GROWTH IN GEOMETRY AND DYNAMICS

VAUGHN CLIMENHAGA

ABSTRACT. *[Rough notes for a colloquium talk at Rice University on Thursday, January 18, 2024, including some extra details and digressions.]*

One difference between hyperbolic and Euclidean geometry is in how certain quantities grow with respect to distance, such as the relationship between circumference and radius of a circle. A similar distinction arises in dynamical systems, where the growth is with respect to time. This talk will explore the connections between these two areas, including a survey of the role that ergodic theory plays in the Margulis asymptotic estimates for closed geodesics on negatively curved manifolds, and their recent extension to a broader setting.

1. CIRCUMFERENCE AND CURVATURE

1.1. **Growth.** What is the perimeter $L(r)$ of a circle with radius r? In the Euclidean plane $\mathbb{R}^2$ it is $2\pi r$; on the unit sphere $S^2 \subset \mathbb{R}^3$, a short trigonometry exercise shows that it is $2\pi|\sin(r)|$. We will mostly be concerned with the story in the hyperbolic plane $\mathbb{H}^2$, which can be represented in the unit disc or upper half-plane model; see Figure 1. It is an exercise in integration (easiest in the disc model) to show that

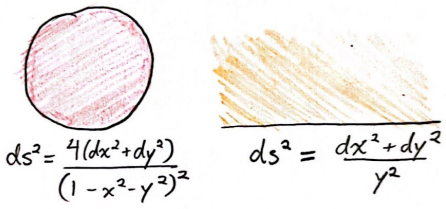

FIGURE 1. Two models of $\mathbb{H}^2$.

$$L(r) = 2\pi \sinh(r) = \pi(e^r - e^{-r}). \tag{1.1}$$

In what follows, the formulas for the metric of $\mathbb{H}^2$ in the two models will not be particularly important; what will be important is that geodesics in both models are (Euclidean) circles and lines that intersect the boundary orthogonally.

These formulas display a trichotomy between exponential growth, polynomial growth, and boundedness, which is related to curvature. The geometries of $\mathbb{R}^2$, S^2, and $\mathbb{H}^2$ have curvature $0, 1, -1$ respectively. More generally, for constant curvature $\kappa > 0$, scale everything by κ to get back to the unit sphere S^2, and deduce that a circle of radius r has circumference

$$L(r) = 2\pi\kappa^{-1}|\sin(\kappa r)|.$$

Similarly, for constant curvature $\kappa < 0$, scale by $h = |\kappa|$ to get to $\mathbb{H}^2$, giving

$$L(r) = 2\pi h^{-1} \sinh(rh) = \pi h^{-1}(e^{rh} - e^{-rh}). \tag{1.2}$$

Date: January 18, 2024.

The author was partially supported by NSF grant DMS-2154378.

Observe that (1.2) lets us recover the curvature from the function $L(r)$: we have $\kappa = -h$, where

$$h = \lim_{r\to\infty} \frac{1}{r} \log L(r). \tag{1.3}$$

Question 1.1. If we perturb the metric on $\mathbb{H}^2$ to have *variable* negative curvature, does the limit in (1.3) exist? If so, what does it represent? Does $L(r)$ satisfy some formula analogous to (1.2)?

To get any sort of meaningful answer to Question 1.1, we will need to impose some constraints on the kinds of perturbations allowed. We will return to this in §3.

1.2. **Product structure.** Before setting Question 1.1 aside temporarily, observe that the exponential growth exhibited in $\mathbb{H}^2$ applies to arcs of the circle as well: if we consider a point $x \in \mathbb{H}^2$ and an arc of vectors subtending an angle $\theta > 0$, then by following each of the corresponding geodesics for a distance r, we obtain an arc with length $\frac{\theta}{2\pi} L(r) = \frac{\theta}{2}(e^r - e^{-r})$. Figure 2 shows this growth for a surface of negative curvature embedded in $\mathbb{R}^3$.

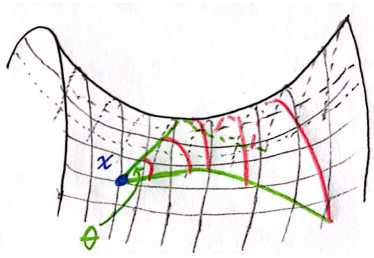

FIGURE 2. Arc growth.

The rapid growth of this arc as r increases means that if we imagine ourselves moving away from x along one of the geodesics determined by the original vectors, but do not know which geodesic specifically we are on, then the range of possible futures available to us becomes very large, very quickly. (Compare this to Euclidean $\mathbb{R}^2$, where the arc length θr grows only linearly.)

A qualitative version of this distinction between $\mathbb{H}^2$ and $\mathbb{R}^2$ is illustrated in Figure 3: given any two geodesics $c_1, c_2 \colon \mathbb{R} \to \mathbb{H}^2$, there is a (unique) geodesic $c \colon \mathbb{R} \to \mathbb{H}^2$ that approaches c_1 in the past and c_2 in the future, in the sense that

$$\lim_{t\to\infty} d(c(-t), c_1(-t)) = 0, \tag{1.4}$$

$$\lim_{t\to\infty} d(c(t), c_2(t)) = 0. \tag{1.5}$$

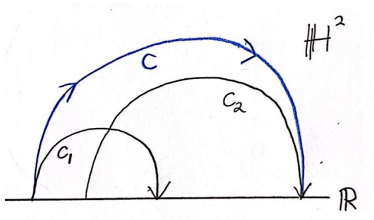

FIGURE 3. Product structure.

We refer to this as having *product structure*: if we interpret a geodesic as a trajectory, and describe a "possible past" as a set of geodesics such that any two of them satisfy (1.4), with a corresponding definition of "possible future" in terms of (1.5), then the set of all (bi-infinite) trajectories is the direct product of the set of possible pasts and the set of possible futures.

Observe that no such product structure is present in Euclidean space: there is no Euclidean geodesic in $\mathbb{R}^2$ that approaches the x-axis in backward time and the y-axis in forward time. As we will see, the presence or absence of product structure plays a crucial role in the distinction between the exponential growth in (1.2) and the slower growth in $\mathbb{R}^2$.

2. Counting configurations

2.1. **A lattice gas model.** For now, we turn our attention to something completely different. Consider a bi-infinite row of boxes, each of which can be either full or empty. We refer to a specific pattern of filled/empty boxes as a *configuration*. Writing 0 for an empty box and 1 for a full one, the set of configurations is $\{0,1\}^{\mathbb{Z}}$.

From now on, we declare that a configuration is *legal* if no two adjacent boxes are full, and restrict our attention to legal configurations. (This is the hard-square lattice gas model from statistical mechanics.)

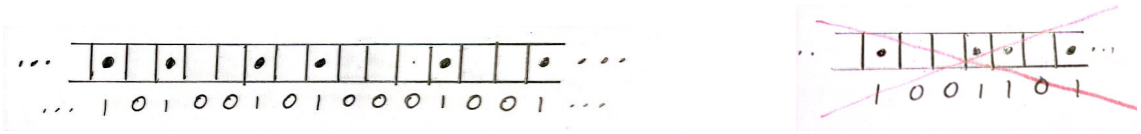

FIGURE 4. Legal and illegal configurations.

Question 2.1. If a configuration is chosen at random, with all legal configurations equally likely, what is the expected proportion of boxes that are filled?

