FUNDAMENTAL PROBLEMS IN STATISTICAL MECHANICS VII
H. van Beijeren, Editor

SCALING BEHAVIOR & THERMODYNAMICS[1]

J.J.P. Veerman[2]
M.J. Feigenbaum

The Rockefeller University
1230 York Avenue
New York, New York 10021

I INTRODUCTION

It is our purpose in this review to explain the thermodynamic formalism, in particular its application to dynamical systems where scaling behavior is present. While intended for students in statistical physics, it is our hope that the material will also be accessible for a somewhat wider audience. We have therefore included some calculations that are standard in statistical physics.

The general setting of the problem is the following: We are given a Cantor set S (in one dimension) generated by some underlying dynamical procedure. In some generality, this means that set S is generated in successive approximations in a level-by-level fashion similar to the way in which one constructs a middle third Cantor set. In sections 2 and 3, we will discuss these recursive procedures in some detail. Those who appreciate a very informal introduction to the notions of Cantor set and dimension can find it in Mandelbrot (1982).

Since most of the sets so constructed have zero measure, their Hausdorff dimension (section 4) is of interest. However, the central issue here is: what are the characteristic properties of the recursive procedure by which the set is naturally generated?

This is of importance for the following reason. In doing an experiment (numerical or otherwise) that creates the Cantor set from an otherwise unknown dynamical system, one has two pieces of information. The first is an approximation of the Cantor set, the second, more important one is the behavior of successive approximations. More detailed information may be hard or impossible to obtain. While this information is not sufficient to completely characterize the Cantor set, it gives a decidedly more detailed description of the Cantor set than just the Hausdorff dimension (see section 5).

It will turn out that the Cantor set is approximated by an ever-growing number of smaller and smaller intervals. The ratios of the length of an interval in one level to that of an interval in the

---

[1] Based on the 1989 NUFFIC Summer School lectures by M.J. Feigenbaum

[2] Current address: Institute for the Mathematical Sciences
SUNY at Stony Brook, New York 11794

previous level are called scalings. The partial characterization we seek is the precise analysis of statistics on the sizes of the scalings (sections 6 and 7). This analysis turns out to be the same as thermodynamics in classical statistical mechanics.

To illustrate the use of this, we will in sections 8, 9 and 10 treat a rather complicated Cantor set which plays an important role in the study of one dimensional dynamical systems. We will set up and solve for an approximate thermodynamics associated with the recursive scheme that generates the set. The notion of phase transition will be discussed with this context as an example.

II TREES

As a first example, we will construct the so-called two scale Cantor set without memory. Throughout this work we will denote Cantor sets by S.

Start with an interval $\Delta \subset \mathbb{R}$, normalized to have length one. So Δ may be taken to be the interval $[0, 1]$. Split this interval in two pieces $\Delta(0)$ and $\Delta(1)$, each of which contains one of the endpoints of Δ. Denote the length of an interval I by $|I|$. The ratios

$$\sigma(\varepsilon_1) = \frac{|\Delta(\varepsilon_1)|}{|\Delta|}, \quad \varepsilon_1 \in \{0,1\}$$

are called "scalings." Clearly,

$$\sigma(0) + \sigma(1) \leq 1 .$$

At the next level of resolution, each of the intervals $\Delta(\varepsilon_1)$ is split in two. We obtain intervals $\Delta(\varepsilon_2\varepsilon_1)$ with the following properties. The four endpoints of $\Delta(\varepsilon_1)$ are also endpoints of the four intervals $\Delta(\varepsilon_2\varepsilon_1)$. For the lengths of the intervals one has

$$|\Delta(\varepsilon_2\varepsilon_1)| = \sigma(\varepsilon_2)\,|\Delta(\varepsilon_1)| . \tag{2.1}$$

If one continues the construction of the intervals in the n-th approximation, their lengths satisfy

$$|\Delta(\varepsilon_n \ldots \varepsilon_1)| = \sigma(\varepsilon_n)\sigma(\varepsilon_{n-1})\ldots\sigma(\varepsilon_1)\,|\Delta| .$$

The set consisting of the 2^n intervals at the n-th level will be called S_n. The Cantor set S consists of those points which are elements of S_n for each n:

$$S = \bigcap_{n>0} S_n .$$

A convenient way of keeping track of the lengths of the intervals $\Delta(\varepsilon_n \ldots \varepsilon_1)$ and their ordering in S_n is to draw a binary tree T. Each node of T has a "0" – and a "1" – branch emanating from it. The nodes to which one assigns the lengths of the intervals $\Delta(\varepsilon_n \ldots \varepsilon_1)$ have been labeled in such a way that the zeros and the ones appear in the order $\varepsilon_n \ldots \varepsilon_1$ when one travels from the node to the top of the tree (see figure 2.1). Notice that T completely determines S.

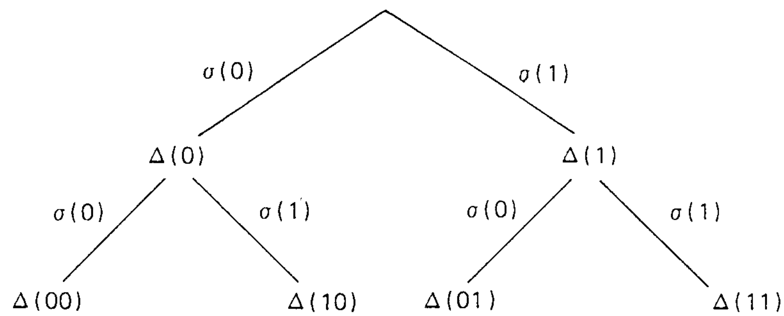


figure 2.1

There are obvious generalizations of the above construction. (First of all, we could have split each interval into $k > 2$ pieces.) We would then get a k-nary tree (each node has k branches). One could also consider sets described by "pruned" k-nary trees. For example if in figure 2.1, consecutive "1" branches are forbidden, one obtains figure 2.2.

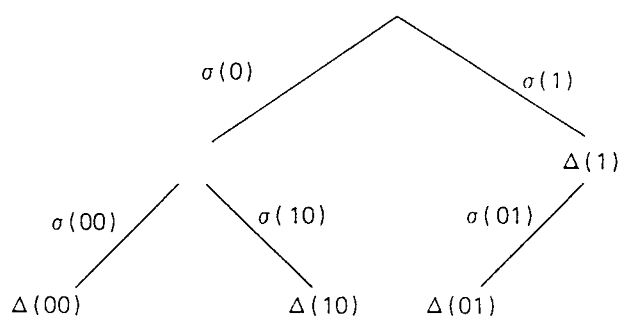


figure 2.2

While scalings without memory serve the useful purpose of giving insight into the combinatorics associated with the trees, for the interesting sets occurring in dynamical systems one has to define scalings more generally:

$$\sigma(\varepsilon_n \ldots \varepsilon_1) = \frac{|\Delta(\varepsilon_n \ldots \varepsilon_1)|}{|\Delta(\varepsilon_{n-1} \ldots \varepsilon_1)|} \tag{2.3}$$

The scaling now depends on the entire path to the node rather than only on the last node. In figure 2.3, we have drawn the tree for the case that the scalings depend on the last two branches leading to the node. However, we will adhere to the convention that the combinatorics of the tree is without memory: the way new branches emanating from a certain branch are labelled depends only on the label of the current branch. This means in particular that there are only four binary trees of this type, because after the level nodes on the first level the choices are exhausted. Notice that these choices determine the spatial ordering on the line of all intervals $\Delta(\varepsilon_n \ldots \varepsilon_1)$. Furthermore, it implies that two nodes are nearby on the tree, when the low epsilons agree (i.e., ε_1, ε_2, etc.). It is clear, then, that the scalings plus the spatial ordering of $\Delta(\varepsilon_n \ldots \varepsilon_1)$ determine the Cantor set.

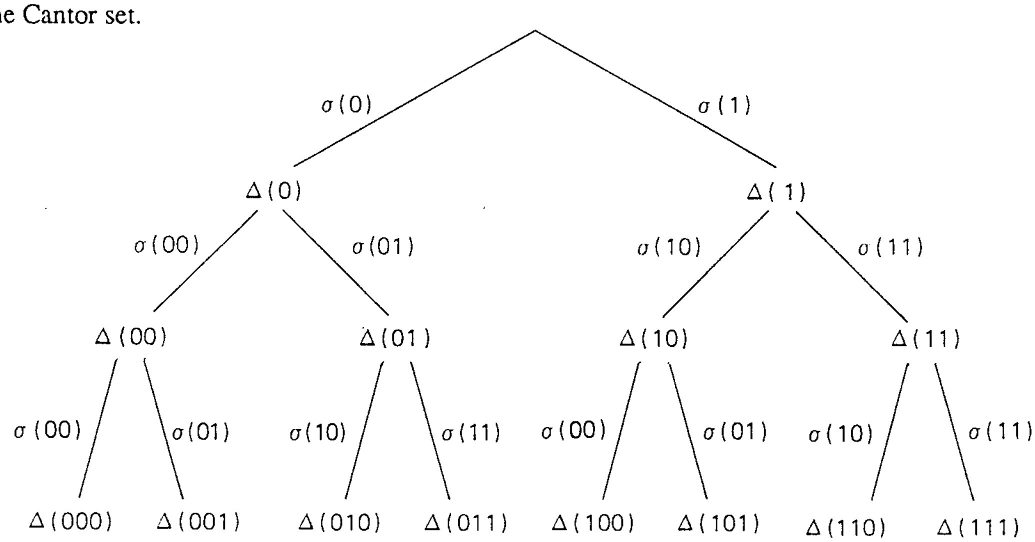


figure 2.3

For the scalings we have the result that they are asymptotically invariant under smooth coordinate transformations. The latter is easily understood. A smooth invertible coordinate transformation restricted to a very small interval is asymptotically affine (i.e., has constant derivative). A high level scaling is the ratio of the lengths of a very small interval and a subinterval of it. Thus, the asymptotic geometry of the Cantor set is completely determined by invariant quantities that are the limit points of the scalings (together with the ordering). The set of limit points is called the scaling function $\sigma(x)$ and can be defined as follows. Let a be the branching ratio of the tree. Then

$$x = \sum_0^\infty \frac{\varepsilon_{n-1}}{a^{i+1}} \Rightarrow \sigma(x) = \lim_n \sigma(\varepsilon_n \varepsilon_{n-1} \ldots \varepsilon_1).$$

The purpose of this labeling scheme is to do it in such a way that the dependence of the scaling upon lower ε's falls off fast. We will argue later that this decay is exponential. The scaling function is then a well-defined notion (the limit exists). Moreover, the fast decay means that the scaling function is very well approximated by $\sigma(\varepsilon_n \ldots \varepsilon_{n-k})$ for some finite k. Hence the importance of figure 2.3.

In order to get the ε-dependence to fall off fast, one may have to change the ordering properties of the intervals. The way it is done in figure 2.4 will be important throughout this work. The first two branches are as usual. The rest of the tree can be constructed recursively by noting that the subtree starting at the node "1" and the entire tree are each others' mirror image.

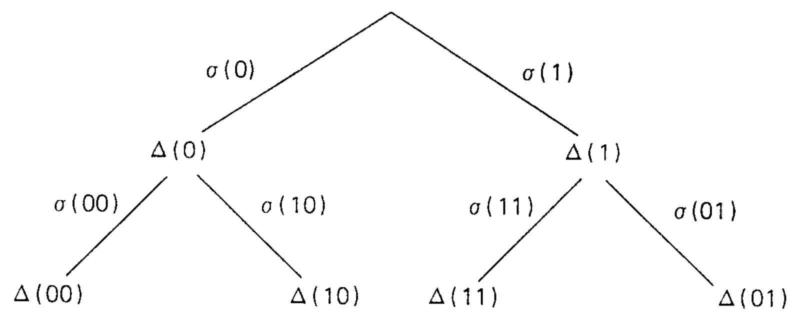


figure 2.4

As remarked before, the tree determines the Cantor set. The converse is not true. One makes the observation that the set determined by the tree of figure 2.4 could also have been generated by the ternary tree of figure 2.5. The first level of this tree consists of the nodes labeled 0, 11, and 01 of figure 2.4. One skips the nodes that end on 10. The trees themselves are equivalent to the set of scalings plus the ordering properties of the intervals.

