Ergod. Th. & Dynam. Sys. (1997), **17**, 531–564
Printed in the United Kingdom © *1997 Cambridge University Press*

Ratio geometry, rigidity and the scenery process for hyperbolic Cantor sets

TIM BEDFORD† and ALBERT M. FISHER‡§

† *Faculty of Mathematics, Delft University of Technology, PO Box 5031, 2600 GA Delft, The Netherlands*
‡ *SUNY at Stony Brook, USA*

(*Received* 29 *July* 1994 *and revised* 8 *June* 1996)

Abstract. Given a $\mathcal{C}^{1+\gamma}$ hyperbolic Cantor set C, we study the sequence $C_{n,x}$ of Cantor subsets which nest down toward a point x in C. We show that $C_{n,x}$ is asymptotically equal to an ergodic Cantor set valued process. The values of this process, called *limit sets*, are indexed by a Hölder continuous set-valued function defined on Sullivan's dual Cantor set. We show the limit sets are themselves $\mathcal{C}^{k+\gamma}, \mathcal{C}^{\infty}$ or $\mathcal{C}^{\omega}$ hyperbolic Cantor sets, with the highest degree of smoothness which occurs in the $\mathcal{C}^{1+\gamma}$ conjugacy class of C. The proof of this leads to the following rigidity theorem: if two $\mathcal{C}^{k+\gamma}, \mathcal{C}^{\infty}$ or $\mathcal{C}^{\omega}$ hyperbolic Cantor sets are $\mathcal{C}^1$ conjugate, then the conjugacy (with a different extension) is in fact already $\mathcal{C}^{k+\gamma}, \mathcal{C}^{\infty}$ or $\mathcal{C}^{\omega}$. Within one $\mathcal{C}^{1+\gamma}$ conjugacy class, each smoothness class is a Banach manifold, which is acted on by the semigroup given by rescaling subintervals. Smoothness classes nest down, and contained in the intersection of them all is a compact set which is the attractor for the semigroup: the collection of limit sets. Convergence is exponentially fast, in the $\mathcal{C}^1$ norm.

0. *Introduction*

Consider the sequence $C_{n,x}$ of Cantor subsets which nest down toward a point x in a hyperbolic Cantor set $C \subseteq [0, 1]$, and which have been affinely rescaled to have left and right endpoints at 0 and 1. We wish to describe how the geometry of these sets changes as n increases. If C is a linear set like the middle-third set this is not so difficult to do (we always get just another copy of C!) but, as we shall see, with nonlinearity the behavior of this 'scenery process' gets much more interesting.

A different way to describe the small-scale structure of C is by the *scaling function*, introduced by Feigenbaum for a specific class of examples, and studied by Sullivan in the present setting of $\mathcal{C}^{1+\gamma}$ hyperbolic Cantor sets.

A third approach is to define a flow, the continuous dynamics of which reflect the geometrical notion of zooming continuously down toward a point. Ergodicity of the flow

§ Current adress: Institute of Mathematics, UFRGS, Av. Bento Goncalves 9500, 91500 Porto Alegre, RS Brazil.

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

implies analogies of the Lebesgue density theorem, proved in [**BeFi1**] for Brownian zero sets and hyperbolic $\mathcal{C}^{1+\gamma}$ Cantor sets; the number that one obtains (order-two density) is a conformal invariant and provides a measure of the *lacunarity* of the fractal; compare [**Mand**]. This *scenery flow* is constructed for hyperbolic Cantor sets in [**BeFi2**], [**BeFi3**], for hyperbolic Julia sets in [**BeFiU**], and for limit sets of geometrically finite Fuchsian and Kleinian groups in [**Fi4**]. The associated translation (rather than dilation) scenery flows are studied for hyperbolic $\mathcal{C}^{1+\gamma}$ Cantor sets and the Fuchsian limit sets in [**Fi4**] and [**BuF**], where theorems like those of [**Fi3**] are proved (an order-two ergodic theorem, and infinite-measure unique ergodicity).

In this paper we will take a viewpoint close to that of Sullivan in [**Su1**], constructing the scenery process in a similar way to the scaling function. It is also possible to work in the other direction; in the later papers [**BeFi2**], [**BeFi3**] we show how to derive the scaling function from the scenery flow, and conversely how to construct the scenery flow from the scaling function. From this point of view the scenery process will be seen as an intermediate object, serving to connect the scenery flow with the scaling function.

Sullivan's main motivation in [**Su1**] was to develop an understanding of the geometry of hyperbolic Cantor sets for use in a new, more 'conceptual' approach to the Feigenbaum–Coullet–Tresser conjectures (see [**Su2**] and [**deMvS**]). (Lanford's proof [**Lan**] uses (rigorous) computer-assisted estimates.)

The first of these conjectures is the existence and uniqueness of a fixed point for the action of the renormalization operator on a certain space of analytic folding maps (see the discussion under 'Related work' below). A further conjecture states that for any folding map which converges to this renormalization fixed point, the small-scale structure of its attracting Cantor set will be the same.

To prove this last statement, Sullivan uses the following observation (which is attributed by Rand [**Ra**] to Misiurewicz): for the particular case of the folding map which is the renormalization fixed point, the Cantor set not only has the folding dynamics, which is a bijection when restricted there, but also a two-to-one $\mathcal{C}^{1+\alpha}$ dynamics, for which it is a *repeller*. This second, hyperbolic, dynamics is much easier to work with in certain respects. Indeed, the existence of this map led Sullivan to a separate question, which is the subject of §§1–3 of [**Su1**]: for general hyperbolic $\mathcal{C}^{1+\alpha}$ Cantor sets, can one classify differentiable structures? He shows that this can be done with a bounded Hölder scaling function providing a complete smooth conjugacy invariant. This, in turn, as will be explained below, is equivalent to knowing the small-scale structure of the Cantor set. In this way the hyperbolic dynamics helps answer a question about the original folding map.

Our own main focus is somewhat different to that of Sullivan in [**Su1**]. We want to describe the *exact* geometry, at small scales, of the Cantor sets, whereas with respect to the differentiable structure, all smoothly equivalent Cantor sets will look the same. However, for this purpose the scaling function also contains precisely the information one needs.

Our main theorems (Theorems 5.4, 7.4 and 7.5) concern the scenery process, the smoothness of limit sets, and $\mathcal{C}^{k+\gamma}$ rigidity respectively. We summarize the totality of the resulting picture. Given one $\mathcal{C}^{1+\gamma}$ hyperbolic Cantor set, consider the collection of

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

all Cantor sets which are $\mathcal{C}^{1+\gamma}$ conjugate to it. Within this collection is a distinguished subcollection, its limit sets. The free semigroup on two generators acts on the conjugacy class (by rescaling subsets of the next level); the limit sets are an attractor for this action, and the scenery process can be described as what one sees when walking out a branch of the tree of the semigroup. Limit sets are exactly the *ratio sets* (see §2) built from the associated scaling function. Within the big collection are subcollections with higher degrees of smoothness. The big collection forms an infinite-dimensional Banach manifold, naturally identified with a factor of the $\mathcal{C}^{1+\gamma}$ diffeomorphisms of the interval, after one Cantor set has been chosen as a base point (the diffeomorphisms are the conjugacies to this set). The subcollections nest down as smoothness increases, and by rigidity these smoothness classes are also conjugacy classes. Contained in the intersection of them all is the collection of limit sets, with the highest possible smoothness. Choosing one of them as a common base point, these subcollections are naturally identified with the $\mathcal{C}^{k+\gamma}$ diffeomorphisms of the interval. Each is a Banach manifold in its own topology, and is dense in a larger collection with respect to its topology. The free semigroup acts on each manifold. Its points are drawn exponentially fast in the $\mathcal{C}^1$ norm toward the common attractor: the collection of limit sets, which form a compact subset of the Banach manifold.

In the course of our paper we give careful proofs of several of Sullivan's theorems ([**Su1**] is extremely sketchy). In some cases our different point of view leads us to different arguments from those indicated in [**Su1**]. We will describe our approach and results more fully after a further explanation of Sullivan's ideas.

Sullivan's differentiable structures. We begin with an ordered topological Cantor set, i.e. a space which is homeomorphic and order isomorphic to the usual middle-third Cantor set. For convenience we use $\Sigma^+ \equiv \Pi_0^\infty\{0, 1\}$, together with the product topology, and with the lexicographic order. *Charts* are defined to be order-preserving homeomorphisms into $\mathbb{R}$; two charts ζ, ξ are $\mathcal{C}^{k+\gamma}$ *compatible* if $\zeta \circ \xi^{-1}$ extends, with that degree of smoothness, to a diffeomorphism defined on neighborhoods of the embedded sets. A *linear* $\mathcal{C}(k, \alpha)$ *differentiable structure* on Σ^+ will be a *maximal atlas* (a maximal compatible collection of charts). Here, following [**Su1**], $\mathcal{C}(k, \alpha)$ denotes all maps which are $\mathcal{C}^{k+\gamma}$ for some $\gamma \in (0, 1]$. Therefore, a $\mathcal{C}(k, \alpha)$ linear differentiable structure determines and is determined by a class of Cantor sets embedded in the real line, equivalent by $\mathcal{C}(k, \alpha)$ changes of coordinates.

For simplicity, we are restricting our attention to charts which are order preserving and globally defined. We mention that the word 'linear' is being used in two ways: when dynamics is introduced on these sets, it will usually be nonlinear; the differentiable structures are called linear because they come from embeddings in the line. (Alternative theories might have charts mapping Σ^+ to a product of Cantor sets, or to a subset of some fractal curve!)

Via the homeomorphism from Σ^+, an embedded set C comes equipped with the dynamics of the shift map σ on Σ^+. The set also inherits from Σ^+ a nested hierarchy of intervals, corresponding to finite words (*cylinder sets*) in Σ^+. Sullivan uses the shift map to define $\mathcal{C}(1, \alpha)$ *hyperbolic Cantor sets* (see §1 below) and the nested intervals to

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

define the *ratio geometry* of a Cantor set. This assigns to each interval the triple (l, g, r) of length ratios of the left subinterval, middle gap and right subinterval respectively. The hypothesis that an embedded Cantor set C is hyperbolic $\mathcal{C}(1, \alpha)$ is enough to show that a limiting ratio geometry, recorded by the *scaling function*, exists. The ratio geometry is bounded away from 0 and 1, a condition called *bounded geometry*; hence the limiting values (a, b, c) are also in the interior of the simplex $\Delta = \{(a, b, c) : a + b + c = 1\}$. Convergence to the scaling function is taken along inverse branches of σ, which are indexed by points of an abstract topological Cantor set called the *dual Cantor set*. Thus the scaling function maps the dual Cantor set to Δ, onto a compact subset of its interior. Convergence is exponentially fast, and the scaling function is Hölder continuous.

Locations in the Cantor set correspond to forward images under σ, since the digits of Σ^+ tell whether the orbit of a point lies in the left or right third of C. To study smoothness of a conjugacy or an expanding map, one expects of course to use the locations to estimate difference quotients. However, since convergence to the scaling function is taken along *inverse* branches of σ, as the scale gets smaller and smaller, the locations jump all over the set.

The first remarkable result from §§1–3 of [**Su1**] is that while indeed one cannot compute the derivative of the shift map from the scaling function, nevertheless this function contains complete information about $\mathcal{C}(1, \alpha)$ differentiable structures.

More precisely, one has the following. As we have already mentioned, (1) a hyperbolic $\mathcal{C}(1, \alpha)$ Cantor set has a bounded Hölder scaling function. Next, (2) this depends only on the differentiable structure, i.e. it is the same for $\mathcal{C}(1, \alpha)$ conjugate Cantor sets. Conversely, (3) an embedded Cantor set which has a bounded Hölder scaling function is in fact $\mathcal{C}(1, \alpha)$ hyperbolic. Finally, (4) in this case the $\mathcal{C}(1, \alpha)$ differentiable structure is determined by the scaling function. In other words, two hyperbolic $\mathcal{C}(1, \alpha)$ Cantor sets with the same scaling function are $\mathcal{C}(1, \alpha)$ conjugate. In summary, the bounded Hölder scaling function gives an intrinsic characterization of the differentiable structure, in the sense that no embedding need be specified.

Furthermore, quoting [**Su1**]: '. . . if the structure admits a $\mathcal{C}(k, \alpha)$ refinement so that the shift is $\mathcal{C}(k, \alpha)$, this structure is also determined uniquely by the same scaling function . . .'. Stated as a result about representatives instead of the entire equivalence class, this can be interpreted as a rigidity theorem: if two $\mathcal{C}(k, \alpha)$ hyperbolic Cantor sets are conjugate by a map which is $\mathcal{C}(1, \alpha)$, then that map (possibly with a different extension to the gaps) is in fact already $\mathcal{C}(k, \alpha)$.

Summary of results. We include in this paper careful statements and proofs, in particular, of (1), (2) and (4) above, and of $\mathcal{C}(k, \alpha)$ rigidity (we mention that our use of the term 'rigidity' is different from that in §5 of [**Su1**]). Since we are interested in the geometry of representatives rather than the equivalence class, all these results are stated in terms of the conjugacy of embedded sets rather than the classification of differentiable structures. As we said above, the reason for this emphasis is that our primary goal is to study the scenery process, and all the sets in the scenery process are the same up to conjugacy.

A small technical difference to [**Su1**] is that we use $\mathcal{C}^{k+\gamma}$ rather than $\mathcal{C}(k, \gamma)$ throughout. We do this because it gives sharper statements. Thus, for example, for

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

rigidity we show that $\mathcal{C}^1$ conjugacy implies $\mathcal{C}^{k+\gamma}$ conjugacy.

In part because of our change in focus, we give a different proof from that suggested in [**Su1**] of (4). Each approach has its own advantages. Sullivan's method, a direct estimate of the derivative by difference quotients using sums of gap lengths, gives a unified way of proving (3) as well as (4). However, one also needs to cite an extension lemma, which is not included in [**Su1**]. On the other hand, our approach gives a unified treatment of $\mathcal{C}^1$ conjugacy and rigidity, and avoids calling on the separate extension lemma. We do not prove (3) here, but will give a full proof (along the lines of [**Su1**]) elsewhere.

Our proof of rigidity is intimately connected to the study of the scenery process. We proceed as follows.

First we use limiting conjugacies to construct a set-valued analogue of the scaling function. This function, $y \mapsto C^y$, is defined for y in the dual Cantor set, is Hölder continuous with respect to a metric derived from the corresponding Hausdorff measures and has as its range a compact subset of the collection of all subsets of $[0, 1]$ in that measure metric, and also in the Hausdorff metric on sets. The scaling dynamics enters by interpreting the dual Cantor set as the past of the natural extension of the expanding map on C; the scenery process $C_{n,x}$ is then asymptotically given by evaluating the shift on any extension $\underline{x} = (y, x)$ of $x \in C$. Since the limit sets were constructed by conjugacies, one can apply a lemma from the appendix of [**Su1**] to help determine their degree of smoothness: we show they have the highest degree of smoothness ($\mathcal{C}^{k+\gamma}$ for some $k \geq 1$, $\mathcal{C}^\infty$ or $\mathcal{C}^\omega$) which occurs in the $\mathcal{C}^{1+\gamma}$ conjugacy class of C.

