CHAPTER 2

Background

2.1. Lattices and Fundamental Domains

Here are some basic geometric objects we consider in this monograph.

Convex Polytopes: Just for the record, we say that a *convex polytope* is the convex hull of (i.e. the smallest closed convex set which contains) a finite union of points in Euclidean space. The intersection of finitely many convex polytopes is again a convex polytope, possibly of lower dimension.

We will not bother to define a general (non-convex) polytope, because all the polytopes we consider are convex, except for several explicit 2-dimensional examples.

Parallelotopes: As a special case, a *parallelotope* in $\boldsymbol{R}^n$ is a convex polytope of the form $T(Q)$, where Q is an n-dimensional cube in $\boldsymbol{R}^n$ and T is an invertible affine transformation of $\boldsymbol{R}^n$.

Euclidean Lattices: We define a *lattice* in $\boldsymbol{R}^n$ to be a discrete abelian group of the form $T(\boldsymbol{Z}^n)$, where T is an invertible linear transformtion. When L is a lattice in $\boldsymbol{R}^n$, the quotient $\boldsymbol{R}^n/L$ is a flat torus.

Fundamental Domains: Let L be a lattice. We say that a convex polytope F is a *fundamental domain* for L if the union

$$\bigcup_{V \in L} (F + V) \tag{2.1}$$

is a tiling of $\boldsymbol{R}^n$. By this we mean that the translates of the interior of F by vectors in L are pairwise disjoint and the translates of F itself cover $\boldsymbol{R}^n$.

Equivalently, we can say that F is a fundamental domain if the following is true.

- For all points $p \in \boldsymbol{R}^n$ there is some vector $V_p \in L$ such that $p + V_p \in F$.
- F and $\boldsymbol{R}^n/L$ have the same volume.

These conditions in turn imply that the choice of V_p is unique unless p lies in a certain countable union of codimension 1 sets – namely the L-translates of the faces of F.

2.2. Hyperplanes

Basic Definition: A *hyperplane* in $\boldsymbol{R}^n$ is a solution V to an equation

$$v \cdot n = d \qquad \forall\, v \in V. \tag{2.2}$$

for some $d \in \boldsymbol{R}$ and some nonzero vector n. The vector n is called a *normal* to V. One might take n to be a unit vector, but this is not necessary. When $d = 0$, the hyperplane V is a vector subspace of $\boldsymbol{R}^n$.

Matrix Action: Suppose that $M : \boldsymbol{R}^n \to \boldsymbol{R}^n$ is an invertible linear transformation. We shall have occasion to want to know how M acts on hyperplanes. Suppose V is the hyperplane satisfying Equation 2.2. Then $M(V)$ is the hyperplane satisfying the equation

$$w \cdot (M^{-1})^t(n) = d \qquad \forall\, v \in M(V) \tag{2.3}$$

This equation derives from the fact that, in general $M(v) \cdot w = v \cdot M^t(w)$.

Parallel Families: We say that a *parallel family* of hyperplanes is a countable discrete set of evenly spaced parallel hyperplanes. The *direction* of the parallel family is specified by giving a normal vector to any hyperplane in the family. All the hyperplanes in the family have the same normal.

Full Families: We say that a collection of n parallel families of hyperplanes is *full* if the corresponding normals form a basis for for $\boldsymbol{R}^n$. The standard example is the collection $C_1, ..., C_n$, where C_i consists of those points $(x_1, ..., x_n)$ with $x_i \in \boldsymbol{Z}$.

Lemma 2.1. *An arbitrary full family is equivalent to the standard example by some affine map. In particular, the complement of a full family is a periodic tiling by parallelotopes.*

Proof: Let $v_1, ..., v_n$ be a basis of normals, where v_i is normal to the ith family. Let M be a matrix such that $(M^{-1})^t$ carries our basis to the standard basis. Then M maps the ith full family to a family parallel to C_i defined above. Further composing with a diagonal matrix and then translating, we get exactly the map we seek. $\square$

2.3. The PET Category

We define a category **PET** whose objects are convex polytopes and whose morphisms are (equivalence classes) of PETs between them.

