CONSISTENT MEASURES OF DEPENDENCE AS A TOOL OF ELICITING NON-LINEAR FEATURES IN COMPLEX SYSTEMS (MILDLY FORMALIZED SYSTEM IDENTIFICATION)

Kirill Chernyshov
Laboratory of Control Systems Identification
V.A. Trapeznikov Institute of Control Sciences
Russia
myau@ipu.rssi.ru

Abstract

The paper presents an approach to the input/output system identification under the condition that no analytical model representation of the system is assumed to be known. Within the approach, the key issue of the problem is a proper handling of inherent dependence between the input and output variables of the system. Using a consistent measure of stochastic dependence of random processes has been proposed within the identification scheme. The measure of dependence is the maximal correlation function. It properly reflects actual nonlinear dependence between random processes, while those of based on the dispersion and, moreover, ordinary product correlation functions do not. In addition, the measure directly leads to determining the input/output relationship of the investigated system. Within the approach, a degree of the system nonlinearity based on the maximal correlation is proposed.

Key words

Complex systems, dada analysis, maximal correlation, measures of dependence, revealing non-linear features, system identification.

1 Introduction

Conventionally, nonlinear system identification significantly uses a priori assumptions with respect to the system model. The assumptions usually consider the system model to belong to a specific model class. However, for a number of cases, body of such knowledge may be considerably restricted. The cases imply elaborating techniques which might decrease dependence of the identification results on the a priori assumptions. Also, such an identification procedure is to lead to a flexible model, i.e. the model is to describe rather rich class of nonlinear systems. The reasons motivate deriving generalized enough and, correspondingly, unified approaches to nonlinear system identification.

The paper presents an approach to nonlinear stochastic system identification based on using a consistent measure of dependence of random processes. The approach is shown to lead to deriving a general input/output system description when no analytical model of the system is available.

2 An Analysis of measures of dependence used within applied problems

The basic concept of the approach proposed within the paper is based on the assumption that identification problem is to be solved for systems which have no completely formalized analytical model. At the same time, from a system theory point of view, the system behavior may be considered in terms of input/output description involving available for observation input and output variables reflecting significant features of the systems. Even if no exact analytical model of the input/output relationship between the variables is stated, obviously, there should always exist an inherent link which reflects dependence of the output variables on the input ones.

When modeling under influence of various uncertainty factors, it is conventional to use stochastic framework assuming the input and output variables to be random processes. In addition, one should also assume the investigated link between input and output processes to be nonlinear. For stochastic case, a natural way to establish an approximate empirical input/output relationship is using measures of stochastic dependence of random processes. 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. In fact [Rajbman, 1981], let a continuous time system subject to identification be as in figure 1, with $f(x(t)) = x^2(t)$, $g(t) = \delta(t)$. Here $\delta(t)$ stands for the delta-function.

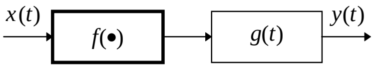

Figure 1. An example leading to vanishing cross correlation.

Let, again, the input $x(t)$ be a stationary Gaussian process with zero mean and unit variance. For the case, ordinary statistical linearization approach to determine the weight function $g(t)$ leads to the correlation equation

$$K_{yx}(t,s) = \int_{t_0}^{t} g(t,v) K_{xx}(v,s)\, dv ,$$

with $K_{yx}(t,s)$ being the cross-correlation function of the processes $y(t)$ and $x(s)$, and $K_{xx}(v,s)$ being the auto-correlation function of the process $x(s)$. However, within the example statement, for the centered random Gaussian processes, $K_{yx}(t,s) = 0$. Hence, the required solution of the correlation equation is zero-valued function $g^*(t,s)$, i.e. $g^*(t,s) \equiv 0$, with mean-squared error of the output approximation by the expression $y(t) = x^2(t)$ corresponding to $g^*(t,s) = 0$ being $\mathbf{M} y^2(t) = 3$. Throughout the paper symbols $\mathbf{M}\bullet$, $\mathbf{D}(\bullet)$, $\mathbf{M}\left\{ \bullet / \bullet \right\}$, and $\mathbf{cov}(\bullet,\bullet)$ will respectively stand for mathematical expectation, variance, conditional expectation, and covariance.

A similar result of linear approximation of the system above may be obtained for odd transformation $f(\bullet)$: $y(t) = 5x^3(t) - 3x(t)$, with marginal distribution of $x(t)$ being uniform at the interval $(-1,+1)$ [Rényi, 1959a].