Next, the proof of rigidity follows as a corollary. Given two hyperbolic $\mathcal{C}^{k+\gamma}$ Cantor sets, if they are $\mathcal{C}^1$ conjugate they have the same scaling function. Hence they have the same limit sets, which are *ratio Cantor sets* constructed from this function. Choosing one of these to act as an intermediary, the composition should also be $\mathcal{C}^{k+\gamma}$. However, the maps may be defined differently on the gaps, which would lead back to the extension problem mentioned before. But now one has a simpler solution: a choice is made on the middle third, and the rest of the definition follows automatically from the dynamics. This completes the proof of rigidity. In summary: if two $\mathcal{C}^{k+\gamma}, \mathcal{C}^\infty$ or $\mathcal{C}^\omega$ hyperbolic Cantor sets are $\mathcal{C}^1$ conjugate, then this conjugacy (with a different extension) is already $\mathcal{C}^{k+\gamma}, \mathcal{C}^\infty$ or $\mathcal{C}^\omega$ respectively.

This leads, then, to the overall picture which is summarized at the end of the first part of the Introduction.

Note on doubling maps. We next discuss, more completely, the relation of [**Su1**] and the present paper to Sullivan's work on renormalization theory. We have already described one aspect of this: for the fixed point folding map, one can introduce a second $\mathcal{C}^{1+\alpha}$ hyperbolic map on its attracting Cantor set, and thus can apply hyperbolic methods (the bounded distortion property, symbolic dynamics, Gibbs theory) in the study of its small-scale geometry.

However, in Sullivan's theory, in addition to this hyperbolic map on the Cantor set, a second expanding map is also of key importance. This is a *doubling map of the circle*, by which we mean a degree-two map of the circle, i.e. one which is topologically conjugate

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

to $z \mapsto z^2$. This map is defined, up to smooth conjugacy, by (roughly speaking) the action on external angles to its Julia set of the complex extension of the folding map (for more precision see §VI.4 of [**deMvS**]).

In fact, all the results of the present paper also hold for $\mathcal{C}^{k+\alpha}$ hyperbolic doubling maps. A scenery process enters the picture as follows. Preimages of the fixed point partition the circle into two intervals. Pulling back this partition gives a hierarchy of nested tilings of the interval, which can be thought of as having a fractal geometry: the tiled interval is like a Cantor set which has been expanded to close the middle gaps, with the tilings marking where these gaps had once been.

A scaling function is defined from the tilings in the obvious way. With an appropriate Hausdorff-type metric on the space of nested tilings, one has an analogue of limit sets and can define the scenery process and flow. See [**BeFi3**] and also the related work of Pinto and Rand discussed below.

The doubling map plays the following role in Sullivan's theory. Sullivan's idea for proving the first Feigenbaum–Coullet–Tresser conjecture on the existence and uniqueness of a renormalization fixed point is to show that the renormalization operator acts locally as a contraction on a Teichmüller space of folding maps. This space is a given topological conjugacy class of folding maps modulo an identification: one associates to each of the folding maps its smooth conjugacy class of doubling maps, which induces an equivalence relation on the folding maps.

The next step is to classify invariant differentiable structures for doubling maps. This classification is in terms of an invariant which can serve as a 'modulus' of the structure. The idea is to then put a complex structure on the set of moduli as in classical Teichmüller theory, and next to use Teichmüller theory in the proof of contraction. For the parallel theory of hyperbolic Cantor sets, this 'modulus' is the bounded Hölder scaling function. For the doubling maps (see [**Su3**]) the modulus can be nicely linked with classical surface theory: it is again given by a scaling function, but equivalently it is given by a conformal structure on a *Riemann surface lamination*: a compact laminated space with solenoidal cross-section.

The point we wish to emphasize is the following. In Sullivan's development of renormalization theory, two very different hyperbolic maps become associated to the renormalization fixed point folding map: a doubling map of the circle, defined up to smooth conjugacy, and a specific $\mathcal{C}^{1+\alpha}$ hyperbolic map on the Cantor set. Then, remarkably, much the same machinery gets used, in different ways, in the proofs of these two very different conjectures.

It is important to mention here that things are not quite so simple away from the fixed point. For this case one still has a single doubling map; however now the hyperbolic Cantor set map should be replaced by a *sequence* of maps along a sequence of Cantor sets: a *Markov family* in the sense of Rand (see the discussion below).

Related work. One knows, from the rigidity theorem, that the maximum degree of smoothness occurring in the $\mathcal{C}^{1+\gamma}$ conjugacy class should be encoded somehow in the scaling function. Work of Tangerman and Przytycki gives one way of recovering that information [**TPr**]. Pinto and Rand ([**PR2**, §5] and personal communication) and Sullivan

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

(personal communication) have suggested other approaches, in a related situation. It would be nice to understand these different points of view in a unified way.

The rigidity theorem is also stated by Tangerman and Przytycki; it is proved as a corollary of their main result. Their approach is quite different from ours and, in particular, does make use of an extension lemma (a version of Whitney's extension theorem).

We have already mentioned the notion of Markov family and its role in renormalization theory. Rand introduced this notion [**Ra**] to help study the relationship between scaling functions and smooth conjugacy in a general situation where one has a sequence of expanding maps of the interval, rather than a single map. See also [**Pi**] and [**PR1, 2**]: our Theorems 5.3, 5.4 and 5.9 below (concerning limit conjugacies) are closely related to Theorem 26 of Pinto's thesis [**Pi**]. What we call the scenery process (see §5 below) corresponds in [**Pi**] to the following. Given a Markov family, Pinto defines a limiting family called a *renormalization Markov family* [**Pi**]. For the special case of a constant Markov family, i.e. for a single hyperbolic map, the sequence of Cantor sets on which this limiting family is defined is a path $C(\sigma^n(\underline{x}))$ in the scenery process. The choice of a point $\underline{x}$ in the shift space determines a particular renormalization family. (We became aware of Tangerman and Przytycki's preprint and of Rand and Pinto's work after the first version of the present paper—an IHES preprint, July 1992—was completed.)

Markov families provide a flexible and general framework for approaching a variety of problems. Interesting examples of Markov families come from both renormalization theory and from random dynamics and, indeed, this notion provides a link between these two fields (see [**AFi**]). In Pinto and Rand's papers, the focus is on constructing smooth conjugacies of Markov families given by the renormalization of one-dimensional maps, including certain circle diffeomorphisms and folding maps: for instance, the folding maps discussed above which are not a fixed or periodic point under renormalization. For an example from random dynamics one can think of a Cantor set generated by a sequence of random perturbations of a single hyperbolic map. In this paper we consider only constant families, i.e. Cantor sets generated by one hyperbolic map. In fact, all the results and methods of this paper hold for (two-sided) Markov families, including such random hyperbolic Cantor sets. This is explained more fully in [**BeFi3**].

These ideas fit into a broader context. As we show in [**AFi**], the basic notion of mapping families (by which we mean a sequence of maps along a sequence of spaces) makes sense for higher dimensions, with much of the standard theory of a single hyperbolic transformation (possibly with singularities, as for pseudo-Anosov maps) carrying through for hyperbolic mapping families: shadowing, the stable manifold theorem, structural stability, and the existence of Markov partitions. These code the family to a *non*-stationary symbolic dynamics; one can then prove the Ruelle–Perron–Frobenius theorem and hence develop Gibbs theory (see also [**BoG**]) and apply this in turn, for families on surfaces, when classifying differentiable structures.

Interesting work on the small-scale geometry of certain fractal sets, in quite different settings, has been done by Furstenberg and Tan Lei. Tan Lei in [**T**] proves the beautiful theorem that certain nonhyperbolic Julia sets, corresponding to Misiurewicz points in the boundary of the Mandelbrot set ∂M, are asymptotically the same as ∂M at that

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

point. These points form a countable dense subset of ∂M, yet the general case is still far from completely understood. This type of asymptotic limit, as well as what we have here called limit sets, provide examples of Furstenberg's general notion of the *microsets* of a subset of Euclidean space (lectures and personal communication). These are by definition all the limiting sets given by rescaling nested subsets by a sequence of affine expansions. Furstenberg applies this in a continuation of the analysis begun in [**Fu**] for determining the Hausdorff dimension of certain sets: intersections of generic translates of linear Cantor sets, and intersections of linear Cantor sets in the plane with foliations of straight lines at a generic slope. His study of these matters is related to the 'times 2 times 3' circle of problems in ergodic theory. An interesting and important area of research is to develop similar results in a nonlinear setting, e.g. for general smooth foliations or for nonlinear Cantor sets.

1. *Two ways of building Cantor sets*

1.1. *Hyperbolic Cantor sets, Hausdorff and Gibbs measures.* We start with the usual middle-third Cantor set. Let S denote the 2-1 map on the middle-third set C defined by $x \mapsto 3x \pmod 1$. The Hausdorff dimension of C is $d = \log 2/\log 3$. Writing H^d for d-dimensional Hausdorff measure and μ for the restriction $\mu = H^d|_C$, we recall that μ is a Borel probability measure (total mass $= 1$) which is invariant under S. The triple (C, S, μ) is canonically isomorphic to the one-sided Bernoulli left shift σ on $\Sigma^+ \equiv \Pi_0^\infty\{0, 1\}$, with infinite $(\frac{1}{2}, \frac{1}{2})$ coin-tossing measure; the correspondence is given by $\pi : (x_0x_1 \ldots) \mapsto x$, where $x \in C$ has ternary expansion

$$x = \sum_{i=1}^{\infty} 2x_i 3^{-i}.$$

A *hyperbolic $\mathcal{C}^{1+\gamma}$ Cantor set* C by definition also has an expanding dynamics $S : C \longrightarrow C$, but now instead of having straight lines as for $3x \pmod 1$, the graph of S may be nonlinear (see Figure 1).

To construct such a set, one defines C as a limit from two contraction mappings $\varphi_0, \varphi_1 : I \longrightarrow I$. We first consider the case where these maps are orientation-preserving, and are *strict* contractions in the sense that the derivatives satisfy $0 < \alpha < D\varphi_i < \beta < 1$. We also require that

$$0 = \varphi_0(0) < \varphi_0(1) < \varphi_1(0) < \varphi_1(1) = 1.$$

This implies that the intervals $I_0 \equiv \varphi_0(I), I_1 \equiv \varphi_1(I)$ are disjoint. We assume that φ_0, φ_1 are $\mathcal{C}^{1+\gamma}$ maps for some $\gamma \in (0, 1]$. Here $\mathcal{C}^{k+\gamma}$ means the kth derivative $D^k\varphi_i$ is Hölder continuous with exponent γ; note that $\mathcal{C}^{1+1}$ means $D\varphi_i$ is Lipschitz, so $\mathcal{C}^2$ implies $\mathcal{C}^{1+1}$ (by compactness) but not conversely. (Exponent $\gamma > 1$ is excluded because in that case, since the domain I is connected, φ_i is identically constant hence immediately of order $\mathcal{C}^\infty$, while the whole purpose of Hölder conditions is to have *intermediate* grades of smoothness.) Our convention for Hölder continuity will be: if we are given that f satisfies $|f(x) - f(y)| \le c_0|x - y|^\gamma$, then we say f is *Hölder continuous* with *Hölder constant* c_0 and *Hölder exponent* γ. We remark that Sullivan in [**Su1**] works instead with $\mathcal{C}(k, \alpha)$ maps; these are defined to be the collection of $\mathcal{C}^{k+\gamma}$ maps for all $\gamma \in (0, 1]$.

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

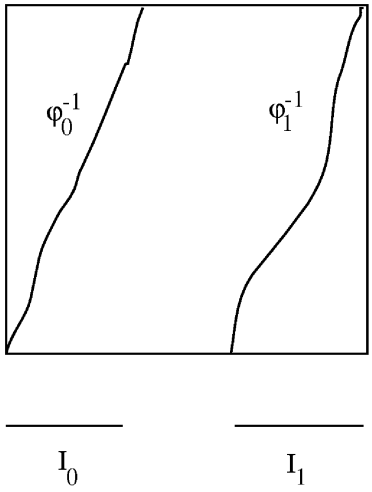


FIGURE 1.

We define $S : I_0 \cup I_1 \to I$ to be the map with inverse branches φ_0, φ_1. Note that since $D\varphi_i$ are bounded away from 0 and ∞, it follows that S is $\mathcal{C}^{1+\gamma}$ with the same Hölder exponent, but with different Hölder constant. Inductively, form

$$I_{x_0 \ldots x_n} = \varphi_{x_0}(\varphi_{x_1} \ldots (\varphi_{x_n}(I))),$$

where $x_k \in \{0, 1\}$; $\bigcup I_{x_0 \ldots x_n}$ (union over all choices, with n fixed) is the nth level approximation to the Cantor set, C, defined as

$$C = \bigcap_{n=0}^{\infty} \bigcup I_{x_0 \ldots x_n}.$$

The restriction of the map S to C maps C to itself and is (just as for the middle-third set) conjugate to the Bernoulli shift (Σ^+, σ) via the map $\pi : (x_0 x_1 \ldots) \mapsto x$, where x is the unique element of $\cap_{n=0}^{\infty} I_{x_0 \ldots x_n}$.

A set C together with map $S : I_0 \cup I_1 \to I$ defined in this way from strict contractions φ_0, φ_1, will be called a *strictly hyperbolic $\mathcal{C}^{1+\gamma}$ Cantor set (with map)*.

Sometimes we are only interested in the dynamics on C itself. Knowledge of this *restricted map* $S|_C$ is equivalent to knowing how C is coded by Σ^+. We will refer to a set together with this labeling as a *marked* Cantor set. When we forget about this coding, C will be referred to as the *underlying* Cantor set of (C, S).

More generally, a *hyperbolic $\mathcal{C}^{1+\gamma}$* Cantor set is defined as follows. (Again we assume that $\varphi_0, \varphi_1 : I \to I$ are order-preserving $\mathcal{C}^{1+\gamma}$ diffeomorphisms such that $\varphi_0(0) = 0, \varphi_1(1) = 1$, and $\varphi_0(1) < \varphi_1(0)$. We also assume, as before, that there exists α with $0 < \alpha < D\varphi_i < 1$. However now the upper bound is replaced by one of the two equivalent conditions which follow. We write $\varphi_{x_0 \ldots x_n} \equiv \varphi_{x_0} \circ \cdots \circ \varphi_{x_n}$.)

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

PROPOSITION 1.1. *The following are equivalent:*
(i) $\exists \beta < 1$ *and* $N \geq 0$ *such that for all* $n > N$, *for any* $x_0 \dots x_n$, $D\varphi_{x_0 \dots x_n} < \beta^n$;
(ii) $\exists c > 0$ *and* $\tilde{\beta} < 1$ *such that for all* n, *for any* $x_0 \dots x_n$, $D\varphi_{x_0 \dots x_n} < c\tilde{\beta}^n$.

