Non-Uniform Chaotic Dynamics with Implications to Information Processing

J. S. Nicolis*, G. Meyer-Kress, and G. Haubs
Institut für Theoretische Physik, Universität Stuttgart

Z. Naturforsch. **38 a**, 1157–1169 (1983); received June 29, 1983

We study a new parameter – the "Non-Uniformity Factor" (NUF) –, which we have introduced in [1], by way of estimating and comparing the deviation from average behavior (expressed by such factors as the Lyapunov characteristic exponent(s) and the information dimension) in various strange attractors (discrete and chaotic flows). Our results show for certain values of the control parameters the inadequacy of the above averaging properties in representing what is actually going on – especially when the strange attractors are employed as dynamical models for information processing and pattern recognition. In such applications (like for example visual pattern perception or communication via a burst-error channel) the high degree of adherence of the processor to a rather small subset of crucial features of the pattern under investigation or the flow, has been documented experimentally: Hence the weakness of concepts such as the entropy in giving in such cases a quantitative measure of the information transaction between the pattern and the processor. We finally investigate the influence of external noise in modifying the NUF.

1. Introduction

Taking averages in physical sciences in general and in communication theory in particular results always in some selective loss of detail. If it happens that a few details account practically for the whole pattern then the averaging process simply "washes out" all the essential information. In statistical mechanics for one the pursuit of evolution of the microscopic probability density function and its moments through the formalism of the Master equation and Fokker-Planck-equation in systems far from equilibrium and near bifurcation points manifest the "break down" of the law of large numbers; this has been amply demonstrated in recent years [2, 3] – together with the ensuing invalidation of the "mean field regime". The entropy for example is just the mean value of the distribution $-\ln p(x)$ where $p(x)$ stands for the *a priori* probability density distribution of a (finite or infinite) set of elements constituting a certain pattern. Some of the elements of the set may be extremely improbable vis-a-vis a certain observer or prone to deliver upon reception a disproportionally large amount of information. Of particular interest is the case where the median value of $p(x)$ is the least probable. In such cases the usual expression(s) for the static entropy:

$$S = -\sum_i p_i(x) \log_2 p_i(x) \tag{1}$$

or

$$\Delta S = -\int p(x) \log_2 p(x)\, \mathrm{d}x \quad \text{(in bits)} \tag{2}$$

is perhaps inadequate in characterizing quantitatively the information transaction.

In this paper we intend to treat dynamical systems where the variety production or information dissipation are given by the dynamical analogs of the entropy and are couched in terms of the Lyapunov-exponents of the flow or the discrete map concerned.

In the following we do three things. First we briefly review some experimental evidence about the dynamics of visual pattern perception and recognition (what are the "crucial features" of the pattern in such a case and how is the processor dealing with them?) as well as the irrelevance of the "law of averages" in certain "coin tossing" and communication problems. Second we discuss the possible use of strange attractors as dynamical models in information processing. Thirdly we calculate how the NUF fares in different attractors as the control parameters change. We provide expressions which under specific circumstances should compliment the Lyapunov exponent(s) and the information dimension of the attractors involved.

* On leave of absence from the Department of Electrical Engineering, University of Patras, Patras, Greece.

Reprint requests to Prof. Dr. J. S. Nicolis, Department of Electrical Engineering, School of Engineering, University of Patras, Patras, Greece.

0340-4811 / 83 / 1100-1157 $ 01.3 0/0. – Please order a reprint rather than making your own copy.

2. The Break Down of the "Law of Large Numbers" in Pattern Recognition and Random Walks

I. Review of Some Experimental Results in Visual Pattern Perception [4, 5, 6]

What are the "crucial features" of a visual pattern and how is the optical (human) cortex (in cooperation with the sensor (fovea)[1] and the optical muscular apparatus) dealing with them? In a remarkable series of experiments conducted in the early seventies, Noton and his associates performed a number of investigations which are discussed below. To begin with just look at Figure 1. The left part represents the famous bust of Queen Nefertiti. At right is displayed the trajectory of the eye movement of a human observer (as recorded by A. L. Yarbus of the Institute for problems of Information Transmission in Moscow, USSR) – as it was scanning Nefertiti's head and neck in the process of perception. The experiments were performed in an attempt to settle the controversy of Gestalt (parallel, global) versus sequential (step by step – or iterative) pattern recognition. (An essential prerequisite for the meaningfulness of the attempted comparison is of course, that the pattern must be extended enough in space in order to allow for scanning by the sensor.) The result of the experiments heavily supports the hypothesis of serial, or piecemeal perception and recognition. Specifically two questions are answered, a) what are the features of the pattern that the optical cortex selects as the key items for identifying the object? and b) how are such features integrated and related to one another to form the complete internal representation of the object?

First of all we have good evidence that the optical system is an hierarchical one [5] and between the successive hierarchical levels mappings take place giving rise to feedforward-feedback loops. The higher levels (cortex) receive information from the lower levels and respond by sending commands to the sensor's muscular apparatus either to move in order to "phaselock" with the feature under investigation or to move away and assume the next algorithmic step of scanning the pattern. The experiments demonstrate that the above scanning

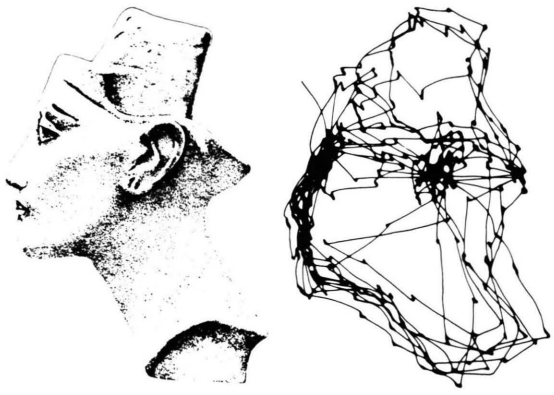

Fig. 1. Scanning Nefertiti's bust (after Noton and Stark, 1971).

algorithm is far from smooth and "homogeneous". There are parts on the pattern that hold for the processor the most information about the subject. The fixation or the "holding time" of the receptor tend to cluster around the parts characterized by sharp curves (small radius of curvature), angles, or in general around areas where the curve is "non-differentiable". It appears that the angles are the principal features the brain uses to store and recognize the pattern. (There is also neurophysiological justification for such a preference revealed through the painstaking experiments of Hubel and Wiesel; it appears that there are angle-detecting neurons in the frog's retina and angle-detecting neurons in the visual cortex of cats and monkeys. Recordings obtained from the human visual cortex by Marg [4] of the University of California at Berkeley give indications that this result can be extended to the human visual cortex as well.) The scanning of Fig. 1 bears witness to the heavy preference of the processor to areas of the face characterized by sharp curves. Such features are *complex* (they need "many bits" for their most laconic description and are therefore endowed with high information content). Dynamical analysis of the sequence of fixation of the receptor at the different states – "features" of the pattern suggests a format for the interconnection of these "states" into the overall internal representation. The result of this experiment reveals that the sensor directed by the brain is essentially involved with two types of "scanning pathways": Regular and irregular. Specifically it appears, as Noton and Stark [4] put, it that "(the) eyes usually scanned it (the pattern) following – intermittently

[1] The fovea constitutes a small central area of the retina and is characterized by the highest concentration of photoreceptors.

