Ergodic Theory Lecture Notes - Jan-Mar. 2011

Lecture 2

April 3, 2011

1 Subadditive Ergodic Theorem

More generally, we can consider a subadditive sequence of functions $f_n(x) \in L^1(X, \mu)$, i.e., $f_{n+m}(x) \leq f_n(x) + f_m(T^n x)$, for $n, m \geq 1$, where $T : X \to X$ preserves a probability measure μ.

Example 1.1. *Let us fix $f \in L^1(X, \mu)$ and then define $f_n(x) := \sum_{k=0}^{n-1} f(T^k x)$. We can then see that for a.e. (μ) $x \in X$:*

$$f_{n+m}(x) = f_n(x) + f_m(T^n x).$$

Example 1.2. *Let $M : X \to GL(d, \mathbb{R})$ be a function taking values in the family (invertible) $d \times d$ matrices. We can define a norm on $GL(d, \mathbb{R})$ by*

$$\|A\| = \sup\{\|Ax\|_2 \, : \, x \in \mathbb{R}^2 \text{ such that } \|x\|_2 \leq 1\}.$$

We assume that $\int \log_+ \|M(x)\| d\mu(x) < +\infty$ (where $\log +(\xi) = \max\{\xi, 0\}$). Let $f_n(x) = \log \|M(x) \cdot M(Tx) \cdots M(T^{n-1}x)\|$ then the subadditivity follows from the standard norm inequality $\|A_1 A_2\| \leq \|A_1\| \|A_2\|$.

Theorem 1.3 (Kingman, 1967)**.** *Assume that μ is ergodic. Let $f_n \in L^1(X, \mu)$ be subadditive. Then the following limit exists*

$$\lim_{n \to +\infty} \frac{1}{n} f_n(x) = l := \inf \left\{ \int f_n d\mu \, : \, n \geq 1 \right\},$$

a.e. (μ) $x \in X$.

Proof. Let us define $\underline{f}(x) := \liminf_{n \to +\infty} \frac{1}{n} f_n(x) \leq \overline{f}(x) := \limsup_{n \to +\infty} \frac{1}{n} f_1(T^i x) = 0$, for a.e. ($\mu$) $x \in X$. Since

$$\underbrace{\liminf_{n \to +\infty} \frac{1}{n+1} f_{n+1}(x)}_{=\underline{f}(x)} \leq \underbrace{\liminf_{n \to +\infty} \frac{1}{n+1} f_1(x)}_{=0} + \underbrace{\liminf_{n \to +\infty} \frac{1}{n+1} f_n(Tx)}_{=\underline{f}(Tx)}$$

we have that $\underline{f}(x) \leq \underline{f}(Tx)$ and thus ergodicity implies that $\underline{f}$ is constant. Similarly, we see that $\overline{f}(x) \leq \overline{f}(Tx)$ and thus ergodicity implies that $\overline{f}$ is constant as well.

Fix $N \geq 1$ and choose $1 \leq i \leq N$. We can then write any $n \geq i$ as $n = i + mN + r$ for unique choices $0 \leq r \leq N-1$ and $m \geq 0$. By subadditivity

$$\begin{aligned} f_n(x) &\leq f_i(x) + f_{nm}(T^i x) + f_r(T^{i+nm}x) \\ &\leq f_i(x) + \sum_{j=0}^{m-1} f_N(T^{jN+i}x) + f_r(T^{(m-1)N+i}x). \end{aligned}$$

Summing this inequality over $1 \leq i \leq N$ gives:

$$Nf_n(x) \leq \sum_{i=1}^{N} f_i(x) + \underbrace{\sum_{i=1}^{N}\sum_{j=0}^{m-1} f_N(T^{jN+i}x)}_{=\sum_{k=0}^{mN-1} f_N(T^k x)} + \sum_{i=1}^{N} f_{\underbrace{n-i-mN}_{=r}}(T^{nN+i}x).$$

Dividing by nN and letting $n \to +\infty$ gives

$$\underbrace{\limsup_{n\to+\infty} \frac{1}{n} f_n(x)}_{=:\overline{f}(x)} \leq 2\underbrace{\limsup_{n\to+\infty} \frac{1}{n}\left(\max_{1\leq i\leq N} \|f_i\|_\infty\right)}_{=0} + \underbrace{\limsup_{n\to+\infty} \frac{1}{nN} f_N^{nN}(x)}_{=\frac{1}{N}\int f_N d\mu}$$

using the Birkhoff theorem for f_N, for any $N \geq 1$. In particular, we see that

$$\overline{f}(x) \leq l =: \inf_N \left\{ \frac{1}{N} \int f_N d\mu \right\}.$$

Next we can turn our attention to $\underline{f}$. Moreover, we may as well assume $l > -\infty$.
Fix $\epsilon > 0$. Let us define $n : X \to \mathbb{N}$ by $n(x) = \inf\{n \geq 1 : f_n(x) \leq n(\underline{f} + \epsilon)\}$.
Fix $M > 0$. Let use define the set of points $A = \{x : n(x) \geq M\}$.

Claim: For $n \geq 1$, $f_n(x) \leq n(\underline{f} + \epsilon) + \sum_{i=0}^{n-1} \chi_A(T^i x)\|f_1\|_\infty + M\|f_1\|_\infty$ for a.e. (μ) $x \in X$.
Proof of Claim. This is similar to the claim in the first proof of the Birkhoff ergodic theorem. We can cover the set $\{1, 2, \cdots, n-1\}$ by sets of the form

1. $\{k : T^k x \in A\}$;
2. $\{l, l+1, \cdots, l + n(T^l x) - 1\}$; or
3. $\{n - M, \cdots, n-1\}$.

This completes the proof of the claim. □

Thus

$$\frac{f_n(x)}{n} \leq (\underline{f} + \epsilon) + \frac{1}{n}\sum_{i=0}^{n-1} \chi_A(T^i x)\|f_1\|_\infty + \frac{M}{n}\|f_1\|_\infty.$$

Integrating this inequality gives that

$$\frac{\int f_n d\mu}{n} \leq (\underline{f} + \epsilon) + \mu(A)\|f_1\|_\infty + \frac{M}{n}\|f_1\|_\infty.$$

and therefore

$$l = \inf_{n \geq 1} \left\{ \frac{\int f_n d\mu}{n} \right\} \leq (\underline{f} + \epsilon) + \mu(A) \|f_1\|_\infty + \frac{M}{n} \|f_1\|_\infty.$$

for any $n \geq 1$.

