1. Ergodicity

**Definition 1.** Let $T : X \to X$ preserve $\mu$. $T$ is ergodic with respect to $\mu$ if $\mu(T^{-1}A\Delta A) = 0$ implies $\mu(A)\mu(A^c) = 0$.

Other authors will often not require that $T$ is $\mu$-measure preserving. Then this is just a condition on a measure class.

Below are some equivalent conditions (assuming this is a Lebesgue space):

1. If $A = T^{-1}A$ and $A$ is measurable then $\mu(A)\mu(A^c) = 0$.
2. Every measurable function $f : X \to \mathbb{R}$ that is almost everywhere $T$ invariant (i.e. $f(Tx) = f(x)$ for $\mu$-almost every $x$) is a.e. constant.
   
   To see this assume not. Then there exists $c$ so that $\mu(f^{-1}(-\infty, c))$ and $\mu(f^{-1}([c, \infty))) > 0$. These are $T$-invariant sets.
3. If $\mu(X) \leq \infty$ and $f(x) \leq f(Tx)$ for all $x$ then $f$ is a.e. constant.
4. If $\mu(X) < \infty$ then $\bigcup_{i=1}^{\infty} T^{-i}A$ has full measure.
5. If $\mu(X) < \infty$, $\mu(A), \mu(B) > 0$ then there exists an $n$ so that $\mu(T^{-n}A \cap B) > 0$.
6. If $\mu(X) < \infty$, $f \in L^1(\mu)$ and $f \circ T = f \mu$ a.e. then $f$ is constant almost everywhere.
7. If $\mu(X) < \infty$ then every measurable $L^2$ function that is a.e $T$ invariant is a.e. constant.

We will show that it is equivalent that $T$ is ergodic and every $\mu$-measurable function is almost everywhere constant and leave the other equivalences as an exercise. Indeed if $T$ is not ergodic then there exists measurable $A$ so that $\mu(A\Delta T^{-1}A) = 0$ and $\mu(A)$ and $\mu(A^c) > 0$. So $\chi_A$ is an almost everywhere $T$-invariant measurable function that is not constant almost everywhere. If $f$ is a measurable function that is not constant almost everywhere then there exists $c$ so that $\mu(f^{-1}(\infty, c)) > 0$ and $\mu(f^{-1}[c, \infty)) = \mu((f^{-1}(-\infty, c))^c) > 0$. Thus if $f$ is also almost everywhere $T$-invariant then $A = f^{-1}(-\infty, c)$ is measurable, satisfies $\mu(T^{-1}A\Delta A) = 0$ by $\mu(A)\mu(A^c) \neq 0$.

The next theorem, which is also called the pointwise ergodic theorem, is one of the central results in ergodic theory. It says that an ergodic measure controls the average behavior of the typical point with respect to that measure.
Theorem 1. (Birkhoff) Let \((X, \mathcal{B}, \mu)\) be a probability space. Let \(T\) be \(\mu\) ergodic and \(f \in L^1(\mu)\). Then

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x) = \int f \, d\mu
\]

for \(\mu\) a.e. \(x\). Moreover, let \(s_n(x) = \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x)\). Then \(s_n\) converges in \(L^1(\mu)\) to \(\int f \, d\mu\).

If \(T\) is just measure preserving then the limit exists for almost every \(x\). The resulting function is \(T\) invariant.

Definition 2. The term \(\sum_{i=0}^{N-1} f(T^i x)\) is called a Birkhoff sum. The term \(\frac{1}{N} \sum_{i=0}^{N-1} f(T^i x)\) is called a Birkhoff average.

We will use this result from now on, though we will prove this later.

1.1. The case of continuous functions. The following is a consequence of the Birkhoff Ergodic Theorem.

Theorem 2. Let \((X, d)\) be a compact metric space, \(\mu\) be a (regular) probability measure on its Borel-\(\sigma\) algebra and \(T\) be \(\mu\) ergodic. There exists \(G_\mu\) so that \(\mu(G_\mu) = 1\) and

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x) = \int f \, d\mu
\]

for every \(f \in C(X)\) and \(x \in G_\mu\).

Observe how the quantifiers have been improved. Rather than having a full measure set that depends on the function, there is a full measure set that works for a (probably uncountable) set of functions simultaneously.

Lemma 1. Let \((X, d)\) be a compact metric space. Then \((C(X), \| \cdot \|_{\text{sup}})\) has a countable dense subset.

Proof. Consider \((C(X), \| \cdot \|_{\text{sup}})\) sitting inside bounded functions with \(\| \cdot \|_{\text{sup}}\). We will show that there exists a countable subset of this space whose closure contains \((C(X), \| \cdot \|_{\text{sup}})\).

