Intermittency from Collatz's itineraries and complexity indicators

2002 J. Phys. A: Math. Gen. 35 4501

(http://iopscience.iop.org/0305-4470/35/21/301)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **35** (2002) 4501–4514 PII: S0305-4470(02)32402-8

Intermittency from Collatz's itineraries and complexity indicators

Mario Casartelli

Dipartimento di Fisica dell'Università—Parco Area Scienze 7a, 43100 PR, Italy
and
Istituto Nazionale di Fisica della Materia, Unità di Parma, Italy

E-mail: casartelli@fis.unipr.it

Received 3 January 2002, in final form 12 April 2002
Published 17 May 2002
Online at stacks.iop.org/JPhysA/35/4501

Abstract

We introduce a new type of discrete map on positive reals, showing peculiar kinds of complexity. They combine a large-scale component (Collatz's itineraries on integers) with a small-scale component (standard chaotic systems, such as logistic maps, or random noise). Usual characterizations, for example by Lyapunov exponents, prove senseless or deceptive. Other indicators are used, and in particular a new entropy is introduced to avoid difficulties regarding invariant measures. All computations converge on the conclusion that the crucial point in qualifying the type of complexity and predictability does not consist in the alternative between deterministic perturbation or random noise, but in the peculiar way the large-scale component reacts on the small-scale component, altering its genericity. Possible connections with systems of physical interest are indicated.

PACS numbers: 05.45.−a, 02.50.Ey

1. Introduction

We shall introduce a class of unidimensional discrete maps, whose mathematical interest consists in their different kinds of high complexity. Moreover, they may be useful in exploring the limits of usual characterizations of chaoticity.

With regard to the latter argument, in [1] the following problem was addressed: is it possible, by bare data analysis, to have a definite answer on the type of stochasticity of a data source? Examples considered there indicated that the answer strongly depends on the criteria taken into consideration, for example the observation window. The observer's interest plays a role in distinguishing 'deterministic chaos' from 'genuine randomness' (see also [2] for a review on related arguments). Time series analysis is concerned with large but finite sets of data, while a definitive distinction between chaoticity and randomness typically involves infinity *in actu* (by continuum resolution or by infinite time observations). Attention focused in [1] on the role of 'microscopic chaos', and the length of the sample had in general the expected good property of stabilizing the results. Such stabilization, however, could fail in the presence of (possibly unbounded) intermittency, since the growth of samples may contribute to confuse qualitative differences between deterministic chaos and noise. Our purpose here is to support and enforce this point, stressing the fact that, also for strictly deterministic systems generating arbitrarily long time series, conclusions may be ambiguous forever. Usual indicators, in particular, may reveal themselves to be deceptive.

This will be shown through a new type of map with particular features of high complexity, obtained by the merging of two distinct applications. The first one is the well known logistic map [3], given by

$$F_\mu(x) = \mu x(1-x), \tag{1}$$

which, for $\mu = 4$, is a typical chaotic map $[0, 1] \to [0, 1]$. In fact, it is topologically conjugated to the unilateral shift on binary sequences, a prototype of deterministic chaos [4]. F_4 has been chosen for its simplicity and numerical reliability, but other chaotic maps $[0, 1] \to [0, 1]$ could work equally well.

The second ingredient is reminiscent of a celebrated conjecture in computation, the Collatz problem (also known as the Syracusae, or Hesse, or Kakutani, or Ulam or $3n + 1$ problem). Briefly, it consists in the following: for a natural number $n > 1$, consider the application

$$C_0(n) = \begin{cases} (3n+1)/2 & \text{if } n \text{ is odd;} \\ n/2 & \text{if } n \text{ is even.} \end{cases} \tag{2}$$

The conjecture states that, for every initial n, the trajectory (iterated application of C_0) reaches 1 in a finite number of steps, falling on the cycle 2, 1, 2, 1, 2, 1, ... Equivalently, every trajectory reaches a power of 2. Itineraries form an infinite graph, the Collatz tree, whose branches are all connected to the powers of 2 branch. This statement (a problem considered a hard one by experts) has been successfully tested on more than 10^{18} integers, but a proof is still lacking (for a review see e.g. [5]). Statistical properties of arithmetical sequences variously related to the Collatz itineraries have often been considered (for recent achievements, see e.g. [6]).

Rule (2) and logistic map (1) may be variously combined in order to define a dynamical system on positive reals. Our reference rule will be denoted C_1. For a given initial condition $x(0) > 0$, the sequence $\{x(t)\}$ is defined iteratively by $x(t+1) = C_1(x(t))$ in the following way: let $I(x(t))$ be the integer part of $x(t)$, and $y(t) = x(t) - I(x(t))$. Then,

$$x(t+1) = C_1(x(t)) = \begin{cases} I((3x(t)+1)/2) + F_4(y(t)) & \text{if } I(x(t)) \text{ is odd;} \\ I(x(t)/2) + F_4(y(t)) & \text{if } I(x(t)) \text{ is even.} \end{cases} \tag{3}$$

C_1 does not consist in the simple addition of two maps, since the logistic function alters the even/odd character of the processed entry in a highly irregular way. Note, however, that $\{y(t)\}$ is exactly the logistic trajectory $y(t+1) = F_4(y(t))$, without any feedback from the integer part.

There are two possible trivial situations: (1) if $0 < x(0) < 1$, the rule (3) reduces to the pure logistic map; (2) if $x(0)$ is a natural number, then one obtains the itinerary to the cycle $\{2, 1\}$ (unless the Collatz conjecture is false: but the system is well defined independently of its validity). Hereafter, we shall exclude such special initial conditions by choosing $x(0) > 1$ and not an integer.

