ON THE STATISTICAL ESTIMATION OF THE FREQUENCY RESPONSE FUNCTION OF A SYSTEM HAVING MULTIPLE INPUT*)

HIROTUGU AKAIKE

(Received Feb. 25, 1965)

1. Introduction and summary

In 1963 a research project for the advancement of the method of estimation of the frequency response function of a system having a single input was undertaken at our Institute and the results were reported in *Supplement III of the Annals of the Institute of Statistical Mathematics* (1964). It was then concluded that it would be necessary to extend the method to the case of multiple input to make the estimation procedure useful in many practical situations [1, p. 15]. In this connection, we note that an interesting paper by Leo J. Tick [7] appeared in 1963, which treats multiple input problem and is full of suggestions.

If we try to apply the method of estimation for the case of single input to the identification of response characteristic to a specified input of a system which has essentially multiple inputs, ignoring the existence of other inputs, we encounter with the following two types of difficulties:

1) decrease of the coherency between the specified input and the output due to the contributions of other inputs which are incoherent with the present input, and

2) contamination of the response characteristic with the ghosts produced by the coherencies between inputs.

To overcome these difficulties we have to develop the method of estimation for the case of multiple input.

In this paper a method of estimation for the case of multiple input is presented, which is a direct extension of the method for the case of

*) By a letter from Mr. A. G. Piersol of the Measurement Analysis Corporation of U.S.A. of 10 Feb. 1964 the author was informed that Mr. L. Enochson, of the Corporation, and Mr. Piersol had written a paper on the analysis of multiple input linear systems subjected to random excitation and it had been submitted to the *Journal of the Acoustical Society of America*. Yet being unable for the present author to have an access to their paper, he decided to present this paper though it may show some overlappings with theirs.

single input. Necessary computation scheme for the estimation procedure is given and the practical applicability of the method is verified by numerical examples. It is seen that a relatively crude approximation to the distribution of a confidence region, given in this paper, can be used quite effectively to see the statistical reliability of the estimate.

2. Preliminary observations of the problem

We shall here consider a time invariant linear system having k inputs $x_1(t)$, $x_2(t)$, $\cdots$, $x_k(t)$ and giving an output $y(t)$ represented by

$$y(t)=\sum_{j=1}^{k}\int_{-\infty}^{\infty} x_j(t-\tau)h_j(\tau)d\tau$$

where for each j the function $h_j(\tau)$ is the impulse response function of the system for the jth input and is supposed to belong to $L_1(-\infty, \infty)$ and $L_2(-\infty, \infty)$, i.e., it satisfies the conditions

$$\int_{-\infty}^{\infty} |h(\tau)|\, d\tau<\infty$$

$$\int_{-\infty}^{\infty} |h(\tau)|^2 d\tau<\infty .$$

The frequency response function $A_j(f)$ of the system for the jth input is the Fourier transform of the impulse response function and is given by

$$A_j(f)=\int_{-\infty}^{\infty} \exp(-i2\pi f\tau)h(\tau)d\tau .$$

If we here restrict the inputs to those which belong to $L_1(-\infty, \infty)$ and $L_2(-\infty, \infty)$, then we can see that the output $y(t)$ also belongs to the same class and we have

$$y(f)=\sum_{j=1}^{k} A_j(f)x_j(f)$$

where $y(f)$ and $x_j(f)$ represent the Fourier transforms of $y(t)$ and $x_j(t)$ respectively. From this relation we can see that if k independent observations of the complete history of the inputs $x_1(t)$, $x_2(t)$, $\cdots$, $x_k(t)$ and output $y(t)$ are available then the value of $A_j(f)$ $(j=1, 2, \cdots, k)$ can be obtained by solving the linear simultaneous equation

$$y^{(\nu)}(f)=\sum_{j=1}^{k} A_j(f)x_j^{(\nu)}(f) \qquad (\nu=1, 2, \cdots, k)$$

where $y^{(\nu)}(f)$ and $x_j^{(\nu)}(f)$ represent the values of $y(f)$ and $x_j(f)$ at the

νth observation and it is assumed that the k-dimensional complex vectors $(x_1^{(\nu)}(f), x_2^{(\nu)}(f), \cdots, x_k^{(\nu)}(f))$ $(\nu=1, 2, \cdots, k)$ are lineary independent. Even in the case where only a single observation is available if we can select a set of frequencies $(f_1, f_2, \cdots, f_k)$ for which the vectors $(x_1(f_\nu), x_2(f_\nu), \cdots, x_k(f_\nu))$ $(\nu=1, 2, \cdots, k)$ are lineary independent and $A_j(f_\nu)$s$(\nu=1, 2, \cdots, k)$ are considered to be approximately equal to the value $A_j(f)$ for each j, we shall be able to get an approximation $\tilde{A}_j(f)$ to $A_j(f)$ by solving the simultaneous linear equation

$$y(f_\nu)=\sum_{j=1}^{k} \tilde{A}_j(f)x_j(f_\nu) \qquad (\nu=1, 2, \cdots, k)$$

for the unknowns $\tilde{A}_j(f)(\nu=1, 2, \cdots, k)$.

This last observation is essential for the understanding of the structure of estimation procedure by a single observation which we shall develop in the following sections.

In practical observations it is more natural to consider that the observed value of the output is contaminated with an additive noise $n(t)$ and thus the observed output, which we shall denote by $x_0(t)$, is represented as

$$x_0(t)=y(t)+n(t)\,.$$

It is also natural to consider here $n(t)$ to be a sample function of a stochastic process. For the moment we shall assume that $n(t)$ belongs to $L_1(-\infty, \infty)$ and $L_2(-\infty, \infty)$ and has its Fourier transform $n(f)$. Then we have the relation

$$x_0(f)=\sum_{j=1}^{k} A_j(f)x_j(f)+n(f)\,,$$

and thus if we solve the following equation for the unknowns $\tilde{\tilde{A}}_j(f)$ $(j=1, 2, \cdots, k)$, in place of the former equation for $\tilde{A}_j(f)$s,

$$x_0(f_\nu)=\sum_{j=1}^{k} \tilde{\tilde{A}}_j(f)x_j(f_\nu) \qquad (\nu=1, 2, \cdots, k)$$

we get an approximation to $A_j(f)$ contaminated with the noise $n(f_\nu)$.

Now if we can assume that there is a set of frequencies $(f_1, f_2, \cdots, f_N)$ for which $A_j(f_\mu)=A_j(\mu=1, 2, \cdots, N\colon j=1, 2, \cdots, k)$ holds and that $n(f_\mu)$s are mutually independently distributed following one and the same distribution, then the relation

$$x_0(f_\mu)=\sum_{j=1}^{k} A_jx_j(f_\mu)+n(f_\mu) \qquad (\mu=1, 2, \cdots, N)$$

suggests that we are just confronted with the problem of estimation of multiple regression coefficients. It should be noted here that the variables are complex-valued.

3. Regression analysis of complex variables

In this section we shall engage ourselves with a necessary extension of Fisher's lemma, on the orthogonal decomposition of real Gaussian random variables [4, pp. 379–381], to the case of complex variables. The result will be used in the later section for the construction of a confidence region of our estimate of the value of frequency response function at a preassigned frequency. Complex random variables we treat in this section are supposed to have real and imaginary part which are mutually independent. Complex random variables are said to be mutually independent when their real and imaginary parts are all mutually independent. Now consider complex random variables $n_\nu(\nu=1, 2, \cdots, N)$ which are mutually independently distributed with the first and the second order moments

$$En_\nu=0$$

$$E(\mathrm{Re}\, n_\nu)^2=E(\mathrm{Im}\, n_\nu)^2=\sigma^2 ,$$

and suppose that our observations are made on the variables $x_{1\nu}, x_{2\nu}, \cdots, x_{k\nu}$ and $x_{0\nu}(\nu=1, 2, \cdots, N)$ satisfying the relation

$$x_{0\nu}=\sum_{j=1}^{k} A_j x_{j\nu}+n_\nu \qquad (\nu=1, 2, \cdots, N)$$

where $x_{1\nu}, x_{2\nu}, \cdots, x_{k\nu}$ are considered to be observable non-random variables.

