Statistical Linearization and Consistent Measures of Dependence: A Unified Approach

Kirill Chernyshov
V. A. Trapeznikov Institute of Control Sciences, 65 Profsoyuznaya Street, Moscow, Russia

Keywords: Cauchy-Schwarz Divergence, Input/Output Model, Maximal Correlation, Measures of Dependence, Mutual Information, Rényi Entropy, Statistical Linearization, System Identification.

Abstract: The paper presents a unified approach to the statistical linearization of input/output mapping of non-linear discrete-time stochastic systems driven with white-noise Gaussian process. The approach is concerned with a possibility of applying any consistent measures of dependence (that is those measures of dependence of a pair of random values, which vanish if and only if these random values are stochastically independent) in statistical linearization problems and oriented to the elimination of drawbacks concerned with applying correlation and dispersion (based on the correlation ratio) measures of dependence, based on linearized representations of their input/output models.

1 PRELIMINARIES

Solving an identification problem of stochastic systems is always based on applying measures of dependence of random values, both within representation of a system under study either by use of an input/output mapping, or within state-space techniques. Most frequently, conventional linear correlation is used as such measures. Its application directly follows from the identification problem statement itself, when it is based on applying conventional least squares approaches. The main advantage of the measure is convenience of its application, involving both a possibility of deriving explicit analytical relationships to determine required system characteristics, and constructing estimation procedures via sampled data, including those of based on applying dependent observations.

However, the linear correlation as a measure of dependence is known to be able to vanish even under existence of a deterministic dependence between random values. In particular, this is valid for the quadratic dependence, $Y = X^2$, when X is the Laplacian random value (Rajbman, 1981), and for an odd transformation of the form $Y = 5X^3 - 3X$, where the random value X has the uniform distribution over the interval $(-1,+1)$ (Rényi, 1959).

Just to overcome such a disadvantage, use of more complicated, non-linear measures of dependence has been involved into the system identification. A key issue of the present paper is applying consistent measures of dependence. In accordance to the A.N. Kolmogorov terminology, a measure of dependence $\mathrm{M}(X,Y)$ between two random values X and Y is referred as consistent, if $\mathrm{M}(X,Y)=0$ if and only if the random values X and Y are stochastically independent.

The statistical linearization of input/output mappings relates to those identification problems, whose solution is most considerably determined by characteristics of dependence of input and output processes of the system subject to identification. Meanwhile, known approaches to the statistical linearization are based either applying conventional correlations, or dispersion (based on the correlation ratio) functions, what, due to the reasons pointed out above, may give rise constructing models, whose output is identically equal to zero. A majority of literature references on correlation based statistical linearization may be found in the books of Roberts and Spanos (2003) and Socha (2008).

The approach presented in the present paper is concerned with a possibility of applying any consistent measures of dependence in statistical linearization problems and is directed to elimination of the drawbacks concerned with applying correlation and dispersion measures of dependence under system identification based on linearized representations of input/output models.

Chernyshov K..
Statistical Linearization and Consistent Measures of Dependence: A Unified Approach.
DOI: 10.5220/0005534805240532
In *Proceedings of the 12th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2015), pages 524-532
ISBN: 978-989-758-122-9

2 PROBLEM STATEMENT

Let in a non-linear dynamic stochastic system $z(t)$ be the output random process assumed to be stationary in the strict sense and ergodic, $u(s)$ be the output random input process assumed within the present problem statement to be the white-noise Gaussian process. Processes $z(t)$ and $u(s)$ are also assumed to be joint stationary in the strict sense, while the dependence of the input and output processes of the system is characterized by the probability distribution densities

$$p_{zu}(z,u,\tau),\ \tau = 1, 2, \ldots. \quad (1)$$

(being of course not known to the user). For sake of simplicity but without loss of the generality, the processes $z(t)$ and $u(s)$ are assumed to be zero-mean and unit-variance, that is

$$\mathbf{E}\{z(t)\} = \mathbf{E}\{u(s)\} = 0,\ \mathbf{var}\{z(t)\} = \mathbf{var}\{u(s)\} = 1 \quad (2)$$

In (2), $\mathbf{E}(\cdot)$ stands for the mathematical expectation, and $\mathbf{var}\{\cdot\}$, for the variance.

A model of the system described by the densities (1) and condition (2) is searched in the form

$$\hat{z}(t;W) = \sum_{\tau=1}^{\infty} w(\tau)u(t-\tau),\ t = 1,2,\ldots, \quad (3)$$

where $\hat{z}(t;W)$ is the model output process, $W = \{w(\tau), \tau \in [1,\infty)\}$, $w(k), k = 1,2,\ldots$ are coefficients of the weight function of the linearized model, subject to identification in accordance to a criterion of the statistical linearization. Such a criterion is the condition of coincidence of the mathematical expectations of the system output process, described by densities (1), and model output process (3), and the condition of coincidence of a given measure of dependence of the input and output processes of the system described by densities (1) and input and output processes of model (3), or mathematically,

$$\mathbf{E}\{z(t)\} = \mathbf{E}\{\hat{z}(t;W)\} = 0, \quad (4)$$

$$\mathrm{M}_{zu}(\tau) = \mathrm{M}_{\hat{z}(W)u}(\tau),\ \tau = 1,2,\ldots, \quad (5)$$

where $\mathrm{M}_{\bullet\bullet}(\tau)$ is some measure of dependence.

