Remark. Thomas O'Hare, Nick Bolle, and Hao-Tong Yan were collaborators.

Problem 1. Let

$$T : [0, 1] \to [0, 1], \quad Tx = 2x \pmod 1$$

Consider the interval

$$E_k^n = \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right),$$

which has dyadic rational endpoints. Let λ be Lebesgue measure.

(1) Compute $T^{-1}(E_k^n)$.
(2) Calculate the Lebesgue measures of E_k^n and $T^{-1}(E_k^n)$. Show they are the same.
(3) Using properties of measure, show that

$$\lambda(T^{-1}(I)) = \lambda(I)$$

for all open intervals I.
(4) Show the same holds for all open sets U.
(5) Conclude that the same holds for all Borel measurable E.

Proof.

(1) On $[0, 1]$, we can rewrite T as

$$Tx = \begin{cases} 2x \text{ if } 0 \leq x \leq 1/2, \\ 2x - 1 \text{ if } 1/2 < x \leq 1. \end{cases}$$

Using this alternative characterization, we see that

$$T^{-1}(E_k^n) = \left[\frac{k}{2^{n+1}}, \frac{k+1}{2^{n+1}}\right) \sqcup \left[\frac{k+2^n}{2^{n+1}}, \frac{k+1+2^n}{2^{n+1}}\right).$$

(2) Since this is a half-open interval, we see that

$$\lambda(E_k^n) = \frac{k+1}{2^n} - \frac{k}{2^n} = \frac{1}{2^n}.$$

Similarly, since we have the disjoint union of two half-open intervals, we see that

$$\lambda(T^{-1}(E_k^n)) = \left(\frac{k+2^n+1}{2^{n+1}} - \frac{k+2^n}{2^{n+1}}\right) + \left(\frac{k+1}{2^{n+1}} - \frac{k}{2^{n+1}}\right)$$
$$= \frac{2}{2^{n+1}} = \frac{1}{2^n}.$$

We see that for all k and n we have that

$$\lambda(T^{-1}(E_k^n)) = \lambda(E_k^n).$$

(3) Let $I = (a, b) \subseteq [0, 1]$ be an open interval. The dyadic rational numbers are dense in $[0, 1]$ (see here), so we can create an increasing sequence of dyadic rational half-open intervals,

say $\{E_j\}_{j=1}^{\infty}$ with $E_1 \subseteq E_2 \subseteq \cdots$, such that $\bigcup_{j=1}^{\infty} E_j = I$. We now use (2) and continuity from below to deduce that

$$\lambda(T^{-1}(I)) = \lambda\left(T^{-1}\left(\bigcup_{j=1}^{\infty} E_j\right)\right) = \lambda\left(\bigcup_{j=1}^{\infty} T^{-1}(E_j)\right) = \lim_{j\to\infty} \lambda(T^{-1}(E_j))$$
$$= \lim_{j\to\infty} \lambda(E_j) = \lambda\left(\bigcup_{j=1}^{\infty} E_j\right) = \lambda(I).$$

(4) In $\mathbb{R}$, an open subset U can be written as a countable union of disjoint open intervals, say $\{I_j\}_{j=1}^{\infty}$. Using this and (3), we have

$$\lambda(T^{-1}(U)) = \lambda\left(T^{-1}\left(\bigsqcup_{j=1}^{\infty} I_j\right)\right) = \lambda\left(\bigsqcup_{j=1}^{\infty} T^{-1}(I_j)\right) = \sum_{j=1}^{\infty} \lambda(T^{-1}(I_j)) = \sum_{j=1}^{\infty} \lambda(I_j)$$
$$= \lambda\left(\bigsqcup_{j=1}^{\infty} I_j\right) = \lambda(U).$$

Remark. Note that if it holds for all open sets U, it holds for all closed sets C. If C is closed, then $C^c = U$ is open, so $U^c = C$. Using the fact that this is a finite measure space, we can then calculate the following:

$$\lambda(T^{-1}(C)) = \lambda(T^{-1}(U^c)) = \lambda(T^{-1}(U)^c) = 1 - \lambda(T^{-1}(U)) = 1 - \lambda(U) = \lambda(U^c) = \lambda(C).$$

Remark. We use the fact that if N is a null-set, then

$$\lambda(F \cup N) = \lambda(F).$$

This follows by subadditivity and monotonicity;

$$\lambda(F) \leq \lambda(F \cup N) \leq \lambda(F) + \lambda(N) = \lambda(F) \implies \lambda(F \cup N) = \lambda(F).$$

(5) A Lebesgue measurable set can be written as the union

$$E = F \cup N,$$

where F is a F_σ set and N is a set of measure zero with respect to Lebesgue measure. We show it will hold for all Lebesgue measurable sets, and then in particular we get it holds for all Borel measurable sets.

Since F is an F_σ set, we have $F = \bigcup_{j=1}^{\infty} C_j$, where $\{C_j\}_{j=1}^{\infty}$ is a sequence of closed sets. Let $D_j = \bigcup_{i=1}^{j} C_j$. Note that D_j is a closed set, since it is a union of a finite number of closed sets. Then $D = \bigcup_{j=1}^{\infty} D_j = \bigcup_{j=1}^{\infty} C_j = F$ and we have

$$\lambda(T^{-1}(F)) = \lambda(T^{-1}(D)) = \lim_{j\to\infty} \lambda(T^{-1}(D_j)) = \lim_{j\to\infty} \lambda(D_j) = \lambda(D) = \lambda(F)$$

by an argument similar to (3).

The goal now is to show that $\lambda(T^{-1}(N)) = 0$. If we can do this, then we have that

$$\lambda(T^{-1}(E)) = \lambda(T^{-1}(F) \cup T^{-1}(N)) = \lambda(T^{-1}(F)) = \lambda(F) = \lambda(F \cup N) = \lambda(E).$$

We now note that T is a continuous map on $[0, 1]$. Since N has measure zero, for every $\epsilon > 0$ we can find an open set U with the property that $N \subseteq U$ and $\lambda(U) = \epsilon$. Thus, we have

$$\lambda(T^{-1}(N)) \leq \lambda(T^{-1}(U)) = \lambda(U) = \epsilon.$$

Since we can do this for every $\epsilon > 0$, this implies that $\lambda(T^{-1}(N)) = 0$. This concludes the proof.

Alternatively, invoke the next theorem or Caratheodory.

□

Problem 2. State **Theorem 1.1** from Peter Walter's book. Use it to prove the prior problem.

Proof. The theorem is as follows:

Theorem (Walters, Theorem 1.1)**.** Suppose $T : (X, \mathcal{M}, \mu) \to (Y, \mathcal{N}, \nu)$ is a measurable transformation of probability spaces. Let $\mathcal{C}$ be a semi-algebra that generates $\mathcal{N}$. If for each $A \in \mathcal{C}$ we have $T^{-1}(A) \in \mathcal{M}$ and $\mu(T^{-1}(A)) = \nu(A)$, then T is measure-preserving.

The conditions for the prior problem are $\mu = \nu = \lambda$ (Lebesgue measure) and the σ-algebras are the Borel σ-algebras. The collection of all intervals form a semialgebra which generates the Borel σ-algebra (see, for example, **Folland Proposition 1.2**), and as we've shown before the measures agree on all intervals (we technically only showed open intervals, but to get a half open interval or a closed interval involves adding points of measure zero, so it doesn't change anything). Invoking the theorem, we have that they agree on all Borel sets, telling us that our map is measurable. □

Problem 3. Prove that a proper subspace of $\mathbb{R}^n$ has zero Lebesgue measure.

Hint. You can use the fact that a proper subspace is the graph of a linear function from $T : \mathbb{R}^k \to \mathbb{R}^{n-k}$ for k the dimension of the subspace and use Fubini's theorem.

Remark. I originally said use **Theorem 2.44** from Folland, which is still technically true but the theorem says invertible matrix, so you have to modify the proof. This ended up being the same thing as Nick Bolle's proof, so credit to him for writing this up.

Proof. Let $V = \mathbb{R}^k \subseteq \mathbb{R}^n$ be a proper subspace. Consider a map $T : \mathbb{R}^k \to \mathbb{R}^{n-k}$ where the graph of T is V. Consider the set

$$\Gamma(T) = \left\{(x, y) \in \mathbb{R}^k \times \mathbb{R}^{n-k} = \mathbb{R}^n : Tx = y\right\}.$$

Now use σ-finitness and Fubini;

$$\lambda(V) = \int_{\mathbb{R}^n} \chi_{\Gamma(T)} d\lambda = \int_{\mathbb{R}^k} \left(\chi_{\Gamma(T)}(x, y) d\lambda(y),\right) d\lambda(x) = \int_{\mathbb{R}^k} \lambda(\{Tx\}) d\lambda(x) = 0.$$

Here, λ is understood to be the Lebesgue measure with respect to whatever $\mathbb{R}^n$ we're integrating over. □

Problem 4. Let G, H be locally compact Hausdorff groups which are also second countable. Let $T : G \to H$ be a continuous surjective endomorphism. Let m be a left Haar measure on G. Let m_G, m_H be Haar measures on G, H respectively. Define a measure μ on H by

$$\mu(E) = m_G(T^{-1}(E)) \text{ for all Borel } E \subseteq H.$$

Prove the following:

(1) μ is a left Haar measure on H;
(2) there is a $c > 0$ such that
$$\mu = c \cdot m_H;$$
(3) if we suppose that $G = H$ is compact, i.e. $m_G(G) < \infty$, and $m_G = m_H$, then prove that $c = 1$.

Note that we can conclude from (3) the following result.

Theorem. If T is an endomorphism of a compact group G, then T preserves the Haar measure on G.