In order to give evidence for the specificity of rule (3), consider the following variations:

- *Rule* C_2. Whole quantities $(3x(t) + 1)/2$ and $x(t)/2$ replace the integer parts used in (3) for C_1, altering the recombination of noninteger parts at the next step. C_2 can check the

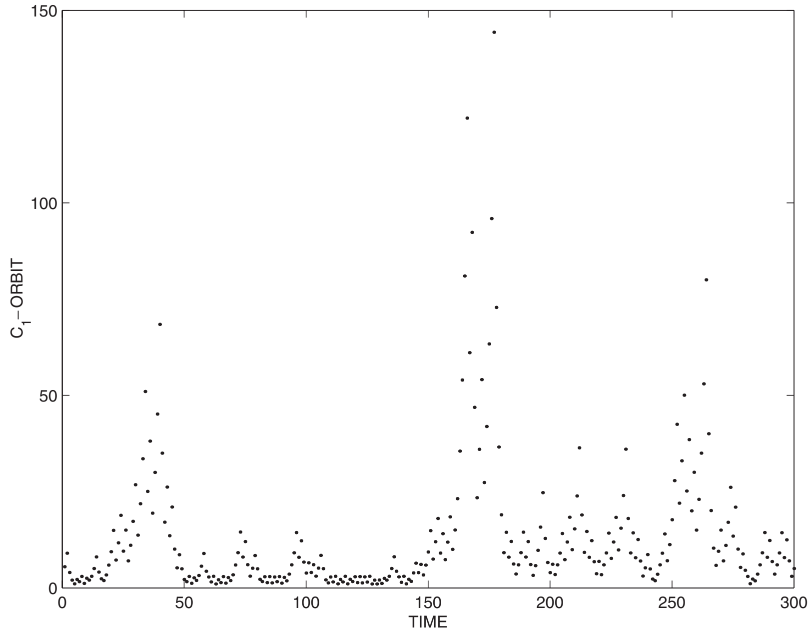

Figure 1. Example of C_1 orbit; first 300 steps.

peculiarity of the reference rule (3), which apparently should be simpler, since this time $\{y(t)\}$ does not reduce to the logistic orbit. Such an irregular recombination could be a source of further chaos.

- *Rule* R_1. The contribution $F_4(y(t))$ in (3)) for C_1 is replaced by 'stochastic noise', for example a uniform random variable $r(t) \in [0, 1]$. This variation should check the influence of a strictly deterministic versus a stochastic rule.
- *Rule* R_2. Analogously, $F_4(y(t))$ in C_2 is replaced by random noise $r(t)$.

These four rules may be classified along two different criteria: (a) deterministic versus random character of the small-scale component (i.e. C_1 and C_2 versus R_1 and R_2) or (b) existence or non-existence of a feedback from the large- to the small-scale component. In this second respect, C_2 is in a distinct class, being the only rule where such a feedback exists.

2. Orbits, densities, dimensions

Starting with C_1, an example of the trajectory is shown in figures 1–3 with different details. In figure 4, a semilog representation of a very long segment shows its great irregularity. Empirically, one sees that trajectories have intermittent peaks, whose maximal height grows irregularly with the length. There is no evidence even for boundedness.

Trajectories $\{C_2(x(t))\}$ (figures 5 and 6) present intermittency in turn but, surprisingly enough, they are much more regular than $\{C_1(x(t))\}$: the range is restricted (in no case did we exceed the value 1000) and concentrated on certain non-uniform attracting regions, which however are not separate attractors (i.e. dynamically invariant distinct sets). Most of the time the trajectory is <10, and a density function would be concentrated in this low range.

We omit figures for R_1 and R_2, which on visual inspection essentially behave as C_1, suggesting that criterion (b) is the effective one.

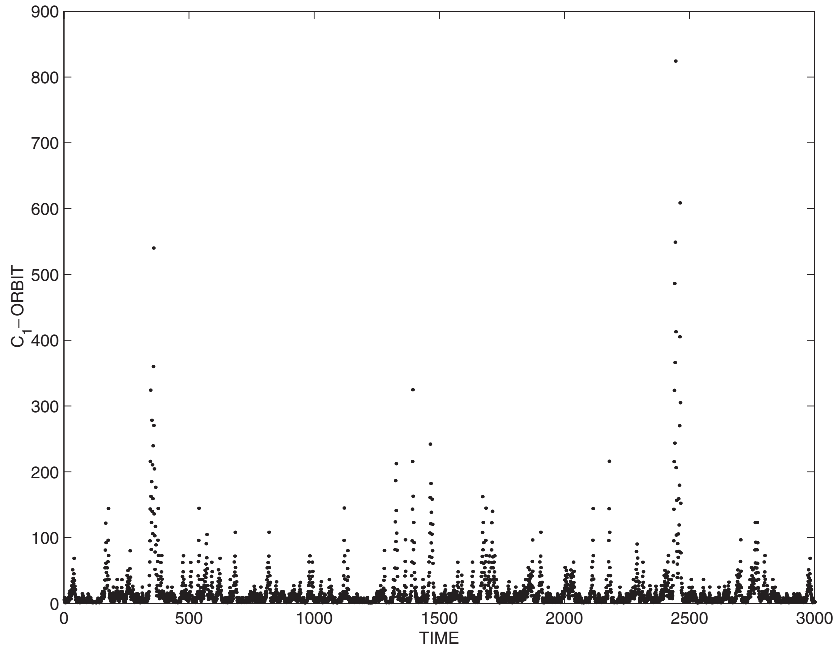

Figure 2. Same C_1 orbit as figure 1 up to 3000 steps.

