mathcal{M}^2 = M y_n^2 = \int y^2\, dG.$$

On the other hand, if $\mathcal{M}_T$ is independent of T, the standard deviation is independent of time. This follows from the relation

$$0 = \Delta \mathcal{M}_T^2 = -\frac{\Delta t}{T + \Delta t} \mathcal{M}_T^2 + \frac{\Delta t}{T + \Delta t} M y_{N+1}^2 \left(N = \frac{T}{\Delta t}\right)$$

i.e.

$$M y_{N+1}^2 = \mathcal{M}_T^2$$

which is independent of time. But therefore it is not sure that the process is stationary in the general sense; for this all statistical moments must be independent of t.

Of course, the output y_n or $y(t)$ depends on the construction of the mechanism the behaviour of which must be regulated by the servo system. In other terms, $y(t)$ is generally the solution of a differential or integro-differential equation which is completely defined as soon as $x(t)$ and the mechanical system are given. Let us for example consider the behavior of an air torpedo. As soon as the torpedo is constructed and the acting forces are given, the path of the torpedo is completely determinable. The problem is to construct the torpedo in order to obtain the best possible stability against disturbances. First we have to try to achieve symmetry in such a way that the probability of a

deviation y_n from the right path in one direction is equal to the probability of the same deviation in the diametrically opposite direction. A necessary and sufficient condition for this is that

$$M y_n = \int y_n \, dG = 0$$

for all $n > 0$. Evidently it is difficult to get this condition exactly fulfilled. In any case it is important to know the mean value of the output at every moment of time. Further we have to minimize the rms error of y. Eventually we also need higher moments.

To be able to use the probability reasoning above we should always have the same "conditions of probability" for the same torpedo path. But, strictly speaking, there exists an infinity of different conditions of probability everyone of which with a certain probability. The distribution function of $x(t)$, $F(u)$, considered above will then give the unconditioned probability distribution (i.e. without our knowing the probability conditions). Thus we have to consider a universe of all disturbances that may happen to the torpedo in question and in that universe in an empirical way determine the probability distribution. Of course, this is a procedure that requires a great many practical experiments but I do not think it to be impossible to realize. Sometimes it may be preferable to employ the distribution for the most risky case.

As previously mentioned only if the Δx_r:s are independent of each other the probability distribution of the change of the input can be determined from the recursion formula (68). In this case we obtain at once by the aid of Laplace-operations

$$\mathcal{L}\{F_n(u)\} = \mathcal{L}\{G_n(u)\} \cdot \mathcal{L}\{F_{n-1}(u)\},$$

$$\mathcal{L}\{G_n(u)\} = \frac{\mathcal{L}\{F_n(u)\}}{\mathcal{L}\{F_{n-1}(u)\}},$$

$$G_n(u) = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} e^{su} \frac{\mathcal{L}\{F_n(u)\}}{\mathcal{L}\{F_{n-1}(u)\}} \, ds. \tag{70}$$

When the time interval Δt tends to zero,

$$\lim_{\Delta t \to 0} G_n(u \, \Delta t)$$

gives the distribution function of the derivative of x, of the limit exists.

In the general case, i.e. when the functions $G_n(u)$ depend on x, $G_n(u) = G_n(u; x)$, we must study $G_n(u; x)$ for different values of x.

There are many random processes which can be expected to become stationary, if they only would be allowed to proceed far enough. But in the case of a torpedo which has to fly through different layers of air the process of disturbing factors can scarcely be waited to reach a stationary state. In any case there will always be a transient state at the beginning of the path. The most difficult question is to treat the problem during this transient state.

Suppose that a stationary state really exists and that the distribution function of Δx at that state is $G(u; x)$. We then try to find a function $G_\nu(u; x)$ which gives the distribution of Δx_ν after the time $\nu \Delta t$ and which for increasing ν tends to $G(u; x)$. The function $G_\nu(u; x)$ may depend on a set of parameters varying with time.

We cannot always wait to be able to describe the evolution of $G_\nu(u; x)$ only with the aid of the mean value $m_\nu(x)$ and the standard deviation $\sigma_\nu(x)$ of Δx_ν. But using further the third and fourth moment, $\mu_{3\nu}(x)$ and $\mu_{4\nu}(x)$, I think we will always have a satisfactory description. Thus in the well-known development

$$G_\nu(u; x) = \sum_{\lambda=0}^{\infty} \frac{c_{\lambda\nu}}{\lambda!} \Phi^{(\lambda)}(u)$$

where

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{z^2}{2}} dz,$$

$$c_{\lambda\nu} = (-1)^\lambda \int_{-\infty}^{\infty} H_\lambda(u)\, d\, G_\nu(u; x)$$

and $H_{\lambda\nu}$ means the λ-th polynomial of Hermite we have to use only the five first terms. Then for the normalized variable

$$\xi_\nu(x) = \frac{\Delta x_\nu - m_\nu(x)}{\sigma_\nu(x)},$$

we can write

$$G_\nu(\xi_\nu; x) = \Phi(\xi_\nu) - \frac{1}{3} S_\nu(x) \cdot \Phi'''(\xi_\nu) + \frac{1}{3} E_\nu(x) \cdot \Phi^{(4)}(\xi_\nu) \qquad (71:\text{a})$$

where

$$S_\nu(x) = -\frac{1}{2} \frac{\mu_{3\nu}(x)}{[\sigma_\nu(x)]^3} \quad \text{(the skewness)}, \qquad (71:\text{b})$$

$$E_\nu(x) = \frac{1}{8} \left\{ \frac{\mu_{4\nu}(x)}{[\sigma_\nu(x)]^4} - 3 \right\} \quad \text{(the excess)}. \qquad (71:\text{c})$$

In the case of symmetri we have $S = 0$.

After having obtained a satisfactory delineation concerning the random process of the input $x(t)$ we have to study the probability distribution of the output $y(t)$. It is very likely that the distributions of Δy_ν and y_ν can be described by formulas similar to that of (71). Of course, we are always able to determine the process of $y(t)$ from that of $x(t)$ by direct computation but this is a very tedious procedure except in the case of linear systems. A possible way to obtain a practical treatment of the problem is to construct an apparatus which produces disturbances according to the probability law adopted for the input. However, I do not intend to deal with this question here.

We begin with the study of the simpliest case,. viz. that of linear systems with constant coefficients.

b) Linear systems with constant coefficients

If the relation between the input and the output can be described by a linear differential equation with constant coefficients

$$\sum_{\nu=0}^{m} a_\nu y^\nu(t) = g[x(t)] \tag{72}$$

where $g[x(t)]$ means a linear function of $x(t)]$ and its derivatives, the mathematical treatment is rather simple compared with what is required in the general case.[1] For the sake of simplicity we assume here that the characteristic equation

$$\sum_{\nu=0}^{m} a_\nu p^\nu = 0$$

has no equal roots. Then the solution of the given differential equation can be written

$$y(t) = \sum_{\nu=1}^{m} \left\{ A_\nu \int_{t_0}^{t} e^{p_\nu(t-\tau)} x(\tau)\, d\tau + C_\nu e^{p_\nu t} \right\} \tag{73}$$

where A_ν are the coefficients of the Heaviside expansion of the transfer function and C_ν the "arbitrary constants" determined by the initial conditions.

The expression

$$\psi(t) = \sum_{\nu=1}^{m} A_\nu e^{p_\nu t}$$

is generally called the weighting function of the system and gives the response of the unit impulse function acting at $t = 0$.

In the following probability consideration the process is supposed to start at $t = 0$. Thus I put $x(0)$ and all C_ν equal to zero. Then, from

$$x_n = \sum_{\nu=1}^{n} \Delta x_\nu$$

we obtain

$$y_n \approx \sum_{r=1}^{m} A_r \int_{0}^{n\Delta t} e^{p_r(n\Delta t-\tau)} \sum_{\nu=1}^{\left[\frac{\tau}{\Delta t}\right]} \Delta x_\nu\, d\tau$$

$$= \sum_{\nu=1}^{n} \sum_{r=1}^{m} A_r \int_{\nu\Delta t}^{n\Delta t} e^{p_r(n\Delta t-\tau)}\, \Delta x_\nu\, d\tau$$

$$\approx \sum_{\nu=1}^{n} \Delta x_\nu \sum_{r=1}^{m} \frac{A_r}{p_r} [e^{p_r(n-\nu)\Delta t} - 1].$$

[1] If $g[x(t)]$ is not a linear function with constant coefficients, we can consider $g[x(t)] = g(t)$ as an input function.

(The sign $\approx$ instead of $=$ is caused by the fact that Δt is not exactly equal to 0.)

Thus, if Δt is small, we have

$$y_n = \sum_{\nu=1}^{n} k_{\nu n} \Delta x_\nu \tag{74: a}$$

where

$$k_{\nu n} \approx \sum_{r=1}^{m} \frac{A_r}{p_r} [e^{p_r (n-\nu) \Delta t} - 1]. \tag{74: b}$$

In the case where the characteristic equation has equal roots, the relation (74: a) still holds, but then, of course, we have to modify the expression of $k_{\nu n}$.

The quantities $\{k_{\nu n}\}$ can be used instead of equation (72) to characterize the relation between the input $x(t)$ and the output $y(t)$. If we do not know the coefficients a_ν of equation (72), we have to determine the numbers $k_{\nu n}$ in an experimental way. This is always possible, as $k_{\nu n}$ are independent of the input signal.

Let us continue considering the torpedo taken as example in the foregoing section assuming that this torpedo will behave linearly. In order to determine the probability distributions of Δx_ν we have to study the qualities of the air especially in respect to the strength and direction of wind gusts. $\overrightarrow{\Delta x_\nu}$ can for instance be the deviation of a variable wird force vector from its mean vector. Then $\overrightarrow{\Delta x_\nu}$ can be used to characterize the atmospheric turbulence. By studying the air it is preferable to use the height above the ground instead of the time as an independent variable. Thus we write $G_h(u; x)$ instead of $G_\nu(u; x)$ for the distribution functions of the disturbance $\Delta x (= \Delta x_h)$ during the height interval Δh. Then, knowing approximately the path of the flying body, it is possible to make a transformation from height to time. In this way we can determine the distribution of Δx for each value of x and for each time interval Δt. The moments of Δx_h up to that of the fourth order may be contained as parameters in the function $G_h(u; x)$ (71: a–c). Higher moments are supposed not to be required.

Now assume $\{\Delta x_\nu\}$ to be normally distributed $(\{m_\nu(\Delta t)\}, \{\sigma_\nu(\Delta t)\})$[1] with the correlation coefficients $r_{\mu\nu}$. Then the frequency function of $\{\Delta x_\nu\}$ will be

$$\varphi(u_1, u_2, \ldots u_n) =$$

$$= \frac{1}{(2\pi)^{n/2} \Pi \sigma_\nu(\Delta t) \cdot \sqrt{R}} \exp. \left\{ -\frac{1}{2R} \sum_{\mu,\nu} \frac{[u_\mu - m_\mu(\Delta t)][u_\nu - m_\nu(\Delta t)]}{\sigma_\mu(\Delta t) \cdot \sigma_\nu(\Delta t)} R_{\mu\nu} \right\}$$

where

$$R = \| r_{\mu\nu} \|$$

and $R_{\mu\nu}$ denotes the minor corresponding to the μ-th row and the ν-th column of R. Of course R must be $\neq 0$. The characteristic function of $\{\Delta x_\nu\}$ becomes

[1] According to a commonly used notation x normal (a, b) means that x is normally distributed with the mean value a and the standard deviation b.

$$M(\exp. j \Sigma u_\nu z_\nu) = \exp. (j \Sigma m_\nu z_\nu - \tfrac{1}{2} \Sigma \sigma_\mu \sigma_\nu r_{\mu\nu} z_\mu z_\nu).$$

Thus the characteristic function of $x = \Sigma \Delta x_\nu$ is

$$M(\exp. j z \Sigma u_\nu) = \exp. (j z \Sigma m_\nu - \tfrac{1}{2} z^2 \Sigma \sigma_\mu \sigma_\nu r_{\mu\nu})$$

which means that x is normally distributed $(\Sigma m_\nu, \Sigma \sigma_\mu \sigma_\nu r_{\mu\nu})$.

We now proceed to the study of the probability law of y_n, thereby assuming *that the mean value and dispersion of y_n be finite for all values of n.* The combined distribution function of $(\Delta x_1, \Delta x_2, \ldots \Delta x_n)$ being $G(u_1, u_2, \ldots u_n)$, the mean value of y_n becomes

$$M y_n = \int \left(\sum_{\nu=1}^{n} k_{\nu n} u_\nu \right) dG = \sum_{\nu=1}^{n} k_{\nu n} \int u_\nu \, dG.$$

But

$$\int u_\nu \, dG = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u \, dG_\nu(u; x) \, dF_{\nu-1}(x) = \int_{-\infty}^{\infty} m_\nu(x) \, dF_{\nu-1}(x)$$

where

$$m_\nu(x) = \int_{-\infty}^{\infty} u \, dG_\nu(u; x)$$

gives the mean value of Δx_ν for $x_{\nu-1} = x$. If the mean value of $m_\nu(x)$ for varying x is denoted by m_ν, we have the simple formula

$$M y_n = \sum_{\nu=1}^{n} k_{\nu n} m_\nu. \tag{75}$$

The mean value of the square of y_n becomes

$$M y_n^2 = \sum_{\nu=1}^{n} k_{\nu n}^2 \int u_\nu^2 \, dG + \sum_{\mu \neq \nu} k_{\mu n} k_{\nu n} \int u_\mu u_\nu \, dG.$$

Then

$$D^2 y_n = \sum_{\nu=1}^{n} k_{\nu n}^2 \sigma_\nu^2 + \sum_{\mu \neq \nu} k_{\mu n} k_{\nu n} \sigma_\mu \sigma_\nu r_{\mu\nu} \tag{76}$$

where σ_ν means the standard deviation of Δx_ν for all values of $x_{\nu-1}$ and $r_{\mu\nu}$ the correlation coefficient between Δx_μ and Δx_ν for all values of $x_{\mu-1}$ and $x_{\nu-1}$. For the numerical treatment we have to tabulate m_ν, σ_ν and $r_{\mu\nu}$ for different indices. Then, in the case of the torpedo for example, we determine, if necessary through interpolation, the corresponding quantities along the path of flight.

If we want the moments of the third and fourth order of y_n, we have to introduce correlation coefficients of the same order.

If we construct an air torpedo in such a way that the disturbing forces become nearly symmetrical, the distribution of $(\Delta x_1, \Delta x_2, \ldots \Delta x_n)$ will probably be approximately normal. Then the distribution of y_n also becomes approximately normal, and we can be completely satisfied with knowing the quantities studied above.

Remark. The autocorrelation function between $x(t)$ and $x(t+\tau)$ is defined by the formula

$$R(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t) \cdot x(t+\tau)\, dt$$

in the continuous case and

$$R(m) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x_n \cdot x_{n+m}$$

in the case of discrete observations and gives the mean value of the product of pairs of values of the random variable which differ from each other by a constant time interval. Of course this concept has its greatest importance, when the ergodic hypothesis is valid. (The ergodic hypothesis states that the time average is equivalent to the ensemble average.) Then, M denoting the ensemble average,

$$R(\tau) = M[x(t) \cdot x(t+\tau)]$$

for every value of t. In this case $r_{\mu\nu}$ in formula (76) will be the normalized autocorrelation function between Δx_μ and Δx_ν, i.e. the autocorrelation function between these quantities divided by the product of the standard deviations.

As a simple example of the method described here we consider the differential equation

$$y'' + y = x(t).$$

We assume all Δx_ν:s to be normally distributed with the same mean value $m(\Delta t)$ and the same standard deviation $\sigma(\Delta t)$. Further we make the assumption that the time series is stationary. Then the correlation coefficient $r_{\mu\nu}$ depends only on the difference $\mu - \nu$ and will be denoted by $r_{\mu-\nu}(\Delta t)$.

The characteristic equation

$$p^2 + 1 = 0$$

has the roots $p_1 = +j$ and $p_2 = -j$. A simple calculation gives

$$k_{\nu n} \approx 2 \sin^2 \frac{(n-\nu)\Delta t}{2}.$$

Hence

$$y_n = 2 \sum_{\nu=1}^{n} \sin^2 \frac{(n-\nu)\Delta t}{2} \cdot \Delta x_\nu,$$

$$M y_n = 2 m (\Delta t) \sum_{\nu=1}^{n} \sin^2 \frac{(n-\nu)\,\Delta t}{2},$$

$$D^2 y_n = 4 [\sigma (\Delta t)]^2 \left\{ \sum_{\nu=1}^{n} \sin^4 \frac{(n-\nu)\,\Delta t}{2} + \right.$$

$$\left. + \sum_{\mu \neq \nu} \sin^2 \frac{(n-\mu)\,\Delta t}{2} \sin^2 \frac{(n-\nu)\,\Delta t}{2} \cdot r_{\mu-\nu} (\Delta t) \right\}.$$

According to the consideration above y_n will be normally distributed.

c) Approximate treatment of general systems

Method 1. Suppose that the input $x(t)$ and the output $y(t)$ be connected by the relation

$$F(x, y, t) \equiv F(x, x', x'', \ldots; y, y', y'', \ldots; t) = 0$$

where $F(\ )$ means a function of x, y, t and the derivatives of the first two variables with respect to t. Let the variations of x and y during the time interval $(\overline{\nu-1} \cdot \Delta t < t < \nu \cdot \Delta t)$ be designed by Δx_ν and Δy_ν. These quantities are random variables, both of them assumed to be continuous functions of t satisfying the equation

$$\Delta F(x, y, t) \equiv F(x + \Delta x_\nu, y + \Delta y_\nu, t + \Delta t) - F(x, y, t) = 0. \qquad (77)$$

If Δt is small, Δx_ν and (for every practically serviceable servomechanism) also Δy_ν generally will take only small values and for decreasing Δx_ν the change of y will also decrease. Of course, Δx_ν can become large, but the probability for this is generally small.

Thus we first choose a region $(-\varepsilon < \Delta x_\nu < +\varepsilon, -\delta < \Delta y_\nu < +\delta)$ where the squares of Δx_ν and Δy_ν and of their derivatives can be neglected assuming thereby that the error of Δy_ν will be less than $\eta\,\%$. In that region we put

$$\frac{\partial F}{\partial x} \Delta x_\nu + \frac{\partial F}{\partial x'} \Delta x'_\nu + \cdots + \frac{\partial F}{\partial y} \Delta y_\nu + \frac{\partial F}{\partial y'} \Delta y'_\nu + \cdots + \frac{\partial F}{\partial t} \Delta t = 0. \qquad (78)$$

For given x, y, t this equation can be solved with the ordinary methods of linear systems. Then the quantities δ and η can be determined (sometimes through practical experiments) for any value of ε and at any time point t.

Provided that Δt is small, the lowest derivatives of Δx_ν and Δy_ν will have greater influence than those of a higher order. The probability for a change of for instance $\Delta x'''_\nu$ is often notably unimportant. If we approximate the curves $\Delta x_\nu = \Delta x_\nu(t)$ by parabolas of the second degree, the third derivative is exactly equal to zero.

For large Δx_ν and Δy_ν we cannot use the equation (78). Then the problem of determining Δy_ν from (77) is identical with the problem of solving the equation $F(x, y, t) = 0$ and this is, of course, a procedure which we want to

avoid. But, as a matter of fact, every servomechanism must have such damping qualities that Δy_ν will never grow large. Thus, for practical reasons, we are often able to write

$$F(x + \Delta x_\nu, y, t) + \frac{\partial F}{\partial y} \Delta y_\nu + \frac{\partial F}{\partial y'} \Delta y'_\nu + \cdots + \frac{\partial F}{\partial t} \Delta t = 0. \tag{79}$$

This is again a linear differential equation of Δy_ν the solution of which sometimes can be used in the whole region of variation $(\Delta x_\nu, \Delta y_\nu, t)$.

The equations (78) and (79) are of the same form as the equation discussed in section b). The only difference is that we now have to deal with the variation of y instead of y itself.

After having solved the equation (79) for a set of values of x, y and t we are able to build up the whole random process. Starting from an arbitrary point (x_0, y_0, t_0) we obtain for every value

$$x_n = x_0 + \sum_{\nu=1}^{n} \Delta x_\nu$$

a corresponding value of y_n

$$y_n = y_0 + \sum_{\nu=1}^{n} \Delta y_\nu$$

where Δy_ν is a function of Δx_ν, $x_{\nu-1}$, $y_{\nu-1}$ and t $(= \nu \cdot \Delta t)$. With regard to formula (74: a) we are going to denote this function $k_{\nu n}(\Delta x_\nu, x_{\nu-1}, y_{\nu-1})$, i.e. for $x_0 = y_0 = 0$

$$y_n = \sum_{\nu=1}^{n} k_{\nu n}(\Delta x_\nu, x_{\nu-1}, y_{\nu-1}). \tag{80}$$

In the linear case we have considered the expression (80) as a linear function of the Δx_ν:s, but we must remember that this is only an approximation. The real difference between the linear and the general case lies in the fact that the right member of formula (74: a) is merely a function of the changes of the input while this function in the general case must be replaced by a combined input-output function.

Method 2. Another perhaps more perspicuous method which also assumes the existence of the derivatives but which does not require any researches on the linearity of the system is the following. Let $\overline{x(t)}$ be the observed mean value and $s(t)$ the standard deviation at the time point t of a certain input signal $x(t)$. Further, let $\overline{x^{(\nu)}(t)}$ and $s_\nu(t)$ be the corresponding values of the ν-th derivative of $x(t)$ $(\nu = 1, 2, \ldots)$. The observations are supposed to have been made at a set of time points $t_1, t_2, \ldots$.

From $F(x, y, t) = 0$ we can always form the corresponding differential equations giving the relations between (x', y', t), (x'', y'', t) and so on. These equations may be denoted $F_1(x', y', t) = 0$, $F_2(x'', y'', t) = 0$ and so on.

In order to obtain a rude conception of $y(t)$ we use a bundle of input signals $\overline{x(t)} + k s(t)$ where k is to be considered as a parameter. In the same way a

conception of $y'(t)$ can be found by considering the inputs $\overline{x'(t)} + k\,s_1(t)$ and so on. Hereby many essential features of the signals will be taken into consideration. For stationary processes all mean values and standard deviations are constants. Then the input signals are step functions.

The signals considered now constitute in one way the most risky cases, as they do not lead to any compensation of the disturbances.

Knowing the probability distribution of the input, we are able to determine the amount of probability mass between the curves corresponding to different values of k. If we let these signals pass through the servo system, we obtain a new distribution of the k-lines, and from the concentration of these lines we can judge the distribution of the output. Of course, this method can be used only when we are not interested in studying the phase relationships. Some more details will be given in section f).

d) Chain processes

Linear case. Suppose that the output y_n obtained by means of formula (74) be used as input in a new linear system characterized by the quantities $\{l_{\nu n}\}$. Then the output z_n may be written

$$z_n = \sum_{\nu=1}^{n} l_{\nu n}\, \Delta\, y_\nu$$

The total process can be symbolized by fig. 10.

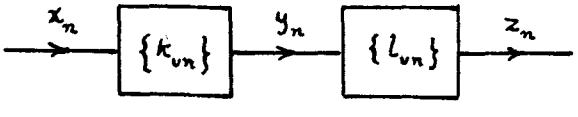

Fig. 10.

Remember that the last relation as well as equation (74) only holds approximately.

From (74: a) we obtain

$$\Delta\, y_\nu = y_\nu - y_{\nu-1} = \sum_{\mu=1}^{\nu} k_{\mu\nu} \cdot \Delta\, x_\mu - \sum_{\mu=1}^{\nu-1} k_{\mu,\nu-1} \cdot \Delta\, x_\mu = \sum_{\mu=1}^{\nu} \delta_\nu k_{\mu\nu} \cdot \Delta\, x_\mu$$

with

$$\delta_\nu k_{\mu\nu} = k_{\mu\nu} - k_{\mu,\nu-1} \quad (\delta_\nu k_{\nu\nu} = k_{\nu\nu}).$$

Thus

$$z_n = \sum_{\nu=1}^{n} l_{\nu n}\, \Delta\, y_\nu = \sum_{\nu=1}^{n} l_{\nu n} \sum_{\mu=1}^{\nu} \delta_\nu k_{\mu\nu} \cdot \Delta\, x_\mu,$$

$$z_n = \sum_{\nu=1}^{n} \left(\sum_{\mu=1}^{n} l_{\mu n} \cdot \delta_\mu k_{\nu\mu} \right) \Delta\, x_\nu. \tag{81}$$

Let us assume that we have another element $\{m_{\nu n}\}$ in the chain (fig. 11).

