# UNDERGRADUATE REVIEW TOPICS - 2019 HOUSTON SUMMER SCHOOL IN DYNAMICAL SYSTEMS

VAUGHN CLIMENHAGA

## Contents

*Date*: May 23, 2019.

## Overview

Section 1 lists some concepts that are typically covered in undergraduate classes: I expect that you have seen many of these ideas before, even if you do not have complete mastery of all of them. We will spend most of our time in the prep sessions going through Sections 2–6, which list some of the main concepts and examples that will appear in the lectures during the summer school. I encourage you to work through as many of the exercises as you are able to, and to ask questions if you get stuck on a particular exercise, or if you want more explanation of one of the concepts that is mentioned.

## 1. Quick review of basic concepts

### 1.1. **Linear algebra.**

1.1.1. *Vector spaces, inner products, and norms.* Let $F$ be either $\mathbb{R}$ or $\mathbb{C}$. A *vector space over $F$* is a set $V$ equipped with an addition operation $V \times V \to V$ and scalar multiplication $F \times V \to V$ such that: addition is associative, commutative, has an identity, and every element has an inverse; scalar multiplication is compatible with multiplication in $F$, multiplication by 1 fixes every $v \in V$, and the distributive laws hold.[1]

An *inner product* on a vector space $V$ is a map $\langle \cdot, \cdot \rangle \colon V \times V \to F$ such that

(1) $\langle v, w\rangle = \overline{\langle w, v\rangle}$ for all $v, w \in V$ (conjugate symmetry);
(2) $\langle av + w, u\rangle = a\langle v, u\rangle + \langle w, u\rangle$ for all $u, v, w \in V$ and $a \in F$ (linearity); this also implies conjugate linearity in the second argument;
(3) $\langle v, v\rangle \geq 0$ for all $v \in V$, with equality if and only if $v = 0$ (positive definiteness).

The standard inner product on $\mathbb{C}^n$ is $\langle v, w\rangle = \sum_{j=1}^n v_j \overline{w_j}$. A *norm* on a vector space $V$ is a map $\|\cdot\| \colon V \to \mathbb{R}$ such that

(1) $\|v\| \geq 0$ for all $v \in V$, with equality if and only if $v = 0$;
(2) $\|\lambda v\| = |\lambda| \|v\|$ for all $v \in V$ and scalars $\lambda$;
(3) $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$.

Given an inner product, $\|v\| = \sqrt{\langle v, v\rangle}$ defines a norm that satisfies the *parallelogram law* $2\|v\|^2 + 2\|w\|^2 = \|v + w\|^2 + \|v - w\|^2$. There are many norms that are not induced by inner products. Given $1 \leq p \leq \infty$, the $\ell^p$-norm on $\mathbb{C}^n$ is given by

$$\|v\|_p = \Big(\sum_{j=1}^n |v_j|^p\Big)^{1/p} \text{ if } p < \infty, \qquad \|v\|_\infty = \max_{1\leq j\leq n} |v_j|. \tag{1.1}$$

These norms are all *equivalent* in the following sense: for every $n \in \mathbb{N}$ there is a constant $C > 0$ such that for every $1 \leq p, q \leq \infty$ and $v \in \mathbb{C}^n$ we have

$$C^{-1}\|v\|_q \leq \|v\|_p \leq C\|v\|_q.$$

The norm $\|\cdot\|_p$ is induced by an inner product if and only if $p = 2$.

1.1.2. *Matrices and linear transformations up through Jordan normal form.* Let $\mathbb{M}(n, \mathbb{C})$ denote the space of $n \times n$ matrices with complex-valued entries. A matrix $L \in \mathbb{M}(n, \mathbb{C})$ defines a linear transformation on $\mathbb{C}^n$ by $x \mapsto Lx$, and we will usually identify a matrix and its linear transformation without further comment; if the entries are real-valued then the linear transformation acts on $\mathbb{R}^n$. An *eigenvalue* of $L$ is a complex number $\lambda$ such that $\lambda I - L$ is not invertible; in other words, there is an *eigenvector* $v \in \mathbb{C}^n$ for which $Lv = \lambda v$.

[1] The same definition holds if $F$ is a more general field, but $\mathbb{R}$ and $\mathbb{C}$ will suffice for our purposes.

The *spectrum* of $L$ is the set of eigenvalues, often written $\sigma(L)$; this is a finite subset of $\mathbb{C}$. The *spectral radius* $r(L) := \max\{|v| : v \in \sigma(L)\}$ can be determined by *Gelfand's formula*:

$$r(L) = \lim_{n\to\infty} \|L^n\|^{1/n}, \text{ where } \|L\| := \sup\{\|Lv\| : \|v\| = 1\}.$$

A matrix $L$ is *diagonalizable* if $\mathbb{C}^n$ has a basis of eigenvectors for $L$; in this case there is an invertible matrix $C \in \mathbb{M}(n,\mathbb{C})$ such that $D = CLC^{-1}$ is a diagonal matrix. This allows for very efficient computation of powers of $L$ since $L^k = C^{-1}D^kC$ for all $k \in \mathbb{Z}$, and powers of a diagonal matrix are easy to compute: if $D = \operatorname{diag}(\lambda_1,\dots,\lambda_n)$, then $D^k = \operatorname{diag}(\lambda_1^k,\dots,\lambda_n^k)$.

Not every matrix is diagonalizable; consider $L = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$. This matrix has $L^2 = 0$; a matrix with $L^k = 0$ for some $k$ is called *nilpotent*. If $L$ is nilpotent then there is an invertible matrix $C$ such that $N = CLC^{-1}$ is strictly upper triangular, meaning that $N_{ij} = 0$ whenever $i \geq j$.

In general, if $L \in \mathbb{M}(n,\mathbb{C})$ has $\sigma(L) = \{\lambda_1,\dots,\lambda_m\}$, then the characteristic polynomial $\det(tI - L)$ factors as $\prod_{j=1}^m (t-\lambda_j)^{n_j}$ for some $n_j \in \mathbb{N}$, which are the *algebraic multiplicities* of the eigenvalues $\lambda_j$. The *geometric multiplicity* of $\lambda_j$ is the dimension of the eigenspace $E_j := \ker(\lambda_j I - L)$; that is, the number of linearly independent eigenvectors for $\lambda_j$. The matrix $L$ is diagonalizable if and only if these multiplicities agree for all eigenvalues. In this case we have $\mathbb{C}^n = E_1 \oplus E_2 \oplus \cdots \oplus E_m$, and each eigenspace $E_j$ is $L$-invariant.

For non-diagonalizable matrices, we can get a decomposition along these lines by letting

$$E_j := \bigcup_{k\geq 1} \ker(\lambda_j I - L)^k = \{v \in \mathbb{C}^n : (\lambda_j I - L)^k v = 0 \text{ for some } k \geq 1\}$$

be the *generalized eigenspace*; then we once again have $\mathbb{C}^n = \oplus_{j=1}^m E_j$, and each $E_j$ is $L$-invariant. Moreover, each $E_j$ has a basis $v_1,\dots,v_{\dim E_j}$ with the property that every $v_i$ has either $Lv_i = \lambda v_i$ (so $v_i$ is an eigenvalue) or $Lv_i = \lambda v_i + v_{\ell(i)}$, where $\ell(i) \in \{1,\dots,\dim E_j\}$ and the map $\ell$ is 1-1. The matrix of the linear transformation $L$ relative to the basis given by the union of all the $v_i$'s is the *Jordan normal form* of $L$.

The *trace* of a matrix $A$ is $\operatorname{Tr} A = \sum_{j=1}^n A_{jj}$, and is equal to the sum of the eigenvalues of $A$, counted with their algebraic multiplicities. The *determinant* of $A$ is the product of the eigenvalues, and its absolute value can be interpreted as the amount by which multiplication by $A$ expands $n$-dimensional volume.

### 1.2. **Real analysis.**

A *metric space* is a set $X$ together with a *metric* $d$, which is a function $d\colon X \times X \to \mathbb{R}$ satisfying the following properties:

(1) $d(x,y) = d(y,x)$ for all $x,y \in X$;
(2) $d(x,y) \geq 0$ for all $x,y \in X$, with equality if and only if $x = y$;
(3) $d(x,y) \leq d(x,z) + d(z,y)$ for all $x,y,z \in X$ (triangle inequality).

A set $U$ in a metric space is *open* if for all $x \in U$ there is $\varepsilon > 0$ such that the ball of radius $\varepsilon$ given by $B(x,\varepsilon) := \{y \in X : d(x,y) < \varepsilon\}$ is contained in $U$. A set is *closed* if its complement is open. It is possible for a set to be both open and closed; it is also possible for a set to be neither open nor closed. A set is *compact* if every open cover has a finite subcover; that is $K \subset X$ is compact if for every collection of open sets $\{U_\alpha\}_{\alpha\in A}$ with $K \subset \bigcup_{\alpha\in A} U_\alpha$, there is a finite subcollection $U_{\alpha_1},\dots,U_{\alpha_m}$ such that $K \subset \bigcup_{i=1}^m U_{\alpha_i}$.

A sequence $x_n \in X$ converges to $x \in X$ if $d(x_n,x) \to 0$; equivalently, if for every open set $U \subset X$ that contains $x$ (a *neighborhood* of $x$) there is $N \in \mathbb{N}$ such that for all $n \geq N$ we have $x_n \in U$. A set $A \subset X$ is closed if and only if every sequence $x_n \in A$ with $x_n \to x \in X$ has $x \in A$. A set $K \subset X$ is compact if and only if every sequence $x_n \in K$ has a convergent subsequences. (These last two statements are true for metric spaces but may fail in more general topological spaces.)

If $V$ is a vector space with a norm $\|\cdot\|$, then $V$ is a metric space with metric given by $d(x,y)=\|x-y\|$. If $d_1,d_2$ are two metrics on $V$ coming from two equivalent norms $\|\cdot\|_1,\|\cdot\|_2$, then $d_1$ and $d_2$ induce the same *topology*: a sequence $x_k\in V$ converges to $x\in V$ w.r.t. $d_1$ if and only if it converges to $x$ w.r.t. $d_2$. Warning: in finite dimensions all norms are equivalent, so they all give the same topology, but in infinite dimensions different norms can induce different topologies.

If $X,Y$ are metric spaces, a map $f\colon X\to Y$ is *continuous* if for every open set $U\subset Y$ the preimage $f^{-1}(U)\subset X$ is open. For metric spaces this is equivalent to the condition that $f(x_n)\to f(x)$ in $Y$ whenever $x_n\to x$ in $X$. The map $f$ is a *homeomorphism* if it a bijection such that $f$ and $f^{-1}$ are both continuous. It is an *isometry* if $d(f(x),f(y))=d(x,y)$ for all $x,y\in X$. Isometric bijections are homeomorphisms but not vice versa.

Another useful example is the set $\Sigma=\{0,1\}^{\mathbb{N}}=\{x_1x_2x_3\cdots : x_k\in\{0,1\}\ \forall k\in\mathbb{N}\}$ of all one-sided infinite binary sequences, equipped with the *symbolic metric*

$$d(x,y)=e^{-\min\{k\in\mathbb{N}:x_k\neq y_k\}},$$

in which $x$ and $y$ are close together if they agree on a long initial segment. The space $\Sigma$ is homeomorphic to the middle-third Cantor set $C\subset[0,1]$ via the map $h\colon\Sigma\to[0,1]$ defined by $h(x)=\sum_{k=1}^\infty 2x_k3^{-k}$. The Cantor set $C$ can be characterized as the set of points in the unit interval $[0,1]$ that have a base-3 expansion in which the digit 1 never appears.

1.3. **Abstract algebra.** Given $n\in\mathbb{N}$, the *symmetric group on $n$ symbols* is the set $S_n$ of permutations (bijections) $\sigma\colon\{1,\dots,n\}\to\{1,\dots,n\}$ together with the binary operation of composition: if $\sigma,\tau\in S_n$ are permutations, then so is $\sigma\circ\tau$. More generally, a *group* is a set $G$ together with a binary operation $\cdot\colon G\times G\to G$, usually written $g\cdot h$ or just $gh$, such that the following axioms hold.

(1) *Associativity*: $g(hk)=(gh)k$ for all $g,h,k\in G$.
(2) *Identity*: There is $e\in G$ such that $eg=ge=g$ for all $g\in G$.
(3) *Inverses*: For every $g\in G$ there is $g^{-1}\in G$ such that $gg^{-1}=g^{-1}g=e$.

A group is *abelian* if $gh=hg$ for all $g,h\in G$. In this case we often write the binary operation as addition. Every vector space (in particular $\mathbb{R}^n$ and $\mathbb{C}^n$) is an abelian group under addition.

Given two groups $G,H$, a map $\varphi\colon G\to H$ is a *homomorphism* if $\varphi(gh)=\varphi(g)\varphi(h)$ for all $g,h\in G$. The *kernel* of a homomorphism is the preimage of the identity element: $\ker\varphi=\{g\in G:\varphi(g)=e_H\}$. The kernel is always a subgroup of $G$. A homomorphism is injective if and only if its kernel is trivial. A bijective homomorphism is an *isomorphism.*

A *subgroup* of a group $G$ is a subset $H\subset G$ that is closed under multiplication and inversion. Equivalently, $H\subset G$ is a subgroup if and only if $gh^{-1}\in H$ for every $g,h\in H$. A *left coset* of a subgroup is a set of the form $gH=\{gh:h\in H\}\subset G$, and a *right coset* is a set of the form $Hg$. If $H$ is the kernel of a homomorphism, then $H$ has the property that every left coset is also a right coset, and vice versa. Equivalently, $gHg^{-1}=H$ for all $g\in G$, and in this case we say that $H$ is *normal.* If $H$ is a normal subgroup of $G$ then the set of left cosets (or the set of right cosets) is a group in its own right, denoted $G/H$, and $H$ is the kernel of the canonical homomorphism $G\mapsto G/H$ given by $g\mapsto gH$.

Most of the groups we are interested in can be described as *matrix groups.* The set $\mathbb{M}(n,\mathbb{C})$ comes equipped with the binary operation of matrix multiplication: $(AB)_{ij}=\sum_{k=1}^n A_{ik}B_{kj}$. This is associative and there is an identity element, but not all matrices have an inverse. The set of invertible matrices $GL(n,\mathbb{C})=\{A\in\mathbb{M}(n,\mathbb{C}):\det A\neq 0\}$ is a group under matrix multiplication, called the *general linear group.*

The set $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ is a group under multiplication, and the map $\det\colon GL(n, \mathbb{C}) \to \mathbb{C}^*$ is a homomorphism since $\det(AB) = \det(A)\det(B)$. The kernel of this homomorphism is the *special linear group* $SL(n, \mathbb{C}) = \{A \in \mathbb{M}(n, \mathbb{C}) : \det A = 1\}$. We can also consider the special linear group over the reals: $SL(n, \mathbb{R}) = \{A \in \mathbb{M}(n, \mathbb{R}) : \det A = 1\}$.

By Cramer's formula for $A^{-1}$, if $A$ has integer entries and $\det A = 1$, then $A^{-1}$ has integer entries as well, so $SL(n, \mathbb{Z}) = \{A \in \mathbb{M}(n, \mathbb{Z}) : \det A = 1\}$ is a subgroup of $SL(n, \mathbb{R})$.

A *ring* is an abelian group with a second binary operation satisfying certain axioms that mimic those satisfied by addition and multiplication of real numbers; usually the abelian group operation is written as addition, and the second binary operation is written as multiplication, and the two are required to satisfy a distributive law. Elements must have additive inverses but need not have multiplicative inverses. Examples of rings include $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, but $\mathbb{N}$ is not a ring since its elements do not have additive inverses. A more sophisticated example of a ring is $\mathbb{R}[x]$, the ring of polynomials in one variable with real coefficients.

A *field* is a ring in which every nonzero element has a multiplicative inverse. Examples of fields include $\mathbb{Q}, \mathbb{R}, \mathbb{C}$, but not $\mathbb{Z}$. Another example of a field is $\mathbb{R}(x)$, the field of rational functions in one variable with real coefficients.

## 2. A crash course in smooth manifolds and hyperbolic geometry

Informally, a *smooth manifold* is something that locally looks like $\mathbb{R}^n$. Instead of giving a general definition we think about two-dimensional examples: surfaces.

### 2.1. **The sphere.**

The two-dimensional sphere $S^2$ can be thought of concretely as the set of all points $x \in \mathbb{R}^3$ for which $\|x\|_2 = 1$. Every point in the northern hemisphere of the sphere (where $z > 0$) is uniquely determined by its $x$ and $y$ coordinates; the upper hemisphere is the graph of the function $z = \sqrt{x^2 + y^2}$ on the open disc $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$. Thus considering $(x, y)$ on this open disc gives *coordinates* on the northern hemisphere of the disc. Formally, the map $\varphi\colon (x, y, z) \to (x, y)$ that takes a point on the northern hemisphere to its $(x, y)$-coordinates is called a *chart*. There is another chart on the southern hemisphere obtained in the same way. These two charts do not quite cover the whole sphere because the equator is not part of either one. The regions determined by $y < 0$, $y > 0$, $x < 0$, and $x > 0$ give four more hemispheres that admit coordinate charts along similar lines, and these six charts together cover the sphere, yielding an *atlas*.

The abstract definition of a smooth manifold $M$ is given via such charts and atlases; one requires that $M$ can be covered by charts that give local coordinates on an open set in $M$, and that the resulting atlas is *smooth* in the sense that the change-of-coordinates maps, which act on a subset of $\mathbb{R}^n$, have infinitely many derivatives.

**Exercise 2.1.** *Let $S^2 \subset \mathbb{R}^3$ be the unit sphere. The* stereographic projection from the north pole *is the map $\varphi\colon S^2 \setminus \{(0, 0, 1)\} \to \mathbb{R}^2$ given by the condition that the three points $(0, 0, 1)$, $\mathbf{x} = (x, y, z)$, and $(\varphi(\mathbf{x}), -1)$ lie on the same line in $\mathbb{R}^3$; see Figure 1. Let $\psi\colon S^2 \setminus \{(0, 0, -1)\} \to \mathbb{R}^2$ be the* stereographic projection from the south pole *given by interchanging the roles of* 1 *and* $-1$ *in the previous sentence. Write down the* change-of-coordinates map $\psi \circ \varphi^{-1}\colon \mathbb{R}^2 \setminus \{\mathbf{0}\} \to \mathbb{R}^2 \setminus \{\mathbf{0}\}$.

