21. MEAN ERGODIC THEOREM

In this section we will establish the existence of "time averages" mentioned in the discussion of the Boltzmann Ergodic Hypothesis.

21.1 Mean Ergodic Theorems.

The first result we will prove is the the Mean Ergodic Theorem of J. von Neumann. This result is based on reinterpreting m.p.t.'s as linear operators on Hilbert (or Banach) spaces via the following observation:

Lemma 21.1 *Let $(\Omega, \mathcal{F}, \mu, \varphi)$ be a measure-preserving system. Then for every $p \geqslant 1$,*

$$Tf(x) = f \circ \varphi(x) \tag{21.1}$$

defines a linear operator on $L^p(\Omega, \mathcal{F}, \mu)$ with the following properties:

(1) $\|Tf\|_p = \|f\|_p$ *(isometry)*
(2) $Tf \geqslant 0$ *if* $f \geqslant 0$ *(positivity-preserving)*

If φ is a measure-preserving bimeasurable bijection, then T is unitary on $L^2(\Omega, \mathcal{F}, \mu)$.

Proof. The positivity-preserving property is immediate from (21.1) so we just need to prove that T is an isometry on L^p. This follows from $\mu \circ \varphi^{-1} = \mu$ via

$$\forall t > 0: \quad \mu\big(|f \circ \varphi| > t\big) = \mu\big(|f| > t\big) \tag{21.2}$$

which gives

$$\begin{aligned}\int |Tf|^p \mathrm{d}\mu = \int |f \circ \varphi|^p \mathrm{d}\mu &= \int_0^\infty pt^{p-1}\mu\big(|f \circ \varphi| > t\big)\mathrm{d}t \\ &= \int_0^\infty pt^{p-1}\mu\big(|f| > t\big)\mathrm{d}t = \int |f|^p \mathrm{d}\mu,\end{aligned} \tag{21.3}$$

via Tonelli's theorem. Finally, if φ is invertible then so is T. An invertible isometry on L^2 is a unitary map. □

Note that an isometry is a *contraction* which (in the lingo of the functional analysis) is a linear map with $\|T\| \leqslant 1$. Now comes the first main result:

Theorem 21.2 (Mean Ergodic Theorem, von Neumann 1932) *Let T be a contraction on a Hilbert space $\mathcal{H}$ and let P be the orthogonal projection onto the space $\{g \in \mathcal{H} : Tg = g\}$ of T-invariant vectors. Then*

$$\forall f \in \mathcal{H}: \quad \frac{1}{n}\sum_{k=0}^{n-1} T^k f \underset{n\to\infty}{\longrightarrow} Pf. \tag{21.4}$$

The proof hinges on the following observation:

Lemma 21.3 *Suppose T is a contraction on a (real or complex) Hilbert space $\mathcal{H}$ and let T^+ be its adjoint. Then*

$$\operatorname{Ker}(1 - T) = \operatorname{Ker}(1 - T^+) \tag{21.5}$$

Proof. We first claim that

$$\forall g \in \mathcal{H}\colon \quad Tg = g \quad \Leftrightarrow \quad (g, Tg) = \|g\|^2 \tag{21.6}$$

Noting that the implication $\Rightarrow$ is trivial, we focus on $\Leftarrow$. Assume g obeys $(g, Tg) = \|g\|^2$. A calculation then shows

$$\|Tg - g\|^2 = \|Tg\|^2 + \|g\|^2 - 2\mathrm{Re}(Tg, g) = \|Tg\|^2 - \|g\|^2 \tag{21.7}$$

Since T is a contraction, $\|Tg\|^2 \leqslant \|g\|^2$ and so $\|Tg - g\|^2 \leqslant 0$. Hence $Tg = g$.

The equivalence (21.6) suffices for the claim. Indeed, $g \in \mathrm{Ker}(1 - T)$ means $Tg = g$ and so $\|g\|^2 = (g, Tg) = (T^+g, g)$ by (21.6). Taking complex conjugate then gives $(g, T^+g) = \|g\|^2$ and so $T^+g = g$ and thus $g \in \mathrm{Ker}(1 - T^+)$ by (21.6) again. This proves $\mathrm{Ker}(1 - T) \subseteq \mathrm{Ker}(1 - T^+)$; the other inclusion follows by $(T^+)^+ = T$. ☐