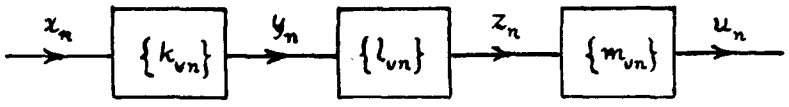

Fig. 11.

Then

$$\Delta z_\nu = \sum_{\lambda=1}^{\nu}\left(\sum_{\mu=\lambda}^{\nu} l_{\mu\nu}\cdot\delta_\mu k_{\lambda\mu}\right)\Delta x_\lambda - \sum_{\lambda=1}^{\nu-1}\left(\sum_{\mu=\lambda}^{\nu-1} l_{\mu,\nu-1}\,\delta_\mu k_{\lambda\mu}\right)\Delta x_\lambda =$$

$$= \sum_{\lambda=1}^{\nu}\left(\sum_{\mu=\lambda}^{\nu}\delta_\nu l_{\mu\nu}\cdot\delta_\mu k_{\lambda\mu}\right)\Delta x_\lambda$$

and

$$u_n = \sum_{\nu=1}^{n} m_{\nu n}\,\Delta z_\nu = \sum_{\nu=1}^{n} m_{\nu n}\sum_{\lambda=1}^{\nu}\left(\sum_{\mu=\lambda}^{\nu}\delta_\nu l_{\mu\nu}\cdot\delta_\mu k_{\lambda\mu}\right)\Delta x_\lambda$$

$$u_n = \sum_{\nu=1}^{n}\left(\sum_{\mu=\nu}^{n}\sum_{\lambda=\nu}^{\mu} m_{\mu n}\cdot\delta_\mu l_{\lambda\mu}\cdot\delta_\lambda k_{\nu\lambda}\right)\Delta x_\nu. \tag{82}$$

In this way we can proceed to an arbitrary number of elements in the chain. If there are N elements, defined by $\{{}_ik_{\nu n}\}$ $(i = 1, 2, \ldots N)$ we write for the last output

$${}_Ny_n = \sum_{\nu=1}^{n} {}_N\varkappa_{\nu n}\cdot\Delta x_\nu. \tag{83: a}$$

The coefficients ${}_N\varkappa_{\nu n}$ are determined successively by the following formulas:

$$\begin{aligned}
{}_1\varkappa_{\nu n} &= {}_1k_{\nu n},\\
{}_2\varkappa_{\nu n} &= \sum_{\mu=\nu}^{n} {}_2k_{\mu n}\cdot\delta_\mu({}_1\varkappa_{\nu\mu}) = \sum_{\mu=\nu}^{n} {}_2k_{\mu n}\cdot\delta_\mu({}_1k_{\nu n}),\\
{}_3\varkappa_{\nu n} &= \sum_{\mu=\nu}^{n} {}_3k_{\mu n}\cdot\delta_\mu({}_2\varkappa_{\nu\mu}) = \sum_{\mu=\nu}^{n}\sum_{\lambda=\nu}^{\mu} {}_3k_{\mu n}\cdot\delta_\mu({}_2k_{\lambda\mu})\cdot\delta_\lambda({}_1k_{\nu\lambda}),\\
{}_4\varkappa_{\nu n} &= \sum_{\mu=\nu}^{n} {}_4k_{\mu n}\cdot\delta_\mu({}_3\varkappa_{\nu\mu}) = \sum_{\mu=\nu}^{n}\sum_{\lambda=\nu}^{\mu}\sum_{\theta=\nu}^{\lambda} {}_4k_{\mu n}\cdot\delta_\mu({}_3k_{\lambda\mu})\cdot\delta_\lambda({}_2k_{\theta\lambda})\cdot\delta_\theta({}_1k_{\nu\theta}),
\end{aligned} \tag{83: b}$$

and so on.

General case. In the more general case treated in section c) (Method 1) we have to replace the sets of numbers $\{{}_ik_{\nu n}\}$ by sets of functions

$$\{{}_ik_{\nu n}(\Delta x_\nu,\ x_{\nu-1},\ y_{\nu-1})\}.$$

Of course, this makes the treatment much more complicated.

e) Continuous determination of the variance of the output

In previous sections of this chapter we have considered the processes going on step by step. We are now going to use a direct continuous reasoning for the special case where the system is linear. Starting from the formula (73) and putting all C_ν equal to zero we can write

$$y(t) = \int_{t_0}^{t} \left\{ \sum_{\nu=1}^{m} A_\nu e^{p_\nu (t-\tau)} \right\} x(\tau)\, d\tau.$$

The integrand contains two factors the first of which being an ordinary function of $(t-\tau)$

$$\psi(t-\tau) = \sum_{\nu=1}^{m} A_\nu e^{p_\nu (t-\tau)} \tag{84}$$

(the weighting function) characterizing the servo circuit and the second $x(\tau)$ a random variable characterizing the disturbances. The function $\psi(t-\tau)$, being the response of the unit-impulse input, can be determined once for all for every servo circuit.

The mean value of $y(t)$ becomes

$$M\,y(t) = \int_{t_0}^{t} \psi(t-\tau)\, M\,x(\tau)\, d\tau. \tag{85}$$

For the variance of $y(t)$ we obtain, writing $R(u, v)$ for the coefficient of correlation between $x(u)$ and $x(v)$ and remembering that $R(u, u) = 1$,

$$D^2 y(t) = \int_{t_0}^{t}\int_{t_0}^{t} \psi(t-u)\cdot\psi(t-v)\cdot D\,x(u)\cdot D\,x(v)\cdot R(u, v)\, du\, dv.$$

For the sake of simplicity I write

$$D^2 y(t) = \int_{t_0}^{t}\int_{t_0}^{t} \Psi(t-u,\, t-v)\cdot M(u, v)\, du\, dv \tag{86: a}$$

where

$$\Psi(t-u,\, t-v) = \psi(t-u)\cdot\psi(t-v) \tag{86: b}$$

and

$$M(u, v) = D\,x(u)\cdot D\,x(v)\cdot R(u, v). \tag{86: c}$$

For $\Psi(t-u,\, t-v) = \text{constant} = 1$ we have the case where $x(t)$ is a velocity and $y(t)$ the distance during the time $t - t_0$. Then

$$D^2 y(t) = \int_{t_0}^{t}\int_{t_0}^{t} M(u, v)\, du\, dv.$$

In section I: e) we have given an approximate treatment of the problem of determining the error of an integral where the integrand is a random variable. The method used here yields the exact expression of the standard deviation.

By using the formulas (86) the greatest trouble concerns the computation of $R(u, v)$. This computation must be carried out by means of observations for a set of points (u, v). But because of the symmetric relation $R(u, v) = R(v, u)$ we only need investigate the case $u < v$.

Now let us return to the chain processes of section d). If there are N links in the chain, we have to determine a function ${}_N\Psi(t-u, t-v)$ similar to that of the formulas (86) and corresponding to all N links. Thus the variance of the output ${}_Ny(t)$ becomes

$$D^2\,{}_Ny(t) = \int_{t_0}^{t}\int_{t_0}^{t} {}_N\Psi(t-u, t-v)\cdot M(u, v)\,du\,dv. \tag{87}$$

f) A method of computing the probability distribution of a Laplace transform from the distribution of a time function and vice versa.

Many times it is much easier to deal with the probability distributions of the Laplace transforms than with the distributions of the time functions. As an example let us consider a linear system where the Laplace transforms of the input and the output are connected by formula 3 of the introduction. Then we have

$$DF_0(s) = |Y(s)|\cdot DF_i(s).$$

On the contrary, the relation between $Df_0(t)$ and $Df_i(t)$ is rather complicated.

An important question in this connection is whether a random function can be Laplace-transformed or not. The answer will be affirmative for each member of the random process for which the Laplace transform exists. It is also clear that every statistical quantity (e.g. the mean value and the standard deviation) can be transformed.

An exact determination of the distributions of Laplace transforms requires studies of integrals of random functions and is often too tedious. Then sometimes, when we do not care of phase relationships, the following method, already used in section c) of this chapter, could be useful.

Suppose we have computed the mean value $\overline{f(t)}$ resp. $\overline{f^{(\nu)}(t)}$ $(\nu = 1, 2, \ldots)$ and the standard deviation $s(t)$ resp. $s_\nu(t)$ of a time function $f(t)$, defined for $t > 0$, and its derivatives for a set of time points $t_1, t_2, \ldots$. The derivatives are supposed to exist almost everywhere. Then in a coordinate system between t and $f(t)$ we can draw the lines

$$f(t) = \overline{f(t)} + k\cdot s(t) \quad (C_k)$$

for different values of k and similar curves for the derivatives. However, in the following I only intend to deal with $f(t)$, the reasoning for the derivatives being completely analogous.

Let us consider all curves lying in the band between $k = k_1$ and $k = k_2$. These curves will of course produce a certain bundle of lines in the Laplace-

plane. The concentration of the k-lines gives a measure of the concentration of the probability mass. For a fixed t-value k is a random variable. If we knew exactly $\overline{f(t)}$ and $s(t)$ we should have $Mk = 0$, $Dk = 1$. Now the characteristics of $f(t)$ are assumed to be computed from observed values and therefore we can only expect to have $Mk \approx 0$ and $Dk \approx 1$. However this question is of no great importance for the following reasoning.

In the case of normally distributed functions the distributions are completely determined by the mean values and the standard deviations. In this case the amount of mass between two k-lines can be taken directly from tables.

For the Laplace transforms of the C_k-curves we have

$$\mathcal{L}\{f(t)\} = F(j\omega) = \int_0^\infty \cos \omega t f(t)\, dt - j \int_0^\infty \sin \omega t f(t)\, dt$$

and

$$\left.\begin{aligned} A(\omega) &= \operatorname{Re}\{F(j\omega)\} = \int_0^\infty \cos \omega t \overline{f(t)}\, dt + k \int_0^\infty \cos \omega t s(t)\, dt \\ B(\omega) &= \operatorname{Im}\{F(j\omega)\} = -\int_0^\infty \sin \omega t \overline{f(t)}\, dt - k \int_0^\infty \sin \omega t s(t)\, dt \end{aligned}\right\} \qquad (\hat{C}_k)$$

After having calculated these expressions for some values of ω we can draw the corresponding curves $(\hat{C}_k)$ in the complex domain $A(\omega)/B(\omega)$.

Two lines $\hat{C}_{k_1}$ and $\hat{C}_{k_2}$ $(k_1 \neq k_2)$ can cut each other only if

$$k_1 \int_0^\infty \cos \omega t s(t)\, dt = k_2 \int_0^\infty \cos \omega t s(t)\, dt,$$

and

$$k_1 \int_0^\infty \sin \omega t s(t)\, dt = k_2 \int_0^\infty \sin \omega t s(t)\, dt,$$

i.e. only if

$$\int_0^\infty \cos \omega t s(t)\, dt = \int_0^\infty \sin \omega t s(t)\, dt = 0.$$

These relations can be replaced by

$$\mathcal{L}\{s(t)\} = 0.$$

Thus all intersections are lying at the points $(A(\omega)/B(\omega))$ for which $\mathcal{L}\{s(t)\} = 0$. Further, through those points all k-lines will pass.

Summing up the foregoing results we could say that to every distribution of probability mass in the $f(t)$-plane there corresponds another distribution in the $F(s)$-plane and both distributions can be represented by bundle of curves (C_k resp. $\hat{C}_k$). This is true even though the values of $\overline{f(t)}$ and $s(t)$ cannot be expected to be correct. Thus I replace the distribution functions commonly

used to represent the distributions by those bundles of k-lines. Of course, the k-lines cannot give the same amount of information as the distribution functions, but many times the conclusions one can draw from those bundles are sufficient for a particular purpose.

A practically important case is when $f(t)$ is normally distributed and the random errors of $\overline{f(t)}$ and $s(t)$ can be expected to be small for all t. Then we have

$$\left.\begin{aligned} A(\omega) \text{ approx. normal } \left(\int_0^\infty \cos \omega t \overline{f(t)}\, dt, \left|\int_0^\infty \cos \omega t\, s(t)\, dt\right|\right), \\ B(\omega) \text{ approx. normal } \left(-\int_0^\infty \sin \omega t \overline{f(t)}\, dt, \left|\int_0^\infty \sin \omega t\, s(t)\, dt\right|\right). \end{aligned}\right\}$$

In many servo problems we have to go from the $F(s)$-plane of the input to the $F(s)$-plane of the output and then to the $f(t)$-plane of the output. But often we can stop at the complex planes and carry out the investigation there.

If we want to come over from a complex plane to a time plane, we can use to the expression (42)

$$f(t) = \frac{2}{\pi} \int_0^\infty A(\omega) \cos \omega t\, d\omega.$$

Thus we once again meet an integral of a random function, i.e. the same problem with which we were confronted when we were going to determine the distribution of a Laplace transform from the distribution of a time function.

In order to obtain the distribution of the k-lines in the complex output plane we write the transfer function

$$Y(j\omega) = R(\omega) + jI(\omega).$$

Then the real part of the output transform becomes

$$A_0(\omega) = U(\omega) + kV(\omega) \tag{88: a}$$

where

$$\left.\begin{aligned} U(\omega) &= R(\omega) \int_0^\infty \cos \omega t \overline{f(t)}\, dt + I(\omega) \int_0^\infty \sin \omega t \overline{f(t)}\, dt, \\ V(\omega) &= R(\omega) \int_0^\infty \cos \omega t\, s(t)\, dt + I(\omega) \int_0^\infty \sin \omega t\, s(t)\, dt. \end{aligned}\right\} \tag{88: b}$$

A similar formula can be obtained for the imaginary part. We have only to change $\cos \omega t$ to $-\sin \omega t$ and $\sin \omega t$ to $+\cos \omega t$. Putting the expression of $A_0(\omega)$ in (42) we obtain a linear function of k and are thus able to study the bundle of k-lines in the $f_0(t)$-plane.

g) Variation of parameters in a servo circuit

Linear case. Firstly, suppose that we have to deal with a linear servo circuit and that we know the form of the transfer function

$$Y(s; a, b, \ldots) = R(\omega; a, b, \ldots) + j I(\omega; a, b, \ldots)$$

where $a, b, \ldots$ are parameters to which can be given arbitrary values in certain intervals. The parameters should be chosen in such a way that we obtain the smallest possible random error of the output. This statement includes the case where we have the possibility of choosing among a number of types of a certain servo link and want to take the best one of these types.

Using the method of section f) we draw the curves $\hat{C}_k$ in the complex output plane for different values of the parameters and study the concentration of the k-lines. The better the concentration the more reason we have to be content. From the formulas defining C_k and $\hat{C}_k$ follows namely that the distance between the curves C_k decreases at the same time as the distance between the $\hat{C}_k$-lines.

If we instead of moving in the complex plane want to carry out the investigation in the time plane, the numerical calculation is a hard work. According to (42) and (88: a and b) we have

$$f_0(t) = \frac{2}{\pi}\left\{\int_0^\infty U(\omega; a, b, \ldots)\cos\omega t\, d\omega + k\int_0^\infty V(\omega; a, b, \ldots)\cos\omega t\, d\omega\right\}. \quad (89)$$

In this expression k varies at random while U and V are supposed to be ordinary variables. Thus

$$Df_0(t) = \frac{2}{\pi}\left|\int_0^\infty V(\omega; a, b, \ldots)\cos\omega t\, d\omega\right| \cdot Dk. \quad (90)$$

For a fixed value of t $Df_0(t)$ will take its smallest value when

$$\int_0^\infty V(\omega; a, b, \ldots)\cos\omega t\, d\omega$$

becomes a minimum.

But these conditions are depending on t and we want a criterion valid in the average. Then we could use the same method as that one which led to formula (69) and consider the expression

$$\mathcal{M}^2_{T_1 T_2} = \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} D^2 f_0(t)\, dt. \quad (91)$$

Here T_2 must be taken so large that all events of interest for us are happening in the considered time interval. The reason that I have not put $T_1 = 0$

is that we have to be cautious concerning the first time, for a servo that will cause a lag during the first time can in spite of that be much better than a more sensitive one.

There are many reasons for believing that we obtain a sort of best values of $a, b, \ldots$ when minimizing the expression (91). But, of course, there are many other expressions about which we can also say that minimizing them gives the best values of our parameters. A further condition will be discussed in the next section. In this connection we could also mention the maximum likelihood method by R. A. Fischer. However, for many purposes I think the integral (91) will be most suitable. Since $Dk \approx 1$ and thus very slowly varies with t we could sometimes omit this factor in (90).

In a case where we do not know the form of $Y(s)$ the formula (91) still holds. But now the function $V(\omega)$ in (90) is entirely unknown and we have to use the calculus of variations in order to determine this function.

Very often we have to choose among a few different constructions of a servo link. Then the computation of (91) always gives us a tool by means of which we are able to decide which of the constructions is the best one.

General case. There remains to say a little about the case where we cannot use the Laplace transforms. Thus, let us assume that the relation between the input $f_i(t)$ and the output $f_0(t)$

$$g(f_i, f_i', \ldots; f_0, f_0', \ldots; t; a, b, \ldots) = 0$$

is entirely arbitrary. Also then, of course, minimizing

$$\frac{1}{T_2 - T_1} \int_{T_1}^{T_2} D^2 f_0(t)\, dt$$

makes it possible to determine the best values of the parameters, but now we cannot always put up an explicit expression of $Df_0(t)$. Generally only approximate methods stay to our disposal. However, in the frame of this investigation falls only that part of the problem which deals with the determination of the relation between $Df_0(t)$ and $Df_i(t)$. The rest of the problem, i.e. the determining of $f_0(t)$ from $f_i(t)$ is a purely mathematical one. The question that interests us now can be concentrated into the following statement. If we know Dx, what is $D\psi(x, x', x'', \ldots)$ and if we know $D\psi(x, x', x'', \ldots)$ what is Dx? This question, being of an utmost importance in the theory of probability, can generally not be solved explicitly. One way to treat the problem is to apply the method 2 of V: c), putting

$$f_i(t) = \overline{f_i(t)} + k s_i(t),$$

$$f_i^{(\nu)}(t) = \overline{f_i^{(\nu)}(t)} + k s_{i\nu}(t) \quad (\nu = 1, 2, \ldots),$$

and solve the differential equation for the output and its derivatives approximately for some values of k. By the solution we have in each case to con-

sider k as a parameter. Thus the output time function also will contain k as a parameter.

From the more or less strong concentration of k-lines in the output time planes we are able to decide which of different cases is the best one for our purpose.

h) Representation in Hilbert spaces

By a Hilbert space we mean a vector space vhere each vector has an infinite but enumerable number of components the sum of its squares being finite. In formula (69) we have met such a sum of squares with the number of terms tending to infinity when $\Delta t \to 0$.

As before, the standard deviation for an observed value, e.g. x_ν, will be denoted by Dx_ν. For the rms error vector corresponding to all observations during the time $(0, T)$ we write $\boldsymbol{X}_T$. Thus $\boldsymbol{X}_T$ has the components

$$\sqrt{\frac{\Delta t}{T}}\,Dx_1,\ \sqrt{\frac{\Delta t}{T}}\,Dx_2,\ \ldots\ \sqrt{\frac{\Delta t}{T}}\,Dx_N \quad (T = N\,\Delta t).$$

We often have to compare the input $x_\nu = x(t_\nu)$ and its observed mean value $\bar{x}_\nu$ with the output $y_\nu = y(t_\nu)$ and its mean value $\bar{y}_\nu$. It is no restriction to assume that $\bar{y}_\nu \to \bar{x}_\nu$ when $\nu \to \infty$, since this result, if it is not valid, for stable systems always can be gained simply by a coordinate transformation. However, this quality of the servo system is almost self-evident and need not be an object of any investigations. The good or bad behaviour of the system can only be characterized by the relations between the deviations of the input and the output. We therefore consider the vectors $\boldsymbol{X}_T$ and $\boldsymbol{Y}_T$ with the components

$$\sqrt{\frac{\Delta t}{T}}\,Dx_1,\ \sqrt{\frac{\Delta t}{T}}\,Dx_2,\ \ldots\ \sqrt{\frac{\Delta t}{T}}\,Dx_N$$

and

$$\sqrt{\frac{\Delta t}{T}}\,Dy_1,\ \sqrt{\frac{\Delta t}{T}}\,Dy_2,\ \ldots\ \sqrt{\frac{\Delta t}{T}}\,Dy_N$$

respectively. The deviations of x are, of course, measured from their observed mean values. Concerning the output there may be a systematical error which we want to get rid of. Let $\bar{\bar{y}}_\nu$ be the mean value of the output without such a systematical error at the time point t_ν. Then, by the computation of Dy_ν we have to use the deviations $y_\nu - \bar{\bar{y}}_\nu$ instead of $y_\nu - \bar{y}_\nu$. The systematical error $\bar{\bar{y}}_\nu - \bar{y}_\nu$ can be studied separately.

An important question is the one concerning the mutual dependence between the components of each vector. The observed values of the standard deviations are random variables and generally not independent whereas the unknown real standard deviations are constant numbers. It could therefore be of interest to try to estimate the deviations of the deviations.

Now let us consider the way by which the process is built up assuming Δt to be fixed. It is immediately seen that $\boldsymbol{X}_1$ lies along the x_1-axis, $\boldsymbol{X}_2$ in the

$(x_1\, x_2)$-plane, $\boldsymbol{X}_3$ in the $(x_1\, x_2\, x_3)$-space and so on. At the same time as the number of components increases, the components themselves will become smaller and smaller. If for increasing T $D x_N$ tends to a limit, the rms error will also tend to a limit $(\boldsymbol{X})$.

The vector representing the rms error of the output moves in a similar way. We have to consider this vector as depending on a certain number of parameters which characterize the servo system and ask for the values of these parameters that can be looked upon as the best ones in some way. Firstly it seems natural to demand that the servomechanism shall minimize the rms error. But then we often have to disregard more or less what is happening during the first time, say T_1. The reason for this was given in section g). Disregarding the time T_1 means in the language of our Hilbert space that we project the vector $\boldsymbol{Y}_T$ on the space corresponding to the remaining $T - T_1$ components. To the vector

$$\boldsymbol{Y}_{T_1 T} = \boldsymbol{Y}_T - \boldsymbol{Y}_{T_1}$$

we have to give a minimum length. Then, if we are going to determine some parameters $a, b, \ldots$ in the mathematical expressions of the servo system, we must have

$$\frac{\partial}{\partial a} |\boldsymbol{Y}_{T_1 T}| = 0 \tag{92}$$

and similarly for the other parameters.

From the nature of the problem we can conclude that there must mostly for each T exist at least one minimum satisfying (92). If there is more than one such minimum value of $\boldsymbol{Y}_{T_1 T}$ and if there exists a systematical error which has not been eliminated, we have not always without further reasons to take the minimum value which is smallest, at least not for large T. The reason that it cannot always be desirable in this case that the dispersion of the output will become smaller than the dispersion of the input depends merely on the following fact. Firstly we have to remember that we have assumed $\lim_{n\to\infty} \bar{y}_n = \lim_{n\to\infty} \bar{x}_n$. Further for large T neither $|\boldsymbol{X}_{T_1 T}|$ nor $|\boldsymbol{Y}_{T_1 T}|$ will vary very much. If now $|\boldsymbol{Y}_{T_1 T}|$ would be considerably smaller than $|\boldsymbol{X}_{T_1 T}|$, it could mean that the servo system would be too insensitive.

