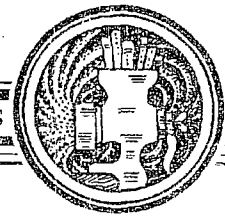
Review of actual problems of nonlinear dynamics



Izv. VUZ «AND», vol.6, № 4, 1998

# CHANNELS AND JOKERS: NEURAL VIEW OF COMPLEX DYNAMICS

*G.G. Malinetskii, A.B. Potapov*

One of the main aspects of brain activity is the ability to predict. Large efforts has been made in nonlinear dynamics to create predicting systems for dynamics of complex objects. One of the main tools for making such predictors is the application of multilayer neural networks. The methods based on the chaos theory prove to be less efficient, and in fact work only for low-dimensional model systems. From our point of view, the problems here are not technical, but related with the applicability of the approach of low-dimension nonlinear dynamics to real systems. Since the brain and some of its very simple models are able to make predictions in real situations, we propose to unify the ideas of nonlinear dynamics and neural networks. From our point of view, in complex real situations it may be possible to find low-dimensional projections, for which the approaches of nonlinear dynamics can be applied, but with serious restrictions. Most concepts, like attractor, its dimension, Lyapunov exponents etc. become inapplicable, and the observed phase space splits into predictable parts («channels») and non-predictable ones («jokers»), where probabilistic description may be more appropriate. We propose some mathematical basis for this idea and its possible application for time series analysis.

## 1. Introduction

One of the key problems in neuroscience is the processing of big information flows. Animal and human brain learned to find good solution for it. It is this fast processing of information about permanently changing situations around us that enabled us to survive in a struggle for life.

This information processing should be able:

- To change behavioural strategies very quickly, taking different features to make proper decision in various situations. That is, to find «the order parameters» of the complex situations.
- To learn not only by trial and error method (there is a lot of situations where there is no second trial), but by training of common sense, intuition, that is the «internal predicting system». Psychologists call this «a forestalling reflection».
- To react quickly. To do it, one must be able to «forget» quickly inessentials, or to send them into long-term storage, leaving in a quick memory only the most important information.

Therefore, the brain possesses very efficient methods of self-organization in information processing, and it is reasonable to look at the problems of current nonlinear dynamics from the viewpoint of neuroscience.

Predicting the future behaviour of a dynamical system and control of chaos is one of the key problems of nonlinear dynamics. Now several techniques of time series processing have been proposed, which enable to measure the main characteristics of

dynamical systems such as fractal dimensions, entropies, Lyapunov exponents, to make predictions of the future behaviour [1-4]. Several estimates of the expected performance of the methods have been proposed, and for rather simple model systems they were well confirmed. In particular, the effect of finite prediction time due to the sensitivity in respect to initial conditions has been analyzed.

But nonlinear dynamics techniques of data processing encountered very serious problems: all methods are efficient only for low-dimensional systems. It is hard to give the precise meaning of the term «low-dimensional», but different estimates [5-9] show, that usually algorithms cease to work for the systems with the attractor of dimension $d>5$. But this inequality leaves almost all practically important situations beyond the capabilities of nonlinear dynamics.

But applications of neural networks sometimes enabled to make predictions in situations which should be hopeless from the point of view of nonlinear dynamics, e.g. for financial time series. This implies two consequences: 1) there should be a reasonable explanation for such facts and probably the ways to overcome the limitations pointed out above; 2) neural networks possess some important feature which enable such data processing. The purpose of this paper is to present such an explanation.

The idea is that the phase space of the dynamical system is nonuniform. And there may be some places where the dynamics requires less variables for its description than in other ones. If a trajectory passes through such region, then at these moments it can be approximately described using the low-dimensional model. That is, the system can be characterized by its low-dimensional projection. We called such projections *Channels*. If the trajectory has visited the regions around this channel often enough, than this projection in principle can be found from a time series, and therefore the ability to make prediction appears. But feed-forward layered neural networks involve operations which resemble projecting: the weighted sum of inputs. Therefore, if a low-dimensional projection exists, a neural network may be capable to find it. Moreover, we shall show below, that projection is an important element of the most predictors, and maybe the ability to work with different projections explains the success of neural networks (and perhaps a real brain).

Therefore, it follows from the ideas of neuroscience that the possible approach for analysis of complex dynamical systems is to search for such local low-dimensional channels. But if we look at the whole dynamics from the viewpoint of such local channel, we shall see, that at some place this channel looses its predictive power. It may look as if from pure (or mainly) deterministic behaviour we quickly change for probabilistic one. Such domains of probabilistic description we shall call *Jokers*. In terms of low-dimensional projections, joker can throw the trajectory of dynamical system into another channel, or even to some part of the previous channel. Therefore, another possible idea is to study the behaviour of low-dimensional systems with jokers.

In these terms we can say that the brain has exceptional abilities in searching of channels, while jokers correspond to very complex situations which are hard to analyze. In such situations brain may activate the mechanisms of emotions which in some aspects may look probabilistic.

Probably, the treatment of complex systems from the channels and jokers viewpoint may give new useful practical ideas. In this paper we would like to show, how this approach emerges from the problem of time series prediction and that it enables to explain some known facts and to propose new ideas for developing numerical algorithms.

## 2. The problem of time-series prediction

Let us consider the general problem of time series prediction. Let $x_1, x_2, \ldots, x_N$ be the values of some observable measured at the moments $t_k=k\tau$. It is necessary to predict the future values $x_{N+1}, x_{N+2}, \ldots$ There are several approaches to this problem. In statistical approach it is assumed that the distribution density for $x_i$ depends on $m$ previous entries, and therefore it is possible to use for predictions the conditional average $E(x_i|x_{i-1}, x_{i-2}, \ldots x_{i-m})$. Nonlinear dynamics enabled to explain the appearing of this dependence and to propose the estimates of the value $m$.

The basic assumption of the nonlinear dynamics approach is that the measured values are the functions of the state of some dynamical system, which is responsible for the effects observed. That is, there exists a dynamical system

$$\mathbf{x}(t+\tau) = f^{\tau}(\mathbf{x}(t)), \qquad \mathbf{x} \in \mathrm{R}^n. \tag{1}$$

(This form enables to consider both mappings $\mathbf{x}_{n+1}=F(\mathbf{x}_n)$ and ODEs of the form $\dot{\mathbf{x}}=F(\mathbf{x})$ in a similar manner.) The second assumption is that the measured observable is a function of the state of this system, i.e. $x_i=h(\mathbf{x}(t_i))$. Then the Takens theorem [1,3] states that for almost all $\tau$, $h$, $f$ and $m \geq 2n+1$ there should be a functional relation between $x_{i-1}, x_{i-2}, \ldots, x_{i-m}$ and $x_i$.