Proof. To pass from (ii) to (i), any β which satisfies $\tilde{\beta} < \beta < 1$ will work. For the converse, take $c = \max\{(D\varphi_{x_0 \dots x_n}) \cdot \beta^{-n} \mid n \leq N\}$ and $\tilde{\beta} = \beta$. □

Two $\mathcal{C}^{1+\gamma}$ hyperbolic Cantor sets $C, \widetilde{C}$ with maps $S, \widetilde{S}$ will be said to be *continuously conjugate*, or $\mathcal{C}^0$ *conjugate*, if there is an order-preserving homeomorphism $\Phi : I \to I$ such that for all $x \in I_0 \cup I_1$, $\widetilde{S} \circ \Phi(x) = \Phi \circ S(x)$. We say $C, \widetilde{C}$ are $\mathcal{C}^{k+\gamma}, \mathcal{C}^\infty, \mathcal{C}^\omega$ *conjugate* if Φ and its inverse have that degree of smoothness. Note that from the definition, for each n, $\Phi(I_{x_0 \dots x_n}) = \widetilde{I}_{x_0 \dots x_n}$. Therefore the conjugacy induces the identity map on the corresponding shift spaces.

When Φ is defined (as above) on all of I we will also call it a *full* conjugacy. A *restricted* conjugacy is a conjugacy between the Cantor sets which can be extended to a full conjugacy.

We immediately have the following lemma.

LEMMA 1.2. *Let* (C, S) *be a hyperbolic Cantor set, and assume* $\Phi : I \to I$ *is a* $\mathcal{C}^{1+\gamma}$ *diffeomorphism. Define sets* $\widetilde{C} \equiv \Phi(C)$, $\widetilde{I}_i \equiv \Phi(I_i)$ *for* $i = 0, 1$, *and define the map* $\widetilde{S} : \widetilde{I}_0 \cup \widetilde{I}_1 \to I$ *by* $\widetilde{S} \equiv \Phi \circ S \circ \Phi^{-1}$. *Then* $(\widetilde{C}, \widetilde{S})$ *is also a hyperbolic* $\mathcal{C}^{1+\gamma}$ *Cantor set.*

Note that in this lemma, using condition (i), $\tilde{\beta}$ stays the same but the constant c may change. Similarly, using (ii), N may change while β remains the same.

Thus, in particular, the $\mathcal{C}^{1+\gamma}$ conjugate of a strictly hyperbolic set is still hyperbolic, though strictness may be lost. We remark without proof that a converse holds: by a well-known theorem due in its original form to Mather, any hyperbolic Cantor set is conjugate to a strictly hyperbolic set (without changing the order of differentiability); see for example [**HP**]. Therefore, if one is studying properties invariant with respect to the equivalence relation given by conjugacy, one might as well begin with the assumption that $c = 1$ in (ii); this situation occurs often in the dynamical systems literature. The new metric on I is referred to as an *adapted metric* for the hyperbolic map S. However, for our purposes it will be important to use the original metric, otherwise the notion of the scenery process will lose its meaning. This will become clear in §5.

We recall from the theory of Bowen and Ruelle ([**Bo1, 2**], [**Ru**]; see also [**Be**]) that the dimension d of a hyperbolic $\mathcal{C}^{1+\gamma}$ Cantor set C is strictly between 0 and 1 and that *Hausdorff (or conformal) measure* $\mu = H^d|_C$ has a unique normalized invariant version ν, called the *Gibbs measure* (or *Gibbs state*). For the middle-third set $\mu = \nu$; in general they are boundedly equivalent (i.e. the Radon–Nikodym derivative is bounded away from 0 and ∞); ν is defined so as to be a probability measure while μ may have total mass $\neq 1$. As in [**BeFi1**] we will need to use both measures.

We remark that all the results in this paper generalize with minor notational changes to the following situation. The maps φ_i are also allowed to be orientation-reversing; there may be more than two maps, $\varphi_1, \dots, \varphi_k$, and the Cantor set is constructed by selecting the maps with respect to some subshift of finite type Σ_A on k symbols instead of the full two-shift Σ.

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

1.2. *Ratio Cantor sets.* As before, Σ^+ denotes $\Pi_0^\infty\{0,1\}$ and we now define: $\Sigma^- \equiv \Pi_{-\infty}^{-1}\{0,1\}$, $\Sigma \equiv \Pi_{-\infty}^{\infty}\{0,1\}$. We will write $y = (\dots y_{-2}y_{-1})$ for $y \in \Sigma^-$, $x = (x_0x_1\dots) \in \Sigma^+$ and $\underline{x} = (y,x) = (\dots x_{-2}x_{-1}.x_0x_1\dots) = (\dots y_{-2}y_{-1}.x_0x_1\dots)$ for a point in Σ. We will let σ denote both the (full) left shift on Σ and the left shift (with truncation) on Σ^+. Σ^+ is known as the *future* of Σ, and Σ^- as its *past*.

Write Δ for the unit simplex in $\mathbb{R}^3$ and $\operatorname{int}\Delta$ for its interior. Let R be a continuous function from Σ^- to $\operatorname{int}\Delta$, and write the components as $R = (R_{\mathrm{l}}, R_{\mathrm{g}}, R_{\mathrm{r}})$. These letters will stand for *left, gap* and *right* respectively; by definition they add to 1 and each is strictly positive. For $\underline{x} = (y,x) \in \Sigma$ we will also think of R as a function on Σ by defining $R(\underline{x}) \equiv R(y)$.

Given the function R, we will define for each $y \in \Sigma^-$ the *ratio Cantor set* $C^y \subseteq [0,1]$ so as to satisfy the following: at each stage subintervals will have length ratios $R(\sigma^n(\underline{x}))$. Thus, we first define $I_0^y = [0, R_{\mathrm{l}}(y)]$, $I_1^y = [1 - R_{\mathrm{r}}(y), 1]$. The left interval I_0^y has subintervals I_{00}^y, I_{01}^y which are defined to have lengths in the ratios

$$\frac{|I_{00}^y|}{|I_0^y|} = R_{\mathrm{l}}(\dots y_{-2}y_{-1}0.), \quad \frac{|I_{01}^y|}{|I_0^y|} = R_{\mathrm{r}}(\dots y_{-2}y_{-1}1.)$$

and left and right endpoints the same as those of I_0^y, respectively. Inductively, for $x \in \Sigma^+$ and $\underline{x} = (y,x)$, $I_{x_0\dots x_{n+1}}^y$ is a subinterval of $I_{x_0\dots x_n}^y$ with length ratio

$$\frac{|I_{x_0\dots x_{n+1}}^y|}{|I_{x_0\dots x_n}^y|} = R_*(\sigma^n\underline{x}).$$

Here $* = \mathrm{l}$, i.e. this is the left subinterval, if $x_{n+1} = 0$, and $* = \mathrm{r}$, i.e. the right subinterval, if $x_{n+1} = 1$.

Note that the fact that R depends only on the past coordinates $y \in \Sigma^-$ of $\underline{x}$ is what makes this well-defined, since therefore the ratio is the same for each other point in that subinterval.

Finally we form the set C^y as before, defining

$$C^y = \bigcap_{n=0}^{\infty} \bigcup I_{x_0\dots x_n}^y.$$

The simplest example is again the middle-third set: taking $R(y) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ for all $y \in \Sigma^-$, C^y is the middle-third set C for each y.

Since R is a continuous map from a compact set into the interior of Δ by assumption, hence strictly into the interior, these intervals $I_{x_0x_1\dots}$ nest down to a single point in $[0,1]$. Hence each C^y inherits from Σ^+ the dynamics of the shift map σ (we will also write σ for this map on C^y).

For R assumed to be Hölder continuous with some exponent $\alpha > 0$ (which will be the case in this paper) it turns out that each ratio Cantor set C^y is also a hyperbolic $\mathcal{C}^{1+\gamma}$ Cantor set. (What needs to be shown is that σ is $\mathcal{C}^{1+\gamma}$ on C^y for some γ with $0 < \gamma < 1$, which follows by a bounded distortion argument, and that it can be extended to a map S on $I_0^y \cup I_1^y$ without losing any smoothness and which is hyperbolic. This can be proved from a lemma of Sullivan, (3) in the Introduction above. For a full proof see [**BeFi3**].) Therefore it also has a Gibbs measure equivalent to Hausdorff measure.

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

The dimensions of C^y and the Gibbs state, viewed as a measure on Σ^+, are the same for each $y \in \Sigma^-$ for the following reason: for $y, w \in \Sigma^-$, C^y and C^w have the same scaling function and hence are $\mathcal{C}^{1+\gamma}$ conjugate (see §7).

2. *Statement of the problem: bounded distortion*

Let C be a hyperbolic Cantor set. We wish to describe the geometry of the sequence of nested Cantor sets one sees along the way when zooming down towards a point.

The notation will be as follows. For $x \in C$ with $x = \pi(x_0 x_1 \ldots)$, let $C_{n,x}$ be the set $C \cap I_{x_0 \ldots x_n}$ affinely rescaled to the unit interval, so as to have endpoints at 0 and 1.

Write $\mu_{n,x}$ for the corresponding Hausdorff measure, the restriction of H^d to the set $C_{n,x}$. Thus we want to see how the sequence of sets, and of the corresponding measures, varies as $n \longrightarrow \infty$.

The first obstacle we encounter is that you do not get from the interval $I_{x_0 \ldots x_n}$ to its subinterval $I_{x_0 \ldots x_{n+1}}$ by one application of φ_0 (or φ_1). Instead you have

$$I_{x_0 \ldots x_{n+1}} = \varphi_{x_0} \ldots \varphi_{x_{n+1}}(I) = \varphi_{x_0} \ldots \varphi_{x_{n+1}}(\varphi_{x_n}^{-1} \ldots \varphi_{x_0}^{-1}(I_{x_0 \ldots x_n})),$$

and since the maps do not commute, you have to go all the way back up and down the tree again, with more nonlinearity introduced each time. To control this nonlinearity we will use the well-known bounded distortion property in the following variation. For a proof see [**ShSu**], [**Mañé**] or Lemma 6.4 below. We learned this version of bounded distortion from M. Urbanski.

THEOREM 2.1. (Classical) *With S as above, $\exists K > 0$ such that for all n, for any $\delta > 0$, if J is an interval such that $S^m|_J$ is one-to-one and the image $S^m(J)$ has diameter less than δ, then for all $x, y \in J$,*

$$e^{-K\delta^\gamma} < \left| \frac{DS^m x}{DS^m y} \right| < e^{K\delta^\gamma}.$$

We mention that one sees from the proof that if c is the Hölder constant for $\log|DS|$, then the constant K is given by $K = c\beta^\gamma/(1-\beta^\gamma)$.

As a consequence of this theorem, since $0 < \alpha < D\varphi_i < \beta < 1$ implies that $\alpha^n < |I_{w_o \ldots w_n}| < \beta^n$ for any w, we have the following corollary.

COROLLARY 2.2. *For any $m, n \geq 0$ and any $w \in \Pi_0^\infty\{0,1\}$ one has for all $x, y \in I_{w_0 \ldots w_{n+m}}$,*

$$e^{-K\beta^{n\gamma}} < \frac{|DS^m(x)|}{|DS^m(y)|} < e^{K\beta^{n\gamma}}.$$

The set $C_{k,x}$ belongs to the collection of 2^k Cantor sets at level k in the tree (rescaled). We want to understand the geometry of the sets in this collection.

A first approximation is the original set itself (at level 0). But by bounded distortion, for n large the 2^n sets at level n provide much better models for the 2^{m+n} sets at level $k = m + n$ and, moreover (and this is the strength of bounded distortion) this is true for *all* m simultaneously. The reason is that since by definition

$$S^m(I_{y_0 \ldots y_m x_0 \ldots x_n}) = I_{x_0 \ldots x_n},$$

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

and since $I_{x_0 \dots x_n}$ has small diameter, by the corollary the derivative of S^m is close to constant, hence S^m is close to linear.

In summary, consider all the Cantor subsets which have this same image under S^m to be grouped in one equivalence class. The 2^k sets at level k are split into 2^n equivalence classes, each with 2^m members which all have approximately the same geometry (but whose locations are scattered throughout the space!). As we scale down toward a point x, we are seeing sets given by these approximations.

Note that the equivalence class of a given interval at level k depends on the immediately previous n branches, rather than on its initial branching structure. We will see in the next section how Sullivan uses this observation to study the asymptotics, associating to the Cantor set a function R like that used to define the ratio Cantor sets in the previous section. Then, in §5, we will show that the sequence of sets one sees in C is asymptotically the same as that for the ratio Cantor sets C^y, and for the set C^y the nested sequence of subsets has an exact description.

Remark. The scenery process of a ratio Cantor set. Each subset in this sequence is itself a ratio Cantor set. Moreover, the sets change in the following way. Writing $C^y_{n,x}$ for the set $C^y \cap I^y_{x_0 \dots x_n}$ affinely rescaled to $[0, 1]$ as above, one has (immediately from the definition) that $C^y_{n,x} = C^{\sigma^n(\underline{x})}$ for all $n \geq 0$.

With the Gibbs measure on the full shift, this gives a stationary, set-valued process which in forward time describes exactly what one sees as one zooms down toward Hausdorff-almost every point in the ratio Cantor set C^y. See §§4 and 5 below.

3. *Sullivan's scaling function*

Now we return to the study of a hyperbolic Cantor set C. Instead of treating the structure of the entire set $I_{x_0 \dots x_n} \cap C$, which is what we have been emphasizing so far, Sullivan focuses on the information contained in the first step of its construction, given by the relative lengths of the subintervals of $I_{x_0 \dots x_n}$. These subintervals are the left third $I_{x_0 \dots x_n 0}$, right third $I_{x_0 \dots x_n 1}$ and middle gap written $G_{x_0 \dots x_n}$. We normalize the lengths of these three intervals, defining for $x \in C$ and $n \geq 0$, where $x = \pi(x_0 x_1 \dots)$,

$$R_{n,x} = (|I_{x_0 \dots x_n 0}|, |G_{x_0 \dots x_n}|, |I_{x_0 \dots x_n 1}|)/|I_{x_0 \dots x_n}|.$$

Sullivan calls this the *ratio geometry function* of C; it maps $\mathbb{N} \times \mathbb{C}$ to the interior of the unit simplex $\Delta \subseteq \mathbb{R}^3$ and determines C uniquely (one simply constructs C to have these ratios).

Next we write, for $y = (\dots y_{-2} y_{-1})$ in $\widetilde{C} \equiv \Pi^{-1}_{-\infty} \{0, 1\}$,

$$R_n(y) = (|I_{y_{-n} \dots y_{-1} 0}|, |G_{y_{-n} \dots y_{-1}}|, |I_{y_{-n} \dots y_{-1} 1}|)/|I_{y_{-n} \dots y_{-1}}|.$$

Following Sullivan, it is nice to think of $\widetilde{C}$ as a distinct Cantor set, dual to C (and called the *dual Cantor set*). Later for the dynamical interpretation we will instead view $\widetilde{C}$ as Σ^-, that is, as the past coordinates of the full shift $\Sigma \equiv \Pi^{\infty}_{-\infty} \{0, 1\}$. We will use whichever symbol ($\widetilde{C}$ or Σ^-) is most appropriate in the context.