The question requires some interpreting since there are uncountably many legal configurations. The set of legal configurations is

$$\Sigma = \{x \in \{0,1\}^{\mathbb{Z}} : x_k x_{k+1} = 0 \text{ for all } k \in \mathbb{Z}\}. \tag{2.1}$$

Ultimately we would like a measure on Σ, but that will come later. To make sense of the condition that all legal configurations are equally likely, one approach is to look at finite parts of the row of boxes, on which only finitely many configurations are possible. This leads to a secondary question.

Question 2.2. Let L_n be the number of legal configurations for n consecutive boxes; what is L_n?

2.2. **Growth estimates.** Start by observing that if all configurations were legal, then L_n would be multiplicative in the sense that $L_{k+n} = L_k L_n$: given any legal $(n+k)$-configuration, its first k symbols and last n symbols are legal as well, and vice versa. In particular we would have $L_n = (L_1)^n = 2^n$.

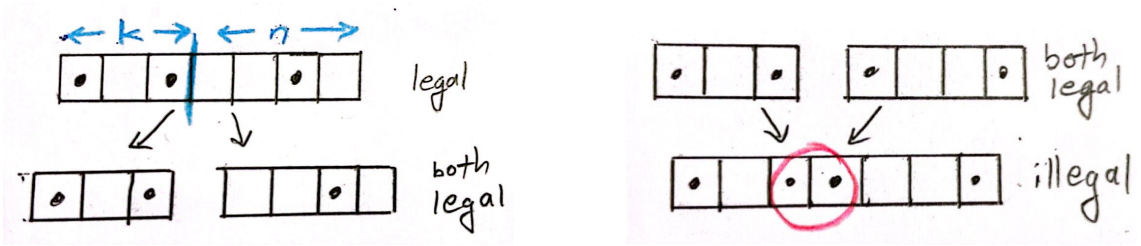

FIGURE 5. L_n is submultiplicative but not multiplicative.

For Σ, however, the requirement that legal configurations have no adjacent full boxes means that the "vice versa" part of the previous paragraph fails: there are some legal configurations $v \in \{0,1\}^k$ and $w \in \{0,1\}^n$ for which vw is not a legal configuration (see Figure 5), so L_n is not multiplicative. However, it is very nearly multiplicative, as we will soon see.

First observe that $L_{k+n} \le L_k L_n$ by the same reasoning as before, and iterating this gives $L_{n\ell} \le (L_n)^\ell$.

Exercise 2.3. Use this fact to prove that $\overline{\lim}_{k\to\infty} \frac{1}{k} \log L_k \le \frac{1}{n} \log L_n$ for all n, and then to deduce that the limit

$$h := \lim_{k\to\infty} \frac{1}{k} \log L_k \tag{2.2}$$

exists, with $L_n \ge e^{nh}$ for all n. This is sometimes called *Fekete's lemma*. Observe that (2.2) is equivalent to $(L_k)^{1/k} \to e^h$.

Remark 2.4. The growth rate h of L_n is the *topological entropy* of Σ.

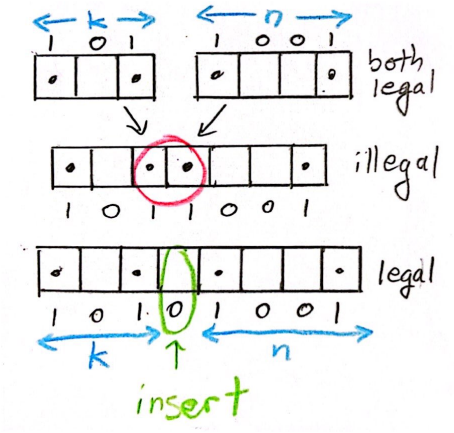

FIGURE 6. L_n is nearly multiplicative.

Given any legal words $v \in \{0,1\}^k$ and $w \in \{0,1\}^n$, the word $v0w$ is always legal (see Figure 6), and thus we have

$$L_k L_n \le L_{n+k+1} \le L_1 L_{n+k} = 2L_{n+k}.$$

Iterating this gives

$$(L_n)^\ell \le 2^\ell L_{n\ell} \quad \Rightarrow \quad L_n \le 2(L_{n\ell})^\ell.$$

Sending $\ell \to \infty$ and using Exercise 2.3 gives

$$e^{nh} \le L_n \le 2e^{nh} \text{ for all } n. \tag{2.3}$$

Remark 2.5. The thing driving the "almost multiplicativity" of L_n here is the *local product structure* of the space Σ: if $0x_1x_2\cdots$ is a legal forward infinite configuration, and $\cdots y_2y_1 0$ is a legal backward infinite configuration, then $\cdots y_2y_1 0x_1x_2\cdots$ is legal, so that locally Σ is the direct product of possible futures and possible pasts (to use more dynamical language). We saw this phenomenon for $\mathbb{H}^2$ in §1.2. This product structure plays a crucial rule in the Margulis argument later on.

From (2.3) we see that the existence of the limit in (2.2), which says that $L_n \approx e^{nh}$, with the ratio of the two sides growing or decaying at most subexponentially, can be strengthened to the statement that $L_n/e^{nh} \in [1,2]$, so that this ratio is in fact bounded away from 0 and ∞. The reader should compare (2.2) to (1.3), and (2.3) to (1.2), observing that the latter is a stronger result than (2.3) since it gives an exact formula. Can we find a corresponding exact formula for L_n?

2.3. **An exact formula.** Yes, we can. A word $w \in \{0,1\}^n$ represents a legal configuration if and only we never have $w_k = w_{k+1} = 1$. At the kth symbol say we are in "state 0" if $w_k = 0$ and "state 1" if $w_k = 1$. State 0 can be followed by either state, but state 1 can only be followed by a 0. This is represented by the directed graph in Figure 7. A word w is legal if and only if it labels a walk on the graph; see Figure 8 for the walk corresponding to the legal word 01001.

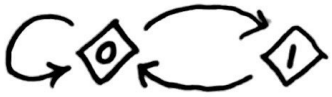

FIGURE 7. Legal transitions.

This situation can be represented by the matrix

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Here $A_{ij} = 1$ if the transition $i \to j$ is allowed, and $A_{ij} = 0$ if it is forbidden. Then a word $w \in \{0,1\}^n$ is legal if and only if $A_{w_1 w_2} A_{w_2 w_3} \cdots A_{w_{n-1} w_n} = 1$. Thus

$$\begin{aligned} L_n &= \#(\text{legal words of length } n) \\ &= \sum_{w \in \{0,1\}^n} A_{w_1 w_2} A_{w_2 w_3} \cdots A_{w_{n-1} w_n} = \sum_{i,j} (A^{n-1})_{ij}. \end{aligned}$$

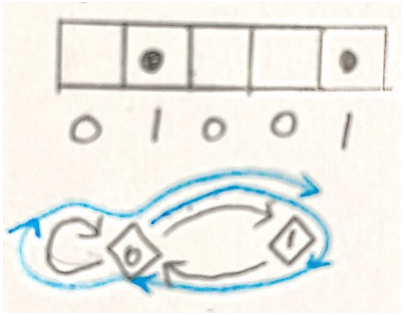

FIGURE 8. A walk on the graph.

