Von Neumann Ergodic Theorem

The standard Birkhoff Ergodic Theorem says that the time average of some function f is equal to its space average (expectation) when the transformation T is measure preserving. Von Neumann's ergodic theorem says that the time average is equal to a *part* of f that is unaffected by—or invariant under—the transformation T. We prove Von Neumann's version below.

Von Neumann Ergodic Theorem: *Let $T : X \to X$ be a measure preserving transformation with respect to measure μ. And let $P : L^2(X, \mu) \to L^2(X, \mu)$ be the orthogonal projection onto the subspace $L^2(X, \mu, I)$ of T-invariant functions. Then for any $f \in L^2(X, \mu)$, we have in L^2* that

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \xrightarrow{N \longrightarrow \infty} Pf.$$

Proof. To start, let f be T-invariant i.e. $f \in L^2(X, \mu, I)$. Then

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n = \frac{1}{N} \sum_{n=0}^{N-1} f \longrightarrow f = Pf$$

in L^2 trivially. Now suppose f is not necessarily T-invariant, but that $f = g - (g \circ T)$ for some $g \in L^2(X, \mu)$. Then we have

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n &= \frac{1}{N} \sum_{n=0}^{N-1} (g - g \circ T) \circ T^n \\ &= \frac{1}{N} \sum_{n=0}^{N-1} g \circ T^n - \frac{1}{N} \sum_{n=0}^{N-1} g \circ T^{n+1} \\ &= \frac{1}{N} g \circ T^{n+1}. && \text{(notice the telescoping terms)} \end{aligned}$$

We know that this function tends to 0 in L^2 because

$$\begin{aligned} \left\| \frac{1}{N}(g - g \circ T^{n+1}) \right\|_2 &\leq \left\| \frac{1}{N} g \right\|_2 + \left\| \frac{1}{N} g \circ T^{n+1} \right\|_2 && \text{(Triangle Inequality)} \\ &= \frac{1}{N}(\|g\|_2 + \|g\|_2) && \text{(T preserves norms)} \\ &= \frac{2}{N} \|g\|_2 \xrightarrow{N \longrightarrow \infty} 0. \end{aligned}$$

Now suppose f is the limit of a sequence of functions of the same form i.e. $f_k \longrightarrow f$ pointwise and $f_k = g_k - (g_k \circ T)$ for $g_k \in L^2(X, \mu)$. We claim that the sequence of averages of $f \circ T^n$ tends to 0 as more terms are included in each average. We choose k such that $\|f - f_k\|_2 \leq \epsilon$. We have

$$\begin{aligned} \left\| \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n - 0 \right\|_2 &= \left\| \frac{1}{N} \sum_{n=0}^{N-1} f_k \circ T^n - \frac{1}{N} \sum_{n=0}^{N-1} (f - f_k) \circ T^n \right\|_2 \\ &\leq \left\| \frac{1}{N} \sum_{n=0}^{N-1} g - g_k \circ T^n \right\|_2 + \frac{1}{N} \sum_{n=0}^{N-1} \|f - f_k\|_2 && \text{(Triangle Inequality and T preserves norms)} \\ &\leq \left\| \frac{1}{N} \sum_{n=0}^{N-1} g - g_k \circ T^n \right\|_2 + \epsilon \xrightarrow{N \longrightarrow \infty} \epsilon. && \text{(chose k accordingly, and first term goes to 0)} \end{aligned}$$

Hence, $\left\| \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \right\|_2 \xrightarrow{N \longrightarrow \infty} 0$ when f is the limit of some sequence of functions of that particular form.

We have just shown if $f \in \overline{\{g - (g \circ T) \mid g \in L^2(X, \mu)\}}$ then its sequence of averages tends to 0 as we include more terms. Now we claim that $L^2(X, \mu, I)$, the set of T-invariant functions, is the orthogonal complement to $\{g - (g \circ T) \mid g \in L^2(X, \mu)\}$. If this is true (we assume it is, then show it is after), then $f = Pf + f_\perp$ for some $f_\perp \in \{g - (g \circ T) \mid g \in L^2(X, \mu)\}$. Thus, we have

$$\begin{aligned}
\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n &= \frac{1}{N}\sum_{n=0}^{N-1}(Pf + f_\perp) \circ T^n \\
&= \frac{1}{N}\sum_{n=0}^{N-1} Pf \circ T^n + \frac{1}{N}\sum_{n=0}^{N-1} f_\perp \circ T^n \\
&\xrightarrow{N \longrightarrow \infty} Pf + 0 = Pf
\end{aligned}$$

where the second term in the second line tends to 0 because what we have shown above.

Now we must prove that the set of T-invariant functions is orthogonal to the set of functions who take the form of $g - (g \circ T)$, i.e. $L^2(X, \mu, I) = \{g - (g \circ T) \mid g \in L^2(X, \mu)\}^\perp$. For the forwards inclusion, let $f \in L^2(X, \mu, I)$ and $g \in L^2(X, \mu)$. Then

$$\begin{aligned}
\langle f, g - (g \circ T)\rangle &= \langle f, g\rangle - \langle f, g \circ T\rangle \\
&= \langle f, g\rangle - \langle f \circ T, g \circ T\rangle \quad && \text{(since } f \text{ is } T\text{-invariant)} \\
&= \langle f, g\rangle - \langle f, g\rangle && \text{(T is measure preserving and so preserves inner products)} \\
&= 0.
\end{aligned}$$

Now, for the opposite inclusion, suppose $\langle f, g - (g \circ T)\rangle = 0$ for $f, g \in L^2(X, \mu)$. Pick f such that $f = g$. Then we have

$$0 = \langle f, g - (g \circ T)\rangle = \langle f, f\rangle - \langle f, f \circ T\rangle$$

which implies $\langle f, f\rangle = \langle f, f \circ T\rangle$. Consider the norm of the difference between f and $f \circ T$. We have

$$\begin{aligned}
\|f - (f \circ T)\|_2^2 &= \langle f - (f \circ T), f - (f \circ T)\rangle \\
&= \langle f, f\rangle - 2\langle f, f \circ T\rangle + \langle f \circ T, f \circ T\rangle \\
&= 2\langle f, f\rangle - 2\langle f, f \circ T\rangle \quad && \text{(because } T \text{ preserves inner products)} \\
&= 2(\langle f, f\rangle - \langle f, f \circ T\rangle) \\
&= 0.
\end{aligned}$$

Thus, $f = f \circ T$, so $f \in L^2(X, \mu, I)$, and we are done.

$\square$