ERGODIC THEOREMS

The ultimate goal of this lecture will be to provide a proof of the Birkhoff ergodic theorem, following the approach via the maximal ergodic inequality initiated by J. Bourgain. To implement this approach we shall also require the more elementary *mean ergodic theorem* due to Von Neumann, which is of interest in its own right. Our presentation here is similar to the one in *Einsiedler-Ward, Ergodic Theory.*

1. The role of $L^2(X, dm)$.

Let $\big(X, m\big)$ be a finite measure space, and consider the Hilbert space $L^2(X, dm)$. Let $T : X \longrightarrow X$ a measure preserving map, not necessarily ergodic. We can then consider the operator

$$U_T f := f \circ T.$$

Since T is measure preserving, this operator is an *isometry*, i. e.

$$\big\|U_T f\big\|^2_{L^2(X,dm)} = \int_X \big|f \circ T\big|^2 \, dm = \int_X \big|f\big|^2 \, dm = \big\|f\big\|^2_{L^2(X,dm)}$$

From now on, we shall simply write $\big\| \cdot \big\|$ instead of $\big\| \cdot \big\|_{L^2(X,dm)}$.
Our goal is to understand the averages

$$A_N(f) := N^{-1} \sum_{j=0}^{N-1} f\big(T^j x\big)$$

We observe that formally, we expect the limit of $A_N(f)$ as $N \to \infty$, to be *invariant under* T. This motivates the introduction of the space

$$V := \{g \in L^2(X, dm), \;\; U_T g = g\}.$$

Since U_T is a bounded, and hence continuous linear operator on $L^2(X, dm)$, the space V is a closed subspace.

Next, we introduce the space

$$W := \overline{\{f - U_T f, \; f \in L^2(X, dm)\}},$$

i. e. the *closure* in $L^2(X, dm)$ of the sub space $\{f - U_T f, \; f \in L^2(X, dm)\}$. These two subspaces actually complement each other, in the sense that

$$L^2(X, dm) = V \oplus W.$$

This follows from

Lemma 1.1. *We have that $V = W^\perp$.*

Proof. We first note that if $g \in V$, we have for any $f \in L^2(X, dm)$ that

$$\begin{aligned}
\langle g, f - U_T f\rangle &= \langle g, f\rangle - \langle g, U_T f\rangle \\
&= \langle g, f\rangle - \langle U_T g, U_T f\rangle \\
&= \langle g, f\rangle - \langle g, f\rangle \\
&= 0.
\end{aligned}$$

Since the space $\{f - U_T f, \; f \in L^2(X, dm)\}$ is dense in W, this implies that $V \subset W^\perp$. Next, assume that

$$\langle g, f - U_T f\rangle = 0 \, \forall f \in L^2(X, dm).$$

This implies that

$$\langle g - U_T^* g, f\rangle = 0 \, \forall f \in L^2(X, dm),$$

which results in

$$g - U_T^* g = 0.$$

But then we have

$$\begin{aligned} 0 &= \left\|g - U_T^* g\right\|^2 \\ &= 2\left\|g\right\|^2 - 2\,\mathrm{Re}\,\langle g, U_T^* g\rangle \\ &= 2\left\|g\right\|^2 - 2\,\mathrm{Re}\,\langle U_T g, g\rangle \end{aligned}$$

The equality case for the Cauchy-Schwarz inequality then implies that

$$U_T g = g.$$

Hence $W^\perp \subset V$. □

2. The Mean Ergodic Theorem of Von Neumann

Here we prove the following result:

Theorem 2.1. *As before let (X, m) be a finite measure space and $T : X \longrightarrow X$ be measure preserving. Let P_T be the orthogonal projection onto the subspace V from the preceding section. The for any $f \in L^2(X, dm)$, we have the limiting relation*

$$\lim_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} f \circ T^j = P_T f.$$

Thus the limit exists in $L^2(X, dm)$ and is invariant under T. In particular, if T is ergodic, we have the relation

$$\lim_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} f \circ T^j = \frac{1}{m(X)} \int_X f\, dm,$$

where the right hand side is a constant, and the limit is in the sense of $L^2(X, dm)$.

Proof. The preceding lemma implies that it suffices to verify the limiting relation for two cases: (i) $f \in V$, and (ii) $f \in W$.

(i) $f \in V$. Here we have $U_T f = f$ and hence $f \circ T^j = f$, $j \geq 0$. It follows that

$$N^{-1} \sum_{j=0}^{N-1} f \circ T^j = f \,\forall N \geq 1,$$

and so the claim is obvious.

(iI) $f \in W$. First assume that $f = g - U_T g$, $g \in L^2(X, dm)$. Then we compute

$$f \circ T^j = g \circ T^j - g \circ T^{j+1},\ j \geq 0.$$

This implies that we arrive at a telescoping sum

$$\begin{aligned} N^{-1} \sum_{j=0}^{N-1} f \circ T^j &= N^{-1}\Big(g - g \circ T + g \circ T - g \circ T^2 + \ldots + g \circ T^{N-1} - g \circ T^N\Big) \\ &= N^{-1}\big(g - g \circ T^N\big). \end{aligned}$$

But then we conclude that

$$\left\|N^{-1} \sum_{j=0}^{N-1} f \circ T^j\right\| \leq \frac{2}{N}\left\|g\right\|,$$

which converges toward 0. Moreover,

$$P_T f = 0.$$

More generally, assume that $f \in W$. Then there is a sequence of $g_i \in L^2(X, dm)$ such that

$$\big\|f - \big(g_i - U_T g_i\big)\big\| \longrightarrow 0$$

as $i \to +\infty$. Thus given $\varepsilon > 0$, there is some i_0 such that

$$\big\|f - \big(g_{i_0} - U_T g_{i_0}\big)\big\| < \varepsilon.$$

Then pick N large enough such that

$$\Big\|N^{-1}\sum_{j=0}^{N-1}\big(g_{i_0} - U_T g_{i_0}\big)\circ T^j\Big\| < \varepsilon.$$

