Information Flow in 1D Maps

K. J. G. Kruscha and B. Pompe

Sektion Physik/Elektronik der Ernst-Moritz-Arndt-Universität, Greifswald, GDR

Z. Naturforsch. **43a**, 93–104 (1988); received April 6, 1987

An information theoretical description is given of the action of 1D maps on probability measures (e.g. on ergodic invariant measures of chaotic maps). On the basis of a detailed analysis of the elements of information flow the problem of optimum measuring of initial states for state predictions is discussed. Moreover, we give an information theoretical description of the relaxation, under the action of a map, of an initial probability distribution to any, not necessarily steady, final distribution. In this connection we formulate an H-theorem for 1D maps.

1. Introduction

There have been rather intensive efforts in recent years to discuss the properties of chaotic dynamical systems with information theoretical methods (see e.g. [1–4]). Some essential quantities characterizing chaotic motions, e.g. the Kolmogorov-Sinai-Invariant, Lyapunov exponents, and several dimension like quantities, are derived from information resp. ergodic theory or have an information theoretical interpretation. Moreover, the predictability problem in such important fields as wheather forecasting, oceanology, engineering, computational modelling etc. has received increasing attention [5, 6] and has stimulated investigations of the information flow in dynamical systems.

The present paper represents an information theoretical approach to deterministic one-dimensional dynamics described by maps of 1D interval on itself. They represent the simplest dynamical systems that can display chaotic behaviour and thus are not quite trivial. More dimensional time-continuous or -discrete systems can sometimes be reduced to a 1D map. (For some more facts supporting the relevancy of the study of 1D maps see e.g. [7].)

In Sect. 2 and 3 we formulate a general scheme of information flow for probability measures that are not necessarily invariant under the action of the map, which allows us to study also relaxation processes.

In our considerations of information flow the so-called initial state information loss due to folding plays the central role. This is the inevitable minimum part of the information obtained in an optimum measurement of an initial state which has no predictive power. The Lyapunov exponent is related to the i.s.-information loss. Some interesting properties of the i.s.-information loss, like e.g. invariance under conjugation and additivity, are explained in Section 4. In Sect. 5 we discuss a relation of the i.s.-information loss to the so-called L-function, which describes the velocity of the relaxation of a non-stationary initial probability density with time under the action of the map. In this connection we formulate an analogon to Boltzmann's H-theorem for 1D maps.

In Sect. 6 we discuss two concepts of measuring the objective information flow, using several partitioning concepts that are adapted to the map. Such adapted partitions are of special interest: On the one hand they enable us to outline the problem of measuring an initial state with a maximum of predictive power on a future state and a minimum of experimental effort (i.e., the problem of optimum measuring for state predictions). On the other hand they allow us to describe some elements of the information flow independent of any partitioning parameters. Such a description is desired because the information flow is an objective process in dynamical systems and thus independent of the observer, i.e., of the measuring instrument.

Reprint requests to Dr. K. J. G. Kruscha, Sektion Physik/Elektronik der Ernst-Moritz-Arndt-Universität, Domstraße 10a, GDR-2200 Greifswald.

0932-0784 / 88 / 0200-0093 $ 01.30/0. – Please order a reprint rather than making your own copy.

2. Principal Mechanisms of Information Flow

Let $f: x \to y$ be a 1D map ($[0,1[\to [0,1[$) which is continuously differentiable ν-a.e. (ν denotes the Lebesgue measure in $\mathbb{R}$). Figure 1 gives some examples of maps which are considered in this paper. Moreover, let ϱ be a probability density defining a measure μ, $\mu(A) = \int_A \varrho \, d\nu$. ϱ can be interpreted as a density of micro-states. After one iteration of f it is transformed to a probability density ϱ_f according to the Frobenius-Perron equation (illustrated e.g. in [1]):

$$\varrho_f(y) = \sum_{x_i:\, f(x_i) = y} \frac{\varrho(x_i)}{\left|\frac{df}{dx}\right|_{x_i}}. \tag{1}$$

In general, ϱ_f defines a new measure μ_f on $[0,1[$. In the special case where $\varrho = \varrho_f \equiv \bar{\varrho}$ (and hence $\mu = \mu_f \equiv \bar{\mu}$) ϱ is invariant with respect to f. In the following ϱ need not necessarily be f-invariant.

Micro-states are not observable because of the limited precision of every measurement. Therefore we consider a finite μ-measureable partition $\alpha \equiv \{A_i\}_{i=1}^n$ which covers $[0,1[$ ($A_i \cap A_j = \emptyset$ if $i \neq j$) and symbolizes a measuring instrument. Each box A_i of α is referred to as an observable macro-state (also called state for short) of the system. α can be chosen on the basis of several criteria, e.g., each box may have equal size ($\nu(A_i) = 1/n \equiv \varepsilon$ for each $A_i \in \alpha$) or equal probability ($\mu(A_i) = 1/n$ for each $A_i \in \alpha$) which we refer to as ε-partition and natural partition, respectively. The states of α are now considered as initial states. An α-measurement provides the (average) initial information

$$H_x \equiv H(\alpha, \mu) = \sum_{i=1}^{n} p_i \operatorname{ld} \frac{1}{p_i} \quad \text{with } p_i \equiv \mu(A_i) \tag{2}$$

using the well-known Shannon formula where $\operatorname{ld} = \log_2$. On the other hand, H_x represents the uncertainty on an initial state before a measurement is carried out.

Consider now a second partition $\beta \equiv \{B_j\}_{j=1}^m$ of $[0,1[$ on the ordinate corresponding to a measurement of a final (macro-)state B_j by another measuring instrument. Such a measurement provides the information

$$H_y \equiv H(\beta, \mu_f) = \sum_{j=1}^{m} q_j \operatorname{ld} \frac{1}{q_j} \quad \text{with } q_j \equiv \mu_f(B_j). \tag{3}$$

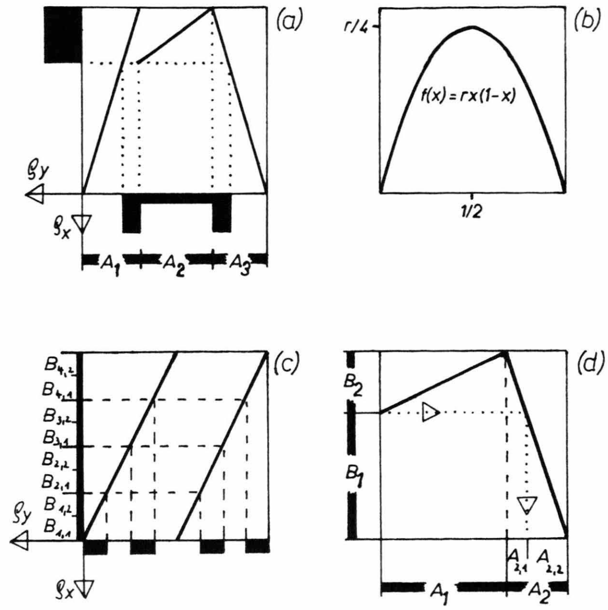


Fig. 1. Examples of 1D maps considered in this paper.

An α-measurement of an initial state A_i may decrease the uncertainty on a final state B_j. The remaining uncertainty on B_j is the result of "final state uncertainty production" ("f.s.-uncertainty production") of the map.

On the other hand, H_x possibly contains a part of information which cannot be attained in a β-measurement of a final state, i.e., due to the mapping this part of initial information is not preserved. It will be called "initial state information loss" ("i.s.-information loss").

Figure 2 illustrates the principal mechanisms of f.s.-uncertainty production and i.s.-information loss neglecting overlappings*. The maps in Fig. 1 are examples in which these mechanisms of information flow can occur. Uncertainty H_p on a final state is produced if one box A_i on the abscissa is mapped on more than one box on the ordinate. Note that f need not necessarily be expanding in A_i, i.e., the absolute value of the slope of f in A_i need not necessarily be greater than one to have uncertainty on a final state.