As in [**Bo1**], for any $\beta \in (0, 1)$, the β-metric on Σ^- (which defines what is meant below by Hölder continuity) is taken to be $d_\beta(y, w) = \beta^n$, where n is the greatest positive

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

integer such that $y_{-n} \dots y_{-1} = w_{-n} \dots w_{-1}$. When Σ^- is thought of as $\widetilde{C}$, i.e. as the dual to a specific hyperbolic Cantor set C, we choose β to be (as before) the upper bound on $D\varphi_i$. We mention that if β is replaced by some other number $\tilde{\beta} \in (0, 1)$, then the metrics are related by

$$d_{\tilde{\beta}} = (d_\beta)^{\log \tilde{\beta}/\log \beta},$$

and the Hölder exponent γ for R in the statement of the next theorem would change to $\gamma \cdot (\log \tilde{\beta}/\log \beta)$.

THEOREM 3.1. (Sullivan) *Let C be a hyperbolic $C^{1+\gamma}$ Cantor set. For every y in the dual Cantor set $\widetilde{C}$,*

$$R(y) \equiv \lim_{n\to\infty} R_n(y)$$

exists. The convergence is of order $O(\beta^{n\gamma})$, uniformly in y, and the function R is Hölder continuous with exponent γ in the β metric. R takes values strictly in the interior of Δ.

Definition. R is called the *scaling function* of C.

Proof. We will first show that for each y, $R_n(y)$, $n = 1, 2, \dots$, is a Cauchy sequence. Since

$$S^m(I_{y_{-(n+m)}\dots y_{-1}}) = I_{y_{-n}\dots y_{-1}}$$

and similarly for the subintervals, applying the mean value theorem and bounded distortion property (Corollary 2.2) we have for all $m \geq 0$ that

$$R_n(y) = R_{n+m}(y)e^{\pm K\beta^{n\gamma}}.$$

Therefore, $R_n(y)$ is Cauchy sequence (i.e. each of its three coordinates is a Cauchy sequence) and hence it converges; call the limit $R(y)$. Next, if $y, w \in \Pi_{-\infty}^{-1}\{0, 1\}$ agree on the coordinates $-n, \dots, -1$, then since $R_n(y) = R(y)e^{\pm K\beta^{n\gamma}}$ and $R_n(y) = R_n(w)$, we have

$$R(y) = R(w)e^{\pm 2K\beta^{n\gamma}}.$$

Writing $\|\cdot\|$ for the sup norm in $\mathbb{R}^3$, this implies that, with the log taken by components,

$$\|\log R(y) - \log R(w)\| \leq 2K(d_\beta(y, w))^\gamma,$$

i.e. $\log R$ is Hölder continuous with exponent γ and therefore so is R. □

Thus (since the normalized lengths add to one) R maps $\widetilde{C}$ onto a compact subset of the interior of the unit simplex in $\mathbb{R}^3$.

4. *Dynamical versions of Sullivan's theorem*

In this and the next two sections we return to the original motivating question: what does the sequence of sets $C_{n,x}$ look like? This is exactly what one sees for the nth level Cantor set as one zooms down toward x. The answer will be given in several forms which can be thought of as dynamical versions of Sullivan's theorem. We begin by describing the behavior of the sequence of ratios $R_{n,x}$ as $n \longrightarrow \infty$.

First we state Theorem 3.1 in a dynamical form. Here it will be crucial to think of the dual Cantor set $\widetilde{C}$ as the past Σ^- of Σ. We extend the function R to Σ by defining $R(\underline{x}) = R(y)$ for $\underline{x} = (y, x)$. This function depends only on the past coordinates y of $\underline{x}$.

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

COROLLARY 4.1. *For each* $x \in C$, *for any choice of* $\underline{w} \in \Sigma$ *such that* $\exists k \geq 0$ *with* $w_k, w_{k+1} \ldots = x_k, x_{k+1} \ldots$ *(where* $x = \pi(x_0 x_1 \ldots)$*) then*

$$\|R_{n,x} - R(\sigma^n \underline{w})\| \longrightarrow 0 \quad \text{as } n \longrightarrow \infty.$$

The proof is immediate from the definitions, and in fact if the Hölder constant for R is $c > 0$ so that

$$\|R(\underline{x}) - R(\underline{z})\| \leq c d_\beta(\underline{x}, \underline{z}),$$

one has

$$\|R_{n,x} - R(\sigma^n(\underline{w}))\| \leq c\beta^{n-k};$$

here the β-metric has been extended to Σ in the natural way, with points $\underline{x}$, $\underline{w}$ having to agree on coordinates from $-n$ to n.

We note that, equivalently, if x and w are in the same stable set in (C, S), then the sequence $R_{n,x}$ (with any past) is in the stable set (in the shift on sequence space) of the sequence given by R sampled along the shift orbit of $\underline{x}$, with an exponential rate of convergence.

Probabalistic interpretation: the scaling process. As we will see, we find that for the question we have asked (how does the sequence of ratios behave), which begins as a geometrical question, the natural answer, in Corollary 4.2, will be stated in the language of probability theory. Indeed the reader with a background in probability theory as well as dynamics will immediately see that the last statement above can be interpreted as giving a pathwise convergence theorem for a stochastic process.

We remark that despite this use of probability concepts, in fact no probabalistic methods at all will be used in the proof; all the actual work is just analysis and hyperbolic dynamics, and has already been carried out in Theorem 3.1 above.

In order to clarify these points for the reader without a background in probability theory, we bring in some basic concepts from probability theory, and then review the fundamental connection of probability theory with ergodic theory. Especially good introductions to probability theory are [**Lam**] and [**GSt**]; for the link with ergodic theory see [**Bi**].

The main objects of study in probability theory are stochastic processes (the word *stochastic* just means *random*). We recall that a *stochastic process* is simply a (one- or two-sided) sequence of measurable functions f_i (known as *random variables*) defined on some probability space (Ω, ν). This is the analyst's interpretation. However, it is also important to understand the probabilist's viewpoint.

A probabilist thinks of the sequence of values $f_i \equiv f_i(\omega)$ as the primary object, with ω unknown but distributed according to the measure ν on Ω. Thus, if one knows the present value of the 'variable' f, that is f_i, one has various possibilities for its next value, f_{i+1}.

From this point of view only the process is important, not the underlying space Ω. In fact, usually one does not give the measure explicitly, or even specify what exactly the space is. Rather one specifies certain attributes of the variables, e.g. the initial

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

distribution of values of f, how that changes in time, and dependence properties such as independence or the Markov property.

However, to put things on a firm foundation one has to move back to the analysis setting and at least prove the existence of such a space and measure which will 'carry' the process. Kolmogorov's consistency theorem gives a way of doing that. Assume for simplicity that the values of f lie in a compact metric space M. We take the space Ω to be the space of *sample paths* $\{(\ldots, f_i(\omega), \ldots)\}$. This is a subset of $\Pi_0^\infty M$, and inherits from that space the product topology and the product Borel σ-algebra. Kolmogorov's theorem says that if the description of the process f_i has specified the finite joint distributions (i.e. the measures of *cylinder sets*) in a consistent way, then this defines a unique measure on Ω. (The proof uses weak-$*$ compactness of the space of probability measures which follows from the compactness of the product space.) In this way one shows that there exists a concrete model for the process. Moreover, this particular choice of Ω is canonical in that it is the *smallest* such space, in the following precise sense: for any other such space, the random variables will factor through Ω.

There is a technical point: note that by the compactness assumption on values, Ω is, in particular, a Polish space (a complete separable metric space). Therefore it is a *Lebesgue space* in the sense of Rochlin, i.e. it is measure-isomorphic to the unit interval, plus (possibly) countably many atoms (points). This becomes important for the ergodic theory interpretation; see the discussion in [**Fi1**].

We remark that of course one reason why dynamical systems has developed into such a rich and interesting area is that so many different fields can be brought to bear: geometry, topology, ergodic theory, probability theory and information theory. Each of these contributes not only its own collection of tools but also its unique developed intuition together with a natural set of questions to ask.

The main objects from the point of view of ergodic theory are of course measure-preserving transformations of a Lebesgue space.

Therefore the basic picture (and hence one's basic intuition) is geometrical. So even if the map is not given to us in a geometrical form, if, say, it does not originate from some algebraic or geometric model, we visualize it as a transformation moving points around in a measure space. Concepts such as generic points, mixing, and entropy follow naturally from this point of view.

So if we can place this probabalistic object (a stochastic process) into the ergodic theory setting, we will have a new set of questions to ask and tools to work with. We recall how this translation is made.

Suppose we are given a stochastic process. For definiteness we take Ω to be path space as described above. Then if the process is invariant in some way, we interpret this invariance as a transformation—and then we have the ergodic theory setup.

We will now give the most basic and important example of an invariance (for other examples, see [**Fi1, 2**], [**BeFi3**]). A stochastic process is, by definition, *stationary* if a time-change does not alter the probability of an event. Equivalently, as is easy to see, the space of paths, acted on by the shift transformation, is a measure-preserving transformation of a probability space.

Now if we are given a measure-preserving transformation, we can conversely move to

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

the probability setting. Indeed, a transformation $T : X \to X$ determines many stationary stochastic processes. Take the underlying probability space Ω to be just X and choose any measurable function f, and evaluate it along orbits; this defines a stationary process $f_i(\omega) \equiv f(T^i(\omega))$.

The example we are interested in here is $R(\sigma^n \underline{w})$, which as we will see is a (Δ-valued) stationary stochastic process. We will call this the *scaling process*. In the next section we will encounter set- and measure-valued versions of this.

We recall the definition of a *generic point* x for an ergodic measure-preserving transformation T on a compact metric space X with probability measure m. For each continuous $f : X \to \mathbb{R}$, which can be thought of as an *observable*, x satisfies

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} f(T^k x) = \int_X f \, dm.$$

That is, x samples each continuous observable well with respect to time averages. If X is a Polish space, complete metric but not necessarily compact—this will occur in the next section—then we instead sample the continuous functions with compact support. By the remarks in the previous paragraphs, this definition also makes sense for a stationary ergodic stochastic process if the path space has been given the topology of a Polish space. In the definition of a generic point we only take time averages toward $+\infty$, so in the case of a two-sided stochastic process, it will also be natural to allow a one-sided sequence as a generic point.

If the measure lives in a compact part of the space (which will always be the case in this paper) then by an ergodic theorem of Kryloff and Boglioubоff (i.e. by the Birkhoff ergodic theorem plus compactness) m-almost every x is a generic point.

Now μ is equivalent to the Gibbs measure ν, which is invariant and has a unique invariant natural extension $\hat{\nu}$ on (Σ, σ). Hence by Kryloff and Boglioubоff, applying the previous corollary, we have the following.

COROLLARY 4.2. *For μ-a.e. $x \in C$, the (one-sided) sequence $R_{1,x}, R_{2,x} \ldots$ is a generic point for the ergodic Δ-valued process $R(\sigma^n(\underline{w}))$ for $n \in \mathbb{Z}$, given by $\underline{w} \in \Sigma$ being distributed like $\hat{\nu}$.*

5. *Conjugacies and the scenery process*

In this section we will construct a set-valued version of the scaling function, and use it to prove analogues of Corollaries 4.1 and 4.2 which will describe how the sequence of sets $C_{n,x}$ approximates the scenery process. We will make use of three different metrics on collections of Cantor sets. One metric, which is derived from the C^1 norm on the space of *conjugacies* of Cantor sets, is well suited to proofs and is natural from an abstract point of view. There we will prove properties (convergence at an exponential rate and Hölder dependence) which will then pass over to two geometrically defined metrics: the Hausdorff metric, and a metric derived from the Hausdorff measures.

Three metrics. Our metrics will be defined on several different spaces. Fix a hyperbolic $\mathcal{C}^{1+\gamma}$ Cantor set with map (C, S). We write $\mathcal{E}^{1+\gamma} \equiv \mathcal{E}^{1+\gamma}(C)$ for the collection of Cantor

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

sets (with maps) which are $\mathcal{C}^{1+\gamma}$ conjugate to (C, S) (from Lemma 1.2 these are also hyperbolic $\mathcal{C}^{1+\gamma}$ Cantor sets). We write $\mathcal{E}_*^{1+\gamma}$ for the quotient space of $\mathcal{E}^{1+\gamma}$, where (C, S) and $(C, \widetilde{S})$ are identified if $S = \widetilde{S}$ on C. This is the collection of marked Cantor sets conjugate to C or, equivalently, the pairs $(C, S|_C)$ with restricted maps. $\mathcal{E}_{**}^{1+\gamma}$ will denote the collection of underlying Cantor sets. We write $\mathrm{Diff}^{1+\gamma}$ for the $\mathcal{C}^{1+\gamma}$ order-preserving diffeomorphisms of I. Given a choice of the pair (C, S), $\mathrm{Diff}^{1+\gamma}$ projects onto $\mathcal{E}^{1+\gamma}(C)$ in a natural way: f is mapped to $(C_f, S_f) \equiv (f(C), f \circ S \circ f^{-1})$. This is many-to-one because there is some freedom given by the gaps; see Proposition 8.3. We note that the projection from $\mathcal{E}^{1+\gamma}$ to $\mathcal{E}_*^{1+\gamma}$ is also many-to-one.

The scenery process can be thought of as taking values in $\mathcal{E}_*^{1+\gamma}$, the marked Cantor sets, or in the space of underlying sets $\mathcal{E}_{**}^{1+\gamma}$. We will first prove convergence in the space of conjugacies, $\mathrm{Diff}^{1+\gamma}$; this will then imply convergence in the other spaces.

First we consider two metrics on $\mathcal{E}_{**}^{1+\gamma}$. We recall the definition of the *Hausdorff metric* on the collection of closed subsets of the interval I:

$$d_H(A, B) = \inf\{\epsilon : A + (-\epsilon, \epsilon) \supseteq B \text{ and } B + (-\epsilon, \epsilon) \supseteq A\}.$$

This defines a metric on $\mathcal{E}_{**}^{1+\gamma}$, and a pseudo-metric on the other spaces defined above.

Next, we define the following metric on the set of finite Borel measures on [0, 1], denoted $\mathcal{M}$. We enumerate in some fixed way, $E_1, E_2, \ldots, E_n \ldots$, the countable collection of binary intervals. For example, taking $k = 1, 2, \ldots$ and $j = 0, \ldots, 2^k - 1$, we set $n = 2^1 + \cdots + 2^{(k-1)} + j$ and define $E_n = [j2^{-k}, (j+1)2^{-k}]$. Then for ν_1, ν_2 in $\mathcal{M}$, set

$$d(\nu_1, \nu_2) = \sum_{n=1}^{\infty} |\nu_1(E_n) - \nu_2(E_n)|/2^{-n}.$$

This metric induces a topology equivalent to the weak topology on $\mathcal{M}$ in the language of probability theory; in analysis terminology this is the weak-$*$ topology on $\mathcal{M}$, the dual of the space of continuous functions.