Remark. If we have these properties, note that G and H are also σ-compact. Since it's second countable, we have a countable basis $\{U_\alpha\}$. Since we are in LCH space, for each $x \in G$ (or H) we can find a neighborhood K_x compact. Taking the interior, we have a precompact neighborhood U_x for every $x \in G$. Since $\{U_\alpha\}$ is a basis, we can find $x \in U_{\alpha_x} \subseteq U_x$. Closure of U_{α_x} is contained in K_x, and a closed subset of a compact set is compact, so U_{α_x} is precompact. We then get $\{U_{\alpha_x}\}_{x\in G}$ covers G (or H). Since the base was countable, we get a refinement of this to $\{U_{\alpha_i}\}_{i=1}^\infty$. Taking closures, we get a countable cover of X by compact sets, implying σ-compact.

Proof. We recall that **left Haar measure** on a topological group G is a nonzero left-invariant Radon measure μ on G. In other words, μ satisfies the following.

(a) For all $x \in G$ and E a Borel set, we have

$$\mu(xE) = \mu(E).$$

(b) For all compact subsets E, we have

$$\mu(E) < \infty.$$

(c) For all Borel subsets E, we have

$$\mu(E) = \inf\{\mu(U) : E \subseteq U, U \text{ open}\}.$$

(d) For all open subsets U, we have

$$\mu(U) = \sup\{\mu(K) : K \subseteq U, K \text{ compact}\}.$$

We now proceed to the problem.

(1) We have that μ is the **pushforward measure**, $\mu = T_*(m_G)$. We show that this is indeed a measure on the Borel σ-algebra.

(a) We see $\mu(\varnothing) = m_G(T^{-1}(\varnothing)) = m_G(\varnothing) = 0$.

(b) For $\{E_j\}_{j=1}^\infty$ a countable disjoint collection of Borel sets, we have

$$\mu\left(\bigsqcup_{j=1}^\infty E_j\right) = m_G\left(T^{-1}\left(\bigsqcup_{j=1}^\infty E_j\right)\right) = m_G\left(\bigsqcup_{j=1}^\infty T^{-1}(E_j)\right) = \sum_{j=1}^\infty m_G(T^{-1}(E_j)) = \sum_{j=1}^\infty \mu(E_j).$$

So this is a measure. We next check that it is a left Haar measure.

(a) For $x \in G$ and E a Borel set, we see that we have

$$\mu(xE) = m_G(T^{-1}(xE)).$$

Since T a surjective endomorphism, there is some $h \in G$ with $T(h) = x$. We now write

$$\begin{aligned} T^{-1}(xE) &= \{g \in G : T(g) = xe \text{ for some } e \in E\} \\ &= \{g \in G : T(g) = T(h)e \text{ for some } e \in E\} \\ &= \{g \in G : T(h^{-1}g) \in E\} \\ &= h\{g \in G : T(g) \in E\} = hT^{-1}(E). \end{aligned}$$

Using the fact that m is a left Haar measure, we can write the above as

$$\mu(xE) = m_G(T^{-1}(xE)) = m_G(hT^{-1}(E)) = m_G(T^{-1}(E)) = \mu(E).$$

Hence we have left invariance of μ.

(b) For this to be true, we need to assume the map is proper. If T is proper, it is clear that we have this property, since for E compact we have

$$\mu(E) = m_G(T^{-1}(E)) < \infty.$$

To get that T is proper from the results, we note that T is a surjective, continuous endomorphism between LCH groups which are σ-compact. We can then use the Open Mapping theorem for topological groups – see here (**Theorem 2.6**) or here. Thus T is an open mapping. Now take $E \subseteq H$ compact. The goal is to show that $T^{-1}(E)$ is compact. Let $\{U_\alpha\}$ be an open cover of $T^{-1}(E)$; that is, suppose we have

$$T^{-1}(E) \subseteq \bigcup_\alpha U_\alpha.$$

Since we have G is a σ-compact LCH space, we can assume that the U_α are precompact; that is, $\overline{U_\alpha}$ is compact (use **Folland Proposition 4.39**, not obvious but follows). We can apply T to get

$$E \subseteq T\left(\bigcup_\alpha U_\alpha\right) = \bigcup_\alpha T(U_\alpha).$$

Since T is an open map, $T(U_\alpha)$ is open. We now use the fact that E is compact to get a finite refinement. We have

$$E \subseteq \bigcup_{i=1}^n T(U_i).$$

We now take preimages to get

$$T^{-1}(E) \subseteq \bigcup_{i=1}^n T^{-1}(T(U_i)) = \bigcup_{i=1}^n \ker(T)U_i \subseteq \bigcup_{i=1}^n \ker(T)\overline{U_i}.$$

Note that the product of compact sets is compact (by Tychonoff and continuity of the product), so $\bigcup_{i=1}^n \ker(T)\overline{U_i}$ is compact. This gives us that $T^{-1}(E)$ is a closed subset of a compact set, hence it is compact.

(c) Regularity follows by **Folland Theorem 7.8**, since G and H are second countable LCH spaces and we've shown that μ is a Borel measure which is finite on compact sets.

(2) This follows by **Folland Theorem 11.9**. Since (1) shows that μ is a Haar measure, we can use that in conjunction with the theorem to find such a constant.

(3) Since G compact, we have $m_G(G) < \infty$, so

$$\mu(G) = m_G(T^{-1}(G)) = m_G(G) < \infty,$$

and we have

$$m_G(G) = c\mu(G).$$

Solving for c gives $c = 1$.

$\square$

Remark. Everybody was a collaborator.

Remark. I was focused on attempting as many problems as possible as opposed to writing completely rigorous arguments. There are plenty of typos and incorrect solutions. Let me know if you find any and I will update it.

Unless otherwise specified, T is invertible and measure preserving.

Problem 5 (Petersen 1.4.1)**.** Show that T if T is a measure preserving transformation, then $U = U_T$ defined on $L^2(X, \mathcal{M}, \mu)$ by

$$U_T(f) := f(Tx)$$

is unitary. What if T is noninvertible?

Proof. Recall an operator $U_T : L^2(\mu) \to L^2(\mu)$ is unitary if it satisfies two conditions:

(1) U_T is surjective.
(2) U_T preserves the inner product.

We check these conditions now.
(1): For surjectivity, we need to show that for all $g \in L^2(\mu)$, there is an $f \in L^2(\mu)$ with $U_T(f) = g$ (at least in L^2 equivalence, so almost everywhere). Since T is invertible and measure preserving, we have $g \circ T^{-1} : X \to \mathbb{R}$ is an L^2 function, and $U_T(g \circ T^{-1}) = g \circ T^{-1} \circ T = g$ (at least almost everywhere). This gives us surjectivity.
(2) : For preserving the inner product, we take $f, g \in L^2(\mu)$ and notice

$$\langle U_T(f), U_T(g)\rangle = \int_X f(T(x))\overline{g(T(x))}d\mu(x).$$

We have $T^{-1}(X) = X$, so preforming a change of variables $y = T^{-1}(x)$ and using measure preserving we have

$$\langle U_T(f), U_T(g)\rangle = \int_X f(y)\overline{g(y)}d\mu(y) = \langle f, g\rangle.$$

Notice that in the proof of (2) we only used measure preserving. So we always get $U_T : L^2(\mu) \to L^2(\mu)$ is an isometry if T is measure preserving. If T is not invertible, the questions is whether U_T is surjective, and it's not true (see the discussion on Carmen). □

Problem 6 (Petersen 1.4.2)**.** Show the following:

(1) The one sided Bernoulli shift $\sigma((x_n)) = y_n$ with $y_n = x_{n+1}$ on $\prod_0^\infty \{0, 1\}$, where $p_0 = 1/2$ and $p_1 = 1/2$, is isomorphic to the doubling map on the circle. That is, it's isomorphic to

$$T : [0, 1) \to [0, 1), \qquad T(x) = 2x \pmod 1.$$

(2) The two sided Bernoulli shift $\sigma((x_n)) = y_n$ with $y_n = x_{n+1}$ on $\{0, 1\}^{\mathbb{Z}}$, where $p_0 = 1/2$ and $p_1 = 1/2$, is isomorphic to Baker's map. That is, it's isomorphic to

$$T(x, y) = \begin{cases} (2x \pmod 1, y/2) \text{ if } 0 \le x \le 1/2 \\ (2x \pmod 1, (y + 1)/2) \text{ if } 1/2 \le x < 1 \end{cases},$$

$T : [0, 1) \times [0, 1) \to [0, 1) \times [0, 1)$.

Proof. We recall what an isomorphism of systems means. Two systems are isomorphic if there exists a σ-algebra isomorphism $\gamma : (X, \mathcal{M}, p) \to ([0, 1), \mathcal{B}, \lambda)$ for which the following diagram commutes:

$$\begin{array}{ccc} (X,\mathcal{M},\mu) & \xrightarrow{T} & (X,\mathcal{M},\mu) \\ \gamma\Big\downarrow & & \Big\downarrow\gamma \\ (Y,\mathcal{N},\nu) & \xrightarrow{S} & (Y,\mathcal{N},\nu) \end{array}$$

(1) Let $X = \prod_{i=0}^{\infty}\{0,1\}$, $\mathcal{M}$ is the σ-algebra generated by cylinders, and $\mu = p$ where p is induced on X by the property $p(\{0\}) = 1/2$ and $p(\{1\}) = 1/2$. Let $Y = [0,1)$, $\mathcal{N} = \mathcal{B}([0,1))$, and λ Lebesgue measure. We have a natural map $\gamma : X \to Y$ given by $\gamma((x_n)) = 0.x_0x_1\ldots$; i.e. a point is mapped to its binary sequence. A maybe more rigorous way to express γ is

$$\gamma : X \to [0,1), \qquad \gamma((x_n)) = \sum_{i=0}^{\infty} \frac{x_i}{2^{i+1}}.$$

Every element in $[0,1)$ has a binary expansion, so it is surjective. Moreover,

$$\gamma(\sigma((x_n))) = \gamma(y_n) = \sum_{i=0}^{\infty} \frac{y_i}{2^{i+1}} = \sum_{i=1}^{\infty} \frac{x_i}{2^i},$$

$$T(\gamma((x_n)) = T\left(\sum_{i=0}^{\infty} \frac{x_i}{2^{i+1}}\right) = \sum_{i=0}^{\infty} \frac{x_i}{2^i} \pmod 1 = \sum_{i=1}^{\infty} \frac{x_i}{2^i}.$$

So $\gamma \circ \sigma = T \circ \gamma$. We see γ is not necessarily injective, since a sequence ending in repeated 0s or repeated 1s gives us issues (akin to base ten and ending in repeated 9s or 0s). However, we note that the collection of points where it is not injective has zero measure (since it is countable), so it is injective off of a set of measure zero. Notice that the set of measure zero is T-invariant in $[0,1)$ and p invariant in X, so removing them doesn't change the dynamics.