Again, in accordance to normalization conditions (2), model (3) is implied to meet the condition

$$\mathbf{var}\{\hat{z}(t;W)\} = 1,$$

Accordingly, the weight coefficients of the model meet the condition

$$\sum_{\tau=1}^{\infty} w^2(\tau) = 1. \quad (6)$$

Thus, expressions (4) and (5) represent a criterion of the statistical linearization of a system described by densities (1).

3 CONSTRUCTING THE UNIVERSAL APPROACH

Let

$$x_t\left(\tau^{out}\right) = \sum_{j=1}^{\tau-1} w(j)u(t-j) + \sum_{j=\tau+1}^{\infty} w(j)u(t-j),$$
$$\tau = 1,2,\ldots$$

be a sequence of random values that are, obviously, Gaussian, zero-mean, and having the variance

$$\mathbf{var}\left\{x_t\left(\tau^{out}\right)\right\} = \sum_{j=1}^{\tau-1} w^2(j) + \sum_{j=\tau+1}^{\infty} w^2(j) = 1 - w^2(k),$$
$$\tau = 1,2,\ldots.$$

Then within the notations introduced and by virtue of model (3) description, one may write the fallowing matrix equalities

$$\begin{pmatrix} \hat{z}(t;W) \\ u(t-\tau) \end{pmatrix} = \begin{pmatrix} 1 & w(\tau) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_t\left(\tau^{out}\right) \\ u(t-\tau) \end{pmatrix}, \quad (7)$$

$$\begin{pmatrix} x_t\left(\tau^{out}\right) \\ u(t-\tau) \end{pmatrix} = \begin{pmatrix} 1 & -w(\tau) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{z}(t;W) \\ u(t-\tau) \end{pmatrix}. \quad (8)$$

As well known, if two n-dimensional random vectors $\mathbf{z}$ and $\mathbf{x}$, having marginal probability distribution densities $p_Y(\mathbf{Y})$ and $p_V(\mathbf{V})$ correspondingly, are connected with a one-to-one mapping $\mathbf{Y} = \varphi(\mathbf{V})$, then

$$p_Y(\mathbf{Y}) = p_V\left(\varphi^{-1}(\mathbf{Y})\right)\left|\frac{D\left(\varphi^{-1}(\mathbf{Y})\right)}{D(\mathbf{Y})}\right|,$$

where $\frac{D\left(\varphi^{-1}(\mathbf{Y})\right)}{D(\mathbf{Y})}$ is the Jacobian of the inverse transformation $\mathbf{V} = \varphi^{-1}(\mathbf{Y})$.

In accordance to this relationship, the joint probability distribution density $p_{\hat{z}(W)u}(\hat{z}(W),u,\tau)$ of the random values $\hat{z}(t;W)$ and $u(t-\tau)$ may be expressed via the joint probability distribution density $p_{x_t\left(\tau^{out}\right)u}\left(x_t\left(\tau^{out}\right),u,\tau\right)$ of the random values $x_t\left(\tau^{out}\right)$ and $u(t-\tau)$. In turn, the density $p_{x_t\left(\tau^{out}\right)u}\left(x_t\left(\tau^{out}\right),u,\tau\right)$ is, evidently, of the form

$$p_{x_t\left(\tau^{out}\right)u}\left(x_t\left(\tau^{out}\right),u,\tau\right)=p_{x_t\left(\tau^{out}\right)}\left(x_t\left(\tau^{out}\right)\right)p_u(u),$$

where $p_{x_t\left(\tau^{out}\right)}\left(x_t\left(\tau^{out}\right)\right)$, $p_u(u)$ are the marginal probability distribution densities of the random values $x_t\left(\tau^{out}\right)$ and $u(t-k)$ correspondingly. Hence, due to relationships (7) and (8), and by virtue of relationship (6), one may write for the density $p_{\hat{z}(W)u}(\hat{z}(W),u,\tau)$:

$$p_{\hat{z}(W)u}(\hat{z}(W),u,\tau)=p_{x_t\left(\tau^{out}\right)}(\hat{z}(W)-w(\tau)u)p_u(u)=$$
$$=\frac{1}{2\pi\sqrt{1-w^2(\tau)}}e^{-\frac{\begin{bmatrix}\hat{z}(W)\\ u\end{bmatrix}^T\begin{bmatrix}1 & -w(\tau)\\ -w(\tau) & 1\end{bmatrix}^{-1}\begin{bmatrix}\hat{z}(W)\\ u\end{bmatrix}}{2}}, \qquad (9)$$

that is this density is Gaussian.