The main idea of this theorem can be put as follows. All $m$ successive values of observable can be related to a single system state

$$x_{i-m+k} = h(f^{k\tau}(\mathbf{x}_{i-m})), \qquad k = 0,1,\ldots,m.$$

If we consider the sequence $x_{i-1}, x_{i-2}, \ldots, x_{i-m}$ as a point in $m$-dimensional Euclidean space

$$\mathbf{z}_{i-m} = (x_{i-1}, x_{i-2}, \ldots, x_{i-m})^T, \qquad \mathbf{z} \in \mathrm{R}^m, \tag{2}$$

then there is the vector function, such that $\mathbf{z}_i=\Lambda(\mathbf{x}_i)$. This function maps the phase space $M$ of the original dynamical system (1) (here $M=R^n$, but in general case it may be an $n$-dimensional manifold) into an $n$-dimensional surface $M_R \in R^m$, $\Lambda: M \to M_R$ or $M_R=\Lambda(M)$. According to the theorems of differential geometry, for $m \geq 2n+1$ and almost all functions this surface is an the embedding of original phase space into $R^m$, and there exists the inverse mapping $\Lambda^{-1}: M_R \to M$. Then it is possible to write $\mathbf{x}_{i-m}=\Lambda^{-1}(\mathbf{z}_{i-m})$, from which it follows that

$$x_i = h(f^{m\tau}(\mathbf{x}_{i-m})) = h(f^{m\tau}(\Lambda^{-1}(\mathbf{z}_{i-m}))) \equiv \Phi(x_{i-1}, \ldots, x_{i-m}). \tag{3}$$

The Takens theorem also enables to make several conclusions about the form of the function. It must consist of two parts: projecting and mapping. The theorem states that $\Phi$ is a component of the mapping of the $n$-dimensional surface $M_R$ into itself. Indeed, let us consider two vectors $\mathbf{z}_i=(x_i, x_{i+1}, \ldots, x_{i+m-1})$ and

$$\mathbf{z}_{i+1} = (x_{i+1}, x_{i+2}, \ldots, x_{i+m}) = (x_{i+1}, x_{i+2}, \ldots, x_{i+m-1}, \Phi(x_i, x_{i+1}, \ldots, x_{i+m-1})) \equiv \Psi(\mathbf{z}_i). \tag{4}$$

Both of them belong to $M_R$, and the function maps $M_R \to M_R$. In fact, (4) may be considered as another representation of the system (1). Then $\Phi$ must be a function of $n$ rather than $m$ arguments. The best choice will be the local coordinates on $M_R$, but they are usually unknown. So the optimal choice is the projection onto the tangent hyperplane to $M_R$ in the neighbourhood of $\mathbf{z}_i$ or to any other plane which is not orthogonal to it. As a rule, such projection (and the needed coordinate system) is only local, for this reason sometimes it is necessary to specify explicitly, which point $\mathbf{z}$ it relates to. Therefore, the general form of the predicting function or predictor must be

$$x_i = \Phi(P_n z_{i-m}) = \Phi(P_n(x_{i-1}, \ldots, x_{i-m})), \tag{5}$$

where $P_n$ denotes the projector onto $n$ local coordinates.

There is another reason, why the projection operator is necessary. In the presence of noise the points $\mathbf{z}_i$ will deviate from the surface $M_R$. But according to the theorem, the mapping is defined only on $M_R$. Then, to make the problem of time series prediction well-posed, instead of a point $\mathbf{z} \in R^m$ it is necessary to take its «reasonable» projection onto $M_R$: $\pi \circ \mathbf{z} \in M_R$. The specific form of the operator is not very important. Therefore, from the viewpoint of nonlinear dynamics, the problem of predicting consists in the approximating the unknown functional dependence by the set of known pairs $\{\mathbf{z}, \Phi(\mathbf{z})\}$. Several methods were reported in the literature [4]:

1) local linear and nonlinear approximations, i.e.

$$\Phi(\mathbf{z}) = \Phi(\mathbf{z}_0) + A_1(\Delta\mathbf{z}) + A_1(\Delta\mathbf{z},\Delta\mathbf{z}) + \ldots, \qquad \Delta\mathbf{z} = \mathbf{z}\text{-}\mathbf{z}_0,$$

here $A_k$ means a polynomial of the order $k$ of its arguments;

2) global polynomial approximations

$$\Phi(\mathbf{z}) = \Phi_0 + A_1(\mathbf{z}) + A_1(\mathbf{z},\mathbf{z}) + \ldots .$$

Note, it is interesting that the Takens theorem does not guarantee the existence of such approximations, but sometimes they prove to be efficient;

3) the method of radial basis functions

$$\Phi(\mathbf{z}) = \Sigma_i a_i \varphi(\|\mathbf{z}\text{-}\mathbf{z}_i\|).$$

Formally the previous note is valid in this case as well, but if the radial basis function $\varphi(r)$ decreases rapidly enough, then the domain where $\Phi(\mathbf{z})$ significantly changes is localized in the vicinity of the surface $M_R$. Perhaps, this is in some sense equivalent to projecting onto the surface;

4) neural networks with feed-forward architecture.

The comparison of different techniques for a number of model examples is presented in [10,4].

According to the results presented in the literature, for simple model systems (Lorenz, Henon attractors and other low-dimensional models) all methods of prediction work well, prediction error and predictability time are in good correspondence with theoretical estimates. But for real data, as experiments show, only local linear predictors, radial basis functions and neural networks are of practical importance (the example of such predictions can be found e.g. in [11], see also references therein).

## 3. Limitations of predicting techniques

The progress of the nonlinear techniques of data processing, such as estimating attractor dimension or Lyapunov exponents, enabled to understand their limitations. In particular, several relations have been obtained for the length of time series $N$ and the greatest dimension of attractor $d$, which can be estimated from this series in the best situation: $N \geq 10^d$ [5-9]. Several estimates use $d/2$ instead of $d$, but even under such conditions for available time series with $N \approx 10^3 \div 10^4$ it is hardly possible to obtain reliable results for the systems with $d>5$.