One can easily check that the eigenvalues of A are $\lambda_1 = (1+\sqrt{5})/2 \approx 1.618$ (the golden ratio) and $\lambda_2 = -1/\lambda_1 \approx -0.618$, and then a little linear algebra shows that there are $a, b > 0$ such that

$$L_n = a\lambda_1^n + b\lambda_2^n \text{ for all } n \in \mathbb{N}. \tag{2.4}$$

This strengthens (2.3) and provides a version of (1.2) in this case. Observing that $|\lambda_2| < |\lambda_1|$, we can deduce from (2.4) that $L_n/(a\lambda_1^n) \to 1$ as $n \to \infty$, a relationship that we record using the notation $L_n \sim a\lambda_1^n$. Together with (2.2) this implies that $h = \log \lambda_1$, so we can also write the asymptotic as

$$L_n \sim a e^{nh}. \tag{2.5}$$

Remark 2.6. If one actually computes a few iterates $A^n = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)^n$ by hand, it quickly becomes apparent that all entries are Fibonacci numbers, and that L_n is in fact the Fibonacci sequence. But our purpose here is to illustrate a more general technique, as §2.5 makes clear.

Remark 2.7. Note that we have seen four different levels of control on the asymptotic behavior of a sequence such as L_n.

(1) Exponential growth rate as in (1.3), (2.2): $h = \lim \frac{1}{n} \log L_n$, so $L_n \approx e^{nh}$ up to a subexponential error term, or equivalently, L_n/e^{nh} does not grow or decay exponentially fast.
(2) Uniform bounds as in (2.3): existence of a constant $C > 0$ such that $C^{-1} e^{nh} \le L_n \le C e^{nh}$ for all n, or equivalently, L_n/e^{nh} is bounded away from 0 and ∞.

(3) Multiplicative asymptotics as in (2.5): $L_n \sim ae^{nh}$, meaning that L_n/e^{nh} actually converges to a limit, in this case a.
(4) Exact formulas as in (1.2) and (2.4), which (among other things) provide information about the rate of convergence in the previous item.

Remark 2.8. Just as in Remark 2.5, the product structure of Σ is the crucial thing in enabling us to compute L_n via powers (products!) of A: given any state $j \in \{0,1\}$, we can join any past ending in j to any future beginning in j.

2.4. **The measure of maximal entropy.** We answered Question 2.2, but not Question 2.1 on the expected proportion of full boxes. One aspect of this is to produce an appropriate measure μ on Σ. Idea is to take $\mu = \lim_n \mu_n$ where μ_n gives equal weight to each legal n-configuration. To define the measures μ_n, one way is to observe that every legal finite configuration whose first and last symbols agree can be extended periodically, and then set

$$\mathrm{Per}_n = \{x \in \Sigma : x_{k+n} = x_k \text{ for all } k \in \mathbb{Z}\}, \tag{2.6}$$

$$\mu_n = \frac{1}{\#\,\mathrm{Per}_n} \sum_{x \in \mathrm{Per}_n} \delta_x. \tag{2.7}$$

Observe that

$$\#\,\mathrm{Per}_n = \sum_i (A^n)_{ii} = \mathrm{Tr}(A^n) = \lambda_1^n + \lambda_2^n \quad \Rightarrow \quad \#\,\mathrm{Per}_n \sim \lambda_1^n. \tag{2.8}$$

Exercise 2.9. Prove that μ_n converges to a Markov measure μ on Σ; that is, if we choose a sequence $x \in \Sigma$ randomly according to μ_n, then in the limit as $n \to \infty$, the probability distribution of the next symbol conditioned on the entire past in fact only depends on the present symbol.

The Markov measure μ is called the *Parry measure* of Σ, and its transition matrix and stationary vector can be written in terms of the eigendata of A. It has several important properties in addition to being the limiting distribution of the periodic orbit measures.

- It is shift-invariant; the μ-probability of seeing a certain finite configuration starting at position $n \in \mathbb{Z}$ does not depend on n.
- The Markov condition encodes a product structure for the measure μ: if $i \in \{0,1\}$ and R_i denotes the set of all $x \in \Sigma$ such that $x_0 = i$, then for any events $A, B \subset R_i$ that only depend on the future and the past, respectively, we have $\mu(A \cap B) = \mu(A)\mu(B)$. This product structure is important in the Margulis argument later.
- It can be shown to be the unique *measure of maximal entropy* for Σ. Just as the topological entropy $h = \lim \frac{1}{n} \log L_n$ gives the growth rate of the number of legal configurations, associated to every shift-invariant probability measure ν on Σ there is a number $h(\nu)$ called the *measure-theoretic entropy*, and there is a *variational principle* stating that $h = \sup_\nu h_\nu$. A measure attaining the supremum is called a measure of maximal entropy (MME).

2.5. **Topological pressure.** Now we answer Question 2.2 and compute the expected proportion of full boxes. Suppose we sample n-periodic configurations at random (giving equal weight to each) and write S_n for the random variable that records how often the symbol 1 appears in a single n-period. Recall from probability theory that for $k \in \mathbb{N}$, the *kth moment* of S_n is $\mathbb{E}[S_n^k]$. In particular the mean and variance are determined by the first and second moments. The moments can be recovered from the *moment-generating function* of S_n, which is

$$M_n(t) = \mathbb{E}[e^{tS_n}] = \frac{1}{\#\operatorname{Per}_n} \sum_{x \in \operatorname{Per}_n} e^{tS_n(x)}.$$

One can quickly check that derivatives of M_n give the moments:

$$\mathbb{E}[S_n^k] = \frac{1}{\#\operatorname{Per}_n} \sum_{x \in \operatorname{Per}_n} S_n(x)^k = M_n^{(k)}(0).$$

To compute $M_n(t)$, observe that if in place of the transition matrix A we use the matrix

$$A(t) = \begin{pmatrix} 1 & 1 \\ e^t & 0 \end{pmatrix},$$

then we have

$$e^{tS_n(x)} = (e^t)^{S_n(x)} = \prod_{j=1}^{n} A(t)_{x_{j-1}x_j},$$

and thus

$$M_n(t) = \frac{1}{\#\operatorname{Per}_n} \sum_{x \in \operatorname{Per}_n} \prod_{j=1}^{n} A(t)_{x_{j-1}x_j} = \frac{1}{\#\operatorname{Per}_n} \sum_i (A(t)^n)_{ii}.$$

Let

$$Z_n(t) = \sum_i (A(t)^n)_{ii} = \operatorname{Tr}(A(t)^n) = \lambda_1(t)^n + \lambda_2(t)^n, \tag{2.9}$$

where $\lambda_1(t)$ and $\lambda_2(t)$ are the roots of the characteristic polynomial $\lambda^2 - \lambda - e^t$. Then we have

$$M_n(t) = \frac{1}{\#\operatorname{Per}_n} Z_n(t) = \frac{Z_n(t)}{Z_n(0)} \quad \Rightarrow \quad M_n'(t) = \frac{Z_n'(t)}{Z_n(0)} = \frac{d}{dt}(\log Z_n(t))|_{t=0}.$$

From this we can deduce that if $E_n = \mathbb{E}[\frac{1}{n} S_n]$ is the expected proportion of 1s with respect to the measure μ_n, then

$$E_n = \frac{d}{dt}\Big(\frac{1}{n} \log Z_n(t)\Big)\Big|_{t=0}. \tag{2.10}$$

Exercise 2.10. Use (2.9) and (2.10) to prove that the expected proportion of full boxes $\lim_{n\to\infty} E_n$ is given by

$$P'(0) = \frac{5 - \sqrt{5}}{10} \approx 0.276 \quad \text{where } P(t) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(t) = \log \lambda_1(t). \tag{2.11}$$

Remark 2.11. The function $t \mapsto P(t)$ is the *topological pressure* function associated to the potential function that gives weight 1 to the symbol 1 and 0 to the symbol 0. The computation in Exercise 2.10 is a specific case of *Ruelle's formula* for the derivative of pressure. See Figure 9 for a graph of the pressure function; the asymptotes with slopes 0 and $\frac{1}{2}$ correspond to the minimum and maximum frequencies of the symbol 1.