Thus, calculating the measure of dependence $\mathrm{M}_{\hat{z}(W)u}(k)$ for density (9) enables one to express the measure as a transformation of $|w(k)|$:

$$\mathrm{M}_{\hat{z}(W)u}(\tau)=\Theta_{\mathrm{M}_{\hat{z}(W)u}}(|w(\tau)|). \qquad (10)$$

Here the lower script "$\mathrm{M}_{\bullet\bullet}$" is used to underline the dependence of the transformation of $|w(k)|$ on a given specific measure of dependence in criterion (5).

Hence, virtue of criterion (5), formula (10) directly implies the relationship for the weight coefficients of linearized model (3)

$$\Theta^{-1}_{\mathrm{M}_{zu}}(\mathrm{M}_{zu}(\tau))=|w(\tau)|,\ \tau=1,2,\ldots . \qquad (11)$$

To "open" the sign of modulo in (11), one should apply the sign of the regression of the output process of the system onto the input one, that is

$$\mathrm{sign}[reg_{zu}(\tau)]=\begin{cases}1, & reg_{zu}(\tau)\geq 0\\ -1, & reg_{zu}(\tau)<0\end{cases},$$

where

$$reg_{zu}(\tau)=\mathbf{E}\left\{z(t)\big/u(t-\tau)\right\},$$

and $\mathbf{E}\left(\cdot/\cdot\right)$ stands for the conditional mathematical expectation. Thus, finally,

$$w(\tau)=\mathrm{sign}[reg_{zu}(\tau)]\Theta^{-1}_{\mathrm{M}_{zu}}(\mathrm{M}_{zu}(\tau)) \qquad (12)$$
$$\tau=1,2,\ldots .$$

Accordingly, relationship (12) just determines the weight function coefficients of linearized model (3).

4 TOWARDS CONSISTENT MEASURES OF DEPENDENCE

As pointed out in Section 1, consistent measures of dependence play a special role in the system identification, first of all, with regard to non-linear systems. A. Rényi (1989) has formulated axioms that were recognized to be the most natural to define a measure of dependence $\mu(X,Y)$ between two random values *X* and *Y*, which is to characterize exhaustively such a dependence. These axioms are as follows:

A) $\mu(X,Y)$ is defined for any pair of random values *X* and *Y*, if none of them is a constant with probability 1.

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

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

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

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

F) If φ and ψ are some one-to-one Borel-measurable functions, then $\mu(\varphi(X),\psi(Y))=$ $=\mu(X,Y)$.

G) If the joint probability distribution of *X* and *Y* is Gaussian, then $\mu(X,Y)=|r(X,Y)|$, where $r(X,Y)$ is the conventional correlation coefficient between *X* and *Y*.

Measures of dependence meeting the Rényi axioms, with the exception, may be Axiom F, will be hereinafter refereed as consistent in the Rényi sense.

The conventional correlation coefficient $r(X,Y)$ is, of course, most widely known among different measures of dependence. More delicate approach to characterize the dependence of random values is concerned with applying the correlation ratio

$$\theta(X,Y)=\frac{\mathbf{var}\left(\mathbf{E}\left(Y/X\right)\right)}{\mathbf{var}(Y)},\quad \mathbf{var}(Y)>0\,,$$

and the maximal correlation coefficient $S(X,Y)$, originally introduced by H. Gebelein (1941), and investigated in papers of O.V. Sarmanov (1963a,b), Sarmanov and Zakharov (1960), A. Rényi, and others

$$S(X,Y)=\sup_{\{B\},\{C\}}\frac{\mathbf{cov}(B(Y),C(X))}{\sqrt{\mathbf{var}(B(Y))\mathbf{var}(C(X))}}\,,$$

$$\mathbf{var}(B(Y))>0,\quad \mathbf{var}(C(X))>0\,.$$

In the formula, the supremum is taken over the sets of Borel-measurable functions $\{B\}$ and $\{C\}$, and $B\in\{B\}$, $C\in\{C\}$, while $\mathbf{cov}(\cdot,\cdot)$ stands for the covariance.

Meanwhile, it was shown in the paper of Rényi (1959) that the maximal correlation coefficient $S(X,Y)$ meets the above axioms only, while the conventional correlation coefficient $r(X,Y)$ and the correlation ratio $\theta(X,Y)$ do not. In particular, Axioms D, E, F are not met for the correlation coefficient, and Axioms D, F are not met for the correlation ratio.

Here one should underline, that consistent in the Kolmogorov sense measures of dependence are not mandatory consistent in the Rényi sense. In the first turn, this is concerned with meeting Rényi Axioms C and G. And, at the same time, the approach to the statistical linearization, presented in the preceding Section, just directs the method of constructing measures of dependence meeting Rényi Axioms C and G. Namely, the approach is as follows:

1) For any measure of dependence M_{XY} *between random values X and Y one should calculated this measure for the two-dimensional Gaussian density depending on the correlation coefficient* $r(X,Y)$.

2) Represent the expression obtained as a function in modulo of the correlation coefficient

$$\Theta_{\mathrm{M}_{XY}}(|r(X,Y)|),\tag{13}$$

and invert this function.