Here we assume that n_ν follows a (complex) Gaussian distribution. Then the likelihood function $L=L(x_{01}, x_{02}, \cdots, x_{0N}; A_1, A_2, \cdots, A_k, \sigma^2)$ is given as follows:

$$L=\left(\frac{1}{2\pi\sigma^2}\right)^N \exp\left(-\frac{1}{2\sigma^2}\sum_{\nu=1}^{N}\left|x_{0\nu}-\sum_{j=1}^{k} A_j x_{j\nu}\right|^2\right) .$$

The maximum likelihood estimate $(\hat{A}_1, \hat{A}_2, \cdots, \hat{A}_k)$ of $(A_1, A_2, \cdots, A_k)$ is obtained by finding the values of A_js which minimize the sum of squares in the second parenthesis of the right hand side. These values are given as follows:

$$\begin{bmatrix} \hat{A}_1 \\ \hat{A}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{A}_k \end{bmatrix} = L^{-1} \begin{bmatrix} (x_0, x_1) \\ (x_0, x_2) \\ \cdot \\ \cdot \\ \cdot \\ (x_0, x_k) \end{bmatrix}$$

where

$$L = \begin{bmatrix} (x_1, x_1) & (x_2, x_1) & \cdots & (x_k, x_1) \\ (x_1, x_2) & (x_2, x_2) & \cdots & (x_k, x_2) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ (x_1, x_k) & (x_2, x_k) & \cdots & (x_k, x_k) \end{bmatrix}$$

$$(x_j, x_l) = \sum_{\nu=1}^{N} x_{j\nu} \bar{x}_{l\nu} \qquad (j, l=0, 1, 2, \cdots, k)$$

and it is assumed that $N \geqq k$ and L^{-1} exists, i. e., that the (column) vectors $x_j = (x_{j1}, x_{j2}, \cdots, x_{jN})^T$*) $(j=1, 2, \cdots, k)$ are linearly independent as an element in the N-dimensional complex Eucledian space where the inner product (x_j, x_l) of two vectors x_j and x_l is defined as above. $\sum_{j=1}^{k} \hat{A}_j x_j$ gives the projection of $x_0 = (x_{01}, x_{02}, \cdots, x_{0N})^T$ on the k-dimensional subspace spanned by the vectors $x_1, x_2, \cdots, x_k$ and thus

$$(x_0 - \sum_{j=1}^{k} \hat{A}_j x_j, x_l) = 0 \qquad (l=1, 2, \cdots, k).$$

By putting $x_0 = \sum_{j=1}^{k} A_j x_j + n$ into above formula for $\hat{A}_j$

$$\begin{bmatrix} \hat{A}_1 \\ \hat{A}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{A}_k \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ \cdot \\ \cdot \\ \cdot \\ A_k \end{bmatrix} + L^{-1} \begin{bmatrix} (n, x_1) \\ (n, x_2) \\ \cdot \\ \cdot \\ \cdot \\ (n, x_k) \end{bmatrix}.$$

From this we can see that the vector $p = \sum_{j=1}^{k} (\hat{A}_j - A_j) x_j$ is the projection of $n = (n_1, n_2, \cdots, n_N)^T$ on the space spanned by the vectors $x_1, x_2, \cdots, x_k$ and thus in the above representation of $\hat{A}_j$s n may be replaced by p to give

*) $(z_1, z_2, \cdots, z_N)^T$ signifies the column vector transposed from the row vector.

$$\begin{bmatrix} \hat{A}_1 - A_1 \\ \hat{A}_2 - A_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{A}_k - A_k \end{bmatrix} = L^{-1} \begin{bmatrix} (p,\ x_1) \\ (p,\ x_2) \\ \cdot \\ \cdot \\ \cdot \\ (p,\ x_k) \end{bmatrix}.$$

This shows that $(\hat{A}_j - A_j)$s are linear functions of the vector p.

Now n can be represented in the form

$$n = n_1 e_1 + n_2 e_2 + \cdots + n_N e_N$$

where e_j denotes a unit vector of which the jth component is equal to 1 and others are all equal to 0. Let us consider another representation

$$n = m_1 d_1 + m_2 d_2 + \cdots + m_N d_N$$

where the set $(d_1, d_2, \cdots, d_N)$ is forming an orthonormal basis, i. e., $(d_j, d_l) = \delta_{jl}$ and δ_{jl} is a Kronecker's delta. Under the present assumption of independence of n_νs we have $En_j\bar{n}_l = 2\sigma^2\delta_{jl}$ and from this and the relation $m_j\bar{m}_l = (d_l, n)(n, d_j) = d_l^T \bar{n} n^T \bar{d}_j$, where $\bar{n}$ and $\bar{d}_j$ denote vectors of which components are complex conjugates of those of n and d_j respectively, we get

$$\begin{aligned} Em_j\bar{m}_l &= 2\sigma^2(d_l,\ d_j) \\ &= 2\sigma^2\delta_{jl}\,. \end{aligned}$$

By using this last equation we can prove, under the assumption of normality of the distribution of n_νs, that m_νs are mutually independently distributed following one and the same type of distribution as that of n_νs. Here we shall consider that the first k vectors $d_1, d_2, \cdots, d_k$ are selected in the subspace spanned by the vectors $x_1, x_2, \cdots, x_k$ which are assumed to be linearly independent and that the residual $N-k$ vectors $d_{k+1}, d_{k+2}, \cdots, d_N$ are forming a basis of the subspace which is orthogonal to the space just mentioned. We have already seen that $(\hat{A}_j - A_j)$s are represented as linear functions of the projection p of n into the space spanned by the vectors $x_1, x_2, \cdots, x_k$ and thus they are, as functionals of the random vector p, distributed following a Gaussian distribution and are independent with any number of linear functionals, which follow a Gaussian distribution, of the residual random vector $r = x_0 - \sum_{j=1}^{k} \hat{A}_j x_j$.

Obviously we have

$$\|n\|^2=\|p\|^2+\|r\|^2$$

where $\|x\|^2=(x, x)$ for any vector x, and by using the above representation we have

$$\|p\|^2=\sum_{j=1}^{k}|m_j|^2$$

$$\|r\|^2=\sum_{j=k+1}^{N}|m_j|^2 .$$

From this last equation it can be seen that $\|r\|^2/2\sigma^2$ is distributed following a χ^2-distribution with d. f. $2(N-k)$ and is independent of any linear functionals of p. Now taking into account the relation

$$\begin{bmatrix} \hat{A}_1-A_1 \\ \hat{A}_2-A_2 \\ \vdots \\ \hat{A}_k-A_k \end{bmatrix} = L^{-1} \begin{bmatrix} (n, x_1) \\ (n, x_2) \\ \vdots \\ (n, x_k) \end{bmatrix}$$

we can see that $E(\hat{A}_j-A_j)=0$ $(j=1, 2, \cdots, k)$ and we can get the variance convariance matrix of $\Delta A_j=\hat{A}_j-A_j (j=1, 2, \cdots, k)$ as follows:

$$\overset{\cdots\cdots l \cdots\cdots}{j \begin{bmatrix} & & \\ & E\Delta A_j \overline{\Delta A_l} & \\ & & \end{bmatrix}} = 2\sigma^2 L^{-1} ,$$

i.e.,

$$E\Delta A_j \overline{\Delta A_l} = 2\sigma^2 \Lambda_{jl}$$

where Λ_{jl} is the (j, l) element of L^{-1}.

From these observations it can be concluded that the following theorem holds, which enables us to construct a confidence region for A_j.

THEOREM. $(N-k)\dfrac{|\Delta A_j|^2}{\|r\|^2}\dfrac{1}{\Lambda_{jj}}$ *is distributed following an F-distribution with d.f.s 2 and 2(N−k), and it holds that*

$$P\left\{\frac{|\Delta A_j|}{|\hat{A}_j|}\leqq R(\delta)\right\}=\delta$$

where

$$R(\delta)=\sqrt{\frac{1}{N-k}\frac{\|r\|^2}{|\hat{A}_j|^2\Lambda_{jj}^{-1}}F(2,\ 2(N-k),\ \delta)}$$

and $F(2,\ 2(N-k),\ \delta)$ *is defined by the relation*

$$P\{F^2_{2(N-k)}\leqq F(2,\ 2(N-k),\ \delta)\}=\delta$$

where $F^2_{2(N-k)}$ *denotes a random variable which is distributed following an F-distribution with d.f.s 2 and* $2(N-k)$.

In this theorem $\|r\|^2$ represents the norm of the residual and $|\hat{A}_j|^2\Lambda_{jj}^{-1}$ represents the contribution of x_j to the norm of the projection which is proper to x_j in the sense that this contribution cannot be ascribed to any linear contributions of x_νs other than x_j. Such an interpretation of the result clarifies under what circumstances we can expect small values of $R(\delta)$.