First we can let $n \to +\infty$. Next we let $M \to \infty$, which in turn implies that $\mu(A) \to 0$. Finally, we let $\epsilon \to 0$. This therefore gives that $l \leq \underline{f}$ a.e. (μ) $x \in X$.

Therefore $\overline{f}(x) = \underline{f}(x) = l$, i.e., $l = \lim_{n \to +\infty} \frac{f_n(x)}{n}$. □

2 Ergodicity of flows and the Hopf method

Let X be a space with associated sigma-algebra $\mathcal{B}$.

Definition 2.1. *We say that μ is T-*invariant *if for any $A \in \mathcal{B}$ we have that $\mu(T^{-t}A) = \mu(A)$ for all $t \in \mathbb{R}$.*

Definition 2.2. *We say that (T^t, X, μ) is* ergodic *if whenever $T^{-t}A = A \in \mathcal{B}$ for all $t \in \mathbb{R}$ then $\mu(A) = 0$ or 1.*

Theorem 2.3 (Birkhoff - Ergodic Theory case)**.** *Let μ be an ergodic measure. Let $f \in L^1(X, \mu)$ then $\lim_{t \to +\infty} \frac{1}{t} \int_0^t f(T^u x) du = \int f d\mu$, a.e. (μ).*

We also need to consider the case that the invariant measure is not necessarily ergodic. This requires a little more notation. Let $\mathcal{I} \subset \mathcal{B}$ be the sub-sigma-algebra of T^t-invariant sets, i.e.,

$$\mathcal{I} = \{B \in \mathcal{B} : T^{-t}B = B \text{ (up to a set of zero measure) } \forall t \in \mathbb{R}\}.$$

Given $f \in L^1(X, \mathcal{B}, \mu)$ there is a unique $g \in L^1(X, \mathcal{I}, \mu)$ such that

1. g is $\mathcal{I}$-measurable (i.e., for any Borel measurable set $B \subset \mathbb{R}$, $g^{-1}B \in \mathcal{I}$);
2. $\int_B f d\mu = \int_B g d\mu$, $\forall B \in \mathcal{I}$.

We then denote $g = E(f|\mathcal{I})$, the conditional expectation of f with respect to $\mathcal{I}$.

Theorem 2.4 (Birkhoff - Invariant measure case)**.** *Let $f \in L^1(X, \mu)$ then $\lim_{t \to +\infty} \frac{1}{t} \int_0^t f(T^u x) du = E(f|\mathcal{B})$, a.e. (μ).*

We want to consider two important examples.

Example 2.5 (Geodesic flows)**.** *We want to show the following for the geodesic flow on $X = SL(2, \mathbb{R})/SL(2, Z)$.*

Theorem 2.6. *The geodesic flow $g_t : X \to X$ defined by*

$$g_t g SL(2, \mathbb{Z}) = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} g SL(2, \mathbb{Z})$$

is ergodic.

The proof uses the so called Hopf method. We define flows $h_t : X \to X$ and $h_t^- : X \to X$ defined by

$$h_s gSL(2,\mathbb{Z}) = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} gSL(2,\mathbb{Z}) \text{ and } h_s^- gSL(2,\mathbb{Z}) = \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} gSL(2,\mathbb{Z}).$$

We begin with the crucial connection between g_t and h_t, and g_t and h_t^-.

Lemma 2.7. *We can write* $g_t h_s g_{-t} = h_{se^t}$ *and* $g_t h_s^{--} g_{-t} = h_{se^{-t}}^-$

Proof. This follows by matrix multiplication

$$g_t h_s g_{-t} = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e^{-t/2} & 0 \\ 0 & e^{t/2} \end{pmatrix} = \begin{pmatrix} 1 & se^t \\ 0 & 1 \end{pmatrix} = h_{se^t}$$

and

$$g_t h_s^- g_{-t} = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} \begin{pmatrix} e^{-t/2} & 0 \\ 0 & e^{t/2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ se^{-t} & 1 \end{pmatrix} = h_{se^{-t}}^-$$

□

We can apply the invariant measure version of the ergodic theorem for flows to deduce that for a.e. (μ) $x \in X$ we have that for any $f \in L^1(X,\mu)$.

$$E(f|\mathcal{I})(x) = \lim_{T\to+\infty} \frac{1}{T} \int_0^T f(g_t x)dt$$

and

$$E(f|\mathcal{I})(x) = \lim_{T\to+\infty} \frac{1}{T} \int_0^T f(g_{-t} x)dt.$$

Lemma 2.8. *If* $E(f|\mathcal{I})(x)$ *is constant if a.e.* (μ) *and all* $f \in C(X)$ *then the flow* g_t *is ergodic.*

Proof. Recall that to show μ is ergodic it suffices to show that for any $f = f \circ T \in L^2(X,\mu)$ we have that f is constant. In terms of the operator $E(\cdot|\mathcal{I})(x)$ this is equivalent to saying that the image of $L^2(X,\mu)$ are the constant functions. However, since $E(\cdot|\mathcal{I}) : L^2(X,\mu) \to L^2(X,\mu)$ contracts the norm, and $C(X) \subset L^2(X,\mu)$ is dense, this equivalent to the same statement for continuous functions. □

Clearly, if $y = g_s(x)$ we have that

$$\begin{aligned} E(f|\mathcal{I})(x) - E(f|\mathcal{I})(y) &= E(f|\mathcal{I})(x) - E(f|\mathcal{I})(g_s x) \\ &= \lim_{T\to+\infty} \frac{1}{T} \int_0^T f(g_t x)dt - \lim_{T\to+\infty} \frac{1}{T} \int_0^T f(g_{s+t} x)dt \\ &= -\lim_{T\to+\infty} \frac{1}{T} \int_{-s}^0 f(g_t x)dt + \lim_{T\to+\infty} \frac{1}{T} \int_T^{T-s} f(g_t x)dt = 0. \end{aligned}$$