3) The expression obtained

$$\Theta^{-1}_{\mathrm{M}_{XY}}(\mathrm{M}_{XY})\tag{14}$$

(as a function of the initial measure of dependence M_{XY}*) defines the measure of dependence between two random values X and Y, meeting Rényi Axioms C and G.*

In particular, for the maximal correlation coefficient $S(X,Y)$ the corresponding function $\Theta_{\mathrm{M}_{XY}}(|r(X,Y)|)=\Theta_{S(X,Y)}(|r(X,Y)|)$ is the identical transformation. Meanwhile, one should be noted that calculation of the maximal correlation coefficient is concerned with the necessity of applying a complex iterative procedure of determining the first eigenvalue and the pair of the first eigenfunctions (corresponding to this eigenvalue) of the stochastic kernel $p_{xy}(x,y)\big/\sqrt{p_y(y)p_x(x)}$.

Along with the maximal correlation coefficient based, in entity, on the comparison of moment characteristics of the joint and marginal probability distributions of the pair of considered random values, a broad class of measures of dependence is constructed by use of the direct comparison of the joint and marginal probability distributions of random values. Such a class is known is the measure of divergence of probability distributions. Most known among them involves (Sarmanov and Zakharov, 1960):

- Contingency coefficient

$$\Delta^2(X,Y)=\mathbf{E}\left\{\frac{(p_{xy}(x,y)-p_x(x)p_y(y))^2}{p_{xy}(x,y)p_x(x)p_y(y)}\right\},\tag{15}$$

- Shannon mutual information

$$I(X,Y)=\mathbf{E}\left\{\ln\frac{p_{xy}(x,y)}{p_x(x)p_y(y)}\right\},\tag{16}$$

Measures of dependence (15), (16) meet all Rényi Axioms with the exception of Axioms C and G.

Correspondingly, the methodology of formulae (13), (14) implies the following transforms:

- for the contingency coefficient (15),

$$\delta^2(X,Y)=\Theta^{-1}_{\Delta(X,Y)}(\Delta(X,Y))=\sqrt{\frac{\Delta^2(X,Y)}{\Delta^2(X,Y)+1}}\,,\tag{17}$$

- for the Shannon mutual information (16)

$$\iota(X,Y)=\Theta^{-1}_{I(X,Y)}(I(X,Y))=\sqrt{1-e^{-2I(X,Y)}}\tag{18}$$

Formulae (17), (18) are known in the literature and in the present paper are presented as illustrative examples confirming the applicability of formulae (13), (14). Measures of dependence (17), (18) meet all the Rényi Axioms and determine solution (12) of the problem of the statistical linearization for the linearization criteria based on a corresponding measure of dependence ((15), (16)).

5 MEASURES OF DEPENDENCE BASED ON THE RÉNYI ENTROPY

Besides the Shannon definition of the entropy, which, in turn, leads to the definition of the Shannon mutual information (16), other ways to define the entropy are known. For a random value *X* having the probability distribution density $p_x(x)$, the Rényi entropy of the order *α* (Rényi, 1961, 1976a,b) is defined as

$$R_\alpha(X) = \frac{1}{1-\alpha}\ln\left(\mathbf{E}\left(p_x(x)\right)^{\alpha-1}\right),\ \alpha > 0, \alpha \neq 1 .$$

Meanwhile, as *α* tends to 1, $R_\alpha(X)$ tends to the expression determining the Shannon entropy that, thus, may be considered as the Rényi entropy of the order 1.

From a computational point of view, especially under the necessity of estimating by use of sample data, the Rényi entropy was recognized as more preferable than that of Shannon, since the Rényi entropy is a "logarithm of integral", what is computationally simpler than an "integral of logarithm" as in the case of the Shannon entropy. Meanwhile, the selection of a specific value of the order *α*, is of a special importance, since the larger the order is, the more complicated the computational procedure becomes.

Also one may be noted that the Rényi entropy for continuous random values takes its magnitude at the whole interval $(-\infty;\infty)$ as well as the Shannon entropy; and for some probability distributions the entropies of Rényi and Shannon may coincide. In that case, of course, the Rényi entropy does not depend on the order *α*. Indeed, this is valid, for instance, for the uniform distribution at the interval $[a;b]$, when both the Shannon and Rényi entropy have the form $\ln(b-a)$. Analytical expressions for a broad class of univariate and multivariate probability distributions are presented in the papers of (Nadarajah and Zografos, 2003, Zografos and Nadarajah, 2005a,b).

As specific values of the order $\alpha > 0, \alpha \neq 1$ any one may be selected, but the problem complexity, meanwhile, grows exponentially with the growth of *α*; at the same time the value of $\alpha = 2$ was recognized in the literature as providing good results (Principe et al., 2000). For $\alpha = 2$ the expression

$$R_2(X) = -\ln\left(\mathbf{E}\left(p_x(x)\right)\right)$$

is known as the quadratic entropy.