Here we shall introduce three quantities $\gamma^2_{0,\,12\cdots k}$, $\gamma^2_{0j,\,12\cdots\hat{j}\cdots k}$ and $\eta^2_{j,\,12\cdots\hat{j}\cdots k}$ which are defined by the relations

$$E\|n\|^2=(1-\gamma^2_{0,\,12\cdots k})E\|x_0\|^2$$

$$E\|n\|^2=(1-\gamma^2_{0j,\,12\cdots\hat{j}\cdots k})E(\|n\|^2+|A_j|^2\Lambda_{jj}^{-1})$$

$$|A_j|^2\Lambda_{jj}^{-1}=\eta^2_{j,\,12\cdots\hat{j}\cdots k}\|\sum_{\nu=1}^{k}A_\nu x_\nu\|^2.$$

By analogy to the multiple regression analysis in real case we shall call $\gamma^2_{0,\,12\cdots k}$ the multiple coherency between x_0 and $x_1, x_2, \cdots, x_k$ and $\gamma^2_{0j,\,12\cdots\hat{j}\cdots k}$ the partial coherency between x_0 and x_j for a given set of values $x_\nu(\nu\neq j,\ \nu=1,\ 2,\ \cdots,\ k)$*). Obviously $\gamma^2_{0,\,12\cdots k}$ and $\gamma^2_{0j,\,12\cdots\hat{j}\cdots k}$ are the direct analogues of the squares of multiple and partial correlations of real variable case, respectively. We shall denote sample values of these quantities by $\hat{\gamma}^2_{0,\,12\cdots k}$, $\hat{\gamma}^2_{0j,\,12\cdots\hat{j}\cdots k}$ and $\hat{\eta}^2_{j,\,12\cdots\hat{j}\cdots k}$. They are given by the relations

$$\|r\|^2=(1-\hat{\gamma}^2_{0,\,12\cdots k})\|x_0\|^2$$

$$\|r\|^2=(1-\hat{\gamma}^2_{0j,\,12\cdots\hat{j}\cdots k})(\|r\|^2+|\hat{A}_j|^2\Lambda_{jj}^{-1})$$

$$|\hat{A}_j|^2\Lambda_{jj}^{-1}=\hat{\eta}^2_{j,\,12\cdots\hat{j}\cdots k}\|\sum_{\nu=1}^{k}\hat{A}_\nu x_\nu\|^2.$$

*) These definitions are in accordance with those in the paper by Tick [7]. The use of the terminology "coherence function" and "partial coherence function" in the paper by Enochson and Piersol was also suggested by Mr. Piersol in the formerly cited letter to the present author. Nevertheless, as the author does not know the strict definition of their coherence functions, the present definition of "partial coherency" may be used in a different context from their "partial coherence function."

These relations and the relations

$$E\,||x_0||^2 = E\,||n||^2 + ||\sum_{j=1}^{k} A_j x_j||^2$$

$$||x_0||^2 = ||r||^2 + ||\sum_{j=1}^{k} \hat{A}_j x_j||^2$$

can be used conveniently to get useful expressions of the results obtained in this section. Taking into account that $E\,||n||^2 = N \times 2\sigma^2$ we can get, by using the present relations, such expressions as

$$\frac{E\,|\Delta A_j|^2}{|A_j|^2} = \frac{1}{N}\left(\frac{1-\gamma^2_{0j,\,12\cdots\hat{j}\cdots k}}{\gamma^2_{0j,\,12\cdots\hat{j}\cdots k}}\right)$$

$$= \frac{1}{N}\left(\frac{1-\gamma^2_{0,\,12\cdots k}}{\gamma^2_{0,\,12\cdots k}\,\eta^2_{j,\,12\cdots\hat{j}\cdots k}}\right)$$

and the corresponding

$$R^2(\delta) = \frac{1}{N-k}\left(\frac{1-\hat{\gamma}^2_{0j,\,12\cdots\hat{j}\cdots k}}{\hat{\gamma}^2_{0j,\,12\cdots\hat{j}\cdots k}}\right) F(2,\,2(N-k),\,\delta)$$

$$= \frac{1}{N-k}\left(\frac{1-\hat{\gamma}^2_{0,\,12\cdots k}}{\hat{\gamma}^2_{0,\,12\cdots k}\,\hat{\eta}^2_{j,\,12\cdots\hat{j}\cdots k}}\right) F(2,\,2(N-k),\,\delta)\,.$$

4. Harmonic analysis of a stationary time series

In practical applications of the method of estimation of the frequency response function, we often encounter with the case where it is natural to consider the additive noise $n(t)$ to be a sample function of a stationary stochastic process, and observations are made only over a time interval of finite length, say $[-T,\,T]$. Thus we shall hereafter consider n(t) to be identical in this interval with a stationary process $s(t)$, defined over the whole line of t, and to be equal to zero outside the interval. We shall assume $s(t)$ to be real and strictly stationary with $Es(t)=0$ and $E\,|s(t)|^2<\infty$. Further it is assumed that $R_s(\tau)=Es(t+\tau)s(t)$ is a continuous function of τ and $\int_{-\infty}^{\infty}|R_s(\tau)|\,d\tau<\infty$. In this case $s(t)$ has a bounded and uniformly continuous spectral density function $p_s(f)$ which is given by $p_s(f)=\int_{-\infty}^{\infty}\exp(-i2\pi f\tau)R_s(\tau)d\tau$. We shall define for any function $z(t)$ which is integrable over any finite interval of t its modified Fourier transform $Z_T(f)$ by

$$Z_T(f) = \frac{1}{\sqrt{2T}}\int_{-\infty}^{\infty}\exp(-i2\pi f t)\,z_T(t)\,dt$$

where $T<\infty$ and $z_T(t)=z(t)$ $(|t|\leqq T)$, $=0$ $(|t|>T)$. Under the present assumptions of $s(t)$, we can expect that the square of the modulus of $s_T(t)$, and a fortiori $s_T(t)$ itself, is integrable with probability one [5, chapter 2, section 2]. Thus for $n(t)$ its modified Fourier transform $N_T(f)$ exists and is equal to $S_T(f)$. Hereafter we shall analyze the statistical properties of $\left\{N_T(f_\nu);\ f_\nu=\frac{\nu}{2T}\ \ \nu=0,\ \pm 1,\ \pm 2,\ \cdots\right\}$. Obviously $N_T(f_\nu)/\sqrt{2T}$s are the Fourier coefficients of a periodic signal which is of period $2T$ and is identical to $n(t)$ almost everywhere in the interval $[-T, T]$. As $En(t)=0$ we have $EN_T(f_\nu)=0$ $(\nu=0, \pm 1, \pm 2, \cdots)$. Here we adopt the spectral representation of the stationary process $s(t)$

$$s(t)=\int_{-\infty}^{\infty}\exp(i2\pi ft)dS(f)$$

where $S(f)$ is an orthogonal process with $E|dS(f)|^2=p_s(f)df$. Taking into account the relation $N_T(f)=S_T(f)$ we have

$$N_T(f_\nu)=\int_{-\infty}^{\infty}W_T(f_\nu-f)dS(f)$$

where

$$W_T(f)=\frac{1}{\sqrt{2T}}\int_{-T}^{T}\exp(-i2\pi ft)dt$$

or the modified Fourier transform of $w(t)=1$.

From this representation of $N_T(f_\nu)$ we can see that

$$EN_T(f_\nu)\overline{N_T(f_\mu)}=\int_{-\infty}^{\infty}W_T(f_\nu-f)\overline{W_T(f_\mu-f)}p_s(f)df$$

where $p_s(f)$ is the power spectral density function of the process $s(t)$. By using Parseval's equality we get

$$\int_{-\infty}^{\infty}|W_T(f)|^2df=\frac{1}{2T}\int_{-T}^{T}dt$$
$$=1,$$

and from this and the relation

$$\int_{-\delta(T)}^{\delta(T)}|W_T(f)|^2df=\int_{-T\delta(T)}^{T\delta(T)}|W_1(f)|^2df$$

we can see that for any $\delta(T)>0$

$$\lim_{T\to\infty}\int_{-\delta(T)}^{\delta(T)}|W_T(f)|^2df=1$$

holds if only $T\delta(T)\to\infty(T\to\infty)$. Now as $p_s(f)$ is uniformly continuous, there exists for any $\varepsilon(>0)$ such a δ that $|p(f_1)-p(f_2)|<\varepsilon$ holds if only f_1 and f_2 satisfy the relation $|f_1-f_2|<\delta$. Here we shall choose a $\delta(T)$ (>0) which satisfies

$$\delta(T)\to 0 \text{ and } T\delta(T)\to\infty \text{ as } T\to\infty.$$

Then we can show that for any $\varepsilon>0$

$$\left| EN_T(f_\nu)\overline{N_T(f_\mu)}-p_s\left(\frac{f_\nu+f_\mu}{2}\right)\int_{-\infty}^{\infty} W_T(f_\nu-f)\overline{W_T(f_\mu-f)}df \right| <\varepsilon$$

holds if only $|f_\nu-f_\mu|<2\delta(T)$ and T is sufficiently large.