* With overlapping we denote a situation where we have an $A_i \in \alpha$ and at least two boxes B_{j_1}, $B_{j_2} \in \beta$ such that $\mu_f(f(A_i) \cap B_{j_1}) \neq 0$ and $\mu_f(f(A_i) \cap B_{j_2}) \neq 0$ and at least one box of β, say B_{j_1}, is only partly covered by $f(A_i)$ (i.e., $\mu_f(f(A_i) \cap B_{j_1}) < \mu_f(B_{j_1})$). Effects of overlappings on the information flow were discussed elsewhere [4].

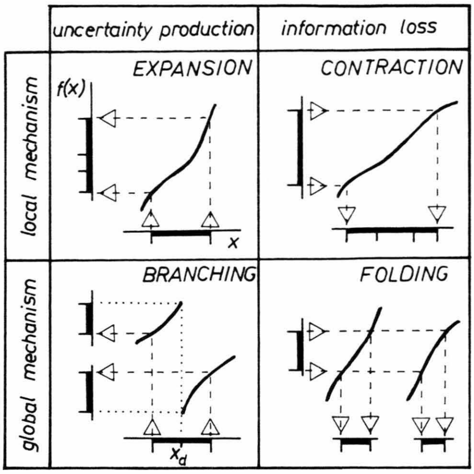


Fig. 2. Global and local mechanisms of information flow.

Thus our present concept of uncertainty production is a generalization of that presented in [4] where uncertainty on a final state was produced only in regions where f has an expanding action (neglecting effects of overlapping). On the other hand, given a final state B_j, we have uncertainty H_l on a former (initial) state if f^{-1} maps B_j to more than one box on the abscissa. Obviously we can distinguish between local and global mechanisms, both, for the uncertainty production and for the information loss as illustrated in Figure 2.

3. Formal Description of Information Flow

To make the notion of f.s.-uncertainty production and i.s.-information loss more precise we now consider conditional entropies (see e.g. [10]). Let $P_{j/i}$ be the probability to find the system after one iteration of f in B_j provided it is at present in A_i,

$$p_{j/i} = \frac{\mu(f^{-1}(B_j) \cap A_i)}{\mu(A_i)}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m \tag{4}$$

with $f^{-1}(B_j) \equiv \{x \mid f(x) \in B_j\}$ ($p_{j/i} = 0$ if $\mu(A_i) = 0$). Then the (average) uncertainty on a final state provided that an initial state is known (called: f.s.-uncertainty production) is given by

$$H_p(\alpha, \mu, \beta, f) = \sum_{i=1}^{n} p_i \sum_{j=1}^{m} p_{j/i} \operatorname{ld} \frac{1}{p_{j/i}}. \tag{5}$$

On the other hand, the uncertainty H_l on an initial state provided that a final state is known (called: i.s.-information loss) is given by

$$H_l(\alpha, \mu, \beta, \mu_f) = \sum_{j=1}^{m} q_j \sum_{i=1}^{n} q_{i/j} \operatorname{ld} \frac{1}{q_{i/j}}. \tag{6}$$

$q_{i/j}$ labels the conditional probability that the final state B_j has been reached from the initial state A_i, provided B_j is given:

$$q_{i/j} = \frac{\mu(f^{-1}(B_j) \cap A_i)}{\mu_f(B_j)}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m, \tag{7}$$

($q_{i/j} = 0$ if $\mu_f(B_j) = 0$). Obviously we have

$$q_{ij} = q_{i/j}\, q_j = p_{j/i}\, p_i = p_{ji}$$

and hence from (2), (3), (5), and (6) follows

$$H_y = H_x - H_l + H_p. \tag{8}$$

This means that H_y consists of two parts: the transinformation (or mutual information) $H_x - H_l$, which is the information we gain about a final state by the measurement of an initial state, and the f.s.-uncertainty production H_p.

From a different point of view, H_x consists of two parts as well: the transinformation $H_y - H_p$ ($= H_x - H_l$) we will gain about a former (initial) state by the measurement of a final state and the i.s.-information loss H_l. Obviously, according to the latter point of view H_l could be considered as "remembrance uncertainty", which represents a direct interpretation of definition (6). However, in this paper we are mainly interested in the predictability problem. Hence we prefer the first interpretation of (8) and call H_l i.s.-information loss. It represents that part of the initial information H_x having no predictive power ($H_l = H_x - (H_y - H_p)$).

For the special case that μ is invariant under f (i.e., $\mu = \mu_f$, which is relevant, e.g., for the stationary behaviour of chaotic maps) and that $\alpha = \beta$, we obtain $H_y = H_x$ and hence $H_l = H_p$. However, in general these equalities do not hold. From Appendix A follows that (8) is invariant under conjugation. Note that instead of f any t-th iterate $f^t = f \circ f^{t-1}$ ($t = 1, 2, 3, \ldots; f^0 = 1$) could be considered by the above formulas.

Though dealing only with 1D maps in this paper, the considerations of this section are valid for many-dimensional systems as well. Moreover, we want to mention that the measure μ need not necessarily be absolutely-continuous with respect to Lebesgue measure. (E.g., if μ is fractal, the considerations are still valid. Of course, in this case we have no densities and hence (1) must be replaced by its more general formulation in terms of measures: $\mu_{f^t}(B) = \mu(f^{-t}(B))$.) Theorem (8) can be referred to as the general scheme of information flow in a dynamical system.

4. Partitioning-Independent Approach to the Information Flow

The information flow as described in Sect. 3 contains objective (f and μ) as well as subjective (α, β) aspects. In this section we consider the objective (partitioning-independent) information flow. The problem of measuring this objective process by chosing suitable partitions is taken into consideration in Section 6.

From the survey of the principal mechanisms of information flow in Fig. 2 we see that all mechanisms can be evaded by a suitable refinement of the partitions α and (or) β on the abscissa and ordinate, respectively, excepting only folding (for details see Section 6). I.s.-information loss due to folding could be evaded only by a suitable coarse-graining of α, which has, however, no practical importance, or by the choice of special initial probability distributions as illustrated in Figure 1 c. Moreover, if we consider only micro-states, then H_p becomes zero because f represents a deterministic system. Consequently (8) reduces to

$$H_x - H_y = H_l, \tag{9}$$

where the i.s.-information loss due to folding represents the remaining essence of the information flow and H_y consists of transinformation only.

In the following we give a derivation of a formula for H_l in terms of micro-states, i.e., without making explicit use of any partitioning concept. Let us start with the Frobenius-Perron equation (1). Using the abbreviation

$$q(x_i, y) \equiv \frac{\varrho(x_i)}{\left|\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_i} \varrho_f(y)}, \tag{10}$$

(1) can be rewritten as

$$1 = \sum_{x_i:\, f(x_i)=y} q(x_i, y).$$

$q(x_i, y)$ can be considered as the "local" probability that a given final micro-state y is reached from the initial micro-state x_i. The corresponding "local" i.s.-information loss is

$$-\sum_{x_i:\, f(x_i)=y} q(x_i, y)\, \mathrm{ld}\, q(x_i, y).$$

Note that this sum extends over all origins x_i of y and hence there is no i.s.-information loss if f is one-to-one. Averaging over all possible micro-states y, we obtain the i.s.-information loss

$$\begin{aligned} H_l &= -\int \sum_{x_i:\, f(x_i)=y} q(x_i, y)\, \mathrm{ld}\, q(x_i, y)\, \mathrm{d}\mu_f \\ &= \int \mathrm{ld}\, \varrho_f\, \mathrm{d}\mu_f - \int \sum_{x_i:\, f(x_i)=y} \frac{\varrho(x_i)}{\left|\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_i}}\, \mathrm{ld}\, \frac{\varrho(x_i)}{\left|\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_i}}\, \mathrm{d}y\,. \end{aligned}$$