So, if the Cantor set does not determine a unique set of scalings, what is the reason we are interested in scalings? This is easily answered. The sets we study come from an underlying dynamical system. This provides us with extra information. Imagine, for example, that there is an orbit that slowly fills out the entire Cantor set. Points on that orbit have a temporal ordering (as well as a spatial ordering).

A natural way to select a tree (and one that is often used) is to insure that each node at level n of the tree corresponds to an interval that is visited equally often (namely with frequency $1/N_n$) by orbits that trace out the Cantor set. It will now be clear that the trees of figure 2.4 and 2.5 cannot both satisfy that requirement, because an interval in level n of figure 2.5 may correspond to an interval in different levels in figure 2.4. Of the class of trees with equal probability in each interval one selects the simplest.

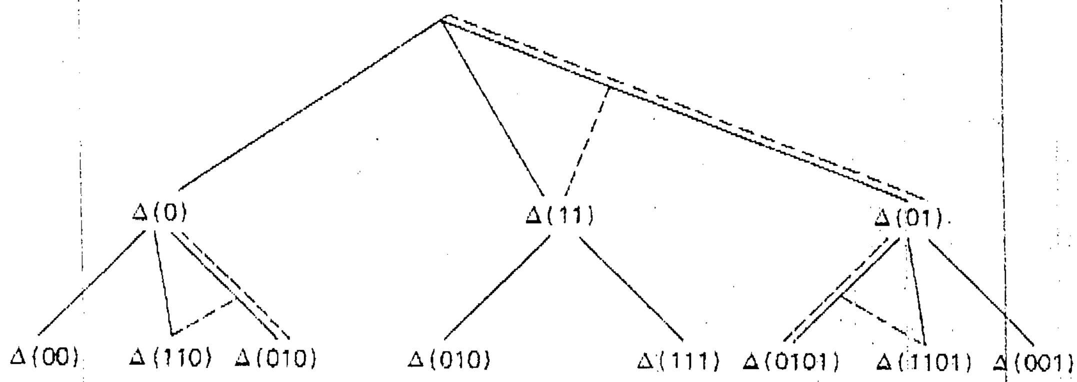


figure 2.5

In conclusion: it is crucial to realize that the sets we are interested in are associated with a dynamical process. The dynamical process selects a tree.

III PRESENTATION FUNCTIONS

There is a more revealing way to present the tree structure of the Cantor sets discussed in the previous section. To this end we construct a map F_ε for each direction in which the tree can branch:

$$F_\varepsilon(\Delta(\varepsilon_n \dots \varepsilon_2)) = \Delta(\varepsilon_n \dots \varepsilon_2\varepsilon) \tag{3.1.a}$$

$$\Delta(\varepsilon_n \dots \varepsilon_1) = F_{\varepsilon_1} \circ F_{\varepsilon_2} \circ \dots \circ F_{\varepsilon_n}(\Delta) . \tag{3.1.b}$$

Let γ and γ^1 be distinct ε-sequences of the same length with $\gamma > \gamma^1$ in the ordering discussed in section 2. Then, for a given tree, the ordering of $\gamma\varepsilon$ and $\gamma^1\varepsilon$ depends only on ε. Therefore the F_ε's are monotone functions from Δ into itself. As a consequence the inverses F_ε^{-1} of the functions F_ε form branches of a single function from Δ into itself. We will name this function the 'presentation function' (Feigenbaum, 1988).

The case we will most thoroughly study is the one drawn in figure 3.1 with F_1 orientation reversing and F_0 orientation preserving. This corresponds to the tree exhibited in figure 2.4.

[Note: strictly speaking, we have defined F_ε only on the endpoints of the intervals $\Delta(\varepsilon_n \dots \varepsilon_1)$. These points are dense in the Cantor set S. Since two nearby nodes x and x′ in the tree are mapped to nearby nodes $\alpha\varepsilon$ and $\beta\varepsilon$ by F_ε, one may extend F_ε to continuous functions on S. The details of F_ε outside the Cantor set will play no role in the analysis to come, and one may just interpolate to get functions defined on Δ.]

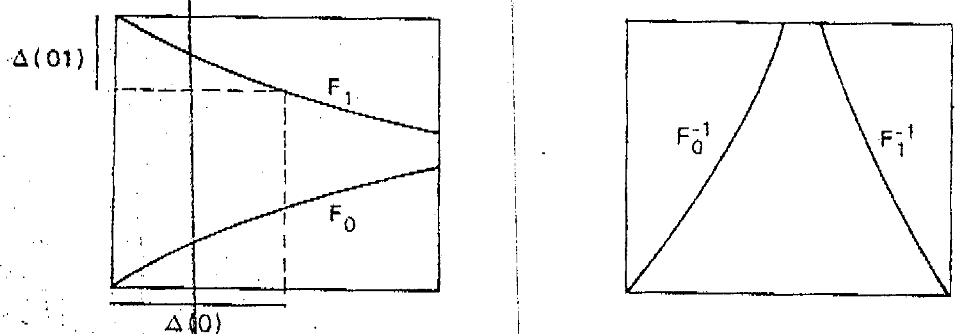


figure 3.1

Putting together equations (2.3) and (3.1), we obtain:

$$\sigma(\varepsilon_n \cdots \varepsilon_1) = \frac{| F_{\varepsilon_1} \circ \cdots F_{\varepsilon_{n-1}}(F_{\varepsilon_n}(\Delta)) |}{| F_{\varepsilon_1} \circ \cdots F_{\varepsilon_{n-1}}(\Delta) |} \tag{3.2}$$

It is also clear from (3.1) that the constructed intervals have the ordering of the nodes on the tree. So the presentation function and the tree are equivalent notions (given one of them, one constructs the other).

We will now show that the exponential decay of the dependence of σ on the ε's is implied by the smoothness of the F_ε (Sullivan, 1988). Since this decay is an assumption that enters at every stage of the theory, it is of paramount importance to understand the smoothness of F_ε.

Assume that the F_ε are contractions and the derivatives are α Holder continuous for some α, that is: there are α, λ, C with

$$0 < 1 - \lambda \le F_\varepsilon'(x) \le \lambda \tag{3.3}$$

and

$$0 < \frac{F_\varepsilon'(\xi) - F_\varepsilon'(\eta)}{|\xi - \eta|^\alpha} < C, \quad 0 < \alpha < 1 \tag{3.4}$$

for all x, ξ and η in Δ. From equation (3.2) one obtains

$$\sigma(\varepsilon_n \ldots \varepsilon_1) = \sigma(\varepsilon_n \ldots \varepsilon_{k+1}) \frac{F_{\varepsilon_1} \ldots F_{\varepsilon_k}(F_{\varepsilon_{k+1}} \ldots F_{\varepsilon_n}\Delta)}{F_{\varepsilon_{k+1}} \ldots F_{\varepsilon_n}\Delta} \frac{F_{\varepsilon_{k+1}} \ldots F_{\varepsilon_{n-1}}\Delta}{F_{\varepsilon_1} \ldots F_{\varepsilon_k}(F_{\varepsilon_{k+1}} \ldots F_{\varepsilon_{n+1}}\Delta)}.$$

Divide by $\sigma(\varepsilon_n \dots \varepsilon_{k+1})$. The right hand side now consists of two terms, which by the mean value theorem are equal to derivatives in some points $\xi_1 \in \Delta(\varepsilon_n \dots \varepsilon_{k+1})$ resp. $\eta_1 \in \Delta(\varepsilon_{n-1} \dots \varepsilon_{k+1})$. Subtracting 1 from each side, one gets:

$$\frac{\sigma(\varepsilon_n \dots \varepsilon_1) - \sigma(\varepsilon_n \dots \varepsilon_{k+1})}{\sigma(\varepsilon_n \dots \varepsilon_{k+1})} = \frac{\frac{d}{dx}\{F_{\varepsilon_1} \dots F_{\varepsilon_k}(\xi_1)\} - \frac{d}{dx}\{F_{\varepsilon_1} \dots F_{\varepsilon_k}(\eta_1)\}}{\frac{d}{dx}\{F_{\varepsilon_1} \dots F_{\varepsilon_k}(\eta_1)\}} .$$

Now set

$$\xi_{i+2} = F_{\varepsilon_{k-i}}(\xi_{i+1})$$
$$\eta_{i+2} = F_{\varepsilon_{k-i}}(\eta_{i+1}),$$

and use the chain rule to obtain

$$\frac{\sigma(\varepsilon_n \dots \varepsilon_1) - \sigma(\varepsilon_n \dots \varepsilon_{k+1})}{\sigma(\varepsilon_n \dots \varepsilon_{k+1})} = \frac{\prod_{i=0}^{k-1} F'_{\varepsilon_{k-i}}(\xi_i) - \prod_{i=0}^{k-1} F'_{\varepsilon_{k-i}}(\eta_i)}{\prod_{i=0}^{k-1} F'_{\varepsilon_{k-i}}(\eta_i)} .$$

Since ξ_i and η_i both are in $\Delta(\varepsilon_n \dots \varepsilon_{k-i})$, by using (3.3) we have

$$F'_{\varepsilon_{k-i}}(\xi_i) = F'_{\varepsilon_{k-i}}(\eta_i) + k_i \lambda^{\alpha(n-k+i)}$$

where $|k_i|$ is bounded by C. Therefore,

$$\left| \frac{\sigma(\varepsilon_n \dots \varepsilon_1) - \sigma(\varepsilon_n \dots \varepsilon_{k+1})}{\sigma(\varepsilon_n \dots \varepsilon_{k+1})} \right| \leq \frac{K\lambda^{\alpha(n-k)}}{1-\lambda} = K\,\beta^{n-k}, \qquad (3.5)$$

for some $\beta < 1$.

IV HAUSDORFF DIMENSION

The sets we consider often have the property that at each level a definite fraction of the length is taken out of the interval. The total length of the remaining intervals at level n then approaches zero. These sets have zero Lebesgue measure. There is a property that can distinguish such sets, namely the Hausdorff dimension.

At this point we should warn the reader that there are many different notions of dimensionality. These notions are certainly not equivalent in general. In spite of its somewhat cumbersome definition, the Hausdorff dimension has in the context of this work convenient properties. For other purposes (e.g., numerical work) other notions may be more fit.

We follow the definitions of Falconer (1985). A δ-cover of a set S is a (countable) collection of intervals $U_i(\delta)$ with length $|U_i(\delta)| \leq \delta$. The β-measure $H^\beta(\delta)$ of S is given by

$$H^\beta(S) = \lim_{\delta\to 0} \inf_{\delta\text{-covers}} \sum_{i=0}^{\infty} |U_i(\delta)|^\beta \quad . \tag{4.1}$$

Here, the infimum is taken over all possible countable collections $\{U_i(\delta)\}_{i=1}^{\infty}$ that cover S. According to this definition, one first covers the set S in an optimal way with intervals that are not bigger than δ. Then one takes a limit as $\delta \to 0$. The Hausdorff dimension is an invariant of the Cantor set (under smooth coordinate transformations).

The β-measure has the following property. There is a number d (the Hausdorff dimension of the set S) such that

$$\begin{aligned} 0 \leq \beta < d \;&\Rightarrow\; H^\beta(S) = \infty \\ \beta > d \;&\Rightarrow\; H^\beta(S) = 0\,. \end{aligned} \tag{4.2}$$

To illustrate these concepts, we will now calculate the Hausdorff dimension of the two scale Cantor set S (figure 2.1). The Hausdorff dimension will be denoted by HD.