The important point is that if $y = h_s x$ then

$$\begin{aligned} E(f|\mathcal{I})(x) - E(f|\mathcal{I})(y) &= E(f|\mathcal{I})(x) - E(f|\mathcal{I})(h_s x) \\ &= \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_{-t}x)dt - \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_{-t}h_s x)dt \\ &= \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_{-t}x)dt - \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(h_{se^{-t}}g_{-t}x)dt \end{aligned}$$

However, given $\epsilon > 0$ for $T_0 > 0$ sufficiently large we have that $\|f(x) - f(h_{se^{-t}x})\|_\infty < \epsilon/2$ for $t > T_0$. In particular, we can bound

$$\left| \frac{1}{T}\int_0^T f(g_t x)dt - \frac{1}{T}\int_0^T f(h_{se^{-t}}g_t x)dt \right| \le \frac{2\|f\|_\infty T_0}{T} + \frac{(T-T_0)\epsilon}{2T}.$$

and deduce that for $y = h_s x$ we have that $|E(f|\mathcal{I})(x) - E(f|\mathcal{I})(y)| < \epsilon$.

Similarly, if $y = h_s^- x$ then

$$\begin{aligned} E(f|\mathcal{I})(x) - E(f|\mathcal{I})(y) &= E(f|\mathcal{I})(x) - E(f|\mathcal{I})(h_s^- x) \\ &= \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_t x)dt - \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_t h_s^- x)dt \\ &= \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(g_t x)dt - \lim_{T\to+\infty} \frac{1}{T}\int_0^T f(h_{se^t}^- g_t x)dt \end{aligned}$$

However, given $\epsilon > 0$ for $T_0 > 0$ sufficiently large we have that $\|f(\cdot) - f(h_{se^t\cdot}^-)\|_\infty < \epsilon/2$ for $t > T_0$. In particular, we can bound

$$\left| \frac{1}{T}\int_0^T f(g_t x)dt - \frac{1}{T}\int_0^T f(h_{se^t}^- g_t x)dt \right| \le \frac{2\|f\|_\infty T_0}{T} + \frac{(T-T_0)\epsilon}{2T}.$$

and deduce that for $y = h_s^- x$ we have that $|E(f|\mathcal{I})(x) - E(f|\mathcal{I})(y)| < \epsilon$.

We next need the following result.

Lemma 2.9. *We can write any element* $\gamma = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in G = SL(2,\mathbb{R})$ *with* $\delta \neq 0$ *in the form* $\gamma = h_{s_1} g_t h_{s_2}^-$.

Proof. It suffices to observe that every matrix in $SL(2,\mathbb{R})$ with $\delta \neq 0$ can be written in the form

$$\begin{pmatrix} 1 & s_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ s_2 & 1 \end{pmatrix} = \begin{pmatrix} e^{t/2} + s_2 s_1 e^{-t/2} & s_1 e^{-t/2} \\ s_2 e^{-t/2} & e^{-t/2} \end{pmatrix}$$

since we can choose t such that $e^{-t/2} = \delta$. □

Thus given almost any two points $x, x' \in G/\Gamma$ we can choose $\gamma \in G$ such that $x' = \gamma x$. In particular, we can write $\gamma x = h_{s_1} g_t h_{s_2}^- x$ and define

$$y_1 = h_{s_2}^- x, y_2 = g_t y_1, x' = h_{s_1} y_2.$$

Thus we can show that the averages $E(f|\mathcal{I})(x) = E(f|\mathcal{I})(x')$.

Example 2.10 (Horocycle flows)**.** *We define probability measures $\nu_t \in \mathcal{M}$, for $t > 0$, by*

$$\int f d\nu_t = \frac{1}{t}\int_0^t f(h_s x)ds \text{ for } f \in C^0(X).$$

Let us denote

$$M_t f(x) = \int_0^1 f(g_{-\log t} h_s x)ds \text{ for } t > 0.$$

We first observe that

$$\begin{aligned} \frac{1}{t}\int_0^t f(h_s x)ds &= \int_0^1 f(h_{st} x)ds \text{ (by a change of variable)} \\ &= \int_0^1 f(g_{-\log t} h_s g_{-\log t} x)ds \text{ (since } h_{st} = g_{-\log t} h_s g_{-\log t}\text{)} \\ &= M_t f(g_{-\log t} x) \text{ (by definition of } M_t \text{ above)} \end{aligned} \tag{1}$$

We begin with the following estimate.

Lemma 2.11. *The family $\{M_t f(x) : t > 0\}$ is equicontinuous, i.e., for all $\epsilon > 0$ there exists $\delta > 0$ such that if $d(x, x') < \delta$ then $|M_t f(x) - M_t f(x')| < \epsilon$ for all $t > 0$.*

Proof. Let us define

$$D_x(\epsilon) := \{g_t h_s^-(x) : |t|, |s| \leq \epsilon\} \text{ and } V_x(\epsilon) := h_{[0,1]} D_x(\epsilon) (= \cup_{0 \leq t \leq 1} h_t D_x(\epsilon)).$$

Then by uniform continuity of f we have

$$\left| M_t f(x) - \frac{1}{\mathrm{Vol}(V_x(\epsilon))}\int_{V_x(\epsilon)} f(g_{-\log t}\cdot)d\mu \right| \to 0 \text{ as } \epsilon \to 0.$$

Moreover, for $x, x' \in X$ sufficiently close we can estimate

$$\begin{aligned} &\left| \frac{1}{\mathrm{Vol}(V_{x'}(\epsilon))}\int_{V_{x'}(\epsilon)} f(g_{-\log t}\cdot)d\lambda - \frac{1}{\mathrm{Vol}(V_x(\epsilon))}\int_{V_x(\epsilon)} f(g_{-\log t}\cdot)d\lambda \right| \\ =& \left| \int f(g_{-\log t}\cdot)\left(\frac{\chi_{V_{x'}(\epsilon)}}{\mathrm{Vol}(V_{x'}(\epsilon))} - \frac{\chi_{V_x(\epsilon)}}{\mathrm{Vol}(V_x(\epsilon))} \right) d\lambda \right| \\ \leq& \underbrace{\|f \circ g_{-\log t}\|_2}_{\|f\|_2} \left\| \frac{1}{\mathrm{Vol}(V_{x'}(\epsilon))}\chi_{V_{x'}(\epsilon))} - \frac{1}{\mathrm{Vol}(V_x(\epsilon))}\chi_{V_x(\epsilon))} \right\|_2 \end{aligned}$$

by the Cauchy-Schwartz theorem (and g_t-invariance of μ). Moreover, it is easy to see the term on the Right Hand Side tends to zero as $d(x, x') \to 0$. □

If X is compact then we can apply the Arzela-Ascoli theorem there exists an accumulation point $\overline{f} \in C(X)$ of $\{M_t f(x) : t > 0\}$, i.e. there exists a subsequence such that $M_{t_k} f \to \overline{f}$ as $t_k \to +\infty$. If X is not compact then we still know that this converges on compact sets (and in L^2).