On $\mathcal{E}_{**}^{1+\gamma}$ we define the *measure metric* d_M from this, setting

$$d_M(C, D) = d(H^d|_C, H^d|_D),$$

where H^d is d-dimensional Hausdorff measure (see [**Fa**] for definitions and background on Hausdorff dimension and measure). On the other spaces, this again defines a pseudo-metric.

Next, recall that the $\mathcal{C}^1$-norm of $f : I \to \mathbb{R}$ is

$$\|f\|_{\mathcal{C}1} = \|f\|_\infty + \|Df\|_\infty.$$

We identify $\mathrm{Diff}^{1+\gamma}$ with the collection of triples (C_f, S_f, f) for $f \in \mathrm{Diff}^{1+\gamma}$, to be written as $\widehat{\mathcal{E}}^{1+\gamma}$ (as we noted above, the map from $\widehat{\mathcal{E}}^{1+\gamma}$ to $\mathcal{E}^{1+\gamma}$ is not one-to-one). The $\mathcal{C}^1$-norm on $\mathrm{Diff}^{1+\gamma}$ determines a metric on $\widehat{\mathcal{E}}^{1+\gamma}$ as follows. For f, g in $\mathrm{Diff}^{1+\gamma}$, we write

$$d_C(C_f, C_g) = \|f - g\|_{\mathcal{C}1}.$$

We call this the $\mathcal{C}^1$ *metric* on $\widehat{\mathcal{E}}^{1+\gamma}$.

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

Note. The metric d_C keeps track of the map S on all of its domain $I_0 \cup I_1$, while d_H and d_M only see the Cantor sets. Thus, for the $\mathcal{C}^1$ metric, writing $\mathcal{I}$ for the identity map, $(C, S, \mathcal{I})$ and (C, S_f, f) will be a positive distance apart unless, in particular, $S = S_f$ on all of $I_0 \cup I_1$ (of course one also needs $f = \mathcal{I}$).

The definition of d_C depends on the initial choice of the set (C, S) (with the identity map). The next proposition shows how the metric varies if we change this 'base point' of $\widehat{\mathcal{E}}^{1+\gamma}$. As we will see in §8, $\mathrm{Diff}^{1+\gamma}$ is a Lie group and this is also a statement about bounded invariance of a metric on that group.

PROPOSITION 5.1. *Let $D \in \mathcal{E}^{1+\gamma}$, with $D = \Phi(C)$ and with $\Phi \in \mathrm{Diff}^{1+\gamma}$. We have*

$$\frac{1}{K} d_D < d_C < K d_D,$$

where $K = 2 \max\{\|D\Phi\|_\infty, \|D(\Phi^{-1})\|_\infty^{-1}\}$.

Proof. We note that from the definition of the $\mathcal{C}^1$-norm one has that if $f : I \to I$ with $f(0) = 0$, then $\|f\|_{\mathcal{C}1} \leq 2\|Df\|_\infty$. Therefore if also $g(0) = 0$, then $\|g \circ f\|_{\mathcal{C}1} \leq 2\|g\|_{\mathcal{C}1}\|f\|_{\mathcal{C}1}$.

Now for f, g and Φ as in the statement of the Proposition, we have

$$\begin{aligned} d_D(C_f, C_g) &= \|f \circ \Phi^{-1} - g \circ \Phi^{-1}\|_{\mathcal{C}1} = \|(f - g) \circ \Phi^{-1}\|_{\mathcal{C}1} \\ &\leq 2\|f - g\|_{\mathcal{C}1}\|\Phi^{-1}\|_{\mathcal{C}1} = d_C(C_f, C_g)\|\Phi^{-1}\|_{\mathcal{C}1} \end{aligned}$$

which gives one of the inequalities. The other is proved in the same way. □

Next we will look at how (on $\widehat{\mathcal{E}}^{1+\gamma}$) the pseudo-metrics d_H and d_M compare to the metric d_C. First we recall how the Hausdorff measure transforms under mappings.

Definitions. Given a one-to-one differentiable map $\Psi : M \longrightarrow N$ between open subsets of $\mathbb{R}$, and given a Borel measure μ on M and a real number $d > 0$, we write $(\Psi^\star \mu)(E) = \mu(\Psi^{-1}E)$ and

$$(\breve{\Psi}\mu)(E) = \int_{\Psi^{-1}E} |D\Psi|^d \, d\mu.$$

Thus, $\Psi^\star\mu$ is the usual *push forward* of μ, and $\breve{\Psi}\mu$ is the (Ψ, d)-*conformal transform* of μ.

Hausdorff measure has the *conformal transformation property* with respect to $\mathcal{C}^1$ maps: for $\Psi : \mathbb{R} \to \mathbb{R}$ a $\mathcal{C}^1$ diffeomorphism,

$$H^d = \breve{\Psi}(H^d).$$

PROPOSITION 5.2. *With d_H, d_M and d_C denoting the Hausdorff, measure, and $\mathcal{C}^1$ metrics respectively, for all C_f, $C_g \in \widehat{\mathcal{E}}^{1+\gamma}$,*

$$d_H(C_f, C_g) \leq d_C(C_f, C_g),$$

and for $\Psi(x) = 5x + 4x^2$, we have for all $C_f \in \mathcal{E}^r(C)$,

$$d_M(C, C_f) \leq \Psi(d_C(C, C_f)).$$

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

Proof. For the Hausdorff (pseudo)-metric this is immediate using the L^∞ norm, since

$$d_H(C_f, C_g) \le \|f - g\|_\infty \le d_C(C_f, C_g).$$

For the second inequality, writing $\mathcal{I}$ for the identity map on I, we have

$$d_C(C, C_f) \equiv \|f - \mathcal{I}\|_{\mathcal{C}1},$$

and writing $\mu = H^d|_C$, $\mu_f = H^d|_{C_f}$,

$$\begin{aligned} d_M(C, C_f) &\equiv \sum_n |\mu E_n - \mu_f E_n| 2^{-n} \\ &= \sum_n \left| \int_{f^{-1}E_n} |Df|^d \, d\mu - \int_{E_n} 1 \, d\mu \right| 2^{-n} \\ &\le 2\|f - \mathcal{I}\|_\infty \|Df\|_\infty^d \sum_n 2^{-n} + \sum_n 2^{-n} \int_{f^{-1}E_n \cap E_n} ||Df|^d - 1| \, d\mu \end{aligned}$$

(here the first term bounds the contribution, for each interval E_n, of its two ends not matching up exactly with $f^{-1}E_n$; we used the fact that since f is a diffeomorphism of I, $\|Df\|_\infty \ge 1$).

Next, we note that for all $x > 0$, $|x^d - 1| \le |x - 1|$. Hence $||Df|^d - 1| \le ||Df| - 1|$, so the above is

$$\begin{aligned} &\le 2\|f - \mathcal{I}\|_\infty (1 + \||Df| - 1\|_\infty) + \||Df| - 1\|_\infty \sum_n |E_n| 2^{-n} \\ &\le 4\||Df| - 1\|_\infty (1 + \||Df| - 1\|_\infty) + \||Df| - 1\|_\infty \\ &\le 4(1 + \||Df| - 1\|_\infty + 1)(\||Df| - 1\|_\infty) = 5\||Df| - 1\|_\infty + 4(\||Df| - 1\|_\infty)^2 \\ &\le \Psi(d_C(C, C_f)), \end{aligned}$$

as claimed. □

We define the (*restricted*) $\mathcal{C}^1$ *metric* on $\mathcal{E}_*^{1+\gamma}$ to be

$$d_{C*}(C_f, C_g) = \|f - g\|_{\mathcal{C}^1, C}$$

where this indicates that the sup norms are taken over the Cantor set C. Obviously, $d_{C*} \le d_C$ on $\widehat{\mathcal{E}}^{1+\gamma}$.

The first theorem that we are aiming for is the following.

THEOREM 5.3. *Given a $\mathcal{C}^{1+\gamma}$ hyperbolic Cantor set (C, S), for every $y = (\dots y_{-n} \dots y_{-1}) \in \Sigma^-$, the limit*

$$C(y) \equiv \lim_{n \to \infty} C_{y_{-n} \dots y_{-1}}$$

exists. Convergence is exponentially fast, and $C(y)$ is Hölder continuous with exponent γ (for both metrics on $\mathcal{E}_{**}^{1+\gamma}$, and for the restricted $\mathcal{C}^1$ metric on the marked Cantor sets $\mathcal{E}_*^{1+\gamma}$). Moreover, $C(y) = C^y$, the ratio Cantor set built from the scaling function of (C, S).

Thus, $C(y)$ is a set-valued version of the scaling function $R(y)$ (compare Theorem 3.1). To prove Theorem 5.3, we first construct, in Theorem 5.9, certain

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

conjugacies, proving convergence in the $\mathcal{C}^1$ metric. Convergence in the Hausdorff and measure metrics then follows from Proposition 5.2. Before giving this construction, we state several further consequences of Theorem 5.9.

In the same way as for the scaling function $R(y)$, we also define $C(\cdot)$ on the full shift space Σ, setting $C(\underline{x}) = C(y)$ for $\underline{x} = (y, x)$.

THEOREM 5.4. *For every $x \in C$, the sequence $C_{n,x} \equiv C_{x_0 \ldots x_n}$ is asymptotic to $C(\sigma^n(\underline{x}))$, with an exponential rate of convergence, for any $\underline{x}$ in Σ with the same future coordinates $(x_0, x_1, \ldots)$.*

This is like Corollary 4.1. We will write $\mathcal{L}_C$ for the (compact) subset of $\mathcal{E}_*^{1+\gamma}$ (with respect to all three metrics) which is the range of the function $\underline{x} \mapsto C(\underline{x})$. This is the collection of (*marked*) *limit sets*. Since the function is continuous and the domain Σ is compact, we then have the following.

PROPOSITION 5.5. *The collection of limit sets is compact.*

Definition. Given a $\mathcal{C}^{1+\gamma}$ hyperbolic Cantor set C, the *set-valued scenery process* of C is the process $C(\sigma^n(\underline{x})) = C^{\sigma^n(\underline{x})}$, with $\underline{x} \in \Sigma$, distributed according to the measure $\hat{\nu}$.

Note that stationarity and ergodicity of this process follow immediately from invariance and ergodicity of the measure $\hat{\nu}$.

The space of paths is a compact subset of the Polish space $\Pi_{-\infty}^{+\infty}\mathcal{E}_{**}^1$ (or $\Pi_{-\infty}^{+\infty}\mathcal{E}_*^1$ for the marked sets) with the product topology determined by the topologies of any of the three metrics; see the proof of Corollary 5.6.

We mention why we use probability terminology—the *scenery process* —for the map $C(\underline{x}) \mapsto C(\sigma\underline{x})$. Note that this dynamics is not in fact given by a map on $\mathcal{L}_C$ itself. Indeed, at every stage you have two choices: the right- or left-hand subsets from the next level of the ratio Cantor set $C(\underline{x})$, with the choice of left or right depending on whether x_0 is 0 or 1. Or, from a different viewpoint, one has the dynamics on $\mathcal{L}_C$ of a semigroup action; see the note at the end of the paper.

The next result is like Corollary 4.2.

COROLLARY 5.6. *For μ-a.e. $x \in C$, the sequence of (rescaled) Cantor sets $C_{n,x}$ which nest to x is a generic point for the stationary ergodic set-valued process $C^{\sigma^n(\underline{x})}$ determined by $(\Sigma, \sigma, \hat{\nu})$.*

In §2 we defined $\mu_{n,x}$ to be the sequence of Hausdorff measures $H^d|_{C_{n,x}}$. Convergence of $C_{n,x}$ to $C(y)$ in the measure metric can be rephrased as follows.

COROLLARY 5.7. *For every $y = (\ldots y_{-n} \ldots y_{-1}) \in \Sigma^-$, the limit*

$$M(y) \equiv \lim_{n \longrightarrow \infty} \mu_{y_{-n} \ldots y_{-1}}$$

exists, and is Hölder continuous with exponent γ.

The support of the measure $M(y)$ is the set $C(y)$, and $M(y)$ is a measure-valued version of the scaling function.

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

As we did for the set-valued process, we define $M(\underline{x}) = M(y)$. We define the *measure-valued scenery process* $M(n, \underline{x}) \equiv M(\sigma^n(\underline{x}))$, again with $\underline{x} \in \Sigma$, distributed like $\hat{\nu}$. We have the following.

COROLLARY 5.8. *For μ-a.e. $x \in C$, the sequence $\mu_{x_0 \ldots x_n}$ is a generic point for the stationary ergodic measure-valued process $M(\sigma^n(\underline{x}))$ determined by $(\Sigma, \sigma, \hat{\nu})$.*

Now we proceed to the proofs. For each interval $I_{w_0 \ldots w_n}$ with $w_i = 0$ or 1, we write $A_{w_0 \ldots w_n}$ for the affine map which expands the interval to the unit interval I. We then set $A_n^y = A_{y_{-n} \ldots y_{-1}}$ for $y \in \widetilde{C}$. This expands the interval $I_{y_{-n} \ldots y_{-1}}$ to I affinely. Next, define maps $\varphi_{k,n}^y : I \to I$ for $k \geq n$ by

$$\varphi_{k,k}^y = \text{identity} \quad \text{and} \quad \varphi_{k,n}^y = \varphi_{y_k} \circ \cdots \circ \varphi_{y-(n+1)}, \quad \text{for } k \geq n.$$

We will also write φ_k^y for $\varphi_{k,0}^y$.

For $0 \leq n \leq k$, we define $\Phi_{k,n}^y : I \to I$ by

$$\Phi_{k,n}^y = A_k^y \circ \varphi_{k,n}^y \circ (A_n^y)^{-1}.$$

It follows that:

(1) for each n, $\Phi_{n,n}^y = \text{identity}$;
(2) for all $m \geq k \geq n$, $\Phi_{m,n}^y = \Phi_{m,k}^y \circ \Phi_{k,n}^y$; and
(3) $\Phi_{n,0}^y = A_n^y \circ \varphi_n^y$.

We will also write Φ_n^y for $\Phi_{n,0}^y$.

The sequence Φ_n^y is, to use Sullivan's words, 'a sequence of ... compositions (of contractions) ... renormalized by post composition with linear maps to obtain mappings between unit intervals...' (Appendix of [**Su1**]). As Sullivan states, and as we will prove in the next section, such a sequence is precompact in $\mathcal{C}(k, \alpha)$ if the original hyperbolic Cantor set is $\mathcal{C}(k, \alpha)$.

This gives convergence along some subsequence $\Phi_{n_k}^y$. Unfortunately, however, this is not enough for our goal of proving an ergodic theorem for Cantor sets and measures (i.e. Corollaries 5.6 and 5.8)—for that purpose we want to prove that the sequence itself converges. We do this in the next theorem, using bounded distortion, and then in the next section we return to Sullivan's idea to prove smoothness of the resulting limiting conjugacy.

