AN INTRODUCTION TO THE COURSE

Let X be a set, and let $T : X \to X$ be a map. This course is about what happens when the map T is applied repeatedly. If one takes a point x and applies T repeatedly, the resulting set $\{x, Tx, T^2x, \dots\}$ is called an orbit. Two basic questions one might ask are: is this orbit dense in X? Is it equidistributed in X?

It turns out that proper formulations of, and solutions to, these questions have applications in number theory and elsewhere, as well as having an intrinsic interest of their own.

One cannot formulate either question unless X comes with some additional structure. To ask whether the orbit $\{x, Tx, T^2x, \dots\}$ is dense one must, at the very least, put a topology on X. If that is done, then to say that the orbit is dense means that the closure $\overline{(T^n x)_{n=1}^{\infty}}$ is all of X. The map T is assumed to be continuous. In this course the set X will in fact have the structure of a compact metric space. This is a convenient assumption for us, since much of the underlying analysis is quite clean in this context. Students should be aware, however, that there are real-world applications in which such assumptions are too strong. This is particularly true of the compactness assumption.

The study of continuous maps $T : X \to X$ and their iterates is called *topological dynamics* and this will be our focus in the first third of the course.

What does it mean to say that the orbit $\{x, Tx, T^2x, \dots\}$ is equidistributed? If $A \subseteq X$ is a set, then the proportion of n for which $T^n x$ lies in A should be $\mathrm{vol}(A)/\mathrm{vol}(X)$. To make any sense of this, one must of course have a notion of the *volume* of a set A. The correct mathematical notion is that of a *measure* μ on subsets of X. The map $T : X \to X$ is assumed to be measurable (usually a somewhat weaker assumption than continuity) and measure-preserving, which means that $\mu(T^{-1}(A)) = \mu(A)$. The triple (X, μ, T) is called a measure-preserving system. The study of such systems, with particular regard to the behaviour of iterates of the map T, is the subject of *ergodic theory* and will be the main focus of the course.

There are strong parallels between topolgical dynamics and ergodic theory, and we will comment on these throughout the course. Roughly speaking ergodic theory is harder, because measures are harder to construct and harder to manipulate than topologies, but the rewards for developing the theory are higher.

1. EXAMPLES

In this section we give some examples of topological dynamical systems and measure preserving systems. Our discussion of measure-preserving systems will be rather minimal at this stage, as we intend to recall appropriate facts from measure theory before commencing a proper development of ergodic theory.

Circle rotations. Let $\alpha \in \mathbb{R}$ be a real number, let $X = \mathbb{R}/\mathbb{Z}$ be the circle, and consider the transformation $R_\alpha : X \to X$ given by $x \mapsto x + \alpha (\mathrm{mod}\, 1)$.

The $\times k$ map. The underlying space is again $X = \mathbb{R}/\mathbb{Z}$. Let $k \in \mathbb{Z}$ and write $S_k : X \to X$ for the map $x \mapsto kx (\bmod 1)$. The map S_k is clearly continuous. Perhaps surprisingly at first sight, it preserves the Lebesgue measure on $\mathbb{R}/\mathbb{Z}$. Indeed given an interval $I = [a, b)$ we see that

$$S_k^{-1}(I) = [\frac{a}{k}, \frac{b}{k}) \cup [\frac{a+1}{k}, \frac{b+1}{k}) \cup \dots \cup [\frac{a+k-1}{k}, \frac{b+k-1}{k}),$$

a collection of intervals with total length $b-a$. Note that the measure of $S_k(I)$ is *not* the same as the measure of I; it is for this reason that the definition of measure-preserving maps uses the inverse T^{-1}.