Lemma 2.12. *$\overline{f}$ is constant on h_s-orbits.*

Proof. We have by (1) that

$$\begin{aligned} \|\frac{1}{t_k}\int_0^{t_k} f(h_s\cdot)ds - \overline{f}\circ g_{-\log t_k}\|_2 &= \|(M_{t_k}f)\circ g_{-\log t_k} - \overline{f}\circ g_{-\log t_k}\|_2 \\ &= \|M_{t_k}f - \overline{f}\|_2 \to 0 \end{aligned} \tag{2}$$

as $t_k \to 0$, by g_t-invariance of the probability measure μ. Moreover, by the von Neumann mean ergodic theorem (for invariant probability measures)

$$\frac{1}{t_k}\int_0^{t_k} f(h_s\cdot)ds \to \widehat{f} = E(f|\mathcal{I}) \in L^2(X) \tag{3}$$

as $k \to +\infty$ in the L^2 norm. Moreover, this limit is naturally h_s-invariant, i.e., $\widehat{f} = \widehat{f}\circ h_s$ for all $s \in \mathbb{R}$. Thus,

1. We have $\|\overline{f} - \widehat{f}\circ g_{\log t_k}\|_2 = \|\overline{f}\circ g_{-\log t_k} - \widehat{f}\|_2 \to 0$ by the invariance of the measure and (2) and (3).
2. We can bound

$$\begin{aligned} &\|\overline{f}\circ h_s - \widehat{f}\circ g_{\log t_k}\|_2 \\ &= \|\overline{f} - \widehat{f}\circ g_{\log t_k}\circ h_{-s}\|_2 \text{ (by } h_s\text{-invariance of } \mu) \\ &= \|\overline{f} - \widehat{f}\circ h_{-st_k}\circ g_{\log t_k}\|_2 \text{ (since } g_{\log t_k}\circ h_{-s} = h_{-st_k}\circ g_{\log t_k}) \\ &= \|\overline{f} - \widehat{f}\circ g_{\log t_k}\|_2 \to 0 \text{ (since } \widehat{f}\circ h_{-st_k} = \widehat{f} \text{ by } h_s\text{-invariance of } \widehat{f}) \end{aligned}$$

Comparing 1 and 2 shows $\overline{f} = \overline{f}\circ h_s$, as required. □

Finally, we need the following simple topological fact (whose proof we omit).

Lemma 2.13. *There exists a dense h_s orbit in X.*

In particular, since $\overline{f}$ is a continuous function and it is constant on h_s-orbits this implies that $\overline{f}$ is a constant. Since we had in the case that X is compact convergence of the averages to a constant along every horocycle orbit, this is enough to deduce the following.

Theorem 2.14. *If X is compact then the horocycle flow $h_t : X \to X$ is (uniquely) ergodic.*

More generally, even if X isn't compact the L^2 convergence allows us to obtain.

Theorem 2.15. *The horocycle flow $h_t : X \to X$ is ergodic.*

3 Continued fractions

Let $0 \leq x \leq 1$ be an irrational number. There is a unique "Continued Fraction Expansion" of the form

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots}}} \text{ where } a_1, a_2, \ldots \in \mathbb{N}.$$

Let us consider $T : [0,1) \to [0,1)$ by

$$Tx = \begin{cases} \{1/x\}(= 1/x - [1/x]) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Then

$$x = \frac{1}{\frac{1}{x}} = \frac{1}{[1/x] + \{1/x\}} = \frac{1}{a_1 + Tx}$$

and

$$Tx = \frac{1}{(Tx)^{-1}} = \frac{1}{\frac{1}{x}} = \frac{1}{[1/Tx] + \{1/Tx\}} = \frac{1}{a_2 + T^2x}.$$

Thus

$$x = \frac{1}{a_1 + \left(\frac{1}{a_2 + T^2x}\right)} = \frac{a_2 + T^2x}{(a_1a_2 + 1) + a_1(T^2x)}$$

In particular, proceeding inductively, for $n \geq 2$

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots \cfrac{1}{a_n + T^nx}}}} \text{ where } a_n = [1/T^{n-1}x] \in \mathbb{N}.$$

We can rearrange this as

$$x = \frac{p_n + p_{n-1}(T^nx)}{q_n + q_{n-1}(T^nx)} \text{ for } n \geq 1.$$

We recall some basic facts.

Lemma 3.1. *1. We define p_n, q_n inductively by:*

$$p_0 = 0, q_0 = 1, p_1 = 1, q_1 = a_1$$

$$p_n = a_np_{n-1} + p_{n-2} \textit{ and } q_n = a_nq_{n-1} + q_{n-2}$$

2. $p_{n-1}q_n - p_nq_{n-1} = (-1)^n$

Proof. 1. Substituting $T^n(x) = \frac{1}{a_{n+1} + T^{n+1}x}$ we can write

$$\frac{p_n + p_{n-1}(T^nx)}{q_n + q_{n-1}(T^nx)} = \frac{p_n + p_{n-1}\left(\frac{1}{a_{n+1} + T^{n+1}x}\right)}{q_n + q_{n-1}\left(\frac{1}{a_{n+1} + T^{n+1}x}\right)}$$

$$= \frac{\underbrace{(a_{n+1}p_n + p_{n-1})}_{=p_{n+1}} + p_n(T^{n+1}x)}{\underbrace{(a_{n+1}q_n + q_{n-1})}_{=q_{n+1}} + q_n(T^{n+1}x)}$$

by induction. It also holds with $n = 1$.