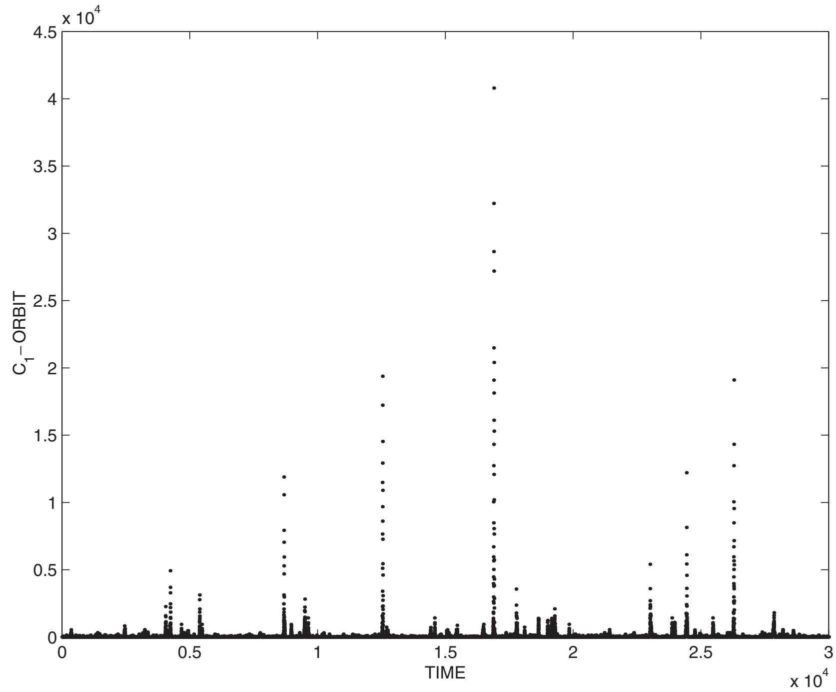

Figure 3. Same C_1 orbit as figure 1 up to 30 000 steps.

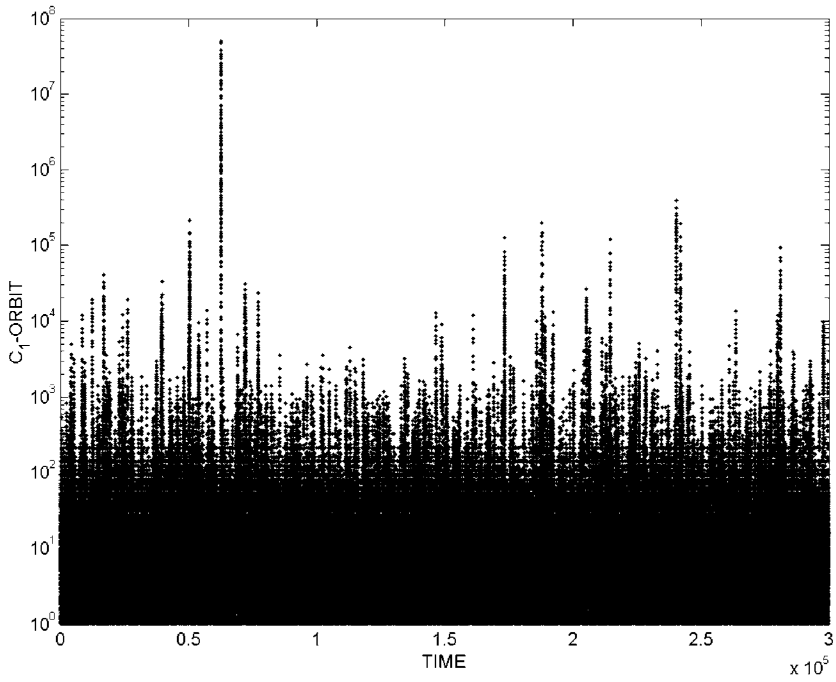

Figure 4. Same C_1 orbit as figure 1 up to 300 000 steps in semilog representation.

The 'natural' invariant measure could be established numerically by averaging densities for a great number of long trajectories. In the range (the 'phase space' of the map) a coarse density would be given by the number of occurrences in each fixed interval, i.e. by the normalized occupation time. In principle, this computation presents some difficulties, especially for C_1, R_1 and R_2: indeed, since the range is not *a priori* established, one cannot know when the orbit's length (for a defined minimum size ε of intervals) is sufficiently long. Different orbits imply averaging over different ranges, and in a ε partition the microscopic scale parameter ε cannot be compared with the natural scale of the range: a deep difference with respect to maps with a known bounded phase space. Also usual relations between such quantities as Lyapunov exponents (LEs) and dynamical or Kolmogorov–Sinai (KS) entropy are not obvious [7].

In practice, on the numerical side, the situation does not appear so desperate: densities are very peaked in the lower part of the range, where they stabilize rather quickly and *independently of initial conditions*, while for high values, and especially for C_1, R_1 and R_2, intermittency makes the stabilization very slow. In short, an *effective* invariant measure seems to exist and to be computable.

This holds in particular for C_2, since $x(t) < 10$ in $95.35 \pm 0.1\%$ and $x(t) < 100$ in $99.97 \pm 0.01\%$ of occurrences (these numbers implicitly estimate the measure of the neglected phase space). By refining the ε partition of the range, the stabilization is obviously slower and slower. Note that coarse densities are referred to the ε intervals, but they are not expected to be continuous with respect to the Lebesgue measure in the attracting regions. In other words, the nature of the attracting regions does not emerge from this numerical approach (even for the C_2 rule, where computations are easier).

All these considerations are summarized in figure 7, showing, at increasing time, the probabilities of having peaks, i.e. pieces of orbits exceeding a certain threshold. Actually, for