but repeatedly – a fixed path which we have termed his 'scan path'...". The occurrences of the scan path were separated by periods in which the fixations were ordered in a less regular manner... Scan paths usually occupied from 25 to 35 percent of the subject's viewing time, the rest being devoted to less regular eye movement." We may conclude then following Noton and Stark that the internal representation (or the mapping) of a pattern in the memory system is an assemblage of features (states) mediated essentially by a feedback loop: A sequence of senso-motor traces recording, abstracting and subsequently reporting to the brain a "Markov chain" as it were, of states-features – whereupon the brain directs the next move of the peripheral activity. The time intervals during which the system holds a given state has obviously to do with the excitation of the directing neuronal tissue – which in turn depends on the degree of curvature of the feature-state involved [7]. This Markov chain, this algorithim, is subsequently stored in the brain isomorphically as, say, a pattern of circulating electrical activity. When the observer is subsequently encountering the same pattern again he recognizes it by matching it with the "feature-ring" or the Markov chain which constitutes the internal representation in his memory, state-by-state. Matching then or recognizing consists in calculating the "distance" or the crosscorrelation between two Markov chains: The memorized one and the one which runs as the observer re-examines the pattern. Obviously the first chain directs the steps of the second through the brain-sensomotor loop and so learning takes place: Beyond a few reruns the object becomes "familiar" as the crosscorrelation above tends to one. It would be then, under such circumstances, rather improper to try to gather the information conveyed from observing the Nefertiti's bust by calculating the "entropy" of this pattern. By the way, this could be done as follows: You partition the two-dimensional pattern in Fig. 1 in $n(\varepsilon)$ squares of size ε through each of which the trace of the scanning path passes at least once. Let p_i be a number proportional to the relative frequency with which the scanner visits the specific square i. Then the "entropy" of the pattern could be calculated as

$$-\sum_{i=1}^{n(\varepsilon)} p_i \log_2 p_i \quad \text{bits} \tag{3}$$

and as the resolution ε becomes finer and finer one may be interested in calculating how the distribution $p_i(\varepsilon)$ scales with resolution. In that case one could obtain the "information dimension", or the bits required for the determination of a point of the scanning curve, as

$$D_1 = \lim_{\varepsilon \to 0} \left\{ \left(-\sum_{i=1}^{n(\varepsilon)} p_i \log_2 p_i \right) \Big/ \left| \log_2 \frac{1}{\varepsilon} \right| \right\}, \tag{4}$$

i.e. the asymptotic value of the slope of entropy versus resolution ($\| \log \varepsilon \|$).

It is implicitly assumed of course that the Nefertiti's bust constitutes indeed a "strangely attracting object" or, that the scanning path remains bounded in space and continuously iterating, on the basis of a dynamical control algorithm dictated by the brain centers – which centers "fire" more or less frequently depending of the curvature of the encountered last detail (state). In view of the shape of the scanning trace on the right side of Fig. 1 instead of relying on (4) one could obtain a more meaningful collective property or quantitative measure of the information gathered by investigating the standard deviation of D_1. However to make calculations feasible we should not start from formula (4) but rather from a relationship between the information dimensionality and the spectrum of Lyapunov-exponents of the flow or the discrete map, and calculate the standard deviation of D_1, ΔD_1, from the NUFs of the corresponding Lyapunov-exponents. The NUF of D_1 would indicate how much fuzziness enters in the degree of compressibility $(N - D_1)/N$ % of information realized by the attractor in N-dimensional space [2].

For example, in the case of the immune system – envisaged as an information processor – a tremendous variety of templates (antibodies) is put forth, either spontaneously or in response to stimulation by a single antigen; the "polyclonicity" of these templates may provoke an autoimmune effect; as a result the organism may suffer from such an excessive response of the processor more damage in terms of entropy production than from the initial intruder. We may say in such a case that the NUF

[2] In the present work we calculate the NUF for the dominant λ of maps and flows. The standard deviation of D_1 will be given in a forthcoming paper.

of D_1 far exceeds the compressibility factor $N - D_1$. In case of "monoclonal" antibodies on the other hand one should expect the NUF of D_1 to be just zero.

II. Information Transfer through Intermittent Channels or "Error Clustering" Media

Let us now change gears and refer to another seemingly unrelated example concerning again the break down of the "law of averages" in Markov chain sequences. The example has essentially to do with coin-tossing games or random walks and has been elevated into the prominence of a new scientific paradigm by Feller [8] and subsequently by Berger and Mandelbrot [9]. According to "widespread beliefs" (the characterization is Feller's) a so-called "law of averages" should ensure that in a long cointossing game each player will be in the winning side for about half the time and that "switching" will take place frequently from one player to the other. Figure 2, taken from Feller represents the result of a computer experiment simulating 10,000 tosses of a fair coin. The top line contains the graph of the first 550 trails; the next two lines represent the entire record. The surprising thing has to do of course with the length of intervals between successive crossings of the "zero"-axis. On a isomorphic basis you can imagine these graphs as standing for a one-dimensional random walk without barriers and ask how often the walker is likely to change sides. Because of the walk's symmetry one expects that in a long walk the man should spend about half of his time on each side of the starting spot. Exactly the opposite is true. Regardless of how long he walks the most probable number of changes from one side to the other is 0, the next probable 1 followed by 2, 3 and so on. If a man walks one step every second say for a year, in one out 20 repetitions of this experiment the walker could expect to go along one direction for more than 364 days and 10 h. (In such simulations one is not using coins of course but decides on the basis of the upcoming digits of to say, 100 decimals: An even digit is "0" and an odd is "1".) In conclusion then it is quite likely that in a long coin-tossing game one of the players remains practically the whole time on the winning side, the other on the loosing side. (In the experiments of Fig. 2 mentioned above, in 10,000 tosses of a perfect coin the lead is at one side for more than 9930 trials and at the other for fewer than 70 with probability greater than 10%.)

On the other hand one "should expect" that in a prolonged coin-tossing game the observed number of changes of lead should increase roughly in proportion to the duration of the game. This again is false.

Feller proves that the number of changes of lead in n trials increases only a $\sqrt{n}$; so in $100\,n$ trials one should expect only $10\,n$ times as many changes of lead as in n trails. Putting it in another way, if N_n is

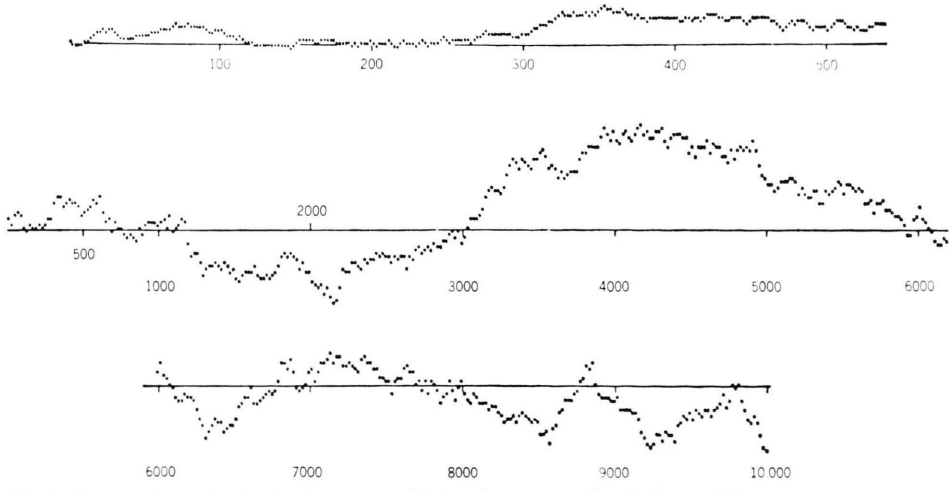

Fig. 2. Fluctuations of gains in the game of fair coin tosses (after Feller, 1970).

