MAXIMAL CORRELATION APPLIED TO THE STATISTICAL LINEARIZATION: AN ANALYSIS AND APPROACHES

Kirill Chernyshov

V.A. Trapeznikov Institute of Control Sciences
65 Profsoyuznaya, Moscow 117997, Russia
E-mail: myau@ipu.rssi.ru

Abstract: The paper presents an approach to the statistical linearization of the input/output mapping of a non-linear discrete-time stochastic system driven by a white-noise Gaussian process. The approach is based on applying the maximal correlation function. At that, the statistical linearization criterion is the condition of coincidence of the mathematical expectations of the output processes of the system and model, and the condition of coincidence of the joint maximal correlation functions of the output and input processes of the system and the output and input processes of the model. Explicit expressions for the weight function coefficients of the linearized model are obtained; an approach to eliminate the influence of unobservable output additive disturbances under conditions when a priori information on the type of their probability distribution is available is proposed.

Keywords: Non-parametric identification, Nonlinearity, Stochastic variables, Correlation coefficients, Single-input/single-output systems

1. AN ANALYSIS OF CONSISTENT MEASURES OF DEPENDENCE WITHIN IDENTIFICATION PROBLEMS

Statistical linearization of input/output mappings of systems under study relates to the class of problems of non-linear identification whose solution is considerably determines by characteristics of stochastic dependence of input and output processes. At that, known approaches are either based on applying conventional correlation functions, or dispersion ones, the dispersion linearization. At the same time, methods of the dispersion linearization leads out the class of linearized models.

Among various measures of dependence, the product correlation functions are well known and commonly used. However, these may vanish even provided that a deterministic dependence between input and output processes exists (Rajbman, 1981, Rényi, 1959). Also, there are known cases when actual dependence between two variables is nonlinear even provided that the regression of a variable onto another one is linear (Sarmanov and Bratoeva, 1967).

Rényi (1959) has formulated seven axioms which are seemed to be the most natural for a measure of dependence $\mu(X,Y)$ between two random variables *X* and *Y*.

A) $\mu(X,Y)$ *is defined for any pair of random variables X and Y, neither of them being constant with probability 1.*

B) $\mu(X,Y)=\mu(Y,X)$.

C) $0\le\mu(X,Y)\le 1$.

D) $\mu(X,Y)=0$ *if and only if X and Y are independent.*

E) $\mu(X,Y)=1$ *if there is a strict dependence between X and Y, i.e. either* $Y=\varphi(X)$ *or* $X=\psi(Y)$ *where* φ *and* ψ *are Borel-measurable functions.*

F) If a Borel-measurable functions φ *and* ψ *map the real axis in a one-to-one way onto itself,* $\mu(\varphi(X),\psi(Y))=\mu(X,Y)$.

G) If the joint distribution of X and Y is normal, then $\mu(X,Y)=|r(X,Y)|$, where $r(X,Y)$ is the ordinary correlation coefficient of X and Y.

Rényi (1959) has shown that a measure of dependence meeting all the above axioms is the maximal correlation:

$$S(X,Y)=\sup_{\{B\},\{C\}}\frac{\mathbf{M}(B(Y)C(X))-\mathbf{M}(B(Y))\mathbf{M}(C(X))}{\sqrt{\mathbf{D}(B(Y))\mathbf{D}(C(X))}},$$
$$\mathbf{D}(B(Y))>0,\quad \mathbf{D}(C(X))>0,$$

with here and below supremum being taken over Borel-measurable functions $\{B\}$ and $\{C\}$, and $B\in\{B\}$, $C\in\{C\}$.

When investigating random processes, the maximal correlation coefficient is transformed to the following function

$$S_{yx}(v)=\sup_{\{B\},\{C\}}\frac{\mathbf{cov}(B(y(t)),C(x(s)))}{\sqrt{\mathbf{D}(B(y(t)))\mathbf{D}(C(x(s)))}},\quad (1.1)$$
$$\mathbf{D}(B(y(t)))>0, \mathbf{D}(C(x(s)))>0,$$

with $v=t-s$. In the two above formulae, the random processes $x(s)$ and $y(t)$ are considered as jointly strongly stationary. The functional $S_{yx}(v)$ is referred as the maximal correlation function of the random processes $y(t)$ and $x(s)$.