So far, the consideration was concerned with the Rényi entropy of one (possibly, multivariate) probability distribution density. Along with such (marginal) entropy, one may in a corresponding manner define the mutual Rényi entropy of the order (α_1, α_2) for a pair of random values *X* and *Y* with a joint and marginal probability distribution densities. Within such an approach, the first probability measure is defined by the joint probability distribution density $p_{xy}(x,y)$ of the random values *X* and *Y*, the second probability measure is defined by the multiplication of the marginal probability distribution densities, correspondingly $p_x(x)$ and $p_y(y)$ of the random values *X* and *Y*. Then the mutual Rényi entropy $R_{\alpha_1,\alpha_2}(X,Y)$ of the order (α_1, α_2) may be defined in this case as follows:

$$R_{\alpha_1,\alpha_2}(X,Y) =$$
$$= \frac{1}{1-\alpha}\ln\left(\mathbf{E}_{p_{xy}}\left(\left(p_{xy}(x,y)\right)^{\alpha_1-1}\left(p_x(x)p_y(y)\right)^{\alpha_2}\right)\right),$$
$$\alpha_1^2 + \alpha_2^2 > 0, \alpha_1 + \alpha_2 = \alpha \neq 1 ,$$

where the mathematical expectation is taken over $p_{xy}(x,y)$. The marginal Rényi entropy is, thus, a partial case of the mutual one, when either $\alpha_1 = 0$, or $\alpha_2 = 0$. In the first case ($\alpha_1 = 0, \alpha_2 = \alpha$) the mutual Rényi entropy takes the form

$$R_{0,\alpha}(X,Y) = \frac{1}{1-\alpha}\ln\left(\mathbf{E}_{p_{xy}}\left(\frac{\left(p_x(x)p_y(y)\right)^{\alpha}}{p_{xy}(x,y)}\right)\right),$$

in the second case ($\alpha_1 = \alpha, \alpha_2 = 0$),

$$R_{\alpha,0}(X,Y) = \frac{1}{1-\alpha}\ln\left(\mathbf{E}_{p_{xy}}\left(p_{xy}(x,y)\right)^{\alpha-1}\right).$$

In the both cases, the mathematical expectation is taken (formally) over $p_{xy}(x,y)$. At the same time, $R_{0,\alpha}(X,Y)$ does not depend on $p_{xy}(x,y)$, while $R_{\alpha,0}(X,Y)$ does not explicitly depend on $p_x(y)$ and $p_y(y)$. So, $R_{\alpha,0}(X,Y)$ and $R_{0,\alpha}(X,Y)$ should be considered as marginal entropies of the probability distribution densities $p_x(x)p_y(y)$ and $p_{xy}(x,y)$ correspondingly. These marginal entropies will be designated as $R_\alpha(p_x p_y)$ and $R_\alpha(p_{xy})$, that is

$$R_\alpha(p_x p_y) = = \frac{1}{1-\alpha}\ln\left(\mathbf{E}_{p_x p_y}\left(p_x(x)p_y(y)\right)^{\alpha-1}\right) = R_{0,\alpha}(X,Y),$$

where the mathematical expectation is taken over $p_x(x)p_y(y)$; and

$$R_\alpha(p_{xy}) = \frac{1}{1-\alpha}\ln\left(\mathbf{E}_{p_{xy}}\left(p_{xy}(x,y)\right)^{\alpha-1}\right) = = R_{\alpha,0}(X,Y),$$

where the mathematical expectation is taken over $p_{xy}(x,y)$.

Of course, one should be noted that $R_\alpha(p_{xy}) = R_\alpha(p_x) + R_\alpha(p_y)$, when the random values *X* and *Y* are stochastically independent.

Also, within the consideration of non-zero cases, when $\alpha_1 \neq 0, \alpha_2 \neq 0$, the "symmetric" case is emphasized, when $\alpha_1 = \alpha_2 = \alpha/2$. It will serve as a basis for constructing the mutual Rényi information of the order *α* for random values.

Again, since the Shannon mutual information *I*(*X*,*Y*) for the pair of random values *X* and *Y* has the representation via corresponding entropies of these random values

$$I(X,Y) = H(X) + H(Y) - H(X,Y),$$

it would be natural to search for the mutual Rényi information $I_{R_\alpha}(X,Y)$ of the order *α* in a similar form, that is

$$I_{R_\alpha}(X,Y) = = c_1 R_{\frac{\alpha}{2},\frac{\alpha}{2}}(X,Y) + c_2 R_\alpha(p_{xy}) + c_3 R_\alpha(p_x p_y), \quad (19)$$

where c_1, c_2, c_3 are normalizing coefficients selected in the manner to provide meeting the condition:

$I_{R_\alpha}(X,Y) \geq 0$, *meanwhile* $I_{R_\alpha}(X,Y) = 0$ *if and only if the random values X and Y are stochastically independent.*	(20)