the number of changes in lead $N_n/n = \sqrt{n}/n = 1/\sqrt{n}$ goes to zero as the number of trails goes to infinity. So the waiting times between successive changeovers of the winner-looser roles between the contestants are likely to be fantastically long. What should be the pertinent probability density function for such a process? Imagine a huge sample of records of ideal coin tossing games each consisting of $2n$ tosses. We pick one at random and observe the number of the last trial at which the accumulated numbers of heads and tails were equal – that is the last changeover. This number is even and we denote it by $2K$; so that $0 \leqq K \leqq n$. Frequent changeovers (that is in the lead of one player) would imply that K is likely to be relatively close to n – but this is not so. Interestingly enough, Feller's calculation gives for the p.d.f. of the variable $x = K/n$ the expression

$$P(x) = 1/n\pi\sqrt{x(1-x)}\,. \tag{5}$$

The symmetry of the distribution implies that the inequalities $K \geqq n/2$ and $K \leqq n/2$ are equally likely. We see that the probabilities near the end points are greatest; the most probable values of K are then the extremes 0 and n, while the median value $K = n/2$ is least likely to occur. The above probability density function is identical (save the proportionality factor $1/n$) to the invariant measure of the logistic map $f(x) = 4rx(1-x)$ for $r = 1$, *i.e.* for the maximum value of the control parameter where the chaotic regime occupies the whole invariant interval and

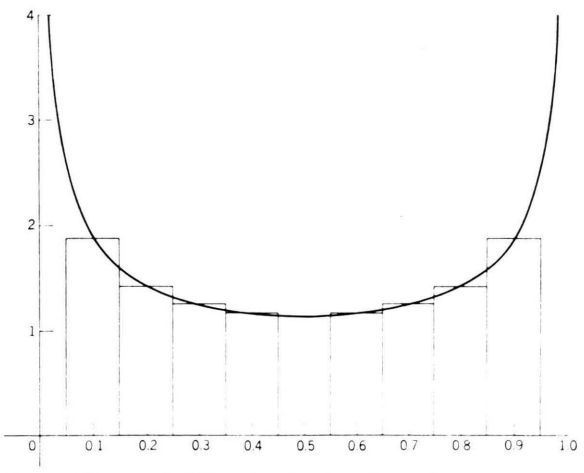

Fig. 3. The probability-density-function of $x = k/n$ (see text, after Feller, 1970).

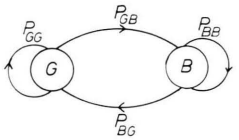

Fig. 4. The Markov-chain model for the burst error channel.

the map becomes exactly two-to-one. (See appendix.)

Let us now turn to the distribution of the occurrence of errors in data transmission on real communication channels, the simplest being the binary symmetric channel (Fig. 4) (Gilbert 1960 [9], Berger and Mandelbrot 1963 [8]). The channel is modelled as a Markov chain with two states G (Good = errorless) and B (Bad). The probability of error depends upon the state. State G is associated with zero probability of an error. In state B a (loaded) coin is tossed to decide whether an error will occur or not. The (biased) coin tossing feature is included because actual bursts or clusters do contain "good" digits interspersed with the errors. To stimulate burst errors, the states G and B must persist; hence the transition probabilities P_{GB} and P_{BG} should be small and the probabilities P_{GG} and P_{BB} should accordingly be large. Berger and Mandelbrot started by considering groups of three successive errors taking the time position t_{n-1} and t_{n+1} fixed and allowing t_n to hover in between; what is the probability density function of the inter-error intervals and the probability density of t_n? When the inter-error intervals are geometrically distributed (like in some cases assumed by the memoryless binary symmetric channel) the distribution of t_n is uniformly distributed between t_{n-1} and t_{n+1}. Berger and Mandelbrot, in order to bring the model as close as possible to the collected data, assumed for the inter-error intervals a Pareto distribution. In that case, for the probability density of t_n the theoretical type of curve best fitting the experimental results is illustrated in Figure 5. It says essentially that the probability of the occurrence of an error increases as we approach t_{n-1} or t_{n+1} symmetrically and becomes greatest in the close neighborhood of the occurrence of another error (the previous or the next). The probability of having an error in between is practically zero. Similarly in the case of a large number of errors a sizeable portion of the total sample length will be

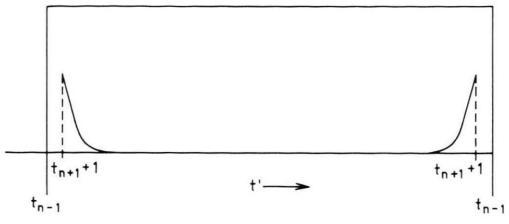

Fig. 5. The p.d.f. of the "intermediate error" in a burst noise channel (after Berger and Mandelbrot, 1963).

found in a few of the longest error intervals, say L of them, thus creating a pattern in which errors are mostly grouped in L clusters. Again the median value of the distribution is the most improbable one while the greatest probability refers to tight error clustering in time. The shape of the probability density function is again of the "hyperbolic" type we have already encountered twice above. The channel has his "Good" days and his "Bad" days – occurring in persistent clusters. This persistance and the scarcity of a possible changeover reminds one of the random walk business or the fair tossing again. In fact one might conjecture that the time axis within the transmission interval displays a fractal property that is the clustering is self-similar down to small scales. The channel spends most likely in the errorless regime, either too long, or too short a time. However the relative number of errors or the average number of digits in errors should be expected to tend to zero as the length of the message n increases to infinity. This conjecture comes for the same reasoning mentioned previously with reference to the average number of changes in the lead of one regime as time outstretches to infinity. It turns out then that as the length of the transmitted message goes to infinity despite the presence of an unbounded number of errors the channel capacity of the burst-error binary symmetric channel tends to one that is the figure one usually gets for the noiseless memoryless channel. Finally in their study of intermittent behavior near a tangent bifurcation of the logistic map, Hirsch, Huberman, and Scalapino [10] derive a similar probability density function for the path lengths that the orbit spends in the laminar (pseudo-limit cycle) regime or the chaotic regime – where it wanders aperiodically before getting a chance of being reinjected in the "channel" formed between the map and the bisector line – that is in the limit cycle regime. In other words in the intermittent regime the orbit spends, say within the limit cycle regime, either too little or too much time. Again the average value of the (symmetrical) distribution function is the least likely to occur.