Existence of the pair of transformations (B,C) in (1.1) is determined by conditions which are equivalent to those of used for random variables stated by Rényi (1959), Sarmanov (1963), Sarmanov and Zakharov (1960), with a basic assumption being the stochastic kernel of the random processes $\mathrm{K}(y,x,v)=\frac{p(y,x,v)}{\sqrt{p(x)p(y)}}$ meeting the condition

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{K}(y,x,v)dydx<\infty \quad (1.2)$$

for any v.

Due to (1.2), the density $p(y,x,v)$ may be represented by a corresponding bilinear eigenfunction expansion converging in mean (Sarmanov and Zakharov, 1960, Chesson, 1976).

In the paper, Section 2 reviews a recent approach oriented to applying the maximal correlation within statistical linearization, and drawbacks of such an approach are demonstrated. In Section 3, a maximal correlation approach to the statistical linearization of the input/output mapping of a non-linear discrete-time stochastic system driven by a white-noise Gaussian process is derived. Finally, in Section 4, an approach to eliminate the influence of unobservable output additive disturbances when a priori information on the type of their probability distribution is available is proposed.

2. REVISING RECENT APPROACHES

An approach applying the maximal correlation functions to the statistical linearization has recently been proposed by Pashchenko (2001, 2006), who considered the following "generalization" of the statistical linearization (the quotes here are used because the models, finally derived, are not linear (with respect to the centered input process), and hence the linearization problem is not solved and substituted by approximation of an initial system's input/output mapping by an approximation of that mapping by a non-linear one from a preliminary given class). Namely, let a non-linear plant

$$Y(t)=F(X(s),t), \quad (2.1)$$

is available, where $Y(t)$ is a random output signal, $X(s)$ is a random input one, $F(\cdot,\cdot)$ is a non-linear inertialless or dynamic transformer which may be represented by the Urysson operator, non-linear differential equation, a non-linear function, may contain δ-functions and their derivatives, logic operators, etc. (Pashchenko, 2001, 2006).

As the plant's model, Pashchenko (2001, 2006) considers the expression

$$By(t)=ACx(s), \quad (2.2)$$

from a class of models unsuccessfully referred in the cited references as "semilinear models". In (2.2) B is a non-linear transformation of the model's output process $y(t)$, C is a non-linear transformation of the model's input process $x(s)$, A is a linear mapping.

As criteria of the statistical linearization, Pashchenko (2001, 2006) considers the following ones: the first criterion is the condition of coincidence of the mathematical expectations of the plant's "output" and model's "output", the condition of coincidence of the functional auto-correlation functions of the plant's "output" and model's output (see (2.4) below and notations therein); the second criterion is the condition of the minimum of the mean square error.

Again, Pashchenko (2001, 2006) "for sake of simplicity" assumes that the class of models (2.2) contains the inverse operator B^{-1}, wile the model of plant (2.1) is searched for in the class of models

$$y(t)=B^{-1}ACx(s). \quad (2.3)$$

Following to Pashchenko (2001, 2006), consider the identification problem in accordance to the first and second criteria of the statistical linearization.

The first criterion takes in that case the form

$$\mathbf{M}y(t)=\mathbf{M}y_M(t);\quad R^{\phi}_{yy}(t,s)=R^{\phi}_{y_M y_M}(t,s), \quad (2.4)$$

where

$$y_M(t) = \\ = K_0(t)\mathbf{M}\left(B^{-1}ACx(s)\right) + K_1(t)B^{-1}AC\overset{\circ}{x}(s) \quad (2.5a)$$

or

$$y_M(t) = K_0(t)A\mathbf{M}Cx(s) + K_1(t)AC\overset{\circ}{x}(s)\,. \quad (2.5b)$$

In (2.4), $R^{\phi}_{zz}(t,s) = \mathbf{cov}\left(B(z(t)), Bz(s)\right)$ is the functional auto-correlation function of the corresponding random process (Pashchenko 2001, 2006). In (2.5), $K_0(t), K_1$ are some non-random coefficients also subject to determination.