Proof of the Mean Ergodic Theorem. Abbreviate $A_n = \frac{1}{n}\sum_{k=0}^{n-1} T^k$ and recall that for any bounded linear operator A on a Hilbert space, $\mathrm{Ker}(A) = \mathrm{Ran}(A^+)^\perp$. In light of the previous lemma,

$$\mathrm{Ker}(I - T) = \mathrm{Ker}(I - T^+) = \mathrm{Ran}(I - T)^\perp \tag{21.8}$$

Hence we get

$$\mathcal{H} = \mathrm{Ker}(I - T) \oplus \overline{\mathrm{Ran}(I - T)} \tag{21.9}$$

or, more explicitly,

$$\mathcal{H} = \overline{\{g + (h - Th)\colon g, h \in \mathcal{H},\, Tg = g\}} \tag{21.10}$$

Now pick $f = g + h - Th$ where $Tg = g$. Then

$$A_n f = \frac{1}{n}\sum_{k=0}^{n-1} T^k f = g + \frac{1}{n}\left\{\sum_{k=0}^{n-1} T^k h - \sum_{k=1}^{n} T^k h\right\} = g + \frac{1}{n}(h - T^n h) \tag{21.11}$$

But $\|h - T^n h\| \leqslant 2\|h\|$ and so $A_n f \to g$ as $n \to \infty$. Since $g = Pf$, we get $A_n f \to Pf$ for a dense set of f's. However, both A_n and P are contractions and so a 3ϵ-argument proves that $A_n f \to Pf$ for all $f \in \mathcal{H}$. ☐

The above approximation argument sits, in one or another way, at the core of many convergence results in ergodic theory, differentiation theory, etc. The point is that the convergence is easy to check on a large class of functions and one only needs a suitable continuity argument to extend it to all the functions of interest. To demonstrate this one more time, we prove a similar Mean Ergodic Theorem in any L^p with $1 < p < \infty$:

Theorem 21.4 (Mean Ergodic Theorem for contractions) *Let $\mathcal{B}$ be a reflexive Banach space (e.g., $\mathcal{B} = L^p$ for $1 < p < \infty$) and let T be a contraction on $\mathcal{B}$ (i.e., a linear map with $\|Tf\| \leqslant \|f\|$). Then there exists a linear map $P\colon \mathcal{B} \to \mathcal{B}$ such that*

$$\forall f \in \mathcal{B}\colon \quad \frac{1}{n}\sum_{k=0}^{n-1} T^k f \underset{n\to\infty}{\longrightarrow} Pf \tag{21.12}$$

Moreover, $P^2 = P$ and $\|P\| \leqslant 1$.

Proof. Using the argument (21.11), it suffices to show that the set

$$\mathscr{L} = \{g + h - Th\colon h, g \in \mathscr{X}, Tg = g\} \tag{21.13}$$

is dense in $\mathscr{X}$. Suppose $\overline{\mathscr{L}} \neq \mathcal{B}$. By the Hahn-Banach theorem, there exists a vector $\ell \in \mathcal{B}$ and a linear functional $\phi \in \mathcal{B}^*$ such that

$$\phi(\mathscr{L}) = \{0\} \;\wedge\; \phi(\ell) \neq 0 \tag{21.14}$$

Using T^* to denote the adjoint of T, the fact that $\phi(h - Th) = 0$ for all $h \in \mathscr{X}$ implies $\phi = T^*\phi$; i.e., $\phi \in \operatorname{Ker}(1 - T^*)$. But then $\phi(\ell) = T^*\phi(\ell) = \phi(T\ell)$ and, recursively, $\phi(\ell) = \phi(A_n\ell)$ where $A_n := \frac{1}{n}(T^0 + \dots + T^{n-1})$. Since $\mathcal{B}$ is reflexive, the Banach-Alaoglu Theorem tells us that the unit ball in $\mathcal{B}$ is weakly compact. From $\|A_n\ell\| \leqslant \|\ell\|$ we thus conclude the existence of a subsequence $n_k \to \infty$ such that $A_{n_k}\ell$ converges weakly to some element g; i.e.,

$$\forall \psi \in \mathcal{B}^*\colon \psi(A_{n_k}\ell) \underset{k\to\infty}{\longrightarrow} \psi(g) \tag{21.15}$$

But then also

$$\forall \psi \in \mathcal{B}^*\colon \psi(TA_{n_k}\ell) = T^*\psi(A_{n_k}\ell) \underset{k\to\infty}{\longrightarrow} T^*\psi(g) = \psi(Tg) \tag{21.16}$$

and $\|A_n g - TA_n g\| \leqslant \frac{1}{n}\|g - T^n g\| \to 0$ thus gives $\psi(Tg) = \psi(g)$ for all $\psi \in \mathcal{B}^*$ and so $g = Tg$. But $\phi(\ell) = \phi(A_n\ell)$ gives $\phi(g) = \phi(\ell) \neq 0$ in contradiction with the fact that ϕ annihilates $\mathscr{L}$.