We next claim that γ is measure preserving. Take the dyadic rational interval

$$I = \left[0, \frac{1}{2^n}\right).$$

The preimage $\gamma^{-1}(I)$ where

$$C_{0,n} \cap \cdots \cap C_{0,0} = \{(x_n) \in X : x_0 = 0, x_1 = 0, \ldots, x_n = 0\}$$

is an intersection of cylinders. We see

$$p(\gamma^{-1}(I)) = p(C_{0,0} \cap \cdots \cap C_{0,n}) = \frac{1}{2^n} = \lambda([0, 1/2^n)).$$

Now the same kind of argument applies to general dyadic rational intervals. Examine

$$I = \left[\frac{p}{2^n}, \frac{p+1}{2^n}\right).$$

Let $\alpha = p \pmod 2$, then

$$\gamma^{-1}(I) = C_{\alpha,n} \cap \cdots \cap C_{0,0}.$$

The same argument now applies. So γ is measure preserving. Thus γ is an isomorphism of dynamical systems.

(2) We claim the same kind of argument as above applies, except now we imagine the map $\gamma : X \to [0,1) \times [0,1)$ is going to send a sequence (x_n) to two binary representations. Let $\pi_1 : [0,1) \times [0,1) \to [0,1)$ be the projection onto the first coordinate and π_2 the projection onto the second. Then

$$\pi_1(\gamma((x_n))) = \sum_{0}^{\infty} \frac{x_{-n}}{2^{n+1}},$$

$$\pi_2(\gamma((x_n))) = \sum_1^\infty \frac{x_n}{2^n}.$$

It's surjective onto each factor map, so surjective onto $[0,1)^2$. The issue now are sequences which are constant in either direction, but those again have measure zero and are preserved by both maps so we can neglect these. It is injective once we throw these out. We just need to check measure preserving, but again this is just a matter of checking on the dyadic intervals. To check that it is a conjugacy, we note

$$T(\gamma((x_n))) = T\left(\sum_0^\infty \frac{x_{-n}}{2^{n+1}}, \sum_1^\infty \frac{x_n}{2^n}\right) = \left(\sum_1^\infty \frac{x_{-n}}{2^n}, \sum_1^\infty \frac{x_{n+1}}{2^n}\right),$$

$$\gamma(\sigma((x_n))) = \gamma(y_n) = \left(\sum_0^\infty \frac{y_{-n}}{2^{n+1}}, \sum_1^\infty \frac{y_n}{2^n}\right) = \left(\sum_1^\infty \frac{x_{-n}}{2^n}, \sum_1^\infty \frac{x_{n+1}}{2^n}\right).$$

So we get that it is an isomorphism of systems.

□

Problem 7 (Petersen 1.4.3 Modified)**.** Let G, H be locally compact Hausdorff groups which are also second countable. Let $T : G \to H$ be a continuous surjective endomorphism. Let m be a left Haar measure on G. Let m_G, m_H be Haar measures on G, H respectively. Define a measure μ on H by

$$\mu(E) = m_G(T^{-1}(E)) \text{ for all Borel } E \subseteq H.$$

Prove the following:

(1) μ is a left Haar measure on H;
(2) there is a $c > 0$ such that
$$\mu = c \cdot m_H;$$
(3) if we suppose that $G = H$ is compact, i.e. $m_G(G) < \infty$, and $m_G = m_H$, then prove that $c = 1$.

Note that we can conclude from (3) the following result.

Theorem. If T is an endomorphism of a compact group G, then T preserves the Haar measure on G.

Remark. If we have these properties, note that G and H are also σ-compact. Since it's second countable, we have a countable basis $\{U_\alpha\}$. Since we are in LCH space, for each $x \in G$ (or H) we can find a neighborhood K_x compact. Taking the interior, we have a precompact neighborhood U_x for every $x \in G$. Since $\{U_\alpha\}$ is a basis, we can find $x \in U_{\alpha_x} \subseteq U_x$. Closure of U_{α_x} is contained in K_x, and a closed subset of a compact set is compact, so U_{α_x} is precompact. We then get $\{U_{\alpha_x}\}_{x \in G}$ covers G (or H). Since the base was countable, we get a refinement of this to $\{U_{\alpha_i}\}_{i=1}^\infty$. Taking closures, we get a countable cover of X by compact sets, implying σ-compact.

Proof. We recall that **left Haar measure** on a topological group G is a nonzero left-invariant Radon measure μ on G. In other words, μ satisfies the following.

(a) For all $x \in G$ and E a Borel set, we have
$$\mu(xE) = \mu(E).$$
(b) For all compact subsets E, we have
$$\mu(E) < \infty.$$
(c) For all Borel subsets E, we have
$$\mu(E) = \inf\{\mu(U) : E \subseteq U, U \text{ open}\}.$$

(d) For all open subsets U, we have

$$\mu(U) = \sup\{\mu(K) : K \subseteq U, K \text{ compact}\}.$$

We now proceed to the problem.

(1) We have that μ is the **pushforward measure**, $\mu = T_*(m_G)$. We show that this is indeed a measure on the Borel σ-algebra.

(a) We see $\mu(\varnothing) = m_G(T^{-1}(\varnothing)) = m_G(\varnothing) = 0$.

(b) For $\{E_j\}_{j=1}^{\infty}$ a countable disjoint collection of Borel sets, we have

$$\mu\left(\bigsqcup_{j=1}^{\infty} E_j\right) = m_G\left(T^{-1}\left(\bigsqcup_{j=1}^{\infty} E_j\right)\right) = m_G\left(\bigsqcup_{j=1}^{\infty} T^{-1}(E_j)\right) = \sum_{j=1}^{\infty} m_G(T^{-1}(E_j)) = \sum_{j=1}^{\infty} \mu(E_j).$$

So this is a measure. We next check that it is a left Haar measure.

(a) For $x \in G$ and E a Borel set, we see that we have

$$\mu(xE) = m_G(T^{-1}(xE)).$$

Since T a surjective endomorphism, there is some $h \in G$ with $T(h) = x$. We now write

$$\begin{aligned} T^{-1}(xE) &= \{g \in G : T(g) = xe \text{ for some } e \in E\} \\ &= \{g \in G : T(g) = T(h)e \text{ for some } e \in E\} \\ &= \{g \in G : T(h^{-1}g) \in E\} \\ &= h\{g \in G : T(g) \in E\} = hT^{-1}(E). \end{aligned}$$

Using the fact that m is a left Haar measure, we can write the above as

$$\mu(xE) = m_G(T^{-1}(xE)) = m_G(hT^{-1}(E)) = m_G(T^{-1}(E)) = \mu(E).$$

Hence we have left invariance of μ.

(b) For this to be true, we need to assume the map is proper. If T is proper, it is clear that we have this property, since for E compact we have

$$\mu(E) = m_G(T^{-1}(E)) < \infty.$$

To get that T is proper from the results, we note that T is a surjective, continuous endomorphism between LCH groups which are σ-compact. We can then use the Open Mapping theorem for topological groups – see here (**Theorem 2.6**) or here. Thus T is an open mapping. Now take $E \subseteq H$ compact. The goal is to show that $T^{-1}(E)$ is compact. Let $\{U_\alpha\}$ be an open cover of $T^{-1}(E)$; that is, suppose we have

$$T^{-1}(E) \subseteq \bigcup_{\alpha} U_\alpha.$$

Since we have G is a σ-compact LCH space, we can assume that the U_α are precompact; that is, $\overline{U_\alpha}$ is compact (use **Folland Proposition 4.39**, not obvious but follows). We can apply T to get

$$E \subseteq T\left(\bigcup_{\alpha} U_\alpha\right) = \bigcup_{\alpha} T(U_\alpha).$$

Since T is an open map, $T(U_\alpha)$ is open. We now use the fact that E is compact to get a finite refinement. We have

$$E \subseteq \bigcup_{i=1}^{n} T(U_i).$$

We now take preimages to get

$$T^{-1}(E) \subseteq \bigcup_{i=1}^{n} T^{-1}(T(U_i)) = \bigcup_{i=1}^{n} \ker(T)U_i \subseteq \bigcup_{i=1}^{n} \ker(T)\overline{U_i}.$$

Note that the product of compact sets is compact (by Tychonoff and continuity of the product), so $\bigcup_{i=1}^{n} \ker(T)\overline{U_i}$ is compact. This gives us that $T^{-1}(E)$ is a closed subset of a compact set, hence it is compact.

(c) Regularity follows by **Folland Theorem 7.8**, since G and H are second countable LCH spaces and we've shown that μ is a Borel measure which is finite on compact sets.

(2) This follows by **Folland Theorem 11.9**. Since (1) shows that μ is a Haar measure, we can use that in conjunction with the theorem to find such a constant.

(3) Since G compact, we have $m_G(G) < \infty$, so

$$\mu(G) = m_G(T^{-1}(G)) = m_G(G) < \infty,$$

and we have

$$m_G(G) = c\mu(G).$$

Solving for c gives $c = 1$.

□

Problem 8 (Petersen 1.4.4)**.** Show that a homomorphism of measure-preserving systems is onto, up to a set of measure 0.