Using the abbreviation

$$\lambda(\mu, f) \equiv \int \mathrm{ld} \left|\frac{\mathrm{d}f}{\mathrm{d}x}\right| \mathrm{d}\mu \tag{11}$$

for the "momentary Lyapunov exponent" (note that μ is not necessarily f-invariant), the i.s.-information loss can be written as

$$H_l = \lambda(\mu, f) + \int \mathrm{ld}\, \varrho_f\, \mathrm{d}\mu_f - \int \mathrm{ld}\, \varrho\, \mathrm{d}\mu\,. \tag{12}$$

For an illustration of (12) we return to the map

$$f(x) = \begin{cases} \dfrac{rx}{1-r} + 1 - r & \text{if} \quad 0 \leqq x < 1 - r \\ \dfrac{1-x}{r} & \text{if}\ 1 - r \leqq x < 1 \end{cases}, \quad 0 < r < 1 \tag{13}$$

of Fig. 1 d, which was already considered in [4]. Figure 3 shows the evolution of the momentary Lyapunov exponent $\lambda(\mu_{f^t}, f)$, the entropy $\int \mathrm{ld}\, \varrho_{f^t}\, \mathrm{d}\mu_{f^t}$ of the probability density ϱ_{f^t}, and the i.s.-information loss

$$H_l(t) = \lambda(\mu_{f^t}, f) + \int \mathrm{ld}\, \varrho_{f^{t+1}}\, \mathrm{d}\mu_{f^{t+1}} - \int \mathrm{ld}\, \varrho_{f^t}\, \mathrm{d}\mu_{f^t}$$

in course of time t $(= 0, 1, 2, \ldots)$ for several values of the control parameter r. (The lower L-curve is explained in Section 5.) In each case we started with a uniform distribution in $[0, 1[$ $(\varrho_{f^0} = 1)$. The dashed lines indicate the asymptotic behaviour which is characterized by the invariant density

$$\bar{\varrho}(x) = \begin{cases} 1/(2-r) & \text{if} \quad 0 \leqq x < 1 - r\,, \\ 1/((2-r)r) & \text{if}\ 1 - r \leqq x < 1\,. \end{cases}$$

The zig-zag-behaviour reflects the alternation of a contractive and expanding action of f. A simple calcu-

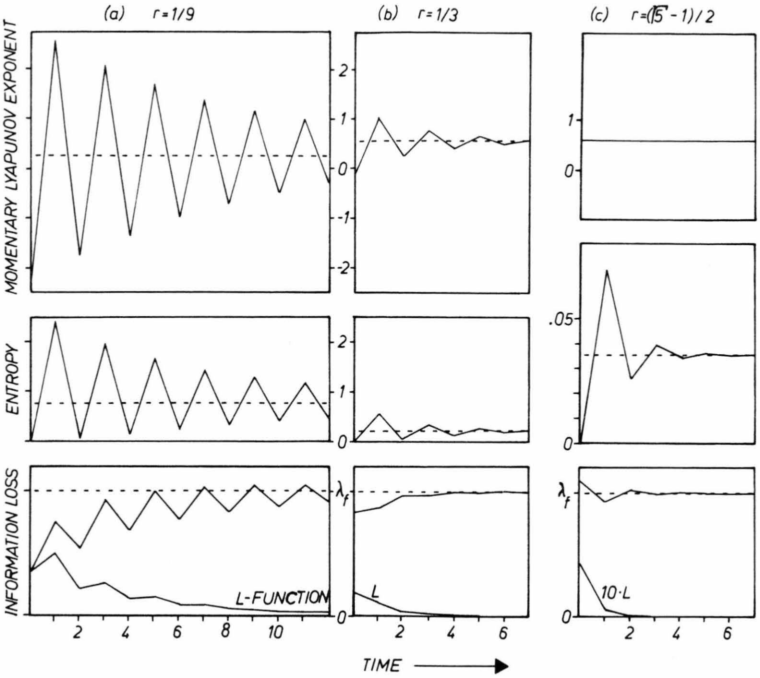


Fig. 3. Relaxation of the momentary Lyapunov exponent $\lambda(\mu_{f^t}, f)$ (Eq. (11)), the entropy $\int \mathrm{ld}\, \varrho_{f^t} \, \mathrm{d}\mu_{f^t}$, the information loss $H_l(t)$ (Eq. (12)), and the L-function (Eq. (19)) for several maps as defined in (13).

lation shows that there is an exponential approach of the momentary Lyapunov exponent to its asymptotic value

$$\lambda_f = \frac{1}{r-1} \, \mathrm{ld}\,(r^r (1-r)^{(1-r)}) \, .$$

Starting with $\varrho_{f^0} = 1$, we obtain (see [9])

$$\lambda(\mu_{f^t}, f) - \lambda_f = \left[\frac{(1-r)^2}{r-2} \, \mathrm{ld} \frac{1-r}{r^2} \right] (r-1)^t \, . \tag{14}$$

Figure 3 demonstrates that, in contrast to H_l, the momentary Lyapunov exponent can considerably deviate from their common asymptotic value λ_f. This is not surprising because in general H_l can vary only through the interval [0, ld k], where k is the number of foldings, whereas $\lambda(\mu_{f^t}, f)$ lies in

$$\left[\inf_{x \in [0,1[} \left\{ \mathrm{ld} \left| \frac{\mathrm{d}f}{\mathrm{d}x} \right| \right\}, \sup_{x \in [0,1[} \left\{ \mathrm{ld} \left| \frac{\mathrm{d}f}{\mathrm{d}x} \right| \right\} \right].$$

For instance, for the map in Fig. 1a there are three foldings for the chosen initial distribution ϱ_x which yields the maximum possible i.s.-information loss of ld 3 bit. On the other hand, any initial density that is concentrated in $A_1 \cup A_3$ (resp. in A_2) yields the maximum (resp. minimum) possible momentary Lyapunov exponent.

If μ is invariant under $f\,(\mu = \bar{\mu})$, the momentary Lyapunov exponent (11) is identical with the "usual" Lyapunov exponent $\lambda_f \equiv \lambda(\bar{\mu}, f)$ and H_l in (12) is given by λ_f alone:

$$H_l \xrightarrow[\max_{A_i \in \alpha} \{\nu(A_i)\} \to 0]{\mu \text{ is } f\text{-invariant}} \lambda_f \, .$$

(For f-invariant measures Chang and Wright [8] have already established the relation between Lyapunov exponent and folding from a different point of view.)

Appendix B shows that the i.s.-information loss (12) is invariant under conjugation and fulfils a additivity property. Moreover, H_l is related to an L-function describing the velocity of the relaxation process of an initial distribution with time, which is considered in the following section.

5. Relaxation of a Non-Stationary Distribution with Regard to the Information Flow

If we consider the Frobenius-Perron equation (1) as an iterative algorithm governing the time evolution of an initial distribution ϱ of an ensemble of micro-states, it is of enormous interest to have a quantity describing the distance from the asymptotic behaviour. Some crucial properties of such a quantity are, e.g., its invariance under conjugation and its monotonous approach to the asymptotic value. In order to derive this quantity, it is useful to consider at first an instructive example.