With the convergence established, we define P by $Pf := \lim_{n\to\infty} A_n f$. Then P is linear and, since $\|A_n\| \leqslant 1$, obeys $\|P\|1$. Using $\|A_n f - TA_n f\| \to 0$ we also conclude $TP = P$. Hence also $A_n P = P$ and, taking the limit, $P^2 = P$ as desired. □

21.2 Weil's equidistribution theorem.

Here are some examples which demonstrate why the previous theorem fails for L^1 and L^∞ (which are generally not reflexive):

- $\underline{p = 1}$: Set $\mathcal{B} := L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ with $\mu :=$ Lebesgue measure and $T := f \circ \varphi$ for $\varphi(x) = x + 1$ and $f = 1_{[0,1)}$. Then $A_n f = \frac{1}{n} 1_{[0,n)}$ which is not Cauchy in L^1.
- $\underline{p = \infty}$: Set $\mathcal{B} = L^\infty([0,1), \mathcal{B}([0,1)), \mu)$ with $\mu :=$ Lebesgue measure, $\varphi(x) = x + \alpha \mod 1$ with $\alpha \notin \mathbb{Q}$. We now construct a function for which L^∞-convergence fails. Define $\{x_k\}_{k\geqslant 0}$ by $x_0 := 0$ and, recursively, $x_{k+1} := x_k - \alpha \mod 1$. Using the periodic structure, abbreviate $B(x,r) := \{y \in [0,1)\colon \min\{|x-y|, |1+x-y|\} < r\}$. Let

$$F := \bigcup_{k\geqslant 0} B(x_k, 2^{-k-3}) \tag{21.17}$$

and observe that

$$\mu(F) \leqslant \sum_{k\geqslant 0} \mu\big(B(x_k, 2^{-k-3})\big) = \sum_{k\geqslant 0} 2 \cdot 2^{-k-3} = \frac{1}{2} \tag{21.18}$$

The construction ensures that $B(x_k, 2^{-n-3}) \subseteq F$ whenever $k \leqslant n$ and so

$$\forall k \leqslant n\colon\; F^{\mathrm{c}} \cap \varphi^{-k}\big(B(0, 2^{-n-3})\big) = F^{\mathrm{c}} \cap B(x_k, 2^{-n-3}) = 0 \tag{21.19}$$

showing that $1_{F^c} \circ \varphi^k = 0$ on $B(0, 2^{-n-3})$ for $k \leqslant n$ and thus $A_n 1_{F^c} = 0$ on $B(0, 2^{-n-3})$. If $A_n 1_{F^c}$ converged in L^∞, Theorem 21.5 below would identify the limit to be the constant $\int 1_{F^c} d\mu = \mu(F^c)$; i.e., $A_n 1_{F^c} \to \mu(F^c)$ in L^∞. But $\mu(B(0, 2^{-n-3})) > 0$ gives $\|A_n 1_{F^c} - \mu(F^c)\|_\infty \geqslant \mu(F^c) > 0$, making L^∞-convergence impossible.

As we will see after we prove Birkhoff's Pointwise Ergodic Theorem, for *finite* measure spaces the theorem is still correct in L^1 and the convergence does hold pointwise a.e., just not uniformly a.e.

In the case of irrational rotations of the unit circle, the above results can be proved and even strengthened by invoking Fourier series representation:

Theorem 21.5 (Weil's Equidistribution Theorem) *Let $\alpha \notin \mathbb{Q}$ and $\varphi_\alpha(x) = x + \alpha$ mod 1.*

(1) *If f measurable with $f \circ \varphi_\alpha = f$ a.e. then f is a constant a.e.*

(2) *For all $f \in L^2([0,1])$,*

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ \varphi_\alpha^k \xrightarrow[n\to\infty]{L^2} \int_0^1 f(x)\mathrm{d}x \tag{21.20}$$

The convergence is uniform if f is 1-periodic and the class $C^2(\mathbb{R})$.