3. Possible Role of Chaos in Information Processing

It appears therefore that in a broad domain of seemingly unconnected physical phenomena a chaotic dynamics or macroscopic "turbulent noise" prevails – in the absence of microscopic multi-dimensional noise – and in fact it is responsible for the break down of the "law of large numbers". (No wonder; chaos is manifested even with 3 macro-variables, coupled via quite deterministic non-linear ordinary differential equations.) Is there any compelling reason to believe that in biological organisms in general and man in particular chaos may play any role in information processing? Reliable information processing rests upon the existences of a "good" code (or map) or language: namely a set of recursive rules which generate variety at a given hierarchical level and subsequently compress it thereby revealing information at a higher level. To accomplish this a language – like good music – should strike at every moment an optimum ratio of variety (stochasticity) versus the ability to detect and correct errors (memory). Is there any dynamics available which might emulate this dual objective in state space? The answer is: In principle yes (Nicolis [11], [12]). We remind ourselves that there are two available dynamical ways of producing information: Either by cascading bifurcations giving rise to broken symmetry or via cascading iterations increasing resolution. The last way is simpler. A three-dimensional strange attractor for example creates variety along the direction of his positive Lyapunov exponent λ_+ and constrains variety (thereby revealing information) along the direction of his negative Lyapunov exponent λ_-, ($|\lambda_-| \geqq \lambda_+$). The issue then of relevance of strange attractors in information processing can first be debated on parsimonial grounds namely what possible evolutionary advantages may the chaotic mode bestow on an organism. Neurophysiological evidence must of course follow. Discussions on the above two topics have been couched by one of us (Nicolis [13], [14]) in a recent series of publications, where for example the thalamocortical pace-

maker of the (human) brain has been tentatively identified as a chaotic processor; we are not going to repeat them here. Enough of emphasizing that a chaotic processor even as an artifact ensures a very rich behavioral repertoire (software) with very simple hardware. What we intend to do in the remaining of this paper is to provide a general discussion on chaotic dynamics and report on a number of calculations on strange attractors (both chaotic maps and flows) which calculations justify, we think, mistrust in the intuitive reliance on the "median" values in characterizing the dynamics of these attractors; we thereby propose some more representative parameters. Let us remind ourselves that basic parameters of a strange attractor are a) his Lyapunov exponents determining the average amount of information produced ($\lambda \geqq 0$) or dissipated ($\lambda \leqq 0$) per iteration and b) the information dimension of the attractor D_1 which essentially determines the average value of the degree of compressibility ensured by the processor namely the average number of bits one needs to determine any point on the attractor after transients subside. For the three-dimensional Lorenz attractor for example, $D_1 = 2.06$ (bits) which means that if it is used as an information processing unit the Lorenz attractor can "save" $3 - 2.06 = 0.94$ bits for any point on its basin that is for any initial condition. Since compressibility of a time series is the necessary prerequisite for subsequent simulation of the dynamical phenomena conveyed by the series, the importance of strange attractors as information processors cannot be overlooked.

4. One-dimensional maps and Markov chains

In the preceding sections we have been referring to a category of dynamical processes – all of them characterized by a probability density function of the "hyperbolic" type; in such a case the median value $\int p(x)\, x\, dx$ is very small and the variance

$$\sigma^2 = \int p(x)\, x^2\, dx - [\int p(x)\, x\, dx]^2 \qquad (6)$$

is large. We now want to test the conjecture that perhaps the median value of the information distribution (that is the entropy) is, under specific circumstances (for example in the vicinity of intermittency) also inadequate in describing the information processing going on by the evolution of a strange attractor. Let us start then with examining the correspondance between one-dimensional maps and Markov chains. Let us consider the asymmetric two-to-one piece-wise linear map of Figure 6. Since matching an environmental (Markov) time series involves the creation of a great number of "templates" by the (chaotic) processor, it is instrumental to start in this simple case by calculating the elements of the Markov chain which represents the iterative process of the map as an information source. We devide the unit interval AE in two parts of length α and $1 - \alpha$, and every time the trace of the iterating trajectory falls on AC we get the symbol 1 while every time the trace falls on CE we get the symbol 0. So we obtain (for every value of the control parameter) strings of 001011011011011101... – which do not constitute, however, just fair coin tosses but rather onesided Bernouilli shifts that are constrained by the (non-linear) shape of the map $f(x)$. What are the transitional probability elements p_{ij} of the two-state Markov chain and what are the probabilities $u_1 = p(0)$ and $u_2 = p(1)$? From Fig. 6 it is clear that $l(\mathrm{AB}) = P_{11}$ since all points on the interval within AB which belongs to α are projected by the map in a portion equal to α. From the geometry we see that $P_{11} = \alpha^2$. Likewise we observe that $l(\mathrm{CD}) = P_{22}$ since all points within this sub-

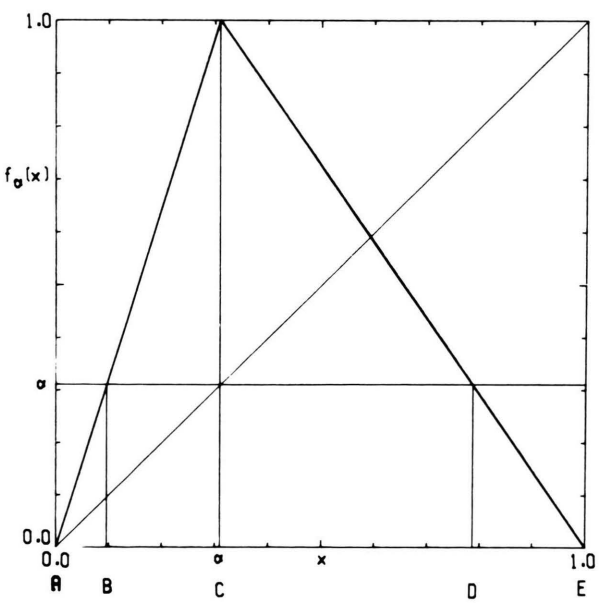

Fig. 6. Graph of $f_\alpha(x)$ for $\alpha = 0.31$ (see (21)).

interval which belong to $1-\alpha$ are projected by the map on the portion $1-\alpha$. $P_{12}=l(\mathrm{BC})$ since all the points within it, belonging to α are projected to $1-\alpha$, and $P_{21}=l(\mathrm{DE})$ since all points within it belonging to $1-\alpha$ are projected to α. From the geometry we get $P_{12}=\alpha-\alpha^2=\alpha(1-\alpha)$, $P_{21}=\alpha(1-\alpha)$ and $P_{22}=(1-\alpha)-\alpha(1-\alpha)=(1-\alpha)^2$. (The values above are not normalized; one should have $P_{11}+P_{12}=1$, $P_{21}+P_{22}=1$.) The probabilities u_1, u_2 are calculated from the relations $u_1=u_1P_{11}+u_2P_{21}$ and $u_1+u_2=1$. We get $u_1=\alpha$, $u_2=1-\alpha$.

We now intend to forward some general discussion on the nature of Lyapunov exponents in multi-dimensional flows and maps.

5. The Non-Uniformity Factor as a Further Characterization of Dynamical Systems

I. The Concept of the Local Divergence Rate

The spectrum of the Lyapunov-exponents constitutes a way to classify dynamical systems in general as well as particular solutions of a dynamical system. It is nowadays widely used in the literature [15], [16], [19].

In this section we want to interpret the maximal Lyapunov-exponent as the time average of a statistical variable we call the local divergence rate $Y(\boldsymbol{x}(t))$ (cf. [1], [29], [30]). Besides the average further information about the system is contained in the higher moments of the statistical variable like the variance. The variance of the local divergence rate $(\Delta Y)^2$ is a measure for the nonuniformity of the dynamical behavior of an attractor in respect to the separation of nearby trajectories. Besides the Lyapunov-exponents the variance $(\Delta Y)^2$ will give additional information about a dynamical system and may turn out to be helpful for further classification of chaotic systems.