5.1. Relaxation of Maps of the Bernoulli Type

For a map $f_k = kx \mod 1$ ($k = 2, 3, 4, \ldots$) of the Bernoulli type the i.s.-information loss (12) is given by

$$H_l = \mathrm{ld}\, k + \int \mathrm{ld}\, \varrho_{f_k} \, \mathrm{d}\mu_{f_k} - \int \mathrm{ld}\, \varrho \, \mathrm{d}\mu \,. \tag{15}$$

Consider now a map $\tilde{f}_k$ that is related to f_k by conjugation as illustrated in Appendix A. Then we have

$$\int \mathrm{ld}\, \varrho \, \mathrm{d}\mu = \int \mathrm{ld} \frac{\tilde{\varrho}}{|\mathrm{d}h^{-1}/\mathrm{d}x|} \mathrm{d}\tilde{\mu} \tag{16}$$

(an analogous formula holds for $\int \mathrm{ld}\, \varrho_f \, \mathrm{d}\mu_f$). The "natural" f_k-invariant measure is given by the Lebesgue measure ($\bar{\varrho}(x) = 1$) and hence from (A1) we see that the "natural" $\tilde{f}_k$-invariant measure is given by the density $\tilde{\bar{\varrho}} (\equiv \bar{\varrho}_{\tilde{f}_k}) = |\mathrm{d}h^{-1}/\mathrm{d}\tilde{x}|$. Thus we obtain from (15) and (16)

$$\mathrm{ld}\, k - H_l = \int \mathrm{ld}\, (\tilde{\varrho}/\tilde{\bar{\varrho}}) \, \mathrm{d}\tilde{\mu} - \int \mathrm{ld}\, (\varrho_{\tilde{f}_k}/\tilde{\bar{\varrho}}) \, \mathrm{d}\mu_{\tilde{f}_k} \,. \tag{17}$$

With (A1) in mind, we immediately see that $\int \mathrm{ld}\, (\tilde{\varrho}/\tilde{\bar{\varrho}}) \, \mathrm{d}\tilde{\mu}$ is invariant under conjugation. It can be considered as the information we gain if the invariant density $\tilde{\bar{\varrho}}$ is substituted by the momentary density $\tilde{\varrho}$ [10]. Obviously, the i.s.-information loss H_l due to folding cannot exceed $\mathrm{ld}\, k$, where k is the number of foldings, and thus the right hand part of (17) is always non-negative. Consequently, the information gain

$$\int \mathrm{ld}\, (\varrho_{\tilde{f}_k^t}/\tilde{\bar{\varrho}}) \, \mathrm{d}\mu_{\tilde{f}_k^t}, \quad t = 0, 1, 2, \ldots$$

fulfils the desired monotony criterion as time t goes to infinity. It approaches zero if $\varrho_{\tilde{f}_k^t} \to \tilde{\bar{\varrho}}$ for $t \to \infty$ *.

To illustrate the relaxation of the information gain we refer to the stimulating numerical simulations of Shaw [1] (Figs. A2 and A3), concerning the logistic map $\tilde{f}_2 = 4x\,(1-x)$, which is conjugate to the Bernoulli map f_2 (see e.g. [12]).

Consider e.g. Fig. A2 in [1]. In the first four iterations there is no folding. Thus the i.s.-information loss H_l in (17) equals zero, from which we conclude that the change of the information gain equals exactly $\mathrm{ld}\, 2 =$ 1 bit for each of the first four iterations. However, with the onset of foldings i.s.-information loss occurs and hence the approach to equilibrium is delayed considerably.

Motivated by the above considerations, we argue that the information gain fulfils the desired properties of a measure for the distance from the stationary distribution also in a more general sense.

5.2. Information Gain as a Relaxation Measure

Generally the i.s.-information loss H_l in (12) can be rewritten as

$$H_l = \int \mathrm{ld}\, (\varrho_f/\bar{\varrho}) \, \mathrm{d}\mu_f - \int \mathrm{ld}\, (\varrho/\bar{\varrho}) \, \mathrm{d}\mu + \int \mathrm{ld}\, |\mathrm{d}f/\mathrm{d}x| \, \mathrm{d}\mu + \int \mathrm{ld}\, \bar{\varrho} \, \mathrm{d}\mu_f - \int \mathrm{ld}\, \bar{\varrho} \, \mathrm{d}\mu \,, \tag{18}$$

supposing that $\mathrm{supp}\, \bar{\varrho} \supseteq \mathrm{supp}\, \varrho, \mathrm{supp}\, \varrho_f$. (18) can be rewritten as

$$\int \mathrm{ld} \left| \frac{\mathrm{d}f}{\mathrm{d}x} \right| \frac{\bar{\varrho}(f(x))}{\bar{\varrho}(x)} \mathrm{d}\mu - H_l = \int \mathrm{ld} \frac{\varrho}{\bar{\varrho}} \mathrm{d}\mu - \int \mathrm{ld} \frac{\varrho_f}{\bar{\varrho}} \mathrm{d}\mu_f \equiv L(\mu, f) \,. \tag{19}$$

From (B1) and (A1) we immediately see that all terms in (19) are invariant under conjugation.

In order to show that the information gain $\int \mathrm{ld}\, (\varrho_{f^t}/\bar{\varrho}) \, \mathrm{d}\mu_{f^t}$ can only decrease or remain constant with time $t = 0, 1, 2, \ldots$, we rewrite the L-function in (19). Using the relation

$$\int \mathrm{ld} \frac{\varrho}{\bar{\varrho}} \mathrm{d}\mu = \int \sum_{x_i:\, f(x_i) = y} \frac{\varrho(x_i)}{\varrho_f(y) \left| \frac{\mathrm{d}f}{\mathrm{d}x} \right|_{x_i}} \mathrm{ld} \frac{\varrho(x_i)}{\bar{\varrho}(x_i)} \mathrm{d}\mu_f$$

and the abbreviations (10) and

$$\bar{q}(x_i, y) \equiv \frac{\bar{\varrho}(x_i)}{\bar{\varrho}(y) \left| \frac{\mathrm{d}f}{\mathrm{d}x} \right|_{x_i}} \,,$$

we obtain for the L-function

$$L = \int \sum_{x_i:\, f(x_i) = y} q(x_i, y) \, \mathrm{ld} \frac{q(x_i, y)}{\bar{q}(x_i, y)} \mathrm{d}\mu_f \,. \tag{20}$$

The sum in (20) can be interpreted as the information we gain if the (discrete) probability distribution $\{\bar{q}(x_i, y)\}_i$ is substituted by $\{q(x_i, y)\}_i$. It is a well-known fact that this information gain is always non-negative (see e.g. [10]). From the local non-negativity follows globally

$$L \geqq 0 \,, \tag{21}$$

* It should be mentioned that the information gain is used in statistical mechanics as an entropy measure that cannot increase for certain stochastic processes [11].

i.e., the information gain $\int \mathrm{ld}(\varrho_{f^t}/\bar{\varrho})\,\mathrm{d}\mu_{f^t}$ cannot increase with time t (supposing that $\mathrm{supp}\,\varrho_{f^t} \subseteq \mathrm{supp}\,\bar{\varrho}$ for all $t = 0, 1, 2, \ldots$). (Note that $q(x_i, y)$ and $\bar{q}(x_i, y)$ must exist only ν-a.e. Especially our considerations are still valid if $\mathrm{d}f/\mathrm{d}x$ exists only ν-a.e.) Obviously, (21) can be considered as an analogon to Boltzmann's H-theorem in thermodynamics. For some illustrations of the evolution of the L-function see figure 3.

5.3. *Interpretation of the L-Function*

According to the definition in (19), the L-function describes the change of the information gain after one iteration of f. However, there is another possible interpretation of L in terms of the i.s.-information loss, which will be given below.