Let \(B_1, ..., B_n\) be a countable family of metric balls which form a base for the topology. Consider sequence of families of functions:

\[
\mathcal{F}_n = \{g : X \to \mathbb{R}| g = \sum_{i=1}^{n} a_i \chi_{B_i} \text{ where } a_1, ..., a_n \in \mathbb{Q}\}.
\]
Observe that each \( F_n \) is countable and the \( \| \cdot \|_{\text{sup}} \) closure of \( \cup_{n=1}^{\infty} F_n \) contains \((C(X),\| \cdot \|_{\text{sup}})\). (See Problem 1.) This implies that \((C(X),\| \cdot \|_{\text{sup}})\) contains a countable dense subset. (See Problem 2.)  

\[ \square \]

**Proof of Theorem 2.** First, by the previous lemma there exists \( S \) a countable dense subset of \((C(X),\| \cdot \|_{\text{sup}})\). For each \( f \in S \) let

\[ G_f = \{ x : \lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x) = \int f \, d\mu \}. \]

The Birkhoff ergodic theorem guarantees that \( \mu(G_f) = 1 \). Now because \( S \) is countable (and we are in a probability space) \( \mu(\cap_{f \in S} G_f) = 1 \). To complete the proof of the theorem we now show that if \( x \in \cap_{f \in S} G_f \) and \( h \in C(X) \) then for any \( \epsilon > 0 \) we have

\[ \limsup_{N \to \infty} \left| \frac{1}{N} \sum_{i=0}^{N-1} h(T^i x) - \int h \, d\mu \right| < \epsilon. \]

There exists \( f \in S \) so that \( \| f - h \|_{\text{sup}} \leq \frac{\epsilon}{3} \) and since \( \mu \) is a probability measure this implies that \( | \int (f - h) \, d\mu | \leq \frac{\epsilon}{3} \). Thus

\[
\left| \frac{1}{N} \sum_{i=0}^{N-1} h(T^i x) - \int h \, d\mu \right| \leq \left| \sum_{i=0}^{N-1} h(T^i x) - \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x) \right| + \left| \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x) - \int f \, d\mu \right| + \left| \int (f-h) \, d\mu \right|
\]

which is at most \( \epsilon \) for all large enough \( N \). Indeed our choice of \( f \) implies that the first and third terms are at most \( \frac{\epsilon}{3} \) for all \( N \). Since \( x \in G_f \), we have that the second term is at most \( \frac{\epsilon}{3} \) for all \( N \) large enough. Since we have this for any \( \epsilon \), we have established (1) and by extension the theorem. \( \square \)

The next example suggests that we should not be able to improve the condition of continuous in the previous theorem to measurable.

**Example 1.** Let \((X,B,\mu)\) be a non-atomic measure space where each point is measurable. Let \( T \) be \( \mu \) measure preserving (possibly ergodic)

\[ \textbf{Briefly, it suffices to show that for every } \epsilon > 0 \text{ there exists a countable subset of } (C(X),\| \cdot \|_{\text{sup}}). \]

For each element of \( g \in \cup_{n=1}^{\infty} F_n \) let \( f_g \) be an element of \( B(g,\frac{\epsilon}{2}) \cap C(X) \) if it exists.

\[ \{ f_g : g \in \cup_{n=1}^{\infty} F_n \text{ and } B(g,\frac{\epsilon}{2}) \cap C(X) \neq \emptyset \} \]

is a countable \( \epsilon \) dense subset of \((C(X),\| \cdot \|_{\text{sup}})\) that is \( \epsilon \) dense.
and $f \in L^1(X, \mu)$. Observe that $f_c = \begin{cases} c & \text{if } x \in T^i p \\ f(x) & \text{else} \end{cases}$ is measurable and

$$\lim_{N \to \infty} \sum_{i=0}^{N-1} f_c(T^i x) = c$$

but $\int f_c d\mu = \int f d\mu$.

What goes wrong with proof of Theorem 2 for these functions?

1.1.1. A brief digression on the weak star topology. Context: We will be restricting our attention to the case when $(X, d)$ is a compact metric space. There are analogues for the current discussion if $(X, d)$ is a $\sigma$-compact metric space or even when $(X, \mathcal{T})$ is a locally compact Hausdorff space.

Recall that given a complex measure $\mu$ there exists an associated measure $|\mu|$, called the total variation of $\mu$. There is $f \in L^1(|\mu|)$ so that $\mu(A) = \int_A |d\mu|$ for all measurable $A$. There is a norm on complex measures, the total variation, where the total variation of $\mu$ is $|\mu|(X)$.