Let, in accordance to Pashchenko (2001, 2006), A be a linear non-stationary integral operator of the form

$$Ax(t,s) = \int_0^T g(t,s)x(s)ds\,, \quad (2.6)$$

where $g(t,s) = 0$ as $s,t \notin [0,T]$; $g(t,s) \neq 0$ as $s,t \in [0,T]$.

Then, from conditions (2.4) the following model has been derived in (Pashchenko 2001, 2006)

$$\mathbf{M}y(t) = K_0(t)\mathbf{M}B^{-1}\int_0^T g(t,\tau)C[x(\tau)]d\tau\,, \quad (2.7a)$$

$$R^{\phi}_{yy}(t,s) = \\ = K_1(t)K_1(s)\int_0^T\int_0^T g(t,\tau)g(s,\lambda)R^{\phi}_{xx}(\tau,\lambda)d\tau d\lambda\,. \quad (2.7b)$$

When applying the second criterion, minimum of the mean square error, Pashchenko (2001, 2006) has written the following system of equations

$$\mathbf{M}\left((By(t)) - K_0(t)\int_T g(t,\tau)C[x(\tau)]d\tau\right) = 0\,, \quad (2.8a)$$

$$R^{\phi}_{yx}(t,s) = K_1(t)\int_T g(t,\tau)R^{\phi}_{xx}(s,\tau)d\tau\,. \quad (2.8b)$$

At that, Pashchenko (2001, 2006) indicates that solving this problem of the statistical "linearization" consists of the following four stages.

At the first stage, the operators B and C are determined in accordance to the criterion

$$(B,C) = \arg\sup_{\{B\},\{C\}} \frac{\mathbf{cov}\left(By(t), Cx(s)\right)}{\sqrt{\mathbf{D}\left(By(t)\right)\mathbf{D}\left(Cx(s)\right)}}\,, \quad (2.9)$$

$$\mathbf{D}\left(By(t)\right) = \mathbf{D}\left(Cx(s)\right) = 1\,.$$

At the second stage, the coefficient $K_1(t)$ is determined. For instance (?!),

$$K_1(t) = R^{\phi}_{yy}(t,t) \big/ R^{\phi}_{xx}(t,t)\,. \quad (2.10)$$

At the third stage, the operator A or its weight function $g(t,s)$, is determined, and, finally, at the fourth stage, the coefficient $K_0(t)$ is determined from equation (2.7) or (2.8) in dependence on the statistical linearization criterion chosen.

As seen, the first two stages of the four stages of (Pashchenko, 2001, 2006) are of declarative nature exclusively and are by no means related to the above criteria of statistical linearization. "Linearization" is of course another unsuccessful term because no linear model is derived within the considered scheme. Going back to the mentioned term "semilinear" models, the approach of Pashchenko (2001, 2006) should be referred as "statistical *semi*linearization". In the scheme presented, condition (2.9) is by no means related to the above criteria of the statistical "linearization" as a measure of association of the plant and model. Condition (2.10) is at all stated as imposed "for instance", from what it follows that the coefficient $K_1(t)$ may be chosen in arbitrary manner, "for instance" one may be set to be equal to 1 or to any other constant, or to any a priori given function, that indicates the inanition of introducing the coefficient $K_1(t)$ in models (2.5a), (2.5b) by Pashchenko (2001, 2006).

More over, equations (2.7), (2.8) themselves are neither derived in (Pashchenko 2001, 2006) no in references cited therein. At the same time, credibility of these formulae should of certainly cause doubts, taking into account at least the non-linearity of the operator B. More specifically, one may argue validity of equation (2.7a) only, which (the equation) is derived as a result of taking mathematical expectations of the left had and right hand parts of equation (2.5) (under the condition $\mathbf{M}B^{-1}AC\overset{\circ}{x}(s) = 0$).

As to resting equations, (2.7b), (2.8a), (2.8b), then one can bee seen that they are *linear* both in coefficients $K_0(t)$, $K_1(t)$, and the weight function $g(t,s)$, while model (2.5) is *linear* in the coefficients $K_0(t)$, $K_1(t)$, but is *non-linear* in the weight function $g(t,s)$ (one should also be noted that *truth* equation (2.7a) is *linear* in the coefficient $K_0(t)$, but is *non-linear* in the weight function $g(t,s)$). This circumstance confirms the assumption on invalidity of equations (2.7b), (2.8a), (2.8b).