We can get an upper estimate of $H^\beta(S)$ easily. Cover S with any convenient cover, such as the intervals $\Delta(\varepsilon_n \ldots \varepsilon_1)$ of the n-th level of construction of the set S. These form a δ-cover of S where

$$\delta = \max_{\{\varepsilon\}_n} |\,\Delta(\varepsilon_n \ldots \varepsilon_1)\,|\,.$$

(In fact, these sets are a closed, not open, cover of S. But by replacing them with open sets $\tilde{\Delta}(\varepsilon_n \ldots \varepsilon_1)$ ever so slightly larger, the calculation would not be changed.) Thus

$$H^\beta(S) \leq \lim_{n\to\infty} \sum_{\{\varepsilon\}_n} |\,\Delta(\varepsilon_n \ldots \varepsilon_1)\,|^\beta = \lim_{n\to\infty} \sum_{\{\varepsilon\}_n} [\sigma(\varepsilon_1)\ldots\sigma(\varepsilon_1)]^\beta = \lim_{n\to\infty} \sum_{\varepsilon} [\sigma(\varepsilon)^\beta]^n\,.$$

Therefore, if d satisfies

$$\sum \sigma(\varepsilon)^d = 1\,, \tag{4.3}$$

then

$$\mathrm{HD}(S) \geq d.$$

To show that HD(S) is equal to d, one would have to show that for some $C > 0$

$$H^d(S) \geq C \lim_{n\to\infty} \sum_{\{\varepsilon\}_n} |\,\Delta(\varepsilon_n \ldots \varepsilon_1)\,|^{\mathrm{HD}(S)}\,.$$

Equation (4.3) then implies that

$$HD(S) = d. \tag{4.4}$$

This is much harder and we will not pursue this any further. A more general statement can be found in Ruelle (1988): if the F_ε satisfy (3.3) and (3.4) and if T is a finitely pruned k-nary tree, then the HD(S) is given by:

$$HD(S) = \lim_{n\to\infty} d_n$$

where

$$\sum_{\{\varepsilon\}_n} |\Delta(\varepsilon_n \dots \varepsilon_1)|^{d_n} = 1.$$

(Finitely pruned means: the pruning of a branch depends on finitely many, say N, previous branches.)

For the two scale Cantor set with equal scaling one obtains

$$2\sigma^{HD(S)} = 1 \Rightarrow HD(S) = -\frac{\ln 2}{\ln \sigma}.$$

V A TANTALIZING EXAMPLE

In this section we will study the statistics of the scalings on the n-th level of construction of the Cantor set. We will do so from first principles by means of an example, the two scale Cantor set. The general theory will be set forth in the next section. Recall from the first two chapters that the scalings and therefore all the statistics associated with them are characteristics of the presentation function rather than of the set itself.

For the Cantor set S one has (figure 2.1)

$$|\Delta(\varepsilon_{2n} \dots \varepsilon_1)| = \sigma(0)^{n-i}\, \sigma(1)^{n+i}, \tag{5.1}$$

where $n - i$ of the ε's are zero, and the rest are ones. The set of arguments of Δ consists of all binary words of length 2n. So:

$$\sum |\Delta(\varepsilon_{2n} \dots \varepsilon_1)|^\beta = \sum_{i=-n}^{n} \binom{2n}{n+i} \sigma(0)^{(n-i)\beta}\, \sigma(1)^{(n+i)\beta} \tag{5.2}$$

or

$$\sum |\Delta(\varepsilon_{2n} \dots \varepsilon_1)|^\beta = (\sigma(0)^\beta + \sigma(1)^\beta)^{2n}. \tag{5.3}$$

The ratio of the (i + 1)-st and the i-th term in (5.2) is given by

$$\frac{C_{i+1}}{C_i} = \frac{n-i}{n+i+1}\left[\frac{\sigma(1)}{\sigma(0)}\right]^{\beta} \cong \frac{1-i/n}{1+i/n}\left[\frac{\sigma(1)}{\sigma(0)}\right]^{\beta} .$$

It is easy to see that the contribution to the sum in (5.2) is maximized when this ratio equals one, or

$$\frac{i}{n} = \frac{\left[\frac{\sigma(1)}{\sigma(0)}\right]^{\beta} - 1}{\left[\frac{\sigma(1)}{\sigma(0)}\right]^{\beta} + 1} . \tag{5.4}$$

Now assume n to be very large. It is clear that for n large enough the contributions of $\frac{i}{n} \in \{[-1, -1+\varepsilon] \cup [1-\varepsilon, 1]\}$ are negligible if ε is small. In fact, the contribution from those intervals is drowned in the contribution of the terms $\frac{i}{n} = -1+\varepsilon$ and $\frac{i}{n} = 1-\varepsilon$, as one concludes readily from (5.2).

Of course, there are many terms for which i/n is approximately constant. It seems handy to replace i by a new variable $\mu = i/n$ (or some affine function of it). Since the factorials are not very convenient for this, we apply Stirling's formula first. (Notice that this is justified by our remark that contributions of the tail are vanishingly small.)

Stirling's formula gives:

$$\sum_{\{\varepsilon\}2n} |\Delta(\varepsilon_{2n} \dots \varepsilon_1)|^{\beta}) = \sum_{i=-n(1-\varepsilon)}^{n(1-\varepsilon)} \frac{(2n)^{1/2}}{(2\pi)^{1/2}(n+i)^{1/2}(n-i)^{1/2}} \exp[2n \ln 2n - (n+i)\ln(n+i) - (n-i)\ln(n-i)] \cdot \sigma(0)^{(n-i)\beta}\, \sigma(1)^{(n+i)\beta}$$

where we have neglected terms of order $\frac{1}{\varepsilon n}$ for fixed ε. With the new variable μ this becomes

$$\dots = \sum_{\mu=-1+\varepsilon}^{1-\varepsilon} \left[\frac{1}{\pi n(1-\mu)(1+\mu)}\right]^{1/2} \exp\Big[2n\,\{\ln 2 - \tfrac{1}{2}(1+\mu)\ln(1+\mu) - \tfrac{1}{2}(1-\mu)\ln(1-\mu) + \tfrac{1}{2}\beta \ln \sigma(0)\,\sigma(1) + \tfrac{1}{2}\mu\beta \ln \frac{\sigma(1)}{\sigma(0)}\}\Big] . \tag{5.5}$$

With obvious changes of notation this will be rewritten as

$$\dots = \sum_{\mu=-1+\varepsilon}^{1-\varepsilon} C(\mu)\, e^{2n\{s(\mu)-\beta(c_1\mu + c_2)\}} . \tag{5.6}$$

Since μ increments in steps of $\frac{1}{n}$, this can be written as an integral (neglecting order $\frac{1}{n}$ terms):

$$\dots = \int n\, d\mu\, C(\mu)\, e^{2n\{s(\mu)-\beta(c_1\mu + c_2)\}} . \tag{5.7}$$

As shown before,

$$-X(\mu, \beta) = s(\mu) - \beta(c_1\mu + c_2)$$

has a unique maximum (see (5.4)) at $\mu = \bar{\mu}$. Further:

$$-\frac{d^2}{d\mu^2} X \Big|_{\mu=\bar{\mu}} = \frac{-1}{(1+\bar{\mu})(1-\bar{\mu})} < 0, \tag{5.8}$$

so that the maximum is non degenerate. This implies that one may use steepest descent methods to evaluate the integral (5.7).

$$\int n\, d\mu\, C(\mu)\, e^{-2nX(\mu,\beta)} = n\, C(\bar{\mu})\, e^{-2nX(\bar{\mu},\beta)} \int d\mu\, e^{-n(\mu-\bar{\mu})^2\, \partial^2 X}$$

$$= n\, C(\bar{\mu}) \left[\frac{\pi}{n\partial_\mu^2 X(\bar{\mu}, \beta)} \right]^{1/2} e^{-2nX(\bar{\mu}, \beta)} . \tag{5.9}$$

In the last term we have again (in applying the steepest descent method) neglected $\frac{1}{n}$ terms. Comparison with (5.8), (5.4) and the definition of $C(\mu)$ shows that all prefactors cancel. One finally obtains (replacing $2n$ by n)

$$\sum |\Delta(\varepsilon_n \dots \varepsilon_1)|^\beta) = e^{-nX(\bar{\mu}, \beta)} . \tag{5.10}$$

(As an exercise, the readers should convince themselves that when one works out what $X(\bar{\mu}, \beta)$ is using equations (5.4) and (5.5), one gets equation (5.3) back.)

This calculation yields

$$X(\bar{\mu}, \beta = HD) = 0,$$

if one combines (5.10) and the results in section 4. This equation then determines the Hausdorff dimension of the set S.

Here then is the first step of the statistics of the problem. The largest single contribution to the sum in (5.2) comes from the terms with $\frac{i}{n} = \bar{\mu}$. There are $C(\bar{\mu})e^{ns(\bar{\mu})}$ of these intervals and each has length $e^{-n(c_1\bar{\mu} + c_2)}$ (i.e., average scaling is $e^{-(c_1\bar{\mu} + c_2)}$). If we add up nearby ($\frac{i}{n} \equiv \bar{\mu}$) contributions, they come from intervals with smaller length or have lengths that occur less abundantly. From (5.10) one concludes that, on the average, one can say that the sum is dominated by $e^{ns(\bar{\mu})}$ intervals with average scaling $e^{-(c_1\bar{\mu} + c_2)}$.

Suppose we were interested in the set $S_n(\mu_0, \delta) \subset S_n$ of intervals which have an average scaling of $e^{-(c_1\bar{\mu} + c_2)}$ where $\mu \in [\mu_0 - \delta, \mu_0 + \delta]$ for some positive, small δ. We will denote the sum over these intervals by $\sum_0$. It is worth pointing out that $|\mu_0|$ should not be taken too close to one, since that would preclude our applying Stirling's formula in what follows: the whole reasoning described below fails miserably if the number of pieces $\Delta(\varepsilon_{2n} \dots \varepsilon_1)$ with scaling μ_0 is not large. This is important to bear in mind when doing numerical or real life experiments.

As before, Stirling's approximation gives equation (5.5) with the summation now running over $[\mu_0 - \delta, \mu_0 + \delta]$. Expressing this as an integral as in (5.7) yields

$$\sum\nolimits_0 |\Delta(\varepsilon_n \dots \varepsilon_1)|^\beta = \int_{\mu_0-\delta}^{\mu_0+\delta} n\, d\mu\, C(\mu)\, e^{ns(\mu)-n\beta(c_1\mu+c_2)} . \tag{5.11}$$

Clearly, then, $S_n(\mu_0, \delta)$ dominates the above sum if β is chosen to be equal to β_0 with

$$\mu_0 = \bar{\mu}(\beta_0),$$

in which case one obtains, in agreement with (5.10),

$$\sum\nolimits_0 |\Delta(\varepsilon_n \dots \varepsilon_1)|^{\beta_0} = e^{-nX(\bar{\mu},\beta_0)} .$$

For $\beta \neq \beta_0$ one clearly has

$$\mu_0 \neq \bar{\mu}(\beta)$$

That is: for small enough δ the maximum of $s(\mu) - \beta(c_1\mu + c_2)$ lies outside the interval of integration of (5.11). To evaluate the integral in this case, set $\mu = \mu_0$ and integrate:

$$\sum\nolimits_0 |\Delta(\varepsilon_n \dots \varepsilon_1)|^\beta = 2\delta n\, C(\mu_0)\, e^{ns(\mu_0)-n\beta(c_1\mu_0+c_2)+\mathcal{O}(n\delta)} . \tag{5.12}$$

The exponential error here comes from the fact that

$$\frac{1}{2\delta e^{-nx}} \int_{x-\delta}^{x+\delta} e^{-n\mu}\, d\mu = \frac{e^{+n\delta} - e^{-n\delta}}{2\delta} = e^{\mathcal{O}(n\delta)} .$$

In analogy with the definition of the Hausdorff dimension, we now define the exponent f as follows:

$$\begin{aligned} &\text{for all } \beta < f \quad \lim_{n\to\infty} \sum\nolimits_0 |\Delta(\varepsilon_n \dots \varepsilon_1)|^\beta = \infty \\ &\text{for all } \beta > f \quad \lim_{n\to\infty} \sum\nolimits_0 |\Delta(\varepsilon_n \dots \varepsilon_1)|^\beta = 0 . \end{aligned} \tag{5.13}$$

Then, of course, (5.12) implies

$$s(\mu_0) - f(c_1\mu_0 + c_2) = 0$$

where

$$\mu_0 = \bar{\mu}(\beta_0)$$

Thus

$$f = \frac{s(\bar{\mu}(\beta_0))}{c_1\bar{\mu}(\beta_0) + c_2}. \tag{5.14}$$

This formula gives the relation between the exponent β_0 such that the μ_0 scalings dominate the sum $\sum |\Delta|^{\beta_0}$ and the exponent f for which the sum $\sum |\Delta|^f$ crosses over as defined by (5.13).

