

On Complexity, Grammars and the Inferring of Models

Peter Wagner

Naturwissenschaftliches und Medizinisches Institut (NMI), Reutlingen, FRG

Z. Naturforsch. **48a**, 651–654 (1993); received March 23, 1993

Various measures of the physical complexity of a dynamical system are considered. Without question, these measures introduce a new quality in the description of nonlinear complex systems, but we think that their advantages have not yet been demonstrated in a convincing manner. We show that the estimation of the complexity of a given system is related to the extraction of models from real-world data.

1. Complexity

A theory concerned with complex systems certainly needs a quantitative measure of complexity in order to obtain a description and a classification of complex systems which is more than qualitative. Currently, there are a number of measures of complexity proposed in the literature [1–4], to name only a few of them. Let us shortly point out two features which such a measure should fulfil, a more detailed discussion can be found in [1, 3]:

- It should be zero in case of total order or total disorder, and nonzero in between.
- The numerical value of the complexity depends on the observer, or more exactly on the method used to describe the system.

The second point can be made more transparent by way of an example [3]. Consider the logistic map at the accumulation point of the period doubling cascade. The generating partition converts any sequence of iterations $\{x_n\}$ into a string $\{s_n\}$:

$$x_{n+1} = r x_n (1 - x_n), \quad s_{n+1} = \begin{cases} 1 & \text{if } x_{n+1} > \frac{1}{2}, \\ 0 & \text{else.} \end{cases}$$

If one tries to analyse such a sequence, one could think of it as generated by a Markov-process of (possibly) high order. This is exactly what most of the algorithms to extract various measures of complexity try [1, 4] when they construct the minimal automaton required to generate this string. This leads to diverging values of the complexity [1, 4] at the Feigenbaum-point. However, it is well known that the string generated by logistic map at this point can be described by

Reprint requests to Dr. P. Wagner, Naturwissenschaftliches und Medizinisches Institut (NMI), Gustav-Werner-Straße 3, W-7410 Reutlingen, FRG.

two rules only, $0 \rightarrow 11$ and $1 \rightarrow 01$, which define a parallel (DOL) grammar.

Therefore we have found a very simple characterization of this system, and if we interpret complexity as the length of the description of the generating rules (the minimal program) for the system under consideration, then the complexity is very small for the dynamics at the Feigenbaum-point. One sees that reducing complexity is a way of “finding the simplest description consistent with the data”. Of course, even if a system has a short description, it may be hard to find it, as we will see in Section 4. Additionally one may ask the question whether it is possible to reduce the description of a given system always to a mere triviality, i.e. to zero complexity or not.

2. Grammars, Regular Languages and Markov-Processes

Let us now turn to the problem of inferring a grammar from given data. The problem, e.g. for the strings generated by the logistic map, consists in learning a grammar from positive examples only, as it is quoted in the literature about formal languages. This problem is NP-hard in the strict formulation [5], so one has to think of restrictions in generality, which allows one to extract approximations of the grammar underlying the system under consideration.

Superfluous to point out that an algorithm for inferring the grammar describing given data has strong theoretical and practical importance. It could be used for data-reduction, for the automatic extraction of models and for making optimal predictions.

Up to now, only the inference of a regular language is well understood, a topic we will from now concentrate on. Finding the grammars characterizing the sys-

0932-0784 / 93 / 0500-0651 \$ 01.30/0. – Please order a reprint rather than making your own copy.



Dieses Werk wurde im Jahr 2013 vom Verlag Zeitschrift für Naturforschung in Zusammenarbeit mit der Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. digitalisiert und unter folgender Lizenz veröffentlicht: Creative Commons Namensnennung-Keine Bearbeitung 3.0 Deutschland Lizenz.

Zum 01.01.2015 ist eine Anpassung der Lizenzbedingungen (Entfall der Creative Commons Lizenzbedingung „Keine Bearbeitung“) beabsichtigt, um eine Nachnutzung auch im Rahmen zukünftiger wissenschaftlicher Nutzungsformen zu ermöglichen.

This work has been digitalized and published in 2013 by Verlag Zeitschrift für Naturforschung in cooperation with the Max Planck Society for the Advancement of Science under a Creative Commons Attribution-NoDerivs 3.0 Germany License.