The left hand term in (19) can be rewritten as

$$\int \mathrm{ld}\left|\frac{\mathrm{d}f}{\mathrm{d}x}\right| \frac{\bar{\varrho}(f(x))}{\bar{\varrho}(x)}\,\mathrm{d}\mu$$

$$= \int \sum_{x_i:\, f(x_i) = y} q(x_i, y)\,\mathrm{ld}\,\frac{1}{\bar{q}(x_i, y)}\,\mathrm{d}\mu_f \equiv \bar{H}_l\,, \quad (22)$$

and thus we have

$$L = \bar{H}_l - H_l\,(\geqq 0)\,. \quad (23)$$

In order to give an interpretation of $\bar{H}_l$ defined in (22), we refer to the Bongard-Kerridge entropy H^B [15] describing on the basis of a hypothesis $\{\bar{q}_j\}_j$ the uncertainty on an ensemble of states occurring according to the probability distribution $\{q_j\}_j$:

$$H^B \equiv -\sum_j q_j\,\mathrm{ld}\,\bar{q}_j\,,$$

i.e., H^B is the subjective uncertainty. Thus, on the basis of the stationarity hypothesis $\{\bar{q}(x_i, y)\}_i$, (22) can be considered as the subjective i.s.-information loss $\bar{H}_l$. It is a well-known fact that the objective uncertainty (here: objective i.s.-information loss H_l) cannot exceed the subjective uncertainty (here: subjective i.s.-information loss $\bar{H}_l$). Hence L in (23) must be non-negative as was already suggested above in a somewhat different way. If ϱ_{f^t} approaches $\bar{\varrho}$ as time t goes to infinity, then the objective i.s.-information loss H_l approaches the subjective i.s.-information loss $\bar{H}_l$ from below ($H_l \uparrow \bar{H}_l$). However, this approach must not necessarily be monotonous, which is in contrast to the monotonous approach of the information gain to its asymptotic value zero (Fig. 3 shows an example). The more the subjective i.s.-information loss differs from the objective one (i.e., the greater L is), the faster the system approaches its asymptotic distribution $\bar{\varrho}$.

5.4. *Relaxation to Non-Stationary Asymptotic Behaviour*

Finally we want to discuss a situation where the initial distribution $\varrho \equiv \varrho_{f^0}$ does not necessarily approach a f-invariant distribution $\bar{\varrho}$ with time. Consider e.g. the family of logistic maps $f = rx(1-x)$, $r \in [0, 4]$. Slightly above an accumulation point of a period doubling cascade chaotic behaviour is likely to occur on a k-band attractor. If the initial measure μ (corresponding to ϱ) is already concentrated on the bands and the measure of at least one band differs from $1/k$, then ϱ_{f^t} cannot approach $\bar{\varrho}$. However, in this situation the asymptotic behaviour of the sequence $\{\varrho_{f^t}\}_{t=0}^{\infty}$ is periodic with a period k, and we have

$$\lim_{t\to\infty} \varrho_{f^{tk+l}} \equiv \bar{\varrho}_l \quad \text{for } l = 1, 2, \ldots, k\,. \quad (24)$$

I.e., each $\bar{\varrho}_l$ is f^k-invariant. Then the distance from the asymptotic behaviour can be described by the information gain

$$\int \mathrm{ld}(\varrho_{f^{tk+l}}/\bar{\varrho}_l)\,\mathrm{d}\mu_{f^{tk+l}}\,, \quad l = 1, 2, \ldots, k\,.$$

But there is another approach to describe the relaxation process. This can be done by considering two (in general different) initial densities ϱ and $\varkappa$, where $\varkappa$ shall play the role of a (not necessarily f-invariant) comparison density that is iterated according to the Frobenius-Perron equation, just as ϱ is iterated. Supposing that $\mathrm{supp}\,\varrho \subseteq \mathrm{supp}\,\varkappa$, we can consider the information gain

$$G(\varrho, \varkappa, f, t) \equiv \int \mathrm{ld}(\varrho_{f^t}/\varkappa_{f^t})\,\mathrm{d}\mu_{f^t}\,. \quad (25)$$

(For the special case that $\varkappa$ is f-invariant and that $\varrho_{f^t} \to \varkappa$ if $t \to \infty$ we have the situation of section 5.2.) From (A 1) follows immediately that (25) is still invariant under conjugation. Moreover, in the present more general case we can define a (generalized) L-function that cannot be negative:

$$L(\varrho, \varkappa, f, t) \equiv G(\varrho, \varkappa, f, t) - G(\varrho, \varkappa, f, t+1) \geqq 0\,. \quad (26)$$

(Note that in Section 5.2. we did not have to make use of the f-invariance of $\bar{\varrho}$ in order to argue that L is always non-negative. Thus we can immediately verify (26).) Now the information gain in (25) possibly approaches asymptotically a positive value, but it still cannot increase with time t. The existence of the limit

$$G_\infty \equiv \lim_{t\to\infty} G(\varrho, \varkappa, f, t) \quad (27)$$

follows from the monotony (26) and the fact that (25) is bounded from below by zero.

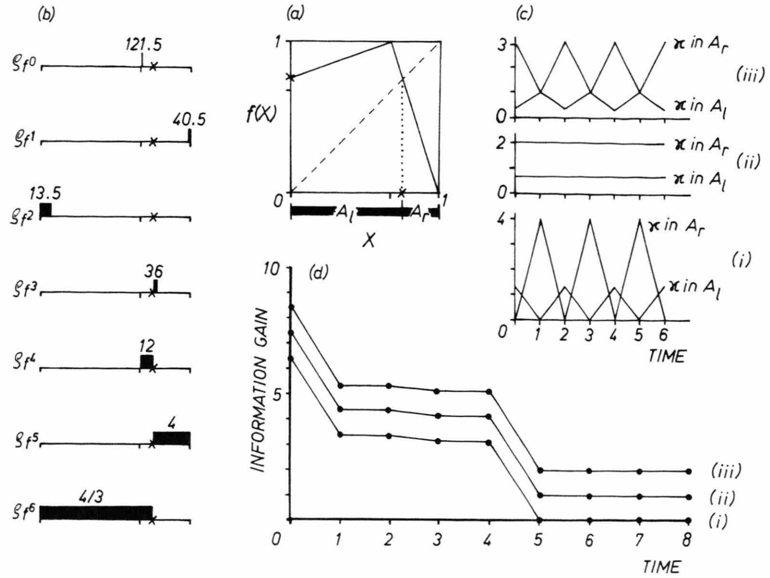


Fig. 4. Illustration of the evolution of the information gain (25) of a density ϱ_{f^t} relative to three different comparison densities $\varkappa_{f^t}$ for the map (30) ($r = 1/3$). Picture (a) shows the graph of the considered map. In (b) the evolution of ϱ_{f^t} is illustrated where the numbers indicate the height of ϱ_{f^t}. (c) shows the evolution of $\varkappa_{f^t}$ in three different cases. Finally, (d) illustrates the monotonous approach of the information gain to the asymptotic value (29) for three different comparison densities.

Presuming e.g. a k-band attractor, we can rewrite (25) as

$$G(\varrho, \varkappa, f, tk + l) = \int \mathrm{ld} \frac{\varrho_{f^{tk+l}}}{\bar{\varrho}_l} \, \mathrm{d}\mu_{f^{tk+l}} - \int \mathrm{ld} \frac{\varkappa_{f^{tk+l}}}{\bar{\varrho}_l} \, \mathrm{d}\mu_{f^{tk+l}}, \quad (28)$$

supposing that $\operatorname{supp} \bar{\varrho}_l \supseteq \operatorname{supp} \varrho_{f^{tk+l}}, \operatorname{supp} \varkappa_{f^{tk+l}}$.

From (24), (27), and (28) we see that the information gain (25) approaches for $t \to \infty$

$$G_\infty = -\int \bar{\varrho}_l \, \mathrm{ld} \frac{\bar{\varkappa}_l}{\bar{\varrho}_l} \, \mathrm{d}x \quad \text{for any } l \in \{1, 2, \ldots, k\}, \quad (29)$$

where we used the abbreviation $\bar{\varkappa}_l \equiv \lim_{t \to \infty} \varkappa_{f^{tk+l}}$, supposing that the limit $\bar{\varkappa}_l$ exists. It should be emphasized that the limit value G_∞ of the information gain is independent of the "phase" l.