THEOREM 5.9. *Let (C, S) be a $C^{1+\gamma}$-hyperbolic Cantor set. For each y in the dual Cantor set $\widetilde{C}$,*

$$\Phi^y \equiv \lim_{n \to \infty} \Phi_n^y$$

exists. This is an order-preserving diffeomorphism from I to I. Convergence is of order $O(\beta^{n\gamma})$ in the C^1 norm, uniformly in y, and the function $y \mapsto \Phi^y$ is Hölder continuous of order γ, in the β-metric.

Proof. We will show, using bounded distortion, that for n large and for $k > n$ arbitrary, $\Phi_{k,n}^y$ is close to the identity. Then since $\Phi_{k,0}^y = \Phi_{k,n}^y \circ \Phi_{n,0}^y$, this will imply convergence.

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

Now, since the maps A are affine they have constant derivative. So for each $a \in I$, for $m = k - n$, we have

$$\begin{aligned}(D(\Phi^y_{k,n}))(a) &\equiv D(A^y_k \circ (S^m)^{-1} \circ (A^y_n)^{-1})(a) = \frac{DA^y_k}{DA^y_n DS^m(z)} \quad \text{for } z \equiv (A^y_n)^{-1}(a) \\ &= \frac{DS^m(z_0)}{DS^m(z)} \quad \text{for some } z_0 \in I_{y_{-k}\dots y_{-1}},\end{aligned}$$

by the mean value theorem. Therefore, by Corollary 2.2

$$e^{-K\beta^{n\gamma}} < D\Phi^y_{k,n}(a) < e^{K\beta^{n\gamma}}$$

for all $k > n$, $y \in \widetilde{C}$ and all $a \in I$. This implies the sequence $D\Phi^y_n$, $n = 0, 1, 2 \dots$, is Cauchy, hence converges. Since $\Phi^y_n(0) = 0$ for all n, it follows from the fundamental theorem of calculus that the limit $\Phi^y \equiv \lim_{n\to\infty} \Phi^y_n$ exists, and that $D\Phi^y = \lim D\Phi^y_n$. By Corollary 2.2, $D\Phi^y$ is bounded away from 0 and ∞ by $e^{\pm K\beta^\gamma}$; in particular, Φ^y is an order-preserving diffeomorphism from I to I, as claimed.

We define for each n,

$$\Phi^y_{\infty,n} = \lim_{k\to\infty} \Phi^y_{k,n};$$

the limit exists by the above arguments, and this map is $\mathcal{C}^1$ close to the identity map $\mathcal{I} : I \to I$. We have for each n that

$$\Phi^y = \Phi^y_{\infty,n} \circ \Phi^y_{n,0},$$

and that:
(1) $\exists k_0 > 0$ such that

$$\|\Phi^y_{\infty,n} - \mathcal{I}\|_{\mathcal{C}1} < k_0 \beta^{n\gamma}.$$

The constant k_0 here only depends on K from Corollary 2.2, which in turn depends on β, the upper bound for $|D\varphi_i|$. Here is the calculation. We have

$$\|\Phi^y_{\infty,n} - \mathcal{I}\|_{\mathcal{C}1} \le \|D\Phi^y_{\infty,n} - 1\|_\infty$$

and we know that

$$e^{-K} \le e^{-K\beta^{n\gamma}} \le D\Phi^y_{\infty,n} \le e^{K\beta^{n\gamma}} \le e^K$$

for all n. Now, since for x in the interval $[e^{-K}, e^K]$ one has $|e^x - 1| < k_0 x + 1$, where we take $k_0 = (\exp(\exp(K)) - 1)/(\exp(K))$, statement (1) holds true.

Now recall from the proof of Proposition 5.1 that for $f,\ g : I \to I$ with $f(0) = 0$ and $g(0) = 0$, then $\|g \circ f\|_{\mathcal{C}1} \le 2\|g\|_{\mathcal{C}1}\|f\|_{\mathcal{C}1}$. From this, it follows that statement (1) is equivalent to:
(2) $\exists k_1 > 0$ such that

$$\|\Phi^y_n - \Phi^y\|_{\mathcal{C}1} < k_1 \beta^{n\gamma}$$

(here we can take $k_1 = (e^K)k_0$).

Both statements express, in different ways, that Φ^y is close to $\Phi^y_{n,0}$, with exponentially fast convergence; (2) is what we stated in the theorem. Finally, it is now also easy to check Hölder continuity:

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

(3) $\exists k_2 > 0$ such that for all $y, w \in \widetilde{C}$,

$$\|\Phi^y - \Phi^w\|_{\mathcal{C}1} \leq k_2 (d_\beta(y, w))^\gamma. \qquad \square$$

Proof of Theorem 5.3. Writing C_n^y for $\Phi_n^y(C) = C_{y_{-n} \ldots y_{-1}}$, and $C^y \equiv \Phi^y(C)$, statement (2) above says exactly that

$$d_C(C_n^y, C^y) < k_1 \beta^{n\gamma}.$$

Hence we have convergence in the $\mathcal{C}^1$ metric. By Proposition 5.2, therefore, d_H has the same bound. For the measure metric, we have

$$\begin{aligned} d_M(C_n^y, C^y) &\leq \Psi(d_{C^y}(C^y, C_n^y)) \\ &= \Psi(\|\Phi_{\infty,n}^y - \mathcal{I}\|_{\mathcal{C}1}) \\ &\leq \Psi(k_0 \beta^{n\gamma}) \end{aligned}$$

by (1), and this is $\leq k_3 \beta^{n\gamma}$, where $k_3 = k_0(5 + 4k_0)$. Next we show Hölder continuity. Now $d_C(C^y, C^w) = \|\Phi^y - \Phi^w\|_{\mathcal{C}1}$, so (3) proves Hölder continuity for the $\mathcal{C}^1$ and Hausdorff metrics. Then, applying Proposition 5.1, $d_M(C^y, C^w) \leq \Psi(d_{C^y}(C^y, C^w)) \leq \Psi(d_C(C^y, C^w)\|\Phi^{-1}\|_{\mathcal{C}1}) \leq \Psi(e^K k_2 \beta^{n\gamma}) \leq k_4 \beta^{n\gamma}$, where $k_4 = 5a + 4a^2$ and $a = e^K k_2$. Finally, it is clear from the constructions that $C(y)$ has ratio geometry given by $R(y)$, hence $C(y)$ is indeed equal to C^y. $\square$

Proof of Theorem 5.4. From the proof of Theorem 5.3, since the exponential bound is uniform over all sets of level n, we have that the d_C-distance from $C(\sigma^n x)$ to $C_{x_0 \ldots x_n}$ is bounded by $k_1 \beta^{n\gamma}$. The bounds for d_H and d_M then follow as above. $\square$

Proof of Corollary 5.6. We give the proof for the process which takes values in the collection of marked sets. Here the space will be $\Pi_{-\infty}^{\infty} \mathcal{E}_*^1(C)$, with the shift transformation; this is a Polish space (since $\mathcal{E}_*^1$ is a Polish space; we use $\mathcal{E}_*^1$ rather than $\mathcal{E}_*^{1+\gamma}$ so as to have a complete space). Therefore we know from our definitions what it will mean for the one-sided sequence $C_{n,x}$ to be a generic point. Now the map $\underline{x} \mapsto (\ldots C(\sigma^{-1}\underline{x}), C(\underline{x}), C(\sigma \underline{x}) \ldots)$ from Σ has as its image a compact invariant subset of $\Pi_{-\infty}^{\infty} \mathcal{L}_C \subseteq \Pi_{-\infty}^{\infty} \mathcal{E}_*^1(C)$; this image is the space of paths of the scenery process, and is the support of the image of the measure $\hat{\nu}$. The ergodicity of $(\Sigma, \hat{\nu}, \sigma)$ passes over to the scenery process, hence a.e. the path $(\ldots C(\sigma^{-1}\underline{x}), C(\underline{x}), C(\sigma \underline{x}) \ldots)$ is generic for the shift on path space. Finally, since by Theorem 5.4 we know the sequence $C_{n,x}$ is forward asymptotic to $C(\sigma^n \underline{x})$, we will compute the same time average for the continuous functions. Thus, $C_{n,x}$ is generic for ν-a.e. x and hence for μ-a.e. x. $\square$

6. *Smoothness of conjugacies*

Now we will see how to prove that the conjugacies of §5 in fact have higher smoothness properties.

The basic idea will be to imitate what one knows about analytic maps for $\mathcal{C}^{k+\gamma}$ or $\mathcal{C}(k, \gamma)$ maps. Thus Lemma 6.2 is a version of Leibnitz' formula, and Lemma 6.3 is one step in showing $\mathcal{C}^{k+\gamma}$ maps are morphisms in a category. This means they can be used to define equivalence relations on sets, and to give the analogue of differentiable structures. One also imitates the Arzela–Ascoli theorem in Lemma 6.4; as Sullivan says

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

in the Appendix of [**Su1**], and quoted in the previous section, the sequence Φ_n^y will be precompact in $\mathcal{C}(k,\alpha)$. (As usual, instead we do our proofs in $\mathcal{C}^{k+\gamma}$.)

Here is the main theorem we are aiming for.

THEOREM 6.1. *Let C be a $\mathcal{C}^{k+\gamma}$ hyperbolic Cantor set, and let $\Phi^y : I \to I$ be defined as in §5. We claim:*

(i) *if $k = 1, 2, \ldots$, then Φ^y is $\mathcal{C}^{k+\gamma}$ (with the same Hölder exponent, but a different Hölder constant);*

(ii) *if $k = \infty$ or ω, then Φ^y is $\mathcal{C}^\infty$ or $\mathcal{C}^\omega$ respectively.*

First we need a few lemmas.

LEMMA 6.2. *For $A \subseteq \mathbb{R}$, if $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$ are bounded, γ-Hölder continuous with Hölder constants c, d, then:*

(i) *$f + g$ is γ-Hölder with constant $c + d$; and*

(ii) *$f \cdot g$ is γ-Hölder with constant $c\|g\|_\infty + d\|f\|_\infty$.*

Proof. (i) is immediate. The argument for (ii) comes by imitating the proof of Leibnitz' rule in the calculus:

$$\begin{aligned} |f(x)g(x) - f(y)g(y)| &= |[f(x) - f(y)]g(x) + f(y)[g(x) - g(y)]| \\ &\leq c|x-y|^\gamma \|g\|_\infty + d|x-y|^\gamma \|f\|_\infty. \end{aligned}$$

□

LEMMA 6.3. *Fix $k \geq 1$. For $A, B \subseteq \mathbb{R}$, let $f : A \to B$ and $g : B \to \mathbb{R}$ be such that $D^k(f)$, $D^k(g)$ are bounded and γ-Hölder. Then $D^k(g \circ f)$ is γ-Hölder.*

Proof. This now follows by induction from the chain rule plus Lemma 6.2. □

The next lemma is basically the same as the 'bounded variation' lemma, Lemma 1.15 from [**Bo1**], except it is written in the reverse direction with the contractions φ_i instead of the inverse map S. For the special case $f_i = \varphi_{w_j}$ and $h_i = \log|D\varphi_{w_j}|$, where $i + j = n$, one gets exactly the bounded distortion property (Theorem 2.1). The formulation given here is from the Appendix of [**Su1**]; the key idea for proving our Theorem 6.1, which is also in that Appendix (the sentence immediately preceding the corollary there) will be how to use this lemma to control higher-order derivatives of the composition. One can summarize the idea as follows: do not look at $\log|D^k\varphi_{x_i}|$, but instead at $D^{k-1}\log|D\varphi_{x_i}|$. Then we are applying the linear operator D^k to a sum, which leads to the proof.

As usual, for notational simplicity, we assume strict hyperbolicity.

LEMMA 6.4. *Consider a composition of contractions $f_n \circ \cdots \circ f_1$, with $f_i : J_i \to J_{i+1}$ for intervals $J_i \subseteq \mathbb{R}$, and such that $|Df_i| < \beta < 1$. For a point $x \in J$, write $x_1 = x$, $x_{i+1} = f_i(x_i)$. Let $h_i : J_i \to \mathbb{R}$ be Hölder continuous with the same exponent $\gamma \in (0, 1]$ and the same constant c. Then for $h(x) \equiv h_1(x_1) + \ldots + h_n(x_n)$, h is also γ-Hölder continuous, with constant $c_0 = c\beta^\gamma/(1 - \beta^\gamma)$ (independent of n).*

Proof. Immediate from the geometric series, since for $x, y \in J_1$ we have $|x_i - y_i| < \beta^i$. □

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

The next lemma is more subtle than one might at first think. We wish to thank Z. Nitecki and M. Urbanski for discussions which resulted in a first proof, and Y. Kifer for then finding the much simpler argument given here.

LEMMA 6.5. *Let $f_n : I \to I$ be continuous functions with continuous kth derivative and assume that there exist functions f, g such that:*
(i) $f_n \to f$; *and*
(ii) $D^k f_n \to g$, *uniformly as* $n \to \infty$.
Then $D^k f = g$.

Proof. We define, for each $0 \le j \le k$, functions g_j, and sequences of functions $f_{n,j}$ and $p_{n,j}$ by

$$g_k = g \quad \text{and} \quad g_{j-1}(t) = \int_0^t g_j,$$
$$f_{n,k} = D^k f_n \quad \text{and} \quad f_{n,j-1}(t) = \int_0^t f_{n,j},$$
$$p_{n,j} = D^j f_n - f_{n,j}.$$

Thus

$$p_{n,k} \equiv 0,$$
$$p_{n,k-1}(t) j = (D^{k-1} f_n)(t) - ((D^{k-1} f_n)(t) - (D^{k-1} f_n)(0)) \equiv (D^{k-1} f_n)(0),$$
$$p_{n,k-2}(t) = D^{k-2} f_n(0) + t(D^{k-1} f_n)(0),$$

and similarly (for each n) $p_{n,j}$ is for all j a polynomial of degree $k - j - 1$, such that $Dp_{n,j} = p_{n,j+1}$. Now for each j, $\lim_{n\to\infty} f_{n,j} = g_j$. In particular,

$$g_0 = \lim_{n\to\infty} f_{n,0} = \lim_{n\to\infty} (f_n - p_{n,0}) = f - \lim p_{n,0}.$$

Hence, $\lim_{n\to\infty} p_{n,0}$ converges (uniformly) to some polynomial p_0 with degree at most $k - 1$, and we have $g_0 = f - p_0$. Therefore,

$$g = D^k g_0 = D^k f - D^k p_0 = D^k f$$

as claimed. □

Proof of Theorem 6.1. Since $D^k \varphi_0$ and $D^k \varphi_1$ are γ-Hölder, by Lemma 6.3 so is $D^{k-1} \log D\varphi_i$, with some Hölder constant c_0. Now we apply Lemma 6.4 to

$$f_n \circ \cdots \circ f_1 = \varphi_{y_n} \circ \cdots \circ \varphi_{y_1} \equiv \varphi_n^y$$

and

$$h_j \equiv D^{(k-1)} \log D\varphi_{y_j}.$$

For $x \in I$, writing $x_1 = x$, $x_2 = h_1(x)$, etc as in Lemma 6.4, since

$$D^{(k-1)} \log D\varphi_n^y = \sum_{j=1}^{n} h_j(x_j),$$

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

we conclude that $D^{(k-1)} \log D\varphi_n^y$ is γ-Hölder, with some different constant c_1 which is, however, independent of n.