Condition (20) implies infinitely many solutions that in a unified form may be written as

$$\begin{cases} c_1 = \nu, c_2 = c_3 = -\frac{\nu}{2}, IF\ \alpha > 1 \\ c_1 = -\nu, c_2 = c_3 = \frac{\nu}{2}, IF\ \alpha < 1 \end{cases}, \quad (21)$$

where $\nu > 0$. Hence, one may just set $\nu = 1$, and taking into account all above considerations, the mutual Rényi information of the order *α* is written in the form

$$I_{R_\alpha}(X,Y) = = \frac{1}{\gamma(\alpha)}\ln\frac{\mathbf{E}_{p_{xy}}\left(\left(p_{xy}(x,y)\right)^{\alpha/2-1}\left(p_x(x)p_y(y)\right)^{\alpha/2}\right)}{\sqrt{\mathbf{E}_{p_{xy}}\left(p_{xy}(x,y)\right)^{\alpha-1}\mathbf{E}_{p_x p_y}\left(p_x(x)p_y(y)\right)^{\alpha-1}}}$$

, where

$$\gamma(\alpha) = \begin{cases} 1-\alpha, IF\ \alpha > 1 \\ \alpha - 1, IF\ \alpha < 1 \end{cases}.$$

As well as the Shannon mutual information, thus obtained $I_{R_\alpha}(X,Y)$ takes its values at the interval $[0,\infty)$. Meanwhile, the fraction, standing inside the sign of logarithm, has an evident interpretation as the cosine of the angle between vectors of corresponding Hilbert space formed by *α* times integrated functions mapping R^2 into R^1, where the inner product of its vectors $\phi_1(x,y), \phi_2(x,y)$ is defined by the natural expression

$$\langle \phi_1(x,y), \phi_2(x,y) \rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi_1(x,y)\phi_2(x,y)dxdy,$$

as well as the Euclidean norm of the vector $\phi(x,y)$ has the form

$$\|\varphi(x,y)\|_2 = \sqrt{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\phi(x,y)\right)^2 dxdy}.$$

Thus following to the notations introduced, the mutual Rényi information may be written as

$$I_{R_\alpha}(X,Y) = = \frac{1}{\gamma(\alpha)}\ln\frac{\left\langle \left(p_{xy}(x,y)\right)^{\alpha/2}, \left(p_x(x)p_y(y)\right)^{\alpha/2}\right\rangle}{\left\|\left(p_{xy}(x,y)\right)^{\alpha/2}\right\|_2 \left\|\left(p_x(x)p_y(y)\right)^{\alpha/2}\right\|_2} = = \frac{1}{\gamma(\alpha)}\ln\left(\cos\left(\left(p_{xy}(x,y)\right)^{\alpha/2}, \left(p_x(x)p_y(y)\right)^{\alpha/2}\right)\right).$$

For a partial case, when $\alpha = 2$, the preceding expression directly implies the sol called Cauchy-Schwartz divergence

$$D_{CS}(X,Y) = -\ln\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p_{xy}(x,y)p_x(x)p_y(y)dxdy}{\sqrt{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(p_{xy}(x,y)\right)^2 dxdy \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(p_x(x)p_y(y)\right)^2 dxdy}}, \quad (22)$$

proposed declaratively in the fullness of time in a number of papers (for instance, (Principe et al., 2000) and subsequent papers), that is by involving the Cauchy-Schwartz inequality, but disregarding condition (19)-(21) imposed on the relationship between the mutual Rényi information and corresponding Rényi entropies. Thus the Cauchy-Schwartz divergence $D_{CS}(X,Y)$ is a partial case of the mutual Rényi information $I_{R_\alpha}(X,Y)$ as $\alpha = 2$.

The Cauchy-Schwartz divergence $D_{CS}(X,Y)$ meets Rényi Axioms A, B, D, E and does not meet Axioms C, F, G. At the same time, $D_{CS}(X,Y)$ meets Axiom F in the case of affine transformations. In accordance to formulae (13), (14), for $D_{CS}(X,Y)$ in (22) one may construct the following transformation:

$$d_{CS}(X,Y) = 2\times\sqrt{1-\frac{2-\sqrt{4-3e^{-4D_{CS}(X,Y)}}}{e^{-4D_{CS}(X,Y)}}}. \quad (23)$$

One may show that the measure of dependence $d_{CS}(X,Y)$ in (23) meets all Rényi Axioms with the exception of Axiom F, but the property of invariance to one-to-one transformations is preserved for any affine transformations of random values. The behavior of the measure (23) in dependence of values of the Cauchy-Schwartz divergence is displayed in Figure 1.

Measure of dependence (23) determines solution (12) of the problem of the statistical linearization for the linearization criteria based on the Cauchy-Schwartz divergence (22).

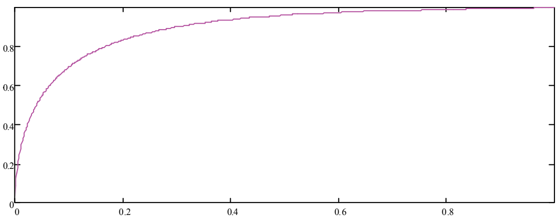


Figure 1: The dependence of the measure $d_{CS}(X,Y)$ in (23) of values of the Cauchy-Schwartz divergence (22).