PETs: A PET from X to Y is a map of the form given in Equation 1.2, when we have partitions

$$X = \bigcup_{i=1}^{m} A_i, \qquad Y = \bigcup_{i=1}^{m} B_i. \tag{2.4}$$

with A_i and B_i being translates.

Composition: If we have PET maps $f : X \to Y$ and $f' : Y \to Z$ we can compose them and get a PET map $f' \circ f : X \to Z$. Concretely, suppose that $f : X \to Y$ is defined in terms of partitions $\{A_i\}$ and $\{B_i\}$ and $f' : X \to X$ is defined in terms of partitions $\{A'_i\}$ and $\{B'_i\}$. Then $f' \circ f$ is defined in terms of the partitions

$$\{f^{-1}(B'_i) \cap A_j\}, \qquad \{f'(B_i \cap A'_j)\}. \tag{2.5}$$

and equals the obvious composition on each piece.

Equivalence: We call $f_1, f_2 : X \to Y$ *equivalent* if the two maps agree on a convex polytope partition which is a common refinement of the partitions defining the maps. It is easy to see that the relation on PETs from X to Y really is an equivalence relation.

The Category: PET is the category whose objects are convex polytopes and whose morphisms are equivalence classes of PETs between the polytopes. It is easy that the composition defined above respects the equivalence classes. Hence **PET** really is a category.

Remark: Sometimes it is annoying to distinguish between a PET and an equivalence class of PETs. We make the distinction here so that we can interpret equation 1.3 precisely. In practice, however, we will have a concrete PET defined in terms of some explicit partitions.

2.4. Periodic Tiles for PETs

Let (X, f) be a PET, as in §1.1. The map $f : X \to X$ is based on the two partitions of X, as in Equations 1.1 and 1.2.

Let $\mathcal{A}_1$ denote the collection $\{A_1, ..., A_n\}$ of polytopes in the first partition of X and let $\mathcal{B}_1$ denote the collection $\{B_1, ..., B_n\}$ of polytopes in the second partition.

For positive integers $n \geq 2$, we inductively define $\mathcal{A}_n$ to be the collection of polyhedra

$$f^{-1}(f(P) \cap A), \qquad P \in \mathcal{A}_{n-1}, \qquad A \in \mathcal{A}_1. \tag{2.6}$$

The partition $\mathcal{A}_n$ refines the partition $\mathcal{A}_{n-1}$. All these partitions consist of convex polytopes, and the power f^n is defined on the complement of the union

$$\bigcup_{P \in \mathcal{A}_n} \partial P. \tag{2.7}$$

LEMMA 2.2. *There is a codimension* 1 *subset* $S \subset X$ *such that every point of* $X - S$ *has a well-defined orbit.*

Proof: Take S' to be the union of all the sets in Equation 2.7. This set has codimension 1, and every point in $X - S'$ has a well-defined forward orbit. There is an analogous set $S'' \subset X$ such that every point in $X - S''$ has a well-defined backwards orbit. We let $S = S' \cup S''$ and we are done. □

LEMMA 2.3. *Let* P *be an open polytope of* $\mathcal{A}_n$. *Suppose that some point of* P *is periodic, with period* n. *Then all points of* P *are periodic, with period* n.

Proof: The first n iterates of f "do the same thing" on the interior of each polytope of $\mathcal{A}_n$. More precisely, if p_1 and p_2 lie in the same open polytope in $\mathcal{A}_n$ then

$$f^k(p_1) - f_k(p_2) = p_1 - p_2, \qquad k = 1, ..., n. \tag{2.8}$$

Our lemma follows immediately from this observation. □

COROLLARY 2.4. Suppose that $p \in X$ is a periodic point. Then there exists a maximal open convex polytope U_p such that f is entirely defined and periodic on U_p.