Further, from Parseval's equality we can show that

$$\int_{-\infty}^{\infty} W_T(f_\nu-f)\overline{W_T(f_\mu-f)}df=1 \qquad (f_\nu-f_\mu)$$

$$=0 \qquad (f_\nu\neq f_\mu).$$

Thus we get a conclusion: if T is sufficiently large then for f_ν, and f_μ satisfying the relation $|f_\nu-f_\mu|<2\delta(T)$ it holds that

$$EN_T(f_\nu)\overline{N_T(f_\mu)}\approx p(f_\nu) \qquad (f_\nu=f_\mu)$$

$$\approx 0 \qquad (f_\nu\neq f_\mu)$$

where $\approx$ denotes that the difference between the both side members of $\approx$ is bounded by an arbitrary but preassigned positive constant which is independent of the choise of f_ν, and f_μ. When $|f_\nu-f_\mu|\geqq 2\delta(T)$ holds we have

$$|EN_T(f_\nu)\overline{N_T(f_\mu)}|\leqq B\int_{-\infty}^{\infty} |W_T(f_\nu-f)|\,|W_T(f_\mu-f)|\,df$$

$$=B\left[\int_1+\int_2\right]$$

where B is the largest value of $p_s(f)$ and

$$\int_1=\int_{|f_\nu-f|<\delta(T)} |W_T(f_\nu-f)|\,|W_T(f_\mu-f)|\,df$$

$$\int_2=\int_{|f_\nu-f|\geqq\delta(T)} |W_T(f_\nu-f)|\,|W_T(f_\mu-f)|\,df.$$

Here we have from Schwartz's inequality

$$\int_2 \leqq \left[\int_{|f_\nu - f| \geqq \delta(T)} |W_T(f_\nu - f)|^2 df \int_{|f_\nu - f| \geqq \delta(T)} |W_T(f_\mu - f)|^2 df\right]^{1/2}$$

$$\leqq \left[1 - \int_{-\delta(T)}^{\delta(T)} |W_T(f)|^2 df\right]^{1/2}.$$

Taking into account that in the present case the condition $|f_\nu - f| < \delta(T)$ implies $|f_\mu - f| \geqq \delta(T)$, we can get

$$\int_1 \leqq \int_{|f_\mu - f| \geqq \delta(T)} |W_T(f_\nu - f)| |W_T(f_\mu - f)| df.$$

The right hand side member of this inequality is just equal to $\int_2$ and by using the above evaluation of $\int_2$ we get the second conclusion: if T is sufficiently large then for any f_μ and f_ν satisfying the relation $|f_\mu - f_\nu| \geqq 2\delta(T)$ it holds that

$$EN_T(f_\nu)\overline{N_T(f_\mu)} \approx 0$$

where the meaning of the symbol $\approx$ is the same as that defined in the former conclusion.

By combining these two conclusions we get the following:

THEOREM. *As T tends to infinity it holds that*

$$EN_T(f_\nu)\overline{N_T(f_\mu)} = p(f_\nu) + o(1) \qquad (f_\nu = f_\mu)$$

$$= 0 + o(1) \qquad (f_\nu \neq f_\mu)$$

where $o(1)$ denotes a quantity which is bounded, uniformly with respect to f_ν and f_μ, by a constant which tends to zero as T tends to infinity.

By using the formerly defined symbol $\approx$ the relation just stated in the theorem can be expressed as

$$EN_T(f_\nu)\overline{N_T(f_\mu)} \approx p(f_\nu) \qquad (f_\nu = f_\mu)$$

$$\approx 0 \qquad (f_\nu \neq f_\mu).$$

Now as our $s(t)$ is assumed to be real it follows that

$$N_T(f_\nu) = \overline{N_T(-f_\nu)},$$

and we can get the relation

$$EN_T(f_\nu)N_T(f_\mu) = EN_T(f_\nu)\overline{N_T(-f_\mu)}$$

$$\approx p_s(f_\nu) \qquad (f_\mu=-f_\nu)$$
$$\approx 0 \qquad (f_\mu\neq -f_\nu)\,.$$

For a set of general complex random variables $X_1=U_1+iV_1$ and $X_2=U_2+iV_2(U_j,\ V_j$ real), the conditions $EX_1\bar{X}_2=EX_1X_2=0$ means $EU_1U_2=EV_1V_2=EV_1U_2=EV_2U_1=0$ and the condition $EX_1X_1=0$ means $EU_1^2=EV_2^2$. Combined with these facts, above observations lead us to the following:

THEOREM. *When T tends to infinity, if the distribution of $\{N_T(f_\nu): \nu=0,\ 1,\ 2,\ \cdots\}$ tends to be Gaussian, the real part and imaginary part of $N_T(f_\nu)$ tend to be mutually independent and with the same variance $p_s(f_\nu)/2$, with the exception of $N_T(f_0)$ which has identically vanishing imaginary part and has a real part with variance tending to $p_s(f_0)$.*

If we can assume the process $s(t)$ to be Gaussian, then the final statement of the theorem holds true. It seems that even if the original process $s(t)$ is not Gaussian, for a wide class of $s(t)$, $N_T(f_\nu)$s may often be expected to be nearly Gaussian, at least in a limited range of frequency, by a type of central limit law which assures the asymptotic Gaussian property of the outputs of very narrow band filters having a wide band noise [6]. Thus, in the practical applications, it is expected that the results in the following sections of the present paper, obtained under the assumption of normality of $N_T(f_\nu)$, will serve as a fairly good approximation to the reality.

5. Estimation of the frequency response function

In this section we shall turn to the estimation of the frequency response function. Let us recall the fundamental relation in section 3

$$x_0(f)=\sum_{j=1}^{k} A_j(f)x_j(f)+n(f)\,.$$

Corresponding to this relation we have for an observation of finite duration

$$X_{0T}(f_\nu)=\sum_{j=1}^{k} A_j(f_\nu)X_{jT}(f_\nu)+J(f_\nu)+N_T(f_\nu) \qquad (\nu=0,\ \pm1,\ \pm2,\ \cdots)$$

where $J(f_\nu)$ represents the effect of the transients in the system at $t=-T$ and T. Assuming that T is sufficiently large compared with the decay time of the transient response of the system, we shall consider the approximate relation

$$X_{0T}(f_\nu)=\sum_{j=1}^{k} A_j(f_\nu)X_{jT}(f_\nu)+N_T(f_\nu) \qquad (\nu=0,\ \pm1,\ \pm2,\ \cdots)\,,$$

which is exact only for the cyclic inputs with period $2T$. It should be remembered that for the practical application of the results of the following analysis $N_T(f_\nu)$ should be understood as $J(f_\nu)+N_T(f_\nu)$ and thus even if $n(t)\equiv 0$ there still remains the effect of transients in the system at $t=-T$ and T as a disturbance to the observation.

In many practical applications, as we have discussed briefly at the end of the last section, it will be a fairly good approximation to the reality to assume that the distribution of $\{N_T(f_\nu)\ \nu=0, 1, 2, \cdots\}$ tends to be Gaussian as T tends to infinity. In this case, it can be seen from the theorem of the last section that $N_T(f_\nu)$s ($\nu=0, 1, 2, \cdots$) tend to be mutually independent. Taking into account of the definition $f_\nu=\dfrac{\nu}{2T}$, we can see that in the relation

$$X_{0T}(f_\nu)=\sum_{j=1}^{k} A_j(f_\nu)X_{jT}(f_\nu)+N_T(f_\nu)$$

$$(\nu=L+1,\ L+2,\ \cdots,\ L+N)$$

$f_{L+N}-f_{L+1}$ can be made so small, by making T sufficiently large, as to make $p_s(f_\nu)$ and $A_j(f_\nu)$ to be nearly equal to some constants p_s and A_j respectively ($j=1, 2, \cdots, k$) for f_νs ($\nu=L+1, L+2, \cdots, L+N$).