If we think of $\mathbb{R}^2$ in Exercise 2.1 as the *complex plane* $\mathbb{C}$ and then add a *point at infinity*, we obtain the *Riemann sphere*.

### 2.2. **The torus.**

Another two-dimensional smooth manifold is the torus. We can visualize the torus as a surface of revolution in $\mathbb{R}^3$, but there is a different description that is often

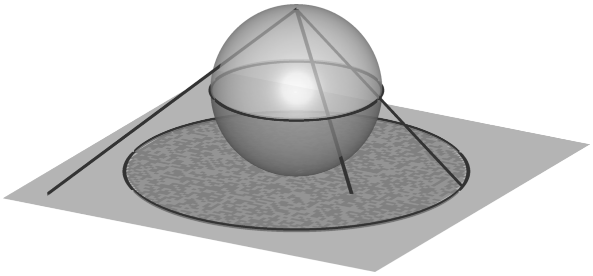

Figure 1. Stereographic projection from the north pole.

more useful. Take a square $[0,1]\times[0,1]$; imagine that it is a sheet of paper (or something even more flexible), and that we identify the left and right edges, gluing them together to obtain a cylinder. The ends of the cylinder are two circles, and if we glue these circles togetherthen we obtain a torus. Thus the torus can be thought of as the square with opposite edges identified.

**Exercise 2.2.** *What kind of surface do we get if we identify opposite edges of a hexagon?*

A different (but related) way of thinking of the torus comes from taking the *quotient space* of $\mathbb{R}^2$ by a certain *equivalence relation.* Given $x,y\in\mathbb{R}^2$, say that $x\equiv y \pmod{\mathbb{Z}^2}$ if $x-y\in\mathbb{Z}^2$; that is, if $x_1-y_1\in\mathbb{Z}$ and $x_2-y_2\in\mathbb{Z}$. The *equivalence class* of $x\in\mathbb{R}^2$ is the set of all $y\in\mathbb{R}^2$ such that $x\equiv y \pmod{\mathbb{Z}^2}$; this is a copy of the integer lattice $\mathbb{Z}^2$ that has been shifted so that it contains $x$. Denote this set by $[x]$ or $x+\mathbb{Z}^2$, and note that $[x]=[y]$ if and only if $x\equiv y \pmod{\mathbb{Z}^2}$. The sets $[x]$ form a partition of $\mathbb{R}^2$. The torus can be identified with the space of all equivalence classes, and it inherits a natural metric from $\mathbb{R}^2$. We often write

$$\mathbb{T}^2=\mathbb{R}^2/\mathbb{Z}^2=\{[x]=x+\mathbb{Z}^2 : x\in\mathbb{R}^2\}, \tag{2.1}$$

$$d([x],[y])=\min\{\|(x+a)-(y+b)\|_2 : a,b\in\mathbb{Z}^2\}=\min\{\|(x-y)+k\|_2 : k\in\mathbb{Z}^2\}. \tag{2.2}$$

Note that each equivalence class $[x]$ intersects the unit square $[0,1]^2$ exactly once if neither $x_1$ nor $x_2$ is an integer; we say that $[0,1]^2$ is a *fundamental domain* for the torus $\mathbb{T}^2$.

The torus is a smooth manifold; its charts arise very naturally since a small neighborhood of $[x]$ is obtained by translating $[x]\subset\mathbb{R}^2$ by some vector $v\in\mathbb{R}^2$ with $\|v\|$ small; thus the change-of-coordinates map between any two overlapping charts is a translation. This construction works in any dimension and we write $\mathbb{T}^n=\mathbb{R}^n/\mathbb{Z}^n$ for the $n$-dimensional torus. Note that when $n=1$ we obtain the circle $S^1=\mathbb{R}/\mathbb{Z}$, which can be viewed either as the unit interval $[0,1]$ with endpoints identified (so $[0,1]$ is a fundamental domain), or as the set of all translations of the set of integers in $\mathbb{R}$.

**Exercise 2.3.** *The map $[x]\mapsto(\cos 2\pi x,\sin 2\pi x)$ gives a homeomorphism from $\mathbb{R}/\mathbb{Z}$ to the unit circle in $\mathbb{R}^2$. Write a similar formula for a homeomorphism from $\mathbb{R}^2/\mathbb{Z}^2$ to the surface of revolution in $\mathbb{R}^3$ obtained by rotating a circle in the $xz$-plane with centre $(R,0)$ and radius $r<R$ around the $z$-axis.*

In this example, $\mathbb{R}^n$ is a *covering space* for $\mathbb{T}^n$. We will not get into the formal definition of covering space; its main utility here is that certain maps on $\mathbb{R}^n$ descend to maps on $\mathbb{T}^n$. For example, the map $F\colon\mathbb{R}\to\mathbb{R}$ given by $F(x)=2x$ has the property that $F(x+n)=2x+2n\equiv 2x \pmod{\mathbb{Z}}$, and thus the map $f\colon S^1\to S^1$ given by $f([x])=[2x]$ is well-defined since $f(y)\in[f(x)]$ whenever $y\in[x]$. This is the *doubling map* on the circle. Another important example is given by taking $L=\left(\begin{smallmatrix}2&1\\1&1\end{smallmatrix}\right)\in SL(2,\mathbb{Z})$ and observing that the map $L\colon\mathbb{R}^2\to\mathbb{R}^2$ has the property that $L(x+\mathbb{Z}^2)=Lx+\mathbb{Z}^2$ since $L$ gives a bijection from $\mathbb{Z}^2$ to itself. Thus $L$ induces a bijection (in fact a homeomorphism) from $\mathbb{T}^2$ to itself by $L[x] = [Lx]$. The same principle works with any $n \in \mathbb{N}$ and $L \in SL(n,\mathbb{Z})$, giving a *toral automorphism* $F_L\colon \mathbb{T}^n \to \mathbb{T}^n$ by $F_L([x]) = [Lx]$. We describe this situation by saying that $SL(n,\mathbb{Z})$ *acts on* $\mathbb{T}^n$ *by toral automorphisms.* We will discuss group actions a little more in §6.3.

**Exercise 2.4.** *Given $L \in SL(n,\mathbb{Z})$, how many periodic points does $F_L\colon \mathbb{T}^n \to \mathbb{T}^n$ have? Recall that a point $[x] \in \mathbb{T}^n$ is* periodic *for $F_L$ if there is $k \in \mathbb{N}$ such that $F_L^k(x) = x$, where $F_L^k$ is the result of composing $F_L$ with itself $k$ times, so $F_L^2 = F_L \circ F_L$, $F_L^3 = F_L \circ F_L \circ F_L$, and so on. If you have difficulty answering this question for a general $L$, start with $L = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, or ask the same question for the doubling map $f\colon S^1 \to S^1$ given by $f([x]) = [2x]$. If you solve the question, think about a more difficult version: given $k$, how many periodic points* of period $k$ *does $F_L$ have?*

2.3. **Tangent spaces.** Suppose $M \subset \mathbb{R}^3$ is a smooth surface. (We didn't define this concept precisely yet, but never mind; if that bothers you, just think of the unit sphere.[2]) Fix a *point* $x \in M$ and a *vector* $v \in \mathbb{R}^3$. Note that we are thinking of $x$ as specifying a location and $v$ as specifying a direction, even though both of them are represented by an ordered triple of real numbers. We say that *$v$ is tangent to $M$ at $x$* if the line through $x$ with direction $v$ (that is, the set $\{x + tv : t \in \mathbb{R}\}$) is tangent to $M$. (Of course, we didn't define what that means yet either; can you write down a good definition?) Equivalently, $v$ is tangent to $M$ at $x$ if there is a curve $\gamma\colon \mathbb{R} \to M \subset \mathbb{R}^3$ such that $\gamma(0) = x$ and $\gamma'(0) = v$. Let $T_xM$ denote the set of all vectors $v$ that are tangent to $M$ at $x$; this is the *tangent space* to $M$ at $x$. We should think of each element of $T_xM$ as carrying two pieces of information: a location on $M$ (the point $x$) together with a direction of motion along $M$ (really, not just a direction but a speed as well).

An abstract smooth manifold (which we still avoid defining) has a similar notion of tangent space at each point. We avoid the precise definition and just say that if $M$ is an $n$-dimensional manifold, then for each $x \in M$, the tangent space $T_xM$ is an $n$-dimensional vector space such that for every smooth curve $\gamma\colon \mathbb{R} \to M$ passing through $x$, the tangent vector to $\gamma$ at $x$ is an element of $T_xM$. Thus a tangent vector $v \in T_xM$ carries two pieces of information: where its footprint is (the point $x$), and a direction (with magnitude) along $M$. In the specific case of the torus $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$, the tangent space can always be identified with $\mathbb{R}^n$ by using the canonical coordinates, and if you have not worked with smooth manifolds before, you should think of this and/or the picture of the sphere in $\mathbb{R}^3$ whenever we discuss tangent spaces here.

It is sometimes useful to talk about the *tangent bundle*, which is the *disjoint* union of all the tangent spaces: $TM = \bigsqcup_{x \in M} T_xM = \{(x,v) : x \in M, v \in T_xM\}$. For $\mathbb{T}^n$, the tangent bundle is $\mathbb{T}^n \times \mathbb{R}^n$, the set of all pairs $(x,v)$, where $x$ specifies a point on $\mathbb{T}^n$ and $v$ specifies a vector based at $x$. For other manifolds, it is not necessarily possible to express the tangent bundle as a direct product.

**Exercise 2.5.** *Convince yourself that $TS^2 \neq S^2 \times \mathbb{R}^2$. (Giving a proper proof of this requires a little machinery, which you may or may not have seen before.)*

2.4. **Riemannian manifolds.** Let $\mathbb{T}^2$ be the torus represented as $\mathbb{R}^2/\mathbb{Z}^2$, and let $M$ be the surface of revolution from Exercise 2.3. That exercise showed that $\mathbb{T}^2$ and $M$ are homeomorphic; they have the same topological properties. However, their metric properties are different: $M$ inherits a natural metric from $\mathbb{R}^3$, while $\mathbb{T}^2$ inherits a metric from $\mathbb{R}^2$ via (2.2).

[2] If you insist: a subset $M \subset \mathbb{R}^3$ is a smooth surface if for every $x \in M$ there is an open set $U \subset \mathbb{R}^3$ containing $x$ and a smooth function $\Phi\colon U \to \mathbb{R}$ such that the gradient of $\Phi$ never vanishes and $M \cap U = \Phi^{-1}(0)$. Equivalently, if for every $x \in M$ there is an open set $U$ such that on $M \cap U$, one of the three coordinates can be written as a smooth function of the other two, so that $M \cap U$ is the graph of this function.

**Exercise 2.6.** *Let $h\colon \mathbb{T}^2 \to M$ be the homeomorphism from Exercise 2.3. Given $a \in \mathbb{R}$, consider the closed curves $\gamma_a = \{[(a,y)] : y \in \mathbb{R}\} \subset \mathbb{T}^2$ and $\eta_a = \{[(x,a)] : x \in \mathbb{R}\} \subset \mathbb{T}^2$. Show that the curves $\gamma_a, \eta_a$ all have the same length on $\mathbb{T}^2$, but that this is not true for the curves $h(\gamma_a), h(\eta_a)$ on $M$.*

Since $\mathbb{T}^2$ is not just locally homeomorphic to $\mathbb{R}^2$, but locally *isometric* to $\mathbb{R}^2$, it is sometimes called the *flat torus.* The torus of revolution $M$, on the other hand, is not flat. We say that $\mathbb{T}^2$ and $M$ are the same as smooth manifolds (they are *diffeomorphic*), but they are different as *Riemannian manifolds.*

Informally, a Riemannian manifold is a smooth manifold in which we are given a little extra information: not only do we have a tangent space at every $x \in M$, but to every $v \in T_xM$ we assign a length $\|v\|$ (then we can define the angle between two vectors by using the law of cosines). Once we know this, we can define the *length* of a curve $\gamma\colon [0,1] \to M$ as $\int_0^1 \|\gamma'(t)\|\,dt$. We will return to this idea below.

2.5. **Surfaces of higher genus and a little hyperbolic geometry.** If one glues together the edges of an octagon in the pattern shown in Figure 2, one obtains the *surface of genus 2* shown there, which we denote by $M$. Can we get this surface via a quotient construction like we did with the torus? It turns out that we can, but first we need a little hyperbolic geometry.

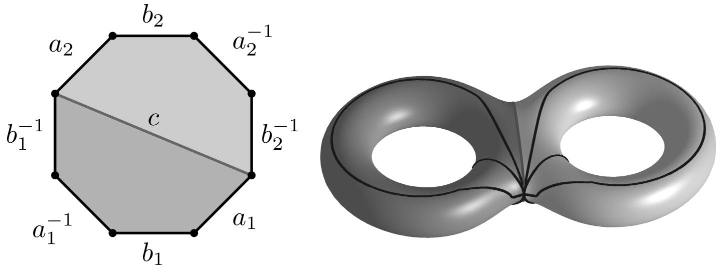

FIGURE 2. Identifying edges of an octagon gives a surface of genus 2.

First we describe one more way to get the torus $\mathbb{T}^2$ from the Euclidean plane $\mathbb{R}^2$. Let $a\colon \mathbb{R}^2 \to \mathbb{R}^2$ be translation by the vector $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$, and let $b\colon \mathbb{R}^2 \to \mathbb{R}^2$ be translation by $\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$. Then $x \equiv y \pmod{\mathbb{Z}^2}$ if and only if we can get from $x$ to $y$ by repeatedly applying $a$, $b$, and their inverses in some order. If we write $F = [0,1]^2$ for the fundamental domain given by the unit square, then the images of $F$ under all iterates of $a$ and $b$ tile the plane. In terms of the planar model given by $[0,1]^2$ with opposite edges identified, we see that $a$ maps the left edge to the right edge, and identifies $x$ with $a(x)$, while $b$ maps the bottom edge to the top edge, and identifies $x$ with $b(x)$.

Nothing quite so simple can work with the octagon, because the angles don't add up. With the edge identifications shown in Figure 2, all the vertices of the octagon become the same point on $M$. This was fine for the torus, because each vertex had an angle of $\frac{\pi}{2}$ in the square, and since there were 4 vertices, the total angle around the resulting point on $\mathbb{T}^2$ was $4 \cdot \frac{\pi}{2} = 2\pi$. But the octagon has 8 vertices, each with an internal angle of $\frac{3\pi}{4}$, so the total angle around this point on the $M$ would be $6\pi$, which is much too big. To resolve this, we need to find a way to draw a regular octagon whose angles are all equal to $\frac{\pi}{4}$. This is impossible in the Euclidean plane, but possible in the hyperbolic plane $\mathbb{H}^2$.

The hyperbolic plane $\mathbb{H}^2$ can be thought of in two ways: in the *upper half-plane model*, it is the set $\{z \in \mathbb{C} : \operatorname{Im} z > 0\}$, and in the *unit disc model*, it is the set $\{z \in \mathbb{C} : |z| < 1\}$. In both cases, we must define a metric on $\mathbb{H}^2$. First recall a (somewhat circuitous) way to define the metric on Euclidean space $\mathbb{R}^2$. Given any smooth path $\gamma\colon [0,1] \to \mathbb{R}^2$, the length of $\gamma$ is $\ell(\gamma) = \int_0^1 \|\gamma'(t)\|\, dt$, where here $\|\cdot\|$ is the usual Euclidean norm. Then the distance between $x, y \in \mathbb{R}^2$ is $\inf\{\ell(\gamma) : \gamma(0) = x \text{ and } \gamma(1) = y\}$. A path achieving this infimum is called a *geodesic*; in Euclidean space, geodesics are just straight lines.

In the upper half-plane model, we can define the length of a curve $\gamma\colon [0,1] \to \mathbb{H}^2$ by

$$\ell(\gamma) = \int_0^1 \|\gamma'(t)\|_H \, dt, \qquad \text{where} \qquad \|\gamma'(t)\|_H := \frac{\|\gamma'(t)\|}{\operatorname{Im}\gamma(t)}, \tag{2.3}$$

and then define distance and geodesics just as above. The real line $\mathbb{R}$ is not part of the upper half-plane; it (together with the point at $\infty$) is called the *ideal boundary*. The equation for $\ell(\gamma)$ shows that the ideal boundary is an infinite distance from every point in $\mathbb{H}^2$. A geodesic in $\mathbb{H}^2$ is either a vertical line or an arc of a circle that intersects the ideal boundary orthogonally. The geodesics in the unit disc model are similar, although in this case we need to use a different formula for the metric.

**Exercise 2.7.** *Find a function $y\colon \mathbb{R} \to (0,\infty)$ such that for each $a \in \mathbb{R}$, the curve $\gamma(t) = a + iy(t)$ has the property that $\|\gamma'(t)\|_H = 1$ for all $t$.*

**Exercise 2.8.** *Define a curve $\gamma\colon (0,\pi) \to \mathbb{H}^2$ by $\gamma(t) = e^{it} = \cos t + i \sin t$. Find a reparametrization $t\colon \mathbb{R} \to (0,\pi)$ such that the curve $\eta(s) = \gamma(t(s))$ has the property that $\|\eta'(s)\|_H = 1$ for all $s \in \mathbb{R}$, hence $\eta$ is a unit speed geodesic.*

**Exercise 2.9.** *Show that the map $\theta(z) = \frac{-z+i}{z+i}$ gives a bijection from the upper half-plane to the unit disc.*

The hyperbolic metric on the unit disc is defined so that the map $\theta$ in Exercise 2.9 is an isometry.