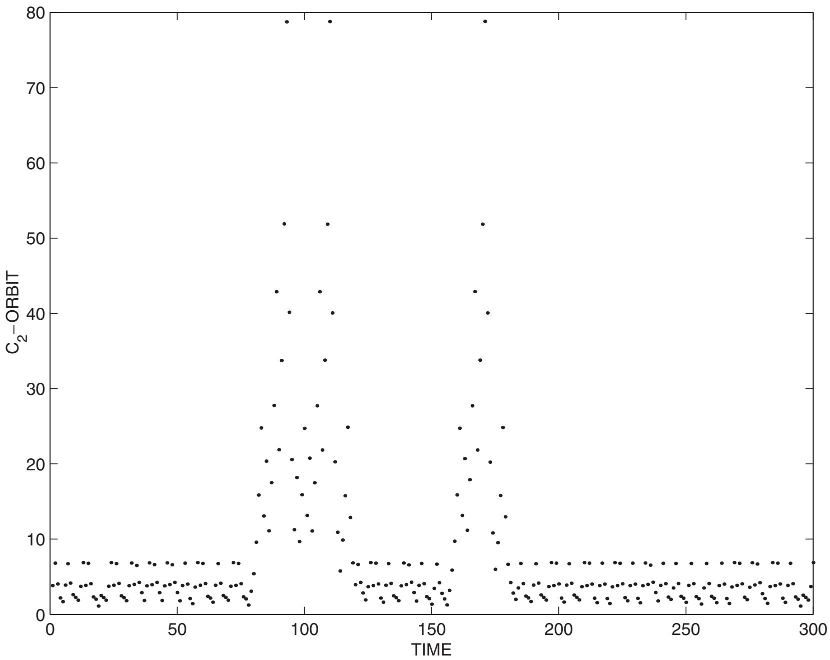

Figure 5. Example of C_2 orbit; first 300 steps.

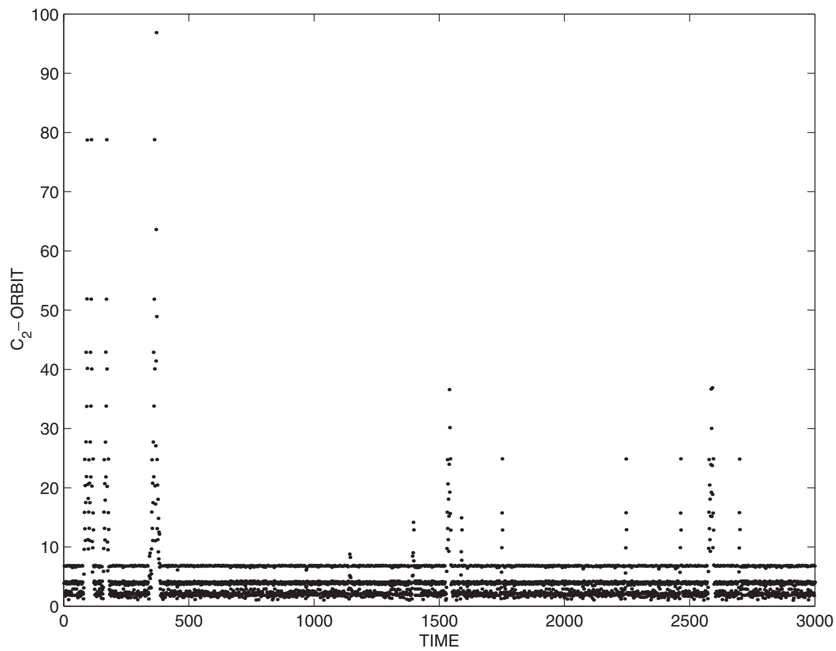

Figure 6. Same C_2 orbit as figure 5; first 3000 steps.

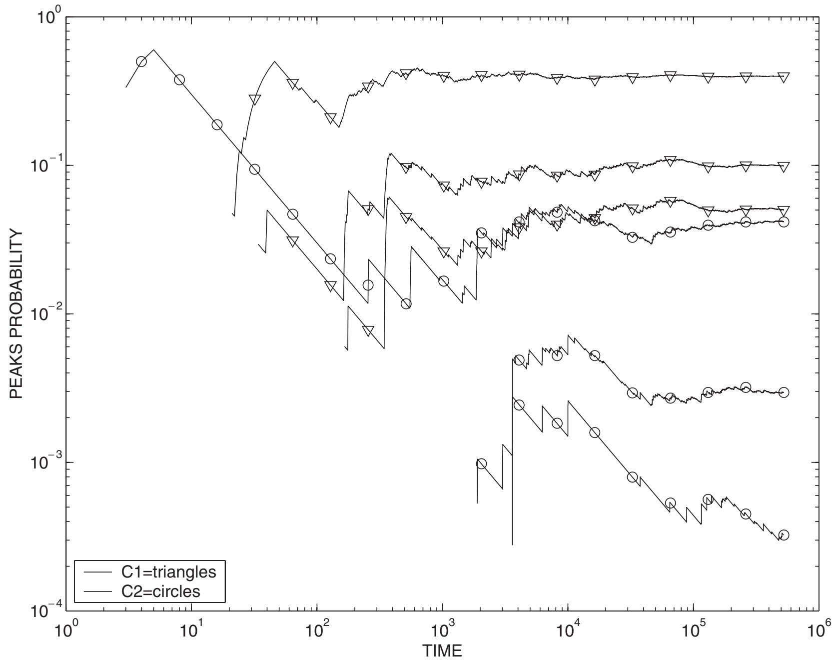

Figure 7. Probabilities, at increasing time, of being over a certain height for C_1 (triangles) and C_2 (circles). Considered cut-off thresholds are, from above, 10, 50 and 100 for both systems.

both C_1 and C_2, the cut-off thresholds have been positioned at heights $10, 50, 100$. These probabilities are therefore also proportional to the neglected phase space, confirming our estimates. In principle, cut-offs may be read as an intentional distinction between spiking and non-spiking regions, following possible resemblances with other intermittent phenomena. However, non-spiking regions have different features for C_1 and C_2: the former (see figures 1–3) indeed presents a rough and non-homogeneous self-similarity, preventing the identification of low-valued regions with a sort of 'laminar' regime interrupted by chaotic intermittencies. Possibly, such an identification could work with C_2.