As an approximation to this situation, we take up here the model

$$X_{0T}(f_\nu)=\sum_{j=1}^{k} A_jX_{jT}(f_\nu)+N_T(f_\nu)$$

$$(\nu=L+1,\ L+2,\ \cdots,\ L+N)\,,$$

where $N_T(f_\nu)$s are assumed to be independent and Gaussian with $EN_T(f_\nu)=0$ and $E(\mathrm{Re}\, N_T(f_\nu))^2=E(\mathrm{Im}\, N_T(f_\nu))^2=p_s/2$. For this simplified model, by using the theorems in sections 3 and 4, we can get an estimate $(\hat{A}_1, \hat{A}_2, \cdots, \hat{A}_k)$ of $(A_1, A_2, \cdots, A_k)$:

$$\begin{bmatrix} \hat{A}_1 \\ \hat{A}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{A}_k \end{bmatrix} = \begin{bmatrix} (X_1, X_1) & (X_2, X_1) & \cdots & (X_k, X_1) \\ (X_1, X_2) & (X_2, X_2) & \cdots & (X_k, X_2) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ (X_1, X_k) & (X_2, X_k) & \cdots & (X_k, X_k) \end{bmatrix}^{-1} \begin{bmatrix} (X_0, X_1) \\ (X_0, X_2) \\ \cdot \\ \cdot \\ \cdot \\ (X_0, X_k) \end{bmatrix}$$

where

$$(X_j, X_l)=\sum_{\nu=L+1}^{L+N} X_{jT}(f_\nu)\overline{X_{lT}(f_\nu)}\,.$$

We get, from the result of section 3,

$$\frac{E\,|\Delta A_j|^2}{|A_j|^2}=\frac{1}{N}\,\frac{1-\gamma^2_{0j,\,12\cdots\hat{j}\cdots k}}{\gamma^2_{0j,\,12\cdots\hat{j}\cdots k}}$$

where $\gamma^2_{0j,\,12\cdots\hat{j}\cdots k}$ denotes the partial coherency between X_0 and X_j. Naturally, this $\gamma^2_{0j,\,12\cdots\hat{j}\cdots k}$ will be called the partial coherency between $x_0(t)$ and $x_j(t)$ at this frequency.

Confidence region for A_j can be obtained using the sample partial coherency entirely analogously as in section 3, and we shall not repeat it here. For practical applications of this last simplified model, the condition $A_j(f_\nu)=A_j(\nu=L+1,\ L+2,\ \cdots,\ L+N;\ j=1,\ 2,\ \cdots,\ k)$ should be satisfied at least approximately, otherwise the bias of the present estimate $(\hat{A}_1,\ \hat{A}_2,\ \cdots,\ \hat{A}_k)$ can not be ignored.

6. Practical estimation procedure

Even though the high speed computers are available today, to follow directly the estimation procedure described in the last section will cost too much to make the procedure practically useless, and we shall propose in this section a substitute for the quantity $(X_j,\ X_l)$. Obviously $X_{jT}(f_\nu)\overline{X_{lT}(f_\nu)}/2T$ $(\nu=0,\ \pm1,\ \pm2,\ \cdots)$ are the Fourier coefficients of $\tilde{C}_{jl}(t)$ which is given by

$$\tilde{C}_{jl}(t)=\frac{1}{2T}\int_{-T}^{T} x_j^*(t+\tau)x_l(\tau)d\tau$$

where $x_j^*(t)$ is defined almost everywhere by the relation $x_j^*(t)=x_j(t)$ for $|t|\leqq T$ and $x_j^*(2nT+t)=x_j^*(t)$ (n: integer). Then for a properly selected function $D(t)$, which is usually called lag window and is periodic with period $2T$ in this case, $D(t)\tilde{C}_{jl}(t)$ has Fourier coefficients

$$\{\sum_{\mu} w_\mu X_{jT}(f_{\nu-\mu})\overline{X_{lT}(f_{\nu-\mu})};\ \nu=0,\ \pm1,\ \pm2,\ \cdots\}$$

where $2Tw_\mu$s are the Fourier coefficients of $D(t)$. By using these $\sum_{\mu} w_\mu X_{jT}(f_{\nu-\mu})\overline{X_{lT}(f_{\nu-\mu})}$ in place of $(X_j,\ X_l)$ of the above formula of $(\hat{A}_1,\ \hat{A}_2,\ \cdots,\ \hat{A}_k)$ we can get an estimate of $(A_1(f_\nu),\ A_2(f_\nu),\ \cdots,\ A_k(f_\nu))$. The advantages of adopting such substitution will be as follows: a) by using a $D(t)$ which vanishes outside an interval $[-L,\ L]$ $(L<T)$ necessary range of computation of $\tilde{C}_{jl}(t)$ is reduced to $[-L,\ L]$ from $[-T,\ T]$, b) undesirable effect of local variation of $A_j(f)$ may be reduced by proper selection of $D(t)$. We shall not go into the details of the analysis of the effect of adopting $D(t)$. Necessary informations will be found in the papers [2, 3].

For practical applications we shall further replace the above circular $\tilde{C}_{jl}(t)$ by a non-circular

$$C_{jl}(t)=\frac{1}{2T}\int_{-T}^{T}x_{jT}(t+\tau)x_{lT}(\tau)d\tau .$$

Adoption of this non-circular $C_{jl}(t)$ corresponds to the model where the observed output $x_{0T}(t)$ is put into correspondence with the non-circular input $x_{1T}(t), x_{2T}(t), \cdots, x_{kT}(t)$. Maximum possible bias due to the transients in the system will be reduced by the adoption of this non-circular model. In practical applications it is expected that the magnitude of the sampling variability of the estimate exceeds by far that of the bias here considered and the main reason for the adoption of the non-circular definition is its expected reduction of sampling variability. By ignoring the effect of the transients in the system we get a (approximate) relation

$$\tilde{C}_{0l}(t)=\sum_{j=1}^{k}\int_{-\infty}^{\infty}h_j(\sigma)\tilde{C}_{jl}(t-\sigma)d\sigma+\tilde{C}_{nl}(t) \qquad (l=1, 2, \cdots, k)$$

where

$$\tilde{C}_{nl}(t)=\frac{1}{2T}\int_{-T}^{T}n^*(t+\tau)x_l(\tau)d\tau ,$$

and a corresponding relation for non-circular $\tilde{C}_{jl}$

$$C_{0l}(t)=\sum_{j=1}^{k}\int_{-\infty}^{\infty}h_j(\sigma)C_{jl}(t-\sigma)d\sigma+C_{nl}(t) \qquad (l=1, 2, \cdots, k)$$

where

$$C_{nl}(t)=\frac{1}{2T}\int_{-T}^{T}n_T(t+\tau)x_{lT}(\tau)d\tau .$$

These relations form the basis of our estimation procedures and the sampling variabilities of our circular and non-circular estimates are caused by the existence of the terms $\tilde{C}_{nl}(t)$ and $C_{nl}(t)$.

We shall here consider the case where the input $\{x_1(t), x_2(t), \cdots, x_k(t)\}$ is a sample function of a k-dimensional stationary process. Such a circumstance will be the most common in practical applications of the present estimation procedure. In this case, it can be expected that in the expressions such as

$$\tilde{C}_{jl}(t)=C_{jl}(t)+\frac{1}{2T}\int_{T-t}^{T}x_j(t+\tau-2T)x_l(\tau)d\tau \qquad (t>0)$$

$$(j, l=1, 2, \cdots, k) ,$$

the contribution of the second term in the right hand side is usually less significant relatively than that in the expressions such as

$$\tilde{C}_{nl}(t)=C_{nl}(t)+\frac{1}{2T}\int_{T-t}^{T} n(t+\tau-2T)x_l(\tau)d\tau \qquad (t>0)$$

$$(l=1, 2, \cdots, k),$$

and analogously for $t\leqq 0$. This is due to the existence and non-existence of the correlation between the variables considered, and we can expect that the sampling variation of $C_{nl}(t)$ relative to $C_{jl}(t)$s is smaller than that of $\tilde{C}_{nl}(t)$ relative to $\tilde{C}_{jl}(t)$s.

Thus to replace $\tilde{C}_{jl}(t)$ by $C_{jl}(t)$ $(j, l=0, 1, 2, \cdots, k)$ will only result to some (minor) reduction of the sampling variability of our estimate. This is the reason why we propose the adoption of non-circular $C_{jl}(t)$ for practical estimation procedures.