On 01.01.2015 it is planned to change the License Conditions (the removal of the Creative Commons License condition “no derivative works”). This is to allow reuse in the area of future scientific usage.

tem under consideration is equivalent to the extraction of a minimal automaton (there exists a correspondence between a grammar and an automaton which understands this grammar [6]).

In physical terms, the identification of an automaton from given data corresponds to the rewriting of a Markov-chain of m -th order into a Markov-chain of first order, but with more states (i.e. different symbols) than the original time-series. In principle this is always possible, similar to the rewriting of an m -th order differential-equation to a set of first order ones. More exactly, one can show that to any Markov-chain there exists a related (minimal) automaton, but not vice versa. We restrict ourselves in the following to the complexity measures related with Markov-chains and therefore with the automata, and mention in passing that there exist measures of complexity which do not use the assumption of Markov-chains [3].

3. Measures of Complexity

Mutual Information

Given a Markov-process of first order, in [2] the following measure of complexity is used

$$^pC(t) = \sum_{\{S_t, S_{t-1}\}} p(S_t, S_{t-1}) \log \frac{p(S_t, S_{t-1})}{p(S_t) p(S_{t-1})}, \quad (1)$$

which is the mutual information between two successive states of the system. Here S_t denotes the state of the system at time t , and $p(S_t)$ is the probability of finding the system in this state. This measure is also applicable to non-stationary systems (see [2] for an example), which is not true for the following measures of complexity, but it has the disadvantage of requiring the definition of an appropriate ensemble.

In case of a nonlinear system with continuous observables and without noise, $^pC(t)$ reduces to the information-dimension [7]:

$$^pC(t) \rightarrow C_1(\varepsilon) = -D_1 \log \varepsilon,$$

where ε is the phase-space resolution.

Restricting in the following to 0/1-strings of observables, we shortly discuss two further measures of complexity:

Effective Measure Complexity (EMC) [1]

Given a string of length n , it measures the information one needs to make an optimal prediction of the

next symbol. From the Shannon entropies of order n , $H_n = - \sum_{\{s_1 \dots s_n\}} p(s_1 \dots s_n) \log p(s_1 \dots s_n)$ one calculates the block entropies of order n , which are approximations to the metric entropy of a dynamical system, $h_n = H_{n+1} - H_n$, where $H_0 = 0$. With the metric entropy $h := \lim_{n \rightarrow \infty} h_n$, the EMC is given by

$$\text{EMC} = \sum_{n=0}^{\infty} (h_n - h). \quad (2)$$

If one has found the minimal automaton, it is possible to calculate also the

Set Complexity (SC) [1, 4]

It determines the effort to affirm that a given string is a member of the grammar represented by the automaton. If one knows already that the string $s_1 s_2 \dots s_n$ is member of the grammar, the SC measures the amount of information to determine whether $s_1 s_2 \dots s_n s_{n+1}$ is also in the grammar. To answer this question, one must know the actual position in the automaton after parsing $s_1 s_2 \dots s_n$, which leads to

$$\text{SC} = - \sum_{\text{nodes}} p(i) \log p(i), \quad (3)$$

where $p(i)$ is the probability that the automaton is in node i .

In the following section we will show by way of some examples how these measures work.

4. Applications

We have used the procedure sketched above to investigate the following problem. In [4] it is conjectured that there exists a kind of phase transition in complex systems at the transition from order to disorder or to chaos. In Fig. 1 we have plotted the complexity (we have used the EMC, which can be calculated with less effort than the SC) against the Kolmogorov entropy h for a number of generalized shift maps (GS-maps) [8]. A similar behaviour for the logistic map is reported in [4], as well as in [9] for a certain class of cellular automata (CA).