Now to prove the theorem, first consider the case $k \geq 2$. Here we have $D^{(k-1)} \log D\Phi_n^y = D^{(k-1)} \log D\varphi_n^y$, since the constant derivative of A_n^y disappears upon higher differentiation. For $k = 1$ these are not equal; they differ by the constant $\log DA_n^y$ (which increases with n). However, these cancel upon subtraction, so in either case we have, for any $a, b \in I$,

$$|\log D\Phi_n^y(a) - \log D\Phi_n^y(b)| = |\log D\varphi_n^y(a) - \log D\varphi_n^y(b)|.$$

Therefore for all $k \geq 1$, $D^{k-1} \log D\Phi_n^y$ is a sequence of bounded functions which is γ-Hölder with the same constant, c_1. Also, this sequence is uniformly bounded. For $k = 1$ this follows from bounded distortion, as in the proof of Theorem 5.4, and in fact a bound is $e^{K\beta^\gamma}$. For $k > 1$, we argue as follows. If it were unbounded, then by Hölder continuity with the same constant, some subsequence goes uniformly to either $+\infty$ or $-\infty$. By integration $(k-1)$ times, by induction this contradicts the boundedness for $k = 1$. This implies equicontinuity. Now by boundedness and equicontinuity, there is some convergent subsequence, using the standard diagonalization argument as in the proof of the Arzela–Ascoli theorem. At the same time, from §5 we know that $\lim D\Phi_n^y = D\Phi^y$ exists which implies $\log D\Phi_n^y$ converges to $\log D\Phi^y$. Calling the subsequence

$$\log D\Phi_{n_j}^y = f_j,$$

we are in the situation of Lemma 6.5: $f_j \to f$, $D^{k-1} f_j \to g$, hence $D^{k-1} f = g$. Thus $D^{k-1} \log D\Phi^y$ is a uniform limit of γ-Hölder functions with the same constant c_o, hence the limit is γ-Hölder with constant c_o. From Lemma 6.3, $D^k \Phi^y$ is also γ-Hölder and we are done for $k = 1, 2, \ldots$.

Finally, note that for $k = \infty$ we are done by part (i), and for $k = \omega$ we can apply Arzela–Ascoli to see that $\{\Phi_n^y\}$ is a normal family, hence the limit Φ^y is also analytic. □

Remark. We emphasize again the subtle point in the logic of this argument: $\mathcal{C}^{1+\gamma}$ convergence of Φ_n^y to Φ^y is *not* known. What we *do* know is convergence in the $\mathcal{C}^1$ norm (from Theorem 5.9) and convergence along a *subsequence* in the $\mathcal{C}^{1+\gamma}$ norm, as just shown. This is enough to prove the claim of the theorem.

7. *Smoothness of limit sets and rigidity*

Given two $\mathcal{C}^{1+\gamma}$ hyperbolic Cantor sets (C, S) and $(\widehat{C}, \widehat{S})$, recall that the (full) conjugacy Φ is an order-preserving map defined on all of I. This map is uniquely determined on C by the conjugacy equation, since, as one sees, the symbolic dynamics is preserved. Note that for any two topological Cantor sets, once they have been coded by the two-shift Σ^+ in an order-preserving way, this conjugacy on the Cantor sets extends to a homeomorphism on I. The issue, therefore, is what types of conjugacies preserve what type of structure. As is well known and not hard to show, for instance, a biLipschitz Φ will preserve the Hausdorff dimension. We noted in [**BeFi1**] that $\mathcal{C}^1$ maps preserve the order-two density. Furthermore, for $\mathcal{C}^1$ conjugacy from [**Su1**] one has the following.

LEMMA 7.1. *If two $\mathcal{C}^{1+\gamma}$ hyperbolic Cantor sets are $\mathcal{C}^1$ conjugate, then they have the same scaling function.*

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

Proof. By uniform continuity of the derivatives, since we already know the scaling functions exist from Theorem 3.1, this is immediate. □

Hence under the same assumption, by Corollary 5.3 we have the following.

COROLLARY 7.2. *They have the same collection of limit sets.*

To prove our rigidity theorem, we will need the following.

LEMMA 7.3. *Let (C, S) be a hyperbolic $\mathcal{C}^{k+\gamma}$ Cantor set. Let $\widehat{S} : I_0 \cup I_1 \to I$ be a $\mathcal{C}^{k+\gamma}$ map such that $S = \widehat{S}$ on C. Then S and $\widehat{S}$ are $\mathcal{C}^{k+\gamma}$ conjugate.*

Proof. The conjugacy is the identity map on C; what we want to do is define it on the gaps. We begin by defining Φ to be the identity also on the gap G between I_0 and I_1. The conjugacy is then uniquely defined from the conjugacy equation by the dynamics. That is, writing $G_{x_0 \ldots x_n} = \varphi_{x_0 \ldots x_n}(G)$, we have for $a \in G_{x_0 \ldots x_n}$,

$$\Phi(a) = \widehat{\varphi}_{x_0 \ldots x_n}(\varphi^{-1}_{x_0 \ldots x_n}(a)) = \widehat{\varphi}_{x_0 \ldots x_n}(S^n(a)).$$

One immediately checks that with this definition Φ is a conjugacy.

This map is $\mathcal{C}^{k+\gamma}$ on the interiors of all the gaps. At points in C, to check $\mathcal{C}^{k+\gamma}$ one must be careful because these points are also limits of interior points in the gaps.

Here is one way of proving Φ is everywhere $\mathcal{C}^{k+\gamma}$. Define a sequence of maps $\Phi_n : I \to I$, where $\Phi_0 =$ the identity, $\Phi_1 = \Phi_0$ on G and $\widehat{\varphi}_{x_0} \circ S$ everywhere else (i.e. on $I_0 \cup I_1$) and, inductively, set Φ_n to be equal to Φ_{n-1} everywhere except on $\bigcup I_{x_0 \ldots x_n}$, where it is defined to be $\widehat{\varphi}_{x_0 \ldots x_n} \circ S^n$.

These maps converge uniformly to Φ. So if we can show that for each n, $D^k \Phi_n$ is γ-Hölder with a constant independent of n, this will carry over to the limit and we will be done (here we will use the fact that the maps $\widehat{\varphi}_{x_0 \ldots x_n} \circ S^n$ are γ-Hölder with a fixed constant).

The advantage of this method is that we must only check smoothness at each stage, and so each time at only finitely many points.

Now consider the map $f = \widehat{\varphi}_{x_0 \ldots x_n} \circ S^n$ on $C \cap I_{x_0 \ldots x_n}$. It is the identity there, and since C is dense in itself, $Df = 1$ on that set. Since it is twice differentiable, $D^2 f = 0$ there and similarly for $D^k f$. Therefore when we define Φ_n by gluing together $\widehat{\varphi}_{x_0, \ldots, x_{n-1}} \circ S^{n-1}$ and $\widehat{\varphi}_{x_0 \ldots x_n} \circ S^n$ at an endpoint p, the two functions agree at p for all derivatives $\leq k$. Also, D^k is γ-Hölder for each piece. Hence, for all n, Φ_n is $\mathcal{C}^{k+\gamma}$ with a fixed Hölder constant, as we wanted to show, and so we are done. □

This produces one conjugacy. In §8 we will return to this proof in order to study *how many* such maps Φ there are.

We are now ready to prove the following.

THEOREM 7.4. (Highest smoothness) *Given a $\mathcal{C}^{1+\gamma}$ hyperbolic Cantor set C, its limit sets have the highest degree of smoothness of any hyperbolic $\mathcal{C}^{1+\gamma}$ Cantor set in the $\mathcal{C}^{1+\gamma}$-conjugacy class of C.*

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

Proof. Let $(\widehat{C}, \widehat{S})$ be a $\mathcal{C}^{k+\gamma}$, $\mathcal{C}^\infty$ or $\mathcal{C}^\omega$ hyperbolic Cantor set which is $\mathcal{C}^1$ conjugate to C. By Corollary 7.2, C and $\widehat{C}$ have the same limit sets, and by Theorem 6.1, the map $\widehat{\Phi}^y : I \to I$ defines a dynamics $S^y : I_0^y \cup I_1^y \to I$ by conjugation with the map $\widehat{S}$, which has that same degree of smoothness. □

THEOREM 7.5. (Rigidity) *If (C, S) and $(\widehat{C}, \widehat{S})$ are two $\mathcal{C}^{k+\gamma}$, $\mathcal{C}^\infty$ or $\mathcal{C}^\omega$ hyperbolic Cantor sets which either (a) are $\mathcal{C}^1$ conjugate by a map Φ, or (b) have the same scaling function R, then they are in fact conjugate by a map $\widetilde{\Phi} : I \to I$ which for (a) agrees with Φ on C or for (b) agrees with the coding; this map is $\mathcal{C}^{k+\gamma}$, $\mathcal{C}^\infty$ or $\mathcal{C}^\omega$ respectively.*

Proof. By either hypothesis they have the same limit sets. Choose one, C^y. Again by Theorem 6.1, the maps Φ^y, $\widehat{\Phi}^y$ have the same smoothness as S, $\widehat{S}$. Now let S^y, $\widehat{S}^y$ denote the maps defined on $I_0^y \cup I_1^y$ by these conjugacies. We are exactly in the situation of Lemma 7.3, and have a conjugacy Φ of S^y and $\widehat{S}^y$. Composing the three maps

$$(\widehat{\Phi}^y)^{-1} \circ \Phi \circ \Phi^y$$

finishes the proof. □

8. *Banach space structure*

Fix a hyperbolic $\mathcal{C}^{1+\gamma}$ Cantor set (C, S). For $r = k + \gamma$, where $\gamma \in (0, 1]$, $k \geq 1$ or for $r = \infty, \omega$, we write $\mathcal{E}^r \equiv \mathcal{E}^r(C)$ for the collection of Cantor sets (with maps) which are $\mathcal{C}^r$ conjugate to (C, S). (From Lemma 1.2 these are also hyperbolic $\mathcal{C}^{1+\gamma}$ Cantor sets.) The spaces Diff^r, $\widehat{\mathcal{E}}^r$, $\mathcal{E}^r_*$ and $\mathcal{E}^r_{**}$ are defined as they were in §5 for the case $r = 1 + \gamma$.

In this section we will see how $\widehat{\mathcal{E}}^r$ can be viewed as a Banach manifold, in fact a Banach Lie group. We will also define a natural topology on $\mathcal{E}^r$, and show that $\widehat{\mathcal{E}}^r$ factors nicely over $\mathcal{E}^r$ as a topological space.

We first define the $\mathcal{C}^\gamma$ norm on the Hölder functions $\mathcal{C}^\gamma(I, \mathbb{R})$ to be

$$\|f\|_{\mathcal{C}^\gamma} = \|f\|_\infty + \sup_{x,y \in I} \frac{|f(x) - f(y)|}{|x - y|^\gamma}.$$

For $r = k + \gamma$, where $\gamma \in (0, 1]$, $k \geq 1$, the $\mathcal{C}^r$ norm will be

$$\sum_{l=0}^{k-1} \|D^l f\|_\infty + \|D^k f\|_{\mathcal{C}^\gamma}.$$

For $\mathcal{C}^\infty$, we define

$$\|f\|_{\mathcal{C}^\infty} = \sup_l \{\|D^l f\|_\infty\},$$

and for $\mathcal{C}^\omega$ we will use the sup norm (since it is equivalent to all the other $\mathcal{C}^r$ norms there).

By definition a Banach manifold is a manifold which is locally modelled on a Banach space, and a Lie group is a group which is also a C^∞ manifold modelled on a complete, locally convex vector space (see e.g. [**Mi**]). Recall that Diff^r denotes the $\mathcal{C}^r$ order-preserving diffeomorphisms of I. Now the choice of a set in $\mathcal{E}^r$ identifies the collection $\widehat{\mathcal{E}}^r$ with Diff^r, as we have seen in §5. Diff^r is an open subset of $\mathcal{C}^r_{0,1}(I, \mathbb{R})$, which is how

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

we will write the set of all $\mathcal{C}^r$ functions from I to $\mathbb{R}$ such that $f(0) = 0$ and $f(1) = 1$. This in turn is a closed affine subspace of $\mathcal{C}^r(I, \mathbb{R})$. To see this, note that, defining

$$\mathcal{B}^r_{0,1}(I, \mathbb{R}) = \{f \in \mathcal{C}^r(I, \mathbb{R}) : \ f(0) = 0 = f(1)\},$$

two functions in $\mathcal{C}^r_{0,1}$ differ exactly by an element of $\mathcal{B}^r_{0,1}$. Now $\mathcal{B}^r_{0,1}$ is a Banach space with the $\mathcal{C}^r$ norm. Hence, $\widehat{\mathcal{E}}^r$ is a Banach manifold: it is identified with Diff^r, which in turn corresponds to an open subset of $\mathcal{B}^r_{0,1}$. Now Diff^r is a group, hence it (and therefore $\widehat{\mathcal{E}}^r$) is a Banach Lie group. Two choices have been made: the choice of a Cantor set in $\mathcal{E}^r$, and of a special point (the identity) in Diff^r. These choices determined the maps to $\mathcal{B}^r_{0,1}$ and hence the metric (inherited from the $\mathcal{C}^r$ norm). Both choices, moreover, amount to the same thing: changing C to D in $\mathcal{E}^r$ (as in Proposition 5.1, for $r = 1$) corresponds to a right translation in the group Diff^r.

Now in a Lie group one ideally would like to work with a (left- or right-) invariant metric. If the group is compact (or, more generally, amenable) one can make a given metric invariant (while keeping equivalence) by averaging over translations. In our case, however, one cannot get an equivalent invariant metric—Diff^r is not only non-compact but non-amenable! The (non-uniform) bounded equivalence proved in Proposition 5.1 is, nevertheless, enough for what we needed for the proof of Theorem 5.3.

In summary, we have the following proposition.

PROPOSITION 8.1. *$\widehat{\mathcal{E}}^r$ is a Banach manifold. It is naturally identified up to right composition with the Banach Lie group* Diff^r*, and with an open subset of a closed affine subspace of* $\mathcal{C}^r(I, \mathbb{R})$.

An estimate similar to that shown in Proposition 5.1 for $r = 1$ holds for $r > 1$. Therefore one has, for $r = k + \gamma, \infty, \omega$, the following proposition.

PROPOSITION 8.2.

(a) *The $\mathcal{C}^r$ metric on* Diff^r *is right-invariant up to (non-uniform) bounded equivalence.*
(b) *The $\mathcal{C}^r$ metric on $\widehat{\mathcal{E}}^r$ is base-point independent up to (non-uniform) bounded equivalence.*

We note that $\mathcal{E}^r_*$, the space of marked Cantor sets, is also a Banach manifold by the same reasoning as for Diff^r: it is an open subset of a closed affine subspace of $\mathcal{C}^r(C, \mathbb{R})$.