Hereafter we consider the case where the sampled data $\{x_0(n\Delta t), x_1(n\Delta t), x_2(n\Delta t), \cdots, x_k(n\Delta t); n=1, 2, \cdots, M\}$ are given. We shall assume that $\Delta t(>0)$ is sufficiently small so that the effects due to aliasing are negligibly small. For this case we adopt the following estimation procedure:

1) $C_{jl}(m)$ $(m=0, \pm 1, \pm 2, \cdots, \pm h; j, l=0, 1, 2, \cdots, k)$ are calculated by the following formulae

$$C_{jl}(m)=\frac{1}{M}\sum_{n=1}^{M-m}\tilde{x}_j(m+n)\tilde{x}_l(n) \qquad (h>m\geqq 0)$$

$$=\frac{1}{M}\sum_{n=1-m}^{M}\tilde{x}_j(m+n)\tilde{x}_l(n) \qquad (-h<m<0)$$

and

$$C_{jl}(h)=\frac{1}{2M}\sum_{n=1}^{M-h}\tilde{x}_j(h+n)\tilde{x}_l(n)$$

$$C_{jl}(-h)=\frac{1}{2M}\sum_{n=1+h}^{M}\tilde{x}_j(-h+n)\tilde{x}_l(n)$$

where

$$\tilde{x}_j(n)=x_j(n\Delta t)-\bar{x}_j$$

$$\bar{x}_j=\frac{1}{M}\sum_{n=1}^{M}x_j(n\Delta t).$$

2) Numerical Fourier transforms $\bar{p}_{jl}(r)$ $(r=0, 1, 2, \cdots, h)$ of $C_{jl}(m)$

are then calculated by the formula

$$\bar{p}_{jl}(r)=\sum_{m=-h}^{h}\exp\left(-i2\pi\frac{r}{2h}m\right)C_{jl}(m) .$$

3) $\bar{p}_{jl}(r)$ are smoothed by a properly selected weights $\{a_n\}$ to give $\hat{p}_{jl}(r)=\sum_{n=-d}^{d}a_n\bar{p}_{jl}(r-n)$.

4) Estimates of $\left\{A_1\left(\frac{r}{2h\Delta t}\right), A_2\left(\frac{r}{2h\Delta t}\right), \cdots, A_k\left(\frac{r}{2h\Delta t}\right)\right\}$ and the necessary quantities for the construction of the confidence regions are obtained by the following procedure:

4.1) We arrange the $\hat{p}_{jl}(r)$ in the following matrix form

$$\begin{array}{c} \text{columns} \\ \text{rows} \end{array} \quad \begin{array}{r} \\ 1 \\ 2 \\ \cdot \\ \cdot \\ \cdot \\ k \\ k+1 \end{array} \begin{array}{cccccccccc} 1 & 2 & \cdots & k & k+1 & k+2 & k+3 & \cdots & 2k+1 \\ \hat{p}_{11}(r), & \hat{p}_{12}(r), & \cdots, & \hat{p}_{1k}(r) & \hat{p}_{10}(r) & 1 & 0 & \cdots & 0 \\ \hat{p}_{21}(r), & \hat{p}_{22}(r), & \cdots, & \hat{p}_{2k}(r) & \hat{p}_{20}(r) & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \hat{p}_{k1}(r), & \hat{p}_{k2}(r), & \cdots, & \hat{p}_{kk}(r) & \hat{p}_{k0}(r) & 0 & 0 & \cdots & 1 \\ \hat{p}_{01}(r), & \hat{p}_{02}(r), & \cdots, & \hat{p}_{0k}(r) & \hat{p}_{00}(r) & 0 & 0 & \cdots & 0 \end{array} .$$

Here it should be noted that $\hat{p}_{jl}(r)=\bar{\hat{p}}_{lj}(r)$ holds, where the bar denotes the complex conjugate.

4.2) Let us divide the 1st row by $\hat{p}_{11}(r)$ and then subtract from the jth row the 1st row multiplied by $\hat{p}_{j1}(r)$ $(j=2, 3, \cdots, k+1)$.

4.3) We then divide the 2nd row by the (2, 2) element of the matrix and then subtract from the jth row the 2nd row multiplied by the $(j, 2)$ element of the matrix $(j=1, 3, \cdots, k+1)$.

4.4) The operation on the rows of the matrix is continued up to the kth step where we get the final form of the matrix

$$\begin{array}{r} \\ 1 \\ 2 \\ \cdot \\ \cdot \\ \cdot \\ k \\ k+1 \end{array} \begin{array}{cccccccccc} 1 & 2 & \cdots & k & k+1 & k+2 & k+3 & & 2k+1 \\ 1 & 0 & \cdots & 0 & \alpha_1 & r_{11} & r_{12} & \cdots & r_{1k} \\ 0 & 1 & \cdots & 0 & \alpha_2 & r_{21} & r_{22} & \cdots & r_{2k} \\ \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & 1 & \alpha_k & r_{k1} & r_{k2} & \cdots & r_{kk} \\ 0 & 0 & \cdots & 0 & \varepsilon & \beta_1 & \beta_2 & \cdots & \beta_k \end{array} .$$

5) Our estimate $\hat{A}_j(r)$ of $A_j\left(\frac{r}{2h\Delta t}\right)$ is given by

$$\hat{A}_j(r) = -\beta_j$$
$$= \bar{\alpha}_j \, .$$

The sample multiple coherency $\hat{\gamma}^2_{0,\, 12\cdots k}(r)$ and the sample partial coherency $\hat{\gamma}^2_{0j,\, 12\cdots \hat{j}\cdots k}(r)$ at the frequency $\frac{r}{2h\Delta t}$ are given by

$$\hat{\gamma}^2_{0,\, 12\cdots k}(r) = 1 - \frac{\varepsilon}{\hat{p}_{00}}$$

$$\hat{\gamma}^2_{0j,\, 12\cdots \hat{j}\cdots k}(r) = \frac{|\alpha_j|^2 r^{-1}_{jj}}{\varepsilon + |\alpha_j|^2 r^{-1}_{jj}} \, .$$

6) It is expected that the following equation holds approximately;

$$P\left(\left| \hat{A}_j(r) - A_j\left(\frac{r}{2h\Delta t}\right) \right| \leqq \left| \hat{A}_j(r) \right| R_{j,\delta}(r) \right) = \delta$$

where

$$R_{j,\delta}(r) = \left(\frac{1}{N-k} \frac{\varepsilon r_{jj}}{|\alpha_j|^2} F(2,\, 2(N-k),\, \delta) \right)^{1/2},$$

$F(2,\, 2(N-k),\, \delta)$ is as defined in section 3, and $N=$ the integer nearest to $\frac{M}{h} \frac{1}{2\sum_{n=-d}^{d} |a_n|^2}$. We use this $R_{j,\delta}(r)$ as an indicator of the sampling variability of $\hat{A}_j(r)$. By assuming the above stated probability relation, we can use this $R_{j,\delta}(r)$ to construct an approximate confidence region for $A_j(r)$.

Comments: a) As a possible selection of the weights $\{a_n\}$ to be used at the stage 3), we shall propose the successive application of the following sets [1]:

	a	$a_1 = a_{-1}$	$a_2 = a_{-2}$	$a_3 = a_{-3}$
W_1:	(0.5132	0.2434	0	0)
W_2:	(0.6398	0.2401	−0.0600	0)
W_3:	(0.7029	0.2228	−0.0891	0.0149).

As to the decision when a significant difference between the results is observed, the statements in the preceding paper [1, pp. 11–12] will be available.

b) For the effective use of the time shift operation, by which is meant to use $C_{0j}(m+K_j)$ in place of $C_{0j}(m)$ for properly selected K_j, to compensate for the rapid change of the phase shift of the frequency response function also confer the paper [3]. Such an operation may be applied even for $C_{lj}(m)$ with $l \neq 0$ to reduce the bias due to the adoption of a lag window.

c) The definition of N for the approximate equality of 6) is obtained following the considerations in the former papers [2, 3]. It is still tentative and may need further theoretical investigation to assure the validity of approximation. Nevertheless, as will be seen in the numerical examples in the following section, such an approximation provides a quite valuable information of the sampling variability of our estimates in practical applications. This fact was also recognized empirically in the applications of the method to the case of single input [1, p. 15].

d) In the definition of $R_{j,\delta}(r)$ at the both side ends of the frequency range which correspond to $r=0$ and h, $F(2, 2(N-k), \delta)$ should be replaced by $F(1, N-k, \delta)$ which is obtained by replacing $F^2_{2(N-k)}$ by F^1_{N-k} in the definition of $F(2, 2(N-k), \delta)$.

e) When someone of the (j, j) elements loses the significant digits during the process of computation of the stage 4), computation should be stopped. In this case somewhat finer analysis of the dependency between the inputs is necessary.