In the problems of time series prediction the obtaining of similar estimates is slightly harder. It is clear, that it is always possible to make *some* prediction with the help of the «zeroth order» method: $\Phi(\mathbf{z}) \cong \Phi(\mathbf{z}_0)$, where $\mathbf{z}_0$ is the nearest point with known value of $\Phi(\mathbf{z}_0)$ or as a weighted sum $\Phi(\mathbf{z}) \cong \Sigma_k w_k \Phi(\mathbf{z}_{0k})$, where $\mathbf{z}_{0k}$ are several nearest neighbours (the method of radial basis functions can be considered as a generalization of this approach). It is the *prediction error* which is of special interest. The «typical» error can be estimated as $|\delta x| \approx \|D\Phi(\mathbf{z}_0)\| \cdot \|\mathbf{z}\text{-}\mathbf{z}_0\|$. To approximately evaluate $\|\mathbf{z}\text{-}\mathbf{z}_0\|$, it is possible to use the hypothesis of homogeneous filling with data points of a $d$-dimensional cube with the edge $l$ (the magnitude of $x$ oscillations). If we denote by $a$ the mean distance between points, then $N \approx (l/a)^d$, and $\|\mathbf{z}\text{-}\mathbf{z}_0\| \cong a/2 \cong lN^{-1/d}/2$. Therefore, the expected relative prediction error can be approximately estimated as $\varepsilon_0 \approx \|D\Phi(\mathbf{z}_0)\| \cdot N^{-1/d}$. It is rather hard to estimate $\|D\Phi(\mathbf{z}_0)\|$, but it can be shown that it is proportional to $\exp(\lambda\tau)$, where $\lambda$ is the largest Lyapunov exponent.

In a similar manner it is possible to estimate the prediction error for local linear predictor («first order» method), $\varepsilon_1 \approx \|D^2\Phi(\mathbf{z}_0)\| \cdot N^{-2/d}$ [4]. The success of the nonlinear techniques, described in [4], can be explained by the use of small $d$ (1...3), and by the fact that for small $\tau$ usually $\|D\Phi(\mathbf{z}_0)\| \approx 1$ and $\|D^2\Phi(\mathbf{z}_0)\|$ also is $O(1)$, or even close to 0. If we take $\|D\Phi(\mathbf{z}_0)\| \approx \|D^2\Phi(\mathbf{z}_0)\| \approx 1$, $N \approx 10^3$, $d=2$, then it is easy to obtain that $\varepsilon_0 \approx 0.03$, $\varepsilon_1 \approx 0.001$. Close values were obtained in numerical experiments for model systems, see [4] and references therein.

For experimental data, such as physiological, economical time series, $d$ is usually

unknown, but scarcely it is less than 5, the estimate of derivatives also hardly can be obtained. But even assuming $\|D\Phi(\mathbf{z}_0)\|\approx\|D^2\Phi(\mathbf{z}_0)\|\approx 1$, $N\approx 10^3$, $d=5$ we get $\varepsilon_0\approx 0.3$, $\varepsilon_1\approx 0.1$. It is hard to say about physiology, but in economical problems such accuracy is absolutely unsatisfactory.

Nonetheless, there are situations, when for real complex systems good predictions were made, usually with the help of neural networks (there is no reliable information about such facts in publications, since all successes in financial predictions are commercial secrets, but brief notes in newspapers and numerous private communications insist that such events do occur). Below we shall try to explain this phenomenon, but first we shall consider the general structure of 3-layer neural network and its relation with the prediction problems and Takens theorem.

## 4.Predictors and 3-layer neural network

Let us consider the principal elements of a 3-layer feed-forward neural network:

$$\text{Inputs}\rightarrow \underset{\text{combinations}}{\text{Linear}} \rightarrow \underset{\text{function }(\sigma)}{\text{Sigmoid}} \rightarrow \underset{\text{combinations}}{\text{Linear}} \rightarrow \underset{\text{function }(\sigma)}{\text{Sigmoid}} \rightarrow \text{Outputs}$$

Since we are interested in predicting the next entry of the time series from $m$ previous ones, let us suppose the following network architecture: $m$ input neurons, a number of hidden neurons and a single output. $m$ previous values $x_{i-1},x_{i-2},\ldots,x_{i-m}$ or $\mathbf{z}_{i-m}$ form the input.

The linear combinations of the form

$$y_j = \Sigma_{k=1}^{m} A_{jk} x_{i-k} + \vartheta_j$$

can be considered as simultaneous calculation of the components of a number of projections, i.e. the components of $P_n\mathbf{z}$. Then the sigmoid function $\sigma(y_j)$ is calculated. As a rule, for this purpose a function with saturation is used, which approximately can be considered as piecewise linear: on the «working» interval $\sigma(x)\equiv cx$, while outside of it $\sigma(x)\equiv\pm 1$. With the help of $\vartheta_j$ the working interval can be shifted, such that for a given set of input parameters the value of $y_j$ fall into it («active component») or outside it («passive component»). This in principle enables to make active only $n$ necessary components $y_j$ and therefore to obtain on the neurons of the hidden layer the necessary projection $P_n\mathbf{z}$.

The next two steps are the calculation of the function $X=\sigma(\Sigma_j B_j y_j)$. This is nothing but a local linear approximation of the unknown function. And all these linear approximations prove to be continuously matched, like splines.

Therefore, a 3-layer network implements the main requirement of the Takens theorem: projection + approximation. This also explains the known fact, that increase of the number of layers usually do not improve approximation (see e.g. [10,12]). Combination of two projectors is equivalent to a single projector, and combination of linear approximations is again a linear approximation. Therefore, one can expect that basic features of multilayer network can be obtained on a 3-layer one with the proper number of hidden units and properly chosen sigmoid function.

## 5. When complex dynamics can be predicted? Channels and Jokers

Now let us consider the problem of predictability of complex dynamics. Above we pointed out the limitations of predictability. But they were related only to «global predictability» or to the ability to reconstruct the *whole* dynamical system in $\mathbf{z}$-representation (4). For a complex system it is indeed impossible. But maybe this can be done *locally*?

The mentioned information about successes of neural networks in making such

predictions is in favour of this idea, but as it has been said in previous section, neural networks possess very high «projecting capabilities». So, it seems reasonable to look for «predictability in projections».

Suppose that locally, in some domain $G$ of $n$-dimensional phase space, the behaviour of a complex system can be approximately but with good enough accuracy described by a low-dimensional model with the dimension of phase space $n_1<n$. Then, if the observed trajectory during the observation time has passed through $G$ several times, it may be insufficient to reconstruct general $n$-dimensional mapping, but enough to reconstruct $n_1$-dimensional function and to make good prediction. Under such circumstances we have no contradiction with the limitations of predicting techniques.

