20. ORIGINS AND SETUP OF ERGODIC THEORY

In the remaining lectures this quarter, we will move attention to the subject of Ergodic Theory. We will start with motivational remarks on the origins of this theory that go back to the work of L. Boltzmann in statistical mechanics.

20.1 Hamiltonian flow and Boltzmann's ergodic hypothesis.

In mid-to-late 17th century, I. Newton cast the laws of classical mechanics in the form of second order differential equations for functions describing motion of physical objects in time. A key concept was that of a *force* which, however, is loosely defined and, in fact, can largely be given a meaning only through Newton's laws themselves. The only kind of force in Newton's theory that is not of such "kinematic type" is the force of gravity. This force is also of a potential form, meaning that the gravitational pull exerted by a celestial body on an object of a unit mass is a gradient of a potential function. (Another such force is the electrostatic force, but that had not yet been discovered.)

For objects interacting via potential forces, in 1833 W.R. Hamilton re-cast Newton's laws (reformulated earlier via variational calculus by J.L. Lagrange) in terms of first-order differential equations for two kinds of quantities:

- *positions* $(q_1,\dots,q_n)$
- *momenta* $(p_1,\dots,p_n)$

Here, typically, $n = 3N$ where N is a number of, say, gas particles in a container. In this interpretation $(q_{3i}, q_{3i+1}, q_{3i+2})$ are the Cartesian coordinates and $(p_{3i}, p_{3i+1}, p_{3i+2})$ are the coordinates of the velocity of the i-th particle multiplied by its mass. The time evolution of such a system is then a trajectory in $\mathbb{R}^{2n}$.

For a general class of system with potential forces, the evolution is determined by *Hamilton's equations*: We assume that there exists a continuously differentiable function $H\colon \mathbb{R}^{2n} \to \mathbb{R}$ such that $q_1(t),\dots,q_n(t)$ and $p_1(t),\dots,p_n(t)$ obey the ODEs

$$\forall k = 1,\dots,n\colon \quad \frac{\mathrm{d}q_k}{\mathrm{d}t} = \frac{\partial H}{\partial p_k} \ \wedge \ \frac{\mathrm{d}p_k}{\mathrm{d}t} = -\frac{\partial H}{\partial q_k} \tag{20.1}$$

A typical example of the *Hamiltonian* H is the function

$$H(q,p) := \sum_{i=1}^{n} \frac{p_i^2}{2} + V(q_1,\dots,q_n) \tag{20.2}$$

The two terms on the right represent the kinetic energy and potential energy.

Under the assumption that $H \in C^2(\mathbb{R}^{2n})$, a solution

$$x(t) \ := \ (q_1(t),\dots,q_n(t),p_1(t),\dots,p_k(t)) \tag{20.3}$$

to (20.1) with initial condition $x(0) = (q_1(0),\dots,q_n(0),p_1(0),\dots,p_n(0))$ is locally unique (meaning that the map $x(0) \mapsto x(t)$ is injective in a neighborhood of $x(0)$ for t sufficiently small). We thus have a one-parameter family of maps $T_t\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ defined by

$$T_t\big(x(0)\big) = x(t) \tag{20.4}$$

We refer to this family as the *Hamiltonian flow*. Here is a key observation:

Theorem 20.1 (Liouville's theorem) *T_t preserves the Lebesgue measure on $\mathbb{R}^{2n}$.*

Proof. This is a beautiful exercise in multivariable calculus. It suffices to show that the Jacobian J_t of the transformation T_t

$$J_t(x(0)) = \det\Big\{\frac{\partial x_i(t)}{\partial x_j(0)}\Big\}_{i,j=1}^n. \tag{20.5}$$

equals 1 at all t. Moreover, because $x(t)$ is a solution to the above equations, the Jacobian has the following semigroup property:

$$J_{t+\delta}(x) = J_t(x)J_\delta(x(t)). \tag{20.6}$$

It thus suffices to show that $\frac{\partial}{\partial t}J_t(x) = 0$ at $t = 0$ and all x. To this end we note that $x_i(t) = x_i(0) + tV_i(x(0)) + o(t)$ where V is the vector field

$$V(x) = \Big(\frac{\partial H}{\partial p_1}, \dots, \frac{\partial H}{\partial p_n}, -\frac{\partial H}{\partial q_1}, \dots, -\frac{\partial H}{\partial q_n}\Big). \tag{20.7}$$