Proof. Consider $(X, \mathcal{M}, \mu, T)$, $(Y, \mathcal{N}, \nu, S)$ measure-preserving systems, $\gamma : (X, \mathcal{M}, \mu) \to (Y, \mathcal{N}, \nu)$. We say γ is a **homomorphism** of these systems if:

(1) γ is measurable (meaning $\varphi^{-1}(\mathcal{N}) \subseteq \mathcal{M}$);
(2) γ is measure preserving (meaining $\mu(\varphi^{-1}(E)) = \nu(E)$ for all $E \in \mathcal{N}$);
(3) We have $\gamma \circ T = S \circ \gamma$ almost everywhere.

The goal is to show that γ is essentially surjective, meaning surjective off of a set of measure zero. Consider

$$Z = \{y \in Y : \text{ there is no } x \in X \text{ with } \varphi(x) = y\}.$$

We see

$$\varphi^{-1}(Z) = \varnothing.$$

Since it is measure preserving, we see

$$\mu(\varphi^{-1}(Z)) = \mu(\varnothing) = 0 = \nu(Z).$$

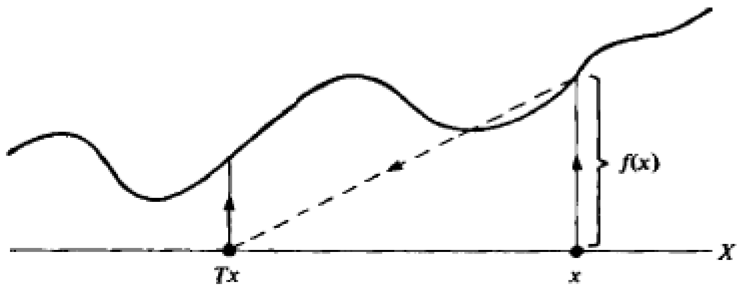
Off of Z, we see that φ is surjective. □

Let's recall some of the terminology.

Let $(X, \mathcal{M}, \mu, T)$ be a measure-preserving system. For $f : X \to (0, \infty)$, consider

$$\Gamma_f = \{(x, t) : 0 \le t < f(x)\}.$$

This is the collection of points "under f." We identify $(x, f(x))$ and $(Tx, 0)$. We have the following picture:



For $n \in \mathbb{Z}$, define

$$S_n(x) = \begin{cases} \sum_{k=0}^{n-1} f(T^k(x)) \text{ if } n > 0 \\ 0 \text{ if } n = 0 \\ -\sum_{k=1}^{-n} f(T^{-k}(x)) \text{ if } n < 0. \end{cases}$$

Problem 9. Use Poincare recurrence to show that $S_n(x) \to \infty$ as $n \to \infty$. The same kind of argument can be used to show $S_n(x) \to -\infty$ as $n \to -\infty$.

Proof. Notice that $f^{-1}((0,\infty)) = X$. By continuity of measures, there must be some $a > 0$ so that $\mu(f^{-1}((a,\infty))) \neq 0$. If we let $E_a = f^{-1}((a,\infty))$, then by Poincare recurrence almost every $x \in E_a$ returns to E_a infinitely often, so $f(T^k(x)) > a$ infinitely often for almost every $x \in E_a$. Consequently, for almost every $x \in E_a$, we have $S_n(x) \to \infty$ as $n \to \infty$. Now this holds for each E_a (technically, even if the set E_a has measure zero it will still hold), and we can write

$$X = \bigcup_{a>0} E_a.$$

The union of sets of measure zero will be measure zero, so it holds for almost every $x \in X$. □

For $x \in X$, $0 \leq t < f(x)$, $s \in \mathbb{R}$, define

$$n(x,t,s) := \min \{k \in \mathbb{Z}_{\geq 0} : s + t < S_{k+1}(x)\}.$$

This is called the **hitting number**.

Problem 10. Show that $n(x,t,s)$ is well-defined, and satisfies the property that

$$S_{n(x,t,s)} \leq s + t < S_{n(x,t,s)+1}.$$

Proof. The fact that $n(x,t,s)$ is well-defined follows from the fact that $S_n(x) \to \infty$, so there must be some n so that $s + t < S_n(x)$, and the minimum will be unique. The fact that it's a minimum tells us that we have the above identity. □

For $t \geq 0$, define

$$T_s^f(x,t) = T^f(x,t,s) := (T^{n(x,t,s)}(x), s + t - S_{n(x,t,s)}(x)).$$

Problem 11. Show that S_n satisfies the cocycle relation; i.e.,

$$S_{n+m} = S_n + S_m \circ T^n.$$

Proof. We see that

$$S_m \circ T^n(x) = \sum_{j=0}^{m-1} f(T^{j+n}(x)) = \sum_{j=n}^{m+n-1} f(T^j(x)) = S_{m+n}(x) - S_n(x).$$

□

Problem 12. Show that $n_s = n(\cdot, s)$ satisfies the coycle relation; i.e.,

$$n(x,t,s+q) = n(x,t,s) + n(T^{n(x,t,s)}(x), s+t-S_{n(x,t,s)}(x), q).$$

Proof. Notice that

$$n(T^{n(x,t,s)}(x), s+t-S_{n(x,t,s)}(x), q) = \min\left\{k \in \mathbb{Z} : s+t+q-S_{n(x,t,s)}(x) < S_{k+1}(T^{n(x,t,s)}(x))\right\}.$$

By the cocyle relation for S_n, this is the same as

$$n(T^{n(x,t,s)}(x), s+t-S_{n(x,t,s)}(x), q) = \min\left\{k \in \mathbb{Z} : s+t+q < S_{n(x,t,q)+k+1}(x)\right\}.$$

After changing variables appropriately, we see

$$n(T^{n(x,t,s)}(x), s+t-S_{n(x,t,s)}(x), q) = \min\left\{\alpha \in \mathbb{Z} : s+t+q < S_{\alpha+1}(x)\right\} - n(x,t,q) = n(x,t,s+q) - n(x,t,q).$$

This gives us the cocycle property. □

Problem 13. Show that T_s^f is a flow.

Proof. There are two things we need to show.

(1) We see that

$$n(x,t,0) = \min\{k \in \mathbb{Z} : t \le S_{k+1}(x)\} = \min\left\{k \in \mathbb{Z} : t \le \sum_{j=0}^{k} f(T^j(x))\right\}.$$

Since $0 \le t < f(x)$, this implies that

$$t \le S_1(x) = f(x),$$

so $n(x,t,0) = 0$. Therefore

$$T_0^f(x,t) = (T^0(x), t - S_0(x)) = (x,t).$$

So T_0^f is the identity.

(2) We next need to check that the $\mathbb{R}$ action is satisfied, meaning

$$T_{s+q}^f(x,t) = T_s^f \circ T_q^f(x,t).$$

Notice

$$\begin{aligned} T_s^f\left(T_q^f(x,t)\right) &= T_s^f(T^{n(x,t,q)}(x), q+t-S_{n(x,t,q)}(x)) \\ &= \Big(T^{n(T^{n(x,t,q)}(x), s+t-S_{n(x,t,q)}(x), s)}(T^{n(x,t,q)}(x)), \\ & s+q+t-S_{n(x,t,q)}(x) - S_{n(T^{n(x,t,q)}(x), s+t-S_{n(x,t,q)}(x), s)}(T^{n(x,t,q)}(x))\Big). \end{aligned}$$

Use the cocycle property for n to get

$$n(T^{n(x,t,q)}(x), s+t-S_{n(x,t,q)}(x), s) = n(x,t,s+q) - n(x,t,q),$$

so

$$T^{n(T^{n(x,t,q)}(x), s+t-S_{n(x,t,q)}(x), s)}(T^{n(x,t,q)}(x)) = T^{n(x,t,s+q)}(x).$$

Now

$$S_{n(T^{n(x,t,q)}(x), s+t-S_{n(x,t,q)}(x), s)}(T^{n(x,t,q)}(x)) = S_{n(x,t,s+q)-n(x,t,q)}(T^{n(x,t,q)}(x)).$$

Plugging in the definition, we get

$$S_{n(x,t,s+q)-n(x,t,q)}(T^{n(x,t,q)}(x)) = \sum_{j=0}^{n(x,t,s+q)-n(x,t,q)-1} f(T^{j+n(x,t,q)}(x)) = \sum_{j=n(x,t,q)}^{n(x,t,s+q)-1} f(T^j(x))$$
$$= S_{n(x,t,s+q)}(x) - S_{n(x,t,q)}.$$

Substituting this in, we have

$$S_{n(x,t,q)}(x) + S_{n(T^{n(x,t,q)}(x),s+t-S_{n(x,t,q)}(x),s)}(T^{n(x,t,q)}(x)) = S_{n(x,t,s+q)}(x).$$

So

$$T_s^f\left(T_q^f(x,t)\right) = \left(T^{n(x,t,s+q)}(x), s+q+t-S_{n(x,t,s+q)}(x)\right) = T_{s+q}^f(x,t).$$

Thus this is actually a flow.

□

We call $\{T_s^f\}_{s\in\mathbb{R}}$ the **induced flow** (or flow built under a function). Note this is the flow going upward with unit speed.

Problem 14 (Petersen 1.4.7)**.** Verify that the flow built under a function is measure preserving.