For an illustration of the relaxation of the (generalized) information gain (25) let us consider the map

$$f(x) = \begin{cases} \dfrac{rx}{1 - r^2} + \dfrac{1}{1 + r} & \text{if } 0 \leqq x < 1 - r, \\ & \qquad \text{for } 0 < r < 1. \\ \dfrac{1 - x}{r} & \text{if } 1 - r \leqq x < 1, \end{cases} \quad (30)$$

The graph of f is shown in Fig. 4a for $r = 1/3$. As was already mentioned in [4], (30) has a two-band attractor. All points of $A_l =]0, 3/4[$ are mapped to $A_r =]3/4, 1[$, and vice versa. Starting with a uniform distribution in the interval $]2/3, 2/3 + 1/121.5[$, Fig. 4b shows the evolution of ϱ_{f^t} according to the Frobenius-Perron eq. (1). After 5 iterations of ϱ_{f^0} the system becomes uniformly distributed in A_r and thus it has reached a periodic situation,

$$\varrho_{f^{2t+1}} = \bar{\varrho}_1 \text{ and } \varrho_{f^{2t+2}} = \bar{\varrho}_2 \quad \text{for } t = 2, 3, 4, \ldots. \quad (31)$$

Figure 4d shows the evolution of the information gain (25) for several comparison densities $\varkappa_{f^t}$. The evolution of $\varkappa_{f^t}$ in A_l and A_r is illustrated in Figure 4c. In case (i) we have chosen

$$\varkappa_{f^{2t+2}} = \bar{\varrho}_2 = \bar{\varkappa}_2 \quad \text{for } t = -1, 0, 1, 2, \ldots$$

(and hence $\varkappa_{f^{2t+1}} = \bar{\varrho}_1 = \bar{\varkappa}_1$ for $t = 0, 1, 2, \ldots$). From (29) we immediately see that $G_\infty = 0$ in this case and from (31) follows that the information gain (25) is zero after 5 iterations. In case (ii) the comparison density $\varkappa_{f^t}$ is f-invariant. Hence the value of $\varkappa_{f^t}$ remains constant in each box A_l and A_r as illustrated in Figure 4c. Now the information gain (25) approaches $G_\infty = 1$ bit, which is the information whether the system is in A_l or A_r. Finally we have chosen the comparison density $\varkappa_{f^1} = 1$ in [0, 1], which develops as illustrated in case (iii) of Figure 4c. From (29) we see now that the information gain approaches

$$G_\infty = \int \varrho_{f^5} \, \mathrm{ld} \, \varrho_{f^5} \, \mathrm{d}x = \int_{(A_r)} 4 \, \mathrm{ld} \, 4 \, \mathrm{d}x = 2 \text{ bit}.$$

6. Measuring the Information Flow

In Sects. 4 and 5 we have considered some aspects of the objective information flow – e.g. the central formula (12) of the i.s.-information loss H_l contains no partitioning parameters. However, micro-states are not observable. Hence the measurement of the information flow is always based on macro-states of partitions corresponding to measuring instruments. Generally, these partitions yield to subjective aspects of the information flow and possibly enormously falsify the objective information flow. This is, e.g., due to the occurrence of overlappings (see *, p. 94), which were investigated elsewhere [4]. Nevertheless, the following considerations demonstrate that we can measure the objective information flow if we use privileged partitions which are "adapted" to the dynamical system.

6.1. *Forward Partitioning*

Let us assume now that we have a given partition α of the abscissa. Obviously the most natural way to evade overlappings in this case is to define a partition β of the ordinate by the images of all divisional points of α under the action of f. More precisely, we have in mind the following "forward partitioning" concept:

Let $\{x_i\}_{i=0}^{n} \equiv D(\alpha)$ be the set of divisional points of α. Then the set $\{f(x_i)\}_{i=0}^{n} \equiv D(\beta)$ defines a partition $\beta = f(\alpha) = \{B_j\}_{j=1}^{m}$ of [0, 1[on the ordinate ($m \leqq n$). To exclude f.s.-uncertainty production due to branching as illustrated in Fig. 2, we suppose that $D(\alpha)$ contains all points of discontinuity of f. If x_d is a point of discontinuity, it is assumed to produce two divisional points $\lim_{\delta \downarrow 0} f(x_d \mp \delta)$ on the ordinate. Obviously, this concept of partitioning also prevents information flow due to contraction as illustrated in Figure 2. Within this concept α can be chosen relatively arbitrarily, whereas β is uniquely determined by α and the map f.

Let α be chosen now in such a way that the slope $\mathrm{d}f/\mathrm{d}x$ as well as the density ϱ are constant in each box of α. (This implies that ϱ_f is constant in each box of β as well.) It should be mentioned that for piecewise linear maps this assumption may hold already for relatively rough partitions on the abscissa. However, if we have no piecewise linear map, this supposition can be approximated by sufficiently fine partitions α. In the latter case the following formulas are meant to be approximations which are assumed to become identities for $\max_i \{v(A_i) \mid A_i \in \alpha\} \to 0$. Using the abbreviations

$$\Delta x_i \equiv v(A_i), \; \Delta y_j \equiv v(B_j), \; s_i \equiv \left| \frac{\mathrm{d}f}{\mathrm{d}x}\bigg|_{x \in A_i} \right|,$$
$$\varrho_i \equiv \varrho(x \in A_i), \; \varrho_{f,j} \equiv \varrho_f(y \in B_j),$$

the transition probabilities (4) and (7) can be rewritten as

$$p_{j/i} = \frac{\Delta y_j}{s_i \Delta x_i} \quad \text{and} \quad q_{i/j} = \frac{\varrho_i}{s_i \varrho_{f,j}}. \tag{32}$$

Thus we obtain via (5) and (32) the f.s.-uncertainty production

$$\begin{aligned} H_p &= \sum_{i=1}^{n} p_i \,\mathrm{ld}\, s_i + \sum_{i,j=1}^{n,m} p_{ji} \,\mathrm{ld}\, \frac{\Delta x_i}{\Delta y_j} \\ &= \lambda(\mu, f) + \sum_{i=1}^{n} p_i \,\mathrm{ld}\, \Delta x_i - \sum_{j=1}^{m} q_j \,\mathrm{ld}\, \Delta y_j . \end{aligned} \tag{33}$$

On the other hand, from (6) and (32) follows the i.s.-information loss

$$H_l = \sum_{i=1}^{n} p_i \,\mathrm{ld}\, s_i + \sum_{i,j=1}^{n,m} p_{ji} \,\mathrm{ld}\, \frac{\varrho_{f,j}}{\varrho_i},$$

which can be rewritten as formula (12). Hence, in the framework of the forward partitioning concept, H_l describes the objective (partitioning-independent) i.s.-information loss.

6.2. *Aspects of Uncertainty Production*

In several recent papers [1, 2, 4] the f.s.-uncertainty production H_p was considered as the most interesting quantity with regard to the predictability problem in chaotic systems. However, in our present consideration we have in mind different partitions α and β on the abscissa and ordinate, respectively, and nonstationary measures. Consequently the f.s.-uncertainty production generally differs from the i.s.-information loss. Moreover, using e.g. the forward partitioning concept of Section 6.1, we see that H_p involves, in contrast to H_l, see (12) and (33), subjective aspects of the information flow, i.e., H_p depends on terms containing partitioning parameters, which can already be found in H_x and H_y: Within the framework of the piecewise linear approximation described above we have $p_i = \varrho_i \Delta x_i$ and we can rewrite the initial information (2) as

$$\begin{aligned} H_x &= \sum_{i=1}^{n} p_i \,\mathrm{ld}\, \frac{1}{\Delta x_i} - \int \mathrm{ld}\, \varrho \,\mathrm{d}\mu \\ &= \mathrm{ld}\, n - \left(\sum_{i=1}^{n} p_i \,\mathrm{ld}\, \frac{\Delta x_i}{\varepsilon} + \int \mathrm{ld}\, \varrho \,\mathrm{d}\mu \right), \end{aligned} \tag{34}$$

($\varepsilon \equiv 1/n$). The final information H_y can be rewritten in a similar way using $q_j = \varrho_{f,j} \Delta y_j$.