Therefore, $\frac{\partial x_i(t)}{x_j(0)} = \delta_{ij} + t\frac{\partial V_i}{\partial x_j}(x(0)) + o(t)$ and so

$$J_t(x) = \det\big(\delta_{ij} + t(\nabla V)_{ij}(x) + o(t)\big). \tag{20.8}$$

But $\det(1 + tA) = 1 + t\mathrm{Tr}A + o(t)$ and $\mathrm{Tr}(\nabla V) = \nabla \cdot V$ which vanishes by equality of second mixed partials. Hence $J_t = 1 + o(t)$, i.e., $\frac{\partial}{\partial t}J_t$ vanishes at $t = 0$. $\square$

Just as Newtonian mechanics, the Hamiltonian mechanics is reversible — solving the equations backwards allows us to recover the initial configuration from the present state. This, however, does not quite match the observation that many processes in the world are quite apparently irreversible. Moreover, the systems of physical interest — say, the gas particles in a container — typically have an extremely large number of constituents and, as later P.T. Ehrenfest interpreted L. Boltzmann's ideas, a typical trajectory of such systems winds around rapidly filling all "corners" of the configuration space. In order to make this consistent with earlier conclusions about "distribution of velocities" made by J.C. Maxwell, the following conjecture was made:

Conjecture 20.2 (Boltzmann's Ergodic Hypothesis) *The time averages*

$$\frac{1}{T}\int_0^T f(q_t, p_t)\mathrm{d}t \tag{20.9}$$

can be replaced, in the limit $T \to \infty$, by space averages

$$\int f(q, p)\mu(\mathrm{d}q\mathrm{d}p) \tag{20.10}$$

for a suitable (invariant) measure.

The departure from the trajectory point of view of classical mechanics gave birth to statistical mechanics (which is a physical theory that regards the state of the system as a random variable or a measure). However, it was soon realized (and excessively debated) that the conjecture has at least two intrinsic problems:

(1) The existence of the limit (20.9) is *a priori* unclear, especially once we learn (as we will soon do) that the system keeps coming back to the vicinity of its initial state.
(2) The invariant measure is generally not unique. The problem is inherent to the definition of the Hamiltonian flow. Indeed, the Hamiltonian H is constant along the trajectories of the system. To see this, let $F \in C^2(\mathbb{R}^n)$. Plugging in the solution $x(t)$ into F we now compute its (total) derivative with respect to t to be

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \sum_i \Big(\frac{\partial F}{\partial q_i}\frac{\mathrm{d}q_i}{\mathrm{d}t} + \frac{\partial F}{\partial p_i}\frac{\mathrm{d}p_i}{\mathrm{d}t}\Big) = \{F, H\} \tag{20.11}$$

where

$$\{A, B\} := \sum_i \Big(\frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q_i}\Big) \tag{20.12}$$

is the so called *Poisson bracket*. Since $\{F, H\} = -\{H, F\}$, we get

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \{H, H\} = 0 \tag{20.13}$$

The flow will thus not leave the set $\{H = E\}$, where E is the value of H at time zero. By a compactness argument, if H has compact level sets, for each E there will be a invariant probability measure concentrated on $\{H = E\}$. (Other "conservation laws" may complicate this further.)

While the full proof of the Ergodic Hypothesis is still elusive, intensive work by mathematicians in the first half of the 20th century shed much new light into the problem. Indeed, now we know that the limit of the time averages exists under very general circumstances and that the hard part of the problem is thus the classification of "irreducible" invariant measures.

20.2 Measure preserving transformations.

We will now move back to mathematics and start developing the foundations of the ergodic theory. In light of Liouville's theorem, much of the theory deals with transformations of measure spaces that preserve the underlying measure.