**Exercise 2.10.** *Determine which curves in Figure 3 are geodesics.*

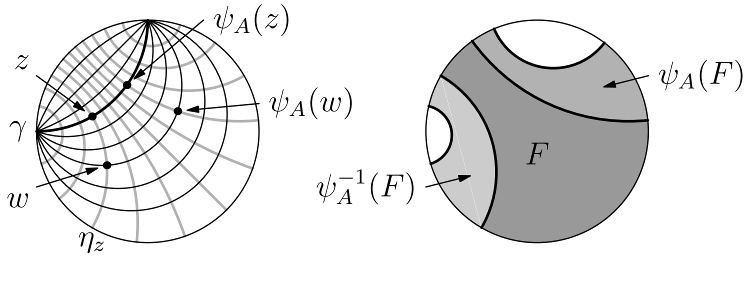

Figure 3. A hyperbolic translation.

One kind of isometry on $\mathbb{H}^2$ is illustrated in Figure 3; the map $\psi_A$ from the disc to itself moves points along $\gamma$ and maps geodesics to geodesics. In fact, a formula for $\psi_A$ can be in terms of a matrix $A \in SL(2,\mathbb{R})$. (We omit the proof that $\psi_A$ is an isometry.)

**Exercise 2.11.** *Given a matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL(2,\mathbb{R})$, define a map $\varphi_A\colon \mathbb{C} \to \mathbb{C}$ by $\varphi_A(z) = \frac{az+b}{cz+d}$; this is called a* fractional linear transformation *(or* Möbius transformation*). Show that $\varphi_A$ is a bijection from the upper half-plane to itself, and that $\ell(\gamma) = \ell(\varphi_A \circ \gamma)$, where $\ell(\gamma)$ is the length of the curve $\gamma$ as in* (2.3). *(We say that $\varphi_A$ is an* isometry.*)*

The map $\psi_A$ can be written as $\psi_A = \theta \circ \varphi_A \circ \theta^{-1}$ for some $A \in SL(2,\mathbb{R})$.

**Exercise 2.12.** *Show that $\varphi_{-A} = \varphi_A$, and that $\varphi_{AB} = \varphi_A \circ \varphi_B$. In particular, show that $A \mapsto \varphi_A$ is a homomorphism from $SL(2,\mathbb{R})$ into the isometry group of $\mathbb{H}^2$, and that the kernel of this homomorphism is $\{\pm I\}$.*

The quotient group $SL(2,\mathbb{R})/\{\pm I\}$ is denoted $PSL(2,\mathbb{R})$ (the *projective special linear group*. The *unit tangent bundle* of $\mathbb{H}^2$, denoted $T^1\mathbb{H}^2$, is the set of pairs $(p,v)$ such that $p$ is a point in $\mathbb{H}^2$, $v$ is a tangent vector at $p$, and $\|v\|_H = 1$ using the definition of hyperbolic length in (2.3). Given $A \in SL(2,\mathbb{R})$ (so $\{\pm A\} \in PSL(2,\mathbb{R})$), the corresponding isometry $\varphi_A \colon \mathbb{H}^2 \to \mathbb{H}^2$ induces an isometry on the unit tangent bundle by

$$D\varphi_A(p,v) = (\varphi_A(p), \varphi_A'(p)v). \tag{2.4}$$

In the last expression we think of $\varphi_A'(p)$ and $v$ as complex numbers. The following exercise gives a bijection between $PSL(2,\mathbb{R})$ and $T^1\mathbb{H}^2$ by identifying $\pm I$ with $(i,1)$, and $\pm A$ with the image of $(i,1)$ under $D\varphi_A$.

**Exercise 2.13.** *Given $(p,v) \in T^1\mathbb{H}^2$, show that there is a unique $\{\pm A\} \in PSL(2,\mathbb{R})$ such that $\varphi_A(i) = p$ and $\varphi_A'(i) = v$.*

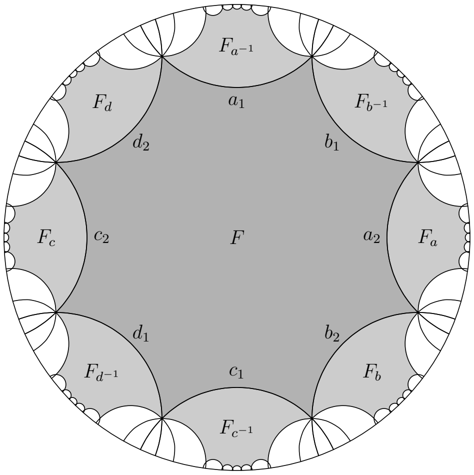

FIGURE 4. A regular octagon in the hyperbolic plane.

Returning to the question of getting the surface of genus 2, consider the disc model and take 8 geodesics that are evenly spaced around the circle, as shown in Figure 4. These geodesics can be chosen so that the octagon they form has angle $\frac{\pi}{4}$ at all 8 vertices. Let $a$ be the hyperbolic isometry that takes the edge labeled $a_1$ into the edge labeled $a_2$, and similarly for $b, c, d$. Then the images of the octagon $F$ under all iterates of $a, b, c, d$ (and their inverses) tile $\mathbb{H}^2$. Say that two points $x, y \in \mathbb{H}^2$ are equivalent if $x$ is mapped into $y$ by the composition of some combination of $a, b, c, d$, and their inverses. The quotient space of $\mathbb{H}^2$ by this equivalence relation gives the surface of genus 2, just as the quotient space of $\mathbb{R}^2$ by the equivalence relation induced by the translations $a, b$ gave the torus.

2.6. **Geodesics and horocycles.** In the previous section we encountered geodesics in the hyperbolic plane as (Euclidean) circles and lines that intersect the ideal boundary orthogonally. Another important class of curves in $\mathbb{H}^2$ is given by *horocycles*, which include (Euclidean) circles that are *tangent* to the ideal boundary.

Horocycles have the following fundamental property. Suppose that $p, q \in \mathbb{H}^2$ lie on the same horocycle, and that $v, w$ are two tangent vectors based at $p, q$, respectively, which $v, w$ have unit length with respect to the hyperbolic norm given in (2.3). Suppose moreover that $v, w$ are orthogonal (perpendicular, normal) to the horocycle and point 'inwards', that is, towards the point at which the horocycle is tangent to the ideal boundary. Let $\gamma_v$ and $\gamma_w$ be the unit speed geodesics with the property that $\gamma_v(0) = p$, $\gamma_v'(0) = v$, $\gamma_w(0) = q$, and $\gamma_w'(0) = w$. Then $\gamma_v(t)$ and $\gamma_w(t)$ approach each other exponentially quickly as $t \to \infty$, in the sense given by the following series of exercises.

**Exercise 2.14.** *Use the result of Exercises 2.8 and 2.13 to write explicit formulas for $\gamma_v$ and $\gamma_w$ by finding $A, B \in SL(2, \mathbb{R})$ such that $D\varphi_A(i, -1) = (p, v)$ and $D\varphi_B(i, -1) = (q, w)$, and then putting $\gamma_v = \varphi_A \circ \eta$ and $\gamma_w = \varphi_B \circ \eta$.*

**Exercise 2.15.** *Use the result of Exercise 2.14 to show that there are $C, \lambda > 0$ such that $d_H(\gamma_v(t), \gamma_w(t)) \leq Ce^{-\lambda t}$ for all $t \geq 0$, where $d_H$ is the hyperbolic distance induced by the length function in* (2.3).

The result of Exercise 2.15 is sometimes summarized as the statement that "the normal vector field to the horocycle is the stable manifold for geodesic flow". Let us explain what the ingredients of this statement mean. The "normal vector field to the horocycle" is the set of pairs $(p, v)$, where $p$ is a point on the horocycle and $v$ is an inward-pointing unit vector at $p$ that is orthogonal to the horocycle. The exercise shows that the geodesics determined by these pairs $(p, v)$ get close together exponentially quickly as $t \to \infty$; this is the "stable manifold" behavior.[3] But what is "geodesic flow"?

As mentioned above, each pair $(p, v) \in T^1\mathbb{H}^2$ determines a unique unit speed geodesic $\gamma$ with the property that $\gamma(0) = p$ and $\gamma'(0) = v$. Given $t \in \mathbb{R}$, let $g_t(p, v) = (\gamma(t), \gamma'(t))$; that is, $g_t \colon T^1\mathbb{H}^2 \to T^1\mathbb{H}^2$ moves each unit tangent vector a distance $t$ along the geodesic that it generates. The family $\{g_t\}_{t\in\mathbb{R}}$ is called the *geodesic flow* on $T^1\mathbb{H}^2$.

**Exercise 2.16.** *Use the result of Exercise 2.13 to show that the map $\pi \colon PSL(2, \mathbb{R}) \to T^1\mathbb{H}^2$ defined by $\pi(A) = D\varphi_A(i, i)$ is a bijection. Show that if $g_t$ is the geodesic flow just defined, then for every $t \in \mathbb{R}$ and $\pm A \in PSL(2, \mathbb{R})$, we have $\pi^{-1} \circ g_t \circ \pi(A) = A\left(\begin{smallmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{smallmatrix}\right)$.*

Exercise 2.16 says that when geodesic flow on $T^1\mathbb{H}^2$ is viewed as a flow on $PSL(2, \mathbb{R})$, it behaves like right multiplication by the diagonal matrix $\left(\begin{smallmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{smallmatrix}\right)$. Another important flow on $T^1\mathbb{H}^2$ is the *stable horocycle flow*, which in terms of $PSL(2, \mathbb{R})$ is given by right multiplication by $\left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right)$.

**Exercise 2.17.** *Show that the horocycle flow corresponds to moving $(p, v)$ to $(q, w)$, where $q$ is a point on the same horocycle as $p$, obtained by moving a distance $t$ along that horocycle, and $w$ is the inward-pointing unit normal vector at $q$.*[4]

Finally, we point out one special horocycle; horizontal lines in the upper half-plane model are also called horocycles, and can be thought of as circles of infinite radius that are tangent to the point at infinity on the ideal boundary. The following mimics Exercise 2.15.

**Exercise 2.18.** *Show that if $a, b \in \mathbb{R}$ and $y(t)$ is as in Exercise 2.7, then there are $C, \lambda > 0$ such that $d_H(a + iy(t), b + iy(t)) \leq Ce^{-\lambda t}$ for all $t \geq 0$.*

[3] If we instead choose the *outward* pointing normal vectors, then this convergence happens as $t \to -\infty$, and we speak of the "unstable manifold".

[4] Again, there is an *unstable horocycle flow* where we use outward-pointing normals instead, and the matrix is lower-triangular instead of upper-triangular.

## 3. A crash course in symbolic dynamics

### 3.1. **General notions.**

Given $p \in \mathbb{N}$, the *full one-sided shift on $p$ symbols* is the metric space

$$\Sigma = \Sigma_p = \{1, \dots, p\}^{\mathbb{N}} = \{x_1 x_2 x_3 \cdots : x_k \in \{1, \dots, p\} \text{ for all } k \in \mathbb{N}\},$$
$$d(x, y) = e^{-\min\{k \in \mathbb{N} : x_k \neq y_k\}},$$

together with the map $\sigma \colon \Sigma \to \Sigma$ given by $\sigma(x_1 x_2 x_3 \cdots) = x_2 x_3 x_4 \cdots$. The full two-sided shift is defined similarly, as $\{1, \dots, p\}^{\mathbb{Z}}$; in this case the metric is defined by taking the minimum of $|k|$ where $x_k \neq y_k$, so that two points are close together if they agree for a long interval of indices on both sides of 0. In these notes we will only consider one-sided shifts, to keep notation simpler.

A *finite word* over the alphabet $A = \{1, \dots, p\}$ is a finite sequence of symbols from $A$. The set of all words is denoted $A^* = \bigcup_{n \geq 0} A^n$. Given a word $w \in A^*$, the *length* of $w$ is the number of symbols in $w$, which we denote $|w|$, and the *cylinder* corresponding to $w$ is

$$[w] = \{x \in \Sigma : x_k = w_k \text{ for all } 1 \leq k \leq |w|\}.$$

That is, $[w]$ is the set of infinite sequences that start with the word $w$. Every cylinder is a ball in the metric $d$, and is both open and closed.

A *subshift* of $\Sigma$ (also called a *shift space*) is a closed subset $X \subset \Sigma$ that is shift-invariant ($\sigma(X) = X$). The *language* $\mathcal{L}(X)$ is the set of finite words that appear in some word in $X$; that is,

$$\mathcal{L}(X) = \{w \in A^* : [w] \cap X \neq \emptyset\}.$$

**Exercise 3.1.** *Show that a set $\mathcal{L} \subset A^*$ is the language of some subshift if and only if it satisfies the following conditions:*

*(1) if $w \in \mathcal{L}$ and $v$ is a subword of $w$, then $v \in \mathcal{L}$;*
*(2) if $w \in \mathcal{L}$, then there exists a symbol $a \in A$ such that $wa \in \mathcal{L}$.*

Given a subshift $X$, write $\mathcal{L}_n$ for the collection of words of length $n$ in the language of $X$.

**Exercise 3.2.** *Let $X$ be the shift space on the alphabet $\{0, 1\}$ defined by the rule that $x \in X$ if and only the symbol 1 never appears twice in a row. Compute $\#\mathcal{L}_n$.*

**Exercise 3.3.** *Prove that every language has $\#\mathcal{L}_{m+n} \leq (\#\mathcal{L}_m)(\#\mathcal{L}_n)$.*

It follows from Exercise 3.3 that the sequence $a_n = \log \#\mathcal{L}_n$ is *subadditive*: it satisfies $a_{m+n} \leq a_m + a_n$ for every $m, n$.

**Exercise 3.4.** *Prove* Fekete's lemma*: if $a_n$ is a subadditive sequence, then $\lim_{n \to \infty} \frac{a_n}{n}$ exists and is equal to $\inf_{n \in \mathbb{N}} \frac{a_n}{n}$.*

We conclude that $\lim_{n \to \infty} \frac{1}{n} \log \#\mathcal{L}_n$ exists; denote this limit by $h(\mathcal{L})$. This is the *topological entropy* of the subshift $X$; it measures the exponential growth rate of the number of words in the language of $X$. The entropy of the full shift on $p$ symbols is $\log p$.

### 3.2. **Markov shifts and positive entropy.**

A *(finite) directed graph* consists of a finite set of vertices, labeled $1, \dots, p$, together with a set of directed edges, which are ordered pairs $(i, j)$, where $i, j$ are both vertices (possibly the same vertex). We will always assume that for any given choice of $i, j$, there is at most one directed edge going from vertex $i$ to vertex $j$. We write $i \to j$ if there is an edge from $i$ to $j$, and $i \not\to j$ if there is not.

The *Markov shift* (or *topological Markov chain*) associated to a directed graph is the subshift $X \subset \Sigma_p$ whose language consists of all words $w = w_1 \cdots w_n$ such that $w_k \to w_{k+1}$

for all $1 \le k < n$. In other words, $X$ consists of all infinite sequences $x = x_1x_2x_3\cdots$ that label a walk along the graph.

Given a directed graph, define a $p \times p$ matrix $T$ of 0s and 1s (the *transition matrix*) by

$$T_{ij} = \begin{cases} 1 & i \to j, \\ 0 & i \not\to j. \end{cases}$$

A word $w$ of length $n$ is in the language of the shift if and only if $T_{w_1w_2}T_{w_2w_3}\cdots T_{w_{n-1}w_n} = 1$.

**Exercise 3.5.** *Let $\mathcal{L}$ be the language of the Markov shift with transition matrix $T$. Given two symbols $i, j \in A = \{1, \dots, p\}$, prove by induction that the total number of words in $\mathcal{L}_n$ that start with $i$ and end with $j$ is $(T^{n-1})_{ij}$.*

It follows from Exercise 3.5 that for a Markov shift with transition matrix $T$, we have

$$\#\mathcal{L}_n = \sum_{i,j=1}^{p} (T^{n-1})_{ij}. \tag{3.1}$$

Say that a transition matrix $T$ is *irreducible* if for every $i, j \in A$ there is some $n \in \mathbb{N}$ such that $(T^n)_{ij} > 0$; equivalently, given any two vertices $i, j$ on the graph, there is a path that goes from $i$ to $j$ (but may take many steps to do so). Say that $T$ is *primitive* if there is a single value of $n$ that works for all $i$ and $j$.

**Theorem 3.6** (Perron–Frobenius)**.** *If $T$ is primitive then*

*(1) it has a positive real eigenvalue $\lambda$ such that every other eigenvalue $\mu$ has $|\mu| < \lambda$;*
*(2) $\lambda$ is a simple eigenvalue (it has geometric and algebraic multiplicity 1);*
*(3) $T$ has a unique (up to a scalar) eigenvector $v$ in the* positive cone *$\{v \in \mathbb{R}^p : v_i > 0$ for all $1 \le i \le p\}$;*
*(4) given every $w \in \mathbb{R}^p$, the sequence $(T^n w)/\lambda^n$ converges to a multiple of $v$.*

This theorem will be proved during the main lecture series at the summer school. The main idea is to observe that $T$ maps the positive cone inside itself and show that the intersection of all the forward images of this cone is a line, which contains the PF eigenvector $v$.

Once the first three parts are shown, it is not hard to show that the convergence in the fourth part is exponential: choosing $\xi < 1$ such that $|\mu| < \lambda\xi$ for every eigenvalue $\mu \ne \lambda$, for every $w \in \mathbb{R}^p$ there are $C > 0$ and $a \in \mathbb{R}$ such that $\|(T^n w)\lambda^{-n} - av\| \le C\xi^n$ for all $n \in \mathbb{N}$.

**Exercise 3.7.** *Use the Perron–Frobenius theorem and* (3.1) *to prove that if $\mathcal{L}$ is the language of a Markov shift with transition matrix $T$, then $h(\mathcal{L}) = \log \lambda$, where $\lambda$ is the Perron–Frobenius eigenvalue of $T$.*

3.3. **Zero entropy shifts.** Markov shifts have positive topological entropy and hence their languages grow exponentially quickly. At the other extreme are shift spaces where the language grows slowly.