2. We can write

$$p_n q_{n+1} - p_{n+1} q_n = p_n(a_{n+1}q_n + q_{n-1}) - (a_{n+1}p_n + p_{n-1})q_n = p_n q_{n-1} - p_{n-1}q_n = -(-1)^{n-1} = (-1)^n$$

by induction. It also holds with $n = 1$.

□

Corollary 3.2. *$q_n \geq 2^{(n-1)/2}$ and $p_n \geq 2^{(n-2)/2}$*

Proof. The proof is by induction. By the above

$$q_n \geq q_{n-1} + q_{n-2} \geq 2^{(n-2)/2} + 2^{(n-3)/2} \geq 2 \cdot 2^{(n-3)/2} = 2^{(n-1)/2}.$$

by induction. It also holds for $n = 1$. Similarly for p_n. □

We now want to study the ergodic properties of T.

Lemma 3.3 (Gauss)**.** *The probability measure μ defined by*

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{dx}{1+x} \text{ for any } A \in \mathcal{B}$$

is T-invariant.

Proof. It suffices to show that for any interval $(0, \alpha)$ we have that $\mu((0,\alpha)) = \mu(T^{-1}(0,\alpha))$. In particular,

$$\begin{aligned}
\mu(T^{-1}(0,\alpha)) &= \sum_{k=1}^{\infty} \mu\left(\left(\frac{1}{k+\alpha}, \frac{1}{k}\right)\right) = \frac{1}{\log 2} \sum_{k=1}^{\infty} \int_{\frac{1}{k+\alpha}}^{\frac{1}{k}} \frac{dx}{1+x} \\
&= \frac{1}{\log 2} \sum_{k=1}^{\infty} \left(\log\left(1 + \frac{1}{k}\right) - \log\left(1 + \frac{1}{k+\alpha}\right)\right) \\
&= \frac{1}{\log 2} \sum_{k=1}^{\infty} \left(\log\left(1 + \frac{\alpha}{k}\right) - \log\left(1 + \frac{\alpha}{k+1}\right)\right) \\
&= \frac{1}{\log 2} \sum_{k=1}^{\infty} \int_{\frac{\alpha}{k+1}}^{\frac{\alpha}{k}} \frac{dx}{1+x} \\
&= \frac{1}{\log 2} \int_0^{\alpha} \frac{dx}{1+x} = \mu\left((0,\alpha)\right).
\end{aligned}$$

□

Theorem 3.4. *The measure μ is ergodic.*

Proof. Consider the intervals

$$I_n(x) := \begin{cases} \left[\frac{p_n}{q_n}, \frac{p_n + p_{n-1}}{q_n + q_{n-1}}\right] & \text{if } n \text{ is even} \\ \left[\frac{p_n + p_{n-1}}{q_n + q_{n-1}}, \frac{p_n}{q_n}\right] & \text{if } n \text{ is odd} \end{cases}$$

then the length is $l(I_n(x)) = \frac{1}{q_n(q_n+q_{n-1})}$ (by Lemma (ii)), where l denotes Lebesgue measure. We can define $\phi : [0,1] \to I_n$ by

$$\phi(t) = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots \frac{1}{a_n+t}}}} = \frac{p_{n-2} + (a_n+t)p_{n-1}}{q_{n-2} + (a_n+t)q_{n-1}}.$$

Note that ϕ is a bijection. Assume that $E = T^{-1}E \subset [0,1]$ then we can write

$$l(E \cap I_n(x)) = \int_0^1 \chi_E(\phi(t))|\phi'(t)|dt$$

and $|\phi'(t)| = \frac{1}{(q_{n-2}(a_n+t)q_{n-1})^2} \in \left[\frac{1}{(q_n+q_{n-1})^2}, \frac{1}{q_n^2}\right]$ by (i) and (ii).

In particular,

$$\frac{l(E \cap I_n)}{l(I_n)} \geq \frac{1}{4} l(E) \tag{$*$}$$

since $l(E \cap I_n) \geq \frac{l(E)}{(q_n+q_{n-1})^2} \geq \frac{l(E)}{(q_n+q_{n-1})^2} \underbrace{(l(I_n)q_n)^2}_{\leq 1} \geq l(E)l(I_n)/4$.

Given $\epsilon > 0$ we can cover E^c by intervals $\cup_n I_n$ of the above form such that

$$\sum_n l(I_n) \geq l(E^c) \geq \sum_n l(I_n) - \epsilon.$$

Then

$$\begin{aligned} l(E)l(E^c) &\leq l(E)\sum_n l(I_n) \leq 8\sum_n l(E\cap I_n) \qquad \text{(by (*))} \\ &= 8\left(\sum_n l(E\cap I_n) + l(\underbrace{E^c \cap (\cup_n I_n)}_{=E^c}) - l(E^c)\right) \\ &\leq 8\left(\sum_n l(E\cap I_n) + \sum_n l(E^c\cap I_n) - l(E^c)\right) \\ &= \left(\sum_n l(I_n) - l(E^c)\right) \leq 8\epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, either $l(E) = 0$ or $l(E^c) = 0$, i.e., $\mu(E) = 0$ or $\mu(E^c) = 0$.

□

By applying the Birkhoff ergodic theorem we have the following.

Theorem 3.5. *For a.e. (μ) $x \in (0,1)$,*

1. *the limit of the geometric averages satisfies*

$$\lim_{n\to+\infty} (a_1 \cdots a_n)^{1/n} = \prod_{k=1}^{\infty}\left(1 + \frac{1}{k^2+2k}\right)^{\frac{\log k}{\log 2}} ; \textit{and}$$

2. the growth rate of the q_n is

$$\lim_{n\to+\infty} \frac{\log q_n}{n} = \frac{\pi^2}{12\log 2}$$

Proof. 1. Let $f(x) = \log[1/y]$. By the Birkhoff ergodic theorem:

$$\begin{aligned}
\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) &= \frac{1}{n}\sum_{k=1}^{n} f(T^{k-1}x) \\
&= \frac{1}{n}\sum_{k=1}^{n} \log\left[\frac{1}{T^{k-1}(x)}\right] \\
&= \frac{1}{n}\sum_{k=1}^{n} \log a_k \\
&\to \frac{1}{\log 2}\int_{\frac{1}{k+1}}^{\frac{1}{k}} \frac{\log k}{1+x} dx \\
&= \sum_{k=1}^{\infty} \frac{\log k}{\log 2} \log\left(1 + \frac{1}{k^2+2k}\right)
\end{aligned}$$

where $f \in L^1(X, m)$ we have that $\int f d\mu \leq \frac{1}{\log 2}\sum_{k=1}^{\infty} \frac{\log k}{k(k+1)}$. Taking exponentials gives the result.