Proof. The goal is to show that for fixed $s \in \mathbb{R}$ that T_s^f is a measure preserving system. The rest can be deduced from 1.4.C. Thus fix some $s \in \mathbb{R}$, and examine

$$T_s^f : \Gamma_f \to \Gamma_f,$$
$$T_s^f(x,t) = \left(T^{n(x,t,s)}(x), s+t-S_{n(x,t,s)}(x)\right).$$

Consider $E \subseteq \Gamma_f$ measurable (under the appropriate product measure). We need to show that

$$\mu_f\left((T_s^f)^{-1}(E)\right) = \mu_f(E),$$

where μ_f denotes the appropriate product measure. Let's examine

$$(T_s^f)^{-1}(E) = \{(x,t) \in \Gamma_f : (T^{n(x,t,s)}(x), s+t-S_{n(x,t,s)}(x)) \in E\}.$$

Consider

$$E_n = \{(x,t) \in \Gamma_f : (T^n(x), s+t-S_n(x)) \in E\}.$$

This is a disjoint collection for $n \geq 0$, and so we can decompose

$$(T_s^f)^{-1}(E) = \bigsqcup_{n\geq 0} E_n.$$

So it suffices to show that

$$\mu_f(E_n) = \mu_f(T_s^f(E_n)).$$

Now we can realize E_n as

$$E_n = \{(x,t) \in \Gamma_f : n(x,t,s) = n \text{ and } (T^n(x), s+t-S_n(x)) \in E\}.$$

Now we can write (abusing Fubini-Tonelli)

$$\mu_f(E_n) = \int_{x\in X}\left(\int_0^{f(x)} \chi_{E_n}(x,t)dt\right) d\mu(x).$$

Notice

$$\mu_f(T_s^f(E_n)) = \int_{x\in X}\left(\int_0^{f(x)} \chi_{T_s^f(E_n)}(x,t)dt\right) d\mu.$$

Since it is a measure preserving transformation with respect to the first variable, we see that this doesn't change. The question remains about whether it is a measure preserving transformation with respect to the second variable. However, for fixed $x \in X$ and $s \in \mathbb{R}$ we see that this is just a translation, which is going to be measure preserving. So for fixed x the integral on the inside remains the same, i.e.

$$\int_0^{f(x)} \chi_{T_s^f(E_n)}(x,t)dt = \int_0^{f(x)} \chi_{E_n}(T^{-n}(x),t)dt,$$

and then we can use the measure preserving property to get that after a change of variables this will be the same for all x, so

$$\mu_f(T_s^f(E_n)) = \mu_f(E_n).$$

Now

$$\mu_f((T_s^f)^{-1}(E)) = \sum_{n\geq 0} \mu_f(E_n) = \sum_{n\geq 0} \mu_f(T_s^f(E_n)) = \mu_f(E).$$

So it's indeed measure preserving. □

Problem 15 (Petersen 1.4.10)**.** Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be Lebesgue spaces. Let $T : L^2(\nu) \to L^2(\mu)$ be an isometry which is multiplicative, meaning

$$T(fg) = T(f) \cdot T(g)$$

whenever f, g and $fg \in L^2(\nu)$. Show that there is a homomorphism $\varphi : X \to Y$ such that $Tf(x) = f(\varphi x)$ almost everywhere.

Proof. The claim is that $T(\chi_E)$ is a characteristic function for every $E \subseteq Y$ with $\mu(E) < \infty$. This is where the multiplicative property comes in. Notice that

$$T(\chi_E)^2 = T(\chi_E)T(\chi_E) = T(\chi_E^2) = T(\chi_E).$$

So almost everywhere we have that $T(\chi_E)$ is either 0 or 1, which means it is a characteristic function for some set $F \subseteq X$. Furthermore, since we have an isometry we have

$$\|\chi_F\|_2 = \|\chi_E\|_2 \implies \mu(F) = \nu(E).$$

Define $\widehat{\varphi} : \mathcal{N} \to \mathcal{M}$ on the sets of finite measure by $\widehat{\varphi}(E) = F$. Then

$$T(\chi_E) = \chi_{\widehat{\varphi}(F)}.$$

This is defined on every measurable set with finite measure, and since we are in a Lebesgue space we can get it for all measurable sets by taking limits. So actually $\widehat{\varphi} : \mathcal{N} \to \mathcal{M}$ is uniquely defined on all measurable sets. One can check that it's a homomorphism. If we take Lebesgue spaces to be probability measure spaces, then $T(1) = 1$. If we check complements, we have

$$\chi_E + \chi_{E^c} = 1 \implies X \setminus \widehat{\varphi}(E) = \widehat{\varphi}(E^c).$$

For finite unions we use the modular equation

$$\chi_{E\cup F} = \chi_E + \chi_F - \chi_E\chi_F,$$

so applying T to both sides we have

$$\widehat{\varphi}(E \cup F) = \widehat{\varphi}(E) \cup \widehat{\varphi}(F).$$

Induction then gives the result. For infinite unions, let $B_n = \cup_{i=1}^n E_i$, $B_n \nearrow B$, then

$$\chi_{B_n} \to \chi_B$$

and we can apply dominated convergence to get $\chi_B \in L^2(\nu)$. T is an isometry, so $T(\chi_B) \to T(\chi_B)$, and that implies

$$\widehat{\varphi}\left(\bigcup_1^\infty E_i\right) = \bigcup_1^\infty \widehat{\varphi}(E_i).$$

This is then a homomorphism of σ-algebras. We can find φ which will induce $\widehat{\varphi}$. This will satisfy the problem. □

We recall some definitions first. A **σ-ring** on a set X is a subset of $\mathcal{P}(X)$ satisfying the following properties:

(1) $\varnothing \in R$.
(2) For all $A, B \in R$, we have $A \cup B \in R$.
(3) For all $A, B \in R$, we have $A \setminus B \in R$.

A function $\mu : R \to [0, \infty]$ is a **premeasure** if it satisfies the following properties:

(1) $\mu(\varnothing) = 0$.
(2) We have

$$\mu\left(\bigsqcup_1^\infty A_n\right) = \sum_1^\infty \mu(A_n).$$

We recall the Caratheodory-Hopf Extension theorem.

Theorem (Caratheodory-Hopf Extension)**.** Let X be a set, R a σ-ring on X, and $\mu : R \to [0, \infty]$ a pre-measure on R. Then there is a measure μ' on $\sigma(R)$ which extends μ. Moreover, if μ is σ-finite, then μ' is unique.

The following problems are credited to Fabrice Baudoin (see here).

Problem 16. Let $B_n \subseteq \mathbb{R}^n$ be a sequence of Borel sets that satisfy $B_{n+1} \subseteq B_n \times \mathbb{R}$. Let us assume that for every n a probability measure μ_n is given on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ and that these probability measures are compatible in the sense that

$$\mu_n(A_1 \times \cdots \times A_{n-1} \times \mathbb{R}) = \mu_{n-1}(A_1 \times \cdots \times A_{n-1}).$$

Suppose they also satisfy

$$\mu_n(B_n) > \epsilon$$

where $0 < \epsilon < 1$. Show that there exists a sequence of compact sets $K_n \subseteq \mathbb{R}^n$ such that we have the following.

(1) $K_n \subseteq B_n$.
(2) $K_{n+1} \subseteq K_n \times \mathbb{R}$.
(3) $\mu_n(K_n) \geq \epsilon/2$.

Proof. We can use the regularity of Lebesgue measure. For every n, there is a compact $K_n^* \subseteq B_n$ satisfying

$$\mu_n(B_n \setminus K_n^*) \leq \frac{\epsilon}{2^{n+1}}.$$

Define

$$K_n = (K_1^* \times \mathbb{R}^{n-1}) \cap \cdots \cap (K_{n-1}^* \times \mathbb{R}) \cap K_n^*.$$

We see that $K_n \subseteq B_n$ by construction. We also see that

$$\begin{aligned} K_{n+1} = (K_1^* \times \mathbb{R}^n) \cap \cdots \cap (K_n^* \times \mathbb{R}) \cap K_{n+1}^* &\subseteq \left((K_1^* \times \mathbb{R}^{n-1}) \cap \cdots \cap (K_{n-1}^* \times \mathbb{R}) \cap K_n^*\right) \times \mathbb{R} \\ &= K_n \times \mathbb{R}. \end{aligned}$$

Finally, we can use the fact that the measure is finite to get

$$\begin{aligned}\mu_n(K_n) = \mu_n(B_n) - \mu_n(B_n \setminus K_n) = \mu_n(B_n) - \mu_n\left(B_n \setminus \left((K_1^* \times \mathbb{R}^{n-1}) \cap \cdots \cap (K_{n-1}^* \times \mathbb{R}) \cap K_n^*\right)\right) \\ \geq \mu_n(B_n) - \mu_n(B_n \setminus (K_1^* \times \mathbb{R}^{n-1})) - \cdots - \mu_n(B_n \setminus (K_{n-1}^* \times \mathbb{R})) - \mu_n(B_n \setminus K_n^*) \\ \geq \mu_n(B_n) - \mu_1(B_1 \setminus K_1^*) - \mu_2(B_2 \setminus K_2^*) - \cdots - \mu_n(B_n \setminus K_n^*) \geq \epsilon/2.\end{aligned}$$

□

Problem 17 (Petersen 1.4.12)**.** Use the Caratheodory-Hopf Extension theorem to prove *Kolmogorov's Consistency Theorem*: Let A be an index set and for each $n = 1, 2, \ldots$ and each n-tuple $(\alpha_1, \ldots, \alpha_n)$ of elements of A, let $\mu_{(\alpha_1,\ldots,\alpha_n)}$ be a Borel probability measure on $\mathbb{R}^n$. Assume that

(1) If $\tau \in \mathrm{Sym}(n)$ and T_τ is the corresponding transformation of $\mathbb{R}^n$ (i.e. sends the basis elements to their shuffled basis elements), then

$$\mu_{\tau(\alpha_1,\ldots,\alpha_n)}(E) = \mu_{(\alpha_1,\ldots,\alpha_n)}(T_\tau^{-1}(E)) \text{ for all Borel } E \subseteq \mathbb{R}^n.$$

(2) If $\Pi_{n+k,n} : \mathbb{R}^{n+k} \to \mathbb{R}^n$ is the projection map defined by

$$\Pi_{n+k,n}(x_1, \ldots, x_{n+k}) = (x_1, \ldots, x_n),$$

then

$$\mu_{(\alpha_1,\ldots,\alpha_n)}(E) = \mu_{(\alpha_1,\ldots,\alpha_{n+k})}(\Pi_{n+k,n}^{-1}(E)) \text{ for all } n, k = 1, 2, \ldots, \text{ and all Borel } E \subseteq \mathbb{R}^n.$$

Then there is a probability space $(\Omega, \mathcal{F}, P)$ and a family $\{f_\alpha : \alpha \in A\}$ of measurable functions on Ω such that we always have

$$\mu_{(\alpha_1,\ldots,\alpha_n)}(E) = P\{\omega : (f_{\alpha_1}(\omega), \ldots, f_{\alpha_n}(\omega)) \in E\}.$$

Proof. Let's see what our second condition is saying. It is saying that

$$\mu_{(\alpha_1,\ldots,\alpha_{n+k})}(E \times \mathbb{R}^k) = \mu_{(\alpha_1,\ldots,\alpha_n)}(E).$$

This is what it means for the measure to be *consistent*.

