PROOF OF THE BIRKHOFF ERGODIC THEOREM

We follow the strategy initiated by Bourgain, relying on the maximal ergodic inequality stated in the last lecture. We prove this first, relying on the Vitali covering lemma. Throughout we assume that g is real valued.

1. Proof of the maximal ergodic inequality

We follow here the presentation of Einsiedler-Ward. To begin with, we start with the following observation, which is a direct application of the Vitali covering lemma:

Lemma 1.1. *Let $\{[a_i, a_i + l_i - 1]\}_{i=1}^N$, $l_i > 0$, be a collection of intervals in $\mathbb{Z}$. Then there is a disjoint sub-collection of intervals*

$$\{[a_{i_k}, a_i + l_{i_k} - 1]\}_{k=1}^K$$

such that

$$\begin{aligned}\bigcup_{i=1}^N [a_i, a_i + l_i - 1] &\subset \bigcup_{k=1}^K [a_{i_k} - l_{i_k} + 1, a_i + 2l_{i_k} - 2] \\ &\subset \bigcup_{k=1}^K [a_{i_k} - l_{i_k}, a_i + 2l_{i_k} - 1].\end{aligned}$$

This is immediate from the Vitali lemma, by interpreting the intervals as subintervals of $\mathbb{R}$.

Let now (X, m) be a finite measure space, $T : X \longrightarrow X$ measure preserving, and $g \in L^1(X, dm)$. For technical reasons, we shall assume that g only takes finite values. This can be achieved by replacing g by $\chi_{|g|<M} \cdot g$ and eventually passing to the limit $M \to +\infty$. As all the bounds derived below will be independent of M, the maximal ergodic inequality will then follow without the cutoff.

Recall that we intend to show that with

$$g^*(x) := \sup_{N\geq 1} N^{-1} \sum_{j=0}^{N-1} g(T^j(x))$$

and $\alpha > 0$, the set

$$E_\alpha := \{g^*(x) > \alpha\}$$

has measure $m(E_\alpha) \leq 3\alpha^{-1} \big\|g\big\|_{L^1(X,dm)}$. The strategy for proving this shall use the following trick: replacing x by $T^a x$ for some $a \geq 0$, we get

$$T^{-a}\big(E_\alpha\big) = \{g^*\big(T^a x\big) > \alpha\} = \{\sup_{N\geq 1} N^{-1} \sum_{j=0}^{N-1} g(T^{a+j}(x)) > \alpha\}.$$

By measure preservation, we have

$$m\big(T^{-a}\big(E_\alpha\big)\big) = m(E_\alpha).$$

Furthermore, we can write

$$m\Big(\{\sup_{N\geq 1} N^{-1} \sum_{j=0}^{N-1} g(T^{a+j}(x)) > \alpha\}\Big) = \int_X \chi_{T^{-a}\big(E_\alpha\big)}\, dm$$

The idea is now to sum over a $a \in [0, 1, \ldots, J-1]$ for some large J, and to observe that *for fixed x*

$$\sum_a \chi_{T^{-a}\big(E_\alpha\big)}(x) = \Big|\{a \mid \sup_{N\geq 1} N^{-1} \sum_{j=0}^{N-1} g(T^{a+j}(x)) > \alpha\}\Big|$$

The right hand side, when viewing $\sup_{N\geq 1} N^{-1}\sum_{j=0}^{N-1} g(T^{a+j}(x))$ as a function of a, is the cardinality of the set where this maximal function is large. This is quite analogous to similar maximal functions used for example to prove the Lebesgue differentiation result, and one can estimate the cardinality by using the Vitali lemma, in terms of an expression *not involving the supremum*. At the same time, by measure preservation, we have that

$$\sum_a m\big(T^{-a}(E_\alpha)\big) = \big|\{a\}\big| \cdot m(E_\alpha).$$

Combining these observations will allow us to bound $m(E_\alpha)$ in terms of $\big\|g\big\|_{L^1(X,dm)}$, after carrying out the x-integration.