2. Since we can write

$$\begin{aligned}
x = \frac{p_n(x)}{q_n(x)} &= [a_1, \cdots, a_n] \\
&= \frac{1}{1 + [a_2, \cdots, a_n]} \\
&= \frac{1}{1 + \frac{p_{n-1}(Tx)}{q_{n-1}(Tx)}} \\
&= \frac{q_{n-1}(Tx)}{p_{n-1}(Tx) + q_{n-1}(Tx)}
\end{aligned}$$

we deduce that $p_n(x) = q_{n-1}(Tx)$, by comparing numerators. In particular,

$$\frac{p_n(x)}{q_n(x)}\frac{p_{n-1}(Tx)}{q_{n-1}(Tx)}\cdots\frac{p_1(T^{n-1}x)}{q_1(T^{n-1}x)} = \frac{1}{q_n(x)}$$

and so

$$\frac{1}{n}\log q_n(x) = -\frac{1}{n}\sum_{k=0}^{n-1} \log\left(\frac{p_{n-k}}{q_{n-k}}(T^k x)\right).$$

If $f(x) = \log x$ then

$$\frac{1}{n}\log q_n(x) = -\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) + \underbrace{\frac{1}{n}\left(\sum_{k=0}^{n-1} \log(T^k x) - \log\left[\frac{p_{n-k}}{q_{n-k}}(T^k x)\right]\right)}_{(*)}.$$

By the Birkhoff ergodic theorem and the expansion $\log(1+x) = \sum_{k=0}^{\infty}(-1)^k \frac{(-1)^k}{k+1}$

$$\begin{aligned}
\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) &= \frac{1}{\log 2}\int_0^1 \frac{\log x}{1+x}dx \\
&= -\frac{1}{\log 2}\int_0^1 \frac{\log(1+x)}{x}dx \text{ (Integrating by parts)} \\
&= -\frac{1}{\log 2}\sum_{k=0}^{\infty}(-1)^k \int_0^1 \frac{x^k}{k+1}dx \text{ (by expanding the logarithm)} \\
&= -\frac{1}{\log 2}\sum_{k=0}^{\infty}\frac{(-1)^k}{(k+1)^2} = -\frac{\pi^2}{12\log 2}.
\end{aligned}$$

It still remains to show that the contribution from (*) is trivial:

$$\sum_{k=0}^{n-1}\left|\log\left(\frac{T^k x}{\left[\frac{p_{n-k}}{q_{n-k}}(T^k x\right]}\right)\right| \leq \sum_{k=0}^{n-1}\left|\frac{T^k x)}{\left[\frac{p_{n-k}}{q_{n-k}}(T^k x)\right]} - 1\right|$$

since $|\log y| \leq |y-1|$. Moreover,

$$\left|\frac{x}{p_k/q_k} - 1\right| = \frac{q_k}{p_k}\left|x - \frac{p_k}{q_k}\right| \leq \frac{1}{p_k q_{k+1}} \leq \frac{1}{2^{k-1}}$$

giving a bound

$$\sum_{k=0}^{n-1}\left|\frac{T^k x}{\left[\frac{p_{n-k}}{q_{n-k}}(T^k x)\right]} - 1\right| \leq \sum_{k=0}^{n-1}\frac{1}{2^{n-k-1}} \leq 1.$$

□

4 Markov Operators

Let $T_1, \cdots, T_n : X \to X$ be measure preserving transformations with respect to a probability measure μ (i.e., for all $B \in \mathcal{B}$ we have that $\mu(T_i^{-1}B) = \mu(B)$). Let $(p_1, \cdots, p_n)$ be a probability vector (with $p_1 + \cdots + p_n = 1$). We define a linear operator, $T : L^1(X,\mu) \to L^1(X,\mu)$ by

$$Tf(x) = \sum_{i=1}^{n} p_i f(T_i x), \forall f \in L^1(X,\mu), a.e.(\mu)x \in X.$$

Definition 4.1. *We call $T : L^1(X,\mu) \to L^1(X,\mu)$ a* Dunford-Schwartz operator *if*

$$\|Tf\|_1 \leq \|f\|_1 \text{ and } \|Tf\|_\infty \leq \|f\|_\infty$$

In addition, we say

1. *T is* positive *if $f_1(x) \leq f_2(x)$ implies that $Tf_1(x) \leq Tf_2(x)$ a.e. (μ)*

2. T *is* Markov *if* $T1 = 1$ *where* 1 *is a constant function.*

Finally, we say that T *is* ergodic *if* $T\chi_A = \chi_A$ *where* $A \in \mathcal{A}$ *implies that* $\mu(A) = 0$ *or* 1.

Definition 4.2. *We say that the family* $\{T_1, \cdots, T_n\}$ *is ergodic whenever* $A \in \mathcal{B}$ *satisfies* $T_i^{-1}A = A$*,* $i = 1, \cdots, n$ *then* $\mu(A) = 0$ *or* $\mu(A) = 1$*.*

If $T\chi_A(x) = \chi_A(x)$ then by convexity $T_i^{-1}A = A$ for $i = 1, \cdots, n$ and thus $\mu(A) = 0$ or 1, i.e., the operator T is ergodic.

Theorem 4.3 (Hopf-Dunford-Schwartz)**.** *If a Dunford-Schwartz operator* $T : L^1(X, \mu) \to L^1(X, \mu)$ *is ergodic then*

$$\frac{1}{n}\sum_{k=0}^{n-1} T^k\phi(x) \to \int \phi d\mu, \quad a.e.(\mu) x \in X$$

and $\phi \in L^1(X, \mu)$*.*

Remark 4.4. In particular, this subsumes the Birkhoff Ergodic Theorem by letting $n = 1$ and $p = (1)$.

The proof is slightly reminiscent of that for the von Neumann ergodic theorem. We begin with the following simple lemma (whose proof we omit).