In [Rényi, 1959a], 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 has been formulated. These are:

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.

Commonly used measures of dependence are the ordinary correlation coefficient, the correlation ratio $\theta(X,Y))$, $\theta(X,Y) = \mathbf{D}\left(\mathbf{M}\left\{ Y / X \right\}\right) / \mathbf{D}(Y)$ with nonzero $\mathbf{D}(Y)$, and the maximal correlation coefficient $S(X,Y)$,

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

$$\mathbf{D}(B(Y)) > 0, \ \mathbf{D}(C(X)) > 0 ,$$

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

Rényi [1959a] has shown the only $S(X,Y)$ to satisfy all the above axioms, while $r(X,Y)$ and $\theta(X,Y)$ do not, in particular, the correlation coefficient does not meet axioms D, E, F, and the correlation ratio does not meet axioms D and F.

When investigating random processes, the above coefficients are transformed to the corresponding functions. As shown above, the ordinary product correlation functions are not the exhaustive tool to be used within the identification problem. The dispersion function $\theta_{yx}(v)$ [Rajbman, 1981] may be considered as a modification and extension of the correlation ratio

$$\theta_{yx}(v) = \mathbf{M}\left(\mathbf{M}\left\{ y(t) / x(s) \right\} - \mathbf{M} y(t) \right)^2 , \ v = t - s .$$

Thus, the function inherits the mentioned disadvantages of the correlation ratio.

In turn, the maximal correlation coefficient is transformed to the following function

$$S_{yx}(v) = \qquad (1)$$

$$= \sup_{\{B\},\{C\}} \frac{\mathbf{cov}(B(y(t)), C(x(s))) - \mathbf{M}(B(y(t)))\mathbf{M}(C(x(s)))}{\sqrt{\mathbf{D}(B(y(t)))\mathbf{D}(C(x(s)))}} .$$

The functional $S_{yx}(v)$ is referred as the maximal correlation function of the random processes $y(t)$ and $x(s)$. The notion of the maximal correlation as a measure of dependence was originally introduced and investigated for random variables in [Rényi, 1959a, 1959b, Sarmanov, 1963a] and then extended to random processes in [Sarmanov, 1963b].

Existence of the pair of transformations (B,C) in (1) is determined by conditions which are equivalent to those of used for random variables stated in [Rényi, 1959a, Sarmanov, 1963b, Sarmanov and Zakharov, 1960, Breiman and Friedman, 1985], with a basic assumption being the stochastic kernel of the random processes $\frac{p(y,x,v)}{\sqrt{p(x)p(y)}}$ meeting the following condition

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{p^2(y,x,v)}{p(x)p(y)}dydx<\infty \qquad (2)$$

for any v. Here $p(x)$, $p(y)$, $p(y,x,v)$, $v=t-s$ are the joint and marginal distribution densities of the input $x(s)$ and output $y(t)$ random processes correspondingly.

Due to (2), the density $p(y,x,v)$ may be represented by the following bilinear eigenfunction expansion converging in mean [Sarmanov and Zakharov, 1960, Chesson, 1976]

$$p(y,x,v)=p(x)p(y)\Big(1+S_{yx}(v)B(y)C(x)+\sum_{i=2}^{\infty}S_i(v)B_i(y)C_i(x)\Big),$$
$$S_{yx}(v)\ge S_2(v)\ge\ldots\ge 0,\ B_i\in\{B\},C_i\in\{C\},$$
$$i=2,\ldots.$$

Here $S_{yx}(v)\ge S_2(v)\ge\ldots\ge 0$ are the eigenvalues of stochastic kernel in (2), with $S_{yx}(v)$ being the largest eigenvalue, and the optimal transformations B and C from (1) being the eigenfunctions corresponding to $S_{yx}(v)$.

In addition to the above examples, there exist 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]. For the example, the dependence can be properly handled only by the maximal correlation. In fact [Sarmanov and Bratoeva, 1967], let the joint distribution density of the input variable x and output variable y of a static model be given by the following relationship

$$p(x,y)=\frac{1}{3\pi\sqrt{3}}\left\{e^{-\frac{3}{2}\left(x^2+y^2+xy\right)}+2e^{-\frac{2}{3}\left(x^2+y^2-xy\right)}\right\}=$$
$$=\frac{e^{-(x^2+y^2)/2}}{2\pi}\left\{1+\sum_{k=1}^{\infty}\frac{2+(-1)^k}{3\cdot 2^k}H_k(x)H_k(y)\right\}=$$
$$\frac{e^{-(x^2+y^2)/2}}{2\pi}\left\{1+\sum_{k=1}^{\infty}c_kH_k(x)H_k(y)\right\}$$

where $H_k(x)=\frac{(-1)^k}{\sqrt{k!}}e^{x^2/2}\frac{d^k}{dx^k}e^{-x^2/2}=$ $=\frac{1}{\sqrt{k!}}\left\{x^k-\frac{k(k-1)}{1!\cdot 2}+\ldots\right\}$, with $H_k(x)$ being the Hermite polynomials.