Proof: Let n be the minimal period of p. Since f^n is defined on p, there is some unique convex open polytope U_p of $\mathcal{A}_n$ which contains p in its interior. By the previous result, all points of U_p have period n. By definition, f^n is not defined on ∂U_p, so U_p is the maximal domain having the advertised properties. □

Remark: There is one subtle point about our previous argument which deserves to be made more clear. Suppose, for instance, that there is some n such that every point on which f is defined has period n. It is tempting to then say that f^n is defined, and the identity, on all of X. However, in order for f^n to be defined, all the iterates f^k must be defined for $k = 1, ..., n$. So, we will not say that f^n is defined, and the identity, on all of X. Rather, we will keep to the original convention and say that f^n is defined, and the identity, on the complement of the set in Equation 2.7.

As mentioned in the introduction, the periodic tiling Δ is the union of the open periodic tiles of f. From the results above, each tile of Δ is an open convex polytope belonging to some $\mathcal{A}_n$. We might have made all of the above definitions and arguments in terms of the inverse map f^{-1} and the analogous partitions $\mathcal{B}_n$. Thus, we can say at the same time that each tile of Δ is an open convex polytope belonging to some $\mathcal{B}_n$. We find it more convenient to work with the forward iterates of f, however.

LEMMA 2.5. *If $\{P_k\}$ is any sequence of tiles in Δ, the period of points in P_k tends to ∞ with k.*

Proof: If this is false, then there is a single n such that P_k is an open polytope of $\mathcal{A}_n$ for all k. But $\mathcal{A}_n$ has only finitely many polytopes. □

Remark: It is worth pointing out that Δ might be empty. If one is willing to disregard a countable set of points, one can view an irrational rotation of the circle as a 2-interval IET. Such an IET has no periodic points at all. D. Genin noticed a similar phenomenon for outer billiards relative to irrational trapezoids [**G**]. We are interested in the opposite case, when Δ is dense.

2.5. The Limit Set

In the introduction we gave a definition of the limit set which works for the systems we study in this monograph. Here we give a more robust definition. We call a point $p \in X$ *weakly aperiodic* if there is a sequence $\{q_n\}$ converging to p with the following property. The first n iterates of f are defined on q_n and the points $f^k(q_n)$ for $k = 1, ..., n$ are distinct. We let Λ denote the union of weakly aperiodic points. We call Λ the *limit set*. Some authors call Λ the *residual set*.

Not all points of Λ need to have well-defined orbits. In fact, the set $\Lambda' \subset \Lambda$ of aperiodic points is precisely the set of points of Λ with well-defined orbits. Now we reconcile the definition here with what we said in the introduction.

LEMMA 2.6. *Suppose that the periodic tiling Δ is dense. Then a point belongs to Λ if and only if every open neighborhood of the point contains infinitely many tiles of Δ.*

Proof: Let Λ^* denote the set of points p such that every neighborhood of p contains infinitely many tiles of Δ. We want to prove that $\Lambda = \Lambda^*$. Every point of Λ^* is the accumulation point of periodic points having arbitrarily large period. Hence $\Lambda^* \subset \Lambda$.

To show the reverse containment, suppose that $p \in \Lambda$. There exists a sequence of points $q_n \to p$ with the following property $f^1(q_n), ..., f^n(q_n)$ are all defined and distinct. Since Δ is dense, we can take a new sequence $\{q'_n\}$ of periodic points converging to p, and we can make $|q_n - q'_n|$ as small as we like. Making these distances sufficiently small, we guarantee that $f^1(q'_n), ..., f^n(q'_n)$ are all distinct. This means that q'_n has period more than n.

Since X is compact, there are only finitely many periodic tiles having diameter greater than ϵ, for any $\epsilon > 0$. Hence, the size of the periodic tile containing q'_n necessarily converges to 0. But then every neighborhood of p intersects infinitely many periodic tiles. □

LEMMA 2.7. *When Δ is dense, Λ is closed.*

Proof: When Δ is dense, we have simply $\Lambda = X - \Delta$. Since Δ is open and X is closed, Λ is also closed. □

Remark: Λ is closed even when Δ is not dense. We leave this an an exercise for the interested reader.