**Exercise 3.8.** *Show that if $X$ is a shift space with infinitely many points, then $\#\mathcal{L}_n(X) \ge n + 1$ for all $n \in \mathbb{N}$.*

A shift space for which $\#\mathcal{L}_n = n + 1$ for all $n$ is called a *Sturmian shift*, and a sequence $x \in \Sigma$ is a *Sturmian sequence* if $X := \overline{\{\sigma^n x : n \in \mathbb{N}\}}$ is a Sturmian shift. Clearly every Sturmian shift has zero topological entropy.

It may not be immediately obvious that any Sturmian sequences exist. One method for producing a Sturmian sequence is to fix an irrational number $\alpha > 0$ and consider the line

$y = \alpha x + \beta$ for some $\beta \in \mathbb{R}$; if we start at $x = 0$ and move to the right along this line, writing down the symbol 0 every time $x$ passes through an integer value and 1 every time $y$ passes through an integer value, then we obtain the *cutting sequence* associated to $\alpha, \beta$. A binary sequence is Sturmian if and only if it is the cutting sequence associated to some $\alpha, \beta$ with $\alpha$ irrational.

An equivalent way to describe this is to consider the map $R_\alpha \colon S^1 \to S^1$ given by rotation by $\alpha$; viewing $S^1$ as the interval $[0, 1]$ with endpoints identified, we can write $R_\alpha(x) = x + \alpha \pmod 1$. Given $x \in [0, 1]$, we can *code* the trajectory of $x$ according to the partition $[0, 1) = [0, \alpha) \sqcup [\alpha, 1)$ by writing the sequence $h(x) \in \Sigma$ given by

$$h(x)_n = \begin{cases} 1 & f^n x \in [0, \alpha), \\ 0 & f^n x \in [\alpha, 1). \end{cases}$$

Then a sequence is Sturmian if and only if it is the coding of the trajectory of some point under some irrational rotation according to this partition. *Aside:* This procedure for coding trajectories of a system in terms of their itineraries relative to some predetermined partition is an extremely useful one for many classes of systems.

One important example of a Sturmian sequence is the *Fibonacci word*, which can be characterized as the cutting sequence of a line of slope $1/\phi$, where $\phi$ is the golden ratio. Alternately, the Fibonacci word is the infinite sequence obtained by starting with the word 0 and iteratively performing the following substitutions: at every step, replace each 0 with 01, and each 1 with 0. Thus we obtain

$$0 \mapsto 01 \mapsto 010 \mapsto 01001 \mapsto 01001010 \mapsto 0100101001001 \mapsto \cdots.$$

Note that at each step the part of the word that we have already written down does not change, so the (infinite) Fibonacci word begins with the symbols $0100101001001 \cdots$.

### 3.4. Symbolic codings and countable-state shifts.

The *Gauss map* is the map $f \colon (0, 1] \to (0, 1]$ defined by $f(x) = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$; that is, $f(x)$ is the fractional part of $1/x$. Note that $f$ maps each interval $I_k := (\frac{1}{k+1}, \frac{1}{k}]$ monotonically onto $(0, 1]$.

**Exercise 3.9.** *Prove that for every sequence $k_0, k_1, k_2, \dots \in \mathbb{N}$, there is a unique $x \in (0, 1]$ such that $f^n(x) \in I_{k_n}$ for every $n \geq 0$.*

This gives a bijection between $(0, 1]$ and the space $\mathbb{N}^{\mathbb{N}}$ of infinite sequences of natural numbers.

**Exercise 3.10.** *Prove that if $x \in (0, 1]$ and $k_0, k_1, \dots \in \mathbb{N}$ are related as in Exercise 3.9, then $k_0, k_1, \dots$ is the continued fraction expansion of $x$.*

## 4. A crash course in measure theory

### 4.1. Basic examples of measures and integration.

Informally, a *measure* on a set $X$ is a function $\mu$ that assigns to each subset $E \subset X$ a weight $\mu(E) \geq 0$, with the property that $\mu(\bigsqcup_{i=1}^\infty E_i) = \sum \mu(E_i)$ whenever the sets $E_i$ are disjoint. In the formal definition, $\mu(E)$ is actually only defined when $E$ comes from the *$\sigma$-algebra of measurable sets*, but the details of this will not concern us here, as all the sets we consider are measurable.

4.1.1. *Lebesgue measure on* $\mathbb{R}$. The first measure to understand generalizes the notion of *length* to subsets of $\mathbb{R}$ that need not be intervals. Given an interval $I \subset \mathbb{R}$, let $\ell(I)$ denote the length of $I$. We can define a measure $\mu$ on $\mathbb{R}$, called *Lebesgue measure*, by declaring that

(1) $\mu(I) = \ell(I)$ when $I$ is an interval;
(2) $\mu(\bigsqcup_{j=1}^k I_j) = \sum_{j=1}^k \ell(I_j)$ when $I_1, \dots, I_k$ are disjoint intervals; and in general,
(3) $\mu(E) = \inf\{\sum_{j=1}^\infty \ell(I_j) : E \subset \bigcup_{j=1}^\infty I_j$ and each $I_j$ is an interval$\}$.

**Exercise 4.1.** *Show that these definitions are consistent and that it does not matter whether the intervals we use are open or closed.*

One important property of Lebesgue measure is that $\mu(\{x\}) = 0$ for every $x \in \mathbb{R}$; in other words, there are no points that carry positive measure. We say that $\mu$ is *non-atomic.*

**Exercise 4.2.** *Show that Lebesgue measure satisfies* $\mu(\bigsqcup_{i=1}^\infty E_i) = \sum_{i=1}^\infty \mu(E_i)$ *whenever the sets* $E_i$ *are closed and disjoint.*

It follows that every countable set $E$ has $\mu(E) = 0$; in particular, $\mu(\mathbb{Q}) = 0$. We could define other measures on $\mathbb{R}$ for which this fails; given any $x \in \mathbb{R}$ we can define the *delta measure* $\delta_x$ by $\delta_x(E) = 1$ if $x \in E$ and $0$ if $x \notin E$. These measures are atomic. We will mostly be concerned with non-atomic measures.

*Remark* 4.3. The *countable additivity* property in Exercise 4.2 actually holds whenever for any collection of disjoint *measurable sets* $E_i$: a set $E$ is said to be measurable if $\mu(X) = \mu(X \cap E) + \mu(X \cap E^c)$ for every $X \subset \mathbb{R}$. I find it helpful to think of such an $E$ as representing an event that can be conditioned on in the sense of probability, so that $\mu(X \cap E)/\mu(E)$ is the conditional probability that $X$ occurs given that $E$ occurred, and similarly with $E^c$; then the criterion for measurability can be thought of as $\mathbb{P}(X) = \mathbb{P}(X|E)\mathbb{P}(E) + \mathbb{P}(X|E^c)\mathbb{P}(E^c)$. Of course $\mu$ is not actually a probability measure so the analogy breaks down somewhat. In any case we will not concern ourselves with these details, as all of the sets we encounter in these notes are measurable.

A property is said to hold *$\mu$-almost everywhere* (usually abbreviated $\mu$-a.e.) if there is a set $E \subset X$ such that $\mu(X \setminus E) = 0$ and the property holds for all $x \in E$. For example, if $\mu$ is Lebesgue measure on $\mathbb{R}$, then $\mu$-a.e. real number is irrational, while $\delta_0$-a.e. real number is equal to 0.

**Exercise 4.4.** *Prove that if we write* $E + t := \{x + t : x \in E\}$, *then* $\mu(E + t) = \mu(E)$ *for every* $E, t$. *This property is called* translation-invariance *of Lebesgue measure.*

4.1.2. *Lebesgue integration.* Given a positive function $\varphi\colon [a, b] \to (0, \infty)$, Riemann integration can be thought of as cutting up the region $\{(x, y) : a \le x \le b, 0 \le y \le \varphi(x)\}$ that lies underneath the graph of $\varphi$ into a large number of very narrow vertical strips, each of which is nearly a rectangle, then adding up the areas of these rectangles and taking a limit as the number of rectangles goes to $\infty$ to get $\int_a^b \varphi(x)\,dx$.

Roughly speaking, the *Lebesgue integral* is defined by cutting that same region up into *horizontal* strips, adding up the (approximate) areas of the strips, and then taking a limit as the width of the strips goes to 0. That is, Riemann slices the cake vertically, while Lebesgue slices it horizontally. To make the definition slightly more formal, say that a partition $\xi$ of $[0, \infty)$ is a sequence $0 = c_0 < c_1 < c_2 < \cdots$ such that $c_n \to \infty$, and the diameter of $\xi$ is $\operatorname{diam} \xi = \sup_n (c_{n+1} - c_n)$. Given a function $\varphi$ and a partition $\xi$, consider the sets

$$A_n^{\varphi,\xi} = \{x : \varphi(x) \ge c_n\}.$$

The set $A_n^{\varphi,\xi} \times [c_n, c_{n+1}]$ is one of the "horizontal strips" mentioned above, and it follows from the definition that

$$\bigcup_{n=0}^{\infty} A_{n+1}^{\varphi,\xi} \times [c_n, c_{n+1}] \subset \{(x,y) : 0 \leq y \leq \varphi(x)\} \subset \bigcup_{n=0}^{\infty} A_n^{\varphi,\xi} \times [c_n, c_{n+1}],$$

as illustrated in Figure 5. Thus the "area under the graph of $\varphi$" should be close to the sum of the "areas of the horizontal strips", and we can define the Lebesgue integral as

$$\int \varphi \, d\mu = \lim_{\operatorname{diam} \xi \to 0} \sum_{n=0}^{\infty} \mu(A_n^{\varphi,\xi})(c_{n+1} - c_n).$$

This last expression is a limit of Riemann sums and thus we obtain

$$\int \varphi \, d\mu = \int_0^{\infty} \mu\{x : f(x) \geq t\} \, dt,$$

where the integral on the right is the usual Riemann integral.

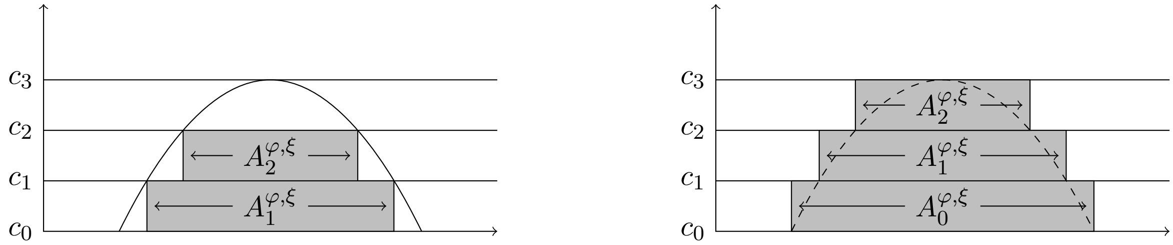

Figure 5. Cutting the cake horizontally.

One difference between this approach and Riemann integration is that Lebesgue integration can handle a large class of functions: for example, if $\varphi = \mathbf{1}_{\mathbb{Q}}$ is the characteristic function of the rational numbers (so $\varphi(x) = 1$ if $x \in \mathbb{Q}$ and 0 if $x$ is irrational), then the Riemann integral of $\varphi$ does not exist (the lower and upper Rieman sums do not approach the same limit), but the Lebesgue integral does. It turns out that there are still some functions that cannot be integrated, but such examples will not concern us here.

**Exercise 4.5.** *Compute $\int \mathbf{1}_{\mathbb{Q}} \, d\mu$.*

The above definition was for nonnegative functions, but extends easily to functions that take both positive and negative values. However, we must be careful that the function does not have positive and negative parts that both integrate to $\infty$, otherwise we would end up with $\infty - \infty$ when we compute the overall integral. To avoid this we usually restrict our attention to the following class of *integrable* functions:

$$L^1(\mu) = \left\{ \varphi : \varphi \text{ is a (measurable) function with } \int |\varphi| \, d\mu < \infty \right\}.$$

Notice that the definition of Lebesgue integration easily adapts to spaces other than $\mathbb{R}$, and measures other than Lebesgue measure.

**Exercise 4.6.** *Let $\delta_0$ be the delta measure at 0, and show that $\int \varphi \, d\delta_0 = \varphi(0)$ for all $\varphi$.*

Given a measure $\mu$ (Lebesgue measure, or otherwise), the set $L^1(\mu)$ is a vector space, on which we can define a norm by

$$\|\varphi\|_1 = \int |\varphi| \, d\mu.$$

Crucially, $L^1(\mu)$ is *complete* with respect to this norm: if $\varphi_n$ is any Cauchy sequence[5] of functions in $L^1(\mu)$, then there is $\varphi \in L^1(\mu)$ such that $\|\varphi_n - \varphi\|_1 \to 0$. It is useful to know that the set of *simple functions* given by

$$L^1_{\text{simple}}(\mu) := \Big\{ \sum_{j=1}^n c_j \mathbf{1}_{E_j} : c_j \in \mathbb{R}, E_j \subset X, \mu(E_j) < \infty \Big\} \tag{4.1}$$

is dense in $L^1(\mu)$. (Warning: the notation in (4.1) is not standard.)

This is as good a place as any to point out that linear algebra in *infinite-dimensional* vector spaces such as $L^1(\mu)$ is often a rather different beast from our familiar finite-dimensional linear algebra. The following two exercises illustrate this.

**Exercise 4.7.** *Show that if $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation with the property that $T^k x \to 0$ for every $x \in \mathbb{R}^n$, then there are $\lambda \in (0,1)$ and $C \geq 1$ such that for every $x \in \mathbb{R}^n$ and $k \in \mathbb{N}$, we have $\|T^k x\| \leq C\lambda^k$.*

**Exercise 4.8.** *Let $\mu$ be counting measure on the natural numbers, and $\ell^1 = L^1(\mu)$, so that $\ell^1$ is the set of all sequences whose sums converge absolutely. Define a linear transformation $T\colon \ell^1 \to \ell^1$ by $(Tx)_i = (1 - \frac{1}{i})x_i$. Prove that $\|T^k x\|_1 \to 0$ for every $x \in \ell^1$, but that given any $\lambda \in (0,1)$ and $C \geq 1$ there is $x \in \ell^1$ such that $\|T^k x\| > C\lambda^k$ for some $k \in \mathbb{N}$.*

4.1.3. *Lebesgue measure on $\mathbb{R}^n$.* Lebesgue measure on $\mathbb{R}$ generalizes the idea of length; similarly, one can construct Lebesgue measure on $\mathbb{R}^2$ generalizing the idea of area, on $\mathbb{R}^3$ generalizing the idea of volume, and so on. As with Lebesgue measure on $\mathbb{R}$, one first defines the function on a (relatively small) collection of sets with 'nice' structure, then extends it to more general sets. In $\mathbb{R}^n$ the collection of 'nice' sets is

$$\mathcal{S} = \{[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n : a_i \leq b_i \text{ for all } 1 \leq i \leq n\},$$

and we write

$$\mu([a_1, b_1] \times \cdots \times [a_n, b_n]) = (b_1 - a_1) \cdots (b_n - a_n).$$

Then we define Lebesgue measure for more general sets as

$$\mu(E) = \inf \Big\{ \sum_{j=1}^\infty \mu(R_j) : E \subset \bigcup_{j=1}^\infty R_j \text{ and } R_j \in \mathcal{S} \text{ for all } j \Big\}. \tag{4.2}$$

To develop all of this completely, one needs to describe precisely the axioms that the collection of 'nice' sets should satisfy, and the conditions under which the definition in the last line agrees with the original definition on elements of $\mathcal{S}$ (this is the *Carathéodory extension theorem*), but we ignore these details.

[5] This means that for every $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that $\|\varphi_m - \varphi_n\| < \varepsilon$ whenever $m, n \geq N$.

4.1.4. *Hausdorff measure and dimension.* Let $C \subset [0,1]$ be the middle-third Cantor set. That is, let $C_0 = [0,1]$ and construct a sequence of sets $C_n \subset [0,1]$ such that

(1) $C_n$ is a union of $2^n$ intervals of length $3^{-n}$;
(2) $C_{n+1}$ is obtained from $C_n$ by deleting the open middle third of each of the $2^n$ component intervals.

Thus $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$, $C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$, and so on. Note that $C_0 \supset C_1 \supset C_2 \supset C_3 \supset \cdots$. The Cantor set is $C = \bigcap_n C_n$.

**Exercise 4.9.** *Prove that the there is a natural homeomorphism between $C$ and the space of infinite binary sequences $\{0,1\}^{\mathbb{N}}$. In particular, $C$ is uncountable. (Note, however, that the set of endpoints of removed intervals is only countable.)*

**Exercise 4.10.** *Use the fact that $\mu(C) \leq \mu(C_n)$ for all $n$ to prove that the Lebesgue measure of $C$ is 0.*

From these exercises we see that the Cantor set is large from the point of view of cardinality (indeed, it is the same cardinality as $\mathbb{R}$), but small from the point of view of Lebesgue measure. So how should we quantify its size?

One way to approach this is to consider a new measure that is obtained by modifying the definition of Lebesgue measure. Recall that Lebesgue measure on $\mathbb{R}$ was defined by

$$\mu(E) = \inf\Big\{ \sum_{j=1}^\infty \ell(I_j) : E \subset \bigcup_{j=1}^\infty I_j \text{ and each } I_j \text{ is an interval}\Big\}. \tag{4.3}$$

Moreover, for every $\varepsilon > 0$ and every interval $I$, we can cover $I$ by a collection of intervals of length $< \varepsilon$, and whose lengths add up to the length of $I$; thus we can require that all the intervals $I_j$ in (4.3) have length $< \varepsilon$, without changing the value of $\mu(E)$. In particular, the following is equivalent to (4.3) (even if the reason for writing it this way is not immediately clear):

$$\mu(E) = \lim_{\varepsilon\to 0} \inf\Big\{ \sum_{j=1}^\infty \ell(I_j) : E \subset \bigcup_{j=1}^\infty I_j \text{ and each } I_j \text{ is an interval with } \ell(I_j) < \varepsilon\Big\}. \tag{4.4}$$

We consider the following variant of (4.4): given $\alpha \geq 0$, the *$\alpha$-dimensional Hausdorff measure on* $\mathbb{R}$ is defined by

$$m_H^\alpha(E) = \lim_{\varepsilon\to 0} \inf\Big\{ \sum_{j=1}^\infty \ell(I_j)^\alpha : E \subset \bigcup_{j=1}^\infty I_j \text{ and each } I_j \text{ is an interval with } \ell(I_j) < \varepsilon\Big\}. \tag{4.5}$$

In particular $m_H^1(E) = \mu(E)$, while $m_H^0(E)$ is just the number of points in $E$.