But there is also another possibility to judge of the best values of the parameters and which can be combined with the foregoing method. It can namely be demanded that the angle ϑ between the vectors $\boldsymbol{X}_{T_1 T}$ and $\boldsymbol{Y}_{T_1 T}$ becomes as small as possible. This can be expressed by saying that

$$\cos \vartheta = \frac{\boldsymbol{X}_{T_1 T} \cdot \boldsymbol{Y}_{T_1 T}}{|\boldsymbol{X}_{T_1 T}| \cdot |\boldsymbol{Y}_{T_1 T}|},$$

i.e. the correlation coefficient between the two vectors (where $\boldsymbol{X} \cdot \boldsymbol{Y}$ means the scalar product of the vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$) must be a maximum. Now $|\boldsymbol{Y}_{T_1 T}|$ is a minimum. Then we must have

$$\frac{\partial}{\partial a}(\boldsymbol{X}_{T_1 T} \cdot \boldsymbol{Y}_{T_1 T}) = 0 \tag{93}$$

and similarly for the other parameters. Since the vector $\boldsymbol{X}_{T_1 T}$ is independent of a, equation (93) can also be written

$$\boldsymbol{X}_{T_1 T} \cdot \frac{\partial}{\partial a} \boldsymbol{Y}_{T_1 T} = 0, \tag{94}$$

i.e. the changes of $\boldsymbol{Y}$ must be orthogonal to $\boldsymbol{X}$.

In the most general case we do not know the mathematical relation between the input and the output. Then we have to study empirically the quantities $\frac{\partial}{\partial a} D y_\nu$ for a set of time points $t_1, t_2, \ldots$ and for different values of a.

i) More than one input function

In a servo circuit there can, of course, exist more than one input random variable. Suppose we have to take into consideration the disturbing functions ${}^1x(t), {}^2x(t), \ldots$ with the distribution functions ${}^1F, {}^2F, \ldots$ and have to determine the distribution function of the output $y(t)$. If the input functions are given, the output will also be given according to the mathematical relation that must always exist. Then the distribution function of the output must be an ordinary function of the distribution functions of the inputs. But, except when the equations are linear with constant coefficients, there is generally no hope of determining the distribution function of the output in an explicit way. Then we must try to obtain satisfactory results by means of approximate methods. One such a method consists of studying the changes of certain statistical characteristics with the time and was described in section c) (Method 2).

As an example of a "device" with a great many of servo links let us consider a human being. Every input signal concerning that device belongs to an infinite universe. The response of an input is the acting of the human being and that acting occurs in a different way for different persons. The deviation from "the right way" should be a measure of the variance. But what is the right way? It can scarcely be the acting of a group of devices, the laws of whose acting being determined by the devices themselves more or less against the nature. The right way must be the average acting of a large collective of persons being influenced by the same *natural* inputs. If there should be something wrong with a link of the servo circuit for an individual the variance of that individual from the right average acting would increase. The same thing is valid for a mechanical servo system. It is therefore very important to know the probability distribution of the *natural* inputs for every link of the system.

VI. Noise in linear systems

a) Non-stationary discrete processes. Two noise components

In a book by Norbert Wiener (The Extrapolation, Interpolation and Smoothing of Stationary Time Series) the noise problem is treated for linear systems[1] in the case of stationary input signals. However, we have many times to deal with time series which are not stationary and therefore we must also try to obtain a solution for those cases. On the other hand, the linear systems are so common also for non-stationary inputs that they deserve a special treatment.

In section V b) was shown that the relation between the output y_n and the input x_n (after a time $n\,\Delta t$) in linear systems can be expressed by the formula

$$y_n = \sum_{\nu=1}^{n} k_{\nu n}\,\Delta x_\nu \qquad (74\text{: a})$$

where Δx_ν means the variation of the input during the time interval $(\nu-1)\,\Delta t < t < \nu\,\Delta t$. If Δt is small, the transfer coefficients $k_{\nu n}$ can be considered as independent of the input. On the contrary, $k_{\nu n}$ is highly dependent on Δt itself.

The discontinuous reasoning used above generates a special form of error of very great importance in noise problems. As a matter of fact, frequencies $> 2\pi/\Delta t$ radians $\cdot$ sec^{-1} cannot be studied in this manner. Thus we have to take care of the highest noise frequencies by special arrangements. The error of the output committed by omitting the highest frequencies can be calculated according to the method of IV a).

Let

$$x_n = \sum_{\nu=1}^{n} \Delta x_\nu$$

be a message[2] which shall be sent through a stable system. By passing the system the signal is assumed to have a time lag $q\,\Delta t$. Thus we have to compare the message x_n at the time point $n\,\Delta t$ with the output y_{n+q} at the time point $(n+q)\,\Delta t$. As an initial condition we assume x_0 and y_q to be zero.

First of all we must now be precise about what we have to investigate. Using a mathematical definition we could say that we have to determine the transfer function in such a way that the accumulated effect of the noise will be as small as possible. Thus we once again have met a problem where we can use the rms error. This way has been followed by Wiener in the book mentioned above.

Since equation (74: a) does not contain the transfer function directly, we have to determine the relation between that function and $\{k_{\nu n}\}$. In order to do so we start with the Laplace transform of the differential equation between x and y:

$$\sum a_\mu\, y^{(\mu)} = \sum b_\nu\, x^{(\nu)}$$

that gives

[1] With linear systems are here meant such systems which can be described by linear differential equations with constant coefficients.

[2] Of course, it need not be a "message" in the ordinary sense. The word is used here only for the sake of convenience.

$$\mathcal{L}\{y_n\} = Y(s) \cdot \mathcal{L}\{x_n\} = \sum_{\nu=1}^{n} Y(s) \cdot \mathcal{L}\{\Delta x_\nu\}$$

where $Y(s)$ as usual denotes the transfer function. *We can assume* $a_0 \neq 0$ ($\because Y(0)$ *finite*), *since in other case we have only to study* y' *instead of* y.

The time function Δx_ν varies only in the interval $(\nu-1)\Delta t < t < \nu \Delta t$ and is equal to zero for $t < (\nu-1)\Delta t$. Therefore its Laplace transform becomes

$$\mathcal{L}\{\Delta x_\nu\} = \int_{(\nu-1)\Delta t}^{\nu \Delta t} e^{-s\tau} \Delta x_\nu(\tau)\, d\tau + \Delta x_\nu \int_{\nu \Delta t}^{\infty} e^{-s\tau} d\tau = \frac{\Delta x_\nu}{s} e^{-s\nu \Delta t} + 0(\Delta x_\nu \cdot \Delta t)^{1}$$

under the assumption that $\mathrm{Re}\,(s) > 0$. In the first integral $\Delta x_\nu(\tau)$ means the change of $x_{\nu-1}$ during the time $(\nu-1)\Delta t \ldots (\nu-1)\Delta t + \tau$.

When talking about random processes, stationary or non-stationary, in this chapter I always presume, as in section V: f), that the process starts from zero, continuously or with a jump, at the moment when we begin to take care of it. Then we can always use the Laplace transform of a random process.

The Laplace transform of y_n can be written

$$\mathcal{L}\{y_n\} = \sum_{\nu=1}^{n} \left\{ \Delta x_\nu \frac{Y(s)}{s} e^{-s\nu \Delta t} + Y(s) \cdot 0(\Delta x_\nu \cdot \Delta t) \right\}.$$

Hence

$$y_n = \sum_{\nu=1}^{n} \frac{\Delta x_\nu}{2\pi j} \int_{b-j\infty}^{b+j\infty} \frac{Y(s)}{s} e^{s(n-\nu)\Delta t}\, ds + \sum_{\nu=1}^{n} \frac{1}{2\pi j} \int_{b-j\infty}^{b+j\infty} Y(s)\, e^{sn\Delta t} 0(\Delta x_\nu \cdot \Delta t)\, ds. \quad (95)$$

Here we have to choose b in such a way that the integrals converge. Since a_0 is assumed to be $\neq 0$, we can let $b \to 0$. Comparing the formulas (74: a) and (95) we see that, when taking

$$k_{\nu n} = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{Y(s)}{s} e^{s(n-\nu)\Delta t} ds, \quad (96)$$

we commit the error of y_n

$$E\, y_n = \sum_{\nu=1}^{n} \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} Y(s)\, e^{sn\Delta t} 0(\Delta x_\nu \cdot \Delta t)\, ds. \quad (97)$$

Omitting $E\, y_n$ is the same as using step inputs Δx_ν at the time points $\nu \Delta t$. Since the system is assumed to be stable, we have $E\, y_n = 0(\Delta t)$.

From (96) it is seen that, if Δt is small, $k_{\nu n}$ depends only on the difference $n - \nu$. The same thing follows from the approximate expression of $k_{\nu n}$ in V b). Therefore we write here $k_{n-\nu}$ instead of $k_{\nu n}$. Hence

$$y_n = \sum_{\nu=0}^{n-1} k_\nu \Delta x_{n-\nu}. \quad (98)$$

[1] The symbol $0(\ )$ means as usual "small of the same order as".

From (96) it is further evident *that* k_ν, *being the response of a unit step input, must be real.*

If we use a continuous reasoning, $k_{n-\nu}$ can be considered as a function of $(n-\nu)\,\Delta t = \tau$. Thus formula (96) can be written

$$k(\tau) = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \frac{Y(s)}{s} e^{s\tau}\, ds.$$

In opposition to (96) this formula is exact for stable systems. Inversion of this formula gives

$$Y(s) = s\int_0^\infty k(\tau)\, e^{-s\tau}\, d\tau. \tag{99}$$

As previously stated we have to build up the system in such a way that the noise will be eliminated to as high an extent as possible. Firstly, we have to consider the internal noise which always exists and depends on the structure of the system. This noise is approximately stationary. But there is also a lot of external noise caused by disturbances which are not generated by the system itself and this kind of noise cannot always be expected to be stationary. In order not to complicate the mathematical treatment too much *we assume here that there is only one kind of external noise with the input function* $u(t)$. *Further suppose that* $u(t)$ *has the same transfer function as* $x(t)$. *The internal noise* $v(t)$ *may also be of only one kind with the transfer function* $Z(s)$ *and the transfer coefficients* $l_{\nu n}$. Then according to (96)

$$l_{\nu n} = l_{n-\nu} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \frac{Z(s)}{s} e^{s(n-\nu)\Delta t}\, ds.$$

For more than two noise components it is not convenient to use the discontinuous method discussed here, because it would be difficult to survey the operations. In the general case a solution can be accomplished, at least theoretically, by means of calculus of variation (see section c). However, the treatment in this section could give us an idea how to get a numerical result.

The output at the time point $n\,\Delta t$ becomes, according to the assumptions made above,

$$y_n = \sum_{\nu=0}^{n-1}[k_\nu(\Delta x_{n-\nu} + \Delta u_{n-\nu}) + l_\nu\,\Delta v_{n-\nu}].$$

If the signal by passing the system will be delayed $q\,\Delta t$ units of time, the error at the time point $(n+q)\,\Delta t$ is

$$e_{n+q} = x_n - y_{n+q} = \sum_{\nu=0}^{n-1}\Delta x_{n-\nu} - \sum_{\nu=0}^{n+q-1}[k_\nu(\Delta x_{n+q-\nu} + \Delta u_{n+q-\nu}) + l_\nu\,\Delta v_{n+q-\nu}] =$$

$$= \sum_{\nu=0}^{n+q-1}({}_q\hat{k}_\nu\,\Delta x_{n+q-\nu} - k_\nu\,\Delta u_{n+q-\nu} - l_\nu\,\Delta v_{n+q-\nu}) \tag{100: a}$$

where

$$ {}_q\hat{k}_\nu = \begin{cases} -k_\nu & \text{for } \nu < q, \\ 1 - k_\nu & \text{» } \nu \geq q. \end{cases} \tag{100: b} $$

We now consider the rms error I_q given by

$$ I_q^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} M \left| \sum_{\nu=0}^{n+q-1} ({}_q\hat{k}_\nu \, \Delta x_{n+q-\nu} - k_\nu \, \Delta u_{n+q-\nu} - l_\nu \, \Delta v_{n+q-\nu}) \right|^2, \tag{101} $$

assuming this expression to be finite. Our task is to determine either the transfer functions $Y(s)$ and $Z(s)$ in the whole or some parameters of these functions by minimizing the expression (101). The signals x, u and v may be real or complex. Firstly, denoting the conjugate quantity of a complex variable z by z^*, we have

$$ \left| \sum_{\nu=0}^{n+q-1} ({}_q\hat{k}_\nu \, \Delta x_{n+q-\nu} - k_\nu \, \Delta u_{n+q-\nu} - l_\nu \, \Delta v_{n+q-\nu}) \right|^2 = $$

$$ = \sum_{\mu,\nu=0}^{n+q-1} ({}_q\hat{k}_\mu \, {}_q\hat{k}_\nu \, \Delta x_{n+q-\mu} \, \Delta x^*_{n+q-\nu} + \cdots - {}_q\hat{k}_\mu \, k_\nu \, \Delta x_{n+q-\mu} \, \Delta u^*_{n+q-\nu} \ldots). $$

For the correlation functions will be used the following notations:

$$ M(\Delta x_{n-\mu} \cdot \Delta x^*_{n-\nu}) = {}_n\alpha_{\mu\nu} $$

$$ M(\Delta u_{n-\mu} \cdot \Delta u^*_{n-\nu}) = {}_n\beta_{\mu\nu} $$

$$ M(\Delta v_{n-\mu} \cdot \Delta v^*_{n-\nu}) = {}_n\gamma_{\mu\nu} $$

$$ M(\Delta x_{n-\mu} \cdot \Delta u^*_{n-\nu}) = {}_n\delta_{\mu\nu} $$

$$ M(\Delta x_{n-\mu} \cdot \Delta v^*_{n-\nu}) = {}_n\eta_{\mu\nu} $$

$$ M(\Delta u_{n-\mu} \cdot \Delta v^*_{n-\nu}) = {}_n\theta_{\mu\nu} $$

for $\mu < n$ and $\nu < n$. (For μ or $\nu \geq n$ these quantities do not exist.) In the actual case we have $n+q$ instead of n as first index. However, in order not to complicate the text too much this index is omitted in the following. In α, β and γ we can permute the indices μ and ν (as a consequence of the fact that the imaginary part of I_q^2 is $\equiv 0$), but concerning δ, η and θ this is generally not permitted. Further, a removal of the conjugate quantity symbol from the second factor to the first does not have any influence on I_q^2.

Let us temporarily introduce the notation

$$ {}_{n+q}F_{\mu\nu} = \operatorname{Re} \{ {}_q\hat{k}_\mu \cdot {}_q\hat{k}_\nu \cdot \alpha_{\mu\nu} + k_\mu \cdot k_\nu \cdot \beta_{\mu\nu} + l_\mu \cdot l_\nu \cdot \gamma_{\mu\nu} - $$

$$ - ({}_q\hat{k}_\mu \cdot k_\nu \cdot \delta_{\mu\nu} + {}_q\hat{k}_\nu \cdot k_\mu \cdot \delta_{\nu\mu}) - ({}_q\hat{k}_\mu \cdot l_\nu \cdot \eta_{\mu\nu} + {}_q\hat{k}_\nu \cdot l_\mu \cdot \eta_{\nu\mu}) + $$

$$ + (k_\mu \cdot l_\nu \cdot \theta_{\mu\nu} + k_\nu \cdot l_\mu \cdot \theta_{\nu\mu}) \}. \tag{102} $$

Then it is immediately seen that ${}_{n+q}F_{\mu\nu} = {}_{n+q}F_{\nu\mu}$.

Using the expressions (102) we obtain

$$M\left|\sum_{\nu=0}^{n+q-1}({}_q\hat{k}_\nu\,\Delta x_{n+q-\nu}-k_\nu\,\Delta u_{n+q-\nu}-l_\nu\,\Delta v_{n+q-\nu})\right|^2=\sum_{\mu=0}^{n+q-1}\sum_{\nu=0}^{n+q-1}{}_{n+q}F_{\mu\nu}$$

and

$$I_q^2=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\sum_{\mu=0}^{n+q-1}\sum_{\nu=0}^{n+q-1}{}_{n+q}F_{\mu\nu}. \tag{103}$$

In order to study the convergence of the series (103) we start from (101) and write this formula

$$I_q^2=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}M\,|e_{n+q}|^2$$

where e_{n+q}, as already said, denotes the error of the message at the time point $(n+q)\,\Delta t$. We see at once that a satisfactory condition is that $M\,|e_{n+q}|^2$ is finite. This condition is assumed to be fulfilled here.

In practical problems we cannot proceed to the limit $N=\infty$. We therefore assume that the highest value of n to be considered is N and thus put

$$I_q^2=\frac{1}{N}\sum_{n=1}^{N}\sum_{\mu=0}^{n+q-1}\sum_{\nu=0}^{n+q-1}{}_{n+q}F_{\mu\nu}. \tag{104}$$

Many times a non-stationary random series after some time becomes more or less stationary. Then I_q according to (101) converges against the same value, as if we do not consider the first non-stationary period. However, this first period is the most important for us now and therefore it is convenient to choose $N\,\Delta t$ equal to this time. Better still would be to carry out the calculations for different values of N and take that value which in the best way corresponds to the actual case.

The expression (104) is a quadratic form of the $3\,(N+q)$ quantities ${}_q\hat{k}_\nu$, k_ν and l_ν with the side conditions between ${}_q\hat{k}_\nu$ and k_ν given in (100: b). The real number of unknown quantities is $2\,(N+q)$. Since all ${}_q\hat{k}_\nu$, k_ν and l_ν cannot be zero according to the definition of ${}_q\hat{k}_\nu$ and since other solutions of $I_q=0$ cannot be expected to exist (this should mean that the noise could be entirely eliminated), I_q^2 is always >0 and the quadratic form is positively definite. It is known that a positively definite quadratic form has always one, and only one, minimum value and the coefficients $\{k_\nu\}$ and $\{l_\nu\}$ corresponding to this value give the solution of our problem.

I am only going to treat here the case where all signals are real. Then the quantities α, β, ... are also real.

Further

$$\frac{\partial\,{}_{n+q}F_{\mu\nu}}{\partial\,{}_q\hat{k}_\mu}={}_q\hat{k}_\nu\,\alpha_{\mu\nu}-k_\nu\,\delta_{\mu\nu}-l_\nu\,\eta_{\mu\nu}\quad\text{for }\mu\neq\nu,$$

$$\frac{\partial\,{}_{n+q}F_{\mu\mu}}{\partial\,{}_q\hat{k}_\mu}=2\,({}_q\hat{k}_\mu\,\alpha_{\mu\mu}-k_\mu\,\delta_{\mu\mu}-l_\mu\,\eta_{\mu\mu}),$$

$$\frac{\partial\, {}_{n+q}F_{\mu\nu}}{\partial\, {}_q\hat{k}_\lambda} = 0 \text{ for } \lambda \neq \mu \text{ and } \nu,$$

$$\frac{\partial\, I_q^2}{\partial\, {}_q\hat{k}_\mu} = \frac{2}{N} \sum_{n=[\mu-q+1,\,1]}^{N} \sum_{\nu=0}^{n+q-1} ({}_q\hat{k}_\nu\, \alpha_{\mu\nu} - k_\nu\, \delta_{\mu\nu} - l_\nu\, \eta_{\mu\nu})$$

and similarly for the derivatives with respect to k_μ and l_μ. From the conditions (100: b) we obtain

$$\frac{\partial\, {}_q\hat{k}_\mu}{\partial\, k_\mu} = -1$$

for all μ and

$$\frac{\partial}{\partial\, k_\mu} I_q^2(k, l) = \frac{\partial}{\partial\, k_\mu} I_q^2(\hat{k}, k, l) - \frac{\partial}{\partial\, {}_q\hat{k}_\mu} I_q^2(\hat{k}, k, l).$$

In the following I consider I_q^2 as a function of only two sets of variables $\{k_\nu\}$ and $\{l_\nu\}$. Then the minimum will be reached for

$$\left.\begin{aligned} &\frac{\partial}{\partial\, {}_q\hat{k}_\mu} I_q^2(\hat{k}, k, l) = \frac{\partial}{\partial\, k_\mu} I_q^2(\hat{k}, k, l) \\ &\frac{\partial}{\partial\, l_\mu} I_q^2(\hat{k}, k, l) = 0 \\ &{}_q\hat{k}_\mu = \left\{ \begin{matrix} -k_\mu \text{ for } \mu < q \\ 1 - k_\mu \;\; » \;\; \mu \geq q \end{matrix} \right\} \\ &\mu = 0, 1, \ldots (N + q - 1) \end{aligned}\right\} \tag{105}$$

Employing the relations (100: b) we obtain

$$\frac{1}{2}\frac{\partial\, I_q^2}{\partial\, {}_q\hat{k}_\mu} = \frac{1}{N} \sum_{n=[\mu-q+1,\,1]}^{N} \sum_{\nu=q}^{n+q-1} \alpha_{\mu\nu} - \frac{1}{N} \sum_{n=[\mu-q+1,\,1]}^{N} \sum_{\nu=0}^{n+q-1} [k_\nu\, (\alpha_{\mu\nu} + \delta_{\mu\nu}) + l_\nu\, \eta_{\mu\nu}].$$

We now have to remember that $\alpha_{\mu\nu}$, $\delta_{\mu\nu}$ and $\eta_{\mu\nu}$ are functions of $n + q$ (more exactly, functions of $n + q - \mu$ and $n + q - \nu$). Then, interchanging the sums in the expression of $\partial\, I_q^2 / \partial\, {}_q\hat{k}_\mu$ we get the result

$$\frac{1}{2} N \cdot \frac{\partial\, I_q^2}{\partial\, {}_q\hat{k}_\mu} = \sum_{\nu=q}^{N+q-1} \sum_{n=\omega}^{N} \alpha_{\mu\nu} - \sum_{\nu=0}^{N+q-1} \left[k_\nu \sum_{n=\omega}^{N} (\alpha_{\mu\nu} + \delta_{\mu\nu}) + l_\nu \sum_{n=\omega}^{N} \eta_{\mu\nu} \right]$$

where $\omega = [\mu - q + 1, \nu - q + 1, 1]$ means the greatest of the numbers $\mu - q + 1$, $\nu - q + 1$ and 1. It is seen that the solution of our problem will not be directly dependent on the correlation functions $\alpha_{\mu\nu}, \beta_{\mu\nu}, \ldots$ but of their sums for n varying from ω to N. If these sums are denoted

$$A_{\mu\nu q} = \sum_{n=\omega}^{N} n+q\alpha_{\mu\nu},$$

$$B_{\mu\nu q} = \sum_{n=\omega}^{N} n+q\beta_{\mu\nu},$$

$$\Gamma_{\mu\nu q} = \sum_{n=\omega}^{N} n+q\gamma_{\mu\nu},$$

$$\Delta_{\mu\nu q} = \sum_{n=\omega}^{N} n+q\delta_{\mu\nu},$$

$$H_{\mu\nu q} = \sum_{n=\omega}^{N} n+q\eta_{\mu\nu},$$

$$\Theta_{\mu\nu q} = \sum_{n=\omega}^{N} n+q\theta_{\mu\nu},$$

we have

$$\frac{1}{2} N \cdot \frac{\partial I_q^2}{\partial \dot{l}_\mu} = \sum_{\nu=q}^{N+q-1} A_{\mu\nu q} - \sum_{\nu=0}^{N+q-1} [k_\nu (A_{\mu\nu q} + \Delta_{\mu\nu q}) + l_\nu H_{\mu\nu q}]$$

and, after a similar reasoning,

$$\frac{1}{2} N \cdot \frac{\partial I_q^2}{\partial k_\mu} = -\sum_{\nu=q}^{N+q-1} \Delta_{\nu\mu q} + \sum_{\nu=0}^{N+q-1} [k_\nu (B_{\mu\nu q} + \Delta_{\nu\mu q}) + l_\nu \Theta_{\mu\nu q}],$$

$$\frac{1}{2} N \cdot \frac{\partial I_q^2}{\partial l_\mu} = -\sum_{\nu=q}^{N+q-1} H_{\nu\mu q} + \sum_{\nu=0}^{N+q-1} [k_\nu (H_{\nu\mu q} + \Theta_{\nu\mu q}) + l_\nu \Gamma_{\mu\nu q}].$$

Using these expressions we may write the equations (105)

$$\begin{cases} \sum_{\nu=0}^{N+q-1} [k_\nu (A_{\mu\nu q} + B_{\mu\nu q} + \Delta_{\mu\nu q} + \Delta_{\nu\mu q}) + l_\nu (H_{\mu\nu q} + \Theta_{\mu\nu q})] = \sum_{\nu=q}^{N+q-1} (A_{\mu\nu q} + \Delta_{\nu\mu q}) \\ \sum_{\nu=0}^{N+q-1} [k_\nu (H_{\nu\mu q} + \Theta_{\nu\mu q}) + l_\nu \Gamma_{\mu\nu q}] = \sum_{\nu=q}^{N+q-1} H_{\nu\mu q} \end{cases} \tag{106}$$

$$(\mu = 0, 1, \ldots \overline{N+q-1})$$

We have seen (p. 200) that our minimum problem has always one, and only one, solution, obtainable by solving the equations (106) with respect to $\{k_\nu\}$ and $\{l_\nu\}$. By means of formula (99) we are then able to determine the transfer functions $Y(s)$ and $Z(s)$ corresponding to $\{k_\nu\}$ and $\{l_\nu\}$. It is easily seen that for a finite time $N\Delta t$ (and all messages must be of finite duration) the transfer functions thus obtained have all the qualities of a stable system.