**Theorem 3.** Borel regular (complex) measures with finite total variation is the dual of $(C(X), \| \cdot \|_{\sup})$.

Frequently, one is really concerned with understanding how measures interact with $C(X)$. In this case one typically puts weak star topology on the set of measures. The open sets for this topology are generated by $N_{f, \epsilon} = \{ \mu : |\int f d\mu - \int f d\nu| < \epsilon \}$ where $f \in C(X)$. It follows from this definition that $\mu_1, ...$ converges to $\mu_\infty$ in the weak star topology if for all continuous functions we have $\lim_{n \to \infty} \int f d\mu_n = \int f d\mu_\infty$. This is a topology of pointwise convergence, where the total variation norm (that is the dual space norm) requests uniform convergence.

Recall that this topology is frequently not metrizable but the set $\{ \mu : |\mu|(X) \leq 1 \}$ is metrizable in this topology. We have the following corollary to Theorem 2:

**Proposition 1.** Let $X$ be a compact metric space, $\mu$ be a probability measure on its Borel-$\sigma$ algebra and $T$ be $\mu$ ergodic. Let $\delta_y$ be the measure point mass at $y$. For $\mu$-a.e. $x$ we have that $\frac{1}{n+1} \sum_{i=0}^{n} \delta_{T^i x}$ converges to $\mu$ in the weak-* topology.

Frequently this set of measures does not converge in the total variation norm. For instance, if we let $\delta_\frac{1}{n}$, Borel measures on $[0, 1]$ with the Euclidean metric, we have that the total variation of $\delta_\frac{1}{n} - \delta_0$ is 2 for all $n \in \mathbb{N}$. So $\delta_\frac{1}{n}$ does not converge to $\delta_0$ in the topology coming from the total variation norm. Similarly, if $\nu_i$ are a sequence of atomic measures then the weak-* limit can be a non-atomic measure, but the
limit in the total variation norm must be an atomic measure (if these limits exist).

Theorem 2 shows that (under its assumptions) \( \mu \) a.e. \( x \) satisfies that the sequence of measure \( \delta_x, \frac{1}{2}(\delta_x + \delta_{Tx}), \ldots \) converges in the weak-* topology to \( \mu \). For context:

**Theorem 4.** (Banach-Aloagolu) The weak star topology restricted to measures with total variation at most 1 is compact.

We won’t be proving this.

Also note that if \((X, d)\) is a compact metric space then \( \chi_X \in C(X) \). So any weak-* limit point of \( \delta_x, \frac{1}{2}(\delta_x + \delta_{Tx}), \ldots \) \( \nu \) has that \( \nu(X) = 1 \).

It follows that probability measures are a closed subset in the weak-* topology in this context.

1.1.2. Back to ergodic theorems for continuous functions.

**Theorem 5.** (Krylov-Bogoliubov) Let \((X, d)\) be a compact metric space (and \( \mathcal{B} \) be its Borel \( \sigma \)-algebra). If \( T : X \to X \) is continuous then \( T \) preserves a probability measure.

**Proof.** Step 1: Pick a point \( x \). There exists a subsequence \( N_1, \ldots \) so that for every continuous function \( \lim_{k \to \infty} \frac{1}{N_k} \sum_{i=0}^{N_k-1} f(T^i x) \) exists.

This is another useful argument which is outlined.

Step a: For any sequence natural numbers \( m_1, \ldots, \) for any function \( f \in C(X) \) and point \( x \) there exists a subsequence \( m_{k_i} \) so the limit \( \lim_{j \to \infty} \frac{1}{m_{k_j}} \sum_{i=0}^{m_{k_j}-1} f(T^i x) \) exists.

Step b: The same results can hold simultaneously for any set finite set \( f_1, \ldots, f_n \).

Step c: The same result can hold simultaneously for any countable set.

Step d: Choosing our countable set in step c to be dense in supremum norm, we can obtain Step 1, (by a similar argument to the proof of Theorem 2).

**Step 2:** Let \( x \) and \( N_1, \ldots \) be as in Step 1. The map \( f \to \lim_{k \to \infty} \frac{1}{N_k} \sum_{i=0}^{N_k-1} f(T^i x) \) is a (positive) continuous linear functional and so it gives us a (positive) measure \( \mu \). This measure has \( \mu(X) = 1 \).

**Step 3:** The linear functional is \( T \)-invariant. Indeed,

\[
\sum_{i=0}^{N-1} f \circ T(T^i x) = \sum_{i=0}^{N-1} f(T^i x) + f(T^N x) - f(x).
\]
Since we are continuous on a compact space the two differing terms are bounded. So \( \lim_{k \to \infty} \frac{1}{N_k} \sum_{i=0}^{N_k-1} f(T^i x) = \lim_{k \to \infty} \frac{1}{N_k} \sum_{i=0}^{N_k-1} f(T^i x) \). So the measure is \( T \) invariant.