(3) *If f is Riemann integrable, then the convergence occurs uniformly on $[0,1]$.*

Proof. (1) Suppose f is measurable with $f \circ \varphi = f$ a.e. Then for each $t \in \mathbb{R}$, $g_t := 1_{\{f \leqslant t\}}$ obeys $g_t \circ \varphi_\alpha = g_t$ a.e. Since g_t is bounded and measurable, its Fourier coefficients obey

$$\hat{g}_t(n) := \int_0^1 g_t(x)\mathrm{e}^{2\pi \mathrm{i} n x}\mathrm{d}x = \int_0^1 g_t(x)\mathrm{e}^{2\pi \mathrm{i} n(x-\alpha)}\mathrm{d}x = \mathrm{e}^{-2\pi \mathrm{i} n\alpha}\hat{g}_t(n) \tag{21.21}$$

As $\alpha \notin \mathbb{Q}$, this only possible if $\hat{g}_t(n) = 0$ for all $n \in \mathbb{Z} \smallsetminus \{0\}$. Since g equals its Fourier representation $\sum_{n\in\mathbb{Z}} c_n \mathrm{e}^{2\pi \mathrm{i} n x}$ for a.e. x, we get that g_t is constant a.e. But this can only hold for all $t \in \mathbb{R}$ if f is constant a.e.

(2) Let $f \in C^2(\mathbb{R})$ with $f(x+1) = f(x)$ for all $x \in \mathbb{R}$. Then f admits a uniformly convergent Fourier representation

$$\forall x \in \mathbb{R}: \quad f(x) = \sum_{n\in\mathbb{Z}} c_n \mathrm{e}^{2\pi \mathrm{i} n x} \tag{21.22}$$

where $c_n := \int_0^1 f(x)\mathrm{e}^{-2\pi \mathrm{i} n x}\mathrm{d}x$ are such that $\sum_{n\in\mathbb{Z}} |c_n| < \infty$. This allows us to write

$$A_n f(x) = \frac{1}{N}\sum_{k=0}^{N-1} f \circ \varphi_\alpha(x) = \sum_{n\in\mathbb{Z}} c_n \left(\sum_{k=0}^{N-1} (\mathrm{e}^{2\pi \mathrm{i} n\alpha})^k\right)\mathrm{e}^{2\pi \mathrm{i} n x} \tag{21.23}$$

But

$$\frac{1}{N}\sum_{k=0}^{N-1} (\mathrm{e}^{2\pi \mathrm{i} n\alpha})^k = \begin{cases} \dfrac{1}{N}\dfrac{1-\mathrm{e}^{2\pi \mathrm{i} N n\alpha}}{1-\mathrm{e}^{2\pi \mathrm{i} n\alpha}} \underset{N\to\infty}{\longrightarrow} 0, & \text{if } n \neq 0 \\ 1, & \text{if } n = 0 \end{cases} \tag{21.24}$$

and so $A_n f(x) \to c_0$ uniformly for any 1-periodic $f \in C^2(\mathbb{R})$. But such f's are dense in L^2 and since the A_n's are contractions in L^2, same is true, albeit only in the sense of L^2-convergence, for all $f \in L^2$.

(3) For each Riemann integrable function $f\colon [0,1] \to \mathbb{R}$ and each $\epsilon > 0$, there exist step functions g and h of the form $\sum_{j=1}^n c_j 1_{[a_j,b_j]}$ such that $g \leqslant f \leqslant h$ and

$$\int_0^1 (f-g)\mathrm{d}x < \epsilon \quad \text{and} \quad \int_0^1 (h-f)\mathrm{d}x < \epsilon \tag{21.25}$$

Since $A_n g \leqslant A_n f \leqslant A_n h$, it suffices to prove the claim for step functions and, by linearity, for functions of the form $1_{[a,b]}$. But then we can find C^2-functions $\tilde{g}, \tilde{h}\colon [0,1] \to \mathbb{R}_+$ such that $\tilde{g} \leqslant 1_{[a,b]} \leqslant \tilde{h}$ and

$$\int_0^1 (1_{[a,b]} - \tilde{g})\mathrm{d}x < \epsilon \quad \text{and} \quad \int_0^1 (\tilde{h} - 1_{[a,b]})\mathrm{d}x < \epsilon \tag{21.26}$$

The argument in (2) shows that $A_n\tilde{g}(x) \to \int_0^1 \tilde{g}\mathrm{d}z$ and $A_n\tilde{h}(x) \to \int_0^1 \tilde{h}\mathrm{d}z$ uniformly in $x \in [0,1]$. It follows that $A_n f \to \int_0^1 f\mathrm{d}z$ uniformly on $[0,1]$. □