For the case, the correlation between y and x is linear, with the correlation coefficient $c_1=K_{yx}=1/6$, the regression functions are linear and have the form $\mathbf{M}\left\{y/x\right\}=x/6$, $\mathbf{M}\left\{x/y\right\}=y/6$. At the same time, $H_1(x)$ is not the first eigenfunction, and $c_1=K_{yx}=1/6$ is not the first eigenvalue. For the example, $c_2=S_{yx}=1/4$ is the first eigenvalue, i.e. namely c_2 is the maximal correlation, with $H_2(x)$ being the first eigenfunction.

In accordance to [Sarmanov and Zakharov, 1960], determining the pair of optimal transformations (B,C) meeting (1) is natural to refer as maximal arithmetization of the probability distribution given by the density $p(y,x,v)$.

The maximal correlation function is a complete measure of dependence of random processes and properly reflects actual nonlinear dependence between them, while, as demonstrated above, those of based on the dispersion and, moreover, ordinary product correlation functions do not.

3 System identification based on the maximal correlation technique

3.1 Deriving system model

For the reasons stated in the previous section, the maximal correlation function approach will be used as a basic tool to derive an approximate system model under the assumption that no analytical model of the system is a priori available. Let us assume that for a system it had been made a preliminary expert selection of significant observable characteristics which will be considered as input and output variables in accordance to the input/output description of the system, with the system's variables being assumed to be discrete time stationary and joint stationary random processes. Also, it will not be a restriction to assume the random processes to have zero means.

Then, for each output process $y_i(t)$, $i=1,\ldots,n_y$, an approximate analytical model describing its dependence on input processes $x_j(s)$, $j=1,\ldots,n_x$ is proposed to be searched in the following form

$$B^{(i)}(y_i(t)) = \kappa^{(ij)}(v) C^{(j)}(x_j(t-v)), \qquad (3)$$
$$\mathbf{M}\left(B^{(i)}(y_i(t))\right) = \mathbf{M}\left(C^{(j)}(x_j(t-v))\right) = 0\,,$$
$$\mathbf{D}\left(B^{(i)}(y_i(t))\right) = \mathbf{D}\left(C^{(j)}(x_j(t-v))\right) = 1\,,$$
$$i=1,\ldots,n_y,\ j=1,\ldots,n_x,\ v=1,\ldots,m,$$

with $B^{(i)}(\bullet)$, $C^{(j)}(\bullet)$ being some nonlinear transformations, $\kappa^{(ij)}(v)$ being a scalar coefficient, and *m* being a preliminary specified integer reflecting the model memory. Corresponding structure scheme is presented in fig. 2, with q^{-1} standing for the backward shift operator, i.e. $q^{-1}x[t] = x[t-1]$.

In model (3), the transformations $B^{(i)}(\bullet)$, $C^{(j)}(\bullet)$, and the coefficient $\kappa^{(ij)}(v)$ are to be determined by observing sampled values of $y_i(t)$ and $x_j(s)$ in accordance with minimization of the following identification criterion *I*

$$I\left(B^{(i)}, C^{(j)}, \kappa^{(ij)}(v)\right) = \mathbf{D}\left(E^{(i,j)}(v)\right), \qquad (4)$$
$$E^{(ij)}(v) = B^{(i)}(y_i(t)) - \kappa^{(ij)}(v) C^{(j)}(x_j(t-v)),$$
$$\left(\mathbf{B}^{(i)}, \mathbf{C}^{(j)}, \mathbf{k}^{(ij)}(v)\right) =$$
$$= \arg \inf_{\{B^{(i)}\},\{C^{(j)}\},\{\kappa^{(ij)}\}} I\left(B^{(i)}, C^{(j)}, \kappa^{(ij)}(v)\right),$$
$$\mathbf{M}\left(B^{(i)}(y_i(t))\right) = \mathbf{M}\left(C^{(j)}(x_j(t-v))\right) = 0\,,$$
$$\mathbf{D}\left(B^{(i)}(y_i(t))\right) = \mathbf{D}\left(C^{(j)}(x_j(t-v))\right) = 1\,,$$
$$i=1,\ldots,n_y,\ j=1,\ldots,n_x,\ v=1,\ldots,m.$$