In Part II of this section we derive the expression for the statistical variable, whose first moment will be the Lyanupov-exponent. We will discuss the meaning of the corresponding variance and define the Non-Uniformity Factor (NUF) ΔY in the context of dynamics. In a third part we will consider analytically and numerically different dynamical systems: 1-dimensional maps, a 2-dimensional map (Hénon-map) and a 3-dimensional flow (Rössler-system).

II. Average Value and Variance of the Local Divergence Rate

We start from a continuous dynamical system which is described by a set of coupled ordinary non-linear differential equations

$$\dot{\boldsymbol{x}}=\boldsymbol{F}(\boldsymbol{x})\,[1]. \tag{7}$$

The maximal Lyapunov exponent is defined as [15], [16], [17], [18]

$$\lambda=\lim\frac{1}{t}\ln\|\boldsymbol{u}(t)\|\,[2], \tag{8}$$

where $\boldsymbol{u}(t)$ is a solution of

$$\dot{\boldsymbol{u}}=\frac{\delta\boldsymbol{F}}{\delta\boldsymbol{x}}(\boldsymbol{x}(t))\,\boldsymbol{u}, \tag{9}$$

which is obtained by linearizing (7) along a trajectory $\boldsymbol{x}(t)$ of (7). Note that (9) is a linear system of differential equations with time dependent coefficients that describes the behavior of a perturbation of the trajectory $\boldsymbol{x}(t)$ of system (7). The solution of (9) to a given initial condition $\boldsymbol{e}(t=0)$ at time t can be written as

$$\boldsymbol{u}(t)=U_0^t(\boldsymbol{x}(0))\,\boldsymbol{e}(0). \tag{10}$$

$U_0^t(\boldsymbol{x}(0))$ is a time dependent matrix and is sometimes called fundamental solution matrix. The argument $\boldsymbol{x}(0)$ shall indicate that $U_0^t(\boldsymbol{x}(0))$ depends on the trajectory $\boldsymbol{x}(t)$ with initial condition $\boldsymbol{x}(0)$.

We will now write (10) and (8) in a form that is very helpful for the intuitive understanding of the Lyapunov-exponent and is more suitable for numerical computation of flows as well as for discrete maps [18]. To this end we formally discretize time into intervals of length τ, such that $t=n\tau$. A solution (10) to (9) can then be written as

$$\boldsymbol{u}(t)=U_{(n-1)\tau}^{n\tau}U_{(n-2)\tau}^{(n-1)\tau}\ldots U_{\tau}^{2\tau}U_0^{\tau}\,\boldsymbol{e}(0).$$

We take τ as the time unit and, changing the notation slightly, we obtain

$$\boldsymbol{u}(n)=U_{n-1}^{n}\,U_{n-2}^{n-1}\ldots U_1^2\,U_0^1\,\boldsymbol{e}(0). \tag{11}$$

Let us read (11) from right to left. $U_0^1\,\boldsymbol{e}(0)=\boldsymbol{u}(1)$ denotes the solution of (9) to the initial condition $\boldsymbol{e}(0)$ after time τ. It is a vector with length d_1 and

[1] Explicit time dependence can be eliminated by increasing the dimension of the system by one.

[2] Rigorously, lim should be replaced by limsup, see [28].

direction $\boldsymbol{e}(1)$:

$$U_0^1 \boldsymbol{e}(0) = d_1 \boldsymbol{e}(1), \quad \| \boldsymbol{e}(1) \| = 1.$$

Now we apply U_1^2 to $\boldsymbol{e}(1)$ and obtain another vector of length d_2 and direction $\boldsymbol{e}(2)$. Finally we obtain

$$\boldsymbol{u}(n) = d_n d_{n-1} \ldots d_2 d_1 \boldsymbol{e}(n), \quad \| \boldsymbol{e}(n) \| = 1. \quad (12)$$

We have changed a product of operators into a product of real numbers. The Lyapunov-exponent (8) now reads

$$\lambda = \lim_{n \to \infty} \frac{1}{n\tau} \ln \| d_n \ldots d_1 \boldsymbol{e}(n) \|$$

or

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\tau} \ln d_k. \quad (13)$$

From the definition of $d_k = \| U_{k-1}^k \boldsymbol{e}(k-1) \|$ we see that $\ln d_k$ is the exponential change of the length of $\boldsymbol{e}(0)$ during the time interval τ when the system (7) moves along the trajectory between $\boldsymbol{x}((k-1)\tau)$ and $\boldsymbol{x}(k\tau)$. Normalizing this quantity to one time-unit we define the *local divergence rate* for discrete or discretized systems

$$Y_k = \frac{1}{\tau} \ln d_k. \quad (14)$$

Equation (13) now suggests to interpret Y_k as a statistical variable. The Lyapunov-exponent λ is the time average (in time-units of τ) of the local divergence rate Y_k along the trajectory $\boldsymbol{x}(t)$:

$$\lambda = \langle Y_k \rangle. \quad (15)$$

Note that although (13) describes a discrete sampling of Y_k along a continuous flow it is still exact. $Y_k = 1/\tau \ln d_k$ is obtained by solving the continuous Eqs. (7) and (9). On the other hand (13) can also be used for discrete maps. Y_k is then obtained by iterating the map and its linearization.

With the interpretation of λ as the time average of the local divergence rate we define the nonuniformity factor (NUF) to a solution $\boldsymbol{x}(t)$ as the corresponding standard deviation

$$\Delta Y = \langle (Y_k - \langle Y_k \rangle)^2 \rangle^{1/2} = (\langle Y_k^2 \rangle - \langle Y_k \rangle^2)^{1/2} \quad (16)$$

or

$$(\Delta Y)^2 = \lim_{n \to \infty} \left(\frac{1}{n} \sum_{k=1}^{n} \left(\frac{1}{\tau} \ln d_k \right)^2 \right) - \lambda^2. \quad (17)$$

The NUF is a measure for the deviation of the local divergence rate from its mean value, the Lyapunov-exponent. It characterizes the nonuniformity of a solution of a dynamical system in respect to the sensitivity to initial conditions, or in other words, how much the local divergence rate changes along the flow. Chaotic flow, *e.g.*, with uniform turbulent behavior all over the attractor, yields a positive Lyapunov-exponent and a zero variance. Intermittent chaos, however, where regular laminar and chaotic phases alternate irregularly with each other, also gives a positive Lyapunov-exponent (possibly the same) but the standard deviation (NUF) should be large, since the local divergence rate varies strongly. So the NUF ΔY can be a very helpful quantity besides the Lyapunov-exponents to characterize the behavior of dynamical systems, especially chaotic ones. Its computation does not take more effort than that of the Lyapunov-exponents.

In the following section we consider some chaotic dynamical systems that are well known in the literature [16].