2.6. Some Hyperbolic Geometry

This material is not used until Part 2 of the monograph. More details about hyperbolic geometry can be found in [**B**], [**BKS**], and [**S1**, §10,12].

The Mobius Group: $SL_2(\boldsymbol{C})$ is the group of 2×2 complex matrices of determinant 1. The matrix

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{2.9}$$

Acts on the Riemann sphere $\boldsymbol{C} \cup \infty$ by linear fractional (or Mobius) transformations

$$T_M(z) = \frac{az+b}{cz+d}. \tag{2.10}$$

This group action is compatible with matrix multiplication:

$$T_A \circ T_B = T_{AB}. \tag{2.11}$$

The two maps T_M and T_{-M} have the same action, and it is customary to work with the group $PSL_2(\boldsymbol{C}) = SL_2(\boldsymbol{C})/\pm I$. This group is often called the Mobius group. We will sometimes confound an element of $PSL_2(\boldsymbol{C})$ with the linear fractional transformation it determines.

Generalized Circles: A *generalized circle* in $\boldsymbol{C} \cup \infty$ is either a round circle or a topological circle of the form $L \cup \infty$ where $L \subset \boldsymbol{C}$ is a straight line. Mobius transformations map generalized circles to generalized circles, and preserve angles between them. See e.g. [**S1**, Theorem 10.1].

The subgroup $PSL_2(\boldsymbol{R})$ consists of those equivalence classes of real matrices. Elements of this subgroup preserve $\boldsymbol{R}\cup\infty$ and both the upper and lower half-planes. In particular, elements of $PSL_2(\boldsymbol{R})$ preserve the set of generalized circles which are either vertical lines or circles having the real axis as a diameter.

Hyperbolic Plane: The *hyperbolic plane* is the upper half plane in $\boldsymbol{C}$. We denote it by $\boldsymbol{H}^2$. We equip $\boldsymbol{H}^2$ with the Riemannian metric

$$\langle v, w\rangle_{x+iy} = \frac{v\cdot w}{y^2}. \tag{2.12}$$

The group $PSL_2(\boldsymbol{R})$ acts on $\boldsymbol{H}^2$ by Mobius transformations.

LEMMA 2.8. *$PSL_2(\boldsymbol{R})$ acts isometrically on $\boldsymbol{H}^2$.*

Proof: The claim is fairly obvious for the maps $z \to az + b$, and a direct calculation shows that the element $z \to -1/z$ also has this property. All other elements of $PSL_2(\boldsymbol{R})$, interpreted as linear fractional transformations, are compositions of the maps just mentioned. Compare [**S1**, §10.5]. $\square$

The isometry group of $\boldsymbol{H}^2$ is generated by $PSL_2(\boldsymbol{R})$ and by (say) reflection in a vertical line. The *geodesics* (i.e. length minimizing paths) in $\boldsymbol{H}^2$ are either vertical rays or semicircles which meet $\boldsymbol{R}$ at right angles. We have already mentioned that $PSL_2(\boldsymbol{R})$ permutes these geodesics.

Ideal Triangles: Each geodesic in $\boldsymbol{H}^2$ has two endpoints in $\boldsymbol{R}\cup\infty$. We say that two geodesics are *asymptotic* if they have a common endpoint. For instance, the vertical geodesics $y = 0$ and $y = 1$ share ∞ as an endpoint. We say that an *ideal triangle* is the closed region bounded by three pairwise asymptotic geodesics. Each ideal triangle has 3 *ideal vertices*. One can find an element of $PSL_2(\boldsymbol{R})$ which maps any 3 distinct points in $\boldsymbol{R}\cup\infty$ to any other 3 distinct points in $\boldsymbol{R}\cup\infty$. For this reason, all ideal triangles are isometric. Ideal triangles are noncompact but have finite area equal to π.