6 EXAMPLE: SYSTEMS WITH ZERO INPUT/OUTPUT CORRELATION

As it was pointed out in Section 1, there exist numerous examples, when applying conventional correlation techniques under model deriving does not provide suitable results. Among such systems, one may emphasize those ones, for which the dependence between input and output variables is described by probability distribution densities belonging to the O.V. Sarmanov class of distributions (Sarmanov, 1967, Kotz et al., 2000). In particular, these involve the following one

$$p(x,y) = \frac{e^{-\frac{x^2+y^2}{2}}}{2\pi}\left(1+\lambda\left(2e^{-\frac{3}{2}x^2}-1\right)\left(2e^{-\frac{3}{2}y^2}-1\right)\right), \quad (24)$$

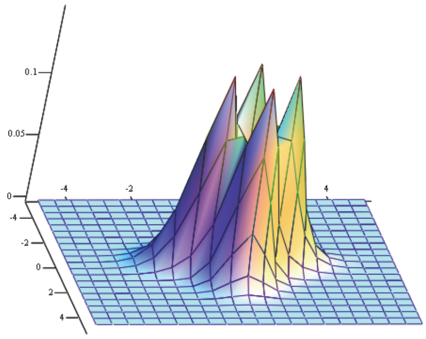
$$-1 \le \lambda \le 1$$

For density (24), the correlation coefficient $r(X,Y)$ and correlation ratio $\theta(X,Y)$ are equal to zero, while the maximal correlation coefficient is of the form:

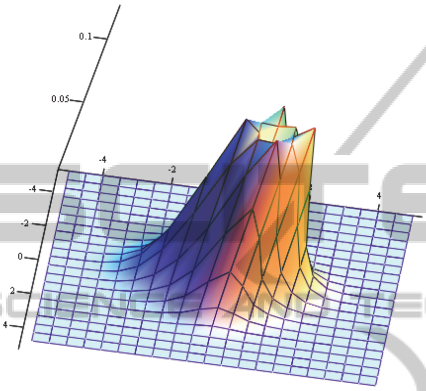
$$S(X,Y) = \left(\frac{4}{\sqrt{7}}-1\right)|\lambda|.$$

Magnitudes of the parameter λ considerably influence the shape of density (24). In Figure 2, the form of probability distribution density (24) under some magnitudes of the parameter λ is presented.

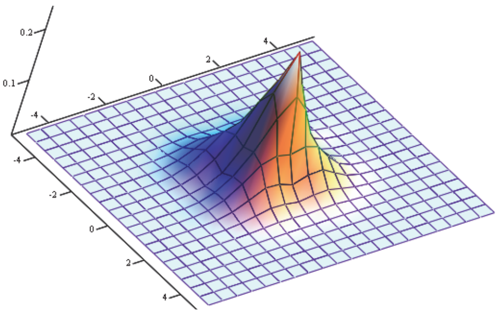
Thus, for instance, stochastic dependence (1) between the output process, $z(t)$, and the input process, $u(s)$, of a non-linear system is defined by a probability distribution density (of course, being not known to the researcher) of form (24) with the parameter $\lambda = \lambda(\tau)$, $\tau = t - s$, then applying both conventional correlation and dispersion techniques of the statistical linearization would lead, under constructing model (3), to the representation to the output system process as the identical zero, what is excluded under applying the approach presented, which is based on consistent measures of dependence.



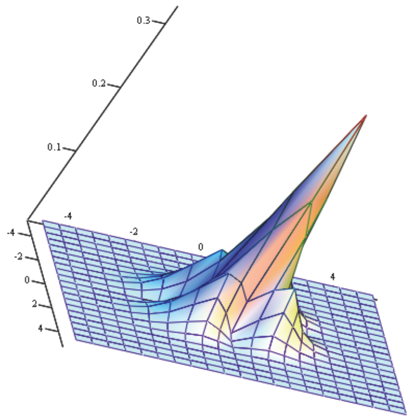
$\lambda = -1$



$\lambda = -1/2$



$\lambda = 1/2$



$\lambda = 1$

Figure 2: The shape of density (24) under various magnitudes of the parameter λ.

In Figure 3 (a, b, c), the dependence of the magnitudes of $\delta^2(Y,X)$ (17), $\iota(Y,X)$ (18), and $d_{CS}(X,Y)$ (23) in the parameter λ is presented correspondingly in the comparison with the magnitudes of the maximal correlation coefficient $S(X,Y)$ (the dotted line).

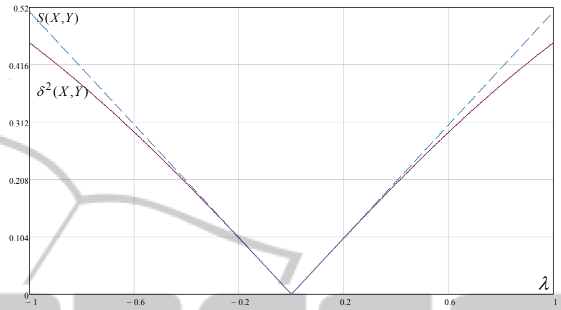


Figure 3a: The comparison of magnitudes of $\delta^2(X,Y)$ and $S(X,Y)$ under various values of the parameter λ in density (24).