**Step 4:** The measure \( \mu \) is \( T \)-invariant.\(^2\) By the outer regularity of Borel measures, it suffices to show that \( \mu(U) = \mu(T^{-1}(U)) \) for all open sets \( \mu \). By the proof of the Riesz representation theorem

\[
\mu(U) = \sup \left\{ \lim_{k \to \infty} \frac{1}{N_k} \sum_{i=0}^{N_k-1} f(T^i x) : f \in C(X), \text{supp}(f) \subset U \text{ and } 0 \leq f \leq 1 \right\}.
\]

Since \( f \) is continuous \( f \circ T \) is as well. So for any \( f \) we are using to approximate \( \mu(U) \) we can use \( f \circ T \) to approximate \( \mu(T^{-1}U) \), which by step 3 means the measure of \( \mu(T^{-1}U) \) is the supremum of a set which contains the numbers that \( \mu(U) \) is the supremum of. That is, because \( f \circ T \) is continuous,

\[
\{ \int f d\mu : f \in C(X), \text{supp}(f) \subset U \text{ and } 0 \leq f \leq 1 \} \subset \\
\{ \int g d\mu : g \in C(X), \text{supp}(f) \subset T^{-1}U \text{ and } 0 \leq g \leq 1 \}.
\]

So \( \mu(T^{-1}U) \geq \mu(U) \).

In our setting Borel measures are also inner regular, so \( \mu(T^{-1}U) = \sup \{ \mu(K) : K \subset T^{-1}U \text{ and } K \text{ is compact} \} \). Now \( T(K) \) is also compact and it is contained in \( U \). So if we show that \( \mu(K) \geq \mu(TK) \) then since \( \mu(U) \geq \sup \{ \mu(TK) : K \subset T^{-1}U, K \text{ compact} \} \) we will have the other inequality. Now

\[
\mu(K) = \inf \{ \int f d\mu : f \in C(X), 0 \leq f \leq 1 \text{ and } f|_K = 1 \}.
\]

For each \( f \in C(X) \) with \( f|_{TK} = 1 \) we have \( f \circ T|_K = 1 \). So \( \mu(K) \leq \mu(TK) \) because it is an infimum over a (possibly) larger set of numbers.

\[ \square \]

This proof

1. needs the compactness of \( X \) and
2. the continuity of \( T \).

This is not just a feature of the proof. See Problem 6.

\(^2\)This is where the continuity of \( T \) is used.
This proof can be interpreted by saying that (if $T$ is continuous) for every $x$ the weak-* limit points of \( \{\delta_x, \frac{1}{2}(\delta_x + \delta_{Tx}), \ldots\} \) are $T$--invariant. We say a measure is quasi-generic if it arises in this way.

The next theorem states a uniform convergence result for certain ergodic measures under a restrictive additional assumption.

**Theorem 6.** Let $(X,d)$ be a compact metric space, $(\mathcal{B}$ is its Borel $\sigma$-algebra), $T : X \to X$ continuous and $\mu$ is the unique preserved probability measure of $T$. If $f \in C(X)$ and $\epsilon > 0$ then there exists $M_{\epsilon,f}$ so that if $N > M_{\epsilon,f}$ then

\[
|\frac{1}{N} \sum_{i=0}^{N-1} f(T^i x) - \int f \, d\mu| < \epsilon.
\]

**Proof.** Step 1: We first claim that for every continuous $f$, $\epsilon > 0$ and $x \in X$ there exists $M$ so that for all $N > M$ we have $|\frac{1}{N} \sum_{i=0}^{N-1} f(T^i x) - \int f \, d\mu| < \epsilon$.

We prove this by contradiction, assuming that it is not true and then using the proof of the Krylov-Bogoliubov theorem to find another invariant measure (to contradict the assumption).

Negating the conclusion of Step 1 says that there exists a continuous function $f$, a point $x \in X$, an $\epsilon > 0$, and a sequence going to infinity $N_1, \ldots$ so that for every $i$ we have $|\frac{1}{N_i} \sum_{i=0}^{N_i-1} f(T^i x) - \int f \, d\mu| > \epsilon$. We can choose a subsequence $N_{i_k}$ so that $\lim_{k \to \infty} \frac{1}{N_{i_k}} \sum_{i=0}^{N_{i_k}-1} f(T^i x)$ exists. It is necessarily different from $\int f \, d\mu$. Now, we can repeat the proof of Step 1 of the Krylov-Bogoliubov theorem and chooses a further subsequence so that the limit exists for all continuous functions. As in the remaining steps of the proof of the Krylov-Bogoliubov Theorem, this gives a continuous linear function on continuous functions and eventually a $T$ invariant probability measure. It disagrees with $\mu$ on $f$ so it is different, contradicting out assumptions.