Definition 20.3 (Measure preserving transformation) *Given a measure space $(\mathscr{X}, \mathcal{G}, \mu)$, a map $\varphi\colon \mathscr{X} \to \mathscr{X}$ is a* measure-preserving transformation*, to be abbreviated m.p.t., if φ is $\mathcal{G}$-measurable ($\varphi^{-1}(\mathcal{G}) \subseteq \mathcal{G}$) and preserves μ (i.e., $\mu \circ \varphi^{-1} = \mu$).*

The object $(\mathscr{X}, \mathcal{G}, \mu, \varphi)$, where φ is an m.p.t. on $(\mathscr{X}, \mathcal{G}, \mu)$, will sometimes be called a *measure preserving system*, to be abbreviated m.p.s. If φ is invertible and both φ and φ^{-1} are m.p.t.'s, we call φ a *measure-preserving bijection*.

Our first examples of measure-preserving systems arise from stationary sequences of random variables defined in:

Definition 20.4 *A sequence $(X_n)_{n\geqslant 0}$ of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be* stationary *if for all $k, n \geqslant 0$, $(X_k, \dots, X_{k+n})$ has the same law as $(X_0, \dots, X_n)$. A similar definition applies to two-sided sequences $(X_n)_{n\in\mathbb{Z}}$.*

Here is the precise sense in which stationary sequences can be interpreted as measure-preserving transformations:

Lemma 20.5 *Let $\{X_n\}_{n\in\mathbb{N}}$ be a stationary sequence on $(\Omega, \mathcal{F}, P)$ taking values in a standard-Borel space $\mathcal{S}$. Let θ denote the left shift on $\mathcal{S}^{\mathbb{Z}}$ defined by*

$$\theta\big(\{x_n\}_{n\in\mathbb{Z}}\big) := \theta\big(\{x_{n+1}\}_{n\in\mathbb{Z}}\big) \tag{20.14}$$

There exists a unique probability measure μ on $(\mathcal{S}^{\mathbb{N}}, \mathcal{B}(\mathcal{S}^{\mathbb{N}}))$ and there exists a measurable map $X\colon \mathcal{S}^{\mathbb{N}} \to \mathcal{S}$ such that

$$\mu \circ \theta^{-1} = \mu \tag{20.15}$$

and

$$\{X_n\}_{n\in\mathbb{N}} \overset{\text{law}}{=} \{X \circ \theta^n\}_{n\in\mathbb{N}} \tag{20.16}$$

In particular, every one-sided stationary sequence of random variables taking values in a standard Borel space is a restriction of a two-sided stationary sequence.

Proof. Abbreviate $[-n, n] := \{-n, \dots, n\}$ and define μ_n on $(\mathcal{S}^{[-n,n]}, \mathcal{B}(\mathcal{S}^{[-n,n]}))$ by

$$\mu_n(A) := \mathbb{P}\big((X_1, \dots, X_{2n+1}) \in A\big), \qquad A \in \mathcal{S}^{[-n,n]} \tag{20.17}$$

By stationarity of $\{X_n\}_{n\geqslant 0}$, the measures $\{\mu_n\}_{n\geqslant 0}$ are consistent and so, by the Kolmogorov Extension Theorem, they are restrictions of a unique probability measure on $(\mathcal{S}^{\mathbb{Z}}, \mathcal{B}(\mathcal{S}^{\mathbb{Z}}))$. The left-shift θ is continuous in product topology and hence measurable; same applies to its inverse, the right-shift. By stationarity, $\mu \circ \theta^{-1}(A) = \mu(A)$ for all cylinder sets which by Dynkin's π-λ Theorem extends to all $A \in \mathcal{B}(\mathcal{S}^{\mathbb{Z}})$. Finally, consider the random variable $X(\omega) = \omega_0$. The construction implies that $(X, X\circ\theta, X\circ\theta^2, \dots, X\circ\theta^n)(\omega) = (\omega_0, \dots, \omega_n)$ has the same law as $(X_0, \dots, X_n)$, for all $n \geqslant 0$, proving (20.16) by Dynkin's π/λ-theorem. $\square$

Let us now give some examples of measure preserving systems arising from stationary sequences of random variables.

***Example* 20.6** $\{X_n\}_{n\geqslant 0}$ i.i.d. The special case of X_n taking a finite number of values, most typically, $X_n \in \{0, 1\}$, is referred to as *Bernoulli shift*.

***Example* 20.7** $\{X_n\}_{n\geqslant 0}$ stationary Markov chain; i.e., the chain started from a stationary measure. (We will discuss these in 275C.)