Such a geometrical difference should be reflected by dimensional properties of curves. Due to intermittency and possible unboundedness, however, the box counting, entropy and correlation dimensions are also generally unstable and difficult to calculate. What may be easily done is the covering of trajectories by smaller and smaller 'rods' (the smallest being 1), allowing for the computation of a finite-size covering dimension. We are aware that this is a rough estimate, because of the discreteness of the curve associated with the time series. Nevertheless, it turns out to be efficient in distinguishing between C_1 and C_2. More precisely, let $\{x(t)\}, t = 0, 1, 2, \ldots, T$ be the time series. Consider a finite rod of decreasing length $s \geqslant 1$ (in our computations $s = 512, 256, \ldots, 4, 2, 1$). The sampled time series $\{X(k)\}$ is defined by $X(k) = x(ks), k = 0, 1, \ldots, m$, where the sampling number m is $m(s) = T/s$. Let $N(s)$ indicate the number of rods necessary to cover the Euclidean length of the curve associated with the sampled time series, i.e.

$$N(s) = \frac{\sum_1^m \sqrt{(X_k - X_{k-1})^2 + s^2}}{s}. \tag{4}$$

Then, an index of the curve variation at sampling s is the exponent $D = D(s)$ defined by $N(s) = m^D$, or

$$D(s) = \frac{\log N(s)}{\log m(s)} \tag{5}$$

as $s \to 1$. Results are deeply different for C_1, R_1 and R_2 on one hand, and C_2 on the other. Precisely, in the tested range $2^{15} \leqslant T \leqslant 2^{19}$:

- For C_2, $D(s)$ is independent of T and initial conditions, varying from $D(512) = 1$ to $D(1) = 1.08 \pm 0.005$.
- For C_1, on the contrary, $D(s)$ strongly depends on T, on initial conditions and also on the considered segment of the same orbit: $D(512) = 1 \pm 0.05$, but $D(1)$ oscillates from 1.20 to 1.85! Of course, one could suspect that $T = 2^{19}$ is still too short; and nevertheless, possible unboundedness suggests that these oscillations, certainly due to the presence of very high peaks, could prove irreducible. Independently of this, the numerical distinction between C_1 and C_2 is striking.
- As to R_1 and R_2, they essentially behave as C_1.

Comments given above about the 'non-laminar' character of C_1 also for low values are therefore confirmed by these dimensional calculations, suggesting (1) that unbounded intermittency makes C_1 a system intrinsically more complex than C_2 and (2) that the distinction between stochastic and deterministic noise is much less important than the coupling rule between large-scale itineraries and small-scale components.

3. Lyapunov exponents, entropy, complexity

We must remark that peaks explode and decay exponentially. This feature (especially when combined with possible unboundedness) has two conflicting effects: it leaves the density associated with peaks negligible, but has an heavy influence on orbits' correlation. The first effect is responsible for the good convergence of densities in the low range, but the second effect says that KS entropy associated with these densities and LEs are not particularly meaningful as indicators of true dynamical properties. Consider, for definiteness, a main ingredient in the construction of the KS entropy, i.e. 'words': such words are strings of symbols (each of them associated with the partition cells) in the order in which these cells are visited during the evolution. The probabilistic weight of 'strange words' corresponding to peaks is negligible, because, over a certain height, orbits are extremely rarefied (as shown by figure 7); nevertheless, as peaks are relatively abundant, 'strange words' are also such. We shall see later how to take into account this behaviour with another entropy functional.

Independently of relations to entropy, it is an instructive exercise to proceed heuristically with the evaluation of the LE, with the usual method [8]. For C_1 and for all initial conditions, averaging over some thousands of 'free flights' is sufficient to obtain LE $= 0.6931\ldots \approx \log(2)$, i.e. the exact LE for F_4. This may be easily understood: when the free flight τ of the second orbit is shorter than the correlation time τ_0 (as necessary to have stable results), the only effective dynamics between the orbits is due to the logistic component. (Choosing an initial distance of 10^{-7}, the correlation time τ_0 is about 20 steps.) For the same reason, however, the independence of initial conditions is a misleading indication: in itself, it cannot guarantee any good 'mixing' property of trajectories over long but finite times. The computation method simply ignores the integer part of orbits, which however contribute heavily to the evolution and correlation properties of the system. This implies that the usual LE has no real meaning for C_1, in accordance with previous observations about invariant measures. For the same reason, the 'finite-size' LE, introduced in [9] and successfully used in [1], cannot say much more either: if the doubling time is within the correlation time, one practically deals with F_4.

(Abusing the method, for curiosity: with longer free flights, when τ is slightly greater than τ_0 the exponent increases, and this is possibly due to the influence of the integer part, but for even greater τ it decreases, as expected, because of the saturation of distance in the mean.)

As for C_2, we obtain LE $\approx 0.43\ldots$. This confirms that, on the small scale, the recombined non-integer parts produce a dynamical system completely different from F_4. It is chaotic too, but more 'ordered' than the logistic map: this is due to the presence of attracting regions concentrated on low values.

These results on LEs, in addition to computational and conceptual problems about invariant measures, discourage us from evaluating the usual KS, Renyi and block entropies, as defined for example in [1] (for a general review, see [10]). Therefore, in order to understand the 'information content' of our trajectories, we shall follow another way, by introducing a new entropy function which jumps over all these difficulties. For a trajectory segment of total length T, consider the partition $\{0 \equiv t_0, t_1, \ldots, t_N \equiv T\}$ of T, where the t_k correspond to extremal points, i.e. times when the sign of $x(t+1) - x(t)$ changes. This partition is further weighted with the oscillation $|x(t_k) - x(t_{k-1})|$. The total weight of the kth semi-oscillation, corresponding to a sort of 'wave energy', is therefore

$$W_k = (t_k - t_{k-1})|x(t_k) - x(t_{k-1})|.$$

Then, a probability distribution $\boldsymbol{p} \equiv \{p_1, \ldots, p_N\}$ may be associated with the trajectory by

$$p_k = W_k \Big/ \sum_{j=1}^{N} W_j \tag{6}$$

and $S(\boldsymbol{p}) \equiv S(\boldsymbol{p}, T)$ will denote the usual Shannon entropy of this distribution:

$$S(\boldsymbol{p}) = -\sum_{k=1}^{N} p_k \log(p_k). \tag{7}$$

Expectations about a correct entropy function can be checked on typical situations (e.g. regular oscillations or irregularly peaked oscillations). Obviously, one has to be careful in using naively the term 'complexity' associated with $S(\boldsymbol{p})$, since, for instance, a regular oscillation has maximal entropy, but the adaptation of terms and concepts is quite immediate.