Step 2: There exists $K < \infty$ so that for every $x$ there exists $N_x \leq K$ with $|\frac{1}{N_x} \sum_{i=0}^{N_x-1} f(T^i x) - \int f \, d\mu| < \epsilon$.

For each $N$ the set of $x$ so that $N$ could be $N_x$ is open. By Step 1 the union over $N$ is $X$. So by compactness the compactness of $X$, there exists a finite subcover. Choose $K$ to be the largest $N$.

Step 3: Completion of proof via breaking up orbits.\[3\]

\[3\]Notice a subtlety: this $K$ is not the $K$ in Step 1. We just want some number less than $K$ where the Birkhoff average is close to the integral, not all beyond this number. This is so that we get an open condition and can use compactness.
\[
\sum_{i=0}^{R} f(T^i x) = \sum_{i=0}^{N_x-1} f(T^i x) + \sum_{i=N_x}^{N_x+N_Tx-1} f(T^i x) + \ldots. \\
\text{The last term with at most } K \text{ summands, and so has size at most } K \|f\|_{\text{sup}} \text{ independent of } R. \text{ A term of the form } |\sum_{i=T^ix}^{N_Tx-1} f(T^i x)| \in (\int f \, d\mu - \epsilon, \int f \, d\mu + \epsilon). \text{ So we have that}
\]

\[
\sum_{i=0}^{R-1} f(T^i x) \in [(R - K)(\int f \, d\mu - \epsilon) - K \|f\|_{\text{sup}}, R(\int f \, d\mu + \epsilon + K \|f\|_{\text{sup}})].
\]

So we have \( \limsup_{R \to \infty} \frac{1}{R} \sum_{i=0}^{R-1} f(T^i x) - \int f \, d\mu \leq \epsilon. \) \qed

Note that Step 1 of the proof of the theorem is interesting in its own right. It says that if \( T \) has only one invariant measure, \( \mu \), then the conclusion of Theorem 2 can be improved to say that for every point \( x \) and every continuous function \( f \) we have that

\[
\frac{1}{N} \sum_{i=0}^{N-1} f(T^i x) \to \int f \, d\mu.
\]

Note this is not true for \( L^1 \) functions because if \( f \) is continuous, \( T \) is aperiodic and \( a_1, \ldots \in \mathbb{R} \) then \( \tilde{f}(x) = \begin{cases} a_i & \text{if } x = T^i p \\ f(x) & \text{else} \end{cases} \) is measurable and its integral agrees with \( f \) but we can proscribe its value on the orbit of \( p \) to be whatever we want.

1.2. Another ergodic theorem (with proof). The following is an ergodic theorem that follows by general Hilbert space arguments.

**Theorem 7.** (Von Neumann Ergodic Theorem) Let \( f \in L^2(\mu) \) and \( T \) be \( \mu \) ergodic. Then \( \frac{1}{N} \sum_{i=0}^{N-1} f \circ T^i \) converges to \( \int f \, d\mu \) in \( L^2 \) norm.

What does this mean? How is it different from saying the Birkhoff Ergodic Theorem holds if we require \( f \in L^1(\mu) \cap L^2(\mu) \)?

**Proof.** Step 0: Let \( M = \{ f : U_T f = f \} \). Notice that \( \frac{1}{R} \sum_{i=0}^{R-1} U_T^i \) acts as the identity on \( M \). Let \( N = \{ g : \exists f \in L^2 \text{ so that } g = f - U_T f \} \).

Notice that the \( \| \cdot \|_2 \) limit of \( \sum_{i=0}^{R-1} U_T^i g \) is 0 for all \( g \in N \).

To see the second claim: \( \frac{1}{N} \sum_{i=0}^{N-1} U_T^i (h - h \circ T) = \frac{1}{N} (h - h \circ T^N) \). The norm of this is at most \( \frac{2}{N} \|h\|_2 \).

**Step 1:** We claim that \( M \) is the orthogonal complement of \( N \).

It is clear that \( M \) is contained in the orthogonal complement of \( N \). Now if \( f \) is in \( N^\perp \) then for every \( h \in L^2 \) we have \( < f, h - U_T h > = 0. \)
So \(< f - U_T f, h > = 0\). Now this implies \(f - U_T^* f = 0\). Now

\[
\|f - U_T f\|^2 = < f - U_T f, f - U_T f > = 2\|f\|^2 - < f, U_T f > - < U_T f, f > = 2\|f\|^2 - 2 < f, f > = 0.
\]

**Step 2:** For any \(f\) in the closure of \(N\) we have that \(\frac{1}{N} \sum_{i=0}^{N-1} U_T^i f\) converges to 0 in \(L^2(\mu)\).