***Example* 20.8** $\{X_n\}_{n\in\mathbb{Z}}$ stationary Gaussian. Let $\mu \in \mathbb{R}$ and let $F\colon \mathbb{Z} \to \mathbb{R}$ be a function of the form

$$F(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{F}(k)\mathrm{e}^{\mathrm{i}kx}\mathrm{d}k \tag{20.18}$$

where $\hat{F}$ is non-negative and integrable. Let $\{X_n\}_{n\in\mathbb{Z}}$ be Gaussian with mean μ and covariance $C(x, y) = F(x - y)$. Then $\{X_n\}_{n\in\mathbb{Z}}$ is stationary with respect to left shifts and thus defines a measure preserving system as specified above.

Next let us discuss examples that do not obviously come from a random sequence:

***Example* 20.9** (Rotation of a circle) Here $\mathscr{X} = [0,1)$ which we think of as a circle, μ is Lebesgue measure and $\varphi(x) := x + \alpha \bmod 1$ for some fixed α. (The most interesting cases will arise from $\alpha \notin \mathbb{Q}$.) As it turns out, φ preserves the Lebesgue measure. Using the Minkowski notation $a + B := \{a + b \colon b \in B\}$ along with the fact that $a + B$ and B have the same Lebesgue measure, this is seen from

$$\varphi^{-1}(A) = \Big(-\alpha + \big(A \cap [\alpha, 1)\big)\Big) \cup \Big(1 - \alpha + A \cup [0, \alpha)\Big) \tag{20.19}$$

and noting that the union is disjoint.

***Example* 20.10** (Continued fractions) Here $\mathscr{X} = [0,1] \smallsetminus \mathbb{Q}$. Each $x \in \mathscr{X}$ is in one-to-one correspondence with an infinite sequence $(a_1, a_2, \dots)$ of positive naturals via the continued-fraction expansion

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \frac{1}{a_3 + \dots}}}. \tag{20.20}$$

Since $a_1 = \lfloor 1/x \rfloor$ and $a_{k+1} = \lfloor 1/x_k \rfloor$ where, inductively, $x_{k+1} = 1/x_k - a_k$, the sequence never "terminates," i.e., $a_k \geqslant 1$ for all k. The left shift $(a_1, a_2, \dots) \mapsto (a_2, a_3, \dots)$ is then pulled onto onto $\mathscr{X}$ as

$$\varphi(x) := x^{-1} - \lfloor x^{-1} \rfloor \tag{20.21}$$

It turns out that φ is measurable and preserves the measure $\mu(\mathrm{d}x) := 1_{\mathscr{X}}(x) \frac{\mathrm{d}x}{1+x} \mathrm{d}x$. This is because

Lemma 20.11 *Let X be sampled from the distribution $\tilde{\mu}(\mathrm{d}x) := \frac{1}{\log 2} 1_{[0,1] \smallsetminus \mathbb{Q}}(x) \frac{\mathrm{d}x}{1+x} \mathrm{d}x$ and let $Z_1, Z_2, \dots$ be the unique positive integers such that*

$$X = \cfrac{1}{Z_1 + \cfrac{1}{Z_2 + \frac{1}{Z_3 + \dots}}} \tag{20.22}$$

Then $\{Z_k\}_{k \geqslant 1}$ are i.i.d.

We leave the proof of this to a homework exercise.

***Example* 20.12** (Baker's transform) Let $\mathscr{X} := [0,1) \times [0,1)$ and let $\varphi \colon \mathscr{X} \to \mathscr{X}$ be defined by

$$\varphi(x,y) = \begin{cases} (2x, \frac{1}{2}y), & \text{if } 0 \leqslant x < 1/2, \\ (2x - 1, \frac{1}{2}(y+1)), & \text{if } 1/2 \leqslant x < 1. \end{cases} \tag{20.23}$$

This transformation deforms the left half of the 1×1-square onto its bottom half and its right half onto its top half. (The name arises from this resembling the move bakers mix bread dough, except that the top half is not cut an stacked up but rather flipped over.) Since the transformation scales x coordinate up by the same number as it scales the y coordinate down, it preserves the Lebesgue measure on $\mathscr{X}$.

Further reading: Durrett, Chapter 6.1