The most difficult part of the problem is perhaps the determination of the correlation functions $\alpha, \beta, \ldots \theta$. Taking for instance $\delta_{\mu\nu}$ we have to put up a correlation table between $\Delta x_{n-\mu}$ and $\Delta u_{n-\nu}$ (Tab. 7).

After having determined the length d of the intervals in a convenient manner we have to reckon the number of cases for which the variations are lying within

Tab. 7.

Correlation between $\Delta x_{n-\mu}$ and $\Delta u_{n-\nu}$.

Δx \ Δu		$-3/2\,d$ $-1/2\,d$	$-1/2\,d$ $+1/2\,d$	$+1/2\,d$ $+3/2\,d$		$\Sigma\,\Delta u$	$\Sigma\,\Delta x\,\Delta u$ Δx const
$-3/2\,d$ $-1/2\,d$			7	2			
$-1/2\,d$ $+1/2\,d$		4	6	10			
$+1/2\,d$ $+3/2\,d$			3	7			
$\Sigma\,\Delta x$							
$\Sigma\,\Delta x\,\Delta u$ Δu const							

the different squares of the table and then determine the sum over all products $\Delta x \cdot \Delta u$. The correlation functions must be computed for a sufficient number of pairs μ, ν. For other pairs these functions may be obtained by interpolation.

b) Stationary processes as a special case

If all the interfering time series are stationary, the formulas will be considerably simplified. Firstly, all the correlation functions $\alpha, \beta \ldots \theta$ will be dependent only on the differences $\mu - \nu$. Thus instead of ${}_n\delta_{\mu\nu}$ we write $\delta_{\mu-\nu}$ (the index n can, of course, be omitted, since n only appears in the differences $\overline{n-\mu} - \overline{n-\nu}$). Further it is easily seen that the correlation functions will become even functions of $\mu - \nu$. This is a consequence of the fact that the frequency functions of the combined variables are even functions of this parameter $\mu - \nu$. Finally, we do not need again the quantities $A_{\mu\nu q}, B_{\mu\nu q}, \ldots$, since all terms of those sums are equal. For the sake of brevity we employ the notation ${}^{N}_{q}C_{\mu\nu}$ for the number of terms divided by N, i.e.

$$ {}^{N}_{q}C_{\mu\nu} = \frac{N+1-\omega}{N} = \begin{cases} 1 \text{ for } \mu \leq q \text{ and } \nu \leq q \\ 1 - \dfrac{\mu - q}{N} \text{ for } \mu \geq q \text{ and } \mu \geq \nu \\ 1 - \dfrac{\nu - q}{N} \text{ for } \nu \geq q \text{ and } \nu \geq \mu \end{cases} $$

The equations (106) become now

$$\begin{cases} \sum_{\nu=0}^{N+q-1} {}_qC_{\mu\nu}\,[k_\nu\,(\alpha_{\mu-\nu}+\beta_{\mu-\nu}+2\,\delta_{\mu-\nu})+l_\nu\,(\eta_{\mu-\nu}+\theta_{\mu-\nu})] = \sum_{\nu=q}^{N+q-1} {}_qC_{\mu\nu}\,(\alpha_{\mu-\nu}+\delta_{\mu-\nu}) \\ \sum_{\nu=0}^{N+q-1} {}_qC_{\mu\nu}\,[k_\nu\,(\eta_{\mu-\nu}+\theta_{\mu-\nu})+l_\nu\,\gamma_{\mu-\nu}] = \sum_{\nu=q}^{N+q-1} {}_qC_{\mu\nu}\,\eta_{\mu-\nu} \end{cases} \tag{107}$$

In his book mentioned above (p. 196) Wiener has given an explicit solution of the noise problem in case of stationary processes employing a continuous reasoning. Thus in Wieners treatment the output has the form

$$y(t) = \int x(t-\tau)\,dK(\tau)$$

instead of

$$y_n = \sum k_\nu\,\Delta x_{n-\nu}.$$

c) Continuous reasoning. An arbitrary number of noise components

Let us now assume that there exists an arbitrary number (Λ) of noise components $v_1(t)$, $v_2(t)$, ... $v_\Lambda(t)$ (stationary or non-stationary) with the transfer functions $Z_1(s)$, $Z_2(s)$, ... $Z_\Lambda(s)$ and the transfer coefficients $k_1(\tau)$, $k_2(\tau)$, ... $k_\Lambda(\tau)$. Thus we have the relations

$$k_\lambda(\tau) = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \frac{Z_\lambda(s)}{s}\,e^{s\tau}\,ds,$$

thereby assuming that $Z_\lambda(0) \neq \infty$ ($\lambda = 1, 2, \ldots \Lambda$).

The message may be $x(t)$ with the transfer function $Y(s)$ and the transfer coefficients $k(\tau)$. In this section *all variables are presumed to be real and to vary continuously as well as their time-derivatives.*[1] Further, *the derivatives are supposed to be finite with the probability* 1.

For the sake of simplicity *we postulate that there is no time delay in the system.* Then the error at the time point t becomes

$$x(t)-y(t) = -\int_0^t \left[\hat{k}(\tau)\frac{dx(t-\tau)}{d\tau} - \sum_{\lambda=1}^{\Lambda} k_\lambda(\tau)\frac{dv_\lambda(t-\tau)}{d\tau}\right] d\tau$$

where $\hat{k}(\tau) = 1 - k(\tau)$. The transfer coefficients $k(\tau)$ and $k_\lambda(\tau)$ ($\lambda = 1, 2, \ldots \Lambda$) are to be determined in such a manner that

$$I^2 = \frac{1}{T}\int_0^T M\,|x(t)-y(t)|^2\,dt \tag{108}$$

---

[1] As before we assume $x(t) \equiv 0$ and $v_\lambda(t) \equiv 0$ for $t \leq 0$.

becomes a minimum; T means the time being of interest in the actual case. Since the integral over the mean value of a function always equals the mean value of the integral, we can also write

$$I^2 = M\left(\frac{1}{T}\int_0^T |x(t) - y(t)|^2\, dt\right). \tag{109}$$

Upon introducing the notation

$$g(\tau, t) = \hat{k}(\tau)\frac{dx(t-\tau)}{d\tau} - \sum_{\lambda=1}^{\Lambda} k_\lambda(\tau)\frac{dv_\lambda(t-\tau)}{d\tau}$$

we obtain

$$I^2 = M\left\{\frac{1}{T}\int_0^T\left[\int_0^t g(\tau, t)\, d\tau\right]^2 dt\right\}. \tag{110}$$

According to its definition $g(\tau, t)$ is a linear and homogeneous function of the transfer coefficients $\hat{k}$ and $\{k_\lambda\}$. The transfer coefficients are continuous functions of the time. Without further restrictive conditions the minimization of I^2 would make all transfer coefficients identically equal to zero and that is of course an impossibility. A sufficient assumption is the natural postulate that one of the noise components has the same transfer function as the message. Then putting $k(\tau) = k_1(\tau)$ we have

$$\hat{k}(\tau) = 1 - k_1(\tau).$$

In order to obtain the most symmetric form of $g(\tau, t)$ we put for a while

$$x = -v_0,$$
$$\hat{k} = k_0$$

and thus obtain

$$g(\tau, t) = -\sum_{\lambda=0}^{\Lambda} k_\lambda(\tau)\frac{dv_\lambda(t-\tau)}{d\tau}$$

with the restrictive condition

$$k_0(\tau) = 1 - k_1(\tau).$$

By the minimization of I^2 let us employ the methods of the calculus of variation. Thus we replace $k_\lambda(\tau)$ by $k_\lambda(\tau) + \alpha\varkappa_\lambda(\tau)$ $(\lambda = 1, 2, \ldots \Lambda)$ and try to determine the k_λ:s in such a way that

$$\frac{\partial I^2}{\partial \alpha} = 0$$

for $\alpha = 0$ and for all possible functions $\varkappa_\lambda(\tau)$.[1]

[1] It is easily seen that the use of different α-values (α_λ) for different $k_\lambda(\tau)$ implies no further generality in this case.

If we introduce the notation

$$\gamma(\tau, t) = -\sum_{\lambda=0}^{A} \varkappa_\lambda(\tau) \frac{d v_\lambda(t-\tau)}{d\tau}$$

with

$$\varkappa_0(\tau) = -\varkappa_1(\tau) \quad .$$

and further put

$$G(t) = \int_0^t g(\tau, t)\, d\tau,$$

$$\Gamma(t) = \int_0^t \gamma(\tau, t)\, d\tau$$

formula (110) changes into

$$I^2 + \alpha\, \delta I^2 + \frac{\alpha^2}{2} \delta^2 I^2 = M\left\{\frac{1}{T}\int_0^T [G(t) + \alpha\, \Gamma(t)]^2\, dt\right\}$$

where δI^2 and $\delta^2 I^2$ are the conventional notations for the first and second variation. Hence $\delta I^2 = 0$ gives

$$M\left\{\int_0^T G(t)\, \Gamma(t)\, dt\right\} = 0. \tag{111}$$

Now it is time again to place the symbol M after the integral sign. According to the relations given above the random factors of the integrands are

$$\frac{d v_\mu(t-\sigma)}{d\sigma} \cdot \frac{d v_\nu(t-\tau)}{d\tau}.$$

Thus, putting

$$M\left(\frac{d v_\mu(t-\sigma)}{d\sigma} \cdot \frac{d v_\nu(t-\tau)}{d\tau}\right) = m_{\mu\nu}(\sigma, \tau, t),$$

we have from (111)

$$\int_0^T dt \int_0^t \int_0^t \sum_{\mu=0}^{A} \sum_{\nu=0}^{A} \varkappa_\mu(\sigma) \cdot k_\nu(\tau) \cdot m_{\mu\nu}(\sigma, \tau, t)\, d\sigma\, d\tau = 0.$$

Since the sums contain only a finite number of terms, they do not cause any troubles. Separating that part of the triple integral which is free from the arbitrary functions $\varkappa_\mu(\sigma)$, i.e.

$$H_\mu(\sigma, t) = \int_0^t \sum_{\nu=0}^{A} k_\nu(\tau)\, m_{\mu\nu}(\sigma, \tau, t)\, d\tau,$$

our equation becomes

$$\sum_{\mu=0}^{\Lambda} \int_0^T dt \int_0^t \varkappa_\mu(\sigma)\, H_\mu(\sigma, t)\, d\sigma = 0.$$

The functions $m_{\mu\nu}(\sigma, \tau, t)$, $k_\nu(\tau)$ and $\varkappa_\mu(\sigma)$ are all assumed to be finite and integrable in the region $(0 \leq \sigma, \tau, t \leq T)$. Then also $H_\mu(\sigma, t)$ must be finite and integrable in the same region and we can interchange the order of integration of the last equation and thus obtain

$$\sum_{\mu=0}^{\Lambda} \int_0^T \varkappa_\mu(\sigma)\, d\sigma \int_\sigma^T H_\mu(\sigma, t)\, dt = 0. \tag{112}$$

In the actual problem we can be satisfied by considering such functions $\varkappa_\mu(\sigma)$ which are continuous and possess continuous derivatives for $0 \leq \sigma \leq T$. Further we shall have $\varkappa_\mu(0) = \varkappa_\mu(T) = 0$. Then, as shown in almost all books on calculus of variation, equation (112) can be fulfilled for all possible functions $\varkappa_\mu(\sigma)$ if, and only if, for all σ $(0 \leq \sigma \leq T)$ the coefficients of these functions disappear, i.e. only if

$$\left.\begin{aligned} -\int_\sigma^T H_0(\sigma, t)\, dt + \int_\sigma^T H_1(\sigma, t)\, dt = 0, \\ \int_\sigma^T H_\mu(\sigma, t)\, dt = 0 \quad (\mu = 2, 3, \ldots \Lambda). \end{aligned}\right\} \tag{113}$$

Introducing the quantities $k(\tau)$ and $m(\sigma, \tau, t)$ we can write the equations (113)

$$\left.\begin{aligned} \sum_{\nu=0}^{\Lambda} \int_\sigma^T dt \int_0^t k_\nu(\tau)\, [m_{1\nu}(\sigma, \tau, t) - m_{0\nu}(\sigma, \tau, t)]\, d\tau = 0, \\ \sum_{\nu=0}^{\Lambda} \int_\sigma^T dt \int_0^t k_\nu(\tau)\, m_{\mu\nu}(\sigma, \tau, t)\, d\tau = 0 \quad (\mu \geq 2). \end{aligned}\right\}$$

Since $k(\tau)$ and $m(\sigma, \tau, t)$ are supposed to be finite, the integral signs may be interchanged, which gives for $\mu \geq 2$.

$$\sum_{\nu=0}^{\Lambda} \left\{ \int_0^\sigma k_\nu(\tau)\, d\tau \int_\sigma^T m_{\mu\nu}(\sigma, \tau, t)\, dt + \int_\sigma^T k_\nu(\tau)\, d\tau \int_\tau^T m_{\mu\nu}(\sigma, \tau, t)\, dt \right\} = 0$$

and similarly for $\mu = 0$ and 1. If we put

$$\mathcal{M}_{\mu\nu}(\sigma, \tau) = \int_{[\sigma, \tau]}^T m_{\mu\nu}(\sigma, \tau, t)\, dt,$$

we obtain

$$\left.\begin{aligned}\sum_{\nu=0}^{\Lambda}\int_0^T k_\nu(\tau)\,[\mathcal{M}_{0\nu}(\sigma,\tau)-\mathcal{M}_{1\nu}(\sigma,\tau)]\,d\tau=0,\\ \sum_{\nu=0}^{\Lambda}\int_0^T k_\nu(\tau)\,\mathcal{M}_{\mu\nu}(\sigma,\tau)\,d\tau=0\quad(\mu\geq 2).\end{aligned}\right\}$$

If we now pass from v_0 back to x, we have to change the signs of $m_{\mu\nu}$ and $\mathcal{M}_{\mu\nu}$ in all cases where one of the indices μ and ν (but not both of them) equals zero. Thus, in the following we let $m_{\mu 0}$ denote the correlation function between $\frac{dv_\mu}{dt}$ and $\frac{dx}{dt}$ instead of, as before, between $\frac{dv_\mu}{dt}$ and $\frac{dv_0}{dt}$. If we further pay regard to the relation between k_0 and k_1, our equations take the form

$$\left.\begin{aligned}&\int_0^T k_1(\tau)\,[\mathcal{M}_{00}(\sigma,\tau)+\mathcal{M}_{10}(\sigma,\tau)+\mathcal{M}_{01}(\sigma,\tau)+\mathcal{M}_{11}(\sigma,\tau)]\,d\tau+\\ &+\sum_{\nu=2}^{\Lambda}\int_0^T k_\nu(\tau)\,[\mathcal{M}_{0\nu}(\sigma,\tau)+\mathcal{M}_{1\nu}(\sigma,\tau)]\,d\tau=\int_0^T[\mathcal{M}_{00}(\sigma,\tau)+\mathcal{M}_{10}(\sigma,\tau)]\,d\tau,\\ &\int_0^T k_1(\tau)\,[\mathcal{M}_{\mu 0}(\sigma,\tau)+\mathcal{M}_{\mu 1}(\sigma,\tau)]\,d\tau+\sum_{\nu=2}^{\Lambda}\int_0^T k_\nu(\tau)\,\mathcal{M}_{\mu\nu}(\sigma,\tau)\,d\tau=\\ &=\int_0^T\mathcal{M}_{\mu 0}(\sigma,\tau)\,d\tau\quad(\mu=2,3,\ldots\Lambda).\end{aligned}\right\}\tag{114}$$

Before studying the equations (114) in details we are going to check these equations by comparing them with the corresponding equations of section a). Then we have to put $\Lambda = 2$ and replace the variables according to the following schedule:

Section c)		Section a)
v_1	→	u
v_2	→	v
k_1	→	k
k_2	→	l
σ	→	$\mu\,\Delta t$
τ	→	$\nu\,\Delta t$
t	→	$n\,\Delta t$
T	→	$N\,\Delta t$

From this follows, with the notations of section a),

$$\left.\begin{aligned} m_{00}(\sigma,\tau,t) &= M\left(\frac{dx(t-\sigma)}{d\sigma}\cdot\frac{dx(t-\tau)}{d\tau}\right) \approx \frac{M(\Delta x_{n-\mu}\cdot\Delta x_{n-\mu})}{\Delta t^2} = \frac{{}_n\alpha_{\mu\nu}}{\Delta t^2} \\ &\cdots\cdots\cdots\cdots \\ m_{12}(\sigma,\tau,t) &= M\left(\frac{du(t-\sigma)}{d\sigma}\cdot\frac{dv(t-\tau)}{d\tau}\right) \approx \frac{M(\Delta u_{n-\mu}\cdot\Delta v_{n-\nu})}{\Delta t^2} = \frac{{}_n\theta_{\mu\nu}}{\Delta t^2} \end{aligned}\right\}$$

$$\left.\begin{aligned} \mathcal{M}_{00}(\sigma,\tau) &= \int_{[\sigma,\tau]}^{T} m_{00}(\sigma,\tau,t)\,dt \approx \frac{1}{\Delta t}\sum_{n=\omega}^{N} {}_n\alpha_{\mu\nu} = \frac{1}{\Delta t}A_{\mu\nu 0} \\ &\cdots\cdots\cdots\cdots \\ \mathcal{M}_{12}(\sigma,\tau) &= \int_{[\sigma,\tau]}^{T} m_{12}(\sigma,\tau,t)\,dt \approx \frac{1}{\Delta t}\sum_{n=\omega}^{N} {}_n\theta_{\mu\nu} = \frac{1}{\Delta t}\Theta_{\mu\nu 0} \end{aligned}\right\}$$

Using these approximate expressions we have

$$\left.\begin{aligned} &\sum_{\nu=0}^{N-1} k_\nu [A_{\mu\nu 0} + A_{\nu\mu 0} + A_{\mu\nu 0} + B_{\mu\nu 0}] + \sum_{\nu=0}^{N-1} l_\nu [H_{\mu\nu 0} + \Theta_{\mu\nu 0}] = \sum_{\nu=0}^{N-1} [A_{\mu\nu 0} + A_{\nu\mu 0}] \\ &\sum_{\nu=0}^{N-1} k_\nu [H_{\nu\mu 0} + \Theta_{\nu\mu 0}] + \sum_{\nu=0}^{N-1} l_\nu \Gamma_{\mu\nu 0} = \sum_{\nu=0}^{N-1} H_{\nu\mu 0} \\ &\qquad (\mu = 0, 1, \ldots \overline{N-1}) \end{aligned}\right\}$$

i.e. exactly the same equations to which we were led in section a), except that there we had not assumed $q = 0$.

The equations of the system (114) are all of the form

$$\sum_{\nu=1}^{A} \int_0^T k_\nu(\tau)\, K_{\mu\nu}(\sigma,\tau)\,d\tau = \int_0^T K_{\mu 0}(\sigma,\tau)\,d\tau \tag{115}$$

where $K_{\mu\nu}(\sigma,\tau)$ are known functions.

If there is only one source of noise with the transfer coefficients $k(\tau)$, these coefficients are obtained from an integral equation of the first kind

$$\int_0^T k(\tau)\, K(\sigma,\tau)\,d\tau = f(\sigma). \tag{116}$$

Some hints concerning the practical solution of this equation will be given in the next section. On the other hand, in case of more than one noise component,

there exists a method by Fredholm by means of which the system of integral equations can be replaced by a unique equation. The following derivation has been taken from Goursat (Cours d'analyse mathématique).

If we put

$$\frac{\sigma}{T} = x,$$

$$\frac{\tau}{T} = y,$$

$$k_\nu(\tau) = k_\nu(Ty) = k_\nu^0(y),$$

$$K_{\mu\nu}(\sigma, \tau) = K_{\mu\nu}(Tx, Ty) = K_{\mu\nu}^0(x, y),$$

$$\int_0^T K_{\mu 0}(\sigma, \tau)\, d\tau = T f_\mu^0(x),$$

our equations get the form

$$\sum_{\nu=1}^{\Lambda} \int_0^1 k_\nu^0(y)\, K_{\mu\nu}^0(x, y)\, dy = f_\mu^0(x) \qquad (117)$$

$$(\mu = 1, 2, \ldots \Lambda).$$

We now introduce a new kernel, defined for $0 < x < \Lambda$, $0 < y < \Lambda$:

$$H(x, y) = K_{\mu\nu}^0(x - \mu + 1, y - \nu + 1) \text{ for } \begin{pmatrix} \mu - 1 < x < \mu \\ \nu - 1 < y < \nu \end{pmatrix}$$

$$(\mu = 1, 2, \ldots \Lambda, \quad \nu = 1, 2, \ldots \Lambda).$$

Let us further consider a function $F(x)$, defined in the interval $(0, \Lambda)$ by the conditions

$$F(x) = f_\mu^0(x - \mu + 1) \text{ for } \mu - 1 < x < \mu \qquad (\mu = 1, 2, \ldots \Lambda).$$

It is evident that the lines $x = 1, 2, \ldots (\Lambda - 1)$; $y = 1, 2, \ldots (\Lambda - 1)$ generally are singular lines of $H(x, y)$. Let $k_1^0(y)$, $k_2^0(y)$, $\ldots$ $k_\Lambda^0(y)$ be a system of solutions of (117). Then we can define an auxiliary function

$$\Phi(y) = k_\nu^0(y - \nu + 1) \text{ for } \nu - 1 < y < \nu,$$

valid in the interval $(0, \Lambda)$.

From the equations (117) we obtain

$$\sum_{\nu=1}^{\Lambda} \int_{\nu-1}^{\nu} k_\nu^0(y - \nu + 1)\, K_{\mu\nu}^0(x - \mu + 1, y - \nu + 1)\, dy = f_\mu^0(x - \mu + 1)$$

which can also be written

$$\int_0^{\Lambda} \Phi(y)\, H(x, y)\, dy = F(x). \tag{118}$$

Having once solved this equation, we obtain the solution of (117) by means of the relations

$$k_\nu(\tau) = k_\nu^0(y) = \Phi\left(\frac{\tau}{T} + \nu - 1\right) \qquad (\nu = 1, 2, \ldots \Lambda). \tag{119}$$

d) Discussion of the solutions. Short summary of results

In the foregoing sections we have assumed that one of the noise components follows the message into the transmission system. Thereby we succeded in avoiding the unnatural solutions $k_\nu(\tau) \equiv 0$ for all ν. The question concerning the solutions of equations (114) or (115) when the kernels are arbitrary functions is a delicate mathematical problem which cannot be treated in the frame of this paper. This is also by no means necessary for our purpose. As a matter of fact I^2 is always ≥ 0 and varies continuously with the transfer coefficients which are always supposed to be finite (in other cases the output would become infinite according to (98)). It is evident that I^2 need not necessarily have a minimum within the range of the transfer coefficients that can be used for practical purposes, but if a minimum exists within this range, we will find it with the methods of sections a–c). Generally we have the possibility to judge of the existence of minimum solutions from the intrinsic physical properties of the system.

Now a few words about the uniqueness of the solution. The system of integral equations (115) can be transformed into a system of linear algebraic equations, as was shown in section a) in case of two noise components, the number of equations of the last system being the same as the number of values of the transfer coefficients, i.e. Λ times the number of time intervals. For each (arbitrarily small) length of the time interval I^2 is a positively definite quadratic form of the transfer coefficients. Thus the algebraic system has always one, and only one, solution, and this solution must approximately satisfy the system (114).