Moreover, this hypothesis enables to explain, why neural networks may occasionally form such local predictors. As it has been said, they form many projections of original phase space, and if it is enough for predictions $n_1<n$ parameters, then the network in principle can detect the presence of such domain $G$ and to form the corresponding low-dimensional predictor.

We shall call such domains $G$ as well as the corresponding low-dimensional models *channels*. When the trajectory enters such a channel, for some time its behaviour becomes predictable and «quasi-low-dimensional». When it leaves the channel, the behaviour becomes more complex again. The low-dimensional deterministic description fails, the situation looks like partially probabilistic. For qualitative description of such situations by low-dimensional models we propose the new class of models - the *systems with jokers*. *Joker* is a domain in phase space, where the behaviour of the deterministic system becomes probabilistic. For example, joker can throw the trajectory to almost any point of phase space, and after it for some time the dynamics become low-dimensional deterministic again. In other words, the approach of channels and jokers is an attempt to apply the low-dimensional ideas to the analysis of high-dimensional systems.

**5.1.** ***How channels can arise.*** Let us consider this idea in more details. Suppose, that for the dynamical system (1) there exist a domain $G$, where the function $f(\mathbf{x})$ has the form

$$f(\mathbf{x}) = f_1(P_{n_1}\mathbf{x}) + \gamma f_2(\mathbf{x}), \qquad \mathbf{x} \in G \subset R^n,$$

where $\gamma<<1$ and $P_{n_1}$ is the projector onto the subspace of the dimension $n_1<n$. This projection can be considered as a plane (or, in general, a surface) $\mathbf{P}$ through some point $\mathbf{x}_0 \in G$, and the operator $P_{n_1}$ - as projector onto it. Let us also denote $n_1$ coordinates on $\mathbf{P}$ by $\mathbf{u}=P_{n_1}\mathbf{x}$, and the rest $n$-$n_1$ - by $\mathbf{v}=(I-P_{n_1})\mathbf{x}$. Then $f(\mathbf{x})=f_1(\mathbf{u})+\gamma f_2(\mathbf{u},\mathbf{v})$, and on the surface $\mathbf{P}$ we obtain the mapping

$$\mathbf{u}_{i+1} = P_{n_1} f_1(\mathbf{u}_i) + \gamma P_{n_1} f_2(\mathbf{u}_i,\mathbf{v}_i), \qquad \mathbf{u}_i,\mathbf{u}_{i+1} \in P. \tag{6}$$

If the accuracy of the desired prediction enables to drop out the second term, for example, if $\gamma$ is very small, then the dynamics can be approximately reduced to $n_1$-dimensional

$$\mathbf{u}_{i+1} = P_{n_1} f_1(\mathbf{u}_i). \tag{7}$$

Therefore, if one needs to predict a component of $\mathbf{u}$, it can be done with the help of (7). The predictions for the components of $\mathbf{v}$ are also possible, but they will depend on $\mathbf{u}$.

Note, that (7) is in fact the equation for the channel associated with the domain $G$.

**5.2.** ***Channels and time series prediction.*** Now let us consider, how the idea of channels can be applied in the problem of time series prediction. In this case we have to deal with the dynamical system (4) and its component (3). We assume that the complete system is high-dimensional, but perhaps somewhere in the reconstructed space $R^m$ there are domains $G_k$, where it is possible to apply the approach of low-dimensional channels. The dimension of the channel we can approximately estimate by means of the methods, applied in the Sect. 2. Let us take as desired accuracy 1%, i.e. ~0.01, and assume $\|D\Phi(\mathbf{z}_0)\| \approx \|D^2\Phi(\mathbf{z}_0)\| \approx 1$ and $N \approx 10^3$. Then for the 1-st order method $\varepsilon_1 \approx N^{-2/d}$ or

$$d \cong -2\lg N/\lg\varepsilon_1 = 3. \qquad (8)$$

Therefore, it is necessary to search for a domain, where the dynamics can be predicted by 3...6 most important components of the vector $\mathbf{z}$, in other words, to determine the projector $P_{n_1}$ (note, that vector $\mathbf{z}$ may or even must be high-dimensional). How to do it is a separate problem, and we shall not discuss it here, though practically it is very important. We suppose, that such domain $G$ and the projector $P_{n_1}$ are found. Let the projection be an $n_1$-dimensional hyperplane $\mathbb{P}$. In the case of scalar time series prediction the component to be predicted is known - it is the last component of $\mathbf{z}$. For this reason we shall suppose that the vector $\mathbf{e}=(0,...,0,1)$ is not orthogonal to $\mathbb{P}$. For practical applications more strong condition must hold: angle between $\mathbf{e}$ and $\mathbb{P}$ should be less than some limiting value, say $\beta<\beta_{max}=60°$. For the components in projection $\mathbf{u}=P_{n_1}\mathbf{z}$ there will be an analog of (7): the approximate reduced system

$$\mathbf{u}_{i+1} = \mathbf{g}(\mathbf{u}_i). \qquad (9)$$

Now let us explicitly express $x_{i+1}$. By definition, $x_{i+1}=(\mathbf{e}, \mathbf{z}_{i+1})$. Let us denote $\mathbf{a}=P_{n_1}\mathbf{e}/\|P_{n_1}\mathbf{e}\|=P_{n_1}\mathbf{e}/\cos\beta$ - the direction on $P$ which contains maximal information about $x_{i+1}$, $\|\mathbf{a}\|=1$. Since $\beta$ is the angle between $\mathbf{e}$ and $\mathbf{a}$, $\mathbf{a}=\mathbf{e}\cos\beta+\mathbf{q}'$, where $\mathbf{q}'$ is orthogonal to $\mathbf{a}$. Then $\mathbf{e}=(\mathbf{a}-\mathbf{q}')/\cos\beta$, and

$$x_{i+1} = (\mathbf{e},\mathbf{z}_{i+1}) = ((\mathbf{a},\mathbf{z}_{i+1}) - (\mathbf{q}',\mathbf{z}_{i+1}))/\cos\beta = ((\mathbf{a},\mathbf{u}_{i+1}) - (\mathbf{q}',\mathbf{z}_{i+1}))/\cos\beta.$$

But because $\mathbf{q}'\perp\mathbf{e}$, it will have nonzero projections only on those components of $\mathbf{z}_{i+1}$, which are present in $\mathbf{z}_i$, that is, there exist such $\mathbf{q}$, that $(\mathbf{q}',\mathbf{z}_{i+1})=(\mathbf{q},\mathbf{z}_i)$. So,

$$x_{i+1} = ((\mathbf{a},\mathbf{g}(\mathbf{u}_i)) - (\mathbf{q},\mathbf{z}_i))/\cos\beta = g(\mathbf{u}_i) + (\mathbf{q},\mathbf{z}_i). \qquad (10)$$

This relation gives the general form of the predictor which uses the channels approach - it is the sum of a nonlinear function of channel coordinates $\mathbf{u}$ and a linear function of the previous state $\mathbf{z}$.