From criterion (4), it directly follows that the optimal set $\left(\mathbf{B}^{(i)}, \mathbf{C}^{(j)}, \mathbf{k}^{(ij)}(v)\right)$ for each $i=1,\ldots,ny$, $j=1,\ldots,nx$, $v=1,\ldots,m$ is to meet the conditions

$$\left(\mathbf{B}^{(i)}, \mathbf{C}^{(j)}\right) = \qquad (5)$$
$$= \arg \sup_{\{B^{(i)}\},\{C^{(j)}\}} \mathbf{M}\left(B^{(i)}(y_i(t)) C^{(j)}(x_j(t-v))\right),$$
$$\mathbf{M}\left(B^{(i)}(y_i(t))\right) = \mathbf{M}\left(C^{(j)}(x_j(t-v))\right) = 0\,,$$
$$\mathbf{D}\left(B^{(i)}(y_i(t))\right) = \mathbf{D}\left(C^{(j)}(x_j(t-v))\right) = 1\,,$$

$$\mathbf{k}^{(ij)}(v) = \sup_{\{B^{(i)}\},\{C^{(j)}\}} \mathbf{M}\left(B^{(i)}(y_i(t)) C^{(j)}(x_j(t-v))\right) =$$
$$= S^{(ij)}_{yx}(v)\,.$$

Here $S^{(ij)}_{yx}(v)$ is the maximal correlation function of the input process $x_j(t-v)$ and the output process $y_i(t)$, corresponding to the optimal pair of transformations $\left(\mathbf{B}^{(i)}, \mathbf{C}^{(j)}\right)$.

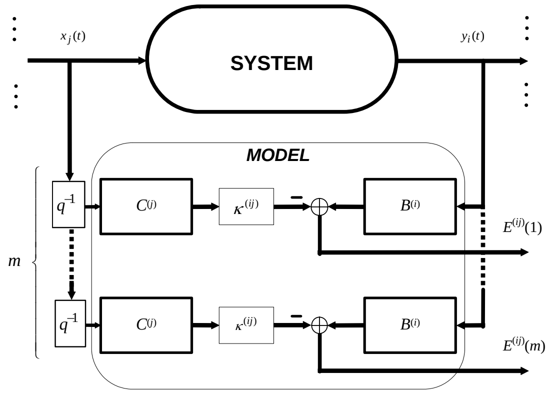

Figure. 2. The identification scheme.

Corresponding methods to estimate the transformations $\left(\mathbf{B}^{(i)}, \mathbf{C}^{(j)}\right)$ in relationship (5) by sampled data and then evaluate $S^{(ij)}_{yx}(v)$ were developed in [Sarmanov, 1963a, Breiman and Friedman, 1985]. These use iterative techniques requiring, in turn, only estimates of the corresponding regression functions. Another way is to obtain at first an estimate of the joint distribution density *p*(*y*,*x*,*v*) and then find the transformations $\left(\mathbf{B}^{(i)}, \mathbf{C}^{(j)}\right)$ from (5) directly.

Thus, procedure (5) combined with a technique of determining $\left(\mathbf{B}^{(i)}, \mathbf{C}^{(j)}\right)$ and $S^{(ij)}_{yx}(v)$ by sampled data leads to obtaining *m* matrices of $ny \times nx$ dimension, whose elements are the triples $\left(\mathbf{B}^{(i)}, \mathbf{C}^{(j)}, \mathbf{k}^{(ij)}(v)\right)$.