One may try "to derive", for instance, equation (2.7b). From criterion (2.4), by virtue of the above considered stages of (Pashchenko, 2001, 2006) it follows:

$$\mathbf{M}\left(B\overset{\circ}{y}(t)B\overset{\circ}{y}(s)\right) \overset{\Delta}{=} R^{\phi}_{yy}(t,s) = \\ = R^{\phi}_{y_M y_M}(t,s) \overset{\Delta}{=} \mathbf{M}\left(B\overset{\circ}{y}_M(t)B\overset{\circ}{y}_M(s)\right) =$$

$$= \mathbf{M}\left(B\left\{ K_1(t)B^{-1}\int_T g(t,\tau)C[x(\tau)]d\tau \right\}\times \right.$$

$$\left. \times B\left\{ K_1(s)B^{-1}\int_T g(s,\lambda)C[x(\lambda)]d\lambda \right\}\right).$$

Hence, equation (2.7b) might be valid under the condition that the coefficient $K_1(t)$ commutes with the *non-linear* transformation *B*, but there are no reasons for such an assumption.

More ambiguity is present in the question on validity of equations (2.8a), (2.8b) because in (Pashchenko, 2001, 2006) under using the second criterion of the statistical linearization (minimum of the mean square error) its analytical expression is not presented. If a common mean square expression may be used as a hypothetical point for further inferences, then by virtue of the above considered four-stage scheme, in accordance to which the transformations *B* and *C* are chosen from condition (2.9), one may conclude that such a criterion is of the form

$$\min_{K_0, K_1, A} \leftarrow$$

$$\leftarrow \mathbf{M}\left(y(t) - K_0(t)\mathbf{M}\left(B^{-1}ACx(s)\right) + K_1(t)B^{-1}AC\overset{\circ}{x}(s \right)^2 \qquad (2.11).$$

Hence, the statement that equations (2.8) follow from criterion (2.11) does no look believable disregarding its mathematical justification.

Thus the above considerations demonstrate the inconsistency (in the common sense of this word) of the method of Pashchenko (2001, 2006) of the statistical “linearization”.

3. STATISTICAL LINEARIZATION BY USE OF THE MAXIMAL CORRELATION

Let a nonlinear discrete-time system be described by an input/output relationship which generically is of the form

$$y(t) = F\left(w(s), s \in I_t\right), \qquad (3.1)$$

where $y(t)$ is the system output process considered as a stationary ergodic random process; $w(s)$, the system input process which, within the problem statement, is considered as a white-noise Gaussian random process; I_t is the set of discrete times, and $t = 1,2,\ldots$. For sake of simplicity but without loss of generality, the above processes $y(t)$ and $w(s)$ are assumed to be zero mean and normalized to unity ones, i.e.

$$\mathbf{M}\{y(t)\} = \mathbf{M}\{w(s)\} = 0\,,\ \mathbf{D}\{y(t)\} = \mathbf{D}\{w(s)\} = 1\,. \qquad (3.2)$$

The processes $y(t)$ and $w(s)$ are also assumed to be joint stationary in the strict sense.

System (3.1) model will be searched for in the following form

$$\hat{y}(t;G) = \sum_{k=1}^{\infty} g(k)w(t-k)\,,\ t = 1,2,\ldots, \qquad (3.3)$$

where $\hat{y}(t;G)$ is the model output process, $G = \{g(k), k \in [1,\infty)\}$, $g(k), k = 1,2,\ldots$ are the coefficients of the transfer function of the linearized model subject to identification in accordance to the condition of coincidence of the above indicated At that, the statistical linearization criterion is the condition of coincidence of the mathematical expectations of the output processes of system (3.1) and model (3.3), and the condition of coincidence of the joint maximal correlation functions of the output and input processes of the system and the output and input processes of the model. Mathematically, such a criterion has the form