Therefore, the use of channels may enable to simplify the structure of the predictors, and to make predictions for high-dimensional systems, which, according to general estimates, are beyond the applicability of low-dimensional nonlinear techniques.

But note, that the accuracy of prediction is limited not only by 1) the errors in data and 2) chaoticity of the dynamical system. The serious source of errors is 3) the discarded term in (6), which imposes the limits of prediction error and can not be diminished within the low-dimensional approach. For this reason, more appropriate model for the channel may be a dynamical system with noise added, but perhaps this «noise» should possess some dynamical features to represent a projection of a high-dimensional trajectory.

**5.3.** ***How to search for channels?*** Searching for channels seems to be a complex problem. At present we would like only to make several brief notes about it.

This problem is related with other techniques, proposed in nonlinear dynamics and statistics earlier. It is possible to mention the techniques of False Nearest Neighbours (FNN) [4], search for dependent variables [13,14], or the attempts to apply the ideas of Principal Components Analysis [15]. But all these techniques were global, while channels require local approach. Therefore, it is necessary to find new technique.

It seems, that the most promising will be the standard approach of searching for a functional dependence between successive reconstructed vectors, which are usually applied for determining proper embedding dimension [16,17,4] (in [4] it is called FNN). The idea is rather simple: if there exist a functional relation between $\mathbf{z}_i$ and $\mathbf{z}_{i+1}$, then if $\|\mathbf{z}_i-\mathbf{z}_j\|$ is small, the same should be true for their iterates by (4), i.e. $\|\mathbf{z}_{i+1}-\mathbf{z}_{j+1}\|$ also should be small. (Another way of using this idea is to compare distances in reconstructions with embedding dimensions $m$ and $m+1$.) To find channels, it is possible to apply similar technique, but only in projection, which in turn to be found.

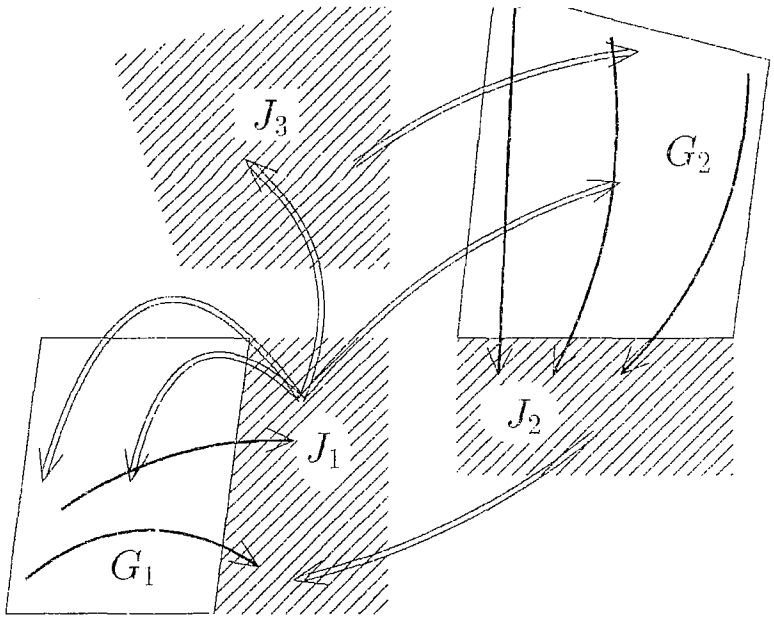



Fig. 1. Scheme of channels-jokers view of a complex dynamics. There are 2 channels ($G_1$ and $G_2$) and 3 jokers ($J_1$, $J_2$, $J_3$). Solid arrows show deterministic description of dynamics (a trajectory of the reduced system), empty arrows show the action of jokers: when trajectory enters the joker area (shaded), it may jump with some probability to some point of a channel or to another joker

That is, we come to the following problem: for a high-dimensional embedding ($m$ may be large) find a low-dimensional projection, i.e. $n_1$=3...6 orthonormal vectors $\mathbf{a}_k$, which define the projector $P_{n_1}\mathbf{x}=\Sigma_k(\mathbf{x},\mathbf{a}_k)\mathbf{a}_k$, and the domain $G$, where it is natural to expect a functional dependence between $P_{n_1}\mathbf{z}_i$ and $P_{n_1}\mathbf{z}_{i+1}$. A possible approach is to study the relation between $\|P_{n_1}(\mathbf{z}_i-\mathbf{z}_j)\|$ and $\|P_{n_1}(\mathbf{z}_{i+1}-\mathbf{z}_{j+1})\|$, look at the distribution of potentially dependent pairs in projection and adjust the vectors $\mathbf{a}_k$. This problem requires very large amount of computations, and probably, it is necessary to find efficient numerical algorithms for this purpose to achieve good performance, comparable with that of neural networks.

**5.4.** ***What is at the end of a channel?*** Within a channel it may be possible to obtain a simple description for a complex system. What will happen when the channel ends (trajectory leaves the domain $G$), while we would like to remain in the frame of low-dimensional description of reality? Simple models can not any more make deterministic prediction, and the only way to remain within low-dimensional paradigm is to admit probabilistic behaviour of the system's model. That is, we suppose that there are domains $J_k$ (jokers), where the behaviour of the trajectory becomes probabilistic. For example, joker can throw the trajectory back into some (or any) point of the same channel $G$ or another channel, or the trajectory may jump between several jokers etc. This scheme is shown in Fig. 1.

Properties of 1-D maps with jokers of different types were studied in [18,19]. It was shown that the presence of joker can drastically change the bifurcation diagram and even suppress arising of dynamical chaos.