Recall that each elementary interval has a 'probability' N_n^{-1} associated with it. The length of the 'average' interval that dominates $\sum |\Delta|^{\beta_0}$ is $N_n^{-(c_1\bar{\mu} + c_2)}$. Define the exponent α as:

$$N_n^{-1/\alpha} = N_n^{-(c_1\bar{\mu} + c_2)}$$

or

$$\alpha = \frac{1}{c_1\bar{\mu} + c_2}. \tag{5.15}$$

Then f can be expressed as a function of α (provided $c_1 \neq 0$):

$$f(\alpha) = \alpha \cdot s\left(\frac{\alpha^{-1} - c_2}{c_1}\right).$$

The definition of s in (5.6) gives:

$$f(\alpha) = \alpha \left\{\ln 2 - \tfrac{1}{2}\left(1 + \frac{\alpha^{-1} - c_2}{c_1}\right) \ln\left(1 + \frac{\alpha^{-1} - c_2}{c_1}\right) - \tfrac{1}{2}\left(1 - \frac{\alpha^{-1} - c_2}{c_1}\right) \ln\left(1 - \frac{\alpha^{-1} - c_2}{c_1}\right)\right\}.$$

Note the word "analogy" in the definition (5.13) of f. If certain conditions are met, one can prove (Bohr and Rand, 1987) that

$$S(\mu_0) = \{\text{all } x \in \Delta \mid X \text{ has average scaling } e^{-(c_1\bar{\mu} + c_2)}\}$$

has Hausdorff dimension $f(\alpha)$ if $\alpha = (c_1\mu_0 + c_2)^{-1}$. Notice that $S(\mu_0)$ is a strange set. Since every point in X can be arbitrarily well approximated by points in $S(\mu_0)$, the closure of $S(\mu_0)$ is S. Furthermore, there are points in S that do not have an average scaling (that is, the limit does not converge). So $S \setminus \bigcup_{\mu} S(\mu) \neq \emptyset$.

VI THERMODYNAMICS (CANONICAL AND MICROCANONICAL ENSEMBLES)

The reader with some training in statistical physics will undoubtedly have recognized the last section as a treatment of the microcanonical ensemble. We will now set forth the general principles behind this.

While we are in the position that we can apply mathematical notions borrowed from statistical physics, the actual 'physics' cannot be transplanted. The difference with the statistical physics of spin systems is that there the 'tangible' objects are the actual spins. Here the spins (or scalings) arise from the presentation function which in turn is generated by an underlying dynamical process. Thus the 'tangible' object, the Cantor set, does not itself determine the scalings (see section 2).

With this in mind we may call $X(\bar{\mu},\beta)$ of equation (5.10) the "free energy" although it is not an energy in the physical sense of the word. To render the description of the theory transparent we will write $F(\beta)$ for that quantity. Similarly we will replace $c_1\mu + c_2$ of section 5 by μ.

Write the number of pieces $\Delta(\varepsilon_n \ldots \varepsilon_1)$ at n-th level of construction of S as

$$N_n \cong a^n \tag{6.1}$$

where

$$a = \lim_{n\to\infty} \frac{1}{n} \ln N_n .$$

Let

$$|\Delta(\varepsilon_n \ldots \varepsilon_1)| = N_n^{-h(\varepsilon_n \ldots \varepsilon_1)} . \tag{6.2}$$

Then

$$\sum_{\{\varepsilon\}} |\Delta(\varepsilon_n \ldots \varepsilon_1)|^{\beta} = \sum_{\{\varepsilon\}} N_n^{-\beta h(\varepsilon_n \ldots \varepsilon_1)}$$

$$= \sum_{\mu>0} N_n^{-\beta\mu} \sum_{\substack{\{\varepsilon\} \\ h(\varepsilon_n \ldots \varepsilon_1)=\mu}} 1 = \sum_{\mu>0} N_n^{-\beta\mu + s(\mu)} . \tag{6.3}$$

This equation defines μ as the value of $h(\varepsilon_n \ldots \varepsilon_1)$ or the "microcanonical energy." The quantity $e^{ns(\mu)}$ equals the cardinality of the configurations with $h = \mu$, that is to say: $s(\mu)$ is the "entropy."

In the previous section we showed for the two scale Cantor set $s(\mu)$ has negative second derivative. (This can be generalized to all finitely pruned k-nary trees satisfying (3.3) and (3.4), see Ruelle (1978). Thus $s(\mu) - \mu\beta$ has a non-degenerate maximum at $\mu = \bar{\mu}$ As in the previous section, the contribution of the maximum dominates the sum if n is very large:

$$\frac{\partial}{\partial\mu}\Big|_{\mu=\bar{\mu}} \{s(\mu) - \mu\,\beta\} = 0$$

$$\frac{\partial^2}{\partial\mu^2}\Big|_{\mu=\bar{\mu}} \{s(\mu) - \mu\,\beta\} < 0 \tag{6.4}$$

$$\sum_{\{\varepsilon\}} |\Delta(\varepsilon_n \dots \varepsilon_1)|^\beta \cong a^{n(s(\bar{\mu})-\beta(\mu))} \qquad (6.5)$$

(note that $\bar{\mu}$ depends on β). We can also regard

$$\sum_{\{\varepsilon\}} |\Delta(\varepsilon_n \dots \varepsilon_1)|^\beta = \sum_{\{\varepsilon\}} a^{-\beta n h(\varepsilon_n \dots \varepsilon_1)} = a^{-nF(\beta)} \qquad (6.6)$$

as a canonical ensemble with $F(\beta)$ the "free energy per particle" (sometimes called "pressure" in the mathematics literature). Therefore, we have

$$F(\beta) = \beta\bar{\mu}(\beta) - s(\bar{\mu}(\beta)) \qquad (6.7)$$

which is the equivalent of (5.10).

We note that by (6.4) $F(\beta)$ is precisely the Legendre transform of the entropy. Therefore $F(\beta)$ satisfies

$$\begin{aligned} F''(\beta) &< 0 \\ F'(\beta) &= \bar{\mu}(\beta) > 0. \end{aligned} \qquad (6.8)$$

We will use the same definitions for f and α as we used before in (5.13) and (5.15). So

$$f = s(\bar{\mu}) / \bar{\mu}$$

$$\alpha = 1/\bar{\mu} \qquad (6.9)$$

By equation (6.8) $F(\beta)$ is strictly monotone and therefore invertible. We will call its inverse $b(F)$. We can now express f as a function of α. Since

$$f = \beta - \frac{\beta\bar{\mu} - s(\bar{\mu})}{\bar{\mu}} = \beta - \alpha(\beta\bar{\mu} - s(\bar{\mu}))$$

and

$$\frac{d\beta}{dF}\left\{\frac{d}{d\beta}[\beta - \alpha(\beta\mu - s(\mu))]\right\} = 0,$$

one finds that $-f(\alpha)$ is the Legendre transform of $b(F)$:

$$f(\alpha) = -\min_F \{\alpha F - b(F)\}. \qquad (6.10)$$

Some papers take $f(\alpha)$, a quantity we derive from $F(\beta)$, as their starting point. They give a direct definition of $f(\alpha)$ in terms of the intervals Δ. We reproduce that statement as a corollary of the above.

Consider the intervals of S_n. As explained in section 2, one can define a probability measure p such that each of these intervals has an equal probability (see (6.1)) $p_n = a^{-n}$. According to (6.5), the 'average' length l_n of the intervals that dominate the sum is $a^{-n\bar{\mu}}$. So

$$p_n = l_n^{1/\bar{\mu}} = l_n^{\alpha}. \tag{6.11}$$

The cardinality of the intervals that dominate the sum is (see (6.5)) $a^{ns(\bar{\mu})}$. One obtains

$$a^{ns(\bar{\mu})} = l_n^{-s(\bar{\mu})/\bar{\mu}} = l_n^{-f} \tag{6.12}$$

The convenient property of $f(\alpha)$ that is often used is the following: Equation (6.10) implies that $f(\alpha)$ has a unique maximum, and moreover at the maximum

$$f(\alpha_{max}) = F(\beta = HD) = 0 \tag{6.13}$$

In that case, as discussed in section 5, f is precisely the Hausdorff dimension of S.

We emphasize that if we had constructed the same set S by means of a different tree, then $F(\beta)$ would have been different and so would $f(\alpha)$. It is left to the reader to show that the tree of figure 2.5 has associated with it:

$$F(\beta) = -\ln\{\sigma(0)^{\beta} + \sigma(0)^{\beta}(\sigma(0)^{\beta} + \sigma(1)^{\beta})\}.$$

To exhibit more clearly the connection between classical statistical mechanics and successive approximation of Cantor sets, we will give the "Hamiltonian" h (see(6.2)) in a few simple cases. For instance, for the two scale Cantor set one concludes from (5.2) that

$$h(\varepsilon_n \dots \varepsilon_1) = \frac{1}{2}\ln\sigma(0)\,\sigma(1) + \frac{1}{2}\frac{i}{n}\ln\frac{\sigma(1)}{\sigma(0)}$$

$$= \ln\sigma(0) + \ln\frac{\sigma(1)}{\sigma(0)}\cdot\frac{1}{n}\sum\varepsilon_i\,. \tag{6.14}$$

This corresponds with a system of n non-interacting spins that assume the values 0 or 1 (Ising) in an external field of strength $\ln\dfrac{\sigma(1)}{\sigma(0)}$.

It is straightforward to derive the Hamiltonian for the binary Cantor set with 1 level of memory (figure 2.3). Note that

$$\sigma(\varepsilon, \varepsilon') = \sigma(00)\left[\frac{\sigma(10)}{\sigma(00)}\right]^{\varepsilon}\left[\frac{\sigma(01)}{\sigma(00)}\right]^{\varepsilon'}\left[\frac{\sigma(11)\sigma(00)}{\sigma(10)\sigma(01)}\right]^{\varepsilon\varepsilon'} .$$

We have

$$|\Delta(\varepsilon_n \ldots \varepsilon_1)| = \sigma(\varepsilon_n\,\varepsilon_{n-1})\,\sigma(\varepsilon_{n-1}\,\varepsilon_{n-2}) \ldots \sigma(\varepsilon_2\,\varepsilon_1).$$

Thus

$$h(\varepsilon_n \ldots \varepsilon_1) = \ln\sigma(00) + \ln\frac{\sigma(10)}{\sigma(00)}\sum_{2}^{n}\frac{\varepsilon_i}{n} + \ln\frac{\sigma(01)}{\sigma(00)}\sum_{1}^{n-1}\frac{\varepsilon_i}{n} + \ln\frac{\sigma(11)\,\sigma(00)}{\sigma(10)\sigma(01)}\sum_{1}^{n-1}\frac{\varepsilon_i\varepsilon_{i-1}}{n}$$

$$\cong \ln\sigma(00) + \ln\frac{\sigma(10)\,\sigma(01)}{\sigma(00)^2}\,\frac{1}{n}\sum_{1}^{n}\varepsilon_i + \ln\frac{\sigma(11)\,\sigma(00)}{\sigma(10)\sigma(01)}\,\frac{1}{n}\sum_{1}^{n-1}\varepsilon_i\varepsilon_{i-1} . \quad (6.15)$$

This now corresponds with an Ising model with nearest neighbor interaction. In general, if we allow k levels of memory, we will get a k-nearest neighbor Ising model. Notice, however, that under assumptions (3.3) and (3.4) these interactions will decay exponentially with distance (because the scalings do so, see (3.5)).

We finally remark that from (6.14) it is immediately apparent that in determining thermodynamic quantities we discard information (Feigenbaum, 1987). In (6.15) only $\{\sigma(00), \sigma(10)\,\sigma(01), \sigma(11)\sigma(00)\}$ is needed as input. The tree, however, is given by 4 scalings plus ordering properties.