Consider the measure space $(\mathrm{Fun}(\Omega, \mathbb{R}), \sigma(\Omega, \mathbb{R}))$ where $\sigma(\Omega, \mathbb{R})$ is the σ-algebra generated by the cylindrical sets

$$\{f \in \mathrm{Fun}(\Omega, \mathbb{R}) : f(t_1) \in I_1, \ldots, f(t_n) \in I_n\},$$

where $I_1, \ldots, I_n$ are intervals and $t_1, \ldots, t_n \in \Omega$.

Consider the cylinders

$$C_{(\alpha_1,\ldots,\alpha_n)}(E) = \{f \in \mathrm{Fun}(\Omega, \mathbb{R}) : (f(\alpha_1), \ldots, f(\alpha_n)) \in E\}.$$

Define

$$\mu(C_{(\alpha_1,\ldots,\alpha_n)}(E)) = \mu_{(\alpha_1,\ldots,\alpha_n)}(E).$$

The consistency assumptions show that μ is well-defined. We also get $\mu(\varnothing) = 0$. We need to establish σ-additivity to win. To do so, let (C_n) be a sequence of pairwise disjoint cylinder with their union C a cylinder as well. Let

$$F_N = \bigcup_0^N C_n,$$

$$D_N = C \setminus F_N.$$

We can write

$$\mu(C) = \mu(D_N) + \mu\left(F_N\right).$$

The goal is to show

$$\lim_{N\to\infty} \mu(D_N) = 0.$$

The fact that D_N is a cylinder implies it only uses a finite sequence of times $(t_n)_{n=1}^N$. We may assume that every D_N can be described as

$$D_N = \{f \in \text{Fun}(\Omega, \mathbb{R}) : (f(\alpha_1), \ldots, f(\alpha_N)) \in B_N\},$$

where B_N a Borel set. Notice that B_N is a sequence which satisfies

$$B_{N+1} \subseteq B_N \times \mathbb{R}.$$

Now suppose the limit converges to some $\epsilon > 0$ for contradiction. Then $\mu(D_N) \geq \epsilon$ for all N. We can use the above problem to get some compact sets (K_N). This is nonempty by assumption, pick $(x_1, \ldots, x_n^n) \in K_n$. Using a diagonal argument, we get a sequence (x_n) such that for every n, $(x_1, \ldots, x_n) \in K_n$. So the sequence

$$\{f \in \text{Fun}(\Omega, \mathbb{R}) : (f(\alpha_1), \ldots, f(\alpha_n)) = (x_1, \ldots, x_n)\} \in D_n.$$

This implies

$$\bigcap_n D_n \neq \varnothing,$$

but this is a contradiction. So we must have the limit is 0. Thus we can invoke Caratheodory to get a probability measure P so that $(\text{Fun}(\Omega, \mathbb{R}), \sigma(\Omega, \mathbb{R}), P)$ is a probability measure space. Choose family $\{\pi_\alpha : \alpha \in A\}$ with $\pi_\alpha(f) = f(\alpha)$. □

Problem 18 (Petersen 2.3.1)**.** State and prove versions of the Maximal Ergodic Theorem and Pointwise Ergodic Theorem for one-parameter measure preserving flows.

Let's recall the two theorems first.

Theorem (Maximal Ergodic Theorem)**.** If $f \in L^1(X, \mathcal{M}, \mu)$ and $(X, \mathcal{M}, \mu, T)$ is a measure preserving system, then

$$\int_{\{f^*>0\}} f d\mu \geq 0, \text{ where } f^*(x) = \sup_{n \geq 1} \frac{1}{n} \sum_0^{n-1} f(T^k(x)).$$

Theorem (Pointwise Ergodic Theorem)**.** Let $(X, \mathcal{M}, \mu)$ be a probability space, $T : X \to X$ an invertible measure preserving transformation, and $f \in L^1(X, \mathcal{M}, \mu)$, then

(1) we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_0^{n-1} f(T^k(x)) = \overline{f}(x) \text{ exists a.e.},$$

(2) we have $\overline{f}$ is T-invariant almost everywhere,

(3) we have that if $A \in \mathcal{M}$ with $T^{-1}(A) = A$, then

$$\int_A f d\mu = \int_A \overline{f} d\mu,$$

(4) we have

$$\frac{1}{n} \sum_0^{n-1} f \circ T^k \to \overline{f} \text{ in } L^1.$$

Proof. Let's consider a one-parameter measure preserving flow on $(X, \mathcal{M}, \mu)$. This is a family of maps $\{T(t, x) : t \in \mathbb{R}\}$ which satisfies for all $t \in \mathbb{R}$ $T(t, x) : X \to X$ is an invertible measure preserving map, $T(t + s, x) = T(t, T(s, x))$, $T(0, x) = x$, and $T(-1, x) = T^{-1}(x)$.

Examine the operator

$$A_r(f)(x) = \frac{1}{|B(r, 0)|} \int_{B(r,0)} f(T(t, x)) d\lambda(t)$$

for $r > 0$. This will be our averaging operator here. Now the goal is to examine

$$f^*(x) = \sup_{r>0} A_r(f)(x).$$

This will play the role of our maximal function. Ideally, we'd like to show that

$$\int_{\{f^*(x)>0\}} f(x)d\mu \geq 0.$$

Let $\epsilon > 0$ be fixed and examine

$$E_\epsilon = \{x : f^*(x) > \epsilon\}.$$

If $x \in E_\epsilon$, then this means there is some $r_x > 0$ so that

$$A_{r_x}(f)(x) > \epsilon.$$

Now since x is fixed, let $T_x : \mathbb{R} \to X$ be defined by $T_x(t) = T(t, x)$. We have

$$A_{r_x}(f)(x) = \frac{1}{|B(r_x, 0)|} \int_{B(r_x,0)} f(T_x(t))dt > \epsilon.$$

In other words, there is some $r_x > 0$ so that

$$\int_{B(r_x,0)} f(T_x(t))dt = \int_{T_x(B(r_x,0))} f(z)d\mu(z) > \epsilon.$$

The claim now is that the set $\{T_x(B(r_x, 0))\}_{x\in X}$ covers E_ϵ, but this is clear since $x \in T_x(B(r_x, 0))$. Now take $F_\epsilon = T_x^{-1}(E_\epsilon) \subseteq \mathbb{R}$. Since T_x is measure preserving and invertible, we have that the balls $B(r_x, 0)$ cover F_ϵ up to a null set (which we exclude anyways). We can then apply a Vitali lemma and use $\{T_x\}$ so that for all $c < \lambda(F_\epsilon) = \mu(E_\epsilon)$ there is a k so that $x_1, \ldots, x_k \in E_\alpha$ and the balls $B_j = B(r_x, x_j)$ are disjoint and $\sum_1^k \mu(T_{x_j}(B_j)) > 3^{-1}c$. The rest of the argument now applies like the Hardy-Littlewood maximal theorem to give us

$$0 \leq \int_{\{f^*(x)>\epsilon\}} f(x)d\mu(x).$$

This applies for all $\epsilon > 0$, so take a limit to get the result.

Being T-invariant is the same thing as saying $f^*(T(1, x)) = f^*(x)$. Notice

$$\begin{aligned} f^*(T(1,x)) = \sup_{r>0} \frac{1}{|B(r,0)|} \int_{B(r,0)} f(T(t, T(1,x)))d\lambda(t) &= \sup_{r>0} \frac{1}{|B(r,0)|} \int_{B(r,0)} f(T(t+1,x))d\lambda(t) \\ &= \sup_{r>0} \frac{1}{|B(r,0)|} \int_{B(r,1)} f(T(t,x))d\lambda(t) = f^*(x) \end{aligned}$$

by the supremum property. The Lebesgue differentiation theorem now kicks in to give us $\overline{f} = f$ in this case. We get all properties except the T-invariance of f (of which I'm not sure is going to hold in the flow case?) □

Problem 19 (Petersen 2.2.5)**.** Identify

$$\overline{f}(x) = \lim_{n\to\infty} A_n(f)(x) \text{ almost everywhere}$$

in each case.

(1) Consider $(X, \mathcal{M}, \mu, T)$ where $X = \{0, \ldots, n-1\}^{\mathbb{Z}}$, $\mathcal{M}$ the σ-algebra generated by cylinders, μ is given by $\mu(j) = p_j$ where $\sum_0^{n-1} p_j = 1$, T is the left shift map, and

$$f((x_n)) = \chi_{\{i\}}(x_0).$$

In other words, f is the function which tells you whether a sequence has i at the 0 index.

(2) Consider $(S^1, \mathcal{B}, \lambda, R_\alpha)$ where $R_\alpha : S^1 \to S^1$ is given by $R_\alpha(x) = x + \alpha \pmod 1$ and $f = \chi_I$ for some interval I.
(3) Consider $(\mathbb{R}, \mathcal{B}, \lambda, T)$ where $f \in L^1$ and $T(x) = x + 1$.

Proof.