Then we conclude that

$$\begin{aligned}
&\big\|N^{-1}\sum_{j=0}^{N-1} f\circ T^j\big\| \\
&\leq \big\|N^{-1}\sum_{j=0}^{N-1}\big(f - (g_{i_0} - U_T g_{i_0})\big)\circ T^j\big\| + \Big\|N^{-1}\sum_{j=0}^{N-1}\big(g_{i_0} - U_T g_{i_0}\big)\circ T^j\Big\| \\
&\leq \varepsilon + \varepsilon \\
&= 2\varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, we conclude that

$$\lim_{N\to\infty} N^{-1}\sum_{j=0}^{N-1} f\circ T^j = 0 = P_T f.$$

□

3. The key tool towards the Birkhoff Ergodic Theorem: maximal ergodic inequality

While the Mean Ergodic Theorem may seem quite close to the Birkhoff one, the *pointwise almost everywhere convergence* is indeed quite a bit more difficult to obtain. To appreciate this, recall that if a sequence of functions $\{f_n\}_{n\geq 1}$ converges in the sense of L^2 towards a limit f in L^2, then it is not in general possible to conclude the convergence of the f_n to f almost everywhere. In fact, this is in general only possible *for a subsequence of the* $\{f_n\}_{n\geq 1}$. For comparison purposes, the convergence of the Fourier series of a function $f \in L^2(S^1)$ towards f in the L^2-sense is quite elementary by comparison to the extremely difficult proof of almost everywhere convergence (due to L. Carleson 1966).

The key tool for the pointwise convergence is typically a so-called *maximal inequality*, which derives bounds not just for individual functions of an approximating sequence, but for the maximal function associated for such a sequence. What this means becomes clear in the following

Theorem 3.1. *Let $\big(X, m\big)$ a finite measure space, and $T : X \longrightarrow X$ measure preserving. For $g \in L^1(X, dm)$ a real-valued function, define*

$$E_\alpha := \{x \in X \mid \sup_{N\geq 1}\frac{1}{N}\sum_{j=0}^{N-1} g\big(T^j x\big) > \alpha\}.$$

Then the following estimate obtains:

$$\alpha\cdot m\big(E_\alpha\big) \leq 3\big\|g\big\|_{L^1(X,dm)}.$$

Remark 3.2. The key point in this inequality is of course the fact that we include the *supremum over all* $N > 0$ on the left. If we simply fixed a particular $N > 0$, then taking advantage of the elementary inequality

$$\big|\{g > \alpha\}\big| < \alpha^{-1}\cdot\big\|g\big\|_{L^1}$$

for a real valued function and $\alpha > 0$, we infer that

$$\Big|\{x \in X \mid \frac{1}{N}\sum_{j=0}^{N-1} g\big(T^j x\big) > \alpha\}\Big| < \alpha^{-1} \cdot \Big\|\frac{1}{N}\sum_{j=0}^{N-1} g\big(T^j x\big)\Big\|_{L^1(X,dm)}$$
$$\leq \alpha^{-1} \cdot \big\|g\big\|_{L^1(X,dm)},$$

where the last inequality results from the fact that T is measure preserving.

The proof of this theorem will follow Bourgain's strategy by taking advantage of a classical result in geometric measure theory, the *Vitali covering lemma.* In fact, we shall use this in a somewhat unusual way.

Here is an abstract version of the Vitali covering lemma in the context of a metric space:

Lemma 3.3. *Let $\big(X, d\big)$ be a metric space and let $\{B(a_i, r_i)\}_{i=1}^N$ a finite collection of closed balls. Then there is a disjoint subcollection of balls*

$$\{B(a_{i_l}, r_{i_l})\}_{l=1}^K \subset \{B(a_i, r_i)\}_{i=1}^N$$

such that

$$\bigcup_{i=1}^N B(a_i, r_i) \subset \bigcup_{l=1}^K B(a_{i_l}, 3r_{i_l}).$$

Proof. We may assume that $r_1 \geq r_2 \geq \ldots \geq r_N$. Then set

$$i_1 = 1.$$

Proceeding inductively, if $i_1, \ldots, i_k$ have been chosen, we continue as follows: either all $B(a_i, r_i)$ with $i > i_k$ intersect one of the $B(a_{i_l}, r_{i_l})$, $1 \leq l \leq k$, or else we pick $i_{k+1} > i_k$ minimal and such that

$$B(a_{i_{k+1}}, r_{i_{k+1}})$$

is disjoint from all the previously chosen balls. This process stops after finitely many steps, resulting in the $\{B(a_{i_l}, r_{i_l})\}_{l=1}^K$.

We claim that this collection of balls has the desired property. For let $B(a_i, r_i)$ be one of the original balls. If it is one of the chosen ones, it is clearly contained in the set

$$\bigcup_{l=1}^K B(a_{i_l}, 3r_{i_l}).$$

If it is not one of the chosen ones, then one of the $B(a_{i_l}, r_{i_l})$ with $i_l < i$ needs to intersect it. For if not, letting $\{i_1, \ldots, i_p\}$ be the chosen indices $< i$, then i would have been chosen at the $p+1$-th stage of the selection process.
Then assuming that

$$B(a_{i_l}, r_{i_l}) \cap B(a_i, r_i) \neq \emptyset, i_l < i,$$

then since $r_{i_l} \geq r_i$, we infer that

$$B(a_i, r_i) \subset B(a_{i_l}, 3r_{i_l}).$$

□