Lemma 4.5. *Let* $Fix(T) = \{g \in L^\infty(X, \mu) : Tg = g\}$ *then* $B = Fix(T) \oplus (I - T)L^1(X, \mu) \subset L^1(X, \mu)$ *is dense (in the* L^1*-norm).*

We use the following notation $S_n\phi(x) = \sum_{k=0}^{n-1} T^k\phi(x)$, $n \geq 1$.

Lemma 4.6. *It* $\phi \in B$ *then* $\frac{1}{n}S_n\phi$ *converges in* $\|\cdot\|_\infty$*.*

Proof. Let $f = g + (I - T)h$ with $g \in \mathrm{Fix}(T)$ and $h \in L^\infty(X, \mu)$. Then $\frac{1}{n}S_n\phi = g + \frac{1}{n}(h - T^n h)$ and

$$\|\frac{1}{n}S_n\phi - g\|_\infty \leq \frac{\|h\|_\infty + \|T^n h\|_\infty}{n}$$

as $n \to +\infty$. □

Moreover, since T is ergodic we see that $\mathrm{Fix}(T)$ consists of constant functions (i.e., $\phi \in B$ implies $\frac{1}{n}S_n\phi \to \int \phi d\mu$).
Aim: We want to extend the convergence to the L^1-closure of B Then the ergodic theorem follows by Lemma **??**).

Definition 4.7. *We associate a maximal "operator" for* $f \in L^1(X, \mu)$ *by*

$$Mf(x) = \sup_n \left|\frac{1}{n}S_n f(x)\right|$$

Lemma 4.8. 1. $Mf \geq 0$ *for* $f \in L^1(X, \mu)$

2. $M(\alpha f) = |\alpha| M(f)$ *for* $f \in L^1(X, \mu)$*,* $\alpha \in \mathbb{C}$

3. *$M(f+g) \le M(f) + M(g)$ for $f, g \in L^1(X,\mu)$*

These follow from the definition of M.

Definition 4.9. *We say that $(\frac{1}{n}S_n f)$ satisfies a maximal inequality if*

$$\mu\{x : Mf(x) > \lambda\} \le \frac{\|f\|_1}{\lambda} \tag{M}$$

for all $\lambda > 0$ and $f \in L^1(X,\mu)$.

To achieve our aim we use the following.

Lemma 4.10 (Banach's Principle)**.** *Assuming (M) holds we have that*

$$C := \left\{ f \in L^1(X,\mu) : \left(\frac{S_n f}{n}\right)_{n=1}^{\infty} \text{ converges a.e.}(\mu) \right\} \subset L^1(X,\mu)$$

is a $\|\cdot\|_1$-closed subspace.

Proof. To see that C is closed: Let $f \in \overline{C} \subset L^1(X,\mu)$ and $\epsilon > 0$. Choose $g \in C$ with $\|g - f\|_1 < \epsilon$. By the triangle inequality: For a.e. (μ) $x \in X$:

$$\begin{aligned} \left|\frac{1}{k}S_k f(x) - \frac{1}{l}S_l f(x)\right| &\le \underbrace{\left|\frac{1}{k}S_k f(x) - \frac{1}{k}S_k g(x)\right|}_{\le M(f-g)(x)} \\ &+ \underbrace{\left|\frac{1}{k}S_k g(x) - \frac{1}{l}S_l g(x)\right|}_{\to 0 \text{ as } k,l\to+\infty} + \underbrace{\left|\frac{1}{l}S_l g(x) - \frac{1}{l}S_l f(x)\right|}_{\le M(f-g)(x)} \end{aligned}$$

Thus

$$h(x) = \limsup_{k,l\to+\infty} \left|\frac{1}{k}S_k f(x) - \frac{1}{l}S_l f(x)\right| \le 2M(f-g)(x).$$

Fix $\lambda > 0$ then

$$\begin{aligned} \mu\{x : h(x) > 2\lambda\} &\le \mu\{x : M(f-g)(x) > 2\lambda\} \\ &\le \frac{\|f-g\|_1}{\lambda} \qquad \text{(by (M))} \\ &\le \frac{\epsilon}{\lambda} \end{aligned}$$

Since $\epsilon > 0$ was arbitrary $\mu\{x : h(x) > \lambda\} = 0$. Since $\lambda > 0$ was arbitrary $\mu\{x : h(x) > 0\} = 0$, i.e., $h = 0$ a.e. (μ). In particular, $f \in C$ (i.e., we deduce C is closed). □

Recall that the Theorem follows from Lemma **??**, Lemma **??** and Lemma **??**.
It remains to show (M) holds. The following is the key.

Definition 4.11. *Given $f \in L^1(X,\mu)$ we define $M_n f(x) = \max_{1\le k\le n}\{S_k f(x)\}$, $n \ge 1$ for a.e. (μ) $x \in X$.*

Lemma 4.12 (Hopf)**.** *We have*

$$\int_{\{x\ :\ M_n f(x) \geq 0\}} f d\mu \geq 0$$

Proof. By definition, for $k = 2, 3, \cdots, n$:

$$S_{k-1} f(x) \leq M_n f(x) \leq \underbrace{\max\{M_n f(x), 0\}}_{=:(M_n f)_+(x)} \tag{$*$}$$

In particular, for $k = 2, \cdots, n$,

$$\begin{aligned} S_k f(x) &= f(x) + T S_{k-1} f(x) \\ &\leq f(x) + T M_n f(x) \text{ (Since } T \text{ is positive and (*))} \\ &\qquad \text{and} \\ S_1 f(x) &= f(x) \leq f(x) + \underbrace{T M_n f(x)}_{\geq 0} \text{ (Since } (M_n f)_+ \geq 0 \text{ and } T \text{ is positive)} \end{aligned}$$

and thus

$$M_n f(x) := \max_{1 \leq k \leq n} S_k f(x) \leq f(x) + T(M_n f)_+(x) \tag{$**$}$$

Therefore,

$$\begin{aligned} \int (M_n f)_+ d\mu &= \int_{\{x\ :\ M_n f \geq 0\}} M_n f d\mu \text{ (by defintion of } (M_n f)_+) \\ &\leq \int_{\{x\ :\ M_n f \geq 0\}} f d\mu + \underbrace{\int_{\{x\ :\ M_n f \geq 0\}} T M_n f d\mu}_{\leq \int (M_n f)_+ d\mu} \text{ (since } T \text{ is positive and contracts)} \end{aligned}$$