Let \(f \in \bar{N}\), the closure of \(N\). For every \(\epsilon > 0\) there exists \(g \in N\) so that \(\|f - g\| < \epsilon\). Now

\[
\|\frac{1}{N} \sum_{i=0}^{N-1} U_T^i(f)\| = \|\frac{1}{N} \sum_{i=0}^{N-1} U_T^i(g) + \frac{1}{N} \sum_{i=0}^{N-1} U_T^i(f - g)\| \leq \|\frac{1}{N} \sum_{i=0}^{N-1} U_T^i(g)\| + \|\frac{1}{N} \sum_{i=0}^{N-1} U_T^i(f - g)\|.
\]

By Step 1, for large enough \(N\) the first term is at most \(\epsilon\). Since \(U_T^i\) is an isometry for all \(i\) the second term is at most \(\|f - g\| \leq \epsilon\).

**Completing the proof:** Now by linearity \(\frac{1}{N} \sum_{i=0}^{N-1} U_T^i\) converges for each \(L^2\) function \(^7\) to projection onto \(M\). Since ergodicity implies that constant functions are the \(T\)-invariant functions, this implies the Von Neumann ergodic theorem.

\[\square\]

### 1.3. Problems.

1. Show that the \(\|\cdot\|_{\text{sup}}\) closure of \(\bigcup_{n=1}^\infty F_n\) contains \(C(X)\) (notation is as in Lemma 1).

2. Show that if \((X, d)\) is a metric space with a countable dense subset and \(Y \subset X\) is closed then \(Y\) has a countable dense

---

4Recall that in Hilbert space if \(< f, h > = 0\) for all \(h\) then \(f = 0\) (consider the case \(h = f\)).

5In jargon, \(\frac{1}{N} \sum_{i=0}^{N-1} U_T^i\) converges to 0 in the strong operator topology on the closure of \(N\). The key fact is if a sequence of operators is uniformly bounded in operator norm and converges on a dense set (to a given operator) in the strong operator topology then it converges on the whole space in the strong operator topology.

6Note that \(N\) is not necessarily closed. Indeed if \(f\) is the \(L^2\) limit of \(g_i - U_T g_i\) then we cannot necessarily find \(h\) so that \(f = h - U_T h\) because there is no reason for the \(g_i\) to have a convergent subsequence (or even to have that their norm is uniformly bounded).

7That is, in the ‘strong operator topology’
subset. Is this true if we remove the assumption that $Y$ is closed?

(3) Write the details of Step 1 of Theorem 5.

(4) Show that continuous functions in a $\sigma$-compact metric space do no necessarily have a countable dense subset. Show that Theorem 2 holds for $(X,d)$ $\sigma$-compact and the set of compactly supported continuous functions.

(5) Can the continuous functions in Theorem 2 be replaced by the characteristic functions of open sets?

(6) Show that the assumptions of the Krylov-Bogoliubov Theorem are proper. That is it is not necessarily true if we do not assume the function is continuous or that the space is compact.

(7) State a theorem for (not necessarily ergodic) measure preserving systems whose proof is the same as the Von Neumann Ergodic Theorem.

1.4. Some consequences of ergodicity.

**Proposition 2.** (Kac Lemma) Let $(X,T,\mathcal{B},\mu)$ be an ergodic system and $\mu(X) < \infty$. If $A$ be measurable with $\mu(A) > 0$ then $\int_A n_x d\mu(x) = \mu(X)$.

Recall $n_x = \min\{n > 0 : T^n x \in A\}$. It exists for a.e. $x \in A$ by the Poincare Recurrence Theorem.

This proof is very similar to the proof that induced maps of measure preserving maps are measure preserving.

**Proof.** Let $C_1 = T^{-1}A \cap A$ and $D_1 = T^{-1}A \setminus A$. Inductively let $C_{n+1} = T^{-1}D_n \cap A$ and $D_{n+1} = T^{-1}D_n \setminus A$. Observe that the $C_i$ and $D_i$ are all disjoint and almost every point in the space is in one of them (by ergodicity). Now $\mu(D_i) = \mu(C_{i+1}) + \mu(D_{i+1}) = \sum_{j > i} \mu(C_j)$. So $\mu(X) = \sum n\mu(C_n)$. Notice $C_n = \{x : n_x = n\}$. $\square$

**Theorem 8.** (Many people but we’ll say Atkinson) Let $(X,T,\mathcal{B},\mu)$ be an ergodic probability system. Let $f : X \to \mathbb{R}$ be in $L^1(\mu)$ and $\int f d\mu = 0$. Then $\lim \inf \frac{1}{N} \sum_{i=0}^{N} f(T^i x) = 0$ for $\mu$ a.e. $x$.