VII MARKOV GRAPHS (GRAND CANONICAL ENSEMBLE)

In the previous section we have described the statistics associated with equation (6.6). This equation says that the β-sum over the lengths of the intervals behaves as a pure exponential if we go down deep enough in the tree. In order to investigate how accurate this statement is for finite, but large n, we have to take into account fluctuations in F. It is natural to follow tradition here and construct a grand canonical ensemble (Feigenbaum 1987).

Suppose

$$\sum_{\{\varepsilon\}_n}|\Delta(\varepsilon_n \ldots \varepsilon_1)|^{\beta} = a^{-nF_n(\beta)} , \quad \lim_{n\to\infty} F_n(\beta) = F(\beta)$$

and define the "Gibbs potential" G

$$e^{-G} = \sum_n z^n \sum_{\{\varepsilon\}_n}|\Delta(\varepsilon_n \ldots \varepsilon_1)|^{\beta} . \quad (7.1)$$

Since the scalings depend exponentially weakly on the low ε's (see (3.5)), we can use equation (2.3) to approximate higher level Δ's very accurately with the ones that are one level lower. Suppose we approximate with k levels of memory:

$$e^{-G} = \sum_n z^n \sum_{\{\varepsilon\}_n} [\sigma(\varepsilon_n \ldots \varepsilon_{n-k})]^\beta |\Delta(\varepsilon_{n-1} \ldots \varepsilon_1)|^\beta$$

$$= \sum_n z^n \sum_{\{\eta_n \cdots \eta_{n-k+1}\}} \sum_{\{\varepsilon\}_n} T_{(\eta_n \cdots \eta_{n-k+1}),(\varepsilon_{n-1} \cdots \varepsilon_{n-k})} |\Delta(\varepsilon_{n-1} \ldots \varepsilon_1)|^\beta , \qquad (7.2)$$

where T is a transition matrix of rank 2^k with

$$T_{(\eta_n \cdots \eta_{n-k+1}),(\varepsilon_{n-1} \cdots \varepsilon_{n-k})} = 0$$

<u>except</u>

$$T_{(\varepsilon_n \cdots \varepsilon_{n-k+1}),(\varepsilon_{n-1} \cdots \varepsilon_{n-k})} = [\sigma(\varepsilon_n \ldots \varepsilon_{n-k})]^\beta .$$

If we now think of

$$\sum_{\{\varepsilon_{n-k-1} \cdots \varepsilon_1\}} |\Delta(\varepsilon_{n-1} \ldots \varepsilon_1)|^\beta = \Psi^{(n-1)}_{(\varepsilon_{n-1} \cdots \varepsilon_{n-k})}$$

as a vector $\Psi^{(n-1)}_\varepsilon$, then (7.2) becomes

$$e^{-G} = \sum_{\{\eta\}} \sum_n z^n (T\Psi^{(n-1)})_\eta = \sum_{\{\eta\}} \left(\sum_n z^n T^n \Psi^{(0)} \right)_\eta \qquad (7.3)$$

for some appropriate $\Psi^{(0)}$. So

$$= \sum_{\{\eta\}} ([1 - zT]^{-1} \Psi^{(0)})_\eta = \sum_{\{\varepsilon\}\{\eta\}} \frac{\text{cof}(\{\eta\}, \{\varepsilon\}) \cdot \Psi^{(0)}_\varepsilon}{\det(1 - zT)} , \qquad (7.4)$$

where $\text{cof}(\{\eta\}, \{\varepsilon\})$ stands for the cofactor of the matrix $(1 - zT)$ centered on the element $(\{\eta\}, \{\varepsilon\})$.

The asymptotic behavior of T (which has positive entries and therefore its largest eigenvalue is real and positive) is dominated by its largest eigenvalue. It is clear from (7.4) that e^{-G} has a pole when z^{-1} equals an eigenvalue of T. We therefore look for

$$\lambda(\beta) = \text{largest eigenvalue of T} = \frac{1}{\text{smallest pole of } e^{-G}} . \qquad (7.5)$$

Then if λ_2 is the second largest eigenvalue

$$e^{-G} = \sum_{\{\varepsilon\}} z^n \lambda(\beta)^n \Psi^{(0)}\left(1 + O\left(\left[\frac{|\lambda_2|}{\lambda}\right]^n\right)\right) .$$

Thus for some α between 0 and 1 (7.1) implies

$$F_n(\beta) = -\frac{\ln [\lambda(\beta)(1 + O(\alpha^n))^{1/n}]}{\ln a} \Rightarrow \lim_{n\to\infty} F_n(\beta) = -\frac{\ln \lambda(\beta)}{\ln a} . \qquad (7.6)$$

The reason we went through this formalism is that T is a sparse matrix and hence one may calculate $[\det(1 - zT)]^{-1}$ by means of Markov graphs. We will now see how one achieves this.

One draws a graph on which the nodes are labelled with $(\varepsilon_n \dots \varepsilon_{n-k+1})$ for all allowed ε-words (in figure 2.2, for instance, "11" is not allowed). One then draws arrows from $(\varepsilon_{n-1} \dots \varepsilon_{n-k})$ to $(\varepsilon_n \dots \varepsilon_{n-k+1})$ respresenting $z[\sigma(\varepsilon_n \dots \varepsilon_{n-k})]^{\beta}$. According to (7.2), e^{-G} is obtained by summing over the contributions of all possible paths through the graph. We are now interested in the poles of the expression thus obtained (see (7.4)).

Before we continue with examples, let us write up the rules by which the summation can be done easily. The following rules must be applied in the order in which they are given here.

1) Links that are in series are multiplied

2) Links that are parallel (i.e., start at node a, end at node b, and $a \neq b$) are added up

3) Links that form loops (i.e., starting and ending on the same node) form a geometric sequence $s \rightarrow \Sigma s^n = \frac{1}{1 - s}$.

These rules can be inferred easily by comparing (7.1) and (7.2) with the definition of a tree. If $a = \frac{1}{2} \log N_n$ (see (6.1)) is unknown, one derives from (7.1):

4) $\Sigma\, |\Delta(\varepsilon_n \dots \varepsilon_1)|^0 = N_n = a^n$.

Thus for $\beta = 0$, the smallest zero of $\det(1 - zT)$ is equal to a^{-1}. Finally, from (6.12)

5) If $\beta = HD(S)$ then $z = 1$ is a zero.

Here are the examples. The Markov diagram drawn in figure 7.n corresponds to the tree drawn in figure 2.n.

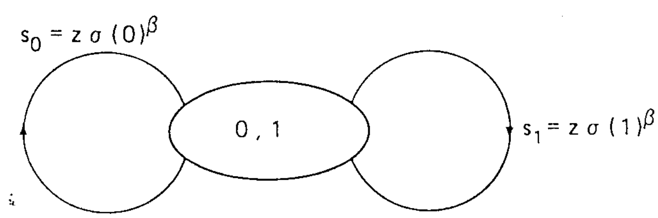


figure 7.1

To sum the contributions in figure 7.1, apply rule 2 first and then rule 3. One obtains

$$e^{-G} = \sum_n (s_0 + s_1)^n = \frac{1}{1 - z(\sigma(0)^\beta + \sigma(1)^\beta)} .$$

As before

$$\sigma(0)^{HD} + \sigma(1)^{HD} = 1$$

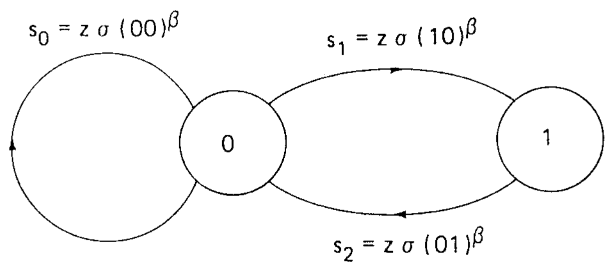


figure 7.2

To sum the contributions, add s_1 and s_2 first. Then figure 7.1 results with s_1 replaced by $s_1 s_2$.

$$e^{-G} = \frac{c}{1 - z\sigma(00)^\beta - z^2\sigma(10)^\beta\sigma(01)^\beta} .$$

The consistency requirements 4 and 5 give

$$1 - a - a^2 = 0 \Rightarrow a = \frac{1+\sqrt{5}}{2}$$

and

$$1 - \sigma(00)^{HD} - \sigma(10)^{HD}\,\sigma(01)^{HD} = 0.$$

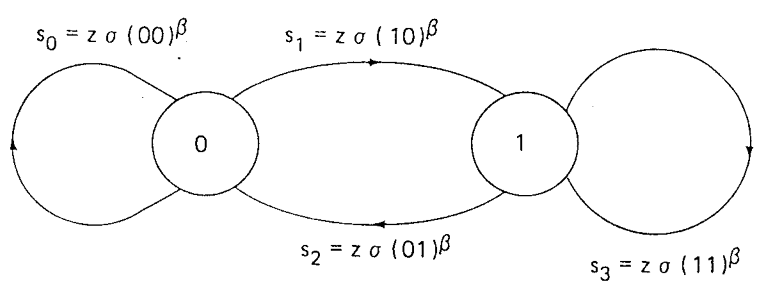


figure 7.3 figure 7.4

As stated before, thermodynamics "forgets" the ordering properties, therefore figure 7.3 and 7.4 are identical. To sum the graphs, first sum the part to the left of node 1. This gives

$$s' = \frac{z^2\, \sigma(01)^\beta\, \sigma(10)^\beta}{1 - z\, \sigma(11)^\beta}\,.$$

Now the figure is reduced to figure 7.1. Thus

$$e^{-G} = \frac{1 - z\, \sigma(11)^\beta}{1 - z(\sigma(00)^\beta + \sigma(11)^\beta) - z^2(\sigma(01)^\beta\, \sigma(10)^{\,\beta} - \sigma(00)^\beta\, \sigma(11)^\beta)}\,.$$

One can check that $a = 2$.

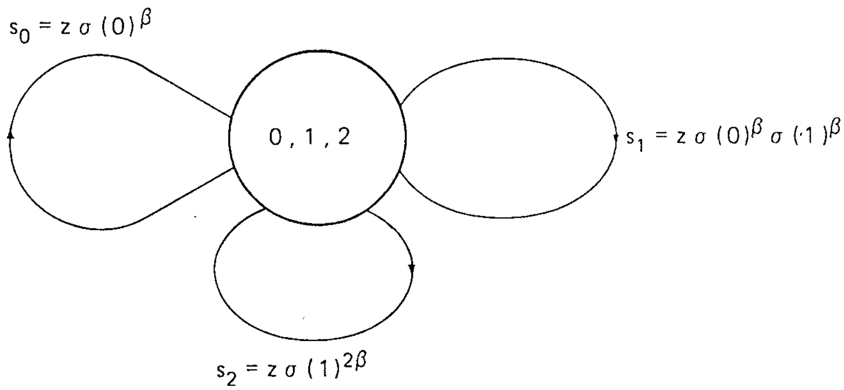


figure 7.5

As in figure 7.1

$$e^{-G} = \frac{1}{1 - z[\sigma(0)^\beta + \sigma(0)^\beta((\sigma(0)^\beta + \sigma(1)^\beta)])}\,.$$

VIII CIRCLE MAPS

In the remaining sections we will discuss an application of the foregoing ideas. This section will provide the reader with some background material needed to appreciate the example. No attempts to give demonstrations have been made in this section.