Cancelling the first and last terms gives $\int_{\{x\ :\ M_n f \geq 0\}} f d\mu \geq 0$, as required. ☐

Finally, this leads to the following

Corollary 4.13. *Let* $A_n f(x) = \max_{1 \leq k \leq n}\{\frac{S_k f(x)}{k}\}$, $n \geq 1$, $f \in L^1(X, \mu)$ *the for any* $\lambda > 0$

$$\mu\{x : A_n f(x) > \lambda\} \leq \frac{\|f\|_1}{\lambda}$$

(i.e, a "finite dimensional" version of (M))

Proof. We can apply the Hopf Lemma (Lemma **??**) to $g = \overline{f} - \lambda 1$ to get

$$\int_{\{x\ :\ M_n g > 0\}} (\overline{f} - \lambda 1) d\mu \geq 0 \tag{$*$}$$

since $\{x : M_n f = 0\} \subset \{x : g(x) \leq 0\}$ by definition of $M_n g$ (and in the set integrated over $\geq$ is replaced by $>$).

Let $1 \leq k \leq n$. Let $0 \leq j \leq k-1$ then

$$T^j g = T^j f - \lambda T^j 1 = T^j f - \lambda 1 \text{ } (T1 = 1, \text{ Markov Property})$$

Averaging over $0 \leq j \leq k-1$ we have:

$$\frac{1}{k} S_k g = \frac{1}{k} S_k f - \lambda 1.$$

Taking the supremum over $1 \leq k \leq n$ gives

$$A_n g = A_n f - \lambda \qquad (**)$$

and thus

$$\{x : A_n f(x) > \lambda\} \subset \{x : A_n g(x) > 0\} = \{x : M_n f(x) > \lambda\} \qquad (***)$$

by (**) and comparing the definitions of $A_n g$ and $M_n g$. Thus

$$\begin{aligned} 0 &\leq \int_{\{x \, : \, M_n g(x) > 0\}} (f - \lambda 1) d\mu \text{ by } (*) \\ &= \int_{\{x \, : \, A_n g(x) > 0\}} (f - \lambda 1) d\mu + \underbrace{\int_{\{x \, : \, M_n g(x) > 0, A_n f(x) \leq \lambda\}} (f - \lambda 1) d\mu}_{\leq 0 \text{ since } f \leq A_n f} \text{ using } (***) \end{aligned}$$

Thus

$$\mu\{x : A_n f(x) > \lambda\} \leq \frac{1}{\lambda} \int_{\{x \, : \, A_n f(x) > \lambda\}} f d\mu \leq \frac{\|f\|_1}{\lambda}$$

as required. □

We can now prove that (M) holds. For $f \in L^1(X, \mu)$ we have that

$$\begin{aligned} \mu\{x : \max_{1 \leq k \leq n} \left\| \frac{1}{k} S_k f(x) \right\| > \lambda\} &\leq \mu\{x : \max_{1 \leq k \leq n} \left\| \frac{1}{k} S_k f(x) \right\| > \lambda\} \text{ (since } T \text{ positive)} \\ &\leq \frac{\|f\|_1}{\lambda} \end{aligned}$$

Letting $n \to +\infty$ then

$$\max_{1 \leq k \leq n} \left\{ \frac{1}{k} S_k f(x) \right\} \nearrow \underbrace{\sup_{k \geq 1} \left\{ \frac{1}{k} S_k f(x) \right\}}_{=: Mf(x)}$$

Thus $\mu\{x : Mf(x) > \lambda\} \leq \frac{\|f\|_1}{\lambda}$ (i.e., (M) holds as required).

4.1 Application to Ergodic Theorems Free groups

There are also some interesting variations of the ergodic theorems to other groups. Recently, there was a nice proof of an ergodic theorem for the free groups. More precisely, assume that $T_1, \cdots, T_d : X \to X$ are each (invertible) measure preserving transformations that generate a free group, that is there is no relationship between the generators, i.e., some combination $T_{i_1}^{n_1} \cdots T_{i_k}^{n_k}$ with $i_1, \ldots, i_k \in \{1, \ldots, d\}$ can only be the identity map when $n_1 = \ldots = n_k = 0$. Unless your choice of transformations is unlucky, this is what one would expect.

Definition 4.14. *We say that the action is ergodic if the only sets $B \in B$ satisfying $T_i^{-1}B = B$, for every $1 \le i \le d$ have either $\mu(B) = 0$ or $\mu(X - B) = 0$.*

For $f \in L^1(X)$ and $x \in X$ we can write

$$\sigma_n f(x) = \frac{\sum_{|n_1|+\cdots+|n_k|=n} f(T_{i_1}^{n_1} \circ \cdots \circ T_{i_k}^{n_k} x)}{\#\left\{(n_1, \cdots, n_k) : |n_1| + \cdots + |n_k| = n\right\}},$$

for the average over points the "shell" corresponding to acting on x a total of n times by different combinations of $T_1, \ldots, T_d$. It is now natural to consider the averages of these terms. In the case that (X, B, μ) is a probability space, we have the following ergodic theorem.

Theorem 4.15 (Nevo-Stein, Bufetov)**.** *Let m be an ergodic measure then, for any $f \in L^1(X)$, we have*

$$\frac{1}{N}\sum_{n=0}^{N-1} \sigma_n f(x) \to \int f(x)dm(x), \ \textit{as } n \to +\infty, \ \textit{a.e.}$$

We can observe that $\#\left\{(n_1, \cdots, n_k) : |n_1| + \cdots + |n_k| = n\right\} = 2k(2k-1)^{n-1}$ (The L^2 version of this was proved much earlier by Guivarc'h). There is also an application of Rota's theorem

Proof. Consider the operator $\widehat{T} : (L^1(X, \mu))^k \to (L^1(X, \mu))^k$ defined by

$$\widehat{T}(f_1, \cdots, f_{2k}(x))_j = \frac{1}{(2k-1)} \sum_{T_i^{-1} \neq T_j} f_i(T_j x).$$

□