It turns out that the location of the maximum in Fig. 1 depends on the measure of disorder used. If one uses instead of h the normalized Shannon entropy $H(n)/n$ as in [4], the location of the maximum is no longer at $h=0$. We think that the use of $H(n)/n$ is not the best choice, because the location of the maximum

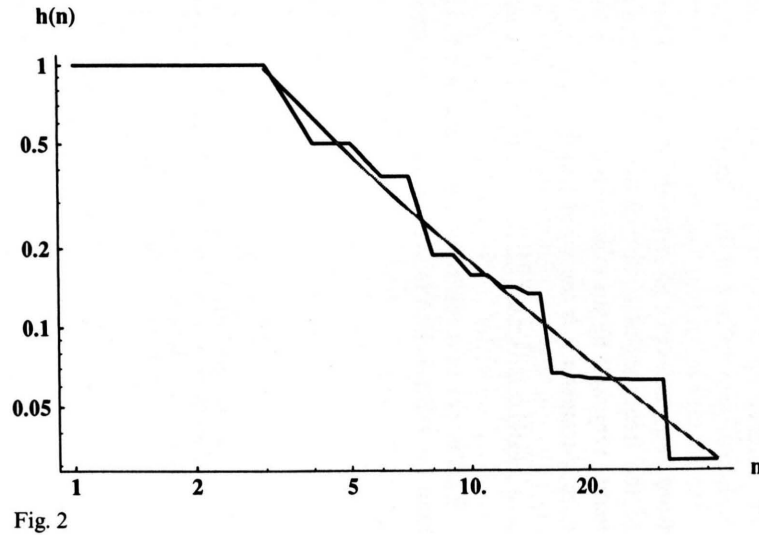
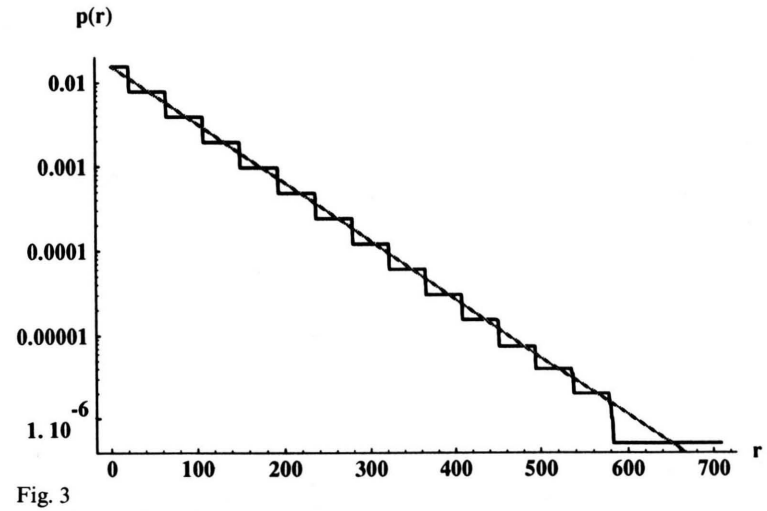
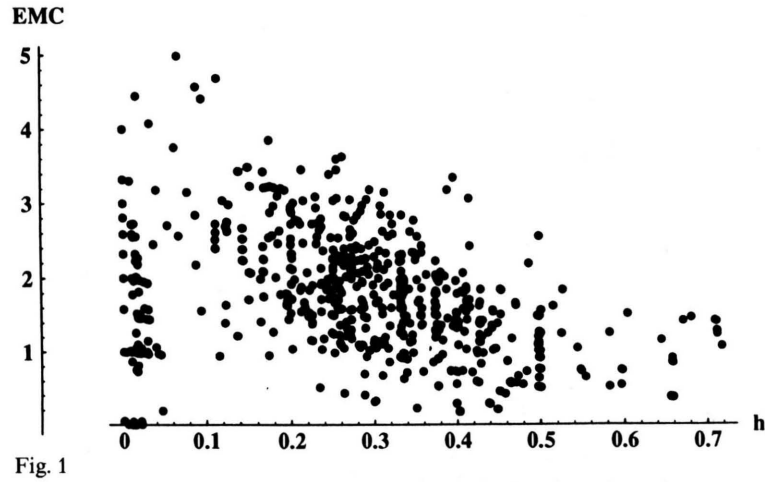


Fig. 1. Plot of EMC, estimated to $m=16$ th order, versus an estimate to the Kolmogorov entropy h for different GS-maps. The rules determining the behaviour of each GS-map are generated randomly.

Fig. 2. Plot of $h(n)$ versus word-length n for a certain GS-map with a large value of EMC. For comparison, also a curve (grey) according to (6) is drawn, which stems from a least squares fit of (6) to the numerically calculated $h(n)$, and which shows the same scaling behaviour.