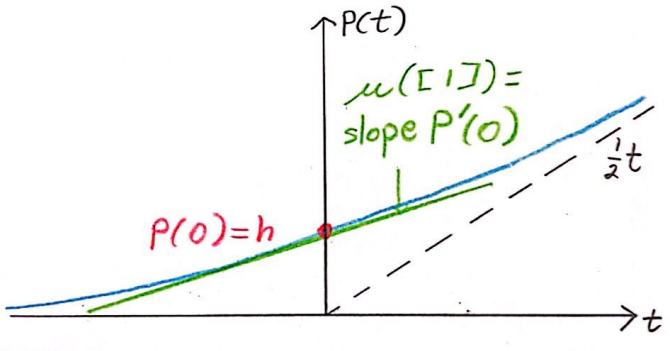

FIGURE 9. Topological pressure.

Exercise 2.12. Recalling Exercise 2.9, prove that the value obtained in (2.11) agrees with the weight that the unique MME gives to the symbol 1, at which point the conclusion about the expected proportion of full boxes can also be deduced from Birkhoff's ergodic theorem.

2.6. **Geometric interpretation.** Returning to the directed graph G, taking the universal cover T (a tree), for each vertex v in T and each $n \in \mathbb{N}$ let us write $S_n(v)$ for the number of vertices in T that lie at a distance exactly n from the reference vertex v; Figure 10 illustrates the case $n = 3$. Thus $S_n(v)$ is a discrete analogue of the "circumference of a circle" from §1.

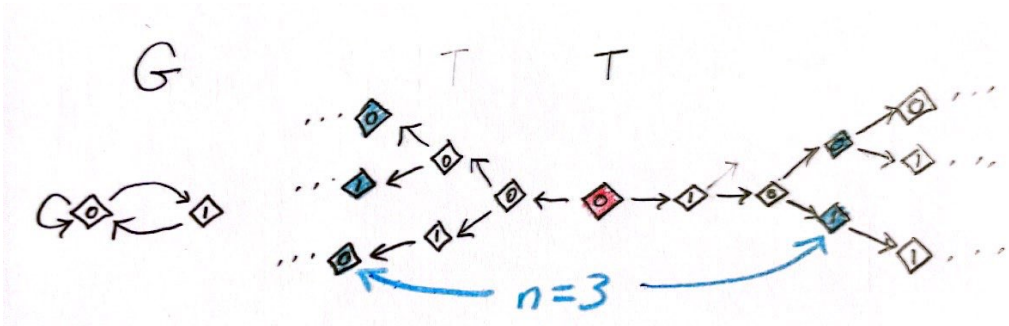

FIGURE 10. The tree covering G.

We have $G = T/\sim$ where $\sim$ represents the equivalence relation of vertices having the same label. If v is labeled with 0, then $S_n(v) = L_n$ since we can go from 0 to both

vertices of G in a single step. If v is labeled with 1 then we have $S_n(v) = L_{n-1}$. In either case, knowledge of the growth of the sequence L_n gives us information about the growth of $S_n(v)$. Observe also that periodic configurations in Σ correspond to periodic paths on G, which lift to walks on T whose initial and final vertices have the same label.

A quick digression: if G' is a different directed graph on finitely many vertices that has the same universal cover T, such as the one shown in Figure 11, then the values of $S_n(v)$ remain the same, since these depend only on the tree T.

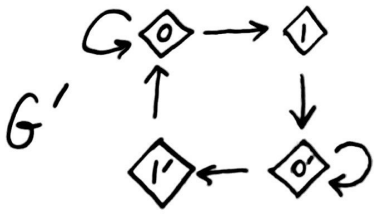

FIGURE 11. A graph with the same universal cover.

Exercise 2.13. Use this fact to prove that there is a constant $C > 0$ such that if L_n and L'_n denote the number of paths of length n on the graphs G and G', respectively, then $C^{-1}L_n \leq L'_n \leq CL_n$, and in particular $\lim_{n\to\infty} \frac{1}{n} \log L_n = \lim_{n\to\infty} \frac{1}{n} \log L'_n$.

From Exercise 2.13 we deduce that the transition matrices for G and G' have the same Perron–Frobenius eigenvalue, and as long as the graphs have the property that the gcd of cycle lengths is equal to 1, then we have $\#\operatorname{Per}_n \sim e^{nh}$ for both G and G', a fact which may be somewhat surprising at first glance.

Exercise 2.14. Verify all of this with the graph in Figure 11, whose transition matrix is

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{2.12}$$

Now we make some observations that set the stage for a return to the geometric setting of §1. In that setting, the mechanism for exponential growth of circumference was negative curvature. Here the growth is driven by the outgoing degree of vertices of T. If every vertex had the same outgoing degree $d \geq 2$, then we would have $S_n(v) = d^n$ for every v; this would be the analogue of the constant negative curvature case. For the graph G from §2.3 and its universal cover T, the outgoing degree is no longer constant but we still get information about the exponential growth of L_n, $\#\operatorname{Per}_n$, and $S_n(v)$. This suggests a way of constraining the metric in Question 1.1 that could lead to meaningful answers:

Question 2.15. If $\sim$ is an appropriate equivalence relation on $\mathbb{H}^2$ and we perturb the metric on the surface $M = \mathbb{H}^2/\sim$, does the limit in (1.3) exist, and does $L(r)$ satisfy some formula analogous to (1.2)? Do periodic geodesics on M grow analogously to (2.8), and does the rate of their growth depend on $\sim$?

3. Counting closed geodesics

3.1. **Quotients of the hyperbolic plane.** In Question 2.15, how should we interpret "an appropriate equivalence relation on $\mathbb{H}^2$"?

Start by thinking about the plane $\mathbb{R}^2$, where the quotient by the action of $\mathbb{Z}^2$ by translations gives the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$, with the unit square as its fundamental domain, whose opposite sides are identified by the translations with displacement vectors $(1,0)$ and $(0,1)$, and these translations generate $\mathbb{Z}^2$. Images of the fundamental domain under $\mathbb{Z}^2$ tile the plane.

Given a geodesic (straight line), if we write 0 whenever it crosses a horizontal boundary and 1 when it crosses a vertical one, we obtain either a periodic sequence (if the slope is rational) or a Sturmian sequence (if the slope is irrational); see Figure 12.

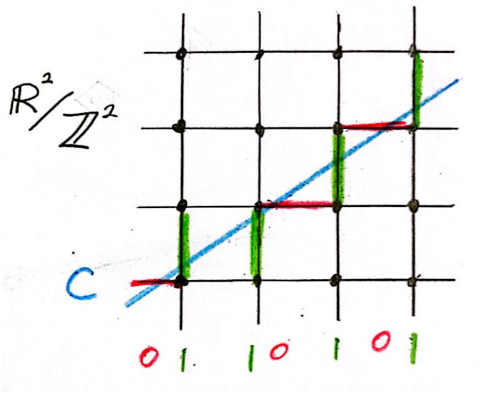

FIGURE 12. Coding a straight line.

For the hyperbolic plane, we will work in the upper half-plane model:

$$\mathbb{H}^2 = \{x + iy \in \mathbb{C} : x, y \in \mathbb{R}, y > 0\}, \text{ with metric } ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

The main thing that we need to remember is that the geodesics in this metric are semi-circles centered on the real line, together with vertical lines.