Obviously only the entropy $\int \mathrm{ld}\,\varrho \, \mathrm{d}\mu$ in (34) is solely determined by μ, i.e., it is independent of the measuring instrument. (The information loss H_l in (12) relates the change of this entropy to the momentary Lyapunov exponent, which is also solely determined by the dynamical system.) The information flow between H_x and H_y that is related to the partitions α and $f(\alpha)$, i.e., to the measuring instruments, is completely described by the f.s.-uncertainty production (33). Indeed, H_p can be rewritten as

$$H_p = \lambda(\mu, f) + \mathrm{ld}\, m - \mathrm{ld}\, n + \sum_{i=1}^{n} p_i \,\mathrm{ld} \frac{\Delta x_i}{\varepsilon} - \sum_{j=1}^{m} q_j \,\mathrm{ld} \frac{\Delta y_j}{\delta}, \tag{35}$$

($\delta \equiv 1/m$). Thus H_p contains three different parts,

(i) the mean expansion (or contraction) rate of f with respect to the measure μ described by $\lambda(\mu, f)$,

(ii) the change of the number of states on the ordinate and abscissa described by $\mathrm{ld}(m/n)$, and

(iii) the change in the valuation of the sizes of the boxes of the partitions α and $f(\alpha) = \beta$ as compared to ε- and δ-partitions, respectively.

However, there are situations where the f.s.-uncertainty production can be expressed by $\lambda(\mu, f)$ alone and thus independent of any parameters describing the underlying partitions. Consider, e.g., a map of the Bernoulli type: $f_k(x) = kx \bmod 1$. Starting with an ε-partition $\alpha = \{[(i-1)/n, i/n[\}_{i=1}^{n}$ on the abscissa, there is induced a partition β on the ordinate according to the forward partitioning concept such that a suitable refinement β^* of β is an ε-partition as well (Figure 1 c).

In this case each box of α is mapped on k boxes of β^*, and the f.s.-uncertainty production H_p is given by the momentary Lyapunov exponent $\lambda(\mu, f_k) = \lambda_{f_k} = \mathrm{ld}\, k$.

There is a general method to make $H_p = 0$, thus making it in a trivial way, independent of any parameters describing the underlying partitions. This method will be described in the following section.

6.3. *Backward-Partitioning and Optimum Measuring*

We consider now a "backward partitioning" concept which involves folding as the only mechanism of information flow and consequently reduces the general scheme (8) to (9). Thus we reconsider the situation of Sect. 4, but now with regard to macro-states.

Let β be a rather arbitrary partition on the ordinate which is characterized by the divisional points $D(\beta) = \{y_j\}_{j=0}^{m}$. Then consider the induced partition $f^{-1}(\beta) = \alpha$ on the abscissa which is defined by the set $\{f^{-1}(y_j)\}_{j=0}^{m} \cup X_d = D(\alpha)$ of divisional points. With X_d we denote the set of all points x_d of discontinuity of f that fulfil the constraints

(i) $\left]\lim_{\delta \downarrow 0} f(x_{d+}^{(-)}\delta), \lim_{\delta \downarrow 0} f(x_{d-}^{(+)}\delta)\right[\cap D(\beta) \neq \emptyset$, and

(ii) if A_{i_1} and A_{i_2} are the boxes on the abscissa that touch each other in x_d, then $\mu(A_{i_1}) > 0$ and $\mu(A_{i_2}) > 0$.

Constraint (i) is to avoid f.s.-uncertainty production by branching as illustrated in Figure 2. Obviously this partitioning concept guarantees (besides $H_p = 0$) that the i.s.-information loss is as small as possible*. Hence in an optimal α-measurement of an initial state, which is arranged to predict a final β-state with a minimum of experimental effort, the information loss is only caused by folding and this loss represents the inevitable minimum of information contained in H_x that has no predictive power. Moreover, for a sufficiently fine partition β (and hence α), such that the piecewise linear approximation of Section 6.1. holds, the (minimum) information loss in an optimal α-measurement is given by (12). Obviously an optimal α-measurement of an initial state also involves the attempt to use natural partitions α, which maximize the initial information H_x for a given number of macro-states.

It should be mentioned that e.g. in the case where there are no points of dicontinuity of f backward partitioning represents a special kind of forward partitioning ($f(f^{-1}(\beta)) = \beta$). Moreover, the refinement $f^{-1}(f(\alpha))$ of a given partition α (which is the partition of forward partitioning) fulfils the conditions of backward partitioning. This refinement avoids i.s.-information loss due to contraction. (For a simple example see Figure 1 d.)

6.4. *Effects of Additional Refinements*

Let us start with a backward partitioning ($\alpha = f^{-1}(\beta)$). Then an additional refinement of α in-

* Possibly there is a box $A_i \in \alpha = f^{-1}(\beta)$ that has measure zero ($\mu(A_i) = 0$). In this situation α is not the roughest partition guranteeing that there is no uncertainty production. To cover this situation we consider all partitions to be equivalent that differ only in boxes of measure zero.

crease, generally, the initial information H_x. However, the excess initial information is equivalent to the additional i.s.-information loss due to contraction (see Fig. 2) having no predictive power. On the other hand, let us start again with backward partitioning and then consider a refinement of β which increases the final information H_y. In this case the excess final information represents, generally, additional f.s.-uncertainty production and surprisingly decreases the i.s.-information loss. If we have in mind, however, the piecewise linear approximation of Section 6.1, the excess final information is equivalent to the additional f.s.-uncertainty production and hence the i.s.-information loss remains unchanged in this case (for more details see [9]).

7. Conclusions

On the basis of well-known information theoretical quantities we have proposed a comprehensive ideology of information flow in one-dimensional maps, including measuring concepts making some elements of information flow available to an experimentator who can only deal with observable macro-states. Moreover, we showed that the relaxation process of initial distribution under the action of the map can well be understood in terms of the information flow. In most considerations of the present paper we have had in mind an asymptotically chaotic behaviour which is characterized by an invariant density absolutely continuous with respect to the Lebesgue measure. Such 1D maps are many-to-one and hence uniquely define the positive time direction. Consequently, our results appear to be quite reasonable – the uncertainty on future states equals zero whereas the i.s.-information loss does not vanish and thus appears to be the central quantity in the scheme of information flow. For invariant (ergodic) measures this uncertainty is given by the Lyapunov exponent of the map. However, in general (for non-invariant measures) the (momentary) Lyapunov exponent does not play such a central role any longer. Moreover, the i.s.-information loss, which is a functional of the dynamics of the system, is related to another important quantity, which is a functional of the momentary "state" of the system. This is the information gain (25) of a present probability density ϱ_{f^t} relative to a comparison density $\varkappa_{f^t}$. It cannot increase with time and thus can be considered as a measure of the distance from asymptotic behaviour which fulfils a kind of Boltzmann's H-theorem for 1D maps. It should be emphasized that $\varkappa_{f^t}$ must not be invariant under the action of the map, which may be highly advantageous in practical computations because, typically, the invariant density defining a unique ergodic measure is not explicity known or even does not exist.

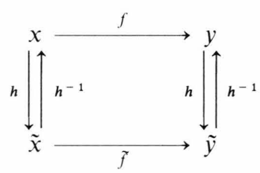
Finally we want to mention that there are far-reaching practical applications of the ideas of information flow presented in this paper. One of us (K.J.G.K.) has used the ideas to establish a breaking off criterion testing the physical relevancy of numerical simulations of plasma-beam interactions [13].

Acknowledgement

We are grateful to the referee for his constructive suggestions and valuable comments.