## 6. A simple example

Now we shall present a simple model example, where the concepts of channel and joker can be shown. Let us consider a chaotic system consisting of two coupled parts, each of them being again a chaotic system:

$$\mathbf{x}_{n+1} = f_1(\mathbf{x}_n) + \mu(\mathbf{x}_n) g_1(\mathbf{y}_n), \qquad \mathbf{x} \in R^{n_1},$$

$$\mathbf{y}_{n+1} = f_2(\mathbf{y}_n) + \mu_0\, g_2(\mathbf{x}_n), \qquad \mathbf{y} \in R^{n_2}.$$

Here $\mathbf{x}$ is supposed to be low-dimensional. The variable coupling we shall choose such that $\mu(\mathbf{x}_n) \equiv 0$ when $\mathbf{x}_n \in G$, $G$ is a domain of $R^{n_1}$, and $\mu(\mathbf{x}_n) \neq 0$ when $\mathbf{x}_n$ is outside it. Then, while $\mathbf{x}$ falls within $G$, we have two almost separate subsystems, and in particular $\mathbf{x}$ obeys the equation $\mathbf{x}_{n+1} = f_1(\mathbf{x}_n)$. This gives a channel. Then we can obtain a time series for some observable $x_n = h(\mathbf{x}_n)$, and see, whether this channel can be detected. It is this *choice of observable* that makes the example simple and enables us to avoid the very complex stage of searching the necessary projection.

In the example below for the mapping $f_1(\mathbf{x}_n)$ we have chosen the modified Hénon mapping

$$\begin{aligned} x_{1,n+1} &= 1 - a[2\sin(x_{1,n}/2)]^2 + b x_{2,n} \equiv \varphi(x_{1,n}, x_{2,n}), \\ x_{2,n+1} &= 2\sin(x_{1,n}/2) \equiv \psi(x_{1,n}), \qquad \mathbf{x}_n = \begin{bmatrix} x_{1,n} \\ x_{2,n} \end{bmatrix}. \end{aligned} \tag{11}$$

The modification is necessary to avoid the escape of the trajectory to infinity, which occurs in original Hénon mapping. For the mapping $f_2(\mathbf{y}_n)$ we used three identical coupled maps (11) with constant coupling. The resulting system has the form

$$\begin{aligned} x_{1,n+1} &= \varphi(x_{1,n}, x_{2,n}) + \mu(x_{1,n}) y_{1,n}, & x_{2,n+1} &= \psi(x_{1,n}) \\ y_{1,n+1} &= \varphi(y_{1,n}, y_{2,n}) + \mu_0 (x_{1,n} + y_{3,n})/2, & y_{2,n+1} &= \psi(y_{1,n}) \\ y_{3,n+1} &= \varphi(y_{3,n}, y_{4,n}) + \mu_0 (y_{1,n} + y_{5,n})/2, & y_{4,n+1} &= \psi(y_{3,n}) \\ y_{5,n+1} &= \varphi(y_{5,n}, y_{6,n}) + \mu_0 y_{3,n}, & y_{6,n+1} &= \psi(y_{5,n}) \end{aligned} \tag{12}$$

$$\mu(x) = \mu_0[1 - \mathrm{th}(5x)]/2, \qquad x_n = h(\mathbf{x}_n) \equiv x_{1,n}$$

$$a = 1.4, \qquad b = 0.3, \qquad \mu_0 = 0.3.$$

It follows from the form of $\mu(x)$, that the domain $G$ corresponds to $x_{1,n} > 0$. This choice has been made on purpose, because in this half-plane the mapping $f_1(\mathbf{x}_n)$ has the fixed point, and the trajectory sometimes spend several iterations near it. In other words, this has been made to ensure, that the trajectory will spend several successive iterations within $G$, which may be important for channel detection in delay reconstruction.

At first glance, the time series for (12) slightly differs from that of unperturbed (11), but the effect of varying coupling is clearly noticeable on the plot of correlation integral, and particularly - its slope (see Fig. 2). From the plots one can conclude, that most probably the processed series is generated by a low-dimensional system, but steady growth of slope with the increase of embedding dimension makes it to look a little «random». Therefore, in the projection onto $\mathbf{x}$-plane we obtain the situation described above: low-dimensional dynamics within $G$ and more complex behaviour outside it. Here the «joker» appears to be rather weak and just adds moderate «noise» to the low-dimension «signal».

To apply our approach to the analysis of this time series, we must find the domain of channel $G$ (again note, that here we do not have to search for the necessary low-dimensional projection, we get it almost automatically because of the choice of observable; in real problems the situation will be worse). To find the channel, we applied rather simple technique, which can be called the *linear predicting test* (LPT).

The idea of LPT can be explained as follows. As it has been mentioned above, the prediction of time series means interpolation of the function $\Phi(\mathbf{z})$ (3) at the necessary

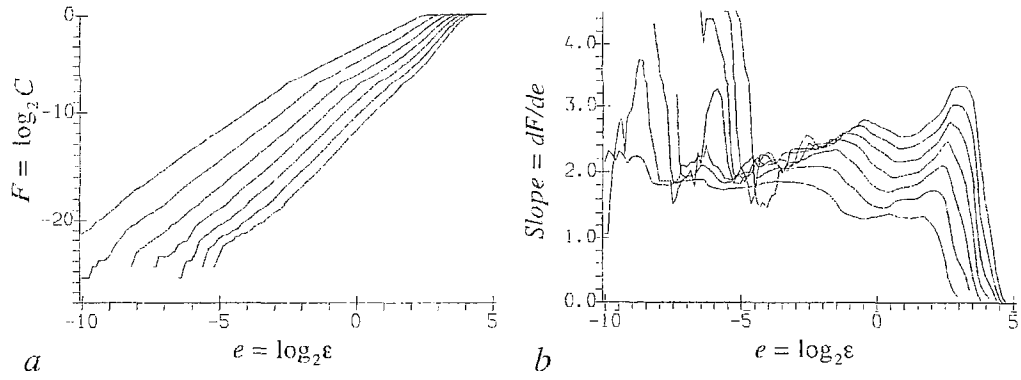



Fig. 2. Plot of correlation integral $\log_2 C(\varepsilon)$ vs $\log_2\varepsilon$ (*a*) and its slope (*b*) for the time series of the observable $x_{1,i}$ for the model system (12). The length of the series $N=10^4$, the embedding dimensions $m$=4, 6, 8, 10, 12, 14, 16. It shows that the set of **z**-vectors does not look random, rather they form «something low-dimensional», but the structure of this set changes with the increase of $m$ and slope gradually grows. Usually this is interpreted as presence of noise. In our case, from the point of view of low-dimensional channel it is the influence of joker, from the point of view of the whole system (12), this is the consequence of the problems with applicability of the Takens theorem: due to varying coupling $\mu(\mathbf{x})$ the observable $x_{1,i}$ does not enable proper reconstruction of the whole system

point **z** from the known values of $\Phi(\mathbf{z}_{0k})$ at neighbouring points $\mathbf{z}_{0k}$. For the simplicity, let us consider one-dimensional case: some function $f(x)$ is known at discrete points $x_i$, $f_i=f(x_i)$, and one has to interpolate its value at some point $x\in[x_i,x_{i+1}]$. Also for the simplicity let $x_{i+1}-x_i=h$ for all $i$, then the linear approximation, which uses the neighbouring points $f_{i+1}$ and $f_i$, is

$$f(x) \equiv \mathrm{Lin}(f_i, f_{i+1}, x) \equiv ax + b = (f_{i+1}+f_i)/2 + (f_{i+1}-f_i)/(2h)(x-x_c), \qquad x_c = (x_{i+1}+x_i)/2.$$