This proof is tricky and so we first state the idea of the proof: Look at the Birkhoff sums $p_j = \sum_{i=0}^{j} f(T^i x) \in \mathbb{R}$. By the Birkhoff Ergodic Theorem for large $N$ we expect the first $N$ of them to be in $(-N\epsilon, N\epsilon)$. So many of the $p_i$ within $k\epsilon$ of another. Now if $|p_i - p_j| < k\epsilon$ we have that $\sum_{i=0}^{j} f(T^i x) \in (-k\epsilon, k\epsilon)$. That is, $T^i x$ looks like it satisfies the theorem. This is true for most of the points in the orbit of the typical point so we have the theorem.
Proof. Step 1: If it is false then there exists $\epsilon > 0$ so that

$$B = \{ x : \inf_{N \geq 0} \sum_{i=0}^{N} f(T^i x) > \epsilon \}$$

has positive $\mu$ measure.

To see this first $B$ is measurable. Second we use downward continuity of measure.

Step 2: Let $p_k = \sum_{i=0}^{k-1} f(T^i x)$. So the $p_k$ are in $\mathbb{R}$ not $X$.

$$\sum_{i=0}^{N-1} \chi_{B^c}(T^i x) \geq \{| \{ k : \exists N > j > k \text{ with } |p_k - p_j| < \epsilon \} |.$$

This follows from the fact that $|p_i - p_j| = | \sum_{i=0}^{j-1} f(T^i x) - \sum_{i=0}^{j-k-1} f(T^i(T^k x))|$

and if $| \sum_{i=0}^{M} f(T^i(T^k x)) | < \epsilon$ for some $T^k x \notin B$.

Step 3: For a.e $x$ for each $\delta > 0$ there exists $M_{x, \delta} := M$ so that for all $N > M$ we have $| \sum_{i=0}^{P-1} f(T^i x) | \in (-N\delta, N\delta)$ for all $P \leq N$ (by the Birkhoff Ergodic Theorem).\(^8\)

Step 4: For every $\epsilon > 0$ and a.e. $x$ there exists $R$ so that

$$\sum_{i=0}^{N-1} \chi_{B^c}(T^i x) > N(1 - \epsilon)$$

for all $N > R$.

Observe by Step 2 if $k < j$ we have $|p_j - p_k| = | \sum_{i=0}^{j-k-1} f(T^i(T^k x))|$

are within $\epsilon$ of each other, this gives a point not in $B$. We have $2N\epsilon^{-1}\delta$ disjoint boxes of size $\epsilon$. All but at most $2N\epsilon^{-1}\delta$ of the points $T^i x$ are in a box with a point $T^j x$ where $j > i$.\(^9\) So if $\delta < \frac{1}{2}\epsilon^2$ then at least $(1 - \epsilon)N$ of the points, $T^i x$, are not in $B$.

Step 5: Completing via the Birkhoff ergodic theorem.

By the Birkhoff ergodic theorem we have $\mu(B^c) = \lim_{n \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} \chi_{B^c}(T^i x)$ for almost every $x$. We have shown that for almost every $x$ the right hand side is at least $1 - \epsilon$, establishing the theorem. \(\square\)

Remark 1. (1) Notice the argument relies on using the hits of orbit of the typical point to estimate the measure of a set. This

\(^8\)Indeed, assume that $x$ is in the full measure set for $f$ given by the Birkhoff ergodic theorem. Let $\delta > 0$ be given. For the Birkhoff ergodic theorem to hold there must exist $R$ so that $| \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x) | < \delta$ for all $N > R$. Now let $k = \max \{| \sum_{i=0}^{j} f(T^i x) : 0 \leq j \leq R \}$. Let $M = \max \{ \frac{k}{2}, R \}$.

\(^9\)Indeed, each box can contain at most one such point.
is a useful argument. More typically one has a set of definite measure and uses ergodicity to put the orbit of the typical point in the set.

(2) Notice how important it is that the range of $f$ is $\mathbb{R}$ and not say $\mathbb{C}$. The point is that in $\mathbb{C}$ there are on the order of $(N\epsilon)^2$ balls of radius 1 in $\{z : |z| \leq N\epsilon\}$. So if $\epsilon$ decays slowly with $N$ we can not get the $p_k$ close to the $p_j$.

(3) We want to draw a conclusion about Birkhoff sums of points in $X$. We observe that these Birkhoff sums are points in $\mathbb{R}$ so this lets us reduce our problem to facts about the geometry of $\mathbb{R}$.

(4) This proof provides recurrence of random walks in $\mathbb{R}$ under mild assumptions.