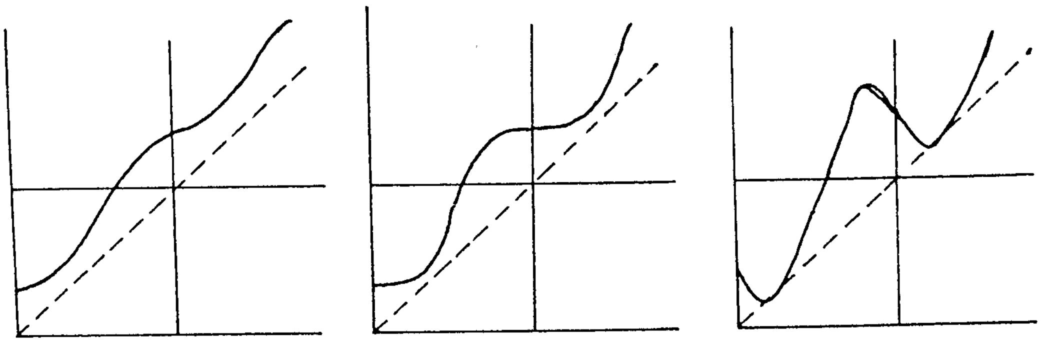
A circle map is a map that takes points of the circle to points of the circle. One can also describe this as a map f of the real line to itself with the additional property that

$$f(x + 1) = f(x) + 1\,.$$

By "forgetting" the integer of the map one obtains the map of the circle to itself. A commonly studied example is

$$x_{i+1} = (x_i + \omega - \frac{k}{2\pi}\sin 2\pi x_i) \bmod 1 \tag{8.1}$$

This is a degree one map, meaning that it wraps around the circle only once. Maps of degree one have the property that if they are monotone their dynamics is fairly simple. In figure 8.1 we have drawn the map (8.1) for $0 \le k < 1$ when it is strictly monotone, for $k = 1$ when it is critical (i.e., it is invertible but there is a point x_0 such that $f'(x_0) = 0$), and for $k > 1$ when it is non-invertible.



$0 \le k < 1$ $k = 1$ $k > 1$.

figure 8.1

The rotation number is defined as the average (angular) velocity along an orbit. More precisely, at a point x we define

$$\rho_f(x) = \lim_{n\to\infty} \frac{f^n(x)}{n} \quad \text{(if the limit exists)} \tag{8.2}$$

(Here f is regarded as a map on the real line.) The rotation number is one of the most important tools in studying the dynamics of circle maps. It describes the average behavior of an orbit.

In the case that $k > 1$ the map may or may not exhibit sensitive dependence on initial conditions ("chaotic behavior"). For $0 \le k \le 1$ there is no sensitive dependence on initial conditions. The case $k = 1$ is sometimes called "critical."

For the purposes of this paper it is sufficient to consider monotone $(0 \le k \le 1)$ maps. These maps have the property that for each point x the limit in (8.2) exists and moreover that its value does not depend on x. However, the rotation number is not constant as we let f vary.

Consider a one-parameter family of circle maps, for constant k, let f_ω be defined as in (8.1) where now $\omega \in [0, 1]$. The question we will address is: how does the rotation number vary as a function of ω. In other words: where do qualitative differences in behavior (namely change of rotation number) occur as we let ω vary through the parameter space? The function that gives, for constant k, the rotation number as a function of ω will be denoted by $\rho(\omega)$. In particular, let f be twice differentiable and monotone of degree one. Note that $f_\omega(x)$ is also a continuous monotone function of ω. The function $\rho(\omega)$ has the following properties (see for example Devaney 1985):

i) $\rho(\omega)$ is monotone increasing and continuous

ii) The inverse image of a rational number $\frac{p}{q}$ is an interval, denoted by $I_{p/q}$.

iii) The inverse image of an irrational number α is a point.

Such a graph (see figure 8.2) is sometimes called the "devil's staircase."

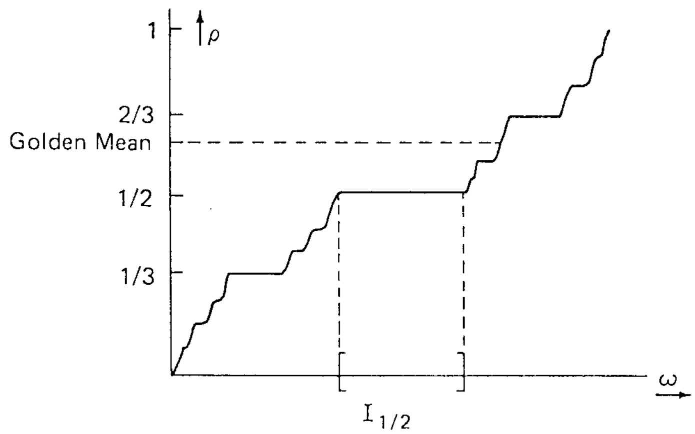


figure 8.2

The object is to study the properties of the Cantor set A defined as follows

$$A = [0,1] \setminus \{ \bigcup_{p/q}^{\infty} \mathring{I}_{p/q} \} , \tag{8.2}$$

where $\mathring{I}$ means the interior of I. All one has to do to apply the theory in the previous sections is to find a natural way to construct that Cantor set by means of a tree.

The simplest example of the set A is obtained when we let k go to zero. The intervals $I_{p/q}$ now collapse to zero length, i.e., to a point. In this case,

$$\rho(\omega) = \omega.$$

To apply the theory we must define a "formal Cantor set" as follows. Replace in the unit interval each rational point $\frac{p}{q}$ by two rational points $\frac{p}{q}+$ and $\frac{p}{q}-$ (the boundary of $I_{p/q}$ if $k > 0$), without changing the distances between any two points. The interval thus modified will henceforth be denoted by A.

To generate the Cantor set, we first generate the rational numbers by means of a tree. The natural way to do that is to use the so-called Farey tree (which also plays an important role in the theory of renormalization of circle maps). The first few levels of the tree are drawn in figure 8.3.

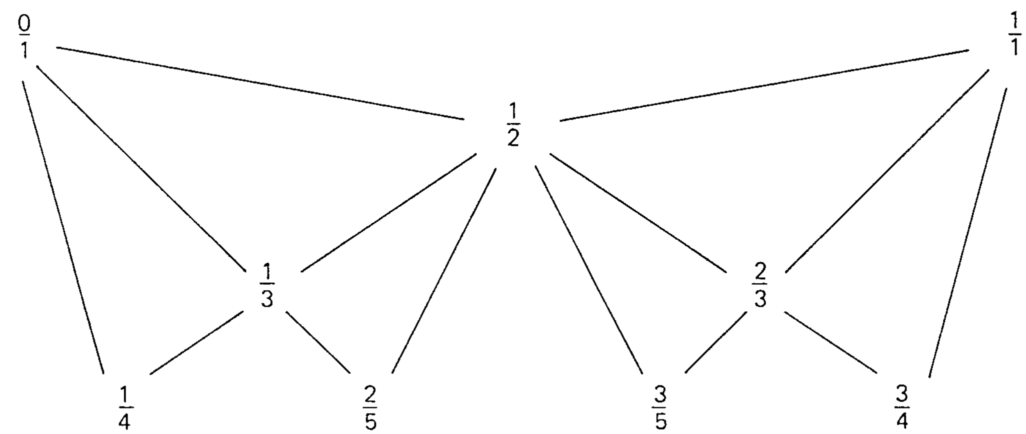


figure 8.3

Without further explanation, we give the rules here (see Hardy and Wright, 1924). The number $\frac{1}{2}$ is the rational between 0 and 1 with the smallest denominator. It is a neighbor of 0 and of 1 in the tree. Its level in the tree is 1. The numbers on level 2 are obtained similarly: between every two neighbors we find the number with the smallest possible denominator. The new pairs of numbers are 0 and $\frac{1}{3}$, $\frac{1}{3}$ and $\frac{1}{2}$, $\frac{1}{2}$ and $\frac{2}{3}$, $\frac{2}{3}$ and 1. One constructs the tree level by level defining the number with the smallest denominator between each pair of neighbors. Notice that each number in the tree now has a "direct parent" and an "indirect parent." In the sequel we will draw only the direct parents, so that a binary tree results (see figure 10.2).

More formally, one can show that $\frac{p}{q}$ on level n and $\frac{r}{s}$ on level $n - \ell$ are Farey neighbors if and only if

$$p \cdot s - q \cdot r = \pm 1. \tag{8.2}$$

Their "child" at level $n + 1$ is given by

$$\text{child} = \frac{p + r}{q + s} . \tag{8.3}$$

The tree enumerates all rational numbers.

The tree that generates the Cantor set can now be defined as follows. Let level 1 consist of

$$\Delta(0) = [0, \tfrac{1}{2}]$$

$$\Delta(1) = [\tfrac{1}{2}, 1] .$$

Level 2 consists of the lengths of

$$\Delta(00) = [\tfrac{0}{1}, \tfrac{1}{3}]$$
$$\Delta(10) = [\tfrac{1}{3}, \tfrac{1}{2}]$$
$$\Delta(11) = [\tfrac{1}{2}, \tfrac{2}{3}]$$
$$\Delta(01) = [\tfrac{2}{3}, \tfrac{1}{1}]$$

We have chosen the labelling so that the topology of the tree is the same as in figure 2.4. The presentation function will be similar to the one drawn in figure 3.1. The reason for this choice is that it is the most natural in terms of the continued fraction expansions, as will be discussed in the next section. We will denote this tree as the "Farey tree of the rational intervals."

IX THERMODYNAMICS OF THE FAREY TREE

In this section the thermodynamics of the Farey tree of rational intervals is discussed. As we have seen in the previous section, this corresponds to the thermodynamics of the set A for the family f_ω given in (8.1) with $k = 0$ (rotations).

There are various reasons to treat this simple case. First of all, the problem can to a large extent be treated analytically and therefore yields insight into the more general case. Further, numerical calculations indicate that the scaling function for $k = 1$ resembles the scaling function for $k = 0$ very closely. Therefore one would expect the thermodynamics to be very similar. The third reason lies a little beyond the scope of this article. Summarized, it is the following. If k is smaller than 1 and if $\rho(\omega)$ assumes an irrational value (a Diophantine number), then (renormalized) high iterates of f_ω will more and more resemble (renormalized) high iterates of rotation by ρ. This suggests that scalings should be the same locally around those irrational values of $\rho(\omega)$. The case $k = 0$ has therefore more general information than one would suppose at first sight.

We first need to discuss some elementary but crucial number theory. Consider a real $\rho \in (0, 1)$ and define the continued fraction coefficients c_i of ρ as follows (see Hardy and Wright, 1924):

$$\rho = \cfrac{1}{c_1 + \cfrac{1}{c_2 + \cfrac{1}{\ddots}}} \stackrel{D}{=} [c_1, c_2, \ldots].. \tag{9.1}$$

The sequence $\{c_i\}$ is finite if and only if ρ is rational. The continued fraction approximate of ρ are given by

$$\frac{p_k}{q_k} \stackrel{D}{=} \cfrac{1}{c_1 + \cfrac{1}{c_2 + \ddots + \cfrac{1}{c_k}}} = [c_1, c_2, \ldots, c_k]. \tag{9.2}$$

Note that without loss of generality we take $c_k \geq 2$. In matrix form this becomes

$$M \stackrel{D}{=} \begin{pmatrix} p_{k-1} & p_k \\ q_{k-1} & q_k \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & c_1 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & c_k \end{pmatrix}, \tag{9.3}$$

as one can check by multiplying out the c_k's in equation (9.2).

The continued fractions are very "natural" in the following sense. If we consider

$$R_\rho(x) = (x + \rho) \bmod 1$$

then the sequence

$$\{R_\rho^{q_k}(x)\}_{k=1}^{\infty}$$

forms a sequence of <u>closest returns</u> to x. (That is, they approximate x better than $R_\rho^q(x)$ with $q < q_k$.)