Note the peculiarity of (7): whereas one usually starts with an ε partition in phase space and estimates the rate of entropy increase for 'words' generated by the trajectory on this partition, here we have to do with a *time* partition, which is intrinsically defined by dynamical properties (change of behaviour), the cell probability being the time interval weighted with the oscillation within. If usual entropy corresponds to the mean information needed to locate the trajectory in the phase space, here we have the information needed to characterize the oscillations pattern in time. In a complexity estimate, these points of view are not exclusive but complementary. The advantages of entropy (7) are:

(1) we may ignore the problem of an invariant measure in the phase space;
(2) we may ignore the problem of 'generating partitions';
(3) while the finite size ε of a cell partition in phase space is an artificial parameter (requiring in principle a limit procedure), the discreteness of the oscillation partition is a meaningful intrinsic feature of the process;
(4) the computation is extremely easy.

Generalizations to piecewise continuous functions are almost immediate. Also for systems with multidimensional phase space generalizations are conceivable (and this would constitute a further advantage, since 'time' to be partitioned is intrinsically one dimensional).

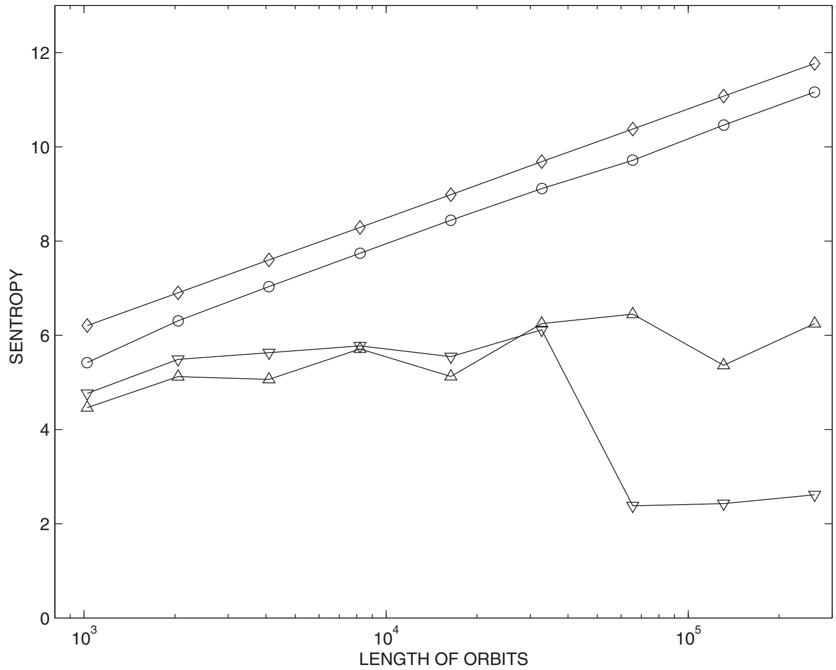

Figure 8. *S* entropy defined in formula (7) as a function of the trajectory length (in number of steps). From above: a random noise (diamonds), a C_2 orbit (circles) and two different C_1 orbits (triangles up and down).

More than actual values of this entropy, it is interesting to follow its behaviour as a function of the total length T. It is obvious (and easy to check by numerical experiments) that for random sources there is a logarithmic growth. As figure 8 shows, this type of growth holds indeed for C_2, independently of initial conditions, at least for sufficiently long trajectories (i.e. $T > 4000$, which is a short time). This means that the small-scale chaoticity, as already measured by the LE, is not substantially influenced by the C_2 intermittency. In contrast, for C_1, the growth is very irregular and heavily dependent on initial conditions. Once again this is due to intermittent high peaks, which irregularly alter the relative weight of the curve behaviour in different trajectory segments. Such a necessity for reiterated adjustments in the observation scale (sensitively registered by $S(\boldsymbol{p})$) may be taken as a significant index of complexity. Moreover, because of the range of the main oscillations, the presence of random or deterministic noise on the small scale proves irrelevant.

Note that our entropy is lower for C_1 than for C_2, while the small-scale chaoticity, as measured by the LE, is higher. Such indications are not contradictory, since for these maps:

(1) LEs are sensitive only to local divergence and therefore are always stable at growing T, while $S(\boldsymbol{p}, T)$ is a global indicator;
(2) high entropy, as previously recalled, is not automatically associated with complexity but with a certain uniformity of behaviour in time (chaotic or not).

These indications should make clear how the initial questions about chaoticity, complexity, predictability etc are intertwined with the type of interest the observer has. We can take the results exhibited in figure 8 as a particularly clear indication of the distinction between chaoticity, complexity and unpredictability.

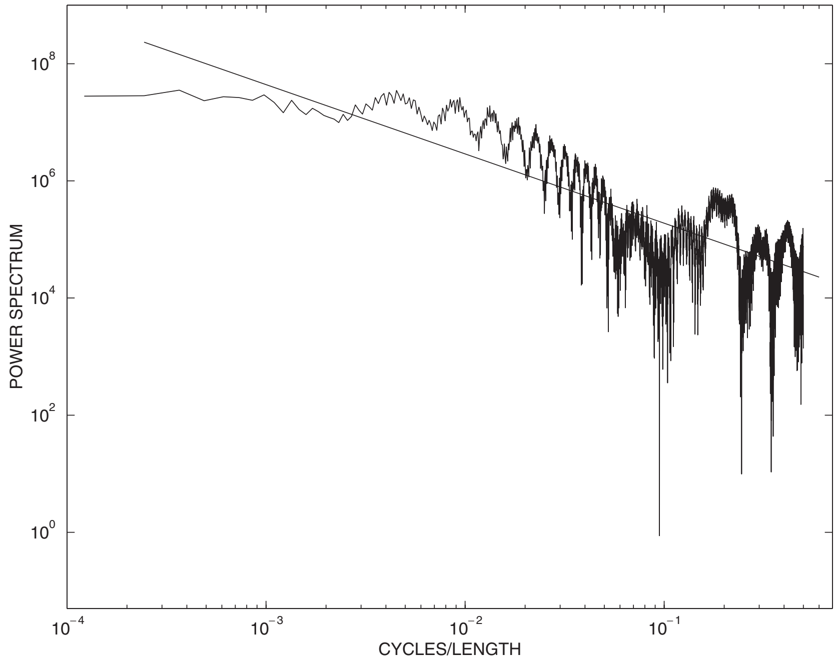

Figure 9. Power spectrum obtained via fast Fourier transform on a segment of C_1 orbit (the same as figure 4). The length is 8192 steps. The slope of the linear interpolation of data is -1.1832.