$$\mathbf{M}\{y(t)\} = \mathbf{M}\{\hat{y}(t;G)\} = 0\,, \qquad (3.4)$$

$$S_{yw}(k) = S_{\hat{y}(G)w}(k)\,,\ k = 1,2,\ldots\ . \qquad (3.5)$$

Again, following to normalization conditions (3.2), it is imposed that $\mathbf{D}\{\hat{y}(t;G)\} = 1$ in model (3.3), and, correspondingly, the model weight coefficients meet the following condition

$$\sum_{k=1}^{\infty} g^2(k) = 1\,. \qquad (3.6)$$

Expressions (3.4) and (3.5) are the criterion of statistical linearization of system (3.1). Correspondingly, in terms of the probability densities, condition (3.5) by virtue of (3.2) takes the form

$$\sup_{\substack{B,C:\\ \mathbf{M}B(\bullet)=\mathbf{M}C(\bullet)=0;\\ \mathbf{D}B(\bullet)=\mathbf{D}C(\bullet)=1}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(y)C(w)p_{y,w}(y,w,k)dydw =$$

$$= \sup_{\substack{B,C:\\ \mathbf{M}B(\bullet)=\mathbf{M}C(\bullet)=0;\\ \mathbf{D}B(\bullet)=\mathbf{D}C(\bullet)=1}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(\hat{y}(G))C(w)p_{\hat{y}(G),w}(\hat{y}(G),w,k)\ d\hat{y}(G)dw\,,\ k = 1,2,\ldots,$$

where $p_{y,w}(y,w,k)$, $p_{\hat{y}(G),w}(\hat{y}(G),w,k)$ are correspondingly the joint distribution densities of the system input and output processes and the model input and output processes, $p_y(y)$, $p_{\hat{y}(G)}(\hat{y}(G))$, and $p_w(w)$ are correspondingly the marginal distribution densities of the system $y(t)$ and model $\hat{y}(t;G)$

output processes, and of the system as well as the model input process $w(s)$, $k = t - s$.

Let

$$v_t\langle -k\rangle = \sum_{j=1}^{k-1} g(j)w(t-j) + \sum_{j=k+1}^{\infty} g(j)w(t-j) ,$$
$$k = 1,2,\ldots$$

be a sequence of random variables which are, obviously, Gaussian zero mean ones, and having the variance

$$\mathbf{D}\{v_t\langle -k\rangle\} = \sum_{j=1}^{k-1} g^2(j) + \sum_{j=k+1}^{\infty} g^2(j) = 1 - g^2(k) ,$$
$$k = 1,2,\ldots .$$

Then, within the notations introduced and by virtue of model (3.3) description, the following matrix equalities may be written

$$\begin{pmatrix} \hat{y}(t;G) \\ w(t-k) \end{pmatrix} = \begin{pmatrix} 1 & g(k) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_t\langle -k\rangle \\ w(t-k) \end{pmatrix}, \quad (3.7)$$

$$\begin{pmatrix} v_t\langle -k\rangle \\ w(t-k) \end{pmatrix} = \begin{pmatrix} 1 & -g(k) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{y}(t;G) \\ w(t-k) \end{pmatrix}. \quad (3.8)$$

Formulae (3.7), (3.8) thus represent linear transformation of a Gaussian random vector, and hence

$$p_{\hat{y}(G),w}(\hat{y}(G), w, k) =$$
$$= \frac{1}{2\pi\sqrt{1-g^2(k)}} e^{-\frac{\begin{bmatrix} \hat{y}(G) \\ w \end{bmatrix}^T \begin{bmatrix} 1 & -g(k) \\ -g(k) & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{y}(G) \\ w \end{bmatrix}}{2}} ,$$

that is the density is Gaussian. Hence, it directly follows, by virtue of the Rényi's axiom G (Section 1), that in condition (3.5)

$$S_{\hat{y}(G)w}(k) = |g(k)|, \; k = 1,2,\ldots . \quad (3.9)$$

"To bare" the modulus in (3.9), one should apply the sign of regression of the output process onto the input one, i.e.

$$\text{sign}[reg_{yw}(k)] = \begin{cases} 1, & reg_{yw}(k) \ge 0 \\ -1, & reg_{yw}(k) < 0 \end{cases},$$

where

$$reg_{yw}(k) = \mathbf{M}\left\{ y(t) \big/ w(t-k) \right\} .$$

Thus, finally,

$$g(k) = \text{sign}[reg_{yw}(k)] S_{yw}(k) , \; k = 1,2,\ldots . \quad (3.10)$$

The latter equation determines the coefficients of the weight function of linearized model (3.3).