7. Numerical examples

In this section we shall show some numerical examples of application of our estimation procedure. Here we treat the case $k=2$. The inputs $x_1(t)$ and $x_2(t)$ were taken from a real record of some physical phenomenon and it was expected that they were highly correlated.

After the time-sampling, the data $\{x_1(n\Delta t)\}$ and $\{x_2(n\Delta t)\}$ were smoothed by numerical filters with frequency response functions $A_1(f)$ and $A_2(f)$ respectively. The outputs of the two filters were added together to produce the output of the (artificial) system.

The first portion of the data was then discarded to simulate the practical observation of the system operating under the stationary inputs and the total of $M=800$ points were used for the estimation. An nearly white noise of rather low level was then added to the output of the system to produce the observed values $\{x_0(n\Delta t)\}$. Numerical values of $C_{jl}(m)$ $(j, l=0, 1, 2)$ are presented in Fig. 1. The power spectra of the inputs and output are shown in Fig. 2. The level of the power of $x_2(t)$ is low compared with that of $x_1(t)$. As will be seen in Fig. 4 the gains $|A_1(f)|$ and $|A_2(f)|$ are of the same order of magnitude and thus it can be expected that to get a good estimate of $A_2(f)$ is fairly difficult. Our estimates $\hat{A}_1(r)$ and $\hat{A}_2(r)$ of $A_1\left(\frac{r}{2h\Delta t}\right)$ and $A_2\left(\frac{r}{2h\Delta t}\right)$ are illustrated in Fig. 3 for the case $h=100$. Two kind of estimates, one obtained by using the window W_1 and the other by W_2, are illustrated with the corresponding theoretical values of the frequency response functions. In Fig. 4 are illustrated $|\hat{A}_j(r)|$s with the corresponding $R_{j,\delta}(r)$s with $\delta=0.95$ for the case $h=60$. In this figure, values of $R_{j,\delta}(r)$ are missing for those r where the sample multiple coherency is greater than 1 or less than 0. As is obvious from the Fig. 2, the power spectra of the inputs show significant variations at some frequencies and $h=60$ is inadequately small to avoid the undesirable effects of the side robes of the lag window. This causes the appearences of some quite irregular numerical values in the result obtained by applying W_2. In the same figure, estimates of the phase shifts are partly illustrated. Obviously we should have introduced some time shift operation to compensate for the rapid change of the phase shifts at the frequencies corresponding to the peaks of the gains. The results show clearly that $R_{j,\delta}(r)$ with $\delta=0.95$ is a fairly good indicator of the sampling variability of $\hat{A}_j(r)$. In Fig. 5 is shown an estimate of $|A_2(f)|$ obtained by using the output before contamination with additive noise. As was mentioned in the former section, besides the distortions induced by the numerical procedure, the transients in the system at the beginning and at the end of the observation introduces distortion into the estimate. It seems that $R_{j,\delta}(r)$ is effective even as an indicator of this type of variability.

As was for the case of single input [1], one of our main observations in these numerical examples will be the practical utility of this $R_{j,0.95}(r)$.

It should be remembered that numerical examples in this section are only intended to show partly the practical applicability of our method under rather bad circumstances and much better results can be expected under better circumstances with fuller exploitations of the results obtained in this and the preceding papers [1, 2, 3].

Acknowledgement

The author expresses his thanks to Prof. H. Shibata of Tokyo University for the suggestion in 1962 of the practical importance of the present problem. Thanks are also due to Mr. T. Hayakawa for reading the draft of section 3 on regression analysis of complex variables, to Miss E. Arahata for programming of the electronic computer and to Miss E. Ozaki for typing.

THE INSTITUTE OF STATISTICAL MATHEMATICS

References

[1] H. Akaike, "Statistical measurement of frequency response function," *Ann. Inst. Stat. Math.*, *Supplement III* (1964), 5-17, with corrigenda in the same *Annals* 17 (1965).

[2] H. Akaike, "On the design of lag window for the estimation of spectra," *Ann. Inst. Stat. Math.*, 14 (1962), 1-21.

[3] H. Akaike and Y. Yamanouchi, "On the statistical estimation of frequency response function," *Ann. Inst. Stat. Math.*, 14 (1962), 23-56.

[4] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1958.

[5] J. L. Doob, *Stochastic Processes*, John Wiley & Sons, Inc., 1953.

[6] V. P. Leonov and A. N. Shiryaev, "Some problems in the spectral theory of higher-order moments, II," *Theory of Probability and Its Applications*, 5 (1960), 417-421, (English translation).

[7] L. J. Tick, "Conditional spectra, linear systems, and coherency," Chapter 13 of *Time Series Analysis* edited by M. Rosenblatt, John Wiley & Sons, Inc. (1963), 197-203.

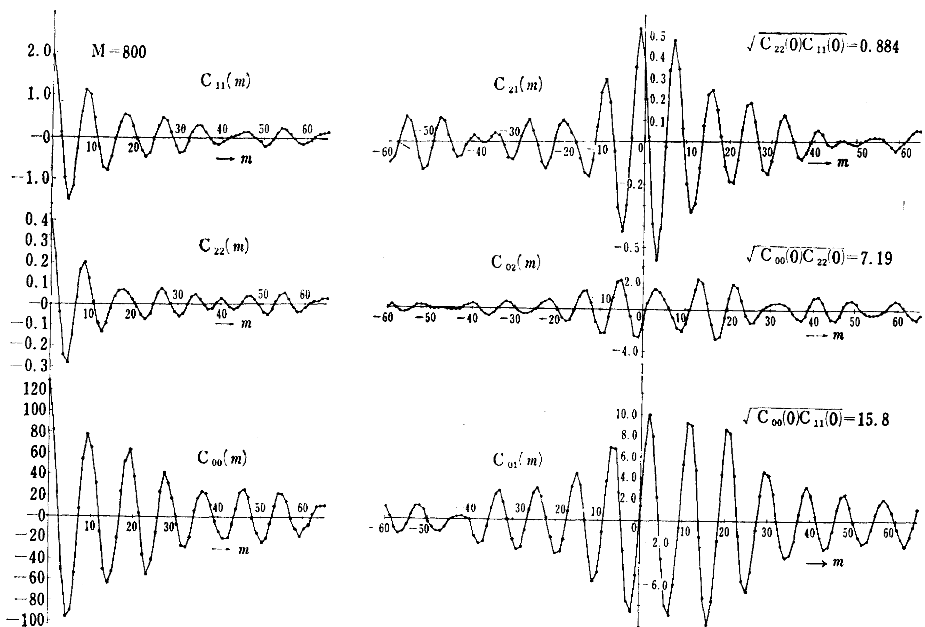

Fig. 1. Observed covariances.

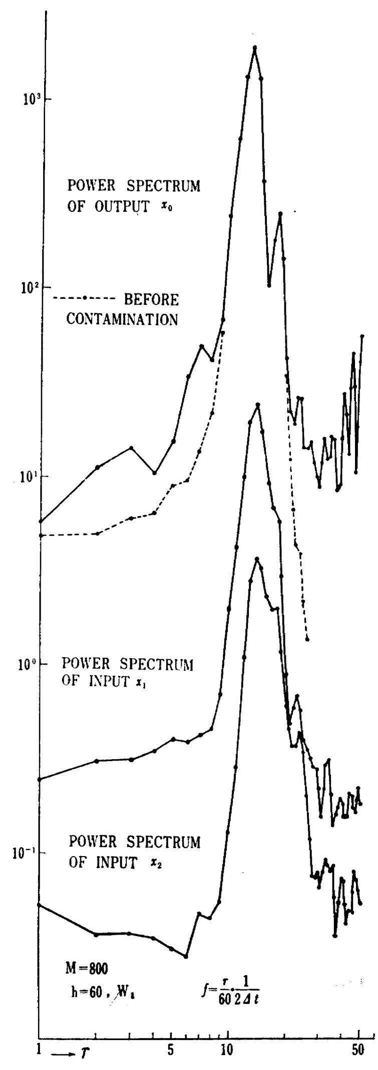

Fig. 2. Power spectra of inputs and output.