Appendix A: Invariance under Conjugation of the Information Flow

In the following we show that (8) is invariant under conjugation. Let h be a continuously differentiable one-to-one map such that the diagram



is commutative. Consider now the partitions α and β and the induced partitions $\tilde{\alpha} \equiv \{h(A_i)\}_{i=1}^{n}$ and $\tilde{\beta} \equiv \{h(B_j)\}_{j=1}^{m}$. Moreover, let $\tilde{\mu}$ be the measure corresponding to the density

$$\tilde{\varrho}(\tilde{x}) = \varrho(h^{-1}(\tilde{x})) \left| \frac{\mathrm{d}h^{-1}(\tilde{x})}{\mathrm{d}\tilde{x}} \right| . \tag{A 1}$$

(Note that if ϱ is f-invariant, then $\tilde{\varrho}$ is $\tilde{f}$-invariant [14].) From (A 1) follows the conservation of probability, $\mu(A_i) = \tilde{\mu}(h(A_i))$, and hence $H(\alpha, \mu) = H(\tilde{\alpha}, \tilde{\mu})$. Moreover, from (1) and (A 1) we see that

$$\varrho_{\tilde{f}}(\tilde{y}) \equiv \sum_{\tilde{x}_i : \tilde{f}(\tilde{x}_i) = \tilde{y}} \frac{\tilde{\varrho}(\tilde{x}_i)}{\left| \frac{\mathrm{d}\tilde{f}}{\mathrm{d}\tilde{x}} \right|_{\tilde{x}_i}} = \varrho_f(h^{-1}(\tilde{y})) \left| \frac{\mathrm{d}h^{-1}(\tilde{y})}{\mathrm{d}\tilde{y}} \right| .$$

Thus $\mu_f(B_j) = \mu_{\tilde{f}}(h(B_j))$ and hence $H(\beta, \mu_f) = H(\tilde{\beta}, \mu_{\tilde{f}})$. On the other hand, we obviously have

$$\mu(f^{-1}(B_j) \cap A_i) = \tilde{\mu}(\tilde{f}^{-1}(h(B_j)) \cap h(A_i))$$

and consequently

$$H_p(\alpha, \mu, \beta, f) = H_p(\tilde{\alpha}, \tilde{\mu}, \tilde{\beta}, \tilde{f}) \quad \text{and}$$
$$H_l(\alpha, \mu, \beta, \mu_f) = H_l(\tilde{\alpha}, \tilde{\mu}, \tilde{\beta}, \mu_{\tilde{f}}) .$$

Thus all components of (8) are invariant under conjugation.

Appendix B: Properties of the I.S.-Information Loss Due to Folding

(i) Invariance Under Conjugation

A substantial property of H_l in (12) is its invariance under conjugation. Using the abbreviation of Appendix A, we can write

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\mathrm{d}h^{-1}}{\mathrm{d}\tilde{y}} \frac{\mathrm{d}\tilde{f}}{\mathrm{d}\tilde{x}} \frac{\mathrm{d}h}{\mathrm{d}x} = \frac{\mathrm{d}\tilde{f}}{\mathrm{d}\tilde{x}} \frac{\mathrm{d}h}{\mathrm{d}x} \Big/ \frac{\mathrm{d}h}{\mathrm{d}y}, \qquad \text{(B1)}$$

from which we can rewrite (12) as

$$H_l = \int \mathrm{ld} \left|\frac{\mathrm{d}\tilde{f}}{\mathrm{d}\tilde{x}}\right| \mathrm{d}\mu + \int \mathrm{ld}\, \varrho_f \,\mathrm{d}\mu_f - \int \mathrm{ld} \left|\left.\frac{\mathrm{d}h}{\mathrm{d}x}\right|_{f(x)}\right| \mathrm{d}\mu - \int \mathrm{ld}\, \varrho \,\mathrm{d}\mu + \int \mathrm{ld} \left|\frac{\mathrm{d}h}{\mathrm{d}x}\right| \mathrm{d}\mu .$$

Using the relation

$$\int \mathrm{ld} \left|\left.\frac{\mathrm{d}h}{\mathrm{d}x}\right|_{f(x)}\right| \mathrm{d}\mu = \int \mathrm{ld} \left|\frac{\mathrm{d}h}{\mathrm{d}y}\right| \mathrm{d}\mu_f$$

and the conservation of probability described by (A1), we obtain

$$H_l = \int \mathrm{ld} \left|\frac{\mathrm{d}\tilde{f}}{\mathrm{d}\tilde{x}}\right| \mathrm{d}\tilde{\mu} + \int \mathrm{ld}\, \varrho_{\tilde{f}} \mathrm{d}\mu_{\tilde{f}} - \int \mathrm{ld}\, \tilde{\varrho}\, \mathrm{d}\tilde{\mu} ,$$

i.e., H_l in (12) is invariant under conjugation. This invariance is not trivial, because each component of H_l is, in general, not invariant. For instance the momentary Lyapunov exponent $\lambda(\mu, f)$ is, in general, not invariant under conjugation. However, in the case where μ is f-invariant we have $\lambda(\mu, f) = \lambda_f = \lambda_{\tilde{f}}$, which is a well-known result (see e.g. [14]).

(ii) Additivity

Another interesting property of H_l in (12) is its additivity. Let $f^2 \equiv f \circ f$, then we obtain

$$H_l(\mu, f^2) = \lambda(\mu, f^2) + \int \mathrm{ld}\, \varrho_{f^2} \,\mathrm{d}\mu_{f^2} - \int \mathrm{ld}\, \varrho \,\mathrm{d}\mu . \qquad \text{(B2)}$$

Noting that

$$\lambda(\mu, f^2) = \int \mathrm{ld} \left|\frac{\mathrm{d}f^2}{\mathrm{d}x}\right| \mathrm{d}\mu = \int \mathrm{ld} \left|\frac{\mathrm{d}f}{\mathrm{d}y}\right| \mathrm{d}\mu_f + \int \mathrm{ld} \left|\frac{\mathrm{d}f}{\mathrm{d}x}\right| \mathrm{d}\mu ,$$

we can rewrite (B2) as

$$H_l(\mu, f^2) = H_l(\mu, f) + H_l(\mu_f, f) .$$

Analogously we obtain in a more general case

$$H_l(\mu, f^{u+v}) = H_l(\mu, f^u) + H_l(\mu_{f^u}, f^v) \quad \text{for } u, v = 1, 2, 3, \ldots$$

(it should be mentioned that, in general, for finite partitions H_l in (6) is not additive).

[1] R. Shaw, Z. Naturforsch. **36a,** 80 (1981).
[2] J. D. Farmer, Z. Naturforsch. **37a,** 1304 (1982).
[3] P. Grassberger, Physica **14D,** 365 (1985).
[4] B. Pompe, J. Kruscha, and R. W. Leven, Z. Nafurforsch. **41a,** 801 (1986).
[5] J. Shukla, Advances in Geophysics **28B,** 87 (1985).
[6] G. Holloway and B. J. West (eds.), Predictability of Fluid Motions, American Institute of Physics, New York 1984.
[7] P. Collet and J.-P. Eckmann, Iterated Maps on the Interval as Dynamical Systems, Birkhäuser, Cambridge MA 1980.
[8] S.-J. Chang and J. Wright, Phys. Rev. **A 23,** 1419 (1981).
[9] J. Kruscha, Untersuchung selbstkonsistenter Welle-Teilchen Wechselwirkungen in Plasmasystemen unter besonderer Berücksichtigung chaotischer Zustände mit Hilfe einer partiellen kinetisch-numerischen Beschreibung der Teilchenbewegung, Dissertation (B), Greifswald 1987.
[10] S. Kullback, Information Theory and Statistics, John Wiley, New York 1959. – A. Rényi, Wahrscheinlichkeitsrechnung – mit einem Anhang über Informationstheorie, 5. Auflage, Deutscher Verlag der Wissenschaften, Berlin 1977.
[11] A. Wehrl, Rev. Mod. Phys. **50,** 221 (1978).
[12] H. Fujisaka, Prog. Theor. Phys. **70,** 1264 (1983).
[13] J. Kruscha, Limits of the Physical Relevance of Numerical Investigations of the Chaotic Beam Relaxation in an Electron Beam-plasma System, Contrib. Plasma Phys. **27,** No. 6 (1987).
[14] S. Grossmann and S. Thomae, Z. Naturforsch. **32a,** 1353 (1977).
[15] M. M. Bongard, Probleme der Kybernetik **6,** 89 (1966).