Power spectra give further information confirming the pattern of results above. For all systems, the prevalent spectra are of the coloured noise type (the so-called $1/\omega^\alpha$ noise). We recall that spectra of this type occur in a great variety of physical phenomena generally classified as 'complex' (see e.g. [11], and the final paragraph of the next section). However, as far as we know, for such noise there is no universal model or explication. Moreover, for C_1, R_1 and R_2, we can only speak of *prevalent* $1/\omega^\alpha$ behaviour, since not only may the exponent α be different for different initial conditions, but there are also frequent spectacular exceptions and differences for different pieces of the same orbit, such as those exhibited in figures 9–11 (only figure 11 is typically $1/\omega^\alpha$). Note, in the linear interpolations of these figures, the large range of exponents α. There is a good correspondence between such fluctuations in the spectral features and the sensitivity to initial conditions already registered by entropy (7) or by covering dimension.

Once again, C_2 proves much more regular, with a $1/\omega^\alpha$ spectrum which, in the whole set of initial conditions we explored, has $\alpha = 1.093 \pm 4\%$. Figures are omitted, being qualitatively identical to figure 11.

4. Final remarks and conclusion

All indicators are coherent in separating C_1, R_1 and R_2 from C_2, while the distinction between random or deterministic small-scale component (noise) seems to be irrelevant. These two groups correspond to distinct classes of complexity, whereas for both of them usual chaoticity, as indicated by the LE, is only an index of mean local behaviour. From experimental evidence, we have frequently argued that the main point qualifying the first class is the unboundedness–intermittency connection. Is it possible to say something more precise?

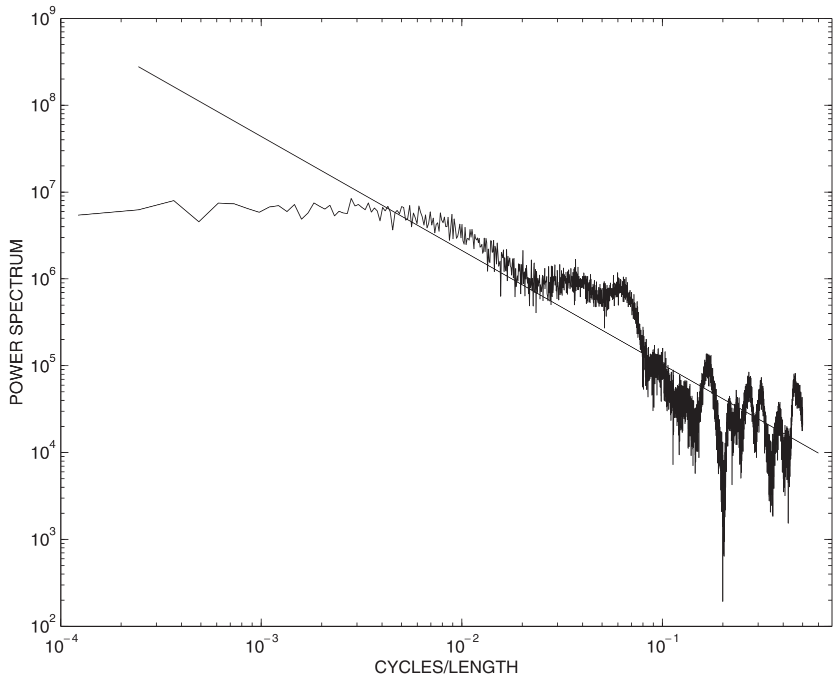

Figure 10. Power spectrum and linear interpolation for the next segment of the same orbit as figure 9. The slope is -1.3124.

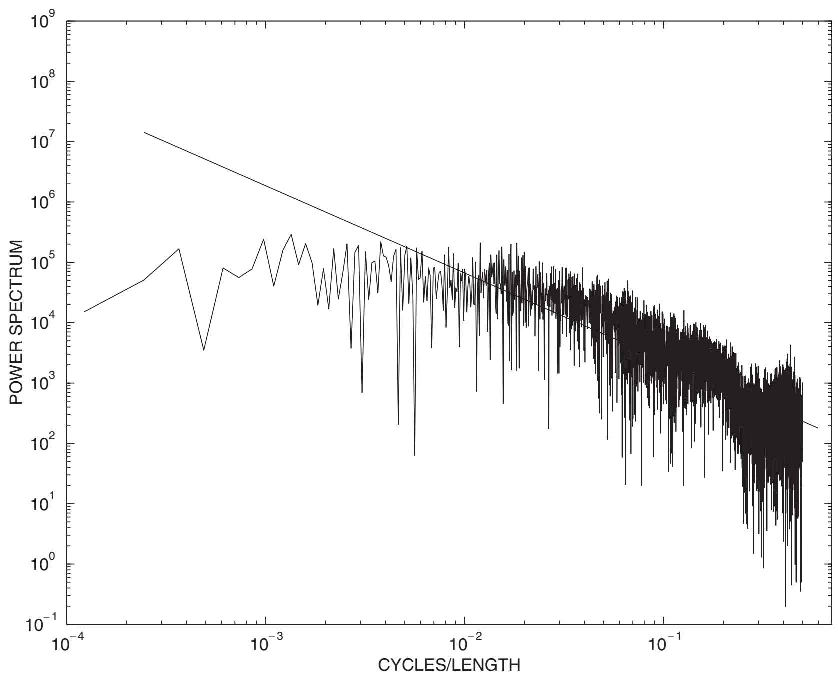

Figure 11. Power spectrum and linear interpolation for the next segment of the same orbit as figure 10. The slope is -1.4467.