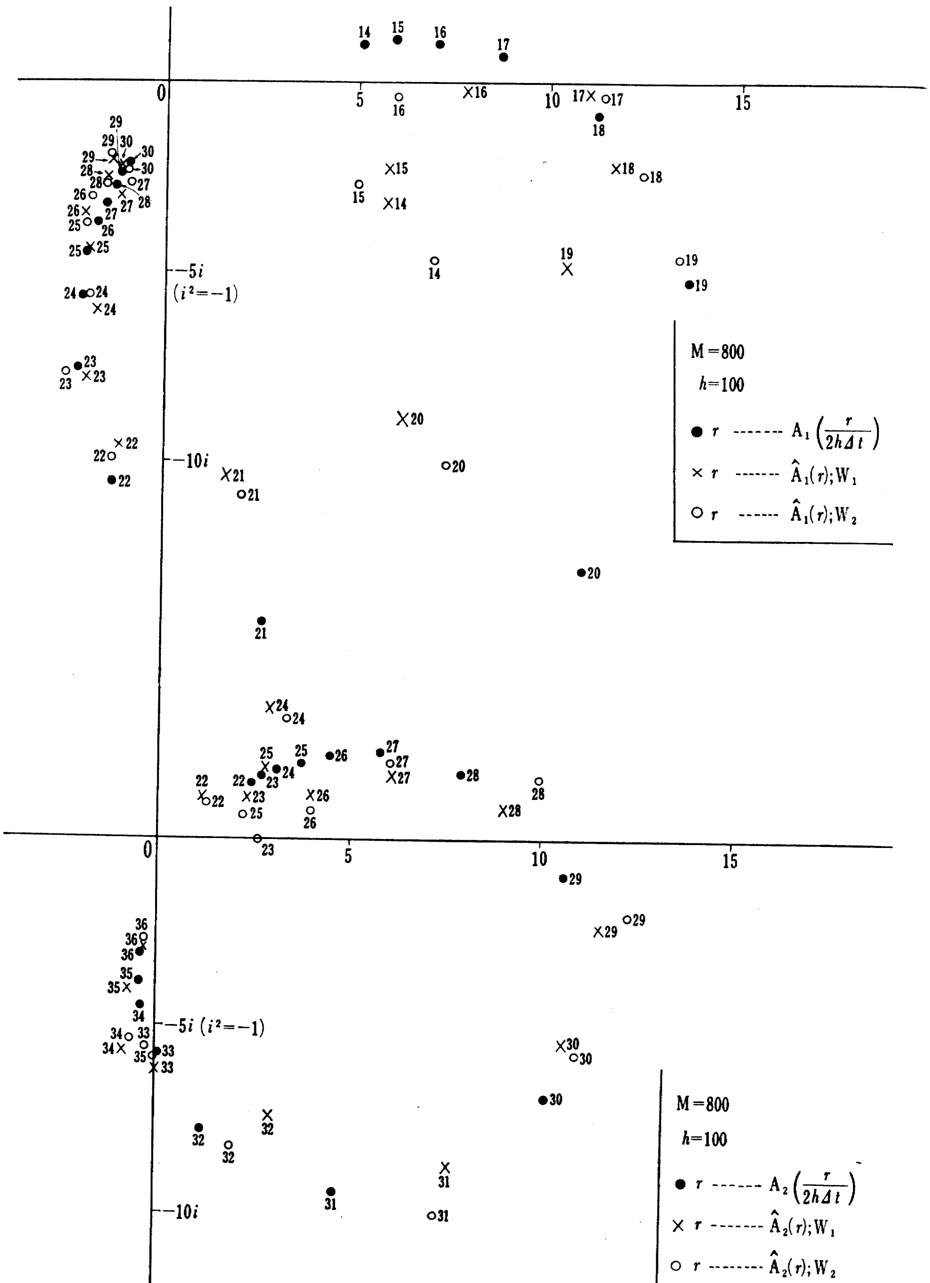

Fig. 3. Estimates of frequency response functions.

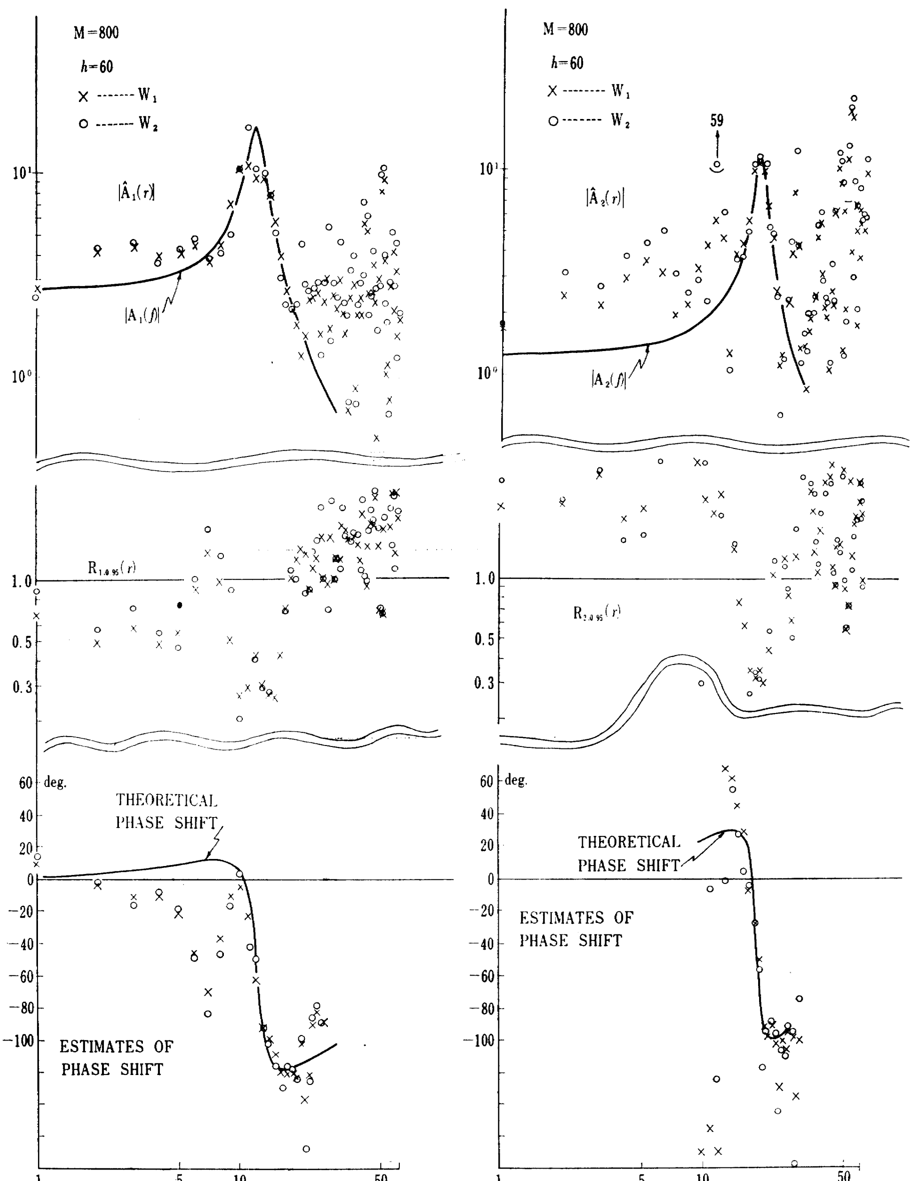

Fig. 4. The estimates and $R_{j,\delta}(r)$.

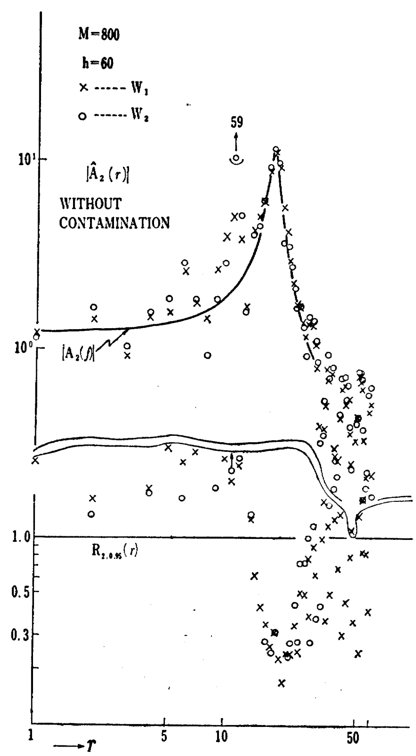

Fig. 5. Estimates when the additive noise is nonexistent.