Discrete Groups and Hyperbolic Surfaces: For us, a *hyperbolic surface* is a geodesically complete Riemannian surface that is locally isometric to $\boldsymbol{H}^2$. Hyperbolic surfaces are closely related to certain subgroup of $PSL_2(\boldsymbol{R})$, as we now explain.

A subgroup $\Gamma \subset PSL_2(\boldsymbol{R})$ is *discrete* if the identity element of Γ is not an accumulation point of Γ. An equivalent definition is that every convergent sequence in Γ is eventually constant. Yet another equivalent definition is that for any compact set $K \subset \boldsymbol{H}^2$, the set

$$\{\gamma\in\Gamma|\ \gamma(K)\cap K\neq\emptyset\} \tag{2.13}$$

is finite.

The group Γ is said to *act freely* if it never happens that $\gamma(p) = p$ for some $\gamma\in\Gamma$ and some $p\in\boldsymbol{H}^2$. When Γ is discrete and acts freely, the quotient $\boldsymbol{H}^2/\Gamma$ is a hyperbolic surface. Conversely, every hyperbolic surface arises this way. The proof just amounts to showing that the universal cover of the surface is $\boldsymbol{H}^2$ and that the deck group Γ is discrete and acts freely on $\boldsymbol{H}^2$.

2.7. Continued Fractions

Here we give a rapid introduction to the theory of continued fractions. See [**Da**] for a more complete exposition.

Gauss Map: The *Gauss map* is the map $G : (0,1) \to [0,1)$ defined by

$$G(x) = \frac{1}{x} - \text{floor}\Big(\frac{1}{x}\Big). \tag{2.14}$$

We have $G(1/n) = 0$ when $n > 1$ is an integer. This map has an invariant measure, the *Gauss measure* $\mu = (1+x)^{-1}dx$. Here, the invariance means that $\mu(G^{-1}(S)) = \mu(S)$ for all measurable sets $S \subset [0,1]$. See [**BKS**, §1,4,5].

Continued Fractions: Given some $s \in (0,1)$ let $s_n = g^n(s)$. There are integers $a_1, ...$ so that

$$a_n = \text{floor}\Big(\frac{1}{s_{n-1}}\Big), \qquad n = 1,2,3,... \tag{2.15}$$

The sequence $(0, a_1, a_2, a_3, ...)$ is called the continued fraction expansion of s. When $s > 1$ we set $a_0 = \text{floor}(s)$ and then define $s = (a_0, a_1, a_2, ...)$ where $(0, a_1, a_2, ...)$ is the continued fraction expansion for $s - a_0$. The auxilliary sequence

$$a_0 + \frac{1}{a_1}, \qquad a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2}}, \qquad a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3}}}, \qquad \ldots \tag{2.16}$$

converges to s.

The Modular Group: The *Modular group* $PSL_2(\boldsymbol{Z}) \subset PSL_2(\boldsymbol{R})$ is the subgroup consisting of integer matrices. As we mentioned above, this group acts on $\boldsymbol{R} \cup \infty$ by linear fractional transformations The modular group is closely related to continued fractions. Suppose $s, s' \in (0,1)$ respectively have continued fractions $(0, n_0, n_1, ...)$ and $(0, n'_0, n'_1, ...)$. Then s and s' are in the same orbit of $SL_2(\boldsymbol{Z})$ if and only if $n_{s+k} = n'_{t+k}$ for some integers $s, t \geq 0$ of the same parity. This result is readily derivable from [**BKS**, Theorem 5.16], which treats the case when $s, s' > 1$. We will not need this result here, but it motivates a similar result involving our renormalization map R.