The continued fraction map. Let $X = [0, 1]$. If $x \in (0, 1]$ then define $Tx := \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$. Define $T(0) = 0$ (the inclusion of the point 0 is just to make the underlying space compact). The map T and its iterates map be used to compute the continued fraction expansion of $x \in (0, 1)$. Indeed if

$$x = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3} + \dots}} \quad \text{then} \quad Tx = \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4} + \dots}},$$

and so the nth partial quotient a_n can be recovered as $a_n = \lfloor 1/T^{n-1}x \rfloor$. The map T is certainly not continuous (it has dramatic discontinuities at $x = \frac{1}{2}, \frac{1}{3}, \dots$) but it does preserve the so-called Gauss measure defined by

$$\mu(E) := \frac{1}{\log 2} \int_E \frac{dx}{1+x}.$$

We will discuss this example in more detail later.

Rotations on tori. The first example can easily be generalised to d dimensions. Set $X = \mathbb{R}^d/\mathbb{Z}^d$, let $\alpha \in \mathbb{R}^d$, and consider the translation $x \mapsto x + \alpha$. We mention this example because orbit closures $\overline{(T^n x)_{n=1}^{\infty}}$ can take many different forms. When $d = 2$, for example, we have the following three different types of behaviour.

- Take $\alpha = (\sqrt{2}, \sqrt{3})$ and $x = (0, 0)$. Then $\overline{(T^n x)_{n=1}^{\infty}}$ is all of X, a consequence of Kronecker's theorem;
- Take $\alpha = (\frac{1}{3}, \frac{2}{5})$ and $x = (0, 0)$. Then $\overline{(T^n x)_{n=1}^{\infty}}$ is a finite set of fifteen points;
- Take $\alpha = (\sqrt{2}, \sqrt{2})$ and $x = (0, 0)$. Then $\overline{(T^n x)_{n=1}^{\infty}}$ is dense in the subtorus $\{(x, x) (\bmod \mathbb{Z}^2) : x \in \mathbb{R}\}$.

On reason we have mentioned this example is that it provides a very simple illustration of a remarkable result called *Ratner's theorem*. This states that under certain conditions, much more general than the above example, orbit closures are always "algebraic" in nature (like subtorii). We hope to at least state the theorem and discuss some of its applications later in the course.

The shift on sequences over a finite alphabet. The three examples we have presented so far have all had, as their underlying space, a compact subset of a manifold. This example is rather different, and it will feature a lot in our course because one uses it in applications to combinatorial number theory.

Take $X = \{0, 1, \dots, k-1\}^{\mathbb{Z}}$, that is to say the space of doubly-infinite words

$$\dots a_{-2} a_{-1} a_0 a_1 a_2 a_3 \dots$$

in which each a_i comes from the finite alphabet $\{0, 1, \dots, k-1\}$. X is a compact and metrizable space (it is compact by Tychonov's theorem and we will construct a metric on it later in the course). Define the shift map $T : X \to X$ by $(Tx)_n = x_{n+1}$. In other words T maps the word

$$\dots a_{-2}a_{-1}a_0a_1a_2a_3 \dots \qquad \text{to the word}$$
$$\dots a_{-1}a_0a_1a_2a_3a_4 \dots .$$

As we shall show later, T is a homeomorphism and so this system is a topological dynamical system. We shall also see that it may be given the structure of a measure-preserving system.

Baker's map. A popular dynamical system, this has underlying space $X = [0,1]^2$ and the transformation $T : X \to X$ is defined by

$$T(x, y) := (2x - \lfloor 2x \rfloor, \frac{y + \lfloor 2x \rfloor}{2}).$$

It is without best to draw a picture; the name comes from the analogy between this map on $[0,1]^2$ and theoperation of kneading bread. This turns out to be very closely related to the previous example: I have included an exercise to this effect on the example sheet.

Toral automorphisms. Take $X = \mathbb{R}^d/\mathbb{Z}^d$ and let $A \in \mathrm{SL}_d(\mathbb{Z})$, the group of $d \times d$ matrices with integer coefficients and determinant 1. Then we get a map $T_A : X \to X$ by defining $T_A(x) := Ax(\mathrm{mod}\, 1)$.