III. Analytical and Numerical Results from Specific Models

a) Maps on the Interval

For a discrete-time dynamical system, which is generated by a map f on the one-dimensional interval I, the Lyapunov-exponent λ as defined in (13) is given by

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \ln \| f'(x_k) \|, \quad (18)$$

where $x_k = f^k(x_0)$ and x_0 is some initial value in the basin of the attractor under consideration. In the case of onedimensional maps we can apply the ergodic hypothesis and replace the time average (18) by the ensemble average of the local divergence rate $Y(x)$ with respect to the probability-density-function $p(x)$ of f [19]. In our case the divergence-rate $Y(x)$ as defined in (14) takes the simple form

$$Y(x) = \ln \| f'(x) \|. \quad (19)$$

The Lyapunov-exponent λ therefore can be expressed as

$$\lambda = \langle Y(x) \rangle = \int_I p(x)\, Y(x)\, \mathrm{d}x. \quad (20)$$

Accordingly, we obtain the NUF as the standard-deviation ΔY of the statistical variable $Y(x)$ (see

(16)). The interpretation of the NUF is especially simple in this case: it just describes the variation of the modulus of the slope of the function f along the attractor. Thus, for the "uniform" maps $f(x) = 2x \bmod 1$ and the "tent"-map

$$f_{0.5}(x) = \begin{cases} 2x, & x \in [0, 0.5] \\ 2(1-x), & x \in [0.5, 1], \end{cases}$$

we have $Y(x) = \ln 2$ and $p(x) = 1$. This is why the NUF vanishes, *i.e.* the separation of trajectories happens uniformly in the whole interval.

We now wish to disturb this uniformity in two different ways by considering:

(i) piece-wise linear but asymmetric maps,
(ii) symmetric but nonlinear maps.

For perturbations of the type (i) we define the family f_α by

$$f_\alpha(x) = \begin{cases} \dfrac{x}{\alpha}, & x \in [0, \alpha], \\ \dfrac{1-x}{1-\alpha}, & x \in [\alpha, 1]. \end{cases} \tag{21}$$

The graph of $f_{0.32}$ is shown in Figure 6. For each α we have constant density $p(x) = 1$, and therefore we can calculate both the Lyapunov-exponent λ_α and the NUF ΔY_α analytically. The Lyapunov-exponent is given by

$$\lambda_\alpha = (\alpha - 1)\ln(1-\alpha) - \alpha \ln \alpha . \tag{22}$$

From

$$Y_\alpha^2(x) = \alpha (\ln \alpha)^2 + (1-\alpha)(\ln(1-\alpha))^2 \tag{23}$$

and

$$\lambda^2 = \alpha^2 (\ln \alpha)^2 + (1-\alpha)^2 (\ln(1-\alpha))^2 + 2\alpha(1-\alpha)\ln\alpha \ln(1-\alpha) \tag{24}$$

we deduce the NUF:

$$\Delta Y_\alpha = (\alpha(1-\alpha))^{1/2} \ln\left(\frac{1-\alpha}{\alpha}\right). \tag{25}$$

In Fig. 7 we have plotted the resulting Lyapunov-exponent λ_α (solid line) and the NUF ΔY_α (broken line) as a function of the asymmetry parameter α. Note that the NUF vanishes in the uniform cases, for which $\alpha \in \{0, \frac{1}{2}, 1\}$. It has two maxima which are given by the transcendental equation

$$\ln\left(\frac{1-\alpha}{\alpha}\right) = \frac{2}{1-2\alpha}. \tag{26}$$

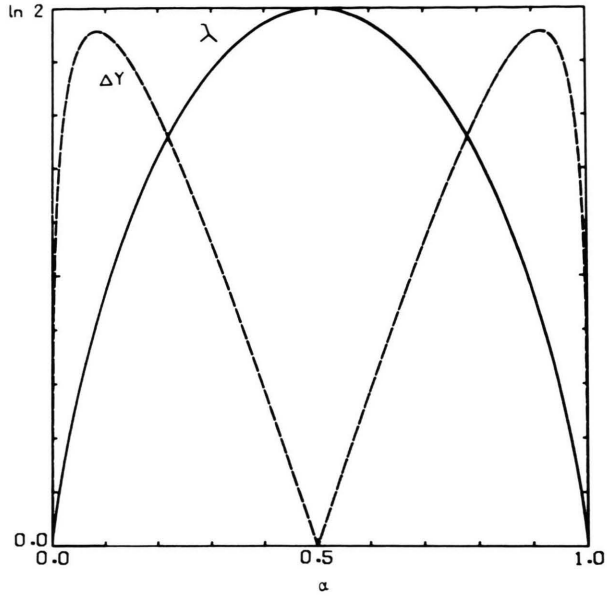

Fig. 7. Lyapunov exponents (λ) and NUF (ΔY) versus the asymmetry parameter α for $f_\alpha(x)$ of (21).

We can see from Fig. 7 that beyond some threshold value of the asymmetry-parameter α, the NUF exceeds the Lyapunov-exponent. From what we have mentioned above, it is clear that all the results also hold for the family of discontinuous functions $\tilde{f}_\alpha$ defined by

$$\tilde{f}_\alpha = (x/\alpha) \bmod 1 . \tag{27}$$

In order to study non-uniform systems of type (ii), we consider the family of nonlinear symmetric functions $g_{\gamma,r}(x)$, defined by [20]:

$$g_{\gamma,r}(x) = r(1 - \| 1 - 2x \|^\gamma) . \tag{28}$$

In Fig. 8 the graph of the "tent"-map $g_{1,1}(x) = f_{0.5}(x)$ (solid line), of the "Lorenz"-map $g_{0.5,1}(x)$ (broken line), and of the "logistic" map $g_{2,1}(x)$ (broken dotted line) are presented. The Lyapunov-exponent λ_γ and the NUF have been numerically determined for various values of the exponent γ and fixed parameter $r = 0.99999$. The results are shown in Fig. 9, where λ_γ (solid line) and ΔY_γ (broken line) are plotted versus the exponent γ. Note that λ_γ is constant except in a neighborhood of $\gamma = 0.5$, where intermittent chaos is observed, which gives rise to the vanishing Lyapunov-exponent [10], [22], [23]. Again we find $\Delta Y_\gamma = 0$ for the uniform case $\gamma = 1$ and two crossings of the λ_γ- and ΔY_γ-curves close to

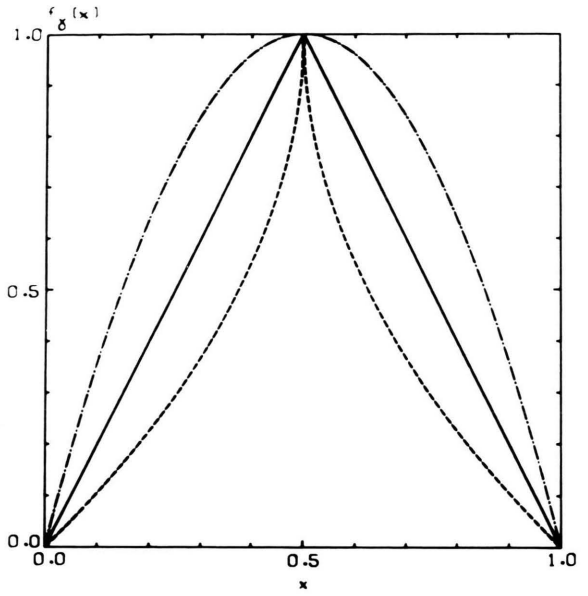

Fig. 8. Graph of $g_{\gamma,r}(x)$ for $r=1$ and $\gamma=\frac{1}{2}$ (---), $\gamma=1$ (——), $\gamma=2$ (–·–).