Observe/recall that $PSL(2,\mathbb{R}) = SL(2,\mathbb{R})/\pm I$ acts isometrically on the upper half-plane model of $\mathbb{H}^2$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : z \mapsto \frac{az+b}{cz+d}.$$

A Fuchsian group is a subgroup $\Gamma \subset PSL(2,\mathbb{R})$ for which the action is discrete ($\Gamma z \subset \mathbb{H}^2$ is discrete for each $z \in \mathbb{H}^2$). The easiest such group to describe algebraically is $\Gamma = PSL(2,\mathbb{Z})$, which is generated by $\alpha(z) = z+1$ and $\beta(z) = -1/z$. (The matrices $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.) As fundamental domain we can take $F = \{z = x+iy \in \mathbb{H}^2 : |z| \geq 1, |x| \leq \frac{1}{2}\}$. The quotient $M = \mathbb{H}^2/PSL(2,\mathbb{Z})$ is the modular surface, corresponding to the tiling shown in Figure 13.

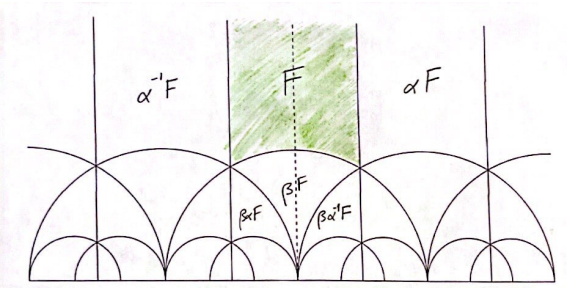

FIGURE 13. Tiling the hyperbolic plane.

3.2. **Coding geodesics (not entirely honestly).** Consider an oriented geodesic in $\mathbb{H}^2$ given by a circle with center on $\mathbb{R}$. It can be broken into segments that cross cells in the modular tiling. Following Adler and Flatto [AF84] (see also [AF91] for

a longer and more in-depth study), we will code the geodesic by recording one of the symbols $1, 2, 3, 4$ each time it crosses from one cell to another.

Our choice of which symbol to record is essentially based on the following idea: each time the geodesic $c\colon \mathbb{R} \to \mathbb{H}^2$ leaves a cell, we can map that cell to the fundamental domain F by an orientation-preserving isometry γ of $\mathbb{H}^2$, and by applying γ to the geodesic c, we obtain a geodesic that leaves the fundamental domain F in one of four ways:[1]

(1) across the line $x = \frac{1}{2}$, moving right;
(2) across the circle $|z| = 1$, moving right;
(3) across the circle $|z| = 1$, moving left;
(4) across the line $x = -\frac{1}{2}$, moving left.

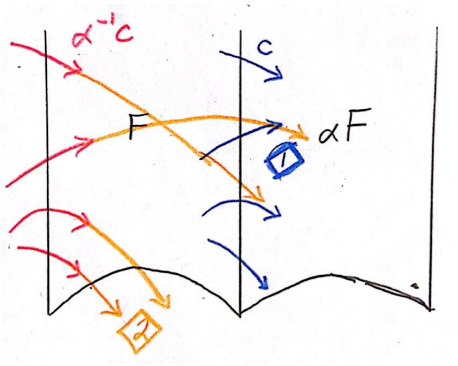

FIGURE 14. Type 1 crossings.

A first attempt would be to record the symbol 1 in the first case, 2 in the second, and so on. Figure 14 illustrates crossings of type 1, and we see that if c exits F through the right-hand boundary, then $\alpha^{-1}c$ enters F through the left-hand boundary, moving to the right, and must therefore exit F as either type 1 or type 2. In other words, if $x \in \{1, 2, 3, 4\}^{\mathbb{Z}}$ codes a geodesic, then x can never contain the word 13 or 14; every instance of 1 must be followed by 1 or 2. Similar considerations mean that every 4 must be followed by 4 or 3.

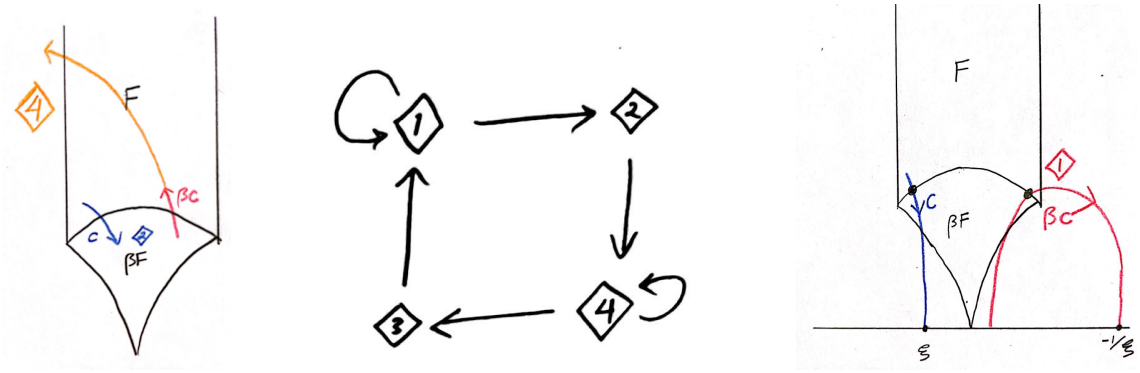

FIGURE 15. An attempted argument, and the problem with it.

Now we might try arguing as follows. (Beware, this contains errors.) If a geodesic c has a type 2 crossing, so that it exits F through the bottom boundary, moving right, then since β is a hyperbolic version of a rotation by π around the fixed point i, we might expect that βc enters F through the bottom boundary, moving left, as illustrated in the first part of Figure 15, and must therefore leave through the

[1]We will only consider geodesics with irrational endpoints on $\mathbb{R}$.

left-hand boundary. This would show that every instance of the symbol 2 must be followed by a 4, and similarly that every 3 must be followed by a 1.

If all of this was correct, then every sequence coding a geodesic would correspond to a walk on the graph shown in Figure 15. (Note that this is the same graph as we saw earlier in Figure 11!) This would in turn suggest that we might understand geodesics on the modular surface $\mathbb{H}^2/PSL(2,\mathbb{Z})$ by studying the corresponding subshift of finite type $\Sigma \subset \{1,2,3,4\}^{\mathbb{Z}}$, which contains all bi-infinite sequences that label infinite walks on this graph. However, there are two problems.

First, we have not yet given any argument that *every* sequence labeling a walk on this graph actually codes a geodesic by the above procedure. Second, it is actually possible for a type 2 crossing to be following by a type 1 crossing, as the third picture in Figure 15 illustrates. So what should we do?