Probably the reader has asked himself what practical use we could have of the formulas derived here. I do not think time is wasted by discussing this question a little. Having once determined the transfer coefficients we have to construct a transmission system with the properties of these transfer coefficients. As I have pointed out before it is not necessary to consider the transfer functions, since the system is as well characterized by the transfer coefficients as by the transfer functions. In most practical noise problems there are some parameters allowed to vary within certain limits. These parameters can be calculated by the method of least squares. Many times the system corresponding to the minimum rms error conditions may not be realizable for practical reasons, but then we have to do the best we can with the situation. The way to do this is highly dependent on the special circumstances and can hardly be argued about in a general way.

e) Some considerations about linear integral equations of the first kind

Some pages now will be devoted to a method by means of which we are able to give a practical solution of the integral equation

$$\int_0^T k(\tau)\, K(\sigma, \tau)\, d\tau = f(\sigma) \tag{116}$$

where $K(\sigma, \tau)$ and $f(\sigma)$ are given functions and $k(\tau)$ is to be determined. This equation appears, according to what is said in section c) of this chapter, in almost all linear noise problems. I think the method described here can be used in most practical cases. The only (almost self-evident) restriction is that $K(\sigma, \tau)$ must be integrable in the sense of Riemann.

We divide the time $(0, T)$ into N parts with the length $\Delta\tau$ and the division points $\tau_0 = 0, \tau_1, \ldots \tau_N = T$. These division points will be used also for the variable σ.[1] In each interval $k(\tau)$ may be approximated by a polynomial. Here we assume that the approximation curves are straight lines

$$k(\tau) = k_\mu + \frac{k_{\mu+1} - k_\mu}{\Delta\tau}(\tau - \tau_\mu)$$

$$(\tau_\mu \leq \tau \leq \tau_{\mu+1};\quad \mu = 0, 1, \ldots \overline{N-1})$$

or more conveniently written

$$k(\tau) = k_\mu \frac{\tau_{\mu+1} - \tau}{\Delta\tau} + k_{\mu+1} \frac{\tau - \tau_\mu}{\Delta\tau}. \tag{120}$$

It is obvious that the right values of $k(\tau)$ could be obtained from the k_μ:s by some sort of smoothing process, but since we cannot get hold of this process, we should at least have a method by means of which we are able to determine the maximum difference between the right and the approximate values of k. A possibility to avoid this question is to use different sets of division points:

$$\{{}_1\tau_\mu\},\ \{{}_2\tau_\mu\},\ \ldots$$

with for instance ${}_2\Delta\tau = \frac{1}{2}\,{}_1\Delta\tau$, ${}_3\Delta\tau = \frac{1}{2}\,{}_2\Delta\tau$, ... Evidently, if equation (116) has one and only one solution, this process always converges. On the other hand, if the process converges, $\lim k_\mu$ gives a solution of (116). It remains to consider the case of more than one solution and the case of no solution. We will be concerned a little with this question below.

Firstly we introduce the notations

$$\left.\begin{aligned} a_\mu(\sigma) &= \int_{\tau_\mu}^{\tau_{\mu+1}} \frac{\tau_{\mu+1} - \tau}{\Delta t} K(\sigma, \tau)\, d\tau \\ b_\mu(\sigma) &= \int_{\tau_\mu}^{\tau_{\mu+1}} \frac{\tau - \tau_\mu}{\Delta\tau} K(\sigma, \tau)\, d\tau \end{aligned}\right\}$$

[1] Observe that σ and τ belong to the same time interval.

It follows

$$a_\mu(\sigma) + b_\mu(\sigma) = \int_{\tau_\mu}^{\tau_{\mu+1}} K(\sigma, \tau)\, d\tau.$$

If, as a special case, the kernel $K(\sigma, \tau)$ can be approximated by straight lines, we have the simple formulas

$$\left.\begin{aligned} a_\mu(\sigma) &= [\tfrac{1}{3} K(\sigma, \tau_\mu) + \tfrac{1}{6} K(\sigma, \tau_{\mu+1})]\, \Delta\tau \\ b_\mu(\sigma) &= [\tfrac{1}{6} K(\sigma, \tau_\mu) + \tfrac{1}{3} K(\sigma, \tau_{\mu+1})]\, \Delta\tau \end{aligned}\right\}$$

With the aid of the functions a_μ and b_μ and the relation (120) equation (116) becomes

$$\sum_{\mu=0}^{N-1} [a_\mu(\sigma) \cdot k_\mu + b_\mu(\sigma) \cdot k_{\mu+1}] = f(\sigma). \tag{121}$$

$$(\sigma = \tau_0, \tau_1, \ldots \tau_N)$$

This is a linear system with $N + 1$ equations and $N + 1$ unknown quantities $k_0, k_1, \ldots k_N$. Disregarding the case where the equations (121) imply an absurdity for all values of N (then equation (116) has no solution) we have to consider two possibilities. If the determinant Δ of the system is $\neq 0$ for all $\Delta\tau$, equation (116) has only one solution. In the case of $\Delta = 0$ the system (121) gives an infinity of solutions from which one is to be taken according to certain side conditions.

We are now going to set up recursion formulas for the computation of a_μ and b_μ for one interval division from another. Hereby it is convenient to begin with the smallest interval which we intend to consider and from this calculate the auxiliary quantities for greater and greater intervals. In this way we obtain, assuming that we are going from $2n$ to n intervals and putting $\tau_{\mu+\frac{1}{2}} = \frac{1}{2}(\tau_\mu + \tau_{\mu+1})$,

$$\begin{aligned} {}_n a_\mu &= \int_{\tau_\mu}^{\tau_{\mu+1}} \frac{\tau_{\mu+1} - \tau}{{}_n\Delta\tau} K\, d\tau = \frac{\tau_{\mu+1}}{{}_n\Delta\tau} \int_{\tau_\mu}^{\tau_{\mu+1}} K\, d\tau - \frac{1}{{}_n\Delta\tau} \int_{\tau_\mu}^{\tau_{\mu+1}} \tau K\, d\tau = \\ &= \frac{\tau_{\mu+1}}{2\,{}_{2n}\Delta\tau} \int_{\tau_\mu}^{\tau_{\mu+\frac{1}{2}}} K\, d\tau + \frac{\tau_{\mu+1}}{2\,{}_{2n}\Delta\tau} \int_{\tau_{\mu+\frac{1}{2}}}^{\tau_{\mu+1}} K\, d\tau - \frac{1}{2\,{}_{2n}\Delta\tau} \int_{\tau_\mu}^{\tau_{\mu+\frac{1}{2}}} \tau K\, d\tau - \frac{1}{2\,{}_{2n}\Delta\tau} \int_{\tau_{\mu+\frac{1}{2}}}^{\tau_{\mu+1}} \tau K\, d\tau = \\ &= \frac{\tau_{\mu+1} - \tau_{\mu+\frac{1}{2}}}{2\,{}_{2n}\Delta\tau} \int_{\tau_\mu}^{\tau_{\mu+\frac{1}{2}}} K\, d\tau + \frac{1}{2}\,{}_{2n}a_\mu + \frac{1}{2}\,{}_{2n}a_{\mu+\frac{1}{2}} = {}_{2n}a_\mu + \frac{1}{2}\,{}_{2n}a_{\mu+\frac{1}{2}} + \frac{1}{2}\,{}_{2n}b_\mu. \end{aligned}$$

In this formula τ_μ means the same time point for both interval divisions. However, to the μ:th point of the great-interval scale ought to correspond the

2μ:th point of the small-interval scale. As a result of that we write, employing the same reasoning for b_μ,

$$\left.\begin{aligned} {}_n a_\mu &= {}_{2n}a_{2\mu} + \tfrac{1}{2}\,{}_{2n}a_{2\mu+1} + \tfrac{1}{2}\,{}_{2n}b_{2\mu} \\ {}_n b_\mu &= \tfrac{1}{2}\,{}_{2n}b_{2\mu} + {}_{2n}b_{2\mu+1} + \tfrac{1}{2}\,{}_{2n}a_{2\mu+1} \end{aligned}\right\} \tag{122}$$

This is no place for a complete mathematical discussion of the validity of the method and the existence of solutions. However, it may be valuable to give a few notes concerning a special case which has turned out to be of particular interest. This case is defined by the assumptions 1–3 below.

Firstly we assume that *the kernel can be developed in a convergent functional series of the form*[1]

$$K(\sigma, \tau) = \sum_{i=0}^{\infty} \alpha_i(\sigma) \cdot \beta_i(\tau)$$

(*assumption 1*) and write

$$K(\sigma, \tau) = \sum_{i=0}^{N} \alpha_i(\sigma) \cdot \beta_i(\tau) + \varepsilon_N(\sigma, \tau).$$

If the kernel is degenerated, $\varepsilon_N = 0$ for N greater than a certain finite value; in other case $\varepsilon_N \to 0$ when $N \to \infty$. Omitting $\varepsilon_N (\neq 0)$, we can only obtain approximate values of $k(\tau)$ which will be denoted by $k'(\tau)$, but *if* $k(\tau)$ *is limited in* $(0, T)$ (*assumption 2*), the approximation will become better and better the more N increases.

If we put

$$\int_0^T \beta_i(\tau)\, k'(\tau)\, d\tau = x_i \quad (i = 0, 1, \ldots N), \tag{123}$$

the integral equation becomes

$$\sum_{i=0}^{N} \alpha_i(\sigma) \cdot x_i = f(\alpha). \tag{124}$$

This relation which must be valid for all values of σ [2] and all $N > 0$ implies that there must exist an expansion of $f(\sigma)$ in a functional series of $\{\alpha_i(\sigma)\}$, if not, equation (124) would convey an absurdity. We write

$$f(\sigma) = \sum_{i=1}^{N} c_i\, \alpha_i(\sigma) + \eta_N(\sigma).$$

If $\eta_N(\sigma)$ is not equal to zero for N greater than a certain number N', we must assume that $\eta_N \to 0$ when $N \to \infty$ (*assumption 3*).

Putting $x_i = c_i$ we commit no error on x_i. On the other hand, omitting the quantity $\eta_N(\sigma)$ implies a reduction of the right member of equation (116) and must have influence on the quantities $k'(\tau)$. The total error of equation (116) becomes

[1] As shown in the theory of functional series this is possible under very general conditions.
[2] At least within a certain interval.

$$\left| \eta_N(\sigma) - \int_0^T \varepsilon_N(\sigma, \tau)\, k(\tau)\, d\tau \right|.$$

Since, according to assumption 2, the integral

$$\int_0^T k(\tau)\, d\tau$$

converges, the error tends to zero when $N \to \infty$. Then, for $N \to \infty$, a solution of (123) must become a solution of (116.)

As the last step we have to consider the solution $k'(\tau)$ of the relations (123), thereby using the approximate expression (120) and the auxiliary quantities

$$\left.\begin{aligned} a_{\mu i} &= \int_{\tau_\mu}^{\tau_{\mu+1}} \beta_i(\tau) \frac{\tau_{\mu+1} - \tau}{\Delta\tau}\, d\tau \\ b_{\mu i} &= \int_{\tau_\mu}^{\tau_{\mu+1}} \beta_i(\tau) \frac{\tau - \tau_\mu}{\Delta\tau}\, d\tau \end{aligned}\right\}$$

One finds

$$\sum_{\mu=0}^{N-1} (a_{\mu i} \cdot k'_\mu + b_{\mu i} \cdot k'_{\mu+1}) = c_i \quad (i = 0, 1, \ldots N). \tag{125}$$

Note that these equations are independent of σ. Here lies a difference with the equation (121) where the coefficients are functions of σ.

Evidently the conditions given above are not enough for the existence of a solution of (125). As shown in the theory of linear equations there must also exist certain relations between the quantities $a_{\mu i}$, $b_{\mu i}$ and c_i. However, I shall not deal with this question here.

In order to see how the method works let us apply it on some simple equations.

1) For the equation

$$\int_0^1 k(\tau) \frac{d\tau}{\sigma^2 + \tau^2} = 1 - \sigma \operatorname{arc\,tg} \frac{1}{\sigma}$$

the method cannot be used when $\sigma = 0$, since

$$\int_0^1 \frac{d\tau}{\sigma^2 + \tau^2}$$

does not converge in this case. As a consequence of that $a_0(0)$ and $b_0(0)$ become infinite.

2) On the contrary, in the case

$$\int_0^1 k(\tau)\,(\sigma^2+\tau^2)\,d\tau = \frac{\sigma^2}{3}+\frac{1}{5}$$

the kernel is integrable in (0, 1). This equation has at least the solution $k(\tau)=\tau^2$.

After some simple calculations we obtain

$$\begin{cases} a_0(\sigma) = \frac{1}{2}\sigma^2\,\Delta\tau + \frac{1}{12}\Delta\tau^3 \\ a_{\mu+1}(\sigma)+b_\mu(\sigma) = \sigma^2\,\Delta\tau+\tau_\mu^2\,\Delta\tau+2\,\tau_\mu\,\Delta\tau^2+\frac{7}{6}\Delta\tau^3 \quad (\mu=0,1,\ldots\overline{N-2}) \\ b_{N-1}(\sigma) = \frac{1}{2}\sigma^2\,\Delta\tau+\frac{1}{2}\Delta\tau-\frac{1}{3}\Delta\tau^2+\frac{1}{12}\Delta\tau^3 \end{cases}$$

Thus the equations (121) become

$$\left(\frac{1}{2}\sigma^2\,\Delta\tau+\frac{1}{12}\Delta\tau^3\right)k_0+\sum_{\mu=0}^{N-2}\left(\sigma^2\,\Delta\tau+\tau_\mu^2\,\Delta\tau+2\,\tau_\mu\,\Delta\tau^2+\frac{7}{6}\Delta\tau^3\right)k'_{\mu+1}+$$

$$+\left(\frac{1}{2}\sigma^2\,\Delta\tau+\frac{1}{2}\Delta\tau-\frac{1}{3}\Delta\tau^2+\frac{1}{12}\Delta\tau^3\right)k_N=\frac{\sigma^2}{3}+\frac{1}{5}$$

$$(\sigma=\tau_0=0,\ \tau_1,\ \ldots\ \tau_N=1)$$

which also can be written

$$\frac{1}{2}k_0\,\sigma^2\,\Delta\tau+\sum_{\mu=0}^{N-2}k_{\mu+1}(\sigma^2+\tau_\mu^2)\,\Delta\tau+\frac{1}{2}k_N(\sigma^2+1)\,\Delta\tau+R_N=\frac{\sigma^2}{3}+\frac{1}{5}$$

where

$$R_N=\frac{1}{12}k_0\,\Delta\tau^3+\sum_{\mu=0}^{N-2}k_{\mu+1}\left(2\,\tau_\mu\,\Delta\tau^2+\frac{7}{6}\Delta\tau^3\right)+k_N\left(-\frac{1}{3}\Delta\tau^2+\frac{1}{12}\Delta\tau^3\right).$$

If k_μ $(\mu=0,1,\ldots N)$ is assumed to be finite, the remainder expression R_N tends to zero, when $\Delta\tau\to 0$.

It is easily seen that the problem has an infinite number of solutions. As a matter of fact, this appears always when the kernel contains no singularities. For $N=4\,(\Delta\tau=0.25)$ four of the quantities $k_0,\ldots k_4$ can be chosen arbitrarily; the fifth is given by the relation

$$k_4 = 2.314\,120-0.548\,022\,k_0-1.163\,840\,k_1-1.367\,229\,k_2-1.706\,212\,k_3.$$

If we take the values of k_0, k_1, k_2, k_3 that corresponds to the solution $k(\tau)=\tau^2$, i.e.

$$k_0=0,\quad k_1=0.0625,\quad k_2=0.25,\quad k_3=0.5625,$$

we obtain $k_4 = 0.94$ (the right value is $k_4 = 1$). When N increases, more and more arbitrariness will be involved in the equations. Now, let $N \to \infty$, i.e. $\Delta\tau \to 0$, and put

$$\int_0^1 k(\tau)\, d\tau = \int_0^1 \tau^2\, d\tau = \frac{1}{3}.$$

Then we have

$$\lim_{\substack{\Delta\tau \to 0 \\ N \to \infty}} \sum_{\mu=0}^{N-1} k_\mu \tau_\mu^2 \Delta\tau = \frac{1}{5}$$

or

$$\int_0^1 k(\tau)\, \tau^2\, d\tau = \frac{1}{5}.$$

Evidently this equation has an infinite number of solutions which may contain an arbitrary number of discontinuities.

3) To show that the number of solutions will be diminished when discontinuities appear in the kernel we consider the equation

$$\int_0^1 k(\tau)\, K(\sigma, \tau)\, d\tau = \sigma^2$$

with

$$K(\sigma, \tau) = \begin{cases} 0 \text{ for } \sigma < \tau \\ \sigma \text{ for } \sigma \geq \tau \end{cases}$$

One finds

$$a_\mu(\tau_\nu) = \begin{cases} 0 \text{ for } \tau_\nu \leq \tau_\mu \\ \tfrac{1}{2}\nu\, \Delta\tau^2 \text{ for } \tau_\nu > \tau_\mu \text{ i.e. } \geq \tau_{\mu+1} \end{cases}$$

$$b_\mu(\tau_\nu) = \begin{cases} 0 \text{ for } \tau_\nu \leq \tau_\mu \\ \tfrac{1}{2}\nu\, \Delta\tau^2 \text{ for } \tau_\nu > \tau_\mu \text{ i.e. } \geq \tau_{\mu+1} \end{cases}$$

$$a_{\mu+1}(\tau_\nu) + b_\mu(\tau_\nu) = \begin{cases} 0 \text{ for } \tau_\nu \leq \tau_\mu \\ \tfrac{1}{2}\nu\, \Delta\tau^2 \text{ for } \tau_\nu = \tau_{\mu+1} \\ \nu\, \Delta\tau^2 \text{ for } \tau_\nu > \tau_{\mu+1} \end{cases}$$

Thus the integral equation will be replaced by the following system of linear equations

$$\begin{cases} k_0 + k_1 = 2 \\ k_0 + 2k_1 + k_2 = 4 \\ k_0 + 2k_1 + 2k_2 + k_3 = 6 \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ k_0 + 2k_1 + \cdots + 2k_{N-1} + k_N = 2N \end{cases}$$

Here we can choose only one of the k-values arbitrarily in order to obtain the others fixed. Putting $k_0 = 1$ we have $k \equiv 1$ which satisfies the given equation.

f) Some examples of linear noise problems

Example 1. The first example which we are going to study is not directly concerned with the formulas given in sections a–d) of this chapter, but in spite of that I think it will be useful for many reasons. We consider the simple differential equation

$$\ddot{y} + a\dot{y} + by = x(t) \tag{126}$$

where $x(t)$ means a random time function given in the form of a uniformly convergent Fourier series

$$x(t) = \tfrac{1}{2}A_0 + \sum_{\nu=1}^{\infty} A_\nu \cos\left(\frac{2\pi\nu t}{L} + \Phi_\nu\right), \tag{127}$$

valid for $t > 0$, while a and b are two constants which are to be determined in some way or other. For $t \leq 0$ we put $x(t) = 0$. L is a positive quantity which may be chosen arbitrarily. If $t < L'$, $x(t)$ is independent of what value $\geq L'$ we take for L. The case $x(t) \equiv 0$ corresponds to the undisturbed movement. This movement is supposed here to be zero.

We assume that the amplitudes A_ν of the series (127) be finite and approximately normally distributed with the mean values m_ν and the standard deviations σ_ν. The distributions of the phases Φ_ν are supposed to be uniform for each turn. Thus we have

$$M\cos\left(\frac{2\pi\tau t}{L} + \Phi_\nu\right) = \frac{1}{2\pi}\int_0^{2\pi} \cos\left(\frac{2\pi\nu t}{L} + \Phi\right) d\Phi = 0,$$

$$M\cos^2\left(\frac{2\pi\nu t}{L} + \Phi_\nu\right) = \frac{1}{2\pi}\int_0^{2\pi} \cos^2\left(\frac{2\pi\nu t}{L} + \Phi\right) d\Phi = \tfrac{1}{2}.$$

Finally we presume that there is *no correlation between the different amplitudes or phases nor between an arbitrary amplitude and an arbitrary phase.* Then we have

$$M\,x(t) = \tfrac{1}{2}m_0.$$

The conditions laid down above can be realized, at least approximately, in many radio circuits.

Let p and q be the roots of the characteristic equation

$$s^2 + as + b = 0.$$

For the sake of stability p and q must have real parts < 0.

According to (74: b) for $\Delta t = 0$

$$k(\tau) = \frac{1}{pq}\left[\frac{qe^{p\tau} - pe^{q\tau}}{p - q} + 1\right]$$

In the case that we have to consider now it is better not to use the limit of (74: a) for $\Delta t = 0$, because it is a difficult problem to determine the derivatives of the random variables A and Φ, but to integrate this limit by parts. According to the assumption that $x(0) = 0$ we then find

$$y(t) = \int_0^t x(t-\tau)\, d\, k(\tau).$$

Introducing the expression of $k(\tau)$ we obtain

$$y(t) = \int_0^t x(t-\tau) \frac{e^{p\tau} - e^{q\tau}}{p-q} d\tau.$$

Since the expansion (127) is assumed to be uniformly convergent, we can integrate term by term and thus get

$$y(t) = \frac{A_0}{2} \int_0^t \frac{e^{p\tau} - e^{q\tau}}{p-q} d\tau + \sum_{\nu=1}^{\infty} A_\nu \int_0^t \cos\left[\frac{2\pi\nu(t-\tau)}{L} + \Phi_\nu\right] \frac{e^{p\tau} - e^{q\tau}}{p-q} d\tau.$$

The integrals become:

$$\int_0^t \cos\left[\frac{2\pi\nu(t-\tau)}{L} + \Phi_\nu\right] e^{p\tau}\, d\tau = \frac{2\pi\nu}{L h_{\nu p}} \sin\left(\frac{2\pi\nu t}{L} + \Phi_\nu\right) -$$

$$- \frac{p}{h_{\nu p}} \cos\left(\frac{2\pi\nu t}{L} + \Phi_\nu\right) - e^{pt}\left[\frac{2\pi\nu}{L h_{\nu p}} \sin\Phi_\nu - \frac{p}{h_{\nu p}} \cos\Phi_\nu\right]$$

$$(\nu = 1, 2, \ldots)$$

where

$$h_{\nu p} = \left(\frac{2\pi\nu}{L}\right)^2 + p^2$$

and similarly for the parameter q. Hence

$$\left.\begin{aligned} y(t) = {} & \frac{A_0}{2pq}\left[\frac{q e^{pt} - p e^{qt}}{p-q} + 1\right] + \\ & + \frac{1}{p-q} \sum_{\nu=1}^{\infty} \frac{2\pi\nu}{L} A_\nu \left(\frac{1}{h_{\nu p}} - \frac{1}{h_{\nu q}}\right) \sin\left(\frac{2\pi\nu t}{L} + \Phi_\nu\right) - \\ & - \frac{1}{p-q} \sum_{\nu=1}^{\infty} A_\nu \left(\frac{p}{h_{\nu p}} - \frac{q}{h_{\nu q}}\right) \cos\left(\frac{2\pi\nu t}{L} + \Phi_\nu\right) - \\ & - \frac{1}{p-q} \sum_{\nu=1}^{\infty} \left[\frac{2\pi\nu}{L} A_\nu \left(\frac{e^{pt}}{h_{\nu p}} - \frac{e^{qt}}{h_{\nu q}}\right) \sin\Phi_\nu - A_\nu \left(\frac{p e^{pt}}{h_{\nu p}} - \frac{q e^{qt}}{h_{\nu q}}\right) \cos\Phi_\nu\right] \end{aligned}\right\} \quad (128)$$

From this formula follows, according to the foregoing assumptions,

$$M\,y(t) = \frac{m_0}{2\,p\,q}\left[\frac{q\,e^{pt} - p\,e^{qt}}{p-q} + 1\right]. \tag{129}$$

We now have to study

$$D^2\,y(t) = M\,[y(t) - M\,y(t)]^2.$$

This expression will contain the following mean values:

$$M\,\sin^2(c_\nu t + \Phi_\nu) = M\,\cos^2(c_\nu t + \Phi_\nu) = \tfrac{1}{2},$$

$$M\,\sin\,(c_\nu t + \Phi_\nu)\cos(c_\nu t + \Phi_\nu) = 0,$$

$$M\,\sin\,(c_\nu t + \Phi_\nu)\sin\Phi_\nu = \tfrac{1}{2}\cos c_\nu t,$$

$$M\,\sin\,(c_\nu t + \Phi_\nu)\cos\Phi_\nu = \tfrac{1}{2}\sin c_\nu t,$$

$$M\,\cos\,(c_\nu t + \Phi_\nu)\sin\Phi_\nu = -\tfrac{1}{2}\sin c_\nu t,$$

$$M\,\cos\,(c_\nu t + \Phi_\nu)\cos\Phi_\nu = \tfrac{1}{2}\cos c_\nu t.$$

In case of different indices of Φ in the two trigonometric factors the mean values become zero. Thus we obtain

$$\begin{aligned} D^2\,y(t) = {} & \frac{\sigma_0^2}{4\,p^2\,q^2}\left[\frac{q\,e^{pt} - p\,e^{qt}}{p-q} + 1\right]^2 + \\ & + \frac{1}{2\,(p-q)^2}\sum_{\nu=1}^{\infty}(\sigma_\nu^2 + m_\nu^2)\left[\left(\frac{2\pi\nu}{L}\right)^2\left(\frac{1}{h_{\nu p}} - \frac{1}{h_{\nu q}}\right)^2 + \left(\frac{p}{h_{\nu p}} - \frac{q}{h_{\nu q}}\right)^2 + \right. \\ & + \left(\frac{2\pi\nu}{L}\right)^2\left(\frac{e^{pt}}{h_{\nu p}} - \frac{e^{qt}}{h_{\nu q}}\right)^2 + \left(\frac{p\,e^{pt}}{h_{\nu p}} - \frac{q\,e^{qt}}{h_{\nu q}}\right)^2 - \\ & - 2\left(\frac{2\pi\nu}{L}\right)^2\left(\frac{1}{h_{\nu p}} - \frac{1}{h_{\nu q}}\right)\left(\frac{e^{pt}}{h_{\nu p}} - \frac{e^{qt}}{h_{\nu q}}\right)\cos\frac{2\pi\nu t}{L} + \\ & + 2\cdot\frac{2\pi\nu}{L}\left(\frac{1}{h_{\nu p}} - \frac{1}{h_{\nu q}}\right)\left(\frac{p\,e^{pt}}{h_{\nu p}} - \frac{q\,e^{qt}}{h_{\nu q}}\right)\sin\frac{2\pi\nu t}{L} - \\ & - 2\cdot\frac{2\pi\nu}{L}\left(\frac{p}{h_{\nu p}} - \frac{q}{h_{\nu q}}\right)\left(\frac{e^{pt}}{h_{\nu p}} - \frac{e^{qt}}{h_{\nu q}}\right)\sin\frac{2\pi\nu t}{L} - \\ & \left. - 2\left(\frac{p}{h_{\nu p}} - \frac{q}{h_{\nu q}}\right)\left(\frac{p\,e^{pt}}{h_{\nu p}} - \frac{q\,e^{qt}}{h_{\nu q}}\right)\cos\frac{2\pi\nu t}{L}\right]. \end{aligned} \tag{130}$$

It is worth while to observe that notations of mean values and standard deviations concerning the phase angles do not appear in formula (130). This depends of course on the special assumptions that we have made about the phase distributions.