**Exercise 4.11.** *Prove that each $m_H^\alpha$ is translation-invariant, just as Lebesgue measure is (recall Exercise 4.4).*

**Exercise 4.12.** *Prove that for every $\alpha \geq 0$, $\lambda > 0$, and $E \subset \mathbb{R}$, we have $m_H^\alpha(\lambda E) = \lambda^\alpha m_H^\alpha(E)$, where $\lambda E = \{\lambda x : x \in E\}$.*

This last scaling property makes $m_H^\alpha$ useful for studying sets with some self-similarity, such as the Cantor set.

**Exercise 4.13.** *Prove that for the middle-third Cantor set $C$, we have*

$$m_H^\alpha(E) = \begin{cases} \infty & \alpha < \log 2/\log 3, \\ 1 & \alpha = \log 2/\log 3, \\ 0 & \alpha > \log 2/\log 3. \end{cases}$$

In fact the phenomenon here, where $m_H^\alpha$ jumps from $\infty$ to 0 at a single critical value of $\alpha$, is not unique to the Cantor set.

**Exercise 4.14.** *Prove that if $E \subset \mathbb{R}$ and $\alpha \geq 0$ are such that $m_H^\alpha(E) < \infty$, then $m_H^{\alpha'}(E) = 0$ for all $\alpha' > \alpha$. Similarly, if $m_H^\alpha(E) > 0$, then $m_H^{\alpha'}(E) = \infty$ for all $\alpha' < \alpha$.*

It follows from Exercise 4.14 that there is *always* a critical value of $\alpha$ (depending on $E \subset \mathbb{R}$) where $m_H^\alpha(E)$ jumps from $\infty$ to 0. This value of $\alpha$ is called the *Hausdorff dimension* of $E$, and written $\dim_H(E)$. More formally, we write

$$\begin{aligned} \dim_H(E) &= \sup\{\alpha \geq 0 : m_H^\alpha(E) = \infty\} \\ &= \inf\{\alpha \geq 0 : m_H^\alpha(E) = 0\}. \end{aligned} \tag{4.6}$$

When $\alpha = \dim_H(E)$, the Hausdorff measure $m_H^\alpha(E)$ can be 0, $\infty$, or anything in between.

**Exercise 4.15.** *Prove that if $E \subset \mathbb{R}$ has positive Lebesgue measure, then $\dim_H(E) = 1$.*

**Exercise 4.16.** *Prove that if $E \subset \mathbb{R}$ has $\dim_H(E) < 1$, then $E$ has Lebesgue measure 0.*

It turns out that there are sets $E \subset \mathbb{R}$ with $\dim_H(E) = 1$ but Lebesgue measure 0. Can you find one?

**Exercise 4.17.** *Fix $\beta \in (0,1)$ and define the "middle-$\beta$" Cantor set $C_\beta$ by mimicking the construction of $C$, except that in going from $C_n$ to $C_{n+1}$, the ratio of the length of the deleted interval to the length of the interval containing it is $\beta$, instead of $\frac{1}{3}$. Thus $C = C_{1/3}$, and for more general $\beta$ we have $C_1 = [0, \frac{1-\beta}{2}] \cup [\frac{1+\beta}{2}, 1]$, etc. Prove that $\dim_H(C_\beta) = \log 2/\log(\frac{2}{1-\beta})$.*

**Exercise 4.18.** *Modify the above construction yet again by allowing $\beta$ to vary at each step, so the resulting Cantor set is determined by a sequence $\beta_1, \beta_2, \dots \in (0,1)$. By choosing $\beta_n \to 0$ quickly enough, construct a set $C' \subset [0,1]$ that is homeomorphic to the middle-third Cantor set $C$ but has positive Lebesgue measure. This set $C'$ is called a* fat Cantor set.

4.1.5. *Bernoulli and Markov measures.* Lebesgue measure is the most important measure on $\mathbb{R}^n$. On the symbolic space $\Sigma = A^{\mathbb{N}} = \{1, \dots, a\}^{\mathbb{N}}$, there are two families of measures that are particularly important. One is the *Bernoulli measures.* A *probability vector* is a vector $p \in \mathbb{R}^a$ such that $p_i \geq 0$ for all $1 \leq i \leq a$ and $\sum_i p_i = 1$. Given a probability vector $p$ and a word $w \in A^n$, define the measure of the cylinder $[w] \subset \Sigma$ as

$$\mu([w]) = p_{w_1} p_{w_2} \cdots p_{w_n}.$$

In other words, the measure of the cylinder is the probability of observing the outcomes $w_1, \dots, w_n$ in that order if we do $n$ independent trials of an experiment with $a$ possible outcomes such that $p_i$ gives the probability of observing the outcome $i$ on any given trial. The measure $\mu$ is then defined in general by (4.2), where $\mathcal{S} = \{[w] : w \in A^*\}$.

**Exercise 4.19.** *Prove that a Bernoulli measure $\mu$ has the property that $\mu(\sigma^{-1}[w]) = \mu([w])$ for every $w \in A^*$.*

Another important family is the *Markov measures.* Let $P$ be an $a \times a$ matrix whose rows are probability vectors (that is, $\sum_{j=1}^{a} P_{ij} = 1$ for all $i$); such a matrix is called a *stochastic matrix.* Assume that some power of $P$ is positive. By (a more general version of) the Perron–Frobenius Theorem 3.6, 1 is an eigenvalue of $P$, with a left eigenvector $\pi$ whose entries are all positive, and scaling $\pi$ we can assume without loss of generality that it is a probability vector. Define a Markov measure $\mu = \mu_P$ by taking the measure of a cylinder $[w]$ to be

$$\mu([w]) = \pi_{w_1} P_{w_1 w_2} P_{w_2 w_3} \cdots P_{w_{n-1} w_n},$$

and then extending by (4.2). In other words, the probability vector $\pi$ gives the probability of starting out in a particular state, and the entries of the stochastic matrix $P$ give the transition probabilities: $P_{ij}$ gives the probably that we will be in state $j$ tomorrow if we are in state $i$ today.

**Exercise 4.20.** *Prove that a Markov measure $\mu$ has the invariance property in Exercise 4.19.*

4.2. **Absolute continuity.** Let $m$ denote Lebesgue measure on $\mathbb{R}$. We can construct many other measures on $\mathbb{R}$ by the following procedure: fix a nonnegative function $\varphi \in L^1(m)$, and then define $\mu$ by $\mu(E) = \int_E \varphi\, dm = \int \varphi \mathbf{1}_E\, dm$. In this case $\varphi$ is sometimes called a *density function*, or more formally the *Radon–Nikodym derivative* of $\mu$ with respect to $m$, and written $\varphi = \frac{d\mu}{dm}$. We also have $\int \psi\, d\mu = \int \psi\varphi\, dm$ for every $\psi \in L^1(\mu)$.

Not every finite measure on $\mathbb{R}$ can be represented in this way, however.

**Exercise 4.21.** *Let $\delta_0$ be the point mass at $0$, and show that there is no $\varphi \in L^1(m)$ such that $\delta_0(E) = \int_E \varphi\, dm$ for every $E$.*

Given measures $\mu, \nu$ on $\mathbb{R}$, say that $\mu$ is *absolutely continuous* with respect to $\nu$ if every set with $\nu(E) = 0$ also have $\mu(E) = 0$; in this case we write $\mu \ll \nu$. Note that if there is $\varphi \in L^1(\nu)$ such that $\mu(E) = \int \varphi\, d\nu$, then $\mu \ll \nu$. The *Radon–Nikodym theorem* says that the converse is true as well; every absolutely continuous measure can be represented by integration against a density function.

One can consider absolute continuity with respect to any reference measure $\nu$, but we will primarily be concerned with the case when $\nu = m$ is Lebesgue measure. It is worth thinking for a little longer about which measures are *not* absolutely continuous with respect to Lebesgue measure. From Exercise 4.21 we see that this includes the "pure point" measures given by $\mu = \sum_k c_k \delta_{x_k}$ (where $c_k \geq 0$ and $x_k \in \mathbb{R}$). But it also includes some more involved examples.

Fix $p \in (0,1)$ and let $\zeta$ be the Bernoulli measure on $\{0,1\}^{\mathbb{N}}$ associated to the probability vector $(p, 1-p)$. Let $\pi\colon \{0,1\}^{\mathbb{N}} \to C$ be the homeomorphism from the full shift to the middle-third Cantor set from Exercise 4.9, and define a measure $\mu$ on $C$ (and hence on $\mathbb{R}$) by $\mu(E) = \mu(\pi^{-1}E)$.

This measure $\mu$ has the property that $\mu(C) = 1$, while $m(C) = 0$, so $\mu$ is not absolutely continuous with respect to Lebesgue measure. In fact, more than this is true: the set $C$ has the property that $m(C) = 0$ and $\mu(\mathbb{R} \setminus C) = 0$. When there is a set $C$ with this property, we say that the measures $\mu$ and $m$ are *mutually singular* and write $\mu \perp m$.

The *Lebesgue decomposition theorem* says that given any measure[6] $\mu$ on $\mathbb{R}$, there are measures $\mu_{\mathrm{ac}}$ and $\nu$ such that $\mu = \mu_{\mathrm{ac}} + \nu$, $\mu_{\mathrm{ac}} \ll m$, and $\nu \perp m$. In other words, any measure can be (uniquely) decomposed into an absolutely continuous part and a singular part.

[6] Technically $\mu$ needs to be *$\sigma$-finite*, but let us not worry about this detail.

**Exercise 4.22.** *Show that given any finite measure $\nu$ on $\mathbb{R}$, we can write $\nu = \eta + \xi$, where $\eta$ is* continuous *(meaning that $\eta\{x\} = 0$ for all $x \in \mathbb{R}$) and $\xi$ is a linear combination of point masses.*

Applying Exercise 4.22 to the singular part $\nu$ from the Lebesgue decomposition theorem and writing $\mu_{\text{sing}} = \eta$, $\mu_{\text{pp}} = \xi$, we obtain a decomposition of our original measure as $\mu = \mu_{\text{ac}} + \mu_{\text{sing}} + \mu_{\text{pp}}$, where $\mu_{\text{ac}}$ is the absolutely continuous part, $\mu_{\text{sing}}$ is the singular continuous part, and $\mu_{\text{pp}}$ is the pure point part.

4.3. **Measure-preserving transformations.** A *measure-preserving transformation* consists of a space $X$ equipped with a measure $\mu$, together with a map $T\colon X \to X$ that leaves the measure invariant, meaning that $\mu(T^{-1}E) = \mu(E)$ for all (measurable) sets $E \subset X$. Exercises 4.19 and 4.20 showed that the shift map $\sigma\colon \Sigma \to \Sigma$ becomes a measure-preserving transformation when it is equipped with any Bernoulli or Markov measure; equivalently, we say that Bernoulli and Markov measures are *shift-invariant*.

At first it may seem a little strange that we look at $T^{-1}E$ instead of $T(E)$ in the definition of invariance. Some explanation for this may be found in the following exercise.

**Exercise 4.23.** *Show that a measure $\mu$ is $T$-invariant if and only if $\int \varphi\, d\mu = \int \varphi \circ T\, d\mu$ for all $\varphi \in L^1(\mu)$. (The expected value of the measurement $\varphi$ is the same whether we perform it today or tomorrow.) Hint: reduce the problem from $L^1(\mu)$ to the set of simple functions from* (4.1) *by using the fact that this set is dense in $L^1(\mu)$.*

Another reason for looking at $T^{-1}E$ is that for a non-invertible transformation, such as the one-sided shift map $\sigma\colon \Sigma \to \Sigma$, there may be many sets whose forward iterates eventually cover the whole space.

**Exercise 4.24.** *Show that given any $w \in A^*$ there is $n \in \mathbb{N}$ such that $\sigma^n([w]) = \Sigma$.*

If $\mu_0$ and $\mu_1$ are two measures on a space $X$ that are both invariant under a transformation $T\colon X \to X$, then for every $t \in [0, 1]$, the measure defined by the *convex combination*

$$\mu_t(E) = t\mu_1(E) + (1 - t)\mu_0(E)$$

is also a $T$-invariant measure on $X$. For example, if $\mu$ is the Bernoulli measure on $\Sigma = \{0, 1\}^{\mathbb{N}}$ such that $\mu[w] = 2^{-|w|}$ for all $w$, and $\delta_0$ is the delta measure sitting on the point $0000\cdots \in \Sigma$, then $\nu = \frac{1}{2}(\mu + \delta_0)$ is also a $\sigma$-invariant measure.

We say that an invariant measure $\mu$ is *ergodic* if it cannot be decomposed as a non-trivial convex combination of two other invariant measures. Thus the measure $\nu$ in the previous paragraph is not ergodic.

**Exercise 4.25.** *Prove that the following are equivalent for a $T$-invariant measure $\mu$.*

*(1) $\mu$ is ergodic.*
*(2) If $E \subset X$ has $T^{-1}E = E$, then either $\mu(E) = 0$ or $\mu(X \setminus E) = 0$.*
*(3) If $\varphi\colon X \to \mathbb{R}$ has $\varphi(Tx) = \varphi(x)$ for $\mu$-a.e. $x$, then there is $c \in \mathbb{R}$ such that $\varphi(x) = c$ for $\mu$-a.e. $x$.*

**Exercise 4.26.** *Prove that every Bernoulli measure is ergodic for the shift map $\sigma$.*

**Exercise 4.27.** *Determine necessary and sufficient conditions on the stochastic matrix $P$ for the associated Markov measure to be ergodic for the shift.*

Usually we are interested in transformations that preserve a *probability* measure; that is, a measure with $\mu(X) = 1$.

**Theorem 4.28** (Poincaré recurrence)**.** *If $T\colon X \to X$ preserves a probability measure $\mu$, then for every $A \subset X$ and $\mu$-a.e. $x \in A$ there is $\tau(x) \in \mathbb{N}$ such that $T^{\tau(x)}(x) \in A$.*

*Proof.* Let $B = \{x \in A : T^n x \notin A \ \forall n \geq 1\}$ be the set of 'bad' points where the conclusion fails. Then given any $j, k \geq 0$ we claim that $T^{-j}B \cap T^{-k}B = \emptyset$; indeed, if $k > j$ and $x \in T^{-j}B \cap T^{-k}B$, then $y = T^j x \in B \cap T^{-(k-j)}B$, so $T^{k-j}y \in B \subset A$, contradicting the fact that $y \in B$. By this disjointness we have $1 = \mu(X) \geq \mu(\bigsqcup_n T^{-n}B) = \sum_n \mu(T^{-n}B) = \sum_n \mu(B)$, and thus $\mu(B) = 0$. □

Given a transformation $T\colon X \to X$ and a function $\varphi\colon X \to \mathbb{R}$, the *nth ergodic sum* is the function

$$S_n\varphi(x) := \sum_{k=0}^{n-1} \varphi(T^k x) = \varphi(x) + \varphi(Tx) + \varphi(T^2x) + \cdots + \varphi(T^{n-1}x).$$

For example, if $X = \{1, \dots, a\}^{\mathbb{N}}$, $T = \sigma$, and $\varphi = \mathbf{1}_{[1]}$ is the characteristic function of the cylinder $[1]$, then $S_n\varphi(x)$ is the number of times that the symbol 1 appears in the first $n$ places of $x$. Suppose $\mu$ is the Bernoulli measure on $X$ associated to a probability vector $p$. Then by the definition of the Bernoulli measure and by the strong law of large numbers, we see that for $\mu$-a.e. $x \in X$ the *ergodic averages* (or *Birkhoff averages*) $\frac{1}{n}S_n\varphi$ have the property that $\lim_{n\to\infty} \frac{1}{n} S_n \mathbf{1}_{[i]}(x) = p_i$ for each symbol $i \in A$. This is a special case of the following foundational result in ergodic theory, which we state but do not prove.

**Theorem 4.29** (Birkhoff ergodic theorem)**.** *Let $(X, T, \mu)$ be an ergodic measure-preserving transformation. Then for every $\varphi \in L^1(X, \mu)$ we have $\lim_{n\to\infty} \frac{1}{n} S_n\varphi(x) = \int \varphi \, d\mu$.*

Given two measures $\mu, \nu$ on the same space $X$, we say that $\mu$ and $\nu$ are *mutually singular* if there is $E \subset X$ such that $\mu(E) = 1$ and $\nu(X \setminus E) = 1$. If $p \neq q$ are probability vectors, then there is $i$ such that $p_i \neq q_i$, and writing $E = \{x : \lim \frac{1}{n} S_n \mathbf{1}_{[i]}(x) = p_i\}$ we see that the corresponding Bernoulli measures $\mu_p, \mu_q$ have $\mu_p(E) = 1$ and $\mu_q(X \setminus E) = 1$, so distinct Bernoulli measures are mutually singular.

**Exercise 4.30.** *Prove that any two distinct Markov measures are mutually singular. More generally, prove that any two distinct ergodic measures are mutually singular.*

At the other extreme, we say that $\mu$ is *absolutely continuous* with respect to $\nu$ if every set with $\nu(E) = 0$ also has $\mu(E) = 0$, and we write $\mu \ll \nu$. Given $\nu$, one way to produce an absolutely continuous measure is to take a function $h \geq 0$, treat it as a *density function*, and define $\mu$ by $\mu(E) = \int_E h \, d\nu$. For example, the Gaussian (normal) probability distribution gives a measure on $\mathbb{R}$ that is absolutely continuous with respect to Lebesgue, with density function proportional to $e^{-x^2/2}$.