We can now define a generating function for the Farey tree (the inverses will turn out to be the presentation function for the Farey tree of the rational intervals)

$$\begin{aligned} F_0 &: [c_1\, c_2 \ldots] \rightarrow [1 + c_1, c_2, \ldots] \\ F_1 &: [c_1\, c_2 \ldots] \rightarrow [1, c_1, c_2 \ldots] \end{aligned} \tag{9.4a}$$

By using the definition of c_k, one finds

$$F_0(x) = \frac{x}{1 + x}$$

$$F_1(x) = \frac{1}{1 + x}. \tag{9.4b}$$

As in section 8, we say that the number $\frac{1}{2}$ is in level 1 of the Farey tree; the numbers $\frac{1}{3}$ and $\frac{2}{3}$ form level 2 and so on. We will show that the n-th level of the Farey tree consists of the numbers

$$\bigcup_{\{\varepsilon_1 \cdots \varepsilon_n\}} F_{\varepsilon_1} \dots F_{\varepsilon_n}(1) = \text{level } n . \tag{9.5}$$

To start the reasoning one shows that p_{k-1}/q_{k-1} and p_k/q_k are Farey neighbors. This comes about because the determinant of the matrix M defined in (9.3) has determinant $(-1)^k$, which is equivalent to criterion (8.2) for being neighbors. To find out who the child of these two numbers is, use (8.3):

$$\begin{aligned}\text{child} &= \frac{p_{k-1}+p_k}{q_{k-1}+q_k} = M\begin{pmatrix}1\\1\end{pmatrix} = M\begin{pmatrix}1&1\\0&1\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}\\ &= \begin{pmatrix}0&1\\1&c_1\end{pmatrix}\cdots\begin{pmatrix}0&1\\1&c_k\end{pmatrix}\begin{pmatrix}1&1\\0&1\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}\\ &= \begin{pmatrix}0&1\\1&c_1\end{pmatrix}\cdots\begin{pmatrix}0&1\\1&c_k+1\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} = [\, c_1, \dots , c_{k-1}, c_k+1] .\end{aligned} \tag{9.6}$$

It follows from this that the "direct" parent (see section 8) of $[c_1 \dots c_k]$ is $r/s = [c_1 \dots, c_k\text{-}1]$. If $c_k > 2$, we have that

$$r/s \neq p_{k-1}/q_{k-1}.$$

Therefore, the second child of p_k/q_k is:

$$\begin{aligned}\text{child} &= \frac{r+p_k}{s+q_k} = \begin{pmatrix}0&1\\1&c_1\end{pmatrix}\cdots\begin{pmatrix}0&1\\1&c_{k-1}\end{pmatrix}\left\{\begin{pmatrix}0&1\\1&c_k\text{-}1\end{pmatrix}+\begin{pmatrix}0&1\\1&c_k\end{pmatrix}\right\}\begin{pmatrix}0\\1\end{pmatrix}\\ &= \begin{pmatrix}0&1\\1&c_1\end{pmatrix}\cdots\begin{pmatrix}0&1\\1&c_{k-1}\end{pmatrix}\begin{pmatrix}0&1\\1&c_k\text{-}1\end{pmatrix}\begin{pmatrix}0&1\\1&2\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} = [c_1 \dots c_k\text{-}1, 2].\end{aligned} \tag{9.7}$$

In the case that $c_k = 2$ (it is not smaller than 2 according to (9.2)), one finds the two parents of $[c_1, \dots, c_{k-1}, 2]$ and then calculates their children. Thus the above implies that the sum of the continued fraction coefficients increases by one as one goes from level n to level n + 1. Since by (9.4a) we have

$$[c_1 \dots c_k] = F_0^{c_1\text{-}1} F_1\ F_0^{c_2\text{-}1}F_1 \dots\ F_0^{c_k\text{-}1}F_1(1), \tag{9.8}$$

equation (9.5) follows.

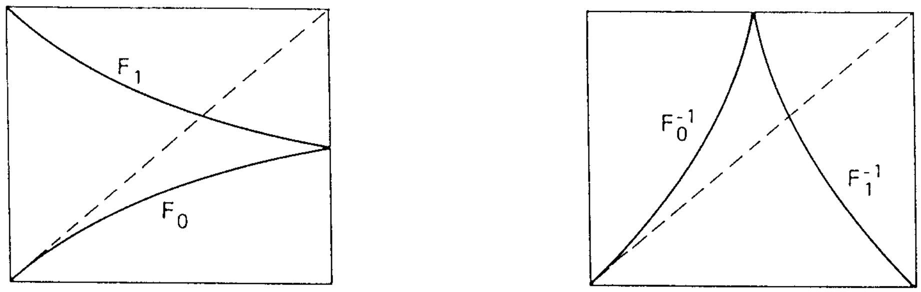


figure 9.1

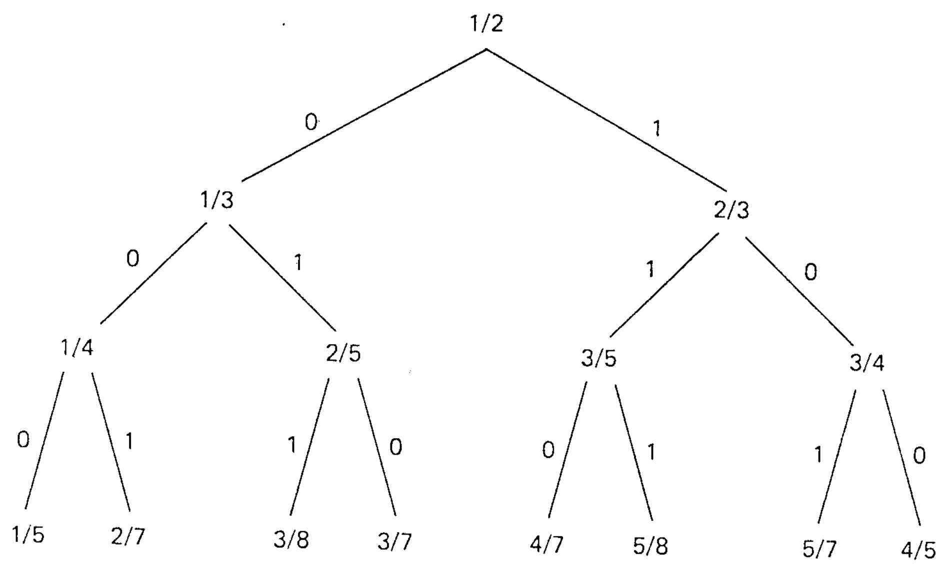


figure 9.2

In figures 9.1 and 9.2 we have drawn the presentation function and the Farey tree. Note that it is of the type of figure 2.4. In figure 9.3 the tree is drawn in its continued fraction representation.

Now all we have to do is to determine the Farey tree T of the rational intervals as defined in (8.4). However, it is not hard to see that the inverses of F_0 and F_1 are precisely the presentation function for T in the way this was described in section 3. (The endpoints of the $\Delta(\varepsilon_n \dots \varepsilon_1)$ are the rationals which are generated by the F_{ε}.)

The intervals $\Delta(0\varepsilon_{n-1} \dots \varepsilon_1)$ and $\Delta(1\varepsilon_{n-1} \dots \varepsilon_1)$ on level n of T are bounded by the Farey neighbors and thus they must have lengths

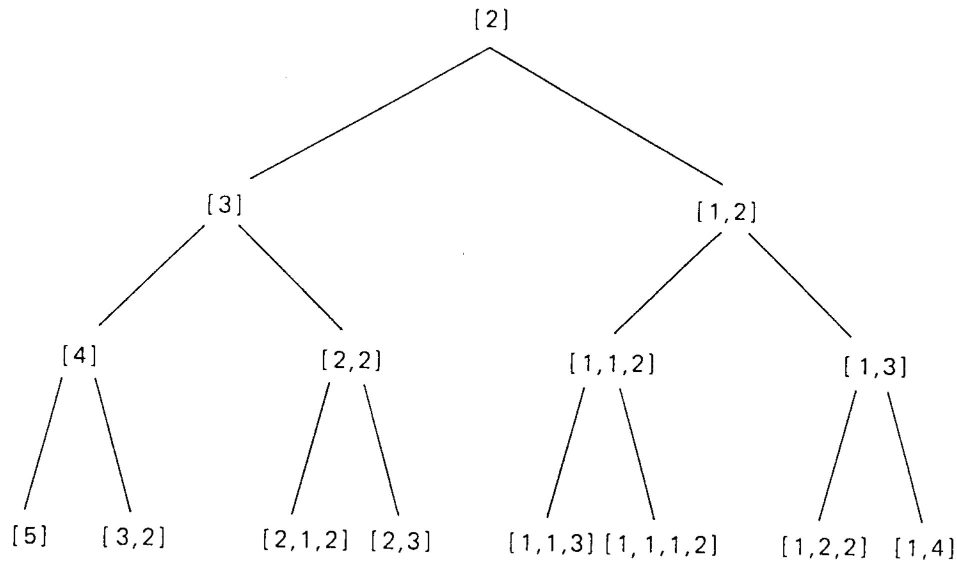


figure 9.3

$$\begin{aligned} A &= \mid [c_1 \dots c_k - 1] - [c_1 \dots c_k] \mid \\ B &= \mid [c_1 \dots c_{k-1}] - [c_1 \dots c_k] \mid \end{aligned} \tag{9.9}$$

where $[c_1 \dots c_k] = F_{\varepsilon_1} \dots F_{\varepsilon_{n-1}}(1)$. (The expressions in (9.9) are valid even if $c_k = 2$ or $c_{k-1} = 1$.)

To calculate the length A and B, use (8.2) to see that the distance between two Farey neighbors equals the reciprocal of their two denominators. It follows from (9.3) that if q_k is the denominator of $[c_1 \dots c_k]$ then

$$\prod_{i=1}^{k} c_i \le q_k \le \prod_{i=1}^{k} (c_i + 1) \; . \tag{9.10}$$

Using (9.9) one obtains:

$$\{c_k(c_k+1) \prod_{i=1}^{k-1} (c_i+1)^2\}^{-1} \le A \le \{(c_k - 1)\, c_k \prod_{i=1}^{k-1} c_i^2\}^{-1}$$

$$\{(c_k+1) \prod_{i=1}^{k-1} (c_i+1)\}^{-1} \le B \le \{c_k \prod_{i=1}^{k-1} c_i^2\}^{-1}$$

$$\Rightarrow \{ \prod_{i=1}^{k} (c_i+1)^{2\beta} \} \le A^\beta + B^\beta \le \{ \prod_{i=1}^{k} c_i^{2\beta} \}^{-1} \cdot c_k^\beta \, ,$$

Forming the β-sum as in equation (6.6) one obtains

$$\sum_{\Sigma c=n} \prod_{i=1}^{k} (c_i + 1)^{-2\beta} \leq 2^{-nF(\beta)} \leq \sum_{\Sigma c=n} c_k \prod_{i=1}^{k} c_i^{-2\beta} \tag{9.11}$$

In the next section we will investigate the approximate thermodynamics.

We close this section with the important observation that our F_ε's do not satisfy (3.4), i.e., they are smooth, but they are not contractions. Our F_ε's have slopes smaller or equal to 1 (in absolute value) and equality is attained for both branches. It is not known how to generalize the proofs of the results of sections 4 to 7 for this case. Numerical results, however, consistently indicate that the thermodynamic formalism still holds. We will therefore proceed with the calculations.

X APPROXIMATE THERMODYNAMICS

In order to solve for the thermodynamics approximately we will define an approximate thermodynamics. The new model then is (compare with (9.11)):

$$2^{-nF(\beta)} = \sum_{\substack{\{c\} \\ \Sigma c=n}}^{n} \prod_{i=1}^{k} \mu^{-\beta} c_i^{-2\beta} \tag{10.1}$$

for some fixed $\mu > 1$. Note that equation (6.2) now implies

$$h(\varepsilon_n \dots \varepsilon_1) = \sum_{i=1}^{k} (\ln \mu + \ln c_i)$$

where $\Sigma c = n$. If k is of order n, then $h(\varepsilon_n \dots \varepsilon_1)$ ("the energy") of the corresponding interval is big and therefore its contribution to the sum in (10.1) is small. One can compare this model with a model in which k particles have a distance c_i/n to each other. Each particle carries an energy $\ln \mu$ with it and furthermore the particles repel each other with a potential energy $\ln c_i$. (Repelling means that the "state" in which they are equidistant with small distance in between contributes very little to the sum for $\beta > 0$.)