(1) We calculate

$$A_n(f)((x_n)) = \frac{1}{n}\sum_0^{n-1} f(T^j(x_n)) = \frac{1}{n}\sum_0^{n-1} \chi_{\{i\}}(x_j).$$

This is the function which measures the frequency of $\{i\}$ in the first n entries of the sequence (x_n). So

$$\overline{f}(x_n) = \begin{cases} 1 \text{ if there are infinitely many occurrences of } i \text{ in } (x_n)_{n\geq 0} \\ 0 \text{ otherwise.} \end{cases}$$

(2) Since rotations by α are dense, regardless of where we start, we have that

$$\overline{f}(x) = \chi_I(x).$$

(3) We again examine

$$A_n(f)(x) = \frac{1}{n}\sum_0^{n-1} f(x+j).$$

Let $f = \chi_I$ for an interval. Then

$$A_n(f)(x) = \frac{1}{n}\sum_0^{n-1} \chi_I(x+j),$$

and we see

$$f^*(x) = \lim_{n\to\infty} A_n(f)(x) = 0.$$

We see that this holds for characteristic functions of compact sets. Let $P : L^1 \to L^1$ be the function $P(f) = f^*$ (we know this works by the pointwise Ergodic theorem). It is linear by the linearity of A_n and the limit. Note that $\|P\| \leq 1$, so it is continuous. We know on all characteristic functions it will be zero, so we can do the usual argument to get that for all L^1 functions it will be zero.

□

Problem 20 (Petersen 2.2.6)**.** Show that if

$$\sum_0^{n-1} f(T^k(x)) \to \infty \text{ almost everywhere,}$$

then

$$\int_x f d\mu > 0.$$

Proof. Notice that this is saying for almost every $x \in X$ there is an n so that $f_n(x) > 0$. Following the argument for the Maximal Ergodic Theorem, we get

$$\int_{\{f^*>0\}} f d\mu = \int_X f d\mu \geq 0.$$

If

$$\int_X f d\mu = 0,$$

then $f = 0$ almost everywhere, which contradicts the property of $\lim_{n\to\infty} f_n$ going to infinity almost everywhere. Thus we have

$$\int_X f d\mu > 0.$$

□

Problem 21 (Petersen 2.3.1)**.** Let $T : X \to X$ be an invertible, measurable, nonsingular transformation on a σ-finite measure space $(X, \mathcal{M}, \mu)$ in that T preserves the σ-ideal of null sets of μ. Recall T is **nonsingular** if $\mu(T(E)) = \mu(T^{-1}(E)) = 0$ for any measurable set E with $\mu(E) = 0$ (a weaker notion of measure preserving).

A set $W \in \mathcal{M}$ of positive measure is called **weakly wandering** if there is a sequence $n_k \to \infty$ such that the sets $T^{n_k}W$ are all pairwise disjoint. Show that if T has a weakly wandering set, then there does not exist a finite invariant measure equivalent to μ.

Recall that two measures m and μ are said to be **equivalent** if for all E measurable with $m(E) = 0$ we have $\mu(E) = 0$, and if E is measurable with $\mu(E) = 0$ then $m(E) = 0$. We write $m \sim \mu$.

Proof. Recall a measure m will be invariant under T if $m(T^{-1}(E)) = m(E)$ for all E measurable. Let $W \in \mathcal{M}$ be a set with $\mu(W) > 0$, m a finite T-invariant measure. Suppose $m \sim \mu$ for contradiction. Notice that if $\{T^{n_k}W\}$ are all pairwise disjoint, $m(T^{n_k}W) = m(W)$ for all k (by T-invariance), and we have

$$\sum_{k=0}^{\infty} m(T^{n_k}W) \le m(X)$$

by monotonicity and disjointness. This can only happen if $m(W) = 0$, but we assumed that $\mu(W) > 0$ and $m \sim \mu$, a contradiction. Since $\mu(W) > 0$, this forces m and μ to not be equivalent. □

We recall a few definition. We work over $(X, \mathcal{M}, \mu)$ a measure space. A collection $\mathcal{J} \subseteq \mathcal{M}$ is a **σ-ideal** if the following are satisfied:

1. $\varnothing \in \mathcal{J}$,
2. when $A \in \mathcal{J}$ and $B \in \mathcal{M}$ with $B \subseteq A$, then $B \in \mathcal{J}$,
3. if $\{A_n\} \subseteq \mathcal{J}$ then $\bigcup A_n \in \mathcal{J}$.

Problem 22. Show that if $\mathcal{J}$ is the collection of all sets of measure zero, then $\mathcal{J}$ is a σ-ideal.

Proof. We see (1) is satisfied, since $\mu(\varnothing) = 0$. Monotonicity gives us (2), and σ-additivity gives us (3). □

A set $W \in \mathcal{M}$ is **wandering** if the collection $\{T^{-n}W\}$ is pairwise disjoint. A map T is **conservative** if every wandering set is in $\mathcal{J}$, the σ-ideal of sets of measure zero.

Problem 23 (Petersen 2.3.2)**.** Let T be as in Petersen 2.3.1. Show that X has a decomposition into disjoint, measurable, invariant conservative and dissipative parts, $X = C \sqcup D$, in the following sense:

1. $T|C$ is conservative.
2. $D = \{T^nW : n \in \mathbb{Z}\}$ for some wandering set W.

Proof. Notice that a measurable subset of any wandering set is a wandering set. Let $B \subseteq W$, W wandering, and take $T^n(B)$, $T^m(B)$ for some $n, m \in \mathbb{Z}$. Then $T^n(B) \subseteq T^n(W)$, $T^m(B) \subseteq T^m(W)$, and

$$T^n(B) \cap T^m(B) \subseteq T^n(W) \cap T^m(W) = \varnothing.$$

The choices of n and m were arbitrary, so B is wandering.

Notice unions of wandering sets from disjoint T invariant sets X and Y will be wandering. Suppose A and B are wandering sets with $A \subset X$, $B \subseteq Y$, X and Y T-invariant and $X \cap Y = \varnothing$. Take $n, m \in \mathbb{Z}$. Then

$$T^n(A \sqcup B) \cap T^m(A \sqcup B) = (T^n(A) \sqcup T^n(B)) \cap (T^m(A) \sqcup T^m(B))$$
$$= (T^n(A) \cap T^m(A)) \cup (T^n(A) \cap T^m(B)) \cup (T^n(B) \cap T^m(B)) \cup (T^n(B) \cap T^m(B)) = \varnothing.$$

Consider now a sequence of increasing wandering sets with respect to $\subseteq$, $\{B_\alpha\}$. The claim is that $B = \bigcup_\alpha B_\alpha$ is a wandering set as well. Take $n, m \in \mathbb{Z}$, then

$$T^n\left(\bigcup_\alpha B_\alpha\right) = \bigcup_\alpha T^n(B_\alpha), \qquad T^m\left(\bigcup_\alpha B_\alpha\right) = \bigcup_\alpha T^m(B_\alpha).$$

We see that

$$T^n(B) \cap T^m(B) = \bigcup_{\alpha,\beta} (T^n(B_\alpha) \cap T^m(B_\beta)).$$

For any α, β, there is a γ so that $B_\alpha, B_\beta \subseteq B_\gamma$ with B_γ a wandering set, so we have

$$T^n(B_\alpha) \cap T^m(B_\beta) \subseteq T^n(B_\gamma) \cap T^m(B_\gamma) = \varnothing.$$

This holds for all n, m, α, β, so in particular B is a wandering set.

We can then consider the collection

$$\Gamma = \{W \in \mathcal{M} : W \text{ is a wandering set}\}.$$

We've just shown that, under the partial ordering $\subseteq$, chains have upper bounds. We invoke Zorn's Lemma to find a maximal $W \in \Gamma$. Consider the set $D = \{T^n W : n \in \mathbb{Z}\}$. This is T invariant, since

$$T^{-1}(D) = \{T^{n-1} W : n \in \mathbb{Z}\} = \{T^n W : n \in \mathbb{Z}\} = D.$$

Let $C = D^c$. This will also be invariant, since

$$T^{-1}(C) = T^{-1}(D^c) = T^{-1}(D)^c = D^c = C.$$

Take $B \subseteq C$ measurable and wandering. Suppose $\mu(B) > 0$, so that B is non-empty. Then $B \subseteq C$, $W \subseteq D$, C and D are T-invariant subsets which are disjoint. By the above, we get that $B \sqcup W$ is going to be a wandering set which contains W, contradicting maximality. We must have $\mu(B) = 0$, so that $B \in \mathcal{J}$. This tells us that $T|C$ is conservative. □

Problem 24 (Petersen 2.3.3)**.** Let T be as in Petersen 2.3.1. Show that T is conservative if and only if

$$P(u) = \sum_0^\infty u \circ T^k$$

takes only the two values 0 and ∞ almost everywhere for each non-negative $u \in L^\infty(X, \mathcal{M}, \mu)$.

Proof. Assume a probability measure space.
($\implies$): If T is conservative, then if E is wandering we have that $\mu(E) = 0$. So taking $E \in \mathcal{M}$, $\mu(E) > 0$, we have that almost every point $x \in E$ goes to E. So $P(u)(x) = \infty$ almost everywhere, where $U = \chi_E$. A linearity argument now applies to get that this holds for all $u \in L^\infty$.
($\impliedby$): Let $E \subseteq X$ be a measurable set which is wandering, and consider $\chi_E \in L^\infty$ positive. Then we see that $P(\chi_E)$ takes either the value 0 or ∞ almost everywhere. If $P(\chi_E)(x) = \infty$, this means that $T^k(x) \in E$ infinitely often. Since E is wandering, this is impossible (off of a set of measure zero, maybe) so $P(\chi_E)(x) = 0$ almost everywhere. Consequently $\chi_E^* = 0$ almost everywhere, and

$$0 = \int_X \chi_E^* d\mu = \int_X \chi_E d\mu = \mu(E).$$

□

Problem 25 (Petersen 2.3.4)**.** Verify that the induced transformations T_A and $\widetilde{T}$ really are measure preserving transformations when T is.

Proof. We break it up into parts.

(1) We show T_A is measure preserving. We define a bunch of sets which have convenient properties and hope things work out. Let $E \subseteq A$ be measurable.