In fact, this is the *only* way to have an absolutely continuous measure: the *Radon–Nikodym theorem* says that if $\mu \ll \nu$, then there is a density function $h \geq 0$ such that $\mu$ is given as above. The function $h$ is called the *Radon–Nikodym derivative* and is denoted $\frac{d\mu}{d\nu}$. Then we have the following relationship between integrals w.r.t. $\mu$ and integrals w.r.t. $\nu$, which explains the notation:

$$\int \varphi \, d\mu = \int \varphi \frac{d\mu}{d\nu} \, d\nu$$

4.4. **Measure-theoretic entropy.** The *entropy* of a probability vector $p = (p_1, \dots, p_a)$ is defined to be

$$H(p) := \sum_{i=1}^{a} -p_i \log p_i. \tag{4.7}$$

**Exercise 4.31.** *Show that* $0 \leq H(p) \leq \log a$ *for all probability vectors* $p \in \Delta_a := \{(p_1, \dots, p_a) : p_i \geq 0, \sum p_i = 1\}$*, with equality if and only if* $p_i = \frac{1}{a}$ *for all* $a$*.* Hint: use convexity. If you struggle, first consider the case $a = 2$.

To understand where (4.7) comes from, we interpret $H(p)$ as the *average information* gained by observing which of the events $1, 2, \dots, a$ occurs, as follows. Let $I(p)$ be the amount of information that we gain if we observe an event whose probability is $p$. We may reasonably expect the function $I$ to satisfy the following properties:

1. $I(p) \geq 0$, with equality if and only if $p = 1$ (we never lose information by making an observation, and we gain information if and only if the event had a positive probability of not occurring);
2. $I$ is continuous and non-increasing (if $p$ changes by just a little bit, then so does $I$, and less likely events carry more information);
3. $I(pq) = I(p) + I(q)$ (if two independent events occur, the amount of information we gain from observing both of them is the sum of the information we gain from observing each one on its own).

One can show that the only function $I \colon (0, 1] \to \mathbb{R}$ satisfying these axioms is the function $I(p) = -\log p$, and thus (4.7) can be rewritten as $H(p) = \sum_i p_i I(p_i)$, which is the expected amount of information we gain by making a single observation of an experiment whose outcomes are distributed according to the probability vector $p$.

Now consider the full shift $\Sigma$ on $a$ symbols. Let $p \in \Delta_a$ be a probability vector, and let $\mu$ be the Bernoulli measure on $\Sigma$ associated to $p$. Given $x \in \Sigma$, keeping track of the first symbol $x_1 \in A$ under iterations of $\sigma$ amounts to conducting successive experiments where each has $a$ possible outcomes; if $x$ is distributed according to $\mu$, then these experiments are independent and identically distributed, so the amount of information we expect to gain per trial is equal to $H(p)$. We call this the *measure-theoretic entropy of* $\mu$ and write $h_\mu(\sigma) = H(p)$.

If $x \in \Sigma$ is distributed according to a shift-invariant measure $\nu$ on $\Sigma$ that is *not* a Bernoulli measure, then the experiments described in the previous paragraph are still identically distributed (this is because $\nu$ is $\sigma$-invariant) but are no longer independent. We continue to define the measure-theoretic entropy of $\nu$ as the amount of information we expect to gain per trial *in the long run*, which can be written as

$$h_\nu(\sigma) := \lim_{n\to\infty} \frac{1}{n} \sum_{w \in A^n} -\nu[w] \log \nu[w]. \tag{4.8}$$

**Exercise 4.32.** *Writing* $c_n = \sum_{w \in A^n} -\nu[w] \log \nu[w]$*, show that* $c_{m+n} \leq c_m + c_n$*, and then use Exercise 3.4 to prove that the limit in* (4.8) *exists and is equal to* $\inf_n \frac{c_n}{n}$*.*

Recall that the topological entropy of the full shift on $a$ symbols is $\log a$; by Exercise 4.31 we therefore have $0 \leq h_\mu(\sigma) \leq \log a$ for every Bernoulli measure $\mu$, with equality if and only if $\mu$ is the Bernoulli measure that gives every 1-cylinder equal weight. In fact, Exercise 4.32 shows that $h_\mu(\sigma) \leq c_1 \leq \log a$ for every invariant measure $\mu$. A similar picture holds for shift spaces that are not the full shift.

**Theorem 4.33** (Variational principle). *Let $X$ be a shift space on a finite alphabet, and $\sigma\colon X \to X$ the shift map. Let $h_{\text{top}}(X,\sigma) = h(\mathcal{L}(X))$ be the topological entropy of $X$, and let $\mathcal{M}(X)$ be the set of all $\sigma$-invariant measures on $X$. Then we have*

$$h_{\text{top}}(X,\sigma) = \sup_{\mu\in\mathcal{M}(X)} h_\mu(\sigma).$$

Note that $\mathcal{M}(X)$ is a convex set because the convex combination of two invariant measures is itself an invariant measure. An *extreme point* of a convex set is a point that cannot be written as a proper convex combination of two other points in the set. In particular, the extreme points of $\mathcal{M}(X)$ are precise the ergodic measures.

In fact $\mathcal{M}(X)$ is a *Choquet simplex*: every element in $\mathcal{M}(X)$ can be written in a unique way as a (possibly infinite) convex combination of its extreme points. This leads to the *ergodic decomposition*; every invariant measure can be decomposed into its ergodic components.

It is often the case that $\mathcal{M}(X)$ is very large (infinite-dimensional); for example, when $X$ is a topological Markov chain given by a transition matrix $T$, we can take any stochastic matrix $P$ such that $P_{ij} \leq T_{ij}$ for all $i,j$ and obtain the corresponding Markov measure, which is supported on $X$. We can also take any periodic sequence $x \in X$, say $x = \sigma^k x$, and define an atomic measure on $X$ by $\mu_x = \frac{1}{k}\sum_{j=0}^{k-1}\delta_{\sigma^j x}$.

**Exercise 4.34.** *Show that the periodic orbit measure $\mu_x$ is $\sigma$-invariant and ergodic.*

In cases like this the variational principle (together with a generalization from entropy to *pressure*, which we do not state here) helps us to select certain distinguished invariant measures. For example, on the full shift, the equidistributed Bernoulli measure is distinguished by the fact that it maximizes the entropy; in fact, one can show that it is the *unique* measure of maximal entropy.

On a topological Markov chain with primitive transition matrix $T$, one can find a measure of maximal entropy via the following procedure: let $\lambda$ be the Perron–Frobenius eigenvalue of $T$, and let $v$ and $w$ be row and column eigenvectors for $\lambda$, so that $vT = \lambda v$ and $Tw = \lambda w$. Assume that $v,w$ are normalized so that $vw = \sum v_i w_i = 1$. Define a probability vector $\pi$ and a stochastic matrix $P$ by

$$\pi_i = v_i w_i, \qquad P_{ij} = \frac{T_{ij} w_j}{\lambda w_i}.$$

**Exercise 4.35.** *Show that the Markov measure $\mu$ given by $\pi, P$ has the following properties:*

*(a) it is a measure (must check that $\forall w \in \mathcal{L}_n(X)$ we have $\sum_{i=1}^n \mu[wi] = \mu[w]$), and*
*(b) it is $\sigma$-invariant.*

The measure $\mu$ constructed in Exercise 4.35 is called the *Parry measure*, and can be shown to be the unique measure of maximal entropy for the Markov shift $X$.

The Variational Principle holds for a broader class of systems (continuous maps on compact metric spaces) but in the interests of brevity, we omit the details of the definitions involved.

4.5. **Basic functional analysis.** We often need to be careful about specifying which class of functions we work with. When $X$ is a metric space it is natural to consider the set

$$C(X) = C(X,\mathbb{R}) = \{\varphi\colon X \to \mathbb{R} : \varphi \text{ is continuous}\}.$$

This is a vector space over $\mathbb{R}$ in a natural way. If $X$ is compact then every $\varphi \in C(X)$ is bounded, and so the following defines a norm on $C(X)$, called the *uniform norm*:

$$\|\varphi\| = \|\varphi\|_u := \sup_{x\in X} |\varphi(x)|.$$

This makes $C(X)$ into a metric space, and this metric space can be shown to be *complete*: every Cauchy sequence $\varphi_n \in C(X)$ converges to some $\varphi \in C(X)$ (ie., $\|\varphi_n - \varphi\|_u \to 0$). A complete normed vector space is called a *Banach space*.

The simplest examples of Banach spaces are finite-dimensional vector spaces, but since all norms are equivalent in this setting, most of the work in Banach space theory goes towards dealing with infinite-dimensional spaces such as $C(X)$.

Given $0 < \alpha < 1$ one can consider the class of *Hölder continuous functions with exponent* $\alpha$ on $X$, which contains those functions $\varphi$ for which there is a constant $C > 0$ such that $|\varphi(x) - \varphi(y)| \leq Cd(x, y)^\alpha$ for all $x, y \in X$. When $\alpha = 1$, such a function is called *Lipschitz*. The space of Hölder continuous functions is denoted $C^\alpha(X)$ and is a Banach space with the norm

$$\|\varphi\|_\alpha := \|\varphi\|_u + |\varphi|_\alpha, \qquad |\varphi|_\alpha := \sup_{x,y\in X} \frac{|\varphi(x) - \varphi(y)|}{d(x, y)^\alpha}.$$

Note that $|\varphi|_\alpha$ is not a norm in its own right, since it vanishes on every constant function; thus it is a *seminorm*.

When $X = \mathbb{R}$ (or more generally when $X$ is a smooth manifold) it is often useful to consider the space of *continuously differentiable functions*

$$C^1 = \{\varphi : \varphi' \text{ exists and is continuous}\}.$$

If we restrict $X$ to be a compact subset of $\mathbb{R}$ (or more generally, a compact smooth manifold), then $C^1$ is a Banach space with norm

$$\|\varphi\|_{C^1} = \|\varphi\|_u + \|\varphi'\|_u.$$

One can define the spaces $C^2, C^3, \ldots$ in a similar way.

Another important type of function space comes from considering not the differentiability properties of functions, but rather their integrability properties. Let $\mu$ be a measure on a space $X$. We already introduced the space $L^1(\mu)$ of all integrable functions; more generally, given $1 \leq p < \infty$, we let

$$L^p(\mu) = L^p(X, \mu) = \Big\{\varphi\colon X \to \mathbb{R} : \int_X |\varphi|^p \, d\mu < \infty\Big\}$$

which becomes a Banach space when equipped with the norm

$$\|\varphi\|_p = \Big(\int_X |\varphi^p| \, d\mu\Big)^{1/p}. \tag{4.9}$$

Note the similarity to the $\ell^p$-norms in (1.1). In fact (4.9) reduces to (1.1) when $X = \{1, \ldots, n\}$ and $\mu$ is the counting measure $\mu(E) = \#E$. A crucial difference is that in general, $L^p(\mu)$ is infinite-dimensional and the $L^p$ norms define different topologies (and indeed, different spaces).

There is a subtlety we are glossing over here, namely that the $L^p$ spaces are actually defined in terms of *equivalence classes* of functions, where $\varphi \sim \psi$ if $\varphi = \psi$ $\mu$-a.e. In practice this distinction will not bother us much here; just remember that $L^p$-functions are defined “almost everywhere”, instead of “everywhere”. This does show up in the definition of the $L^p$ space for $p = \infty$:

$$\|\varphi\|_\infty = \inf_{\psi\sim\varphi} \|\psi\|_u, \qquad L^\infty(\mu) = \{\varphi\colon X \to \mathbb{R} : \|\varphi\|_\infty < \infty\}.$$

## 5. A crash course in linear algebra

5.1. **Hilbert spaces and tensor products.** Recall the definition of an inner product space from §1.1.1. Let $H$ be an inner product space; then $H$ is a metric space with $d(x,y) = \|x-y\|$, where $\|\cdot\|$ is the norm induced by the inner product. If $H$ is complete in this metric, then it is called a *Hilbert space*. Every Hilbert space is a Banach space, but not vice versa.

**Exercise 5.1.** *Prove that every finite-dimensional inner product space is complete.*

We will primarily be concerned with the finite-dimensional case, so it is perfectly reasonable for you to picture $\mathbb{C}^n$ whenever you read the words 'Hilbert space'. Many of the things we say below work in infinite dimensions also; the biggest difference is that one needs to be careful what a 'basis' is when the space is no longer finite-dimensional, and so anything that involves a basis needs to be treated with significant care beyond the finite-dimensional setting.

A *linear operator* on a Hilbert space $H$ is a linear map $T$ from $H$ to itself. The *operator norm* of $T$ is

$$\|T\| = \sup\{\|Tx\| : x \in H, \|x\| = 1\}. \tag{5.1}$$

The operator $T$ is *bounded* if $\|T\| < \infty$. Write $L(H)$ for the set of all bounded linear operators on $H$. When $H = \mathbb{C}^n$, each $T \in L(H)$ is represented by an $n \times n$ matrix of complex numbers.

**Exercise 5.2.** *Prove that if $H$ is finite-dimensional, then every linear operator is bounded.*

In quantum mechanics, the set of all possible *quantum states* of a system is a Hilbert space, and a specific state is often denoted $|\psi\rangle \in H$. When $H = \mathbb{C}^n$, $|\psi\rangle$ can be interpreted as a column vector. In this case, the notation $\langle\psi|$ refers to the row vector with the same entries,[7] and $\langle\phi|\psi\rangle$ is the inner product of two vectors, which can be interpreted as the product of a $1 \times n$ matrix (the row vector) with an $n \times 1$ matrix (the column vector). Note that if the order is reversed, then $|\psi\rangle\langle\phi|$ is the product of an $n \times 1$ matrix with a $1 \times n$ matrix, giving an $n \times n$ matrix that represents an operator on $\mathbb{C}^n$.

To describe a quantum system in terms of its subsystems, we need *tensor products.* If $V, W$ are finite-dimensional vector spaces with bases $\{v_1, \dots, v_m\}$ and $\{w_1, \dots, w_n\}$, respectively, then their *tensor product* is the vector space consisting of all formal linear combinations of pairs of elements from the two bases. That is, given $1 \le i \le m$ and $1 \le j \le n$, we write $v_i \otimes w_j$ for the ordered pair $(v_i, w_j)$; the tensor product $V \otimes W$ is the set of all formal linear combinations of the $v_i \otimes w_j$:

$$V \otimes W = \Big\{ \sum_{i=1}^m \sum_{j=1}^n c_{ij} v_i \otimes w_j : c_{ij} \in \mathbb{C} \Big\}. \tag{5.2}$$

We can also take the tensor product of vectors in $V$ and $W$: given $x = \sum_i x_i v_i \in V$ and $y = \sum_j y_j w_j \in W$, the tensor product of $x$ and $y$ is

$$x \otimes y = \sum_{i=1}^m \sum_{j=1}^n x_i y_j v_i \otimes w_j. \tag{5.3}$$

One can easily check that this is linear in both $x$ and $y$, so $(cx + x') \otimes y = c(x \otimes y) + (x' \otimes y)$, and so on. When $V, W$ are inner product spaces, their tensor product carries an inner product defined by

$$\langle v \otimes w, v' \otimes w' \rangle = \langle v, v' \rangle \langle w, w' \rangle. \tag{5.4}$$

[7] More formally, $\langle\psi|$ is an element of the *dual space* of $H$, which is naturally isomorphic to $H$ itself.

Note that $\mathbb{C}^m \otimes \mathbb{C}^n$ is isomorphic to $\mathbb{C}^{mn}$, so dimension is multiplicative. This is different from the *direct* product, where $\mathbb{C}^m \times \mathbb{C}^n$ is isomorphic to $\mathbb{C}^{m+n}$, and dimension is additive.

Finally, we can take the tensor product of operators. Given Hilbert spaces $H_1, H_2$ and linear operators $T_j \in L(H_j)$ for $j = 1, 2$, define their tensor product $T_1 \otimes T_2 \in L(H_1 \otimes H_2)$ by

$$(T_1 \otimes T_2)(v \otimes w) = (T_1(v)) \otimes (T_2(w)). \tag{5.5}$$

**Exercise 5.3.** *If $H_1 = H_2 = \mathbb{C}^2$ and $T_1, T_2 \in L(\mathbb{C}^2)$, then $T_1, T_2$ are represented by $2 \times 2$ matrices, while $T_1 \otimes T_2 \in L(\mathbb{C}^4)$ is represented by a $4 \times 4$ matrix. Find the relationship between these matrices.*

5.2. **Various flavors of operators.** Given a Hilbert space $H$, the set $L(H)$ is a vector space in its own right because operators can be added together, and multiplied by scalars. But more than this is true; operators can also be multiplied together via composition (which corresponds to matrix multiplication in the finite-dimensional case). Thus $L(H)$ is more than a vector space, it is an *algebra*; a vector space $V$ equipped with a product $V \times V \to V$ satisfying the usual expected associative and distributive laws (but not necessarily commutativity).

For every $v \in H$ and $T \in L(H)$, it can be shown that there is a unique element $T^* v \in H$ such that $\langle T^* v, w\rangle = \langle v, Tw\rangle$ for all $w \in H$. Moreover, the map $T^* \colon H \to H$ is linear and bounded; this is the *adjoint* of $T$, and is defined by the condition

$$\langle T^* v, w\rangle = \langle v, Tw\rangle \text{ for all } v, w \in H. \tag{5.6}$$

**Exercise 5.4.** *Show that if $H = \mathbb{C}^n$, then the matrix representing $T^*$ is the conjugate transpose of the matrix representing $T$. That is, if $T$ is represented by the matrix $A \in \mathbb{M}(n, \mathbb{C})$, then $T^*$ is represented by the matrix $B$ with $B_{ij} = \overline{A_{ji}}$.*

It follows immediately from the definition (or from Exercise 5.4) that $T^{**} = T$. Thus the map $T \mapsto T^*$ is an *involution* on $L(H)$ (a bijection that is its own inverse). With this involution, $L(H)$ becomes an example of a *$C^*$-algebra*; that is, it is an algebra (vector space with multiplication) equipped with a complete norm (the operator norm) and an involution satisfying the following properties.