The coefficient of $(\sigma_\nu^2 + m_\nu^2)$ in (130) can be developed in powers of $\frac{2\pi\nu}{L}$ which development converges at the same time as

$$\frac{1}{h_{\nu p}} = \frac{1}{\left(\frac{2\pi\nu}{L}\right)^2 + p^2} = \frac{1}{p^2}\left[1 - \left(\frac{2\pi\nu}{pL}\right)^2 + \left(\frac{2\pi\nu}{pL}\right)^4 - \cdots\right]$$

and the corresponding series of $\frac{1}{h_{\nu q}}$, i.e. for

$$\left|\frac{2\pi\nu}{pL}\right| < 1$$

and

$$\left|\frac{2\pi\nu}{qL}\right| < 1.$$

From these two inequalities follows

$$\nu < \frac{L|p|}{2\pi}, \quad \nu < \frac{L|q|}{2\pi}.$$

Thus for a given finite L we cannot use the expansions for such values of ν which are greater than the least of the numbers $L|p|/2\pi$ and $L|q|/2\pi$. If we let L tend to ∞, the expansions become valid for all ν, but then, of course, we have no use of them. In this case ($L = \infty$) $D^2 y(t)$ becomes

$$D^2 y(t) = \frac{1}{2p^2q^2}\left[\frac{q e^{pt} - p e^{qt}}{p - q} + 1\right]^2 \left[\frac{1}{2}\sigma_0^2 + \sum_{\nu=1}^{\infty}(\sigma_\nu^2 + m_\nu^2)\right]. \tag{131}$$

The expansion (131) converges at the same time as $\sum(\sigma_\nu^2 + m_\nu^2)$. Let us for for instance take

$$A_\nu = \frac{A}{1 + k\nu^2} \quad (k > 0)$$

with $MA = m$, $DA = \sigma$. Then

$$\sum_{\nu=1}^{\infty}(\sigma_\nu^2 + m_\nu^2) = (\sigma^2 + m^2)\sum_{\nu=1}^{\infty}\frac{1}{(1 + k\nu^2)^2}.$$

This series is obviously convergent. Employing frequencies up to $\nu = N$ we disregard the remainder term

$$R_N \approx (\sigma^2 + m^2)\int_N^{\infty}\frac{d\nu}{(1 + k\nu^2)^2} = \left[\frac{\pi}{4\sqrt{k}} - \frac{\operatorname{arc\,tg} N\sqrt{k}}{2\sqrt{k}} - \frac{N}{2(1 + kN^2)}\right](\sigma^2 + m^2).$$

We now have to consider the mean dispersion during the time $(0, T)$, defined by

$$I^2 = \frac{1}{T}\int_0^T D^2 y(t)\, dt,$$

but since the calculations are trivial and contain nothing of interest, I exclude them here. I only give some notes concerning the limit case $L = \infty$.

The only critical points of I^2 are $p = q$. In this case we obtain for $L = \infty$

$$D^2 y(t) = \frac{1}{2p^4}[1 + (pt - 1)e^{pt}]^2 \left[\frac{1}{2}\sigma_0^2 + \sum_{\nu=1}^{\infty}(\sigma_\nu^2 + m_\nu^2)\right]$$

and

$$I^2 = \frac{1}{2p^4}\left[1 + 2e^{pT} + \frac{4}{pT}(1 - e^{pT}) + \frac{pT}{2}e^{2pT} - \frac{3}{2}e^{2pT} - \frac{5}{4pT}(1 - e^{2pT})\right] \cdot$$

$$\cdot \left[\frac{1}{2}\sigma_0^2 + \sum_{\nu=1}^{\infty}(\sigma_\nu^2 + m_\nu^2)\right] \cdot$$

This expression has obviously no minimum, for $I^2 \to 0$ for increasing $|p|$. It is easily seen that this statement is true also for $p \neq q$. The more remote the roots are situated from the origin and from each other, the less $D^2 y(t)$ and, since $D^2 y(t) > 0$ for all t, also the rms error I becomes.

Example 2. In many noise problems the derivatives of a random function appear at the same time as the random function itself. Let us for instance take the equation

$$\dddot{y} + a\ddot{y} + b\dot{y} = k_1 \dot{z}(t) + k_2 z(t) \tag{132}$$

where the left member is the same as in example 1 but where the right member is a linear function of the disturbing function and its first derivative instead of the disturbing function alone.

We are going to prove that the probability distribution of

$$x(t) = k_1 \dot{z}(t) + k_2 z(t)$$

is of a form very similar to that of $z(t)$, when $z(t)$ is distributed as $x(t)$ in example 1. Assuming that

$$z(t) = \frac{1}{2}B_0 + \sum_{\nu=1}^{\infty} B_\nu \cos\left(\frac{2\pi\nu t}{L} + \Psi_\nu\right) \tag{133}$$

where the amplitudes B_ν are finite and approximately normally and the phase angles ψ_ν uniformly distributed and that

$$x(t) = \frac{1}{2}A_0 + \sum_{\nu=1}^{\infty} A_\nu \cos\left(\frac{2\pi\nu t}{L} + \Phi_\nu\right), \tag{134}$$

its distribution being unknown, we must have the identity

$$\frac{1}{2}A_0 + \sum_{\nu=1}^{\infty} A_\nu \cos\left(\frac{2\pi\nu t}{L} + \Phi_\nu\right) \equiv$$

$$\equiv \frac{1}{2}k_2 B_0 + \sum_{\nu=1}^{\infty} k_2 B_\nu \cos\left(\frac{2\pi\nu t}{L} + \Psi_\nu\right) - \sum_{\nu=1}^{\infty} \frac{2\pi\nu k_1}{L} B_\nu \sin\left(\frac{2\pi\nu t}{L} + \Psi_\nu\right).$$

This identity can also be written

$$\frac{1}{2}A_0 + \sum_{\nu=1}^{\infty} A_\nu \cos\Phi_\nu \cos\frac{2\pi\nu t}{L} - \sum_{\nu=1}^{\infty} A_\nu \sin\Phi_\nu \sin\frac{2\pi\nu t}{L} \equiv$$

$$\equiv \frac{1}{2}k_2 B_0 + \sum_{\nu=1}^{\infty} B_\nu \left(k_2 \cos\Psi_\nu - \frac{2\pi\nu k_1}{L}\sin\Psi_\nu\right)\cos\frac{2\pi\nu t}{L} -$$

$$- \sum_{\nu=1}^{\infty} B_\nu \left(k_2 \sin\Psi_\nu + \frac{2\pi\nu k_1}{L}\cos\Psi_\nu\right)\sin\frac{2\pi\nu t}{L}$$

and can be satisfied only if

$$\begin{cases} A_0 = k_2 B_0 \\ A_\nu \cos\Phi_\nu = \left(k_2 \cos\Psi_\nu - \dfrac{2\pi\nu k_1}{L}\sin\Psi_\nu\right) B_\nu \quad (\nu \geq 1) \\ A_\nu \sin\Phi_\nu = \left(k_2 \sin\Psi_\nu + \dfrac{2\pi\nu k_1}{L}\cos\Psi_\nu\right) B_\nu \quad (\nu \geq 1) \end{cases}$$

From this follows

$$A_\nu = B_\nu \sqrt{k_2^2 + \left(\frac{2\pi\nu k_1}{L}\right)^2} \tag{135}$$

which proves that A_ν and B_ν obey the same distribution law, i.e. they are both approximately normal. If the dispersion of B_ν is denoted by s_ν, we have

$$\sigma_\nu^2 = s_\nu^2 \left[k_2^2 + \left(\frac{2\pi\nu k_1}{L}\right)^2\right]. \tag{136}$$

If we introduce the auxiliary quantities α_ν and ϱ_ν, defined by

$$\varrho_\nu \cos\alpha_\nu = \frac{k_2}{\sqrt{k_2^2 + \left(\frac{2\pi\nu k_1}{L}\right)^2}},$$

$$\varrho_\nu \sin\alpha_\nu = \frac{\frac{2\pi\nu k_1}{L}}{\sqrt{k_2^2 + \left(\frac{2\pi\nu k_1}{L}\right)^2}},$$

we obtain

$$\varrho_\nu = 1,$$

$$\operatorname{tg} \alpha_\nu = \frac{2\pi\nu k_1}{L k_2}$$

and

$$\cos \Phi_\nu = \cos (\Psi_\nu + \alpha_\nu),$$

$$\sin \Phi_\nu = \sin (\Psi_\nu + \alpha_\nu).$$

It follows that if Ψ_ν is uniformly distributed, so is Φ_ν and vice versa.

For the rest the reasoning is the same as in example 1.

Example 3. As a third example we take the differential equation

$$\ddot{y} + a\dot{y} + by = x(t) + v(t) \tag{137}$$

where $x(t)$ means the time function of a "message" and $v(t)$ the time function of the noise; a and b are two constants which are to be determined in such a way that the effect of the noise on the message will be as small as possible. *We put* $x(0) = 0$, $v(0) = 0$. Both the message and the noise may be given in the form of uniformly convergent Fourier series

$$x(t) = \frac{1}{2} A_0 + \sum_{\nu=1}^{\infty} A_\nu \cos\left(\frac{2\pi\nu t}{L} + \Phi_\nu\right), \tag{138: a}$$

$$v(t) = \frac{1}{2} a_0 + \sum_{\nu=1}^{\infty} a_\nu \cos\left(\frac{2\pi\nu t}{L} + \varphi_\nu\right) \tag{138: b}$$

where L can be chosen arbitrarily (> 0). We assume that the amplitudes be approximately normally distributed, while the distributions of the phases may be uniform. Further we presume that there is no correlation between message and noise or between amplitude and phase for each one of the two signals nor between different amplitudes or different phases, in other words, that there is *no correlation at all.* Finally we put

$$M A_0 = M a_0 = 0, \quad M A_0^2 = M a_0^2 = 0.$$

From the above assumptions follows

$$M x(t) = M v(t) = 0.$$

As in the first example we do not use the derivatives of the signals. In the formulas of VI: c ($\Lambda = 1$) we therefore replace $\frac{dx(t-\sigma)}{d\sigma}$ by $x(t-\sigma)$ and $\frac{dv(t-\tau)}{d\tau}$ by $v(t-\tau)$. Instead of that we have to take the derivative of $k(\tau)$. Apart from these modifications we use the same notations as in VI: c. Thus

$$\left.\begin{aligned} m_{00}(\sigma,\tau,t) &= M[x(t-\sigma)\,x(t-\tau)] = \frac{1}{2}\sum_{\nu=1}^{\infty} M A_\nu^2 \cdot \cos\frac{2\pi\nu(\tau-\sigma)}{L} \\ m_{01}(\varrho,\tau,t) &= m_{10}(\sigma,\tau,t) = 0 \\ m_{11}(\sigma,\tau,t) &= M[v(t-\sigma)\,v(t-\tau)] = \frac{1}{2}\sum_{\nu=1}^{\infty} M a_\nu^2 \cdot \cos\frac{2\pi\nu(\tau-\sigma)}{L} \end{aligned}\right\}$$

From this we see that $m_{00}(\sigma,\tau,t)$ and $m_{11}(\sigma,\tau,t)$ only depend on $\tau-\sigma$. The series are uniformly convergent.

Since the correlation functions $m(\sigma,\tau,t)$ are independent of t, it is easy to determine the functions $\mathcal{M}(\sigma,\tau)$. We find

$$\left.\begin{aligned} \mathcal{M}_{00}(\sigma,\tau) &= \frac{1}{2}(T-[\sigma,\tau])\sum_{\nu=1}^{\infty} M A_\nu^2 \cdot \cos\frac{2\pi\nu(\tau-\sigma)}{L} \\ \mathcal{M}_{01}(\sigma,\tau) &= \mathcal{M}_{10}(\sigma,\tau) = 0 \\ \mathcal{M}_{11}(\sigma,\tau) &= \frac{1}{2}(T-[\sigma,\tau])\sum_{\nu=1}^{\infty} M a_\nu^2 \cdot \cos\frac{2\pi\nu(\tau-\sigma)}{L} \end{aligned}\right\}$$

In this case ($\Lambda = 1$) the system of integral equations (114) is reduced to one single equation

$$\int_0^T K(\sigma,\tau)\,k'(\tau)\,d\tau = f(\sigma)$$

of the type studied in section e) with

$$K(\sigma,\tau) = (T-[\sigma,\tau])\sum_{\nu=1}^{\infty}(M A_\nu^2 + M a_\nu^2)\cos\frac{2\pi\nu(\tau-\sigma)}{L}$$

and

$$f(\sigma) = (T-\sigma)\sum_{\nu=1}^{\infty} M A_\nu^2 \cdot \cos\frac{2\pi\nu\sigma}{L}.$$

The kernel $K(\sigma,\tau)$ is apparently symmetric and has all its singularities on the line $\sigma=\tau$. It is convenient to write it in the form

$$K(\sigma,\tau) = \sum_{\nu=1}^{\infty} \varkappa_\nu[\alpha_{\nu1}(\sigma)\,\beta_{\nu1}(\tau) + \alpha_{\nu2}(\sigma)\,\beta_{\nu2}(\tau)]$$

where

$$\left.\begin{aligned} \varkappa_\nu &= M A_\nu^2 + M a_\nu^2 \\ \alpha_{\nu1}(\sigma) &= (T-\sigma)\cos\frac{2\pi\nu\sigma}{L} \\ \alpha_{\nu2}(\sigma) &= (T-\sigma)\sin\frac{2\pi\nu\sigma}{L} \\ \beta_{\nu1}(\tau) &= \cos\frac{2\pi\nu\tau}{L} \\ \beta_{\nu2}(\tau) &= \sin\frac{2\pi\nu\tau}{L} \end{aligned}\right\} \text{ for } \tau \le \varrho$$

$$\left.\begin{aligned}
\alpha_{\nu 1}(\sigma) &= \cos\frac{2\pi\nu\sigma}{L}\\
\alpha_{\nu 2}(\sigma) &= \sin\frac{2\pi\nu\sigma}{L}\\
\beta_{\nu 1}(\tau) &= (T-\tau)\cos\frac{2\pi\nu\tau}{L}\\
\beta_{\nu 2}(\tau) &= (T-\tau)\sin\frac{2\pi\nu\tau}{L}
\end{aligned}\right\} \text{ for } \tau \geq \sigma$$

If we further put

$$\gamma_\nu(\sigma) = (T-\sigma)\,M\,A_\nu^2 \cdot \cos\frac{2\pi\nu\sigma}{L}$$

and *remember that the necessary conditions for integration term by term are fulfilled,* the integral equation becomes

$$\sum_{\nu=1}^{\infty} \varkappa_\nu \int_0^T [\alpha_{\nu 1}(\sigma)\cdot\beta_{\nu 1}(\tau) + \alpha_{\nu 2}(\sigma)\cdot\beta_{\nu 2}(\tau)]\,k'(\tau)\,d\tau = \sum_{\nu=1}^{\infty}\gamma_\nu(\sigma). \tag{139}$$

This equation can be solved by the method of e). Thus we put

$$\left.\begin{aligned}
c_{\mu\nu i} &= \int_{\tau_\mu}^{\tau_{\mu+1}} \frac{\tau_{\mu+1}-\tau}{\Delta\tau}\,\beta_{\nu i}(\tau)\,d\tau,\\
d_{\mu\nu i} &= \int_{\tau_\mu}^{\tau_{\mu+1}} \frac{\tau-\tau_\mu}{\Delta\tau}\,\beta_{\nu i}(\tau)\,d\tau.
\end{aligned}\right\} \quad (i = 1, 2)$$

For facilitating the solution we further introduce the auxiliary quantities:

$$C_{\mu\nu}(\sigma) = \varkappa_\nu[c_{\mu\nu 1}\,\alpha_{\nu 1}(\sigma) + c_{\mu\nu 2}\,\alpha_{\nu 2}(\sigma)],$$
$$D_{\mu\nu}(\sigma) = \varkappa_\nu[d_{\mu\nu 1}\,\alpha_{\nu 1}(\sigma) + d_{\mu\nu 2}\,\alpha_{\nu 2}(\sigma)].$$

The equation (139) becomes

$$\sum_{\mu=0}^{N-1}\left[k'_\mu\sum_{\nu=1}^{\infty}C_{\mu\nu}(\sigma) + k'_{\mu+1}\sum_{\nu=1}^{\infty}D_{\mu\nu}(\sigma)\right] = \sum_{\nu=1}^{\infty}\gamma_\nu(\sigma). \tag{140}$$

If in this equation we put $\sigma = \tau_0, \tau_1, \ldots \tau_N$, we have a linear system with $N+1$ equation and $N+1$ unknown quantities: $k'_0, k'_1, \ldots k'_N$.

VII. Some problems in the theory of autocorrelation functions and spectral densities

a) Random errors in autocorrelation functions and spectral densities calculated from an empirical material

In the foregoing sections we have been concerned only incidentally with autocorrelation functions (this function is defined in a remark in section V: b), whereas spectral densities have not yet been considered in this treatise. The theory of these concepts is treated very carefully in the servotechnic literature[1] and will not be dealt with here. However, the errors committed by using empirical material by the computation of autocorrelation functions and spectral densities seem not to have been studied so much.

Autocorrelation functions

Suppose that the time function $y(t)$ has been observed for $0 \leq t \leq T$ and that the result is given in the form of an oscillogram. Further we presume that the process can be considered as *stationary*. Then a great deal of information can be obtained from the autocorrelation function

$$R(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} y(t)\, y(t+\tau)\, dt. \tag{141}$$

If T is not too small and τ not too large, we can use the approximation formula

$$R(\tau) \approx R_T(\tau) = \frac{1}{T-\tau} \int_{0}^{T-\tau} y(t)\, y(t+\tau)\, dt \tag{142}$$

or

$$R(\tau) \approx R_N(m) = \frac{1}{N-m+1} \sum_{n=0}^{N-m} y_n\, y_{n+m} \tag{143}$$

where $y_n = y(n\,\Delta t) = y(t)$. We are going to estimate the error committed by the use of these approximation formulas. Thereby it is always assumed *that the mean value of $y(t)$ is zero*. Of course, this does not mean any loss of generality.

By the calculation of $R_N(m)$ instead of $R_T(\tau)$ we commit a *computational error* $E\,R_N(m)$, depending on the fact that the interval Δt has a finite length. This error can easily be estimated and will not be considered here. It is more difficult to master the *random error* represented by the standard deviation $D\,R_N(m)$;

$$D^2 R_N(m) = \frac{1}{(N-m+1)^2} \left\{ M\left(\sum_{n=0}^{N-m} y_n\, y_{n+m} \right)^2 - \left[M\left(\sum_{n=0}^{N-m} y_n\, y_{n+m} \right) \right]^2 \right\}.$$

[1] See for instance the treatise by R. S. PHILLIPS in "Radiation Laboratory Series 25". Many of Phillips' notations are used in this chapter.

Since the process is assumed to be stationary, $M(y_n y_{n+m})$ is independent of n. Thus

$$M\left(\sum_{n=0}^{N-m} y_n y_{n+m}\right) = (N-m+1)\,R(m).$$

When computing $M(\sum y_n y_{n+m})^2$ we have to consider the correlation not only between two quantities y_n and y_{n+m} but also between the products $y_n y_{n+m}$ and $y_{n+p} y_{n+p+m}$. Let the last correlation function be denoted by $R(m, p)$, i.e.

$$R(m, p) = M(y_n y_{n+m} \cdot y_{n+p} y_{n+p+m}) = R(p, m).$$

Then

$$D^2 R_N(m) = \frac{1}{N-m+1}\left\{R(m,0) + \right.$$

$$\left. + 2\sum_{p=1}^{N-m}\left(1 - \frac{p}{N-m+1}\right) R(m,p) - (N-m+1)\,[R(m)]^2\right\}. \qquad (144)$$

As a limit for $\Delta t = 0$ we obtain

$$D^2 R_T(\tau) = \frac{2}{T-\tau}\int_0^{T-\tau}\left(1 - \frac{t}{T-\tau}\right)[R(\tau,t) - R(\tau)^2]\,dt. \qquad (145)$$

The formulas (144) and (145) will now be applied to a couple of hypothetical distributions.

$y(t)$ *normal* $(0, \sigma)$. In this case $R(m, p)$ can be expressed as a function of σ, $R(m)$, $R(p)$, $R(m-p)$ and $R(m+p)$. This follows from the form of the frequency function of the combined variable $\{y_n, y_{n+m}, y_{n+p}, y_{n+p+m}\}$:

$$\varphi(\xi_1, \xi_2, \xi_3, \xi_4) = \frac{1}{(2\pi)^2\sqrt{D}}\, e^{-\frac{1}{2D}[D_{11}\xi_1^2 + D_{22}\xi_2^2 + \cdots + 2D_{12}\xi_1\xi_2 + \cdots]}$$

where

$$D = \begin{vmatrix} \sigma^2 & R(m) & R(p) & R(m+p) \\ R(m) & \sigma^2 & R(m-p) & R(p) \\ R(p) & R(m-p) & \sigma^2 & R(m) \\ R(m+p) & R(p) & R(m) & \sigma^2 \end{vmatrix}$$

and $D_{\mu\nu}$ means the minor of the μ:th row and the ν:th column of D.