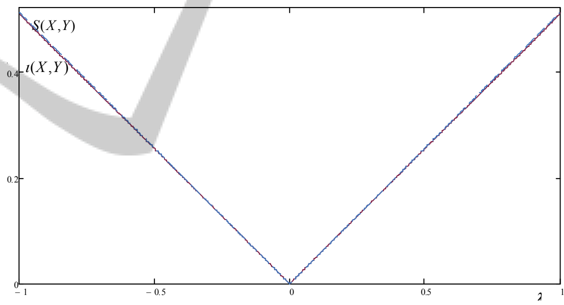


Figure 3b: The comparison of magnitudes of $\iota(X,Y)$ and $S(X,Y)$ under various values of the parameter λ in density (24).

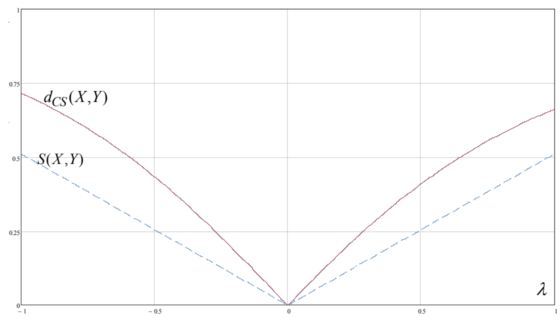


Figure 3c: The comparison of magnitudes of $d_{CS}(X,Y)$ and $S(X,Y)$ under various values of the parameter λ in density (24).

7 CONCLUSIONS

An approach to the statistical linearization of input/output mappings of stochastic discrete-time systems driven with a white-noise Gaussian input process has been considered. The approach is based on applying consistent measures of dependence of random values. Within the approach, the statistical linearization criterion is the condition of coincidence of the mathematical expectations of output processes of the system and model, and the condition of coincidence of a consistent, in the Kolmogorov sense, measure of dependence of output and input processes of the system and the same measure of dependence of the model output and input processes. Explicit analytical expressions for the coefficients of the weight function of the linearized input/output model were derived as a function of this (forming the statistical linearization criterion) consistent measure of dependence of output and input processes of the system. Meanwhile, such a function defines the form of a transformation that enables one to construct corresponding consistent in the Rényi sense measure of dependence from a consistent in the Kolmogorov sense measure of dependence. In the paper, a consistent in the Kolmogorov sense measure of dependence was referred as consistent in the Rényi sense, if such a measure meets all Rényi Axioms (Rényi, 1959) with the exception, may be, the axiom of invariance with respect to one-to-one transformations of random values under study. In particular, such a consistent in the Rényi sense measure of dependence has been constructed from the Cauchy-Schwartz divergence, being a consistent measure of dependence in the Kolmogorov sense.

REFERENCES

Gebelein, H., 1941. "Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichungsrechnung", *Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 21, no. 6, pp. 364-379.

Kotz, S., Balakrishnan, N., and N.L. Johnson, 2000. *Continuous Multivariate Distributions. Volume 1. Models and Applications* / Second Edition, Wiley, New York, 752 p.

Nadarajah, S. and K. Zografos, 2003. "Formulas for Rényi information and related measures for univariate distributions", *Information Sciences*, vol. 155, no. 1, pp. 119-138.

Nadarajah, S. and K. Zografos, 2005a. "Expressions for Rényi and Shannon entropies for bivariate distributions", *Information Sciences*, vol. 170, no. 2-4, pp. 173-189..

Principe, J., Xu, D., and J. Fisher, 2000. "Information Theoretic Learning", In: *Unsupervised Adaptive Filtering* / Haykin (Ed.). Wiley, New York, vol. 1, pp. 265-319.

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.

Rényi, A., 1961. "On measures of information and entropy", in: *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability* (June 20-July 30, 1960). University of California Press, Berkeley, California, vol. 1, pp. 547-561.

Rényi, A., 1976a. "Some Fundamental Questions of Information Theory", *Selected Papers of Alfred Rényi, Akademiai Kiado*, Budapest, vol. 2, pp. 526-552.

Rényi, A., 1976b. "On Measures of Entropy and Information", *Selected Papers of Alfréd Renyi, Akademiai Kiado*, Budapest, vol. 2, pp. 565-580.

Roberts, J.B. and P.D. Spanos, 2003. *Random Vibration and Statistical Linearization*, Dover, New York, 464 p.

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., 1963a. "Investigation of stationary Markov processes by the method of eigenfunction expansion", *Sel. Trans. Math. Statist. Probability*, vol. 4, pp. 245-269.

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

Sarmanov, O.V., 1967. "Remarks on uncorrelated Gaussian dependent random variables", *Theory Probab. Appl.,* vol. 12, issue 1, pp. 124-126.

Socha. L., 2008. *Linearization Methods for Stochastic Dynamic Systems*, Lect. Notes Phys. **730**, Springer, Berlin, Heidelberg, 383 p.

Zografos, K. and S. Nadarajah, 2005b. "Expressions for Rényi and Shannon entropies for multivariate distributions", *Statistics & Probability Letters*, vol. 71, no. 1, pp. 71-84.