(5) Notice that just because $x$ is in the full measure set of points that BET holds for $f$ does not mean it is in the good points for Atkinson’s theorem. It means most points in its orbit are good for an effective version of Atkinson’s Theorem (those that land within fixed $\epsilon$ of 0 at least once).

**Corollary 1.** Let $(T, X, \mathcal{B}, \mu)$ be an ergodic system and $\mu(X) < \infty$. Let $f : X \to \mathbb{R}$ be in $L^1(\mu)$ and $\int f d\mu = 0$. Let $T_f : X \times \mathbb{R} \to X \times \mathbb{R}$ preserves $\mu \times \lambda_\mathbb{R}$ and is conservative with respect to this measure.

1.5. **Problem.**

(1) Let $(X, d)$ be a metric space, $\mu$ be a Borel measure and $T : X \to X$ be $\mu$-ergodic. Let $s_1, ...$ be an increasing sequence of positive real numbers and $F_y : X \to [0, +\infty]$ by $F_y(x) = \liminf_{n \to \infty} s_n d(T^n x, y)$. Show that for each $y$ we have that $F_y$ is constant almost everywhere.

1.6. **Space of invariant measures.** Let $(X, d)$ be a compact metric space and $T : X \to X$ be continuous. Let $\mathcal{M}_T = \{T$ invariant probability measures$\}$. Let $\mathcal{M}_{T_{\text{er}}} = \{T$ ergodic probability measures$\}$.

**Lemma 2.** $\mathcal{M}_T$ is a convex set in the space of measures. It is closed in the weak-*topology.

This is an exercise.

**Theorem 9.** The set of preserved probability measures of $T$ is the convex hull of the ergodic measures. The extreme points are exactly the ergodic measures.
Let $\mu$ be $T$-invariant and $f$ be measurable. \{ $x : f(x) > f(Tx)$ \} and \{ $x : f(x) < f(Tx)$ \} are measurable and if one has positive measure iff the other does.

Proof. Step 1: The $T$ invariance of Radon-Nikodym derivatives of invariant measures:

Let $\mu$ and $\nu$ be $T$-invariant measures and $\nu$ be absolutely continuous with respect to $\mu$. By the Radon-Nikodym Theorem there exists a function \( g = \frac{d\nu}{d\mu} \) so that for every $A \in \mathcal{B}$ we have $\nu(A) = \int_A g d\mu$. We claim $g$ is $T$ invariant. Let $c \in \mathbb{R}$ and $E = \{ x : g(x) < c \}$. Notice that $g(x) < c$ for all $x \in E \setminus T^{-1}E$ and $g(x) \geq c$ for all $x \in T^{-1}E \setminus E$. So either both sets have measure 0 or $\int_{T^{-1}E \setminus E} g d\mu > \int_{E \setminus T^{-1}E} g d\mu$. If the latter then $\nu(T^{-1}E \setminus E) > \nu(E \setminus T^{-1}E)$ contradicting the $T$-invariance of $\nu$.

Step 2: A non-trivial convex combination of (distinct) invariant probability measures is not ergodic.

Let $\mu = c\nu_1 + (1 - c)\nu_2$. $\frac{d\nu_1}{d\mu}$ and $\frac{d\nu_2}{d\mu}$ are nontrivial $T$ invariant functions.

Step 3: Extreme points are ergodic.

We show if an invariant probability measure is not ergodic then it is not extreme. Let $A$ be a $T$ invariant set with positive $\mu$ measure. Let $\eta = \frac{1}{\mu(A)}\mu|_A$ and $\zeta = \frac{1}{\mu(A^c)}\mu|_{A^c}$. Observe $\mu = \mu(A)\eta + \mu(A^c)\zeta$.

Going further:

**Theorem 10.** (Krein-Milman) A convex set in a vector space is the hull of its extreme points.

**Theorem 11.** (Ergodic decomposition) Let $(X, d)$ be a compact metric space and $T : X \to X$ be borel. Let $\mathcal{M}_T$ denote the $T$ preserved probability measures. Let $\mathcal{M}_e^T$ denote the ergodic ones. There exists a measurable map $\Pi$ from $\mathcal{M}_T$ to measures on $\mathcal{M}_e^T$ so that $\mu(A) = \int_{\mathcal{M}_e^T} \nu(A) d(\Pi \mu)(\nu)$.

Examples: $T : [0, 1) \to [0, 1)$ by $T(x) = x$. Lebesgue is preserved. The point masses are the ergodic measures. So $\Pi(\lambda) = \lambda$.

$T : [0, 1)^2 \to [0, 1)^2$ by $T(x, y) = (x, y + x - \lfloor y + x \rfloor)$. $\Pi(\lambda^2)$ is $\lambda$ on the 1st coordinate.

1.7. Problems.

(1) Prove Lemma 2.