To calculate the maximal correlation $S_{yw}(k)$, the following procedure directly followed from (Sarmanov, 1963) should be applied.

First of all, some definitions are required. The densities $p_{y,w}(y,w,k)$, $p_y(y)$, and $p_w(w)$ generate two symmetric distribution densities

$$F_1(y,w,k) = \int_{\infty}^{\infty} \frac{p_{y,w}(y,z,k) p_{y,w}(w,z,k)}{p_w(z)} dz ;$$

$$F_2(y,w,k) = \int_{\infty}^{\infty} \frac{p_{y,w}(y,z,k) p_{y,w}(w,z,k)}{p_y(z)} dz , \quad (3.11)$$

and two symmetric kernels

$$\mathrm{K}_1(y,w,k) = \frac{F_1(y,w,k)}{\sqrt{p_y(y) p_w(w)}} ;$$

$$\mathrm{K}_2(y,w,k) = \frac{F_2(y,w,k)}{\sqrt{p_y(y) p_w(w)}} . \quad (3.12)$$

In accordance to (Sarmanov, 1963), kernels (3.12) are positive and have equal spectra of eigenfunvalues

$$1 < \lambda_1^2(k) \le \lambda_2^2(k) \le \ldots \lambda_l^2(k) \le \ldots$$

and, generically, different spectra of eigenfunctions

$$\{\varphi_i(y)\}, \{\psi_i(w)\}, i = 1,2,\ldots$$

Again,

$$R^{\bullet}(k) = \frac{1}{\lambda_1(k)} \quad (3.13)$$

is referred as the maximal (in absolute value) correlation corresponding to the density $p_{y,w}(y,w,k)$, i.e. $R^{\bullet}(k) = S_{yw}(k)$ (in the present notations), and $R^{\bullet\bullet}(k) = \frac{1}{\lambda_1^2(k)}$ is the maximal correlation for symmetric densities (3.11).

Then, calculation of the maximal correlation $S_{yw}(k)$ is based on the following process of sequential approximations.

As an initial approximation $h_0(w)$ one may take any function having a variance (without limitation of generality, $h_0(w)$ is set to be zero mean). Let

$$h_{2l+1}(y) = \int_{\infty}^{\infty} h_{2l}(w) \frac{p_{y,w}(y,w,k)}{p_y(y)} dw , \quad (3.14a)$$

$$h_{2l+2}(w) = \int_{\infty}^{\infty} h_{2l+1}(y) \frac{p_{y,w}(y,w,k)}{p_w(w)} dy , \quad (3.14b)$$

$$l = 0,1,2,\ldots .$$

Then, with an accuracy up to normalizing factors $e_1(k)$, $g_1(k)$, the first pair of the eigenfunctions is determined by the relationships

$$e_1(k)\psi_1(w) = \lim_{l\to\infty} h_{2l}(w)\lambda_1^{2l},$$
$$g_1(k)\varphi_1(y) = \lim_{l\to\infty} h_{2l+1}(y)\lambda_1^{2l}.$$

If l is large enough, then

$$R^{\bullet\bullet}(k) = \frac{1}{\lambda_1^2(k)} \approx \frac{h_{2l}(w)}{h_{2l-2}(w)} \approx \frac{h_{2l+1}(y)}{h_{2l-1}(y)}.$$

The latter relationship implies the values of $S_{yw}(k)$, $k=1,2,\ldots$ by virtue of (3.13). In turn, the densities $p_{y,w}(y,w,k)$ and $p_y(y)$ in (3.14) are to be preliminary restored by applying a procedure of non-parametric density estimation via sampled data; issues of deriving such estimates, including the strongly consistent ones, are at present profoundly developed, involving such practically important cases as estimating distribution densities by use of dependent observations (e.g. (Györfi and Masry, 1990) and related papers).