**Exercise 5.5.** *Prove that for every $S, T \in L(H)$ and $c \in \mathbb{C}$, we have $(S + T)^* = S^* + T^*$, $(ST)^* = T^* S^*$, $(cT)^* = \bar{c} T^*$, and $\|T^* T\| = \|T\| \|T^*\|$.*

We turn now to some specific classes of operators that play a crucial role. An operator $T \in L(H)$ is called *Hermitian*, or *self-adjoint*, if $T^* = T$. Note that when $H = \mathbb{C}^n$, this condition says that the matrix representing $T$ is equal to its own conjugate transpose.

**Exercise 5.6.** *Prove that if $T$ is self-adjoint, then all of its eigenvalues are real.*

The *spectral theorem* in the finite-dimensional setting says that if $T \in L(\mathbb{C}^n)$ is self-adjoint, then there is an orthonormal basis $\{v_1, \dots, v_n\}$ for $\mathbb{C}^n$ such that each $v_j$ is an eigenvalue for $T$.

**Exercise 5.7.** *Is the basis $\{v_1, \dots, v_n\}$ determined uniquely by $T$?*

Another way to state the spectral theorem is as follows: writing $U$ for the linear transformation taking $e_j \to v_j$, where $\{e_1, \dots, e_n\}$ is the standard orthonormal basis for $\mathbb{C}^n$, we have $T = U^{-1} D U$, where $D$ is a diagonal matrix. The matrix of $U$ has the property that its columns are the vectors $v_j$, while the rows of $U^*$ are the complex conjugates of these vectors, and hence $(U^* U)_{ij} = \sum_k \overline{v_{ik}} v_{jk} = \langle v_j, v_i\rangle = \delta_{ij}$, so $U^* U = I$ is the identity matrix. A matrix $U$ with this property (its conjugate transpose is its inverse) is called *unitary*, and so the spectral theorem says that self-adjoint matrices can by *unitarily* diagonalized.

Using (5.6), the property of being unitary can be interpreted as follows: if $U \in L(H)$ is unitary, then $U^*U = I$, so $\langle v, w\rangle = \langle U^*Uv, w\rangle = \langle Uv, Uw\rangle$ for all $v, w \in H$. Conversely, if $\langle Uv, Uw\rangle = \langle v, w\rangle$ for all $v, w \in H$, then $\langle U^*Uv, w\rangle = \langle v, w\rangle$ for all $v, w \in H$, and thus $U^*U = I$, so $U$ is unitary. Thus the unitary operators are exactly those operators that preserve the inner product. In other words, the unitary operators are the isometries of $H$.

Let $H$ be a finite-dimensional Hilbert space and $V \subset H$ a subspace. Then $V^\perp = \{w \in H : \langle v, w\rangle = 0 \text{ for all } v \in V\}$ is a subspace of $H$ as well, called the *orthogonal complement* of $V$. We have $H = V \oplus V^\perp$, meaning that for every $x \in H$ there is a unique choice of $v \in V$ and $w \in V^\perp$ such that $x = v + w$. The map $T \colon x \to v$ is called *orthogonal projection onto* $V$. It is a linear operator on $H$ and has the property that $T^2 = T$ (it is *idempotent*). Moreover, it is self-adjoint: indeed, given $x \in H$ we have $Tx \in V$ and $x - Tx \in V^\perp$, so for every $x, y \in H$ we have

$$\langle Tx, y - Ty\rangle = \langle x - Tx, Ty\rangle = 0 \quad \Rightarrow \quad \langle Tx, y\rangle = \langle Tx, Ty\rangle = \langle x, Ty\rangle = \langle T^*x, y\rangle,$$

which implies that $T^* = T$.

**Exercise 5.8.** *Prove that if $T \in L(H)$ is idempotent and self-adjoint – that is, $T^2 = T = T^*$ – then there is a subspace $V \subset H$ such that $T$ is orthogonal projection onto $V$. (Consider the kernel and range of $T$.)*

The set of unitary $n \times n$ matrices is denoted $U(n)$. A unitary matrix with real entries is often called an *orthogonal* matrix, and the set of such $n \times n$ matrices is denoted $O(n)$. This is the set of matrices for which the corresponding linear operator on $\mathbb{R}^n$ preserves the usual Euclidean norm.

It is occasionally useful to consider the set of real $n \times n$ matrices preserving a different quantity. A *quadratic form* on $\mathbb{R}^n$ is a function

$$\begin{aligned} q \colon \mathbb{R}^n &\to \mathbb{R} \\ (x_1, \dots, x_n) &\mapsto \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j, \end{aligned}$$

where $a_{ij} \in \mathbb{R}$ are the coefficients of the form. Note that by writing $A$ for the $n \times n$ matrix with entries $a_{ij}$, this formula becomes $q(x) = \langle x, Ax\rangle$. We can assume without loss of generality that $a_{ij} = a_{ji}$ since they both contribute to the same term; in this case $A$ is symmetric (the real version of self-adjoint). Then all of its eigenvalues are real by Exercise 5.6. The *signature* of $q$ is the triple $(a, b, c) \in \{0, 1, \dots, n\}^3$ where $a$ is the number of times 0 is an eigenvalue, $b$ is the number of positive eigenvalues, and $c$ is the number of negative eigenvalues (with multiplicity). Note that $a + b + c = n$. If all eigenvalues are positive then $x \mapsto \sqrt{q(x)}$ is a norm and $(x, y) \mapsto \langle x, Ay\rangle$ is an inner product (not necessarily the standard one). But it is sometimes interesting to consider quadratic forms with other signatures, and in this case one may still work with the *isometry group* associated to the form $q$; that is, the set of real $n \times n$ matrices $A$ such that $q(Ax) = q(x)$ for all $x \in \mathbb{R}^n$.

**Exercise 5.9.** *Describe the isometry groups of the quadratic forms $q(x_1, x_2) = x_1^2 + x_2^2$ and $q'(x_1, x_2) = x_1^2 - x_2^2$ on $\mathbb{R}^2$.*

5.3. **Spectral theory and functional calculus.** Let $V$ be a finite-dimensional vector space and $T \colon V \to V$ a linear operator. Then the *spectrum* of $T$, denoted $\sigma(T)$, is the set of eigenvalues of $T$. The open set $\mathbb{C} \setminus \sigma(T)$ is called the *resolvent set* of $T$.

**Exercise 5.10.** *Prove that $\lambda \in \mathbb{C}$ is in the resolvent set if and only if $\lambda I - T$ is invertible.*

We often work with functions of matrices in the sense that we consider expressions $T^2$; but what should we make of an expression like $\sqrt{T}$? A moment's thought suggests that $\sqrt{T}$ should be a matrix $A$ such that $A^2 = T$. When does such a matrix exist? Can we make sense of $f(T)$ for other functions $f\colon \mathbb{C} \to \mathbb{C}$? This is the motivation behind the *functional calculus*, which uses the spectrum of $T$ to provide an answer in certain cases.

As a first example, observe that if $T = \left(\begin{smallmatrix} 4 & 0 \\ 0 & 9 \end{smallmatrix}\right)$, then $\sigma(T) = \{2, 3\}$ and the matrix $A = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 3 \end{smallmatrix}\right)$ has $A^2 = T$, so it seems reasonable to write $A = \sqrt{T}$. Moreover, $A$ is just the diagonal matrix we obtain by applying the square root function to each of the eigenvalues (points of the spectrum) of $T$, which lay on the diagonal of $T$. This suggests that the spectrum should play a role in a more general theory.

Suppose that $\Omega \subset \mathbb{C}$ is an open set containing $\sigma(T)$, and that $f\colon \Omega \to \mathbb{C}$ is holomorphic (differentiable as a complex function). Let $\gamma$ be a simple closed curve in $\Omega$, and $z \in \Omega$ a point such that $\gamma$ winds once around $z$ in the counterclockwise direction. Then the *Cauchy integral formula* from complex analysis says that

$$f(z) = \frac{1}{2\pi i} \int_\gamma \frac{f(\zeta)}{\zeta - z}\, d\zeta.$$

The utility of this formula for our purposes is that if $\zeta$ is in the resolvent set of $T$, then we can replace $z$ by $T$ in the integrand and obtain the meaningful expression $f(\zeta)(\zeta I - T)^{-1}$ (see Exercise 5.10). Thus if $\gamma$ wraps once around the spectrum $\sigma(T)$, then we can define

$$f(T) = \frac{1}{2\pi i} \int_\gamma f(\zeta)(\zeta I - T)^{-1}\, d\zeta. \tag{5.7}$$

An important property of this definition is that when $f(z) = \sum_{j=0}^n c_j z^j$ is a polynomial, the definition via (5.7) agrees with the natural definition $f(T) = \sum_{j=0}^n c_j z^j$. Moreover, if $f, g$ are suitable functions, then $f(T)g(T) = (f \cdot g)(T)$, where we note that 'multiplication' on the left-hand side is composition of operators, while on the right-hand side it is pointwise multiplication of functions.

There is a corresponding theory in infinite dimensions, though we will not go into it here. The starting point is to take a Banach space $X$ and a bounded linear operator $T$ on $X$, then to define the spectrum $\sigma(T)$ by first defining its complement, the resolvent set, following Exercise 5.10: we define

$$R(T) = \{z \in \mathbb{C} : zI - T \text{ has a bounded inverse}\}.$$

Then the spectrum is $\sigma(T) = \mathbb{C} \setminus \mathbb{R}(T)$. Note that $\sigma(T)$ may be composed of more than just the eigenvalues of $T$, since it is possible that $zI - T$ is bijective but that its inverse is not bounded, or that $zI - T$ is injective (hence has no eigenvalues) but not surjective.

**Exercise 5.11.** *Consider the Hilbert space $\ell^2 = \{(x_1, x_2, \dots) : \sum_k |x_k|^2 < \infty\}$ with the inner product $\langle x, y\rangle = \sum_k x_k \overline{y_k}$. Define the right shift operator $R\colon \ell^2 \to \ell^2$ by $R(x_1, x_2, \dots) = (0, x_1, x_2, \dots)$. Prove that $R$ is bounded, has no eigenvalues, and that $\sigma(R) \neq \emptyset$. What is $\sigma(R)$?*

## 6. A crash course in Lie groups

### 6.1. **Matrix Lie groups and basic examples.**

Abstractly, a *Lie group* is a smooth manifold that also has a binary operation making it a group, such that multiplication and inversion are both smooth maps. For our purposes it will be enough to think of Lie groups concretely, as follows: A *real (matrix) Lie group* is a subgroup of $GL(n, \mathbb{R})$ that is closed in the topology induced by the operator norm. In other words, a Lie group is a subset $G \subset GL(n,\mathbb{R})$ such that

(1) $G$ is a subgroup, so $AB^{-1} \in G$ whenever $A, B \in G$;
(2) $G$ is closed, so if $A_n \in G$ for all $n$ and $\lim A_n = A \in GL(n,\mathbb{R})$, then $A \in G$.

We already saw the example of $SL(n,\mathbb{R}) = \{A \in GL(n,\mathbb{R}) : \det A = 1\}$. This is a subgroup because it is the kernel of homomorphism $A \mapsto \det A$. It is closed because the map $A \mapsto \det A$ is continuous, and so the preimage of the closed set $\{1\}$ is closed.

Another important example is the set of invertible diagonal matrices

$$D_n = \{A \in GL(n,\mathbb{R}) : A_{ij} = 0 \text{ for all } i \neq j\}.$$

Let $D_n^+ = \{A \in D_n : A_{ii} > 0 \text{ for all } i\}$.

**Exercise 6.1.** *Prove that $D_n^+$ is a Lie group and that it is isomorphic to the additive abelian group $\mathbb{R}^n$. Then prove that $D_n^+ \cap SL(n,\mathbb{R})$ is a Lie group that is isomorphic to $\mathbb{R}^{n-1}$.*

A matrix $A$ is *orthogonal* if $AA^T = I$; equivalently, the rows and columns of $A$ both form orthonormal bases for $\mathbb{R}^n$.

**Exercise 6.2.** *Show that the* orthogonal group *$O(n) = \{A \in GL(n,\mathbb{R}) : A$ is orthogonal$\}$ is a Lie group, as is $SO(n) = O(n) \cap SL(n,\mathbb{R})$.*

From Exercise 6.1, we know that $\mathbb{R}$ can be realized as a Lie group in two different ways: it is isomorphic to $D_1^+ = \{\left(\begin{smallmatrix} e^x \end{smallmatrix}\right) : x \in \mathbb{R}\}$ and also to $D_2^+ \cap SL(2,\mathbb{R}) = \{\left(\begin{smallmatrix} e^x & 0 \\ 0 & e^{-x} \end{smallmatrix}\right) : x \in \mathbb{R}\}$. Here is another way to obtain $\mathbb{R}$ as a Lie subgroup of $SL(2,\mathbb{R})$: consider

$$\mathcal{U}_2 = \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbb{R} \right\}.$$

**Exercise 6.3.** *Show that $\mathcal{U}_2$ is a Lie group and that the map $\varphi \colon \mathbb{R} \to \mathcal{U}_2$ given by $\varphi(x) = \left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right)$ is an isomorphism.*

This last example has an important generalization: let $\mathcal{U}_3 \subset SL(3,\mathbb{R})$ be the set of all $3 \times 3$ matrices of the form

$$\begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix}. \tag{6.1}$$

We refer to $\mathcal{U}_3$ as the *Heisenberg group.*

**Exercise 6.4.** *Verify that $\mathcal{U}_3$ is a Lie group.*

Note that $O(n)$ and $SO(n)$ are compact Lie groups, while $SL(n,\mathbb{R})$, $\mathbb{R}^n$, and $\mathcal{U}_3(\mathbb{R})$ are non-compact. We conclude this section by mentioning one non-example.

**Exercise 6.5.** *Let $\alpha$ be irrational and consider the set $G$ of all $4 \times 4$ matrices of the form*

$$\begin{pmatrix} \cos t & \sin t & 0 & 0 \\ -\sin t & \cos t & 0 & 0 \\ 0 & 0 & \cos \alpha t & \sin \alpha t \\ 0 & 0 & -\sin \alpha t & \cos \alpha t \end{pmatrix}$$

*for some $t \in \mathbb{R}$. Show that $G$ is a subgroup of $SL(4,\mathbb{R})$ but that it is not closed.*

6.2. **Lattices and quotients.** Earlier we described the $n$-dimensional torus $\mathbb{T}^n$ as the *quotient* of $\mathbb{R}^n$ by the integer lattice $\mathbb{Z}^n$. A similar process works in other Lie groups. For example, let $\mathcal{U}_3(\mathbb{Z})$ be the set of all matrices in $\mathcal{U}_3 = \mathcal{U}_3(\mathbb{R})$ such that the entries $x, y, z$ in (6.1) all take integer values.

**Exercise 6.6.** *Show that $\mathcal{U}_3(\mathbb{Z})$ is a subgroup of $\mathcal{U}_3(\mathbb{R})$.*

To obtain the torus, we used the lattice $\mathbb{Z}^n$ to put an equivalence relation $\sim$ on $\mathbb{R}^n$; points on the torus $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ are identified with equivalence classes of $\sim$. We can use $\mathcal{U}_3(\mathbb{Z})$ to define an equivalence relation on $\mathcal{U}_3(\mathbb{R})$ in an analogous manner: given $A, B \in \mathcal{U}_3(\mathbb{R})$, say that $A \sim B$ if there is a matrix $C \in \mathcal{U}_3(\mathbb{Z})$ such that $AC = B$.

**Exercise 6.7.** *Give necessary and sufficient conditions on $x, y, z$ and $x', y', z'$ in order to have*

$$\begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & x' & y' \\ 0 & 1 & z' \\ 0 & 0 & 1 \end{pmatrix}.$$

Warning: the correct answer is **not** "$x - x' \in \mathbb{Z}$, $y - y' \in \mathbb{Z}$, $z - z' \in \mathbb{Z}$".

Each equivalence class under $\sim$ is a subset of $\mathcal{U}_3(\mathbb{R})$ – in fact, a left coset of the subgroup $\mathcal{U}_3(\mathbb{Z}) \subset \mathcal{U}_3(\mathbb{R})$ – and the set of equivalence classes is the quotient space $\mathcal{U}_3(\mathbb{R})/\mathcal{U}_3(\mathbb{Z})$. As on the torus, we can define a metric on the quotient space by

$$d([A],[B]) = \min_{C \in \mathcal{U}_3(\mathbb{Z})} \|AC - B\|,$$

where $[A] = A\mathcal{U}_3(\mathbb{Z}) = \{AC : C \in \mathcal{U}_3(\mathbb{Z})\}$ is the equivalence class of $A$. This makes $\mathcal{U}_3(\mathbb{R})/\mathcal{U}_3(\mathbb{Z})$ into a metric space; in fact, it is also a smooth manifold, just as with the torus. (This is an example of what is called a *nilmanifold*, since the group $\mathcal{U}_3(\mathbb{R})$ is *nilpotent*).

**Exercise 6.8.** *Let $F \subset \mathcal{U}_3(\mathbb{R})$ be the set of matrices as in* (6.1) *such that $x, y, z \in [0, 1]$. Show that $F$ is a fundamental domain in the sense that the sets $\{FC : C \in \mathcal{U}_3(\mathbb{Z})\}$ cover $\mathcal{U}_3(\mathbb{R})$ and overlap only along their boundaries. Conclude that the quotient space $\mathcal{U}_3(\mathbb{R})/\mathcal{U}_3(\mathbb{Z})$ can also be described by starting with the unit cube and making appropriate identifications of pairs of faces (or subsets of faces); describe these identifications.*

Rather than giving the precise general definition of "lattice", we simply remark that in both the previous examples, we were taking a quotient of a Lie group $G$ by a subgroup $\Gamma$ that was discrete in the sense that there is no element $g \in \Gamma$ that is the limit of a sequence $g_n \in \Gamma \setminus \{g\}$. Another extremely important example of a lattice is the subgroup $SL(n, \mathbb{Z}) \subset SL(n, \mathbb{R})$, and we can once again consider the quotient space $SL(n, \mathbb{R})/SL(n, \mathbb{Z})$ whose elements are the left cosets of $SL(n, \mathbb{Z})$ in $SL(n, \mathbb{R})$.