We shall now rigorously implement this strategy. In order to do this, we shall for now *fix* $x \in X$ and consider a function on $\mathbb{Z}$ introduced as follows:

For a very large $J \geq 1$, where eventually we shall let $J \longrightarrow +\infty$, we set

$$\phi(j) := g\big(T^j x\big),\ j = 0, \ldots, J$$

and $\phi(j) = 0$ otherwise. This is a well-defined function on $\mathbb{Z}$ taking real values and such that $\phi \in l^1(\mathbb{Z})$. The following is then the essential technical ingredient, namely the *maximal function estimate*:

Lemma 1.2. *(Maximal function) Let*

$$\phi^*(a) := \sup_{N\geq 1} N^{-1}\sum_{j=0}^{N-1}\phi(a+j).$$

Define (for $\alpha \geq 0$)

$$E_\alpha^\phi := \{a \in \mathbb{Z} \,|\, \phi^*(a) > \alpha\}.$$

Then we have

$$\alpha\big|E_\alpha^\phi\big| \leq 3 \cdot \big\|\phi\big\|_{l^1(\mathbb{Z})}.$$

Proof. Let $a_1, a_2, \ldots, a_K$ be distinct integers in E_α^ϕ. We show that αK is bounded by $3 \cdot \big\|\phi\big\|_{l^1(\mathbb{Z})}$, using the Vitali lemma 1.1. By definition, for each $i \in \{1, \ldots, K\}$, there exist $l_j \geq 1$ and such that

$$l_j^{-1} \cdot \sum_{i=0}^{l_j-1}\phi(a_j + i) > \alpha.$$

By lemma 1.1, we can pick a subcollection $a_{i_1}, \ldots, a_{i_R}$ such that the intervals $[a_{i_p}, a_{i_p} + l_{i_p} - 1]$, $p = 1, \ldots, R$ are disjoint, and such that

$$\bigcup_{i=1}^K [a_i, a_i + l_i - 1] \subset \bigcup_{p=1}^R [a_{i_p} - l_{i_p}, a_{i_p} + 2l_{i_p} - 1]$$

In particular, this implies that

$$K \leq 3 \cdot \sum_{p=1}^R l_{i_p}.$$

By choice of the a_i, l_i we have

$$\alpha \cdot l_i < \sum_{j=0}^{l_i-1}\phi(a_i + j).$$

Using this for the $i = i_p$, $p = 1, 2 \ldots, R$, and summing, while exploiting the disjointness of the intervals $[a_{i_p}, a_{i_p} + l_{i_p} - 1]$, we find

$$\alpha \cdot \sum_{p=1}^{R} l_{i_p} < \sum_{p=1}^{R} \sum_{j=0}^{l_{i_p}-1} \phi(a_{i_p} + j) \leq \big\|\phi\big\|_{l^1(\mathbb{Z})}.$$

Combining this with the above bound for K, we infer that

$$\alpha \cdot K \leq 3\alpha \cdot \sum_{p=1}^{R} l_{i_p} \leq 3\big\|\phi\big\|_{l^1(\mathbb{Z})}.$$

□

We can now implement the strategy outlined at the beginning of this section to prove the maximal ergodic inequality. We just have to be careful to take into account the truncations involved in the definition of ϕ. We shall prove the maximal ergodic inequality for the slightly modified maximal function

$$g_P^*(x) := \sup_{P \geq N \geq 1} N^{-1} \sum_{j=0}^{N-1} g(T^j(x)),$$

where $J \gg P \gg 1$. By letting $J, P \to \infty$, the desired result will follow. We analogously define the modified maximal function

$$\phi_P^*(a) := \sup_{P \geq N \geq 1} N^{-1} \sum_{j=0}^{N-1} \phi(a + j).$$

The point is that then restricting $a \in [0, \ldots, J - P]$, we have

$$\phi_P^*(a) = g_P^*(T^a(x)).$$

Also, introduce the set

$$E_\alpha^P := \{ \sup_{P \geq N \geq 1} N^{-1} \sum_{j=0}^{N-1} g(T^j(x)) > \alpha\}$$

According to the preceding lemma, we know that

$$\big|\{a \in [0, J - P],\, |\, \phi_P^*(a) > \alpha\}\big| \leq \big|\{a,\, |\, \phi^*(a) > \alpha\}\big| \leq 3\alpha^{-1}\big\|\phi\big\|_{l^1(\mathbb{Z})} \leq 3\alpha^{-1} \sum_{j=0}^{J} \big|g\big(T^j(x)\big)\big|.$$

We now go back to the identity

$$\sum_{0 \leq a \leq J-P} \chi_{T^{-a}\left(E_\alpha^P\right)}(x) = \Big|\{0 \leq a \leq J - P \,|\, \sup_{P \geq N \geq 1} N^{-1} \sum_{j=0}^{N-1} g(T^{a+j}(x)) > \alpha\}\Big|$$