In order to determine $R(m, p)$ we introduce the characteristic function

$$g(u_1, u_2, u_3, u_4) = M(e^{j(u_1\xi_1 + \cdots + u_4\xi_4)}) = e^{-\frac{1}{2}[\sigma^2 u_1^2 + \cdots + 2R(m)u_1u_2 + \cdots]}.$$

One finds

$$R(m, p) = \left(\frac{\partial^4 g}{\partial u_1 \partial u_2 \partial u_3 \partial u_4}\right)_{\substack{\mu_\nu = 0 \\ \nu = 1, \ldots 4}} = R(m)^2 + R(m+p)\,R(m-p) + R(p)^2$$

and specially

$$R(m, 0) = 2\,R(m)^2 + \sigma^2.$$

With these expressions we have

$$D^2 R_N(m) = \frac{1}{N-m+1}\left\{R(m)^2 + \sigma^2 + \right.$$

$$\left. + 2\sum_{p=1}^{N-m}\left(1 - \frac{p}{N-m+1}\right)[R(m+p)\,R(m-p) + R(p)^2]\right\} \qquad (146)$$

and

$$D^2 R_T(\tau) = \frac{2}{T-\tau}\int_0^{T-\tau}\left(1 - \frac{t}{T-\tau}\right)[R(\tau+t)\,R(\tau-t) + R(t)^2]\,dt. \qquad (147)$$

Modified normal distribution. Many times the distribution of $y(t)$ is not exactly normal but can be represented by a frequency function of the form

$$f(\xi) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\xi^2}{2\sigma^2}} P(\xi)$$

where $P(\xi)$ is a polynomial in ξ. $D^2 R_N(m)$ can also then be given in terms of the autocorrelation function. Firstly, the frequency function of the combined variable can be written

$$f(\xi_1, \xi_2, \xi_3, \xi_4) = \varphi(\xi_1, \xi_2, \xi_3, \xi_4)\,Q(\xi_1, \xi_2, \xi_3, \xi_4)$$

where $Q(\xi_1, \xi_2, \xi_3, \xi_4)$ means a polynomial of $\xi_1, \xi_2, \xi_3, \xi_4$. The characteristic function of $\{\xi_1, \xi_2, \xi_3, \xi_4\}$ is derived from $g(u_1, u_2, u_3, u_4)$ by operations of differentiation, multiplication by constants and addition. Making further the operation $\dfrac{\partial^4}{\partial u_1 \partial u_2 \partial u_3 \partial u_4}$ we obtain $R(m, p)$. If for instance

$$Q = a_1\xi_1 + a_2\xi_2 + a_3\xi_3 + a_4\xi_4 + b\,\xi_1^3\xi_2^3\xi_3^3\xi_4^3,$$

we have

$$R(m, p) = \frac{\partial^4}{\partial u_1 \partial u_2 \partial u_3 \partial u_4}\left(\frac{1}{j}\sum_{\nu=1}^{4} a_\nu \frac{\partial}{\partial u_\nu} + \frac{1}{j^{12}} b \frac{\partial^{12}}{\partial u_1^3 \partial u_2^3 \partial u_3^3 \partial u_4^3}\right) g(u_1, u_2, u_3, u_4) =$$

$$= b\frac{\partial^{16} g}{\partial u_1^4 \partial u_2^4 \partial u_3^4 \partial u_4^4} \text{ (for } u_1 = u_2 = u_3 = u_4 = 0).$$

There is no point, in this case, to give a compact expression for $D^2 R_N(m)$.

Spectral densities

We define the spectral density of a time function $y(t)$ as

$$G(f) = \lim_{T\to\infty} \frac{1}{T} |A_T(f)|^2 \tag{148: a}$$

where $A_T(f)$ means the Fourier transform of the function

$$y_T(t) = \begin{cases} y(t) & \text{for } -T \leq t \leq T \\ 0 & \text{elsewhere} \end{cases}$$

i.e.

$$A_T(f) = \int_{-\infty}^{\infty} y_T(t)\, e^{-2\pi j f t}\, dt. \tag{148: b}$$

In the case of $G(f)$ being infinite at a frequency f_1 one usually puts

$$G(f_1) = \left[\lim_{T\to\infty} \frac{1}{2T^2} |A_T(f_1)|^2\right] \delta(f - f_1) \tag{149}$$

where $\delta(\ )$ means the Dirac delta function.

One of the most important relations states that

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [y(t)]^2\, dt = \int_0^{\infty} G(f)\, df.$$

In the case of the ergodic hypothesis we obtain

$$M(y^2) = \int_0^{\infty} G(f)\, df.$$

Further, between the input and the output spectral density exists the simple relation

$$G_0(f) = |Y(2\pi j f)|^2 G_i(f),$$

$Y(\ )$ being the transfer function of the system.

The spectral density can be derived by determining the autocorrelation function $R(\tau)$ and then using the well-known formula

$$G(f) = 4 \int_0^{\infty} R(\tau) \cos 2\pi f \tau\, d\tau.$$

This way, however, is not appropriate, if $R(\tau)$ converges slowly, when $\tau \to \infty$.[1] Besides, in most cases it is difficult to judge of the convergence of $R(\tau)$. Therefore it seems preferable to start directly from

[1] It is easily shown that $\lim_{\tau\to\infty} R(\tau) = [My]^2$.

$$G(f) = \lim_{T\to\infty} \frac{1}{T}\left| \int_{-T}^{T} e^{-2\pi i f t}\, y(t)\, dt \right|^2$$

and omit the limiting process. Further the integrals are approximated by sums. Thus the following expression will be studied:

$$G_N(f) = \frac{2\,\Delta t}{N+1}\left[\left(\sum_{\nu=0}^{N} y_\nu \cos 2\pi f t_\nu\right)^2 + \left(\sum_{\nu=0}^{N} y_\nu \sin 2\pi f t_\nu\right)^2\right]$$

where $t_\nu = \left(\nu - \frac{N}{2}\right)\Delta t$. We find easily

$$G_N(f) = \frac{2\,\Delta t}{N+1}\left[\sum_{\nu=0}^{N} y_\nu^2 + 2\sum_{m=1}^{N}\sum_{\nu=0}^{N-m} y_\nu\, y_{\nu+m} \cos 2\pi f m\,\Delta t\right] \tag{150}$$

and

$$M\,G_N(f) = 2\sigma^2\,\Delta t + 4\,\Delta t \sum_{m=1}^{N}\left(1 - \frac{m}{N+1}\right) R(m) \cos 2\pi f m\,\Delta t. \tag{151}$$

When $\Delta t \to 0$ one finds

$$M\,G_T(f) = 4\int_0^T \left(1 - \frac{\tau}{T}\right) R(\tau) \cos 2\pi f \tau\, d\tau. \tag{152}$$

For the computation of $M[G_N(f)]^2$ we introduce the autocorrelation function

$$C(m, p, q) = M(y_n\, y_{n+m}\, y_{n+m+p}\, y_{n+m+p+q})$$

related to $R(m, p)$ by

$$C(m, p-m, m) = R(m, p).$$

Then the following powers and autocorrelation functions are to be considered.

Power	C-function	R-function
y_ν^4	$C(0,0,0)$	$R(0,0)$
$y_\mu^3 y_{\mu+m}$	$C(0,0,m)$	Not existing
$y_\mu y_{\mu+m}^3$ $(m \geq 1)$	$C(m,0,0)$	»
$y_\mu^2 y_{\mu+m}^2$	$C(0,m,0)$	$R(m,0) = R(0,m)$
$y_\mu^2 y_{\mu+m} y_{\mu+m+p}$	$C(0,m,p)$	Not existing
$y_\mu y_{\mu+m}^2 y_{\mu+m+p}$ $(m, p \geq 1)$	$C(m,0,p)$	»
$y_\mu y_{\mu+m} y_{\mu+m+p}^2$	$C(m,p,0)$	»
$y_\mu y_{\mu+m} y_{\mu+m+p} y_{\mu+m+p+q}$ $(m, p, q \geq 1)$	$C(m,p,q)$	»

In the case of a normal process we have

$$C(m, p, q) = R(m)\,R(q) + R(p)\,R(m + p + q) + R(m + p)\,R(p + q). \quad (153)$$

The mathematical expression for $M[G_N(f)]^2$ does not look very fascinating.

$$\begin{aligned}
\frac{N+1}{4\,\Delta t^2} M[G_N(f)]^2 = {} & C(0,0,0) + 2\sum_{m=1}^{N}\left(1 - \frac{m}{N+1}\right) C(0,m,0)\,(1 + 2\cos^2 2\pi f m\,\Delta t) + \\
& + 4\sum_{m=1}^{N}\left(1 - \frac{m}{N+1}\right)[C(0,0,m) + C(m,0,0)]\cos 2\pi f m\,\Delta t + \\
& + 4\sum_{m=1}^{N-1}\sum_{p=1}^{N-m}\left(1 - \frac{m+p}{N+1}\right)[C(0,m,p)\cos 2\pi f p\,\Delta t + \\
& \qquad + C(m,0,p)\cos 2\pi f (m+p)\,\Delta t + C(m,p,0)\cos 2\pi f m\,\Delta t] + \\
& + 8\sum_{m=1}^{N-1}\sum_{p=1}^{N-m}\left(1 - \frac{m+p}{N+1}\right)[C(0,m,p)\cos 2\pi f m\,\Delta t\cdot\cos 2\pi f (m+p)\,\Delta t + \\
& \qquad + C(m,0,p)\cos 2\pi f m\,\Delta t\cdot\cos 2\pi f p\,\Delta t + \\
& \qquad + C(m,p,0)\cos 2\pi f (m+p)\,\Delta t\cdot\cos 2\pi f p\,\Delta t] + \\
& + 8\sum_{m=1}^{N-2}\sum_{p=1}^{N-m-1}\sum_{q=1}^{N-m-p}\left(1 - \frac{m+p+q}{N+1}\right) C(m,p,q)\,[\cos 2\pi f m\,\Delta t\cdot\cos 2\pi f q\,\Delta t + \\
& \qquad + \cos 2\pi f (m+p)\,\Delta t\cdot\cos 2\pi f (p+q)\,\Delta t + \\
& \qquad + \cos 2\pi f (m+p+q)\,\Delta t\cdot\cos 2\pi f p\,\Delta t]. \quad (154)
\end{aligned}$$

Generally, however, only those terms for which m, p and q are *all* small have any importance.

Since it is a tedious procedure to compute $DG_N(f)$, we try to avoid, if possible, this computation. As a matter of fact some information can always be drawn from $MG_N(f)$ by comparing the curve of this function with the curve of $G_N(f)$. If for large N the two curves do not differ much from each other, we have reasons to believe that the values obtained for $G_N(f)$ are tolerable.

In practical problems we do not know the theoretical autocorrelation functions. If we then use the empirical values of these functions, the error formulas will give more or less wrong values — how wrong can never be said exactly, as not even a probable value of this new error can be given, since we cannot arrive at the probability distributions of the computed quantities by studying a time series during a relatively short time. By prolonging the observation time gradually we obtain more and more correct values only if *all the influencing circumstances are unchanged*, but about this invariance we can never be absolutely convinced. Thus the judgement becomes more or less a matter of experience and common sense.

Let us now study the limit of $M[G_N(f)]^2$ for $\Delta t \to 0$, i.e. the value corresponding merely to the abbreviation of the observation time (T instead of ∞). One finds

$$M[G_T(f)]^2 = \frac{32}{T}\int_0^T d\xi \int_0^{T-\xi} d\eta \int_0^{T-\xi-\eta} d\zeta \left(1 - \frac{\xi+\eta+\zeta}{T}\right) C(\xi,\eta,\zeta)\cdot$$

$$\cdot[2\cos 2\pi f\xi \cdot \cos 2\pi f\zeta + \cos 2\pi f(\xi + 2\eta + \zeta)]. \quad (155)$$

If C always (or at least in a considerable part of the observation domain) is ≥ 0, we have

$$M[G_T(f)]^2 \leq \frac{96}{T}\int_0^T d\xi \int_0^{T-\xi} d\eta \int_0^{T-\xi-\eta} d\zeta \left(1 - \frac{\xi+\eta+\zeta}{T}\right) C(\xi,\eta,\zeta) = M[G_T(0)]^2.$$

Generally the factor $1-(\xi+\eta+\zeta):T$ has only a little influence. Thus we obtain a rather good conception of the error by employing the formula

$$M[G_T(0)]^2 \approx \frac{96}{T}\iiint_{(\Omega)} C(\xi,\eta,\zeta)\, d\xi\, d\eta\, d\zeta \quad (156)$$

where the integral is to be taken over a region bounded by the planes through the points $(0,0,0)$, $(T,0,0)$, $(0,T,0)$ and $(0,0,T)$.

For further simplification of the computational work we can consider a series of concentric spheres with the origin at $(0,0,0)$ and denote the mean correlation at the distance τ from the origin by $\bar{R}(\tau)$. We then find

$$\iiint_{(\Omega)} C(\xi,\eta,\zeta)\, d\xi\, d\eta\, d\zeta \approx \int_0^T \frac{4\pi\tau^2}{8}\bar{R}(\tau)\, d\tau$$

and

$$M[G_T(0)]^2 \approx \frac{150}{T}\int_0^T \tau^2 \bar{R}(\tau)\, d\tau. \quad (157)$$

It is to be observed here that $\bar{R}$ has the dimension of C.

If the studied distribution is approximately normal, the triple integral of (155) is by means of (153) transformed into a sum of double integrals.

By carrying out calculations of that kind studied above it is a good help to dispose of an autocorrelator, i.e. a device which automatically evaluates the correlation coefficients.

Numerical examples

We consider a stationary and normally distributed random process with the mean value 0, the standard deviation σ and the autocorrelation function

$$R(\tau) = \sigma^2 e^{-24\tau}\cos 40\,\tau.$$

I have chosen $\Delta\tau = 0.02$ sec, $N = 100$. Thus $T = 2$ sec. In this case the series of (146) converges very rapidly and using the expression

$$R_{100}(m) = \sigma^2 e^{-0.48\,m} \cos 0.8\,m$$

we have

$$D R_{100} \quad (0) = 0.17\,\sigma^2,$$
$$D R_{100} \quad (5) = 0.12\,\sigma^2,$$
$$D R_{100} \quad (10) = 0.13\,\sigma^2,$$
$$\cdots\cdots\cdots$$
$$D R_{100}(100) = 1.00\,\sigma^2.$$

The spectral density corresponding to the theoretical autocorrelation function given above becomes

$$G(f) = 4\int_0^\infty R(\tau)\cos 2\pi f\tau\, d\tau = 48\,\sigma^2\left[\frac{1}{24^2 + (2\pi f + 40)^2} + \frac{1}{24^2 + (2\pi f - 40)^2}\right].$$

In the following table $G(f)$ for some values of f is compared with the corresponding values of $M G_{100}(f)$ and $M G_\infty(f)$ calculated by means of (151) for $\Delta t = 0.02$ sec.

f (cps)	$G(f)/\sigma^2$	$M G_{100}(f)/\sigma^2$	$M G_\infty(f)/\sigma^2$
0	0.044	0.048	0.047
4	0.070	0.073	0.073
6	0.090	0.091	0.093
8	0.076	0.079	0.080

In this case, the convergence of $R(\tau)$ for $\tau \to \infty$ being very rapid, the greatest part of the error comes from the choise of the interval Δt. Thus the choise of $T = 2$ sec is adequate.

As a simple example of the calculation of $M[G_N(f)]^2$ we use the formula (157) thereby assuming that $\mathfrak{R}(\tau)$ can be written in the form

$$\mathfrak{R}(\tau) = K e^{-\alpha|\tau|}\cos\beta\tau.$$

One then finds

$$M[G_T(0)]^2 \approx \frac{150\,K}{T(\alpha^2+\beta^2)^3}\{2\alpha(\alpha^2 - 3\beta^2) -$$
$$- e^{-\alpha T}[\langle \alpha T^2(\alpha^2+\beta^2)^2 + 2T(\alpha^4 - \beta^4) + 2\alpha(\alpha^2 - 3\beta^2)\rangle \cos\beta T -$$
$$- \langle \beta T^2(\alpha^2+\beta^2)^2 + 4\alpha\beta T(\alpha^2+\beta^2) + 2\beta(3\alpha^2 - \beta^2)\rangle \sin\beta T]\}.$$

$\beta = 0$ gives

$$M[G_T(0)]^2 \approx \frac{150\,K}{T\alpha^3}[2 - e^{-\alpha T}(\alpha^2 T^2 + 2\alpha T + 2)].$$

For the values used above ($T = 2$ sec, $\alpha = 24$, $K = \sigma^2$) the last formula gives

$$M[G_2(0)]^2 \approx 0.011\,\sigma^2.$$

b) Spectral densities for small time intervals. Instantaneous spectral densities

We will now make some remarks concerning spectral densities for small time intervals.

We assume that the servo circuit can be described by a system of linear differential equations with constant coefficients the solution of which may be written

$$z(t) = \sum_{\nu=1}^{n} \int_0^\infty y_\nu(t-s)\, W_\nu(s)\, ds \tag{158}$$

where $y_\nu(t)$ are the different "signals" from which $z(t)$ is built up and $W_\nu(t)$ the corresponding weighting functions. These weighting functions are related to the transfer functions $Y_\nu(f)$ by the wellknown formula[1]

$$Y_\nu(f) = \int_0^\infty W_\nu(t)\, e^{-2\pi j f t}\, dt.$$

We now introduce the Fourier transforms

$$\int_{-s}^{T-s} y_\nu(t+u)\, e^{-2\pi j f u}\, du = {}_T A_\nu(f, t/-s). \tag{159}$$

For $s = t = 0$ this formula gives the ordinary Fourier transform.

Taking the Fourier transform of (158) for the time interval $(t, t+T)$ we obtain

$${}_T A(f, t/0) = \sum_{\nu=1}^{n} \int_0^\infty W_\nu(s)\, e^{-2\pi j f s}\, ds \cdot {}_T A_\nu(f, t/-s).$$

If T is small, the formula (159) cannot be expected to be representative for the whole process. But instead of choosing a longer observation time we can consider other time curves in the same interval. Thus we have to study the mean values of the Fourier transforms (159)

$$M[{}_T A_\nu(f, t/-s)] = {}_T \bar{A}_\nu(f, t/-s) = \int_{-s}^{T-s} \bar{y}_\nu(t+u)\, e^{-2\pi j f u}\, du.$$

It follows

$${}_T \bar{A}(f, t/0) = \sum_{\nu=1}^{n} \int_0^\infty W_\nu(s)\, e^{-2\pi j f s}\, ds \cdot {}_T \bar{A}_\nu(f, t/-s). \tag{160}$$

Let us for finite T-values define the spectral densities as

$${}_T G_\nu(f, t/-s) = \frac{2}{T} |{}_T \bar{A}_\nu(f, t/-s)|^2 \tag{161}$$

[1] For the sake of simplicity I write $Y(f)$ instead of $Y(2\pi j f)$.

and the cross-spectral densities as

$$ {}_T G_{\mu\nu}(f, t/-s) = \frac{2}{T}\, {}_T\bar{A}^*_\mu(f, t/-s) \cdot {}_T\bar{A}_\nu(f, t/-s). \tag{162} $$

This new definition of the spectral densities is in many respects more natural than a definition that is based on observations during an infinite time.

If the quantities ${}_T\bar{A}_\nu(f, t/-s)$ are independent of s and t, the integrals in (160) become

$$ Y_\nu(f) \cdot {}_T\bar{A}_\nu(f). $$

In this case, by taking the square of the absolute value of ${}_T\bar{A}(f)$ and then dividing by $T/2$ we obtain

$$ {}_T G(f) = \sum_{\nu=1}^{n} |Y_\nu(f)|^2\, {}_T G_\nu(f) + \sum_{\mu \neq \nu}\sum Y^*_\mu(f)\, Y_\nu(f)\, {}_T G_{\mu\nu}(f). \tag{163} $$

In the following we write ΔT instead of T. When $\Delta T \to 0$, we have

$$ {}_{\Delta T}\bar{A}_\nu(f, t/-s) = 0\,(\Delta T), $$

under the condition that My_ν is finite in the interval in question. Then

$$ {}_{\Delta T} G_\nu(f, t/-s) = 0\,(\Delta T). $$

It follows that

$$ \frac{1}{\Delta T}\, {}_{\Delta T} G_\nu(f, t/-s) = \frac{2}{\Delta T^2}\, {}_{\Delta T}\bar{A}^*_\nu(f, t/-s) \cdot {}_{\Delta T}\bar{A}_\nu(f, t/-s) $$

tends to a limit when $\Delta T \to 0$. This limit may be defined here as the *instantaneous spectral density* and denoted by $g_\nu(f, t/-s)$.

In the following we have to consider the case where A^* and A are referable to different times. We therefore introduce the quantities

$$ g_\nu(f, t/-u, -v) = \lim_{\Delta T \to 0} \frac{2}{\Delta T^2}\, {}_{\Delta T}\bar{A}^*_\nu(f, t/-u) \cdot {}_{\Delta T}\bar{A}_\nu(f, t/-v). \tag{164} $$

Similarly the instantaneous cross-spectral densities are given by

$$ g_{\mu\nu}(f, t/-u, -v) = \lim_{\Delta T \to 0} \frac{2}{\Delta T^2}\, {}_{\Delta T}\bar{A}^*_\mu(f, t/-u) \cdot {}_{\Delta T}\bar{A}_\nu(f, t/-v). \tag{165} $$

The spectral densities defined in this way are generally complex.

Knowing the instantaneous spectral densities for the different input signals we can put up an expression for the instantaneous spectral densities of the output, but, of course, this expression will not have such a simple form as (163). We find

$$ g(f, t) = \sum_{\mu,\nu} \int_0^\infty \int_0^\infty W_\mu(u)\, W_\nu(v)\, e^{-2\pi j f(u+v)}\, g_{\mu\nu}(f, t/-u, -v)\, du\, dv. \tag{166} $$

It is difficult to see how to have any practical use of formula (166), the most important reason for that being the impossibility to get rid of the weighting functions. One way is of course to define two mean durations, $\bar{u}$ and $\bar{v}$, such that

$$g(f, t) = \sum_{\mu,\nu} Y_\mu(f)\, Y_\nu(f)\, g_{\mu\nu}(f, t/-\bar{u}, -\bar{v}),$$

but this way is not practical, because $\bar{u}$ and $\bar{v}$ are generally depending on t.

In connection with the problem treated above it is of interest to consider also the instantaneous autocorrelation functions defined as the ensemble averages of the products $y_\mu(t)\, y_\nu(t+\tau)$. Thus

$$R_{\mu\nu}(\tau, t) = M[y_\mu(t) \cdot y_\nu(t+\tau)].$$

Then it follows immediately from (158)

$$R(\tau, t) = \sum_{\mu,\nu} \int_0^\infty \int_0^\infty W_\mu(u)\, W_\nu(v)\, R_{\mu\nu}(\tau + u - v, t-u)\, du\, dv. \tag{167}$$

VIII. On the influence of noise on the quantity of information

a) Derivation of a mathematical expression for the quantity of information in linear systems

In this chapter the influence of disturbances on the possibility to send messages through an information channel will be dealt with. This problem does not belong to the servo theory in the general sense but is intimately connected with it. If we for instance have to steer a rocket from the ground, information must be sent to the rocket in some way, and this information is always affected by noise. The steering is assumed to be carried out according to the given orders by means of a servo system.

The mathematical aspects of the theory of information have been treated by SHANNON, GABOR, TULLER[1] and others. However, unique mathematical definitions of the basic concepts seem still to be missed. Further, only rude idealizations have been objects to treatment. In this chapter an attempt is made to restrict the idealizations in some respects, thereby as much as possible following the conceptions which have been accepted by the authors mentioned above.

We consider here the transmission of messages only through that part of an information channel which can be assumed to be at least *approximately linear*. In what concerns the noise this is often the most important link. The message is supposed to be given in the form of a time function $y(t)$, defined for $0 \leq t \leq T$.