3.2 Degree of nonlinearity

In practice, a primary problem associated with nonlinear system identification is investigating the system to be actually nonlinear. Corresponding techniques are known as “tests on nonlinearity” [Rajbman, 1981, Billings and Voon, 1986]. These are based on various numerical characteristics quantitatively reflecting the system nature. In particular, such a general degree of nonlinearity has been proposed in [Rajbman, 1981]. It is based on the dispersion functions and has the form

$$\eta_{disp}(v) = \sqrt{1 - K_{yx}^2(v) \big/ \theta_{yx}^2(v)}\ , \qquad (6)$$

with $K_{yx}(v)$, $\theta_{yx}(v)$ standing for the product correlation and cross-dispersion functions which are as defined above. A generalization of (6) is that of [Durgaryan and Pashchenko, 1985] based on the maximal correlation function

$$\eta_{\max corr}(v) = \sqrt{1 - K_{yx}^2(v) \big/ S_{yx}^2(v)}\ .$$

Within the approach derived in Section 3.1, a natural way to define the degree of nonlinearity of such a system is as follows. Let

$$\overline{K_{yx}} = \max_{\substack{i,\ i=1,\dots,n_y \\ j,\ j=1,\dots,n_x \\ v,\ v=1,\dots,m}} K_{yx}^{(ij)}(v)\ ,\ \overline{S_{yx}} = \max_{\substack{i,\ i=1,\dots,n_y \\ j,\ j=1,\dots,n_x \\ v,\ v=1,\dots,m}} S_{yx}^{(ij)}(v)$$

where $K_{yx}^{(ij)}(v)$ stands for the corresponding ordinary correlation. Then, let us define the degree of nonlinearity of model (3) as

$$\overline{\eta_{\max corr}} = \sqrt{1 - \left(\overline{K_{yx}}\right)^2 \Big/ \left(\overline{S_{yx}}\right)^2}\ ,$$

with $\overline{\eta_{\max corr}}$ vanishing if and only all the transformations $\left(\mathbf{B}^{(i)},\ i = 1,\dots,n_y,\ \mathbf{C}^{(j)},\ i = 1,\dots,n_x\right)$ are linear ones.

Obviously, $\overline{\eta_{\max corr}} \geq \sqrt{1 - \left(\overline{K_{yx}}\right)^2 \Big/ \left(\overline{\theta_{yx}}\right)^2}$

where

$$\overline{\theta_{yx}} = \max_{\substack{i,\ i=1,\dots,n_y \\ j,\ j=1,\dots,n_x \\ v,\ v=1,\dots,m}} \theta_{yx}^{(ij)}(v)$$

in accordance to the above notations.

The approach developed also enables one to introduce another quantitative characteristic, the degree of nonlinearity in mean, defined as

$$\overline{\eta_{mean}} = \sqrt{1 - \left(\overline{\theta_{yx}}\right)^2 \Big/ \left(\overline{S_{yx}}\right)^2}\ .$$

Such a value reflects how the nonlinear system under study differs from that of linear in mean. A system is said to linear in mean [Rajbman, 1985], if it is linear with respect to conditional mathematical expectation of the output process with respect to the input one.

Consider as an illustrative example a single input /single output system having as the joint input/output distribution density that of presented in Section 2. Then, for such a system, $\overline{\eta_{\max corr}} = \sqrt{5}\big/3$.

Also, as an additional example presenting significance of use of the maximal correlation consider the same system. Let, however, the joint distribution density of the input and output variable of such a system to be of the form [Sarmanov, 1960]

$$p_{-1/2,\lambda}(x,y) = \frac{1}{4\pi\sqrt{1-\lambda^2}} \exp\left[-\frac{x^2 + y^2 - 2\lambda xy}{2\left(1-\lambda^2\right)}\right] + \\ + \frac{1}{4\pi\sqrt{1-\lambda^2}} \exp\left[-\frac{x^2 + y^2 + 2\lambda xy}{2\left(1-\lambda^2\right)}\right],$$

with $|\lambda| < 1$.

From the form it follows that both the input and output variables of the system have normal marginal distribution densities but theirs ordinary correlation is zero as well as the correlation ratio. At the same time, theirs maximal correlation is equal to λ^2. Consequently, for such a system, $\overline{\eta_{disp}}$ is indefinable, while $\overline{\eta_{\max corr}} = \overline{\eta_{mean}} = 1$.

Thus, the technique proposed enables one:

- to split the nonlinear system identification scheme onto simpler sequential stages, i.e. determining nonlinear transformations and linear coefficients;
- to achieve completely formalized choice of nonlinear input and output transformations without any heuristics, and a priori assumptions on distributions of the random processes, or the transformations to belong to a parameterized family;
- to use a measure of dependence of random vector valued processes, which properly reflects the actual inherent stochastic linking between the processes;
- to derive a nonlinearity measure of the system under study, with the measure being more accurate in comparison with those of based on ordinary product correlation functions, dispersion functions, or high order cumulants.