As before we conclude that

$$\big(J - P + 1\big) \cdot m\big(E_\alpha^P\big) = \int_X \Big|\{0 \leq a \leq J - P \,|\, \sup_{P \geq N \geq 1} N^{-1} \sum_{j=0}^{N-1} g(T^{a+j}(x)) > \alpha\}\Big|\, dm$$

At the same time, we have

$$\Big|\{0 \leq a \leq J - P \,|\, \sup_{P \geq N \geq 1} N^{-1} \sum_{j=0}^{N-1} g(T^{a+j}(x)) > \alpha\}\Big| = \big|\{0 \leq a \leq J - P \,|\, \phi_P^*(a) > \alpha\}\big|$$

$$\leq 3\alpha^{-1} \sum_{j=0}^{J} \big|g\big(T^j(x)\big)\big|,$$

according to the preceding.
Carrying out the x-integral and taking advantage of the measure preservation property, we infer that

$$\big(J-P+1\big)\cdot m\big(E_\alpha^P\big)\leq 3\alpha^{-1}\int_X\sum_{j=0}^{J}\big|g\big(T^j(x)\big)\big|\,dm = 3\alpha^{-1}J\cdot\big\|g\big\|_{L^1(X,dm)}.$$

Dividing by J and letting $J\to+\infty$ for fixed P leads to the conclusion

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

Finally letting $P\to\infty$, the maximal ergodic inequality follows.

2. Proof of the Birkhoff ergodic theorem

Let $\big(X,m\big)$ be a finite measure space, and $T: X\longrightarrow X$ a measure preserving map. Further, let $f\in L^1(X,dm)$. By an exercise in the last problem set, we know that

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

converges toward a function $f_*\in L^1\big(X,dm\big)$ as $N\longrightarrow+\infty$. We intend to show that $A_N(f)$ converges almost everywhere to f_*. In fact, we shall show that

$$\Big|\{x\in X\mid \limsup_{N\to\infty}\big|A_Nf(x)-f_*(x)\big|>\varepsilon\}\Big|<12\varepsilon \tag{2.1}$$

for any $\varepsilon>0$. This of course implies that $ANf(x)\longrightarrow f_*(x)$ almost everywhere. In the particular case that T is ergodic, the Birkhoff theorem formulated in the last lecture follows.
We accomplish this in two steps:

Step 1: *Proof of* (2.1), *assuming it is true for* $f\in L^\infty(X,m)\subset L^1(X,m)$. Given $\varepsilon>0$, pick a function $f_0\in L^\infty(X,m)$ such that

$$\big\|f-f_0\big\|_{L^1}<\varepsilon^2.$$

Then, letting $f_{0,*}$ the L^1-limit of A_Nf_0, we have that

$$\begin{aligned}\{x\in X\mid \limsup_{N\to\infty}\big|A_Nf(x)-f_*(x)\big|>\varepsilon\}\subset&\{x\in X\mid \limsup_{N\to\infty}\big|A_Nf_0(x)-f_{0,*}(x)\big|>\frac{\varepsilon}{3}\}\\ &\cup\{x\in X\mid \limsup_{N\to\infty}\big|A_N(f-f_0)(x)\big|>\frac{\varepsilon}{3}\}\\ &\cup\{x\in X\mid \big|(f_*-f_{0,*})(x)\big|>\frac{\varepsilon}{3}\}\end{aligned}$$

By assumption, we have

$$\Big|\{x\in X\mid \limsup_{N\to\infty}\big|A_Nf_0(x)-f_{0,*}(x)\big|>\frac{\varepsilon}{3}\}\Big|=0.$$

By the maximal ergodic inequality, we have

$$\Big|\{x\in X\mid \limsup_{N\to\infty}\big|A_N(f-f_0)(x)\big|>\frac{\varepsilon}{3}\}\Big|\leq\frac{3\big\|f-f_0\big\|_{L^1}}{\frac{\varepsilon}{3}}<9\varepsilon.$$

Finally, we have

$$\big|\{x\in X\mid \big|(f_*-f_{0,*})(x)\big|>\frac{\varepsilon}{3}\}\big|\leq\frac{\big\|f_*-f_{0,*}\big\|_{L^1}}{\frac{\varepsilon}{3}}<3\varepsilon.$$

Inequality (2.1) follows.