[1] SHANNON: *A mathematical theory of communication.* Bell Syst. Tech. J. July 1948, October 1948. — SHANNON: *Communication in the presence of noise.* Proc. IRE. Jan. 1949. — GABOR: *Theory of communication.* J. IEE. Nov. 1946. — TULLER: *Theoretical limitations on the rate of transmissions of information.* Proc. IRE. May 1949.

For stationary processes this time function can be uniquely represented by its Fourier series or, if we are not interested in phase relationships, by the corresponding spectral density function

$$G_T(f) = \frac{2}{T} |A_T(f)|^2$$

where $A_T(f)$ denotes the Fourier transform of $y(t)$;

$$A_T(f) = \int_0^T e^{-2\pi j f t} y(t)\, dt.$$

We assume that only those frequencies lying in the interval $0 \leq f \leq f_m$ will be used for the transmission of information.

The greater T and f_m are, the more information can be transmitted. It is natural to assume that, by unchanging noise conditions, the quantity of information is proportional to T, but it is not as natural to presume that the same is valid concerning f_m. If we say that the quantity of information which in the optimal case *can* be transmitted is proportional to the length of the frequency band, the pronouncement is, of course, correct, but as soon as noise appears, we ought to have a new formulation. Let us for example consider the case where the spectral density of the message is represented by the curves of fig. 12 *a*) and *b*).

If the information is measured by the amount of $G_T(f)$, it is evident that, in the case of noise independent of the frequency, the relative effect of the noise will be less for the frequency band $(0, f_m)$ than for the frequency band $(f_m, 2f_m)$. Thus the quantity of valuable information is greater for the low-frequency band than for the high-frequency band. On the contrary, if the information is obtained by the change of $G_T(f)$, the value of information is the same for all frequencies.

From this example follows that it depends on the coding how the quantity of information shall be measured. We therefore introduce a weighting function $V(f, t)$ measuring the ability of transmitting information at different frequencies and different times.[1] Furthermore we put

Quantity of information $$H = \int_0^T dt \int_0^{f_m} c(f, t)\, V(f, t)\, df.$$

For stationary processes $V(f, t)$ is independent of time ($= V(f)$). Concerning the original messages we put $V(f) \equiv 1$ (no disturbances). Then the quantity of information can be written

$$H = c\, T\, f_m \tag{168}$$

where c is a quantity depending on the coding. But c must also depend on the greatest possible amplitude of oscillations (the quantity of information must increase at the same time as this amplitude). As shown by TULLER (Proceedings

[1] For the sake of simplicity V has not been made dependent on the amplitude, but it is easy to extend the reasoning to this more general case.

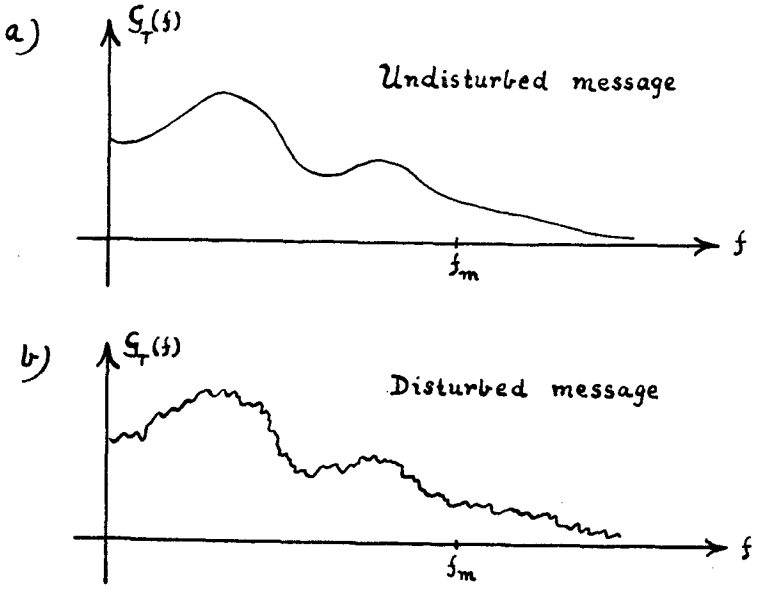

Fig. 12.

of the IRE May 1949) c can be made arbitrarily large. If we choose a finite value for c, this is caused by practical and economic not by theoretical reasons.

If no disturbances at all occur, the quantity of information goes unchanged through the channel. Our task is to determine the influence of disturbances. Another important problem is to construct the channel in such a way that the loss of information will be reduced as much as possible.

To every point in the transmission system exists a time curve corresponding to the message in question. Beside these time curves we consider the corresponding spectral density curves. The following notations will be used to define the spectral densities.

$$\text{Inputs:} \begin{cases} G_0(f) & \text{spectral density of message} \\ G_\nu(f) & \text{spectral density of noise } (\nu = 1, 2, \ldots) \\ G_{\mu\nu}(f) & \text{cross spectral densities } (\mu \neq \nu) \end{cases}$$

$$\text{Output:} \quad z(f) \quad \text{spectral density of the outcoming message}$$

$$\text{Error:} \quad \varepsilon(f) = |G_0(f) - z(f)|$$

The spectral densities are to be calculated for finite time intervals. It is assumed here that *all information or loss of information is contained in these spectral densities.*

Knowing the probability distributions of the inputs and the corresponding transfer functions $Y_\nu(f)$ one can derive the probability distribution of $z(f)$ and $\varepsilon(f)$. In fact, in linear transmission systems $z(f)$ becomes a linear function of the input spectral densities the coefficients of which in an easy way are determined by the transfer functions. If there are n noise components, we have according to (163)

$$z(f) = \sum_{\nu=0}^{n} |Y_\nu(f)|^2 G_\nu(f) + \sum_{\mu \neq \nu} \sum Y_\mu^*(f)\, Y_\nu(f)\, G_{\mu\nu}(f)$$

where Y means a transfer function and Y^* its conjugate.

As is said before the quantity of information depends on the magnetude of the largest amplitudes. The relation between amplitude $\varrho_T(f)$ and spectral density $G_T(f)$ for finite T-values is expressed by the formula

$$\varrho_T(f) = \sqrt{\frac{2}{T}}\sqrt{G_T(f)}.$$

The amplitudes of the Fourier coefficients are obtained by putting $f = \frac{\nu}{T}$ ($\nu = 1, 2, \ldots$). In the following T will be omitted as index and replaced by indices with other meanings.

In order to derive, in a rational way, a mathematical expression for the quantity of information in the case of noise we consider a coordinate system with three dimensions (t, f, ϱ) where t means the time, f the frequency and ϱ the amplitude. The phase will, for the present, not be employed as a source of information. In accordance with the foregoing t is assumed to be able to vary from 0 to T and f from 0 to f_m. The amplitude ϱ is supposed to be limited upwards by a quantity depending on f and t, $\varrho \leq \varrho_m(f, t)$. The space of variation is divided into small parts by the intervals ΔT, Δf_m and $\Delta \varrho_m$. These intervals will be more carefully defined below.

Two neighbouring information lines in a frequency plane (f = const.) can be looked upon as representing different messages only if the difference between the amplitudes for a certain time ΔT amounts to or exceeds the noise amplitude for this time. Thus we have to determine the probability distribution of $\varepsilon(f)$ during the time intervals $(t, t + \Delta T)$. By stationary processes this distribution is independent of t. On the contrary, the length of ΔT must always affect $\varepsilon(f)$ and also its distribution. During the intervals ΔT all signals are supposed to be stationary. The division into the intervals ΔT furnishes a possibility to take into consideration the fact that different parts of a message may be affected by noise in a different degree.

Let the statistical frequency function of $\varepsilon(f)$ during the time interval $(t, t + \Delta T)$ be denoted by $\psi(\varepsilon; f, t, \Delta T)$, i.e.

$$\psi(\varepsilon; f, t, \Delta T)\, d\varepsilon = P\{\varepsilon \leq \varepsilon(f) \leq \varepsilon + d\varepsilon \text{ for } t \leq t \leq t + \Delta T\}.$$

Further we put

$$\varrho_m(f) = \sqrt{\frac{2}{T}} \cdot \sqrt{z_m(f)}.$$

The probable number of distinguishable information bands[1] for every time point in the interval $(t, t + \Delta T)$ becomes[2]

[1] To speak about information *lines* would, of course, be inadequate.

[2] There exist, of course, many other ways to introduce the influence of the noise on the quantity of information.

$$\int_0^{z_m(f)} \frac{\sqrt{z_m(f)}}{\sqrt{\varepsilon}} \psi(\varepsilon; f, t, \Delta T)\, d\varepsilon$$

which conveniently can be written

$$\sqrt{z_m(f)} \cdot M\left[\frac{1}{\sqrt{\varepsilon(f)}}\right]_{t, \Delta T}.$$

The number of information bands will, according to well-known reasons, be given in logarithmic units. If we use binary units, we have to take logarithms with the base 2.

Let us assume that the weighting function $V(f, t)$ be constant in $(\Delta f_m, \Delta T)$. This assumption does not imply any loss of generality, since Δf_m can be taken arbitrarily small. In the case where the information is concentrated to discrete frequencies (Example: The Fourier coefficients), one can consider Δf_m as the difference between the frequencies employed.

The quantity of information which can be transmitted over the region $(\Delta T; \Delta f_m)$ may according to the above principles be defined as[1]

$$c\, \Delta T\, \Delta f_m \left\{\log \sqrt{z_m(f)} + \log M\left[\frac{1}{\sqrt{\varepsilon(f)}}\right]_{t, \Delta T}\right\}.$$

As is already noticed the frequencies have to be weighted with a convenient weighting function $V(f; t)$, depending on t. Hence the total quantity of information becomes

$$H = c\, \Delta T\, \Delta f_m \sum_t \sum_f V(f; t) \left\{\log \sqrt{z_m(f)} + \log M\left[\frac{1}{\sqrt{\varepsilon(f)}}\right]_{t, \Delta T}\right\}. \tag{169}$$

The derivation of formula (169) is in a few words based on the following train of thought. All signals (messages as well as disturbances and covariation signals) are for each time interval ΔT coded as spectral densities. The probability distributions of these spectral densities are put up. The information- and noise-channels are defined by the transfer functions $Y_\nu(f)$. The capacities of these channels are assumed to be limited in such a way that neither the amplitudes nor the frequencies are allowed to exceed certain values. Generally the maximum amplitudes depend on the frequencies. Finally, to the parts of a message transmitted at different frequencies have been given different weights defined by the weighting function $V(f; t)$. The weighting function ought to be determined with respect to the method of coding and the kind of noise.

In the case of all processes being stationary the formula (169) changes into

$$H = c\, T\, \Delta f_m \sum_f V(f) \left\{\log \sqrt{z_m(f)} + \log M\left[\frac{1}{\sqrt{\varepsilon(f)}}\right]_{\Delta T}\right\}. \tag{170}$$

b) Determination of the maximum quantity of information

An extremely important problem is to construct the information channel in such a way that it transmits the greatest possible quantity of information per

[1] c is here a calibration factor.

unit time. From purely mathematical point of view the task is to determine the transfer functions $Y_\nu(f)$ so that the expression (169) or (170) becomes maximum. This is equivalent with maximizing the sum

$$J = \sum_t \sum_f V(f;t) \log M \left[\frac{\sqrt{z_m(f)}}{\sqrt{\varepsilon(f)}} \right]_{t,\, \Delta T} \tag{171}$$

or

$$J = \sum_f V(f) \log M \left[\frac{\sqrt{z_m(f)}}{\sqrt{\varepsilon(f)}} \right]_{\Delta T} \tag{172}$$

respectively. Hereby we can assume that $J > 0$.

In practice it will be most convenient to let the form of the transfer functions be given and to determine in the best possible way some parameters $\alpha, \beta, \gamma, \ldots$.

Let us now assume that information be transmitted only for discrete frequencies: $f_1, f_2, \ldots f_m$, but that the number of these frequencies can be made arbitrarily large. Looking at formula (171) or (172) one easily finds associations with R. A. Fishers "maximum likelihood-method", though the definition of the concept "likelihood-function" in the present case must be modified in several respects. It is easy to give an interpretation of Fishers likelihood-function in terms of information theory. However, I do not intend to use Fishers reasoning here.

As an example we are going to treat the case where all input spectral densities are approximately normally distributed for every time interval ΔT and for all frequencies.[1] (That the distribution cannot be exactly normal is evident, since the spectral densities are always positive.) Then according to (163) also $\varepsilon(f)$ becomes approximately normally distributed. Further we assume that only one parameter (α) can be varied and put

$$M\,\varepsilon(f) = m_f(\alpha) \quad (\text{always} > 0),$$
$$D\,\varepsilon(f) = \sigma_f(\alpha).$$

Generally we do not know the mathematical expressions of m_f and σ_f. Then these quantities must be assumed to be given by a table of numerical values for different values of f and α. For every pair of f and α we have

$$M \left[\frac{1}{\sqrt{\varepsilon(f)}} \right]_{\Delta T} \approx \frac{1}{\sigma_f \sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{\varepsilon}} e^{-\frac{(\varepsilon - m_f)^2}{2\sigma_f^2}} d\varepsilon = \frac{1}{\sqrt{2\pi}\sigma_f} \int_0^\infty \frac{1}{\sqrt{\zeta}} e^{-\frac{1}{2}(\zeta-\mu)^2} d\zeta$$

where $\mu = m_f/\sigma_f$. In the easiest case μ is independent of α. This condition is realized, if $\varepsilon(f)$ changes into $k\,\varepsilon(f)$ (k = constant), when α changes into another value, say α'. Then we have to maximize the sum

$$\sum_\nu V(f_\nu) \log \frac{1}{\sqrt{\sigma_{f_\nu}(\alpha)}}$$

[1] Further we assume that $P\{\varepsilon(f) = 0\}$ be practically zero.

(under the assumption, of course, that this sum is positive) which is the same as minimizing

$$L(\alpha) = \sum_{\nu} V(f_\nu) \log \sigma_{f_\nu}(\alpha). \tag{173}$$

Hitherto we have used two error-reducing methods: the rms method treated in chapter VI and the maximum information method considered in this chapter. Concerning the rms method there is no fundamental difference between considering the spectral densities and the squares of the time functions. But the rms method deals with the standard deviations themselves, whereas the maximum information method considers their logarithms (in the last case). Thus by the former method we have given much greater weight to the greatest amplitudes than by the latter one.

IX. Servo circuits defined by linear differential equations with non-constant coefficients containing random parameters

In the former chapters we have mostly been occupied with the problem of studying servo circuits characterized by linear differential equations with *constant* coefficients. Let us now consider the case where the coefficients are functions of the time containing one or more random parameters the variation of which being due to disturbances. Further these functions are supposed to be depending on one or more adjustable parameters.

For many practical purposes it is convenient to write the equation in the form

$$\sum_{\nu=0}^{n} [a_\nu(t) + \alpha_\nu(t) \cdot \beta_\nu(t)] \frac{d^\nu u}{dt^\nu} = f(t) + \varphi(t) \cdot \psi(t) \tag{174}$$

where $\beta_\nu(t)$ and $\psi(t)$ contain the random parameters whereas $a_\nu(t)$, $\alpha_\nu(t)$, $f(t)$ and $\varphi(t)$ are ordinary functions; $\beta_\nu(t)$ and $\psi(t)$ may be considered as disturbances.

We assume that the probability distribution of the random parameters be such that the time functions $\beta_\nu(t)$ and $\psi(t)$ during a certain time interval (t_1, t_2) describe stationary processes, i.e. that the statistical characteristics during this interval are constant. In the case of the coefficients $a_\nu(t) + \alpha_\nu(t) \cdot \beta_\nu(t)$ being constants the influence of the disturbances on the solution of the equation has been studied by several authors[1] by considering the spectral densities. We shall try to use a similar reasoning also in the more general problem announced above.

In the present treatment only the spectral density values corresponding to the harmonics will be considered. Thus, after having chosen the time interval (t_1, t_2) the functions are assumed to be developed in Fourier series for this interval. In order to have a full description of the processes we must of course take the time $t_2 - t_1$ at least so long that the correlation between the values at the limit points can be neglected.

---

[1] See for example R. S. PHILLIPS; Radiation Laboratory Series 25.

Suppose that the solution of the non-stochastic part of the equation be $y(t)$, i.e.

$$\sum_{\nu=0}^{n} a_\nu(t) \frac{d^\nu y}{dt^\nu} = f(t). \tag{175}$$

For the solution of (174) we write

$$u(t) = y(t) + \eta(t) \cdot \zeta(t)$$

where $\zeta(t)$ is assumed to be a stochastic and $\eta(t)$ an ordinary function. The variation of the conditions of disturbances has been taken account of by the function $\eta(t)$. Many times it ought to be possible to determine this function approximately by practical experiments. In this investigation $\eta(t)$ is assumed to be constant in (t_1, t_2).

The stochastic function $\zeta(t)$ must satisfy the equation

$$\eta \sum_{\nu=0}^{n} [a_\nu(t) + \alpha_\nu(t) \cdot \beta_\nu(t)] \frac{d^\nu \zeta}{dt^\nu} = \varphi(t) \cdot \psi(t) - \sum_{\nu=0}^{n} \alpha_\nu(t) \cdot \beta_\nu(t) \frac{d^\nu y}{dt^\nu}. \tag{176}$$

Postulating that $\zeta(t)$ represents a stationary process we try to write

$$\zeta = \sum_{\mu} \left(X_\mu \cos \frac{\mu t}{T} + Y_\mu \sin \frac{\mu t}{T} \right) \tag{177}$$

where $T = t_2 - t_1$.[1] Of course, without further assumptions it is not sure that (176) has a solution of this form with a *finite* number of terms. Let us for instance consider the simple equation

$$\frac{d^2 y}{dt^2} + k y = \cos t \qquad (k \text{ stochastic variable})$$

which has the solution

$$y = \begin{cases} \dfrac{\cos t}{k-1} & \text{for } k \neq 1 \\ \frac{1}{2} t \sin t & \text{for } k = 1 \end{cases}$$

In the case $k = 1$, however, the solution can be represented in every time interval (t_1, t_2) by a Fourier expansion.

If a solution of the type (177) exists, the derivatives become

$$\frac{d^\nu \zeta}{dt^\nu} = \sum_{\mu} \left[X_\mu \left(\frac{\mu}{T} \right)^\nu \cos \left(\frac{\mu t}{T} + \nu \cdot \frac{\pi}{2} \right) + Y_\mu \left(\frac{\mu}{T} \right)^\nu \sin \left(\frac{\mu t}{T} + \nu \cdot \frac{\pi}{2} \right) \right].$$

A criterion for the possibility to use (177) must be identical with a criterion for the existence of the quantities $\{X_\mu, Y_\mu\}$, giving a convergent series. We shall come back to this question in the following.

We can proceed on two ways. Either we use the Fourier expansions for all of the functions contained in (176) and compare the Fourier coefficients of both members of the equation, or we choose certain time points in the interval

[1] μ = number of periods.

(t_1, t_2) and observe the system functions and the inputs only at these time points. In both cases we have to compute X and Y from a system of linear equations.

In the first case we have to deal with products of sines and cosines which must be written as sums, e.g.

$$\cos\left(\frac{kt}{T} + \nu\cdot\frac{\pi}{2}\right)\cos\frac{it}{T} = \frac{1}{2}\left\{\cos\left[\frac{(k+i)t}{T} + \nu\cdot\frac{\pi}{2}\right] + \cos\left[\frac{(k-i)t}{T} + \nu\cdot\frac{\pi}{2}\right]\right.$$

By comparing the coefficients of the corresponding sine and cosine terms of both members of the resulting equation we obtain a system of equations which will be linear in X_μ and Y_μ. The solution of this system gives X_μ and Y_μ as rational functions of the Fourier coefficients of the observed quantities. Knowing the probability distributions of these Fourier coefficients we can also determine the distributions of X_μ and Y_μ. However, this method being very tedious, I am not going to discuss it in detail.

By the use of the second method we introduce the following auxiliary quantities:

$$A_\nu(t) = a_\nu(t) + \alpha_\nu(t)\cdot\beta_\nu(t)$$

$$\varkappa_\mu(t) = \sum_{\nu=0}^{n} A_\nu(t)\cdot\left(\frac{\mu}{T}\right)^\nu \cos\left(\frac{\mu t}{T} + \nu\cdot\frac{\pi}{2}\right) =$$

$$= \cos\frac{\mu t}{T}\sum_{\nu=0}^{\left[\frac{n}{2}\right]}(-1)^\nu A_{2\nu}(t)\left(\frac{\mu}{T}\right)^{2\nu} + \sin\frac{\mu t}{T}\sum_{\nu=1}^{\left[\frac{n+1}{2}\right]}(-1)^\nu A_{2\nu-1}(t)\left(\frac{\mu}{T}\right)^{2\nu-1}$$

$$\lambda_\mu(t) = \sum_{\nu=0}^{n} A_\nu(t)\left(\frac{\mu}{T}\right)^\nu \sin\left(\frac{\mu t}{T} + \nu\cdot\frac{\pi}{2}\right) =$$

$$= \sin\frac{\mu t}{T}\sum_{\nu=0}^{\left[\frac{n}{2}\right]}(-1)^\nu A_{2\nu}(t)\left(\frac{\mu}{T}\right)^{2\nu} - \cos\frac{\mu t}{T}\sum_{\nu=1}^{\left[\frac{n+1}{2}\right]}(-1)^\nu A_{2\nu-1}(t)\left(\frac{\mu}{T}\right)^{2\nu-1}$$

$$\eta\,\psi(t) = \varphi(t)\cdot\psi(t) - \sum_{\nu=0}^{n}\alpha_\nu(t)\cdot\beta_\nu(t)\frac{d^\nu y}{dt^\nu} = f(t) + \varphi(t)\cdot\psi(t) - \sum_{\nu=0}^{n} A_\nu(t)\frac{d^\nu y}{dt^\nu}.$$

Then we have

$$\sum_\mu [\varkappa_\mu(t)\cdot X_\mu + \lambda_\mu(t)\cdot Y_\mu] = \Psi(t). \tag{178}$$

The number of coefficients X and Y which can be determined must of course be the same as the number of t-values taken into consideration. Thus, the more observations being made, the better our knowledge about the output error becomes. But since the number of observations necessarily must be finite, we always commit an error which will influence on X and Y. Unfortunately this error cannot be determined in advance, but we can obtain some information about its importance by carrying out the computations for different numbers of observations. The condition for the existence of a solution of the form

(177) is of course that for an arbitrary set of t-values in the interval (t_1, t_2) the linear system (178) has a solution $\{X_\mu, Y_\mu\}$. When talking about a solution $\{X_\mu, Y_\mu\}$ I mean a series of number pairs (X_μ, Y_μ) for which the sum (177) converges.

Suppose that (178) have a solution. This solution gives the Fourier coefficients of the output disturbance. As already pointed out the spectra of X_μ and Y_μ are random functions for which we have to put up representative expressions. However, in order to obtain a representative description it is necessary to repeat the computations for several input processes. This is of course a lack but can hardly be avoided, since there exists no simple relation between the spectral densities of the inputs and the output in the case studied now.

Our goal has not been reached when we have determined the spectra for the output disturbance, though these spectra are of great importance by judging the construction of the servo system. The kernel of the problem is the designing of the system in an optimal way. In order to make this design possible we have already assumed that there are a number of adjustable parameters in the functions defining the servo circuit. We are going to determine these parameters in such a way that the rms disturbance of the output becomes a minimum or, what is the same, that the integral of the spectral density of this disturbance becomes as small as possible. Of course, the adjustable parameters may be involved also in $y(t)$, i.e. the non-stochastic part of the solution of equation (174). In this rough treatment, however, I think there is no reason to let this fact influence on the minimization procedure.

Since we know the spectral density only for the harmonic frequencies, we have to minimize the expression

$$S = \sum_\mu (X_\mu^2 + Y_\mu^2).$$

For practical purposes it is no sense to treat this minimization in an exact way. It ought to be satisfactory to solve the equations (178) for a set of parameter combinations and then try to reach at the best combination by some method of interpolation. By solving the system of linear equations there is a need of mathematical machines.

One of the reasons for avoiding exact methods is that S only contains the harmonic components. By minimizing this expression we therefore have assumed that the integral of the spectral density takes its smallest value at the same time as S. In most cases this assumption will lead to acceptable results.

Tryckt den 2 september 1952

Uppsala 1952. Almqvist & Wiksells Boktryckeri AB