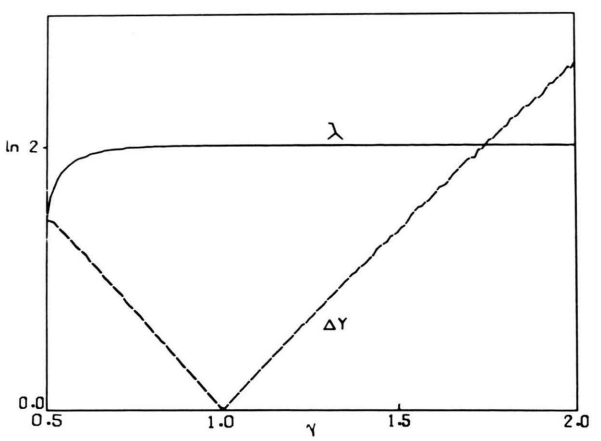

Fig. 9. Lyapunov exponent (λ) and NUF (ΔY) of $g_{\gamma,r}(x)$ vs. γ for 150 values of γ and $r=1.0$. The numerical results of Figs. 9 and 10 were obtained by averaging over 5×10^4 iterations with an accuracy of 16 decimal digits.

the "Lorenz"- and the "logistic" functions. An interesting feature in the intermittency case $\gamma \geqq 0.5$ is the singularity of the relative standard-deviation $\Delta Y_\gamma/\lambda_\gamma$, which indicates the strong non-uniformity, stemming from the alternation of long laminar phases and chaotic bursts. This observation is confirmed in Fig. 10, where we have plotted λ (solid line) and ΔY (broken line) as a function of the height-parameter r in the "logistic" system $g_{2,r}(x)$. At each value of r at which the system exhibits intermittent chaos, *i.e.* where λ_r drops from positive to negative values [19], [21], [10], [23], we find large values of the NUF, just like in the case of the "Lorenz"-map.

We have also performed numerical simulations at parameters r at which the system is fully chaotic and where we have perturbed the system by additive fluctuations. Our results indicate that the NUF does not change very much (a few percent), when the noise-level is increased to its maximal value. This insensitivity against external noise has also been observed for the Lyapunov-exponent [23], [24], although the probability-density-function becomes more and more flat with increasing noise-level.

b) The Hénon-Map

We have used (13) and (17) to compute the maximal Lyapunov-exponent and the NUF for the Hénon-map [25], [16]

$$x_{n+1} = y_n + 1 - a x_n^2,$$

$$y_{n+1} = b x_n .$$

We picked the usual parameter choice $b=0.3$ and varied the parameter a from 0.0 to 1.42. Figure 11 shows the largest Lyapunov-exponent (lower curve) and the corresponding NUF (upper curve). Figure 12 shows a magnification of the chaotic regime of Fig. 11. In the case of period 1 the NUF, of course, is zero up to the bifurcation point to

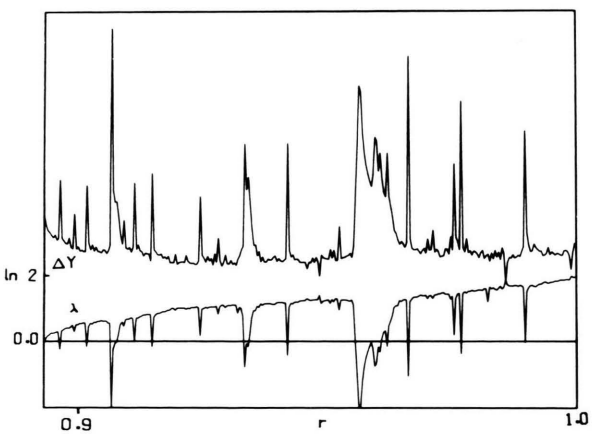

Fig. 10. Lyapunov exponent (λ) and NUF (ΔY) of the "logistic system" $g_{2,r}(x)$ vs. parameter r for 300 values of r.

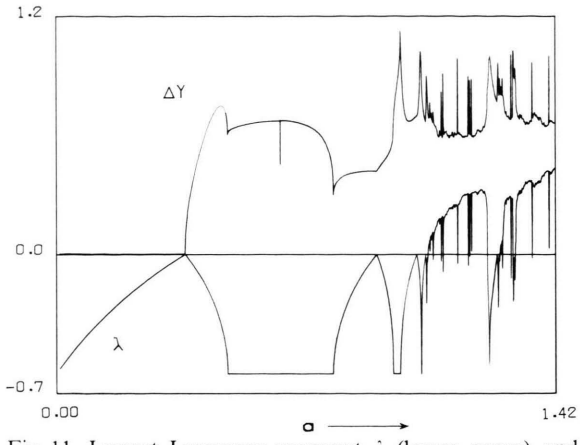

Fig. 11. Largest Lyapunov-exponent λ (lower curve) and corresponding NUF ΔY (upper curve) of the Hénon-system vs. parameter a for $b = 0.3$. Computations were performed with an accuracy of 16 decimal digits.

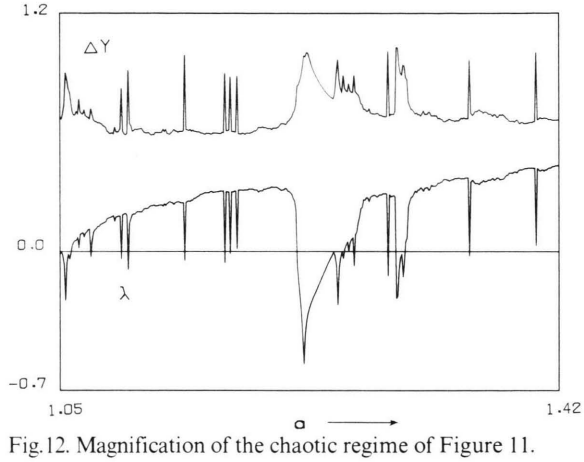

Fig. 12. Magnification of the chaotic regime of Figure 11.

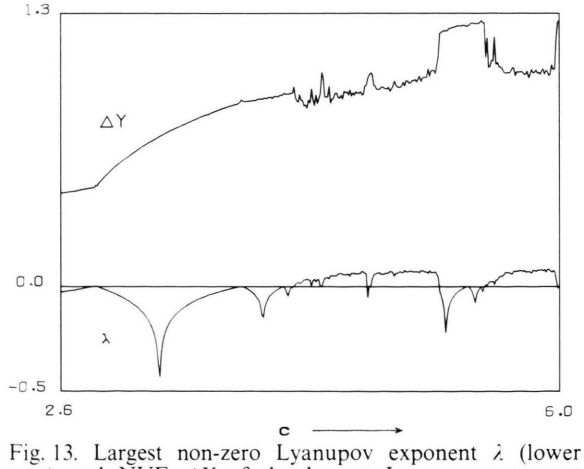

Fig. 13. Largest non-zero Lyanupov exponent λ (lower part) and NUF ΔY of the largest Lyapunov-exponent (upper part) of the Rössler-attractor vs. parameter c for $a = b = 0.2$, $\tau = 0.1$.

period 2. In the period 2 regime λ and the NUF display interesting features: Between $a = 0.49$ and $a = 0.79$ the two Lyapunov-exponents are equal and constant $\lambda_1 = \lambda_2 = \frac{1}{2} \ln b$. In contrast to the Lyapunov-exponents the NUF shows more structure in this parameter range. There are three spikes which correspond to special values of the complex eigenvalues of the fundamental solution matrix (12), whose logarithms of the modulus are the Lyapunov-exponents. At $a = 0.49$ the eigenvalues change from purely real to complex conjugate and start moving on a circle in the complex plane with radius $\frac{1}{2} \ln b$. At $a = 0.64$ the eigenvalues become purely imaginary and for $a = 0.79$ they "leave" the circle and become real again. A similar behavior occurs at higher period doublings.