For the unboundedness of C_1, we have no formal proof. However, a plausibility argument could run as follows: F_4 is topologically conjugated to the shift on binary sequences (see [4]); this means that for every binary sequence $\underline{s} = (s_0, s_1, s_2, \ldots)$ there exists y_0 such that

$$F_4^n(y_0) = y_n \in J_{s_n},$$

where $J_{s_n} = [0, 1/2)$ if $s_n = 0$ and $J_{s_n} = [1/2, 1)$ if $s_n = 1$. It is therefore possible to choose $x_0 = I(x_0) + y_0$ in such a way that the evolution of the non-integer part (that coincides with the logistic evolution) is smaller or larger than $1/2$, allowing for the correct influence on the integer part in (3). This is not rigorous, because the fact of being smaller or larger than $1/2$ does not ensure complete control of the integers' evolution; in a statistical sense, however, it should work. In this sense, it is plausible that, for every M, there are such initial conditions that the $x(t) > M$ for some t. Note that, at fixed M, a binary sequence with the required control properties is also finite, so that the previous argument, if true, holds for uncountably many initial conditions, probably with positive measure. This explains why, in our numerical experiments, very high peaks are visible.

As to R_1 and R_2, the randomness of $y(t)$ at every step is sufficient to ensure the existence of 'luck' sequences of non-integer contributions, allowing for unboundedness. The aperiodic appearance of such luck sequences in generic reals, although different in origin for C_1, R_1 and R_2, seems to be a good hint at the observed intermittency of these systems.

Clearly, such an argument does not work for C_2: the recombination of non-integer parts, indeed, distinguishes $\{y(t)\}$ from the logistic trajectory, destroying both equivalence to the shift and genericity on reals. Appearance of aperiodic privileged regions in the C_2 range is symptomatic for this highly non-generic recombination (and may be that these 'strange' regions deserve further mathematical attention). Once again, these considerations are not a proof that unbounded trajectories cannot also exist for C_2; only, it is plausible that such trajectories have probability zero, or at least that they are substantially less than in previous cases.

A simple experiment may test the correctness of our argument: consider the C_1 rule (3) where, instead of F_4, there is F_μ, with μ around the value of the Feigenbaum constant [3]. Then, since the non-integer part is now concentrated on a strange attractor, C_1 should recover, at least qualitatively, the behaviour of C_2: and this is precisely what happens.

In conclusion, we are aware that the maps introduced here are not physical but mathematical speculative examples. However, their analysis may prove instructive for more general dynamical systems. All this phenomenology seems indeed to depend on the following points:

(a) there is an unbounded system coupled with a small-scale noise or an external perturbation influencing its local behaviour;
(b) the large-scale system has a set of basins (in the present case the branches of the Collatz graph) very sensitive to initial conditions, and therefore to noise or perturbations;
(c) the decisive feature in qualifying the class of complexity is not the random/deterministic alternative, but the form of coupling between large and small scales (more precisely the presence/absence of a feedback altering the genericity of the perturbation, or the continuity of the associated density with respect to the reals).

This characterization of our results suggests an interesting link with sandpile models, where avalanches play the role of Collatz's itineraries, and the random addition of sand grains is the perturbation shifting from one basin to another. Actually, the $1/\omega^\alpha$ behaviour observed in the Bak–Tang–Wiesenfeld and Manna models [12] is compatible with our C_2 spectra. It may be that, in order to obtain also the higher C_1 complexity, one should elude the bounds imposed on avalanches by the finiteness of lattice size.

Acknowledgments

I thank A Vulpiani and one of the referees for their comments on an earlier version of this paper.

References

[1] Cencini M, Falcioni M, Olbrich E, Kantz H and Vulpiani A 2000 *Phys. Rev.* E **62** 427
[2] Boffetta G, Cencini M, Falcioni M and Vulpiani A 2002 *Phys. Rep.* **356** 367
[3] Collet P and Eckmann J P 1980 *Iterated Maps on the Interval as Dynamical Systems* (Boston, MA: Birkhauser)
[4] Devaney R L 1986 *An Introduction to Chaotic Dynamical Systems* (Menlo Park, CA: Benjamin–Cummings)
[5] Lagarias J 1985 *Am. Math.* **92** 3; see also http://www.cecm.sfu.ca/organics/papers/lagarias/
[6] Sinai Ya G 2002 Statistical $(3N + 1)$-problem *Preprint* math.DS/0201102
Rouet J L and Feix M R 2002 *J. Stat. Phys.* **107** 1283
[7] Pesin Ya B 1977 *Russ. Math. Surv.* **32** 55
Sinai Ya G 1991 *Dynamical Systems* vol 1 (Singapore: World Scientific) (reprint)
[8] Benettin G, Galgani L and Strelcyn J M 1976 *Phys. Rev.* A **14** 2338
Lichtenberg A J and Liebermann M A 1983 *Regular and Stochastic Motion* (New York: Springer)
[9] Aurell E, Boffetta G, Crisanti A, Paladin G and Vulpiani A 1996 *Phys. Rev. Lett.* **77** 1262
Aurell E, Boffetta G, Crisanti A, Paladin G and Vulpiani A 1997 *J. Phys. A: Math. Gen.* **30** 1
[10] Badii R and Politi A 1997 *Complexity* (Cambridge: Cambridge University Press)
[11] Press W H 1978 *Comments Mod. Phys.* C **7** 103
Dutta P and Horn P M 1981 *Rev. Mod. Phys.* **53** 497
[12] Casartelli M and Zerbini M 2000 *J. Phys. A: Math. Gen.* **33** 863