4. NOISE CANCELLATION ISSUES

Let now the output process of the initial non-linear system of form (3.1) be disturbed by an unobservable zero mean strongly stationary noise $\xi(t)$ with *known* probability distribution density $p_\xi(\xi)$, i.e.

$$\tilde{y}(t) = F\big(w(s),\, s\in I_t\big) + \xi(t). \quad (4.1)$$

Also, the processes $\xi(t)$ and $w(s)$ are stochastically independent.

At that, the statistical linearization problem statement is to be reformulated in order to achieve statistical coincidence of the linearized system's model in (3.3) and system (4.1) but taken as "noise free", i.e. as $\xi(t)\equiv 0$ almost surely. In that case, the criterion expressed by conditions (3.4), (3.5) remains to be applicable but with taking into account that $y(t)$ therein is to be a "noise free" process, i.e. within notations of system (4.1):

$$y(t) \equiv \tilde{y}(t) \text{ as } \xi(t)\equiv 0 \text{ almost surely}. \quad (4.2)$$

To apply the inferences of Section 3, one should to receive an estimation of the joint distribution density $p_{y,w}(y,w,k)$ of the observable input process $w(s)$ and the unobservable "noise free" output process $y(t)$ from (4.2). Let $p_{\tilde{y},w}(\tilde{y},w,k)$ be the joint distribution density of the observable input process $w(s)$ and the observable disturbed output process $\tilde{y}(t)$ of system (4.1). Then, due to stochastic independence of the processes $\xi(t)$ and $F\big(w(s),\, s\in I_t\big)$, one can finally write

$$p_{y,w}(y,w,k) = $$
$$= \int_{-\infty}^{\infty} p_{\tilde{y},w}(\tilde{y}+\xi, w, k)\, p_\xi(\xi)\, d\xi. \quad (4.3)$$

Thus applying a non-parametric estimation of the joint distribution density $p_{\tilde{y},w}(\tilde{y},w,k)$ via sample data leads to the possibility of using formulae (3.14) and followed by them to find the weight function coefficients in (3.10) for linearized model (3.3) of system (4.1). At that, one just may be noted that with regard to system (4.1), in (3.10) $reg_{yw}(k) = reg_{\tilde{y}w}(k)$.

REFERENCES

Chesson, P.L. (1976) "The canonical decomposition of bivariate distributions", *J. Multivariate Analysis*, vol. 6, pp. 526-537.

Györfi, L. and E. Masry (1990), "The L_1 and L_2 strong consistency of recursive kernel density estimation from dependent samples", IEEE Transactions on Information Theory, vol. IT-36, no. 3, pp. 531-539.

Pashchenko, F.F. (2001) "Determining and modeling regularities via experimental data", In: *System Laws and Regularities in Electrodynamics, Nature, and Society*. Chapter 7, "Nauka", Moscow, 2001, pp. 411-521. (in Russian)

Pashchenko, F.F. (2006) "Introduction to consistent methods of systems modeling". Part 1. Mathematical foundations of systems modeling. "Finansy i statistika", Moscow, 2006, pp. 328 p. (in Russian)

Rajbman, N.S. (1981) "Extensions to nonlinear and minimax approaches", *Trends and Progress in System Identification*, ed. P. Eykhoff, Pergamon Press, Oxford, pp. 185-237.

Rényi, A. (1959) "On measures of dependence", *Acta Math. Hung.*, vol. 10, no. 3-4, pp. 441-451.

Sarmanov, O.V. (1963) "The maximal correlation coefficient (nonsymmetric case)", *Sel. Trans. Math. Statist. Probability*, vol. 4, pp. 207-210.

Sarmanov, O.V and E.K. Zakharov (1960) "Measures of dependence between random variables and spectra of stochastic kernels and matrices", *Matematicheskiy Sbornik*, vol. 52(94), pp. 953-990. (in Russian)

Sarmanov, O.V. and Z.N. Bratoeva, (1967) "Probabilistic properties of bilinear expansions of Hermite polynomials", *Theor. Probability Appl.*, vol. 12, pp. 470-481.