The error of this approximation can be roughly estimated as $\varepsilon\sim f''(x)h^2$, $h=x_{i+1}-x_i$. The second derivative can be approximated by the difference

$$f''h^2 \cong (f_{i+1}-2f_i+f_{i-1})h^2/(2h^2) = (f_{i+1}-2f_i+f_{i-1})/2 = (f_{i+1}+f_{i-1})/2 - f_i = \mathrm{Lin}(f_{i-1}, f_{i+1}, x_i) - f_i.$$

Therefore, the approximation error is of the same order as the difference between the value of $f$ at some point and its linear approximation from the nearest neighbours. This form enables easy generalization for higher dimensions. Then we come to the LPT. Let us estimate the «quality» of the point $\mathbf{z}_i$ as follows.

1. Take $k=k_0>m+2$ nearest neighbours of $\mathbf{z}_i$, $\mathbf{z}_{js}$, $s=1,\ldots,k$, $\Phi(\mathbf{z}_{js})$ are known, and make a linear approximation $L_k(\mathbf{z})$ by these $k$ neighbours (but not using $\Phi(\mathbf{z}_i)$). This gives the values $\varepsilon_1=|L_k(\mathbf{z}_i)-\Phi(\mathbf{z}_i)|$, and $\varepsilon_2=\max_s|L_k(\mathbf{z}_{js})-\Phi(\mathbf{z}_{js})|$.

2. Let us decrease $k$: we shall discard one of the neighbours, $\mathbf{z}_{js}$, for which this operation will give the smallest value of $\varepsilon=|L_{k-1}(\mathbf{z}_i)-\Phi(\mathbf{z}_i)|$. We shall repeat this operation until (i) $k\geq m+2$ and (ii) the decrease of $\varepsilon$ is at least 2%.

3. Finally we obtain the new $k^*<k_0$, and the values

$$\varepsilon_3=|L_{k^*}(\mathbf{z}_i)-\Phi(\mathbf{z}_i)|,\ \varepsilon_4=\max_s|L_{k^*}(\mathbf{z}_{js})-\Phi(\mathbf{z}_{js})|.$$

In most cases it is possible to classify the points $\mathbf{z}_i$ with the help of $\varepsilon_1,\ldots,\varepsilon_4$. Fig. 3 shows the results. Dots show «good» points, crosses - «bad» ones. It is obvious, that the domain of channel can be extracted from these data rather easily. That is, we get the domain of good predictability for our time series and a simple rule for testing, whether $x$ belongs to it. We can conclude, that at least in some cases the approach of channels and jokers can be useful.

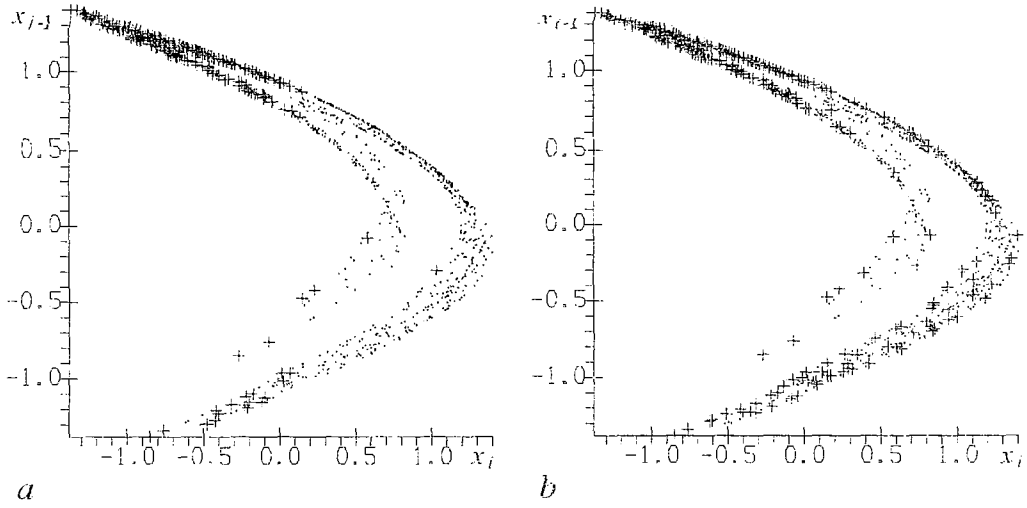



Fig. 3. The results of linear prediction test (LPT) for the short ($N=10^3$) time series of the observable $x_i$ for the model system (12). Crosses show «bad» points, where predictability is bad, and dots show «good» points. For application of LPT for the low-dimensional projection ($m=2$, panel *a*) the domain of channel is clearly visible, it corresponds to half-plane $x_i>0$. For higher-dimensional projection ($m=6$, panel *b*, the coordinates of the plot are the last two components of 6-dimensional **z**-vector) LPT almost fails to detect channel. Therefore the technique of channels search requires searching for proper projection. In our case this search is not necessary because of the good choice of observable

## 7. Conclusions and hypotheses

Our hypothesis is that for complex high-dimensional systems such channels-jokers description may prove useful. In some aspects it can be considered as a generalization of symbolic dynamics, in other aspects - as a system of matched simple models. In natural sciences such description is not popular, but it may be appropriate for social and life sciences, where usually many simple models are used to describe different aspects of the same complex object. Probably, those models can be considered as different channels. Then, for example, the dynamics of society can be represented as a series of channels (regular development) and jokers (abrupt changes, revolutions etc.). Such view can also help in discussions on the correctness of different models: several descriptions may coexist as different possible channels. The problem is which one corresponds to the present situation , how close is the nearest joker, can it be avoided and so on.

This approach also can give some hints on the problems of description of complex objects. In principle, channels need not always be mathematical models, they may be some typical situation, combination of indications, most important details. Therefore, an object can be characterized by an «album» of such typical situations with most probable consequences.

It is interesting that brain possesses great capabilities of finding such important details, creating «channels» and making predictions. If the set of important parameters is incomplete, mistakes will occur, therefore, the previous experience may correspond to the ability of creating proper projections of reality. It seems that this capability is partially inherited by artificial neural networks, and this may be one of the reasons for their successes.