Following equation (7.1), we define the grand canonical ensemble

$$e^{-G} = \sum_{\Sigma c=2}^{\infty} z^{(\Sigma c)} \sum_{\{c\}} \prod_{i=1}^{k} \mu^{-\beta} c_i^{-2\beta} . \tag{10.2}$$

Therefore

$$e^{-G} = -\frac{z}{\mu^\beta} + \sum_{\Sigma c=1}^{\infty} \sum_{\{c\}} \prod_{i=1}^{k} z^{c_i} \mu^{-\beta} c_i^{-2\beta} . \tag{10.3}$$

Because of (10.1) for

$$| z \cdot 2^{-F(\beta)} | < 1 ,$$

the above series converges absolutely and we may therefore change the order of summation in (10.3). Summing first the terms of order $\mu^{-\beta}$ then $\mu^{-2\beta}$ and so on, one obtains

$$e^{-G} = - z\mu^{-\beta} + \sum_{k=1}^{\infty} \left\{ \sum_{n=1}^{\infty} z^n \mu^{-\beta} n^{-2\beta} \right\}$$

$$= \sum_{k=1}^{\infty} \left\{ \Lambda_\mu(z, \beta) \right\}^k - z \mu^{-\beta} = \frac{\Lambda}{1 - \Lambda} - z \mu^{-\beta} \tag{10.4}$$

where

$$\Lambda_\mu = \sum_{n=1}^{\infty} \frac{z^n}{\mu^\beta n^{2\beta}} .$$

According to (7.4) and (7.5), the free energy is determined by the pole $[\lambda(\beta)]^{-1}$ of e^{-G}:

$$F(\beta) = - \ln \lambda(\beta) / \ln 2 , \quad \Lambda(\lambda(\beta), \beta) = 1 \tag{10.5}$$

Since we know that the Hausdorff dimension of the set whose thermodynamics we are trying to calculate equals one (see section 9), we have the constraint:

$$\lambda(1) = 1 \Rightarrow \Lambda(1, 1) = \mu^{-1} \sum_{i=1}^{\infty} \frac{1}{n^2} = 1 .$$

Denote the Riemann zeta function $\sum_{i=1}^{\infty} n^{-s}$ by $\zeta(s)$, then this implies that

$$\mu = \zeta(2) = \pi^2/6.$$

Thus from (10.5) one concludes that the (approximate) thermodynamics is determined by

$$\sum_{n=1}^{\infty} [\lambda(\beta)]^{-n} n^{-2\beta} [\zeta(2)]^{-\beta} = 1. \tag{10.6}$$

To investigate this thermodynamics, let us first consider the behavior of $\lambda(\beta)$ as β goes to $-\infty$. We claim that the asymptotic behavior of $\lambda(\beta)$ is given by:

$$\lambda(\beta) = \left[2\sqrt{\zeta(2)}\right]^{-\beta} = [2.565...]^{-\beta}. \tag{10.7}$$

To verify this, one first observes that (10.7) implies that the n-th term C_n of the series is given by

$$C_n = \left[2\sqrt{\zeta(2)}\right]^{n\beta} n^{-2\beta}[\zeta(2)]^{-\beta} = [2^{-n}\, n^2\, \zeta(2)^{(1-n/2)}]^{|\beta|}.$$

One now checks that asymptotically as $\beta \to -\infty$

$$C_n \le \delta^{n|\beta|}, \quad \text{for some } \delta < 1, \ \underline{\text{except}}\ C_2 = 1.$$

Using (10.5), one deduces that asymptotically

$$\sum_{\{\varepsilon\}_n} |\Delta(\varepsilon_n \dots \varepsilon_1)|^\beta = 2^{-nF(\beta)} = [2.565...]^{-n\beta}.$$

If one reasons along the line of the remark below (10.1), it is clear that for the original model described in section 9, we expect the golden mean to dominate. That is, asymptotically as $\beta \to \infty$

$$\sum |\Delta(\varepsilon_n \dots \varepsilon_1)|^\beta \sim [\sigma(11\dots)]^{n\beta}.$$

Here $\sigma(11\dots) = \dfrac{2}{3+\sqrt{5}} \approx \dfrac{1}{2.618}$ is the derivative of F_1 at its fixed point. This differs less than 2% from the prediction of the approximate model.

We will now determine the nature of the free energy when β is very close to one. This requires some algebra.

$$\int_0^\infty \frac{x^{2\beta-1}}{e^{x-F}-1}\,dx = \sum_1^\infty e^{nF}\int_0^\infty x^{2\beta-1}e^{-nx}\,dx. \tag{10.8}$$

By substituting $u = nx$

$$= \sum_1^\infty \frac{e^{nF}}{n^{2\beta}}\Gamma(2\beta),$$

which according to equation (10.6) $= \zeta(2)^{\beta}\,\Gamma(2\beta)\,.$ (10.9)

Now set

$$2\beta = 2 - \varepsilon.$$

Note that the left hand side of (10.8) and (10.9) are infinitely differentiable with respect to β for $\beta > 0$. One thus (asymptotically as ε goes to zero) solves F from:

$$\int_0^{\infty} \frac{x}{e^{x-F}-1}\,dx + \varepsilon C_1 = \zeta(2)\,\Gamma(2) + \varepsilon C_2\,.$$

From (10.8) and (10.9) one then finds that

$$\int_0^{\infty}\left[\frac{x}{e^{x-F}-1} - \frac{x}{e^{x}-1}\right]dx = \varepsilon C.$$

Writing $e^x = 1 + u$, rationalizing the integrand and splitting the integral in two pieces:

$$\int_0^{b} \frac{\ln(1+u)}{u}\,\frac{(1-e^{+F})(1+u)}{(1+u)(e^{-F}-1)}\,du + \int_b^{\infty} \ldots\, du = \varepsilon C.$$

Taking $-F \ll b \ll 1$, we see that the second integral does not depend on F, and its outcome is a constant K. In the remaining integral the terms $\frac{\ln(1+u)}{u}\,(1+u)$ cancel approximately. The result is an elementary integral which yields

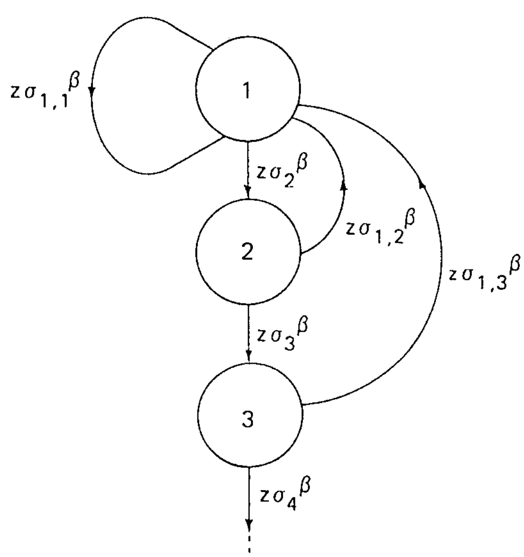


figure 10.1

$$(1 - e^F)\ln[(1+u)e^{-F} - 1]\Big|_0^b + K = \varepsilon C.$$

It is straightforward (neglecting F^2 terms) to show that this implies that asymptotically

$$F \ln(-F) = (\beta - 1)C - K. \tag{10.10}$$

Therefore $F(\beta)$ has a branch point at $\beta = 1$.

It is also possible to draw the Markov diagram associated with (10.3). While we have not formally written down a transition matrix, it is not hard to see how the Δ's change from one level to the next. From (10.1) we have:

$$\sigma_n = \frac{|\Delta(c_k = n+1)|}{|\Delta(c_k = n)|} = \left(\frac{n}{n+1}\right)^2, \quad n \geq 2. \tag{10.11a}$$

When $c_{k+1} = 2$, one has

$$\sigma_{1,n} = \frac{|\Delta(c_k = n-1,\, c_{k+1} = 2)|}{|\Delta(c_k = n)|} = \frac{n^2}{(n-1)\, 2^2 \mu}. \tag{10.11b}$$

These describe all the possible transitions in this approximate model. The Markov diagram is given in figure 10.1. Note that the diagram is infinite. As remarked before, at $\beta = 1$ the term that contributes most is $c_k \to \infty$. This is reminiscent of Bose condensation in statistical physics.

To end this section, a remark about numerical work with circle maps. From the theoretical point of view taken in this work, it is appropriate for us to think about the δ's as iterates under the

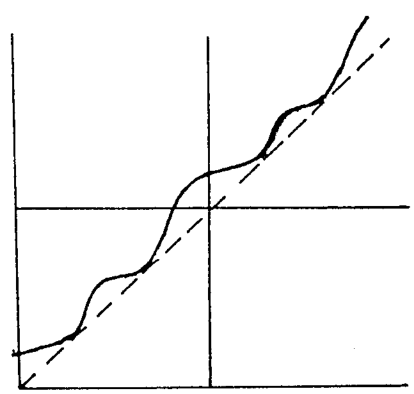


figure 10.2

presentation function as is done throughout this text. Since this presentation function in general is not given this would mean that one has to find the boundaries of the rotation intervals.
Numerically this is not a good scheme. The reason is that the boundaries are very hard to locate with good accuracy due to the marginally unstable character of the dynamics at such parameter values (see figure 10.2). To illustrate this point we have drawn a circle map at the upper boundary of the rotation interval $I_{1/2}$.

There is a much stabler method at hand. This is to find the midpoints $\omega(p/q)$ of the rotation interval $I_{p/q}$. This is dynamically very stable (derivative zero for critical maps). One can now set up the thermodynamics using only intervals $\delta(\varepsilon_n \ldots \varepsilon_1)$ whose endpoints are the midpoints $\omega(p/q)$ (see figure 10.3).

In this scheme, somewhat different from the foregoing, one takes out only one level of δ's (that is: the previous level δ's are first restored). The thermodynamics obtained from such a scheme is identical to the one discussed above. We leave as an ambitious exercise for the readers to convince themselves of this.

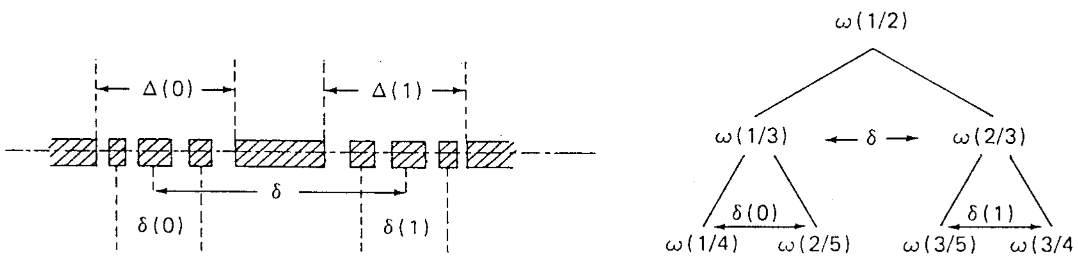


figure 10.3

REFERENCES

T. Bohr, D. Rand, *Physica* 25D, 387-398, 1987.

R.L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Benjamin/Cummings, 1986.

K.M. Falconer, *The Geometry of Fractal Sets*, Cambridge Univ. Press, 1985.

M.J. Feigenbaum, *J. Stat. Phys.*, 46:5/6, 919-924, 1987.

M.J. Feigenbaum, *J. Stat. Phys.*, 46:5/6, 925-932, 1987.

M.J. Feigenbaum, *J. Stat. Phys.*, **52**:3/4, 527-569, 1988.

G.M. Hardy, E.M. Wright, *An Introduction to the Theory of Numbers*, Oxford Univ. Press, 1960.

B.B. Mandelbrot, *The Fractal Geometry of Nature*, W.H. Freeman, 1982.

D. Ruelle, Thermodynamic Formalism, *Encyclopedia of Mathematics and its Applications*, **5**, Addison-Wesley, 1978.

D. Sullivan, *Differentiable Structures on Fractal Like Sets*, preprint CUNY, 1988.

ACKNOWLEDGMENTS

Both authors express their gratitude to the organizers of the Summer school for their hospitality. One of the authors (JJPV) is especially indebted to Eddie Cohen for his encouragement to participate in the Summer school and to write this article. He also acknowledges useful discussions with Donald Spear. While writing this work he was partially supported by DOE grant DE-AC-02-83-ER-13044. Finally, JJPV also expresses his admiration for the excellent research atmosphere in the Physics Department of Rockefeller University, of which he was a part for two years.