Recurrence formula: Suppose that p/q has continued fraction expansion

$$(a_0, a_1, ..., a_n).$$

We introduce numbers $p_{-2}, p_{-1}, p_0, ..., p_n$ and $q_{-2}, q_{-1}, q_0, ..., q_n$ such that

$$p_{-2} = 0, \quad p_{-1} = 1, \quad q_{-2} = 1, \quad q_{-1} = 0. \tag{2.17}$$

We then define the recurrence relation

$$p_k = a_k p_{k-1} + p_{k-2}, \qquad q_k = a_k q_{k-1} + q_{k-2}, \qquad k = 0, ..., n, \tag{2.18}$$

This recurrence relation gives $p = p_n$ and $q = q_n$. One shows inductively that

$$|p_k q_{k+1} - q_k p_{k+1}| = 1, \qquad \forall k. \tag{2.19}$$

This equation in turn implies that

$$\left|\frac{p_k}{q_k} - \frac{p_{k+1}}{q_{k+1}}\right| \leq \frac{1}{q_k^2}. \tag{2.20}$$

This inequality comes from the fact that $q_k \leq q_{k+1}$, as is easily seen from the recurrence relation.

Signed Continued Fractions: We define a *signed continued fraction* exactly like an ordinary continued fraction, except that we allow

$$a_k \in \boldsymbol{Z} - \{0, -1\}. \tag{2.21}$$

Thus, for instance, the S.C.F. $(0, 4, -3, 5, -7)$ means

$$\cfrac{1}{4 + \cfrac{1}{-3 + \cfrac{1}{5 + \cfrac{1}{-7}}}}$$

The recurrence relation described above works for S.C.F.s just as it does for ordinary continued fractions, and so do Equations 2.19 and 2.20. To get the basic fact $|q_k| \leq |q_{k+1}|$ we need to use the fact that $a_{k+1} \neq -1$.

2.8. Some Analysis

Hausdorff Convergence: Given a metric space M and two compact sets $S_1, S_2 \subset M$, one defines the *Hausdorff distance* $d(S_1, S_2)$ to be the infimal ϵ such that S_j is contained in the ϵ-neighborhood of S_{3-j} for $j = 1, 2$. By compactness, the infimum is actually realized. This puts a metric on the space of compact subsets of a metric space. We say that a sequence $\{S_n\}$ of closed (but not necessarily compact) subsets of M converges to S if, for every compact set K, we have $d(S_n \cap K, S \cap K) \to 0$ as $n \to \infty$. We call this the *Hausdorff topology* on the set of closed subsets of a metric space.

Hausdorff Dimension: In this section, we review some basic properties of the Hausdorff dimension. See [**F**] for more details.

Let M be a metric space. We let $|J|$ denote the diameter of a bounded subset $J \subset M$. Given a bounded subset $S \subset M$, and $s, \delta > 0$, we define

$$\mu(S, s, \delta) = \inf \sum |J_n|^s. \tag{2.22}$$

The infimum is taken over all countable covers of S by subsets $\{J_n\}$ such that $\mathrm{diam}(J_n) < \delta$. Next, we define

$$\mu(S, s) = \lim_{\delta \to 0} \mu(S, s, \delta) \in [0, \infty]\,. \tag{2.23}$$

This limit exists because $\mu(S, s, \delta)$ is a monotone function of δ. Note that $\mu(S, n) < \infty$ when $M = \boldsymbol{R}^n$. Usually we'll work in $\boldsymbol{R}^2$. Finally,

$$\dim(S) = \inf\{s |\ \mu(S, s) < \infty\}. \tag{2.24}$$

The number $\dim(S)$ is called the *Hausdorff dimension* of S.

We will mainly be concerned with upper bounds on Hausforff dimension.

LEMMA 2.9. *Suppose that there are infinitely many integers* $m > 0$ *such that* S *has a cover by at most* m^D *sets, all of which have diameter at most* C/m*. Here* C *is some constant that does not depend on* m*. Then* $\dim(S) \leq D$*.*

Proof: Choose and $s > D$. The existence of our covers tells us that $\mu(S, s, \delta) \to 0$ as $\delta \to 0$. Hence $\mu(X, s) = 0$. But then $\dim(S) \leq D$ by definition. $\square$