Besides such «philosophical» outcome, we expect that the approach of channels and jokers may be useful in more common problems, such as time series predictions. One of such possible applications is proposed in this paper. Note, that when one uses such local low-dimensional models, the global invariant characteristics like attractor, fractal dimension, Lyapunov exponents, entropy and so on, can not characterize the model any more. Only local ones, such as local divergence rate of trajectories can be used.

*This work was partially supported by RFBR by grants № 96-01-01161, 96-02-18689, 97-01-00396.*

## References


1. *Takens F.* Detecting strange attractors in turbulence // Dynamical systems and turbulence. Lect. Notes in Math. Berlin: Springer, 1981. Vol. 898. P. 336.

2. *Eckmann J.P., Ruelle D.* Ergodic theory of chaos and strange attractors // Rev. Mod. Phys. 1985. Vol. 57. P. 617.

3. *Sauer T., Yorke J.A., Casdagli M.* Embedology // J. Stat. Phys. 1991. Vol. 65. P. 579.

4. *Abarbanel H.D.I., Brown R., Sidorowich J.J., Tsimring L.S.* The analysis of observed chaotic data in physical systems // Rev. Mod. Phys. 1993.Vol. 65. P. 1331.

5. *Smith L.A.* Intrinsic limits on dimension calculations // Phys. Lett. A. 1988. Vol. 133. P. 283.

6. *Eckmann J.-P., Ruelle D.* Fundamental limitations for estimating dimensions and Liapunov exponents in dynamical systems // Physica D. 1992. Vol. 56. P. 185.

7. *Ruelle D.* Deterministic chaos: the science and the fiction // Proc. Roy. Soc. London A. 1990. Vol. 427. P. 241.

8. *Malinetskii G.G., Potapov A.B., Rakhmanov A.I.* Limitations of delay reconstruction for chaotic dynamical systems // Phys. Rev. E. 1993. Vol. 48. P. 904.

9. *Hegger R., Kantz H., Olbrich E.* Problems in the reconstruction of highdimensional deterministic dynamics from time-series // Nonlinear analysis of physiological data / Eds. H.Kantz, J.Kurths, G.Mayer-Kress. Springer, 1998 (in press).

10. *Casdagli M.* Nonlinear prediction of chaotic time series // Physica D. 1989. Vol. 35. P. 335.

11. *Lefebvre J.H., Goodings D.A., Kamath M.V., Fallen E.L.* Predictability of normal heart rhythms and deterministic chaos // Chaos. 1993. Vol. 3. P. 267.

12. *Gencay R., Dechert W.* An algorithm for the $n$ Lyapunov exponents of an $n$-dimensional unknown dynamical system // Physica D. 1992. Vol. 59. P. 142.

13. *Savit R., Green M.* Time series and dependent variables // Physica D. 1991. Vol. 50. P. 95.

14. *Wu K., Savit R., Brock W.* Statistical tests for deterministic effects in broad band time series // Physica D. 1993 Vol. 69. P. 172.

15. *Broomhead D.S., King G.P* Extracting qualitative dynamics from experimental data // Physica D. 1986. Vol. 20. P. 217.

16. *Cenys A., Pyragas K.* Estimation of the number of degrees of freedom from chaotic time series // Phys. Lett. A. 1988. Vol. 129. P. 227.

17. *Cremers J., Hubler A.* Construction of differential equations from experimental data // Z. Naturforsh. A. 1987. Vol. 42. P. 797.

18. *Bellachook L.V., Malinetskii G.G.* Tricks of jokers on one-dimensional maps // Proc. 5 Int. Spec. Workshop: Nonlinear Dynamics of Electronic Systems. Moscow, 1997.

19. *Bellachook L.V., Malinetskii G.G.* Tricks of jokers on one-dimensional maps. Preprint. M.V. Keldysh Inst. Appl. Math. 1997. № 24.



*Keldysh Institute of Applied Mathematics, RAS*



*Received 15.05.98*



УДК 517.9


# РУСЛА И ДЖОКЕРЫ
# НЕЙРОСЕТЕВОЙ ВЗГЛЯД НА СЛОЖНУЮ ДИНАМИКУ


*Г.Г. Малинецкий, А.Б. Потапов*


Способность к предсказаниям - один из важнейших аспектов деятельности мозга. В нелинейной динамике были вложены большие усилия в разработку методов прогноза поведения сложных систем. Одним из основных средств при этом служат многослойные нейронные сети. Методы, основанные на идеях

нелинейной динамики оказываются не столь эффективными и работают, фактически, только для модельных систем небольшой размерности. С нашей точки зрения, проблемы здесь не технические, а связанные с применимостью подходов маломодовой нелинейной динамики к реальным системам. Поскольку мозг и некоторые простейшие модели нейронных сетей способны к прогнозированию в реальных ситуациях, мы предлагаем объединить идеи нелинейной динамики и нейронных сетей. С нашей точки зрения, в сложных жизненных ситуациях может существовать возможность обнаружить проекции малой размерности, для которых подходы нелинейной динамики могут быть использованы, но с серьезными ограничениями. Большинство понятий, таких как аттрактор, его размерность, ляпуновские показатели и т.п. становятся неприменимыми, а фазовое пространство распадается на области предсказуемости («русла») и области непредсказуемости («джокеры»), где более адекватным является вероятностное описание. Мы предлагаем некоторое математическое обоснование этой идеи и возможное ее использование в задачах анализа временных рядов.

*Malinetskii Georgyi Gennadievich* was born in Ufa (1956), graduated from the Physical Department of the Moscow State University (1979). His Ph. D. work was devoted to nonstationary dissipative structures in nonlinear media. He got his Professor degree (1990) working in the field of diffusional chaos and new types of order in nonlinear media in Institute of Applied Mathematics RAS. Now he is a Head of Nonlinear Dynamics Section in the mentioned above institute. He is the author of the multitude of papers in the field of chaos and nonlinear phenomena investigation, and the text book «Structures, chaos, numerical experiment. Introduction into nonlinear dynamics». E-mail: gmalin@spp.keldysh.ru

*Potapov Alexei Borisovich* was born in 1960. He graduated from the Physical Department of the Moscow State University in 1983. Now he is a Senior Science Fellow of the Institute of Applied Mathematics RAS. The main fields of scientific activity - nonlinear dynamics, time series processing, quantitative characteristics of chaos, modelling of complex systems.
E-mail: copal@spp.keldysh.ru