Step 2: *Proof of* (2.1) *for* $f \in L^\infty\big(X, dm\big)$. As before denote by $f_* \in L^1(X, dm)$ the L^1-limit of $A_N(f)$. Then we have

$$A_M f_* = f_*$$

for any $M \geq 1$. Given $\varepsilon > 0$, pick M large enough such that

$$\big\|f_* - A_M f\big\|_{L^1} < \varepsilon^2.$$

Then consider

$$\{x \in X \mid \limsup_{N\to\infty} \big|f_*(x) - A_N f(x)\big| > \varepsilon\}.$$

Use the estimate

$$\begin{aligned}\limsup_{N\to\infty} \big|f_*(x) - A_N f(x)\big| &\leq \limsup_{N\to\infty} \big|f_*(x) - A_N \circ A_M f(x)\big| + \limsup_{N\to\infty} \big|A_N f - A_N \circ A_M f(x)\big| \\ &= \limsup_{N\to\infty} \big|A_N f_*(x) - A_N \circ A_M f(x)\big| + \limsup_{N\to\infty} \big|A_N f - A_N \circ A_M f(x)\big|.\end{aligned}$$

We show that

- Using the L^∞-bound for f, we have $\limsup_{N\to\infty} \big|A_N f - A_N \circ A_M f(x)\big| = 0$.
- Using the maximal ergodic inequality, we have

$$\Big|\{x \in X \mid \limsup_{N\to\infty} \big|A_N f_*(x) - A_N \circ A_M f(x)\big| > \varepsilon\}\Big| < 3\varepsilon.$$

Combining these facts, we have

$$\begin{aligned}\Big|\{x \in X \mid \limsup_{N\to\infty} \big|f_*(x) - A_N f(x)\big| > \varepsilon\}\Big| &\leq \Big|\{x \in X \mid \limsup_{N\to\infty} \big|A_N f_*(x) - A_N \circ A_M f(x)\big| > \varepsilon\}\Big| \\ &< 3\varepsilon,\end{aligned}$$

which proves (2.1) in our present situation.

Proof that $\limsup_{N\to\infty} \big|A_N f - A_N \circ A_M f(x)\big| = 0$. Observe that[1]

$$\begin{aligned}A_N \circ A_M f(x) &= \frac{1}{NM} \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} f\big(T^{j+i}(x)\big) \\ &= \frac{1}{M} \sum_{i=0}^{M-1} \frac{1}{N} \Big[\sum_{j=0}^{N-1} f\big(T^j(x)\big) - \sum_{j=0}^{(i-1)} f\big(T^j(x)\big) + \sum_{j=N}^{N+(i-1)} f\big(T^j(x)\big)\Big] \\ &= A_N(f)(x) - \frac{1}{NM} \sum_{i=1}^{M-1} \sum_{j=0}^{i-1} f\big(T^j(x)\big) + \frac{1}{NM} \sum_{i=1}^{M-1} \sum_{j=N}^{N+i-1} f\big(T^j(x)\big).\end{aligned}$$

Since we have the easily verified bounds

$$\Big|\frac{1}{NM} \sum_{i=1}^{M-1} \sum_{j=0}^{i-1} f\big(T^j(x)\big)\Big| \leq \frac{M}{N} \cdot \|f\|_{L^\infty}, \quad \Big|\frac{1}{NM} \sum_{i=1}^{M-1} \sum_{j=N}^{N+i-1} f\big(T^j x)\big)\Big| \leq \frac{M}{N} \cdot \|f\|_{L^\infty},$$

we conclude that

$$\limsup_{N\to\infty} \big|A_N \circ A_M f(x) - A_N f(x)\big| = 0.$$

[1] We nterpret sums of the form $\sum_{i=a}^{b}$ with $b < a$ as equal to zero.

Proof that $\Big|\{x \in X \mid \limsup_{N\to\infty} \big|A_N f_*(x) - A_N \circ A_M f(x)\big| > \varepsilon\}\Big| < 3\varepsilon$. This is a direct consequence of the maximal ergodic inequality:

$$\begin{aligned}\Big|\{x \in X \mid \limsup_{N\to\infty} \big|A_N f_*(x) - A_N \circ A_M f(x)\big| > \varepsilon\}\Big| &\leq \Big|\{x \in X \mid \sup_{N\geq 1} \big|A_N f_*(x) - A_N \circ A_M f(x)\big| > \varepsilon\}\Big| \\ &\leq 3\varepsilon^{-1} \cdot \big\|f_* - A_M f\big\|_{L^1} \\ &\leq 3\varepsilon.\end{aligned}$$

The proof of the Birkhoff ergodic theorem is now complete.