Fig. 3. Plot of the ranked word-distribution for words of length $n=42$ for the GS-map mentioned in the text, together with a theoretical curve (grey) according to (5), where q is in this case $q=(158 \pm 1) \cdot 10^{-4}$.

of the complexity depends on the order n for which the Shannon entropy is determined. The fact that in Fig. 1 the maximum is at $h=0$ indicates that complex behaviour may exist only at the “border of chaos” [9].

It is, however, not clear whether one could call such a behaviour a phase-transition, but we find it interesting that three totally different systems show such a similar behaviour. But one should be aware of, as pointed out above, that the kind of inferring procedure one uses influences the complexity one gets.

Some other important points related to the characterization of complex systems can be found in [10]. There it is discussed that an interesting measure is the decrease of the block-entropy $h_n = H_{n+1} - H_n$ with the word-length n . In [10] it is proposed that the general form of this function can be parametrized by

$$h(n) = h_\infty + g n^\mu (\log(n))^\nu + e \frac{1}{n}. \quad (4)$$

In Fig. 2 we have calculated h_n for a certain GS-map with high complexity. For this map, it turns out that the number of words grows with the square of the word-length n . Therefore it is impossible to determine the block-entropies for very long words numerically, because the number of occurrences which one needs for calculating the word-probabilities decreases rapidly to zero. In case of this map one may get rid of these limitations. In Fig. 3 we have plotted the ranked word frequencies of a given length. One sees the shape of the frequency-rank distribution can be approximated very well by an exponential function (r is the rank, parameter n is the word-length):

$$p_n(r) = q_n (1 - q_n)^{r-1}. \quad (5)$$

This is not new, distributions of this kind can be found in typical chaotic systems [11]. Unfortunately,

we were not successful in determining the generating rules underlying this system despite our feeling that there may exist a set of rules determining the grammar of these map.

However, we are interested only in the scaling of $h(n)$ for large n . Estimating the parameter q_n of the exp-distribution as shown in Fig. 2 as function of length, one finds that it scales numerically as

$$q_n \propto n^{-\gamma},$$

where $\gamma = 0.92 \pm 0.03$. This can be used to get the following result for the block-entropy:

$$h(n) \propto n^{-1-\gamma} + n^{-1}, \quad (6)$$

which gives the theoretical curve also drawn in Figure 3. As mentioned already above, we are not interested in the very details of this function, because the procedure used to find this result is very insensitive to the details of the $p_n(r)$ -distribution. The results above apply to all cases where one finds a ranked probability distribution of exponential type, provided the parameter q_n scales similar. Needless to say that care is needed, because in complex systems one has to be aware of surprises – may be that at very large word length the behaviour of the system changes dramatically, making the procedure sketched above senseless.

Acknowledgement

This article is a concise summary of a talk given in the “3rd Annual Meeting of ENGADYN – Workshop on Nonlinearities, Dynamics, and Fractals, Grenoble ’92”. I would also like to thank R. Günther and A. Hoff for many stimulating discussions and for the patience necessary to accompany the process of understanding the topics presented here.

- [1] P. Grassberger, Intern. J. Theor. Phys. **25**, 907 (1986).
- [2] R. Günther, B. Schapiro, and P. Wagner, Intern. J. Theor. Phys. **31**, 600 (1992).
- [3] R. Badii, M. Finardi, and G. Broggi, Unfolding Complexity and Modelling Asymptotic Scaling Behaviour, in: Chaos, Order and Patterns (P. Cvitanovic, ed.), Plenum Press, New York 1991.
- [4] J. P. Crutchfield and K. Young, Phys. Rev. Lett. **63**, 105 (1986).
- [5] E. M. Gold, Inform. and Control **10**, 447 (1967).
- [6] J. E. Hopcroft and J. D. Ullman, Introduction to Automata Theory, Languages and Computation, Addison-Wesley, Reading, MA 1979.
- [7] K. Pawelzik, Nichtlineare Dynamik und Hirnaktivität, Dissertation, Verlag Harry Deutsch, Frankfurt 1991.
- [8] C. Moore, Phys. Rev. Lett. **64**, 2354 (1990).
- [9] C. Langton, Physica D **42**, 12 (1990).
- [10] W. Ebeling and G. Nicolis, Europhys. Lett. **14**, 191 (1991).
- [11] G. Nicolis and C. Nicolis, Physica A **163**, 215 (1990).