Heisenberg nilrotation. Consider the Heisenberg group $G = \begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}$ together with the discrete subgroup $\Gamma = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}$. The quotient $X = G/\Gamma$ is called a nilmanifold, and topologically it looks like the cube $[0,1]^3$ with sides identified. To see this note that, given any matrix $g := \begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix}$, there is an integer matrix $\gamma \in \begin{pmatrix} 1 & k & l \\ 0 & 1 & m \\ 0 & 0 & 1 \end{pmatrix}$ such that $g\gamma$ lies in the fundamental domain

$$\mathcal{F} := \{ \begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} : 0 \leqslant x, y, z \leqslant 1\}.$$

Indeed one may take $k = -\lfloor x \rfloor$, $m = -\lfloor z \rfloor$ and $l = -\lfloor y - x\lfloor z \rfloor \rfloor$. It is not hard to see that if g, g' are distinct points in the interior $\mathcal{F}^\circ$ then g, g' are inequivalent under right multiplication by elements of Γ.

Each $g \in G$ gives rise to a *nilrotation* $T_g : X \to X$ simply by mapping $x\Gamma$ to $gx\Gamma$. This map be given the structure of a topological dynamical system and also that of a measure-preserving system; such systems are of great importance for applications in combinatorial number theory.

Geodesic and horocycle flows on quotients of $\mathrm{SL}_2(\mathbb{R})$. (*Sketch*) Suppose that Γ is a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$; the most obvious example is $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. Then we may form the quotient G/Γ as in the previous example, and consider rotations $T_g : X \to X$ for any given $g \in G$. In general the quotient G/Γ will not be compact, and indeed when $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ it will not be as we shall recall in a moment.

In this setting one more often considers, instead of a single map T_g and its iterates T_g^n, $n \in \mathbb{Z}$, a 1-parameter flow. This is an $\mathbb{R}$-action on G/Γ rather than a $\mathbb{Z}$-action. Two important examples are

- The *horocycle flow* in which T_g^t is left multiplication by $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$;
- The *geodesic flow* in which T_g^t is left multiplcation by $\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$.

**As the names suggest, there is a geometric interpretation of these flows. This comes by recalling that $\mathrm{SL}_2(\mathbb{R})$ acts on the hyperbolic upper half-plane $\mathbb{H}$ by Möbius maps,

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

In fact the action is by isometries, and so $\mathrm{SL}_2(\mathbb{R})$ acts on the unit tangent bundle $T^1(\mathbb{H})$. The stabliser of any $v \in T^1(\mathbb{H})$ is just $\{\pm I\}$, and so $T^1(\mathbb{H})$ may be identified with $\mathrm{PSL}_2(\mathbb{R})$. Ignoring the $\pm I$ for the moment, the geodesic flow may now be visualised as taking unit tangent vectors in $\mathbb{H}$ and moving them along geodesics, which are semicircles perpendicular to $\mathbb{R}$. The horocycle flow may be visualised as moving tangent vectors along horocycles, which are circles tangent to $\mathbb{R}$.

The quotienting by Γ makes the situation interesting, because G/Γ can be thought of as the unit tangent bundle of a Riemann surface. By selecting Γ suitably, this can be arranged to be compact. When $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ the picture of the fundemental domain corresponding to such a surface is familiar (it is not compact, but it does have finite volume). With this quotienting operation in place, the flows repeatedly reach the boundaries of the fundamental domain, and the dynamics become interesting. In particular it is possible, in many cases, to establish that the flows are dense and equidistributed.**

2. Dynamical properties of circle rotations.

Although we have not made the proper definitions, it is possible and quite fun to give an answer of sorts to both of our basic questions – density and equidistribution – for the circle rotations right now.