3.3. **Coding geodesics (honestly).** It turns out that we can resolve things by using a slightly different coding. Given a geodesic $c\colon \mathbb{R} \to \mathbb{H}^2$, let $c(-\infty)$ and $c(\infty)$ denote the endpoints of c on the ideal boundary $\mathbb{R}$, and consider the following four sets of geodesics:

$$\begin{aligned} R_1 &= \{c : c(-\infty) < 0 \text{ and } c(\infty) > 1\},\\ R_2 &= \{c : c(-\infty) < -1 \text{ and } 0 < c(\infty) < 1\},\\ R_3 &= -R_2, \qquad R_4 = -R_1. \end{aligned}$$

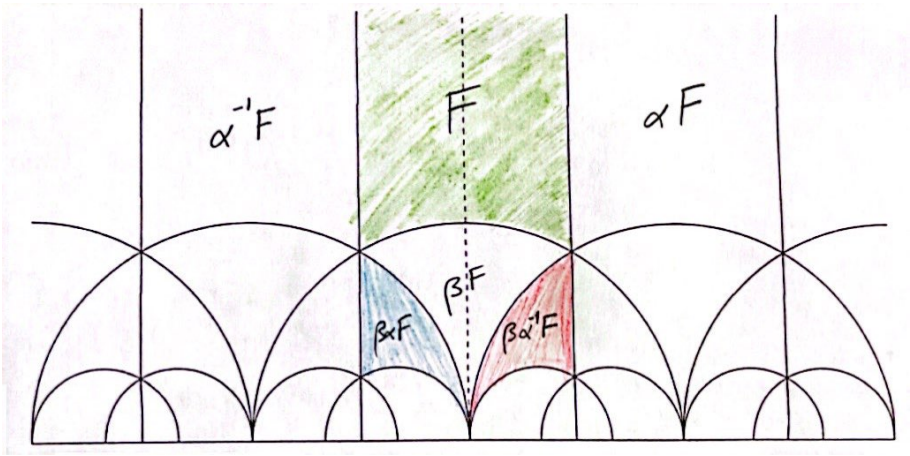

FIGURE 16. Representing geodesics.

Exercise 3.1. Prove that given a geodesic c that leaves a cell F' at time t, there exists a unique isometry γ that carries c into one of the sets R_j and carries $c'(t)$ to a vector that is leaving either F, or $\beta\alpha F$, or $\beta\alpha^{-1}F$.

Using Exercise 3.1 we can associate to each geodesic c with irrational endpoints a sequence in $\{1,2,3,4\}^{\mathbb{Z}}$. Let c_n denote the image of c under the isometry associated

to the nth coding, and let $\xi_n = c_n(\infty)$ and $\eta_n = c_n(-\infty)$. Then $(\xi_{n+1}, \eta_{n+1}) = f(\xi_n, \eta_n)$, where

$$f(\xi,\eta) = \begin{cases} (\xi - 1, \eta - 1) & \text{on } R_1, \\ (-1/\xi, -1/\eta) & \text{on } R_2 \cup R_3, \\ (\xi + 1, \eta + 1) & \text{on } R_4. \end{cases}$$

The regions R_i map onto each other as shown in Figure 17. Moreover, f maps horizontal lines to horizontal lines, and vertical lines to vertical lines. Using this, together with the fact that each of the regions R_i is a rectangle – a direct product of a horizontal interval and a vertical interval – it is possible to show that a sequence in $\{1,2,3,4\}^{\mathbb{Z}}$ codes a geodesic if and only if it labels a walk on the graph in Figure 15. In other words, the set of legal sequences exactly corresponds to the subshift of finite type coded by the matrix A in (2.12).

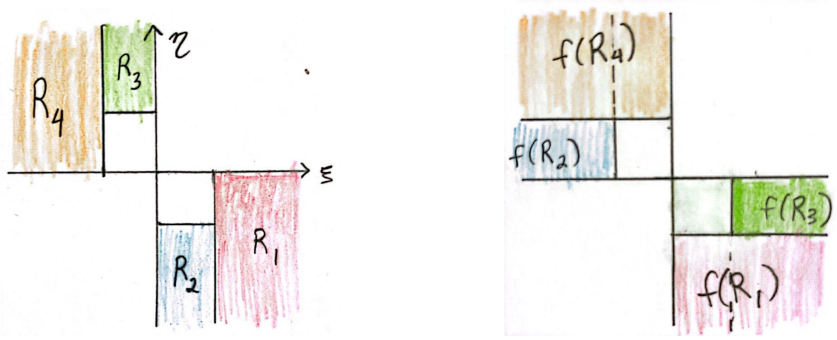

FIGURE 17. A Markov partition for f.

We conclude this section by observing that since f maps vertical lines into vertical lines, we can quotient out the η-direction and get a one-dimensional map $g\colon \mathbb{R} \to \mathbb{R}$ as shown in Figure 18. This map tracks the evolution of the future endpoint $c(\infty)$ as we move the geodesic to intersect the appropriate fundamental domain. Studying this map provides information about the geodesic flow on the modular surface, and this map is closely connected to the Gauss map $x \mapsto \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$ that plays a central role in the study of continued fractions.

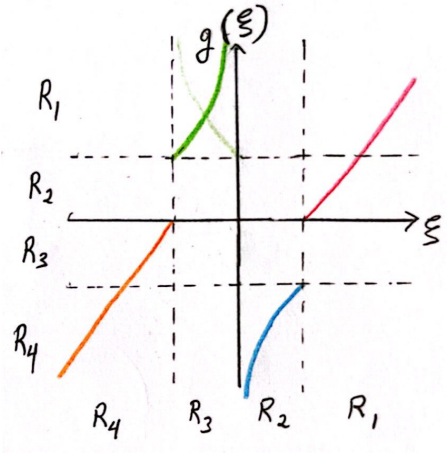

FIGURE 18. A map on $\mathbb{R}$.

3.4. **Asymptotics.** Can use the symbolic representation from the previous section to apply the linear algebra / operator theory / spectral / eigendata approach in §2 and deduce the following.[2]

[2]The details of this are rather more involved and require us to use not just the entropy h but also the topological pressure (as in §2.5) associated to an appropriate "roof function" that measures how long a geodesic takes to cross the fundamental domain between each edge crossing.

Theorem 3.2 ("Prime number theorem for geodesics")**.** *Let* $\mathrm{Per}(T)$ *denote the set of periodic geodesics on* $\mathbb{H}^2/\Gamma$ *with length* $\leq T$*. Then* $\#\mathrm{Per}(T) \sim \frac{e^T}{T}$*.*

Various proofs: Huber [Hub59], Margulis [Mar69, Mar04], Parry and Pollicott [PP83]. See below for comparison. Sharp's survey in [Mar04] provides a much more comprehensive account.

Remark 3.3. Use length $\leq T$ instead of $= T$ because T varies continuously but number of periodic geodesics is countable. Factor of T in the denominator (compared to (2.8)) comes for the following reason: in (2.8) we can group together elements of Per_n that are related by a shift, and if n is the least period then there are n elements in each such grouping, so the number of such groupings is $\sim ce^{nh}/n$.

Question 2.1 on expected proportion of full boxes in lattice gas translates to a question about the proportion of "cusp windings" for a typical closed geodesic on the modular surface. (With a different numerical answer.)

Remark 3.4. The analogy to the prime number theorem is as follows: writing $\pi(N)$ for the number of primes $\leq N$, the PNT says that $\pi(N) \sim N/\log N$, and writing $T = \log N$ this is equivalent to $\#\{p \text{ prime: } \log p \leq T\} \sim \frac{e^T}{T}$.

Remark 3.5. In fact Theorem 3.2 applies to any discrete $\Gamma \subset PSL(2,\mathbb{R})$ such that $\mathbb{H}^2/\Gamma$ has finite area, such as the surface of genus 2 in Figure 19. (We get the same asymptotics!) The original proof of Huber [Hub59] involved the Selberg trace formula, which relates lengths of closed geodesics to the spectrum of the Laplacian.

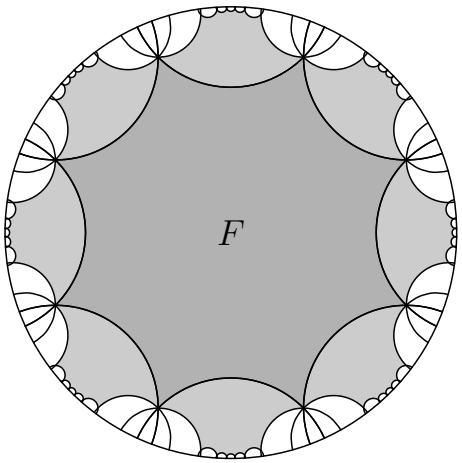

FIGURE 19. A surface of genus 2.