Next we will describe more fully the relationship between the spaces $\widehat{\mathcal{E}}^r$ and $\mathcal{E}^r$. For $r = k + \gamma$, we write $\mathrm{Diff}^r_0(I)$ for the collection of $\mathcal{C}^r$ diffeomorphisms of the unit interval I whose first k derivatives are $1, 0, \ldots, 0$ at the endpoints. This is also a Banach manifold.

PROPOSITION 8.3. *Given the choice of a Cantor set C, $\widehat{\mathcal{E}}^r$ factors naturally, set theoretically and topologically, as*

$$\widehat{\mathcal{E}}^r = \mathcal{E}^r \times \mathrm{Diff}^r_0(I),$$

with the topology on $\mathcal{E}^r$ defined below.

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

Proof. First, let us consider how many $\mathcal{C}^r$ maps from I to I there are, conjugating (C, S) with $(C, \widehat{S})$ (the Cantor sets are the same, but the maps may be different off of the Cantor sets). In the proof of Lemma 7.3, note that instead of starting with Φ equal to the identity on the gap G we could have taken any $\mathcal{C}^{k+\gamma}$ diffeomorphism from G to itself which has derivatives that agree with the identity at the endpoints, up to order k. Conversely, any conjugacy is specified by its values on G, since elsewhere it is then determined by the dynamics. Therefore we see that the set of $\mathcal{C}^r$ conjugacies from (C, S) to $(C, \widehat{S})$ correspond naturally to $\mathrm{Diff}_0^r(I)$ (this is also true when $S = \widehat{S}$!).

Next we consider how many different $\mathcal{C}^r$ conjugacies are possible in the rigidity theorem from (C, S) to $(\widehat{C}, \widehat{S})$. Given the existence of one such map and hence a restricted conjugacy, we can define (all the) other extensions by a method like that used in the proof just given. That is, we first define the conjugacy arbitrarily on the first-level gaps (but with the correct derivatives of order $\leq k$ at the endpoints), then we extend by the dynamics. Or we can quote that statement directly, making use of a ratio Cantor set as intermediary as in the proof of Theorem 7.5, and now replacing Φ by one of the more general maps described above.

This shows we have a product of sets. $\mathcal{E}^r$ has not yet been given a topology. However, from the product decomposition, we can define a family of metrics as follows. Choosing one element of Diff_0^r defines an embedding into $\widehat{\mathcal{E}}^r$, and we just use the $\mathcal{C}^r$ metric there. (One would like to get a more natural definition by taking the infimum over all such choices; however, it is then not clear that the triangle inequality will hold.) At any rate, the metrics are equivalent and so this defines a natural topology on $\mathcal{E}^r$.

We will show that the $\mathcal{C}^r$ metric on $\widehat{\mathcal{E}}^r$ is equivalent to the product of the metrics on $\mathcal{E}^r$ and Diff_0^r.

It is easy to see that the map from $\mathcal{E}^r$ to each factor is continuous (to Diff_0^r it is also affine). For the converse, given the base point (C, S), first let $f, g \in \mathrm{Diff}^r$ be such that $(C_f, S_f) = (C_g, S_g)$. Write $f_0,\ g_0$ for the corresponding elements of Diff_0^r, i.e. the restrictions of f and g to the middle gap of C (rescaled in the range). We claim that if f_0 and g_0 are close in Diff_0^r, then f and g are close in Diff^r. The formula for f on an nth level gap of C is

$$f(a) = \varphi^f_{x_0 \ldots x_n} \circ f_0 \circ \varphi^{-1}_{x_0 \ldots x_n}(a).$$

Here, φ, φ^f denote inverse branches for S and S_f respectively. By assumption, $\varphi^f = \varphi^g$. Now by bounded distortion (Lemma 6.4) for $k = 1$, and for general k by the proof of Theorem 6.1, $D^k \varphi^f_{x_0 \ldots x_n}$ is uniformly γ-Hölder with constant independent of n. This proves $\|f - g\|_{\mathcal{C}1}$ is small, which is what we wanted to show.

Next we drop the assumption that f and g give the same maps. We have chosen an element of Diff_0^r to define the metric on $\mathcal{E}^r$. Let $\tilde{f}$, $\tilde{g}$ denote the maps in Diff^r such that $\tilde{f}_0 = \tilde{g}_0$ is that element with $(C_f, S_f) = (C_{\tilde{f}}, S_{\tilde{f}})$ and similarly for g. Now, by definition, the distance between the pairs in $\mathcal{E}^r$ is $\|\tilde{f} - \tilde{g}\|_{\mathcal{C}1}$. So we just apply the triangle inequality using the previous case to conclude that $d_C(C_f, C_g) \equiv \|f - g\|_{\mathcal{C}1}$ is also small. □

Remark. In a conversation about the proof of Lemma 7.3, Yair Minsky pointed out to us an interesting parallel between that argument and Sullivan's 'flexibility and rigidity'

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

theorem for Kleinian groups. Sullivan showed, for a finitely generated Kleinian group Γ, that the limit set Λ of the group itself is 'rigid', i.e. a quasiconformal conjugacy (to another Kleinian group) which lives (Lebesgue almost-surely) on Λ must be Möbius. This is a consequence of Sullivan's lemma that Λ carries no measurable Γ-invariant line fields. (Note that by contrast, for hyperbolic Cantor sets, quasisymmetric conjugacy does not imply smooth conjugacy; as we have seen, one also needs to know the scaling function.)

Sullivan used this to show that a quasiconformal conjugacy is determined by a Beltrami differential on Ω/Γ, where Ω is the domain of discontinuity. Thus, in a sense one has rigidity on the limit set and flexibility off it. The group actions correspond to the two (*not* restricted) expanding maps, and the surface Ω/Γ or, equivalently, a fundamental domain for the action of Γ on Ω is analogous to the gap G of the Cantor set. As is the case there, the conjugacy is then specified elsewhere by the dynamics. Sullivan's theorem can then be stated as follows: $\mathrm{Teich}(\Gamma) = \mathrm{Teich}(\Omega/\Gamma)$, where this refers to the Teichmüller space of a group and of a surface respectively; this formulation led to the statement in Proposition 8.3.

Concluding remarks: limit sets as the attractor of a semigroup action. $\mathcal{E}^{1+\gamma}$ denotes the $\mathcal{C}^{1+\gamma}$ equivalence class of a given $\mathcal{C}^{1+\gamma}$ hyperbolic Cantor set. The nested subclasses $\mathcal{E}^r$ for maps of smoothness $r = k + \gamma,\ \infty,\ \omega$, are also conjugate with that higher degree of smoothness. Thus, smoothness classes are also conjugacy classes. Choosing one set in $\mathcal{E}^r$ as a base point, $\mathcal{E}^r$ is naturally identified with a topological factor of the $\mathcal{C}^r$ orientation-preserving diffeomorphisms of the interval, Diff^r, which is a Banach manifold. Moreover, we can choose one Cantor set as a common base point for all the $\mathcal{E}^r$, since by Theorem 7.4 the smoothest Cantor sets exist. Then the nested collections $\mathcal{E}^{1+\gamma} \supseteq \cdots \supseteq \mathcal{E}^r \cdots$ are naturally identified with factors of $\mathrm{Diff}^{1+\gamma} \supseteq \cdots \supseteq \mathrm{Diff}^r \cdots$. (Each is a Banach manifold with its own topology, and is a dense subset of the larger collections with respect to their topologies.) The spaces of marked Cantor sets $\mathcal{E}^r_*$ are also Banach manifolds. The free semigroup on two generators FS_2 acts on each submanifold $\mathcal{E}^r_*$ by replacing it with its left or right Cantor subset. From Theorem 7.4, the limit sets are in the intersection of the $\mathcal{E}^r_*$. From Theorem 5.3, because the bounds are uniform over all Cantor subsets of level n, the collection of limit sets is an attractor for this action. This convergence is exponentially fast in the $\mathcal{C}^1$ norm. (Warning: we have only shown convergence in *this* norm; see the Remark at the end of §6.) The semigroup action on the attractor itself, and the relationship of that action to the scenery process, can be described symbolically very simply as follows. Recall the map $y \mapsto C^y$ for y in the dual Cantor set Σ^- and C^y, the corresponding ratio Cantor set. Now just concatenate y on the right with a finite string of symbols $x_0 x_1 \ldots x_n$. A path in the scenery process is determined by infinitely many such choices of 0 and 1; these are successively added onto y as the bi-infinite word $\underline{x} \equiv (y, x)$ (where $x \equiv x_0 x_1 \ldots$) is shifted to the left in the process $C^{\sigma^n(\underline{x})}$. The geometrical meaning of $x_0 x_1 \ldots$ is this: when zooming down toward the location $\pi(x)$ specified by those digits, in any limit set or, indeed, in any set in that $\mathcal{C}^1$-equivalence class, by Theorem 5.4 this is asymptotically the sequence of sets which we see.

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

Acknowledgements. We wish to thank our colleagues and friends for conversations, encouragement and inspiration regarding this and related projects. We give special thanks to M. Urbanski, B. Mandelbrot, S. Kakutani, and D. Sullivan. The second-named author would also like to thank the Dynamical Systems seminar at Memphis State University for their encouragement to give a series of lectures on these topics (Spring 1990) and Yale University, MSRI, IHES, CUNY, the CNRS, Université Paris-Nord, and SUNY at Stony Brook for their support while the paper was being written.

REFERENCES

[AFi] P. Arnoux and A. M. Fisher. Anosov families: renormalization, random dynamics and the Teichmüller mapping flow (in preparation).

[Be] T. Bedford. Applications of dynamical systems theory to fractal sets: a study of cookie cutter sets. *Proceedings of the Séminaire de Mathématiques Supérieures 'Fractal Geometry and Analysis' (Université de Montréal, NATO ASI Series)*. Kluwer, Amsterdam, 1991.

[BeFi1] T. Bedford and A. M. Fisher. Analogues of the Lebesgue density theorem for fractal sets of reals and integers. *Proc. London Math. Soc.* **64** (1992), 95–124.

[BeFi2] T. Bedford and A. M. Fisher. On the magnification of Cantor sets and their limit models. *Monatsh. Math.* **121** (1996), 11–40.

[BeFi3] T. Bedford and A. M. Fisher. Models for the scenery flow of a hyperbolic Cantor set (in preparation).

[BeFiU] T. Bedford, A. M. Fisher and M. Urbanski. The scenery flow for hyperbolic Julia sets (in preparation).

[Bi] P. Billingsley. *Ergodic Theory and Information.* Wiley, 1965. Reprinted by Krieger, New York, 1978.

[Bo1] R. Bowen. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms (Lecture Notes in Mathematics 470)*. Springer, 1975.

[Bo2] R. Bowen. Hausdorff dimension of quasi-circles. *Publications Mathématiques (Institut des Hautes Etudes Scientifiques, Paris)* **50** (1979), 11–25.

[BoG] T. Bogenschütz and M. Gundlach. Ruelle's transfer operator for random subshifts of finite type. *Ergod. Th. & Dynam. Sys.* **15** (1995), 413–447.

[BuF] M. Burger and A. M. Fisher. Infinite measure unique ergodicity for the horocycle flows of certain Fuchsian groups (in preparation).

[Fa] K. J. Falconer. *The Geometry of Fractal Sets (Cambridge Tracts in Mathematics 85)*. Cambridge University Press, 1985.

[Fi1] A. M. Fisher. Convex-invariant means and a pathwise central limit theorem. *Adv. in Math.* **63** (1987), 213–246.

[Fi2] A. M. Fisher. A pathwise central limit theorem for random walks. *Ann. Prob.* to appear.

[Fi3] A. M. Fisher. Integer Cantor sets and an order-two ergodic theorem. *Ergod. Th. & Dynam. Sys.* **13** (1992), 45–64.

[Fi4] A. M. Fisher. Analogues of geodesic and horocycle flows for certain fractal sets (in preparation).

[Fu] H. Furstenberg. Intersections of Cantor sets and transversality of semigroups. *Problems in Analysis (a symposium in honor of S. Bochner).* Princeton University Press, 1970, pp. 41–59.

[HP] M. Hirsch and C. Pugh. Stable manifolds and hyperbolic sets. *Proc. Sym. in Pure Math.* **14** (1970), 133–163.

[Lam] J. Lamperti. *Probability*. Benjamin-Cummings, 1966.

[GSt] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, 1992.

[Lan] O. Lanford. A computer assisted proof of the Feigenbaum conjectures. *Bull. Amer. Math. Soc.* **6** (1982), 427–434.

[deMvS] W. de Melo and S. J. van Strien. *One-Dimensional Dynamics*. Springer, 1993.

[Mand] B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman, 1983.

[Mañé] R. Mañé. *Ergodic Theory and Differentiable Systems*. Springer, 1987.

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press

[Mi] J. Milnor. Remarks on infinite-dimensional Lie groups. *Relativity, Groups and Topology II*. Elsevier, 1984.

[Pi] A. Pinto. Convergence of renormalisation and rigidity of dynamical systems. *PhD Thesis*, Warwick University, 1991.

[PR1] A. Pinto and D. A. Rand. A characterisation of the moduli space of smooth Markov maps by horizontal scaling functions. *Preprint*, Warwick University.

[PR2] A. Pinto and D. A. Rand. A classification of $\mathcal{C}^{1+\alpha}$ structures on embedded trees. *Preprint*, Warwick University.

[Ra] D. A. Rand. Global phase space universality, smooth conjugacies and renormalization: I. The $\mathcal{C}^{1+\alpha}$ case. *Nonlinearity* **1** (1988), 181–202.

[Ru] D. Ruelle. Repellers for real analytic maps. *Ergod. Th. & Dynam. Sys.* **2** (1982), 99–107.

[ShSu] M. Shub and D. Sullivan. Expanding endomorphisms of the circle revisited. *Ergod. Th. & Dynam. Sys.* **5** (1985), 285–289.

[Su1] D. Sullivan. Differentiable structures on fractal-like sets, determined by intrinsic scaling functions on dual Cantor sets. *Amer. Math. Soc. Proc. Symp. Pure Math.* **48** (1987), 15–23.

[Su2] D. Sullivan. Bounds, quadratic differentials, and renormalization conjectures. *Amer. Math. Soc. Centennial Publications* **2** (1991).

[Su3] D. Sullivan. Linking the universalities of Milnor-Thurston Feigenbaum and Ahlfors-Bers. *Topological Methods in Modern Mathematics*. Publish or Perish, 1992, pp. 543–564.

[T] T. Lei. Similarity between the Mandelbrot set and Julia sets. *Comm. Math. Phys.* **134** (1990), 587–617.

[TPr] F. Tangerman and F. Przytycki. Cantor sets on the line: scaling functions and smoothness of the shift map. *Nonlinearity* **9** (1996), 403–412.

https://doi.org/10.1017/S0143385797079194 Published online by Cambridge University Press