Theorem 2.1 (Density of irrational rotations)**.** *Let $x \in X$. Then the orbit $(R_\alpha^n x)_{n=1}^\infty$ is dense if, and only if, $\alpha \notin \mathbb{Q}$.*

Proof. It is obvious that no orbit is dense if $\alpha \in \mathbb{Q}$. Indeed if $\alpha = a/q$ with $(a, q) = 1$ then the orbit $(x + n\alpha)_{n\in\mathbb{N}}$ visits precisely the points $x + r/q$, $r = 0, 1, \dots, q-1$. Suppose conversely that α is irrational. By rotational symmetry it suffices to consider the case when the basepoint x of the orbit is 0. By Dirichlet's theorem there is, for any $\varepsilon > 0$, an integer $m \geqslant 1$ such that $\|m\alpha\|_{\mathbb{R}/\mathbb{Z}} \leqslant \varepsilon$. Since α is irrational we cannot have

$\|m\alpha\|_{\mathbb{R}/\mathbb{Z}} = 0$. The iterates $m\alpha (\bmod 1), 2m\alpha (\bmod 1), 3m\alpha (\bmod 1), \dots$ clearly form a subset of $\mathbb{R}/\mathbb{Z}$ with no gaps of length ε. Since ε was arbitrary, the orbit $(n\alpha (\bmod 1))_{n\in\mathbb{N}}$ is dense.

Theorem 2.2 (Equidistribution of irrational rotations)**.** *Let $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ be a smooth function and suppose that $\alpha \in \mathbb{R}$ is irrational. Then*

$$\lim_{N\to\infty} \mathbb{E}_{n\leqslant N} f(R_\alpha^n x) = \int_0^1 f d\theta.$$

Remarks. Here, and throughout the course, we have written $\mathbb{E}_{n\leqslant N}$ for the average $\frac{1}{N}\sum_{n\leqslant N}$. From this statement one may easily prove that the proportion of $n \leqslant N$ for which $R_\alpha^n x$ returns to $[a, b)$ tends to $b-a$ as $N \to \infty$, by approximating the characteristic function $\chi_{[a.b)}$ above and below by smooth functions.

Proof. The idea is to expand f as a Fourier series:

$$f(\theta) \sim \sum_{r\in\mathbb{Z}} \widehat{f}(r) e^{2\pi i r\theta}. \tag{2.1}$$

The fact that f is smooth implies that the Fourier coefficients $\widehat{f}(r) = \int_0^1 f(\theta)e^{-2\pi i r\theta}\, d\theta$ decay very rapidly; in particular, integrating by parts twice, we have

$$\widehat{f}(r) = \frac{1}{(2\pi i r)^2}\int_0^1 f''(\theta)e^{-2\pi i r\theta}\, d\theta$$

and so $|\widehat{f}(r)| \leqslant C/|r|^2$.

This fact may be used to show that the expansion (2.1) converges uniformly to f; a proof is indicated on the first example sheet.

Once this is shown, we have, for any $\varepsilon > 0$,

$$f(\theta) = \sum_{|r|\leqslant C/\varepsilon} \widehat{f}(r)e^{2\pi i r\theta} + O(\varepsilon).$$

It follows that

$$\mathbb{E}_{n\leqslant N} f(R_\alpha^n x) = \sum_{|r|\leqslant C/\varepsilon} \widehat{f}(r)e^{2\pi i r x}\mathbb{E}_{n\leqslant N} e^{2\pi i r n\alpha} + O(\varepsilon).$$

The contribution from the term $r = 0$ is precisely $\widehat{f}(0) = \int_0^1 f\, dx$. If $r \neq 0$ we have, by summing a G.P.,

$$|\mathbb{E}_{n\leqslant N} e^{2\pi i r n\alpha}| \leqslant \frac{2}{N|1 - e^{2\pi i r\alpha}|}.$$

This tends to zero as $N \to \infty$.

It follows that

$$\lim_{N\to\infty} |\mathbb{E}_{n\leqslant N} f(R_\alpha^n x) - \int_0^1 f\, d\theta| \leqslant \varepsilon;$$

since ε was arbitrary, the result follows. □