So the NUF can given additional information about a dynamical system and can display features which are undetectible by the Lyapunov-exponents alone, e.g. changes in the imaginary parts of the eigenvalues.

c) The Rössler-Attractor

The Rössler-system is one of the simplest examples for chaotic 3-dimensional flows [26], [16]

$$\dot{X} = -(Y + Z), \; \dot{Y} = X + aY, \; \dot{Z} = b + Z(X - c).$$

We choose $a = b = 0.2$ and varied c from 2.6 to 6.0 [27]. The results of the computations are shown in Figure 13. Again the upper curve denotes the standard deviation of the local divergence rate (NUF) while the lower curve indicates the values of the mean value (Lyapunov-exponent). For the numerical computation of the NUF we choose $\tau = 0.1$ (14). Further refinements of the discretization step width τ yield the same results indicating that the computed NUF of the discretized system approached the NUF of the "actual" continuous system.

Acknowledgements

The authors thank Professor H. Haken for his interest and support which made the present work possible. One of us (J.S.N.) expresses also his gratitude to Prof. Haken for his generous hospitality in his institute while J.S.N. served as a Gast-professor for the academic year 1982–83. We also thank G. Nicolis and R. Kapral for stimulating discussions.

Appendix:

The Hyperbolic p. d. f. for the Logistic Map

We give here the derivation of the ("hyperbolic") probability density function for the logistic map $f(x) = 4rx(1-x)$ – when the control parameter r takes its maximum possible value $r = 1$, that is when the whole interval (0, 1) becomes a chaotic invariant set.

Let us consider first the symmetric "tent" piecewise linear map ($\alpha = \frac{1}{2}$) for a height $h < 1$. The slope of the map is $2h$ for the left half and $-2h$ for the right half. The orbit is always chaotic for $h > \frac{1}{2}$ since all trajectories diverge exponentially.

The p. d. f. $p(x)$ of this tent map is calculated as follows: The number of trajectories $p(x)\,\mathrm{d}x$ within the interval $\mathrm{d}x$ equals the number of trajectories within the corresponding intervals at the inverse points of the mapping. Since there are two inverse points x_1, x_2 for a one-hump map

$$p(x)\,\mathrm{d}x = p(x_1)\,\mathrm{d}x_1 + P(x_2)\,\mathrm{d}x_2\,,$$

where

$$\frac{\mathrm{d}x}{\mathrm{d}x_i} = \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_i} = 2h \quad (i = 1, 2)\,.$$

So

$$p(x) = \frac{1}{2h}\left\{p\left(\frac{x}{2h}\right) + p\left(1 - \frac{x}{2h}\right)\right\}\,.$$

For $h = 1$ (two-to-one tent map), the above relation gives $p(x) = 1$. Now consider the transformation $x' = (2/\pi)\sin^{-1}\sqrt{x}$ under which the logistic map is transformed to the symmetric tent map $f_{0.5}$ of (21) with $h = 1$ and $p'(x') = 1$. Since $p'(x')\,\mathrm{d}x' = p(x)\,\mathrm{d}x$ (which implies the conservation the number of trajectories within any small interval under any invertible transformation $x' = g(x)$) we get for the p. d. f. of the logistic map

$$p(x) = \frac{\mathrm{d}x'}{\mathrm{d}x} = 1/\pi\sqrt{x(1-x)}\,,$$

and the Lyapunov exponent for $r = 1$ equals

$$\lambda = \frac{1}{\pi}\int_0^1 \log_2 \| 4(1-2x) \| /\sqrt{x(1-x)}\,\mathrm{d}x = 1 \text{ bit}\,.$$

[1] J. S. Nicolis, G. Mayer-Kress, G. Haubs, in: Proc. of UNESCO-Conf. on "Stochastic Phenomena and Chaotic Attractors in Complex Systems", P. Schuster (ed.), Flattnitz, June 1983.
[2] H. Haken, Synergetics-An Introduction, 2nd ed., Springer-Verlag, Berlin 1978.
[3] G. Nicolis and I. Prigogine, Self-Organization in Non-Equilibrium Structures, Wiley, New York 1977.
[4] D. Noton and L. Stark, Sci. Amer. **224,** 34 (1971), June.
[5] D. Noton, IEEE Trans. on Systems Science and Cybernetics, SSC-6 **4,** 349 (1970).
[6] D. Noton and L. Stark, Science **171,** 308 (1971).
[7] W. Feller, An Introduction to Probability Theory and Applications, Third ed., Vol. 1, Wiley, New York 1970.
[8] J. M. Berger and B. Mandelbrot, IBM Journal **1963,** July, p. 224.
[9] E. N. Gilbert, The Bell Syst. Techn. J. **1960,** Sept., p. 1253.
[10] J. E. Hirsch, B. A. Huberman, and D. J. Scalapino, Phys. Rev. A **25,** 519 (1982).
[11] J. S. Nicolis, The Role of Chaos in Reliable Information Processing, J. Franklin Inst., in press 1983, also in "Synergetics" (Ed.: H. Haken), Proceedings of the Elmau Conference, May 1983.
[12] J. S. Nicolis, Kybernetes **11,** 269 (1982).
[13] J. S. Nicolis, Kybernetes **11,** 123 (1982).
[14] J. S. Nicolis, Dynamics of Hierarchical Systems, to appear in Springer Series of Synergetics. H. Haken (ed.), 1983.
[15] H. Haken, Advanced Synergetics, to appear in Springer-Verlag, Berlin.
[16] A. J. Lichtenberg and M. A. Liebermann, Regular and Stochastic Motion, Springer-Verlag, Berlin 1983.
[17] I. Shimada and T. Nagashima, Progr. Theor. Phys. **61,** 6 (1979).
[18] G. Benettin, L. Galgani, A. Giorgilli, and J. M. Strelcyn, C. R. Acad. Sci. Paris **286,** A-431 (1978).
[19] R. S. Shaw, Z. Naturforsch. **36 a,** 80 (1981).
[20] J. P. Crutchfield and N. Packard, Int. J. Theor. Phys. **21,** 433 (1982).
[21] G. Mayer-Kress and H. Haken, Phys. Lett. **82 A,** 151 (1981).
[22] J. P. Eckmann, L. Thomas, and P. Wittwer, J. Phys. **A 14,** 3153 (1982).
[23] G. Mayer-Kress and H. Haken, J. Stat. Phys. **26,** 149 (1981).
[24] D. Farmer, J. Crutchfield, and B. Huberman, Phys. Rep. **92,** 45 (1982).
[25] M. Henon, Commun. Math. Phys. **50,** 69 (1976).
[26] O. E. Rössler, Phys. Lett. **57 A,** 397 (1976).
[27] J. Crutchfield, D. Farmer, N. Packard, R. Show, G. Jones, and R. J. Donnelly, Phys. Lett. **76 A,** 1 (1980).
[28] H. Haken, Phys. Lett. **94 A,** 2 (1983).
[29] P. Grassberger, I. Procaccia, Dimensions and Entropies of Strange Attractors from a Fluctuating Dynamics Approach, Preprint Wuppertal 1983.
[30] G. Benettin, Mean-Field Approximation for Lyapunov-Exponents in some Conservative Dynamical Systems, Preprint 1983.