First, notice that
$$T_A^{-1}(E) = \bigsqcup_{n \geq 1} (A_n \cap T^{-n}(E)),$$
so
$$\mu_A(T_A^{-1}(E)) = \frac{1}{\mu(A)} \sum_{n \geq 1} \mu(A_n \cap T^{-n}(E)).$$
Next, let
$$F_0 = A, \quad F_k = \{x \in X : T^k x \in A, T^j x \notin A \text{ for } 0 \leq j < k\} \text{ for } k \geq 1.$$
Notice that
$$\begin{aligned} T^{-1}(F_k) = \{x \in X : T^{k+1}x \in A, T^j \notin A \text{ for } 1 \leq j < k+1\} \\ = A_{k+1} \sqcup F_{k+1}. \end{aligned}$$
Now we see that
$$\mu(E) = \mu(E \cap A) = \mu(E \cap F_0).$$
Since T is measure preserving, we have
$$\mu(E \cap F_0) = \mu(T^{-1}(E \cap F_0)) = \mu(T^{-1}(E) \cap T^{-1}(F_0)) = \mu(T^{-1}(E) \cap A_1) + \mu(T^{-1}(E) \cap F_1).$$
We can continue this inductively; that is, we have
$$\mu(T^{-n}(E) \cap F_n) = \mu(T^{-n-1}(E) \cap T^{-1}(F_n)) = \mu(T^{-(n+1)}(E) \cap F_{n+1}) + \mu(T^{-(n+1)}(E) \cap E_{n+1}).$$
Letting this go to infinity gives
$$\mu(E) = \sum_{n \geq 1} \mu(T^{-n}(E) \cap A_n).$$
Thus
$$\begin{aligned} \mu_A(T_A^{-1}(E)) &= \frac{1}{\mu(A)} \sum_{n \geq 1} \mu(A_n \cap T^{-n}(E)) \\ &= \frac{1}{\mu(A)} \mu(E) = \mu_A(E). \end{aligned}$$
So T_A is measure preserving.

(2) We show $\widetilde{T}$ is measure preserving. If the set is in A', then the preimage will just be the set again, so the measures are the same. If the set is in A, then the preimage is divided evenly into the complement and A'. If the set is in $X \setminus A$, then the preimage lies evenly in $X \setminus A$ and A' (evenly here doesn't mean actually evenly, just means that the sum of the measures will be equal to the measure).

□

Problem 26 (Petersen 2.3.5)**.** Describe the action of T_A in these cases.

(1) $X = [0, 1)$, $R_\alpha = x + \alpha \pmod 1$, and $A = [0, 1/2)$.
(2) $X = \{0, 1\}^{\mathbb{Z}}$, σ is the shift, and $A = \{x : x_0 = 0\}$.

Proof.

(1) Will just end up being $R_\alpha = x + \alpha \pmod{1/2}$.

(2) Shift a bunch? Unsure.

□

Problem 27 (Petersen 2.3.6)**.** Prove directly that if T is a measure preserving transformation on a finite measure space and $\mu(A) > 0$ then $E = \{n \geq 1 : \mu(T^{-n}A \cap A) > 0\}$ has bounded gaps.

Proof. Proceed by contrapositive. We can enumerate $E = \{n_k\}$ in increasing order. No bounded gaps implies that for all $i \geq 1$ there is a k so that $n_{k+1} - n_k > i$. Moreover, for each i there is a j_i so that $\mu(T^{-n_{j_i}}A \cap A) = 0$. So we've found an infinite collection $\{T^{-n_{j_i}}(A)\}_{i \geq 1}$ of almost disjoint sets, and by the usual recurrence argument this forces $\mu(A) = 0$. This establishes if E does not have bounded gaps, then $\mu(A) = 0$. □

Problem 28 (Petersen 2.4.1)**.** Prove that if T is ergodic, then so are the induced transformations T_A and $\widetilde{T}$.

See here.

Proof. Let's first show T_A is ergodic if T is ergodic. Recall that if T is ergodic, then $T^{-1}(E) = E$ if and only if E is either null or conull. Let's try the usual trick of decomposing our space into nice sets. As usual, let

$$A_n = \{x \in A : n_A(x) = n\}.$$

Recall we can write

$$T_A^{-1}(E) = \bigsqcup_{n \geq 1} (A_n \cap T^{-n}(E)).$$

Suppose $F \subseteq A$ is such that $T_A^{-1}(F) = F$. Then we have

$$F = \bigsqcup_{n \geq 1} (A_n \cap T^{-n}(F)).$$

Let $\{F_k\}$ be as last time –

$$F_k = \{x \in X : T^k(x) \in A, T^j(x) \notin A \text{ for } 0 \leq j < k\}.$$

Let

$$E = \bigsqcup_{n \geq 1} (F_n \cap T^{-n}(F)).$$

Let

$$K = E \sqcup F.$$

Recall that we had

$$T^{-1}(A) = E_1 \sqcup F_1, \qquad T^{-1}(F_k) = A_{k+1} \sqcup F_{k+1}.$$

So

$$\begin{aligned} T^{-1}(K) &= T^{-1}(E) \sqcup T^{-1}(F) \\ &= \bigsqcup_{n \geq 1} [(F_{n+1} \sqcup E_{n+1}) \cap T^{-(n+1)}(F)] \cup [(E_1 \cup F_1) \cap T^{-1}(F)] \\ &= \bigcup_{n \geq 1} (F_n \cap T^{-n}(F)) \cup \bigcup_{n \geq 1} (E_n \cap T^{-n}(F)) = E \cup F = K. \end{aligned}$$

Thus K is T-invariant, so K is null or conull. If it is null, then $\mu_A(F) = 0$. If it is conull, then

$$K^c = (A \setminus F) \cup ((X \setminus A) \setminus E),$$

so

$$A \setminus F \subseteq K^c.$$

Then $\mu_A(F^c) = 0$ so that it is conull.

The other part is similar. □

Problem 29 (Petersen 2.4.2)**.** Prove that if T is ergodic (on a space of finite measure), $f \geq 0$, and

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{0}^{n-1} f(T^j(x)) < \infty \text{ almost everywhere}$$

then $f \in L^1$.

Proof. Notice that

$$f^*(x) := \limsup_{n\to\infty} \frac{1}{n}\sum_{0}^{n-1} f(T^j(x))$$

is a function so that $f^* \circ T = f^*$ almost everywhere. Since T is ergodic, the only T-invariant functions are the constant functions (almost everywhere), so f^* is constant. That is, $f^* = C < \infty$ almost everywhere for some $C \in \mathbb{R}_{>0}$.

Now let

$$f_k = f\chi_{\{f\leq k\}} + k\chi_{\{f>k\}}.$$

Then this is a bounded function, so $f_k \in L^1(\mu)$, and moreover we see that

$$\frac{1}{n}\sum_{0}^{n-1} f_k(T^j(x)) \leq \frac{1}{n}\sum_{0}^{n-1} f(T^j(x)).$$

Thus we have $f_k^* \leq C$, and the Birkhoff Ergodic Theorem tells us that

$$\int f_k d\mu = \int f_k^* d\mu \leq C\mu(X) \text{ for all } k.$$

Now we can apply the monotone convergence theorem (since $f_k \nearrow f$) to get

$$\lim_{k\to\infty} \int f_k d\mu = \int f d\mu \leq C\mu(X) < \infty.$$

Thus $f \in L^1(\mu)$. □

Problem 30 (Petersen 2.4.3)**.** Which of the equivalent characterizations of ergodicity fail when X has infinite measure?

Proof. Let's recall the equivalent characterizations. Throughout, $(X, \mathcal{M}, \mu, T)$ is a measure preserving system of a probability measure space. The following are equivalent (all results from Petersen):

(1) T is ergodic.
(2) For all measurable f we have $f \circ T = f$ implies f is constant. [Proposition 2.4.1]
(3) 1 is a simple eigenvalue of the transformation U induced on $L^2(X, \mathcal{M}, \mu)$ by T. [Theorem 2.4.2]
(4) For every $f \in L^1(X, \mathcal{M}, \mu)$, we have

$$\overline{f}(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{0}^{n-1} f(T^j(x)) = \int_X f d\mu \text{ almost everywhere.}$$

[Theorem 2.4.4]

(5) For every $f, g \in L^2(X, \mathcal{M}, \mu)$ we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{0}^{n-1} (U^j f, g) = (f, 1)\overline{(g, 1)}.$$

[Proposition 2.4.5]

Let's recall what ergodic means when X has infinite measure. A measure preserving system $(X, \mathcal{M}, \mu, T)$ is ergodic if $T^{-1}(E) = E$ implies $\mu(E) = 0$ or $\mu(X \setminus E) = 0$.

The goal now is to examine $(X, \mathcal{M}, \mu, T)$ a measure preserving system which is σ-finite and figure out which of the following are still true.

(1) $\implies$ (2): Assume T is ergodic. Let f be a measurable function which is T-invariant. Consider the set

$$E_r^f = E_r := \{x \in X : f(x) > r\}.$$

We claim this set is invariant under T. Notice that since $f \circ T = f$, we have

$$T^{-1}(E_r) = \{x \in X : f(T(x)) > r\} = \{x \in X : f(x) > r\} = E_r.$$

Since T is ergodic, this forces $\mu(E_r) = 0$ or $\mu(X \setminus E_r) = 0$. If f is not constant, there is an r so that $0 < \mu(E_r) < \infty$. Notice this means that $\mu(X \setminus E_r) \neq 0$ as well, and thus we have a contradiction.
(2) $\implies$ (1): Let $E \in \mathcal{M}$ be a T-invariant set and consider $f = \chi_E$. This is a measurable function, and we see that

$$\chi_E \circ T = \chi_{T^{-1}(E)} = \chi_E.$$

This implies χ_E is constant almost everywhere, so $\chi_E = 0$ or 1 almost everywhere. If $\chi_E = 0$ almost everywhere, we have

$$\mu(E) = \int \chi_E d\mu = 0.$$

If $\chi_E = 1$ almost everywhere, then $\chi_{X \setminus E} = 0$ almost everywhere, so

$$\mu(X \setminus E) = \int \chi_{X \setminus E} d\mu = 0.$$

This holds for every T-invariant set, so T is ergodic.

(1) $\implies$ (3): Notice that 1 