4 Conclusions

A non-parametric approach to input/output system identification has been presented under the assumption that no analytical model of the system is a priori available. Within the approach, choosing the system input/output relationship is to be done by a “reliable” manner eliminating cases when some important input/output links might be omitted. To meet the requirement a consistent measure of stochastic dependence of random processes has been proposed as a mathematical tool to obtain the system model. The measure is the maximal correlation function. Consistency of the measure of dependence $S_{yx}(v)$ of random processes $y(t)$ and $x(t–v)$ means it to vanish, $S_{yx}(v) = 0$, if and only if the random variables $Y = y(t)$ and $X = x(t)$ are stochastically independent. In contrast, the ordinary product correlation function $K_{yx}(v) = \mathbf{cov}(y(t), x(s))$, $v = t - s$ of the random processes $y(t)$ and $x(s)$ may vanish even provided that the random variables $Y = y(t)$ and $X = x(t)$ are completely dependent, i.e. if there exists a deterministic function $f(\bullet)$ such that $Y = f(X)$ with probability 1. As to the maximal correlation as a measure of dependence, under the condition of complete dependence of the random variables, it is equal to 1.

The maximal correlation function properly reflects actual nonlinear dependence between random processes, while those of based on the dispersion functions and, moreover, ordinary product correlation functions do not. In addition, there exist examples when actual dependence between two variables is nonlinear even provided that the regression of a variable onto another one is linear. For such a case, the dependence can be properly handled by the maximal correlation ultimately. The reasons justify using the consistent measures of dependence. Within the system identification methodology, the measure is shown to be a suitable tool to handle the system input/output relationship, with the body of data required within the technique not exceeding that of used for estimating the joint distribution density or regression functions of the random processes.

As a final remark, one should be noted that of course the subject discussed within the paper has a variety of interesting and significant narrow aspects, so the present paper is by no means an exhaustive approach to such a kind of problems. Just as a supplement, papers [Chernyshov 2002, 2003a, 2003b, 2005, 2007] may be mentioned.

References

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

Rényi, A. (1959a). On measures of dependence. *Acta Math. Hung.,* **10**(3-4), pp. 441-451.

Rényi, A. (1959b). New version of the probabilistic generalization of the large sieve. *Acta Math. Hung.*, **10**(1-2), pp. 217-226.

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

Sarmanov, O.V. (1963b). Investigation of stationary Markov processes by the method of eigenfunction expansion. *Sel. Trans. Math. Statist. Probability*, **4**, pp. 245-269.

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

Breiman, L. and Friedman, J.H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.*, **80**(391), pp. 590-598.

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

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

Billings, S.A. and Voon, W.S.F. (1986). Correlation based model validation test for nonlinear models. *Int. J. Control.*, 1986, **44**, pp. 235-244.

Durgaryan, I.S. and Pashchenko, F.F. (1985). Nonparametric identification of non-linear systems. In *Proc. of the 7th IFAC/IFORS Symposium on Identification and System Parameters Estimation*. New York, vol. 1, pp. 433-437.

Sarmanov, O.V. (1960). Pseudonormal correlation and its various generalizations. *Dokl. AN SSSR*, **132**, 2, pp. 299-302. (in Russian)

Chernyshev, K.R. (2002). A Paper on the Identification of Stochastic Plants. *Automation and Remote Control*, vol. 63, issue 4, p. 687.

Chernyshov, K.R. (2003a). An essay on some delusions in system identification. In *Proc. of the II International conference "System Identification and Control Problems" SICPRO '03. Moscow, 29-31 January 2003.* Moscow: V.A. Trapeznikov Institute of Control Sciences, pp. 2660-2698. (in Russian)

Chernyshov, K.R. (2003b). *Questions of identification: consistent measures of dependence*. Moscow: V.A. Trapeznikov Institute of Control Sciences, 60 p. (in Russian)

Chernyshov, K.R. (2005). Theses on correlation. In *Proc. of the IV International Conference "System Identification and Control Problems" SICPRO '05. Moscow, January 25-28, 2005.* Moscow: V.A. Trapeznikov Institute of Control Sciences, pp. 2071-2077. (in Russian)

Chernyshov, K.R. (2007). Towards the educational process support in the stochastic modeling (the essay no.2). In *Proc. of the VI International Conference "System Identification and Control Problems" SICPRO '07. Moscow, January 29 - February 1, 2007.* Moscow: V.A. Trapeznikov Institute of Control Sciences, pp. 1632-1671. (in Russian)