One difference between this last example and the previous two is that with $\mathbb{R}^n/\mathbb{Z}^n$ and $\mathcal{U}_3(\mathbb{R})/\mathcal{U}_3(\mathbb{Z})$, the quotient space is compact, while $SL(n, \mathbb{R})/SL(n, \mathbb{Z})$ is not compact for $n \geq 2$. To see that the first two are compact, observe that in both cases, the unit cube is a compact set that contains a representative of every coset. We say that $\mathbb{Z}^n$ is *cocompact* in $\mathbb{R}^n$, and $\mathcal{U}_3(\mathbb{Z})$ is *cocompact* in $\mathcal{U}_3(\mathbb{R})$.

**Exercise 6.9.** *Prove that $SL(n, \mathbb{R})/SL(n, \mathbb{Z})$ has infinite diameter in the quotient metric and hence is not compact.* Hint: show that $d([A], [I]) = \inf\{\|A - C\| : C \in SL(n, \mathbb{Z})\}$ is unbounded by considering the matrices $\left(\begin{smallmatrix} e^x & 0 \\ 0 & e^{-x} \end{smallmatrix}\right)$.

6.3. **Group actions.** When $n = 2$, the example $SL(2,\mathbb{R})/SL(2,\mathbb{Z})$ admits a nice visualization if we work in the hyperbolic plane $\mathbb{H}^2$. First recall that in the upper half-plane model, isometries of $\mathbb{H}(2)$ can be encoded by elements of $SL(2,\mathbb{R})$, where the matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ represents the fractional linear transformation $\varphi_A \colon z \mapsto \frac{az+b}{cz+d}$. In fact, as Exercise 2.11 showed, we have $\varphi_{AB} = \varphi_A \circ \varphi_B$.

This extra structure deserves a name: given a set $X$, write $\mathrm{Bij}(X)$ for the set of all bijections $\varphi\colon X \to X$, which is a group under composition. An *action* of a group $G$ on the space $X$ is a homomorphism $G \to \mathrm{Bij}(X)$; that is, a rule that assigns to each $g \in G$ a bijection $\varphi_g \colon X \to X$ with the property that

$$\varphi_{gh} = \varphi_g \circ \varphi_h. \tag{6.2}$$

In this case we often write $\varphi_g(x) = g.x$ (or just $\varphi_g(x) = gx$) so (6.2) becomes $(gh).x = g.(h.x)$. Technically speaking this is a *left action*; a *right action* satisfies $\varphi_{gh} = \varphi_h \circ \varphi_g$ and is often written $\varphi_g(x) = x.g$ so that $x.(gh) = (x.g).h$.

Usually we are interested not in arbitrary bijections, but in bijections preserving a particular structure; for example, if $\varphi_g$ is an isometry of $X$ for every $g \in G$, then we say "$G$ acts on $X$ by isometries". With $SL(2,\mathbb{R})$ and $\mathbb{H}^2$, we say that $SL(2,\mathbb{R})$ acts on $\mathbb{H}^2$ by fractional linear transformations. (In fact, these are isometries of $\mathbb{H}^2$.)

Recall also from Exercise 2.11 that $\varphi_A = \varphi_{-A}$. So the group of fractional linear transformations is not actually $SL(2,\mathbb{R})$, but the quotient group $PSL(2,\mathbb{R}) = SL(2,\mathbb{R})/\{\pm I\}$, where each element of $PSL(2,\mathbb{R})$ is a coset $\{A, -A\}$ for some $A \in SL(2,\mathbb{R})$.

The subgroup $PSL(2,\mathbb{Z}) = SL(2,\mathbb{Z})/\{\pm I\}$ inherits this action on $\mathbb{H}^2$, and thus it places an equivalence relation on $\mathbb{H}^2$ by saying that $x \sim y$ if and only if there is $A \in PSL(2,\mathbb{Z})$ such that $\varphi_A(x) = y$. The equivalence classes are thus the *orbits* of the $PSL(2,\mathbb{Z})$ action; subsets of $\mathbb{H}^2$ of the form $[x] = \{\varphi_A(x) : A \in PSL(2,\mathbb{Z})\}$. Figure 6 shows a fundamental domain for this action: the dark area bounded on the sides by the vertical lines $\mathrm{Re}(z) = \pm\frac{1}{2}$ and on the bottom by the unit circle $|z| = 1$. Also shown are the images of this fundamental domain under the action of various products of $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $A = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \in PSL(2,\mathbb{Z})$, which act by the fractional linear transformations $\varphi_T(z) = z + 1$ and $\varphi_A(z) = -\frac{1}{z}$. (It can be shown that $A$ and $T$ generate $SL(2,\mathbb{Z})$, but we omit this here.)

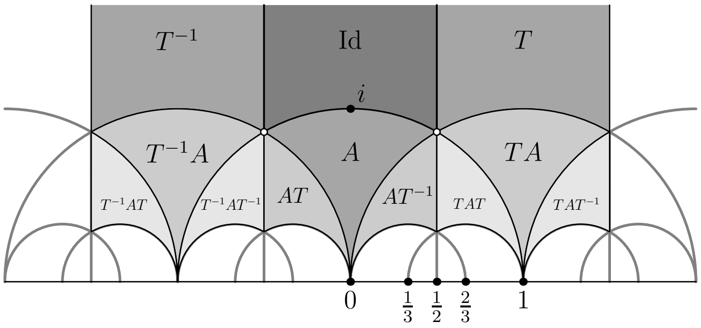

FIGURE 6. Tiling $\mathbb{H}^2$ under the action of $PSL(2,\mathbb{Z})$.

One can take the quotient of $\mathbb{H}^2$ by the action of the *modular group* $PSL(2,\mathbb{Z})$ to obtain the *modular surface*. A point on the modular surface is an orbit of $PSL(2,\mathbb{Z})$; that is, a subset of $\mathbb{H}^2$ of the form $[z] = \{\varphi_A(z) : A \in PSL(2,\mathbb{Z})\}$. Equivalently, one can view the modular surface as the fundamental domain from Figure 6 with vertical edges identified by $\varphi_T$ and bottom edges (the two arcs to the left and right of $i$) identified by $\varphi_A$. It is worth noticing that when we do this, there are two *cone points* corresponding to $i$ and $e^{i\pi/3}$; the total angle around $i$ is $\pi$, and the total angle around $e^{i\pi/3}$ is $\frac{2\pi}{3}$, instead of $2\pi$. Thus the modular surface is an example of an *orbifold*, rather than a manifold.

Note that our construction of the surface of genus 2 as a quotient space of $\mathbb{H}^2$ following Exercise 2.11 is of the same type as the modular surface construction just described here. In that case the group acting on $\mathbb{H}^2$ is not the modular group, but rather the group generated by the isometries that perform the prescribed edge identifications; this group turns out to be the fundamental group of the surface of genus 2, but as we do not discuss fundamental groups here, we set this aside for now.

6.4. **Haar measure.** Earlier, we defined Lebesgue measure $\mu$ on each Euclidean space $\mathbb{R}^n$. This measure has the property of being *translation-invariant*: given $E \subset \mathbb{R}^n$ and $x \in \mathbb{R}^n$, we have $\mu(E + x) = \mu(E)$. This condition gives some compatibility between the measure-theoretic structure on $\mathbb{R}^n$ provided by $\mu$ and the algebraic structure on $\mathbb{R}^n$ provided by addition. Moreover, Lebesgue measure turns out to be uniquely specified (up to a constant) by this condition: if $\nu$ is any translation-invariant measure on $\mathbb{R}^n$, then there is $c > 0$ such that $\nu(E) = c\mu(E)$ for all $E$.

There is a similar algebraically-significant measure on every Lie group.[8] That is, there is a measure $\mu$ on $G$ that is[9]

(1) *left-invariant*: $\mu(gE) = \mu(E)$ for all $g \in G$ and $E \subset G$, where $gE = \{gh : h \in E\}$;
(2) finite on compact sets: $\mu(K) < \infty$ for all compact $K \subset G$.

The measure $\mu$ is called *(left) Haar measure* on $G$. For $G = \mathbb{R}$, Haar measure is just Lebesgue measure (up to a scaling constant). To see how to construct Haar measure on more general groups, recall that to determine the Lie measure of a set $E$ we covered $E$ by smaller and smaller 'rectangles' whose measure was determined in a natural way. The key property that gave us translation-invariance in that setting was that two rectangles obtained from each other by translation had the same measure; thus in particular one could imagine computing the Lebesgue measure of a set $E$ by taking a very small rectangle $R$, counting how many copies of $R$ it takes to cover $E$, dividing that number by the number of copies of $R$ it takes to cover a set of known volume, and then taking a limit as the size of $R$ decreases to 0.

*Warning: the construction in the previous paragraph is rather loosely defined and trying to make it work out precisely involves some technicalities that we will not get into, since it really defines an object called a* ***content*** *from which we must then construct the measure itself. Thus you should take it rather as a general intuitive guide to motivate the following paragraph.*

To carry out the same procedure for a general matrix Lie group, we can fix a compact set $K$ with non-empty interior (think of a closed ball around the origin), and then consider a small open set $U$ that contains the identity; write $n_U(K)$ for the number of translates of $U$ that it takes to cover $K$, so

$$n_U(K) = \min\Big\{n \in \mathbb{N} : \text{there are } g_1, \dots, g_n \in G \text{ such that } K \subset \bigcup_{j=1}^n g_j U\Big\}.$$

Then given a set $E \subset G$, one may consider the ratio $n_U(E)/n_U(K)$ as somehow measuring the size of $E$, and then define $\mu_K(E) = \lim_{i\to\infty} n_{U_i}(E)/n_{U_i}(K)$, where $\operatorname{diam} U_i \to 0$ and

[8] As always, "Lie group" in this document means "matrix Lie group"; the general setting for the following construction is a "locally compact topological group".

[9] There are also some regularity conditions on the measure, but we ignore these technicalities.

we choose $U_i$ such that the limit exists; one must invoke some machinery to produce such a sequence $U_i$, and then to produce the Haar measure $\mu$ from $\mu_K$, but these are relatively standard arguments in topology and measure theory, which we omit here.

6.5. **Lie algebras.** Our main examples of Lie groups up to this point are $\mathbb{R}^n \cong D_n^+$, $SL(n,\mathbb{R})$, $SO(n)$, and $\mathcal{U}_3(\mathbb{R})$. The last of these, the Heisenberg group $\mathcal{U}_3(\mathbb{R})$, naturally contains 3 "curves" – continuous one-parameter families of matrices – given by

$$A_1(t) = \begin{pmatrix} 1 & t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad A_2(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}, \quad A_3(t) = \begin{pmatrix} 1 & 0 & t \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{6.3}$$

**Exercise 6.10.** *Show that for each $i \in \{1,2,3\}$ and $s,t \in \mathbb{R}$ we have $A_i(s)A_i(t) = A_i(s+t)$.*

**Exercise 6.11.** *Show that every $X \in \mathcal{U}_3(\mathbb{R})$ can be written as $X = A_1(x)A_2(y)A_3(z)$ for some $x,y,z \in \mathbb{R}$.*

From Exercise 6.10 we see that $A_i \colon \mathbb{R} \to \mathcal{U}_3(\mathbb{R})$ is a homomorphism. It is natural to ask if our other examples, such as $SL(n,\mathbb{R})$ and $SO(n)$, contain homomorphic images of $\mathbb{R}$. We see relatively quickly that $D_2^+$ contains the following:

$$B_1(t) = \begin{pmatrix} e^t & 0 \\ 0 & 1 \end{pmatrix}, \qquad B_2(t) = \begin{pmatrix} 1 & 0 \\ 0 & e^t \end{pmatrix}. \tag{6.4}$$

This suggests that exponentials might have something to do with a general answer.

6.5.1. *Matrix exponentials.* To define the exponential of a matrix $A$, we recall that for $x \in \mathbb{R}$, the Taylor series of $e^x$ around 0 gives

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Thus one might hope to define the exponential of a matrix by

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!} = I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \cdots.$$

One can show that this series converges for every $A \in \mathbb{M}(n,\mathbb{R})$.

**Exercise 6.12.** *Let $E_{ij}$ denote the matrix with a '1' in the $i,j$th position and all other entries equal to 0. Show that the matrices $A_i, B_i$ from (6.3) and (6.4) satisfy*

$$A_1(t) = e^{tE_{12}}, \quad A_2(t) = e^{tE_{23}}, \quad A_3(t) = e^{tE_{13}}, \quad B_1(t) = e^{tE_{11}}, \quad B_2(t) = e^{tE_{22}}.$$

*Show more generally that if $A = \left(\begin{smallmatrix} x & 0 \\ 0 & y \end{smallmatrix}\right)$, then $e^A = \left(\begin{smallmatrix} e^x & 0 \\ 0 & e^y \end{smallmatrix}\right)$.*

We conclude from the last part of the exercise that every matrix $X \in D_2^+$ is of the form $X = e^A$ for some $A \in D_2$, the set of diagonal $2 \times 2$ matrices. Notice that $D_2$ is a vector space inside $\mathbb{M}(2,\mathbb{R})$, so we obtain a subgroup of $GL(n,\mathbb{R})$ as the image of a subspace of $\mathbb{M}(2,\mathbb{R})$ under the exponential map.

Can we do the same thing for $\mathcal{U}_3(\mathbb{R})$? Exercise 6.12 suggests that we should start by letting $\mathcal{N}_3$ denote the set of all $3 \times 3$ matrices of the form

$$xE_{12} + yE_{23} + zE_{13} = \begin{pmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{pmatrix}.$$

It would be natural at this point to argue that "By Exercises 6.11 and 6.12, every $X \in U_3(\mathbb{R})$ can be written as $A_1(x)A_2(y)A_3(z) = e^{xE_{12}}e^{yE_{23}}e^{zE_{13}} = e^{xE_{12}+yE_{23}+zE_{13}}$ for some $x, y, z$, and hence every $X \in U_3(\mathbb{R})$ is of the form $X = e^A$ for some $A \in \mathcal{N}_3$." **This argument is wrong**. The problem is in the second equals sign: while the exponential function for real numbers has the property that $e^x e^y = e^{x+y}$, this is no longer true for general matrices.

**Exercise 6.13.** *Find a pair of matrices $A, B \in \mathbb{M}(2, \mathbb{R})$ such that $e^A e^B \neq e^{A+B}$.*

6.5.2. *Lie brackets.* Despite the failure of the naive approach to multiplying matrix exponentials, the conjecture that every $X \in \mathcal{U}_3(\mathbb{R})$ is of the form $X = e^A$ for some $A \in \mathcal{N}_3$ is in fact true. To prove it, we need to express $e^A e^B$ as $e^C$ for some $C$ that is given in terms of $A$ and $B$.

**Exercise 6.14.** *Prove that if $A, B$ commute then $e^A e^B = e^{A+B}$.*

Since matrices in $\mathcal{N}_3$ may not commute, we must dig deeper.

**Exercise 6.15.** *Prove that given any $A, B \in \mathcal{N}_3$, we have $e^A e^B = e^{A+B+\frac{1}{2}(AB-BA)}$.*

The key to the exercise is the observation that $E_{12}, E_{23}, E_{13}$ span $\mathcal{N}_3$, and the only non-commuting pair among these is $E_{12}, E_{23}$, for which we have $E_{12}E_{23} - E_{23}E_{12} = E_{13}$. This also means that $AB - BA \in \mathcal{N}_3$ whenever $A, B \in \mathcal{N}_3$, and we conclude that every $X \in \mathcal{U}_3(\mathbb{R})$ is of the form $X = e^A$ for some $A \in \mathcal{N}_3$.

What about our other examples, $SL(n, \mathbb{R})$ and $SO(n)$? Can they be written as $e^V$ for some subspace $V \subset \mathbb{M}(n, \mathbb{R})$? In light of Exercise 6.15, the *commutator* $[A, B] := AB - BA$ of two matrices $A, B \in \mathbb{M}(n, \mathbb{R})$ would seem to play an important role; in particular, it seems useful to require that $V$ contains $[A, B]$ whenever it contains $A$ and $B$. A linear subspace of $\mathbb{M}(n, \mathbb{R})$ satisfying this property is called a *Lie algebra*, and the commutator $[A, B]$ is often called the *Lie bracket* of $A$ and $B$.

As with Lie groups, this is really just a concrete case of a general definition: a Lie algebra is a vector space $V$ equipped with a binary operation $[\cdot, \cdot]\colon V \times V \to V$ that satisfies a list of axioms mimicking the properties of the matrix commutator.

Returning to the question of finding a Lie group as the image of a Lie algebra under the exponential map, one might hope that Exercise 6.15 holds in general. But it doesn't.

**Exercise 6.16.** *Find matrices $A, B$ such that $e^A e^B \neq e^{A+B+\frac{1}{2}[A,B]}$.*

Nevertheless, we have the following theorem (whose proof we omit).

**Theorem 6.17** (Baker–Campbell–Hausdorff)**.** *If $V$ is a Lie algebra, then for every $A, B \in V$ there is $C = C(A, B) \in V$ such that $e^A e^B = e^C$.*

In fact one can write an explicit formula for $C(A, B)$ that depends only on $A$, $B$, and iterated commutators (which must all lie in $V$ by the definition of Lie algebra), but the form of this expression is not important for us here. The important thing is that given a connected Lie group such as $SL(n, \mathbb{R})$ or $SO(n)$, there is a Lie algebra $V$ such that the Lie group is the set of all matrices of the form $e^A$ for some $A \in V$.

**Exercise 6.18.** *Show that $\det(e^A) = e^{\operatorname{Tr} A}$, and deduce that the Lie algebra for $SL(n, \mathbb{R})$ is the set of all $n \times n$ matrices with trace equal to $0$. (This Lie algebra is often written $\mathfrak{sl}(n, \mathbb{R})$.)*

**Exercise 6.19.** *Show that $\mathfrak{so}(n)$, the Lie algebra of $SO(n)$, is the set of all skew-symmetric $n \times n$ matrices.*

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