3.5. **Variable curvature.** Now let M be any compact surface with variable negative curvature, so its universal cover X is a topological disc on which $\Gamma = \pi_1(M)$ acts freely and discretely by isometries.

Theorem 3.6 (Margulis asymptotics)**.** *There is* $h > 0$ *such that* $\#\mathrm{Per}(T) \sim \frac{e^{hT}}{hT}$*, and moreover there is a continuous function* $c\colon M \to (0,\infty)$ *such that if we write* $L(x,r)$ *for the perimeter of the circle in* X *with radius* r *and center at (a lift of)* x*, then* $L(x,r) \sim c(x)e^{hr}$*.*

Remark 3.7. The result holds in higher dimensions too, provided M has negative sectional curvatures.

One proof of Theorem 3.6 is due to Parry and Pollicott [PP83] and follows the "symbolic coding" approach above. The geodesic flow on (the unit tangent bundle of) M is a "uniformly hyperbolic" flow, and it is a fundamental fact of hyperbolic

dynamical systems (Sinai, Ratner, Bowen, Ruelle) [Sin68, Rat69, Bow73, BR75] that such flows can be coded by (suspension flows over) subshifts of finite type. This opens the door to use zeta functions and spectral theory, and in fact provides more precise asymptotic estimates than those stated in Theorem 3.6 [PS98].

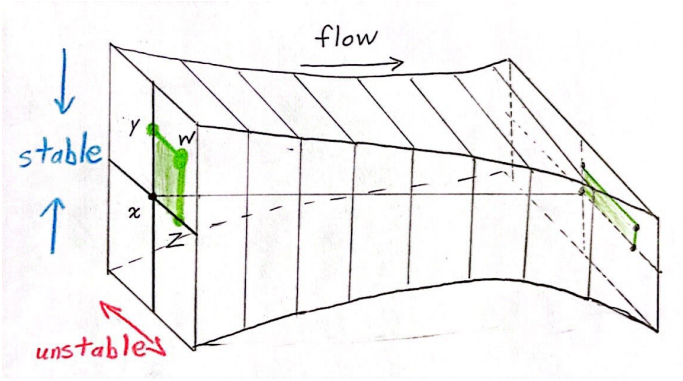

FIGURE 20. A uniformly hyperbolic flow; product structure.

The original proof due to Margulis, on the other hand, does not use a symbolic coding. It relies on the following ingredients.

(1) There is a unique measure of maximal entropy m, to which the periodic orbit measures equidistribute.
(2) m has product structure with respect to stable and unstable directions.
(3) The conditional measures of m along stable and unstable leaves have appropriate scaling properties.

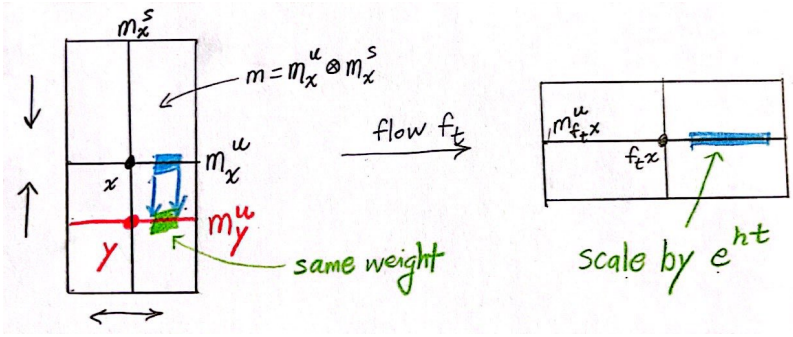

FIGURE 21. The scaling properties requires for the Margulis argument.

Remark: in constant curvature, MME is Liouville. For surfaces at least, this is the only time it happens; [Kat82, BCG95].

4. Beyond negative curvature

Margulis asymptotics for closed geodesics have been proved for the following classes of manifolds.

- Surfaces with negative curvature outside radially symmetric caps: Bryce Weaver [Wea14].
- Rank 1 manifolds of nonpositive curvature, in fact CAT(0): Russell Ricks [Ric22].
- Rank 1 manifolds without focal points: Weisheng Wu [Wu23].
- Surfaces of genus ≥ 2 without conjugate points: C.–Knieper–War [CKW22].

In last 3 cases, Weisheng Wu also proved volume asymptotics [Wu23].

One of the main innovations in CKW [CKW22] was the use of C–Thompson specification machinery [CT12, CT13, CT14, CT16, CT21] to prove uniqueness of the MME (done in [CKW21]). This approach builds on the kinds of computations in §2.2 and goes back to work of Rufus Bowen in the 1970s [Bow75].

It is also interesting to consider systems that are not geodesic flows. Two rather important examples include:

- Sinai billiards: flat $\mathbb{T}^2$ with scatterers removed, straight-line flow reflecting at boundaries. Uniformly hyperbolic, but singularities prevent SFT coding so Parry–Pollicott approach doesn't immediately go through.
- Lorenz attractor: "butterfly" in $\mathbb{R}^3$, chaotic attractor associated to ODEs $\dot{x} = \sigma(y - x)$, $\dot{y} = x(\rho - z) - y$, $\dot{z} = xy - \beta z$ with $\sigma = 10$, $\beta = 8/3$, $\rho = 28$. Non-uniformly hyperbolic; some trajectories spend arbitrarily long time near fixed point at origin.

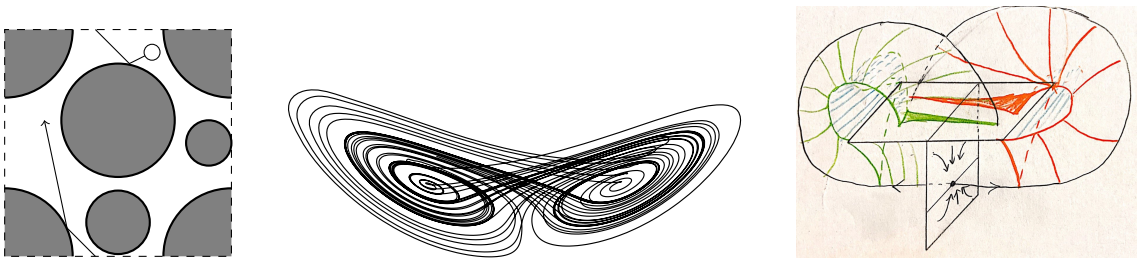

FIGURE 22. The Sinai billiard and Lorenz attractor.

For Sinai, unique MME recently obtained by Baladi–Demers [BD20]. To get leaf measures with proper scaling properties, C.–Day [CD23, CD24] followed approach of C.–Pesin–Zelerowicz [CPZ19] in adapting the definition of Hausdorff measure from geometric to dynamical. Derivation of actual Margulis asymptotics is still work in progress.

For Lorenz, unique MME recently obtained by Pacifico–Yang–Yang [PYY22a, PYY22b], using improved version of C.–Thompson non-uniform specification. No work yet on product structure, idea is to use CPZ approach here too.

Remark that for Lorenz, there is also the SRB measure, which is the "physical" measure.

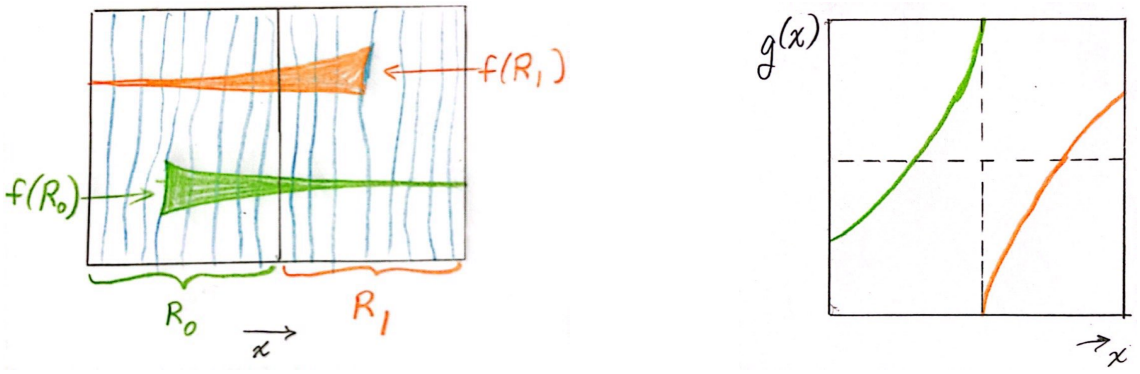

FIGURE 23. Return map for the geometric Lorenz flow; interval map obtained by quotienting out the stable direction.

REFERENCES

[AF84] Roy L. Adler and Leopold Flatto, *Cross section map for the geodesic flow on the modular surface*, Conference in modern analysis and probability (New Haven, Conn., 1982), Contemp. Math., vol. 26, Amer. Math. Soc., Providence, RI, 1984, pp. 9–24. MR 737384

[AF91] Roy Adler and Leopold Flatto, *Geodesic flows, interval maps, and symbolic dynamics*, Bull. Amer. Math. Soc. (N.S.) **25** (1991), no. 2, 229–334. MR 1085823

[BCG95] G. Besson, G. Courtois, and S. Gallot, *Entropies et rigidités des espaces localement symétriques de courbure strictement négative*, Geom. Funct. Anal. **5** (1995), no. 5, 731–799. MR 1354289

[BD20] Viviane Baladi and Mark F. Demers, *On the measure of maximal entropy for finite horizon Sinai billiard maps*, J. Amer. Math. Soc. **33** (2020), no. 2, 381–449. MR 4073865

[Bow73] Rufus Bowen, *Symbolic dynamics for hyperbolic flows*, Amer. J. Math. **95** (1973), 429–460. MR 339281

[Bow75] ———, *Some systems with unique equilibrium states*, Mathematical Systems Theory **8** (1975), no. 3, 193–202.

[BR75] Rufus Bowen and David Ruelle, *The ergodic theory of Axiom A flows*, Invent. Math. **29** (1975), no. 3, 181–202. MR 380889

[CD23] Vaughn Climenhaga and Jason Day, *Equilibrium measures for two-sided shift spaces via dimension theory*, Preprint, arXiv:2309.11410.

[CD24] ———, *Product structure of the measure of maximal entropy for dispersing billiards*, Preprint.

[CKW21] Vaughn Climenhaga, Gerhard Knieper, and Khadim War, *Uniqueness of the measure of maximal entropy for geodesic flows on certain manifolds without conjugate points*, Adv. Math. **376** (2021), Paper No. 107452, 44. MR 4178924

[CKW22] ———, *Closed geodesics on surfaces without conjugate points*, Commun. Contemp. Math. **24** (2022), no. 6, Paper No. 2150067, 35. MR 4466167

[CPZ19] Vaughn Climenhaga, Yakov Pesin, and Agnieszka Zelerowicz, *Equilibrium states in dynamical systems via geometric measure theory*, Bull. Amer. Math. Soc. (N.S.) **56** (2019), no. 4, 569–610. MR 4007162

[CT12] Vaughn Climenhaga and Daniel J. Thompson, *Intrinsic ergodicity beyond specification: β-shifts, S-gap shifts, and their factors*, Israel J. Math. **192** (2012), no. 2, 785–817. MR 3009742

[CT13] ———, *Equilibrium states beyond specification and the Bowen property*, J. Lond. Math. Soc. (2) **87** (2013), no. 2, 401–427. MR 3046278

[CT14] ———, *Intrinsic ergodicity via obstruction entropies*, Ergodic Theory Dynam. Systems **34** (2014), no. 6, 1816–1831. MR 3272773

[CT16] ———, *Unique equilibrium states for flows and homeomorphisms with non-uniform structure*, Adv. Math. **303** (2016), 745–799. MR 3552538

[CT21] ———, *Beyond Bowen's specification property*, Thermodynamic formalism, Lecture Notes in Math., vol. 2290, Springer, Cham, [2021] ©2021, pp. 3–82. MR 4436821

[Hub59] Heinz Huber, *Zur analytischen Theorie hyperbolischen Raumformen und Bewegungsgruppen*, Math. Ann. **138** (1959), 1–26. MR 109212

[Kat82] A. Katok, *Entropy and closed geodesics*, Ergodic Theory Dynam. Systems **2** (1982), no. 3-4, 339–365. MR 721728

[Mar69] G. A. Margulis, *Certain applications of ergodic theory to the investigation of manifolds of negative curvature*, Funkcional. Anal. i Priložen. **3** (1969), no. 4, 89–90. MR 257933

[Mar04] Grigoriy A. Margulis, *On some aspects of the theory of Anosov systems*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2004, With a survey by Richard Sharp: Periodic orbits of hyperbolic flows, Translated from the Russian by Valentina Vladimirovna Szulikowska. MR 2035655

[PP83] William Parry and Mark Pollicott, *An analogue of the prime number theorem for closed orbits of Axiom A flows*, Ann. of Math. (2) **118** (1983), no. 3, 573–591. MR 727704

[PS98] Mark Pollicott and Richard Sharp, *Exponential error terms for growth functions on negatively curved surfaces*, Amer. J. Math. **120** (1998), no. 5, 1019–1042. MR 1646052

[PYY22a] Maria Jose Pacifico, Fan Yang, and Jiagang Yang, *Existence and uniqueness of equilibrium states for systems with specification at a fixed scale: an improved Climenhaga-Thompson criterion*, Nonlinearity **35** (2022), no. 12, 5963–5992. MR 4500887

[PYY22b] Maria Jose Pacifico, Fan Yang, and Jiagang Yang, *Uniqueness of equilibrium states for lorenz attractors in any dimension*, arXiv:2201.06622.

[Rat69] M. E. Ratner, *Markov decomposition for an u-flow on a three-dimensional manifold*, Mat. Zametki **6** (1969), 693–704. MR 260977

[Ric22] Russell Ricks, *Counting closed geodesics in a compact rank-one locally CAT(0) space*, Ergodic Theory Dynam. Systems **42** (2022), no. 3, 1220–1251. MR 4374971

[Sin68] Ja. G. Sinaĭ, *Markov partitions and u-diffeomorphisms*, Funkcional. Anal. i Priložen. **2** (1968), no. 1, 64–89. MR 233038

[Wea14] Bryce Weaver, *Growth rate of periodic orbits for geodesic flows over surfaces with radially symmetric focusing caps*, J. Mod. Dyn. **8** (2014), no. 2, 139–176. MR 3277199

[Wu23] Weisheng Wu, *Volume asymptotics, margulis function and rigidity beyond nonpositive curvature*, Math. Ann. (2023).

Dept. of Mathematics, University of Houston, Houston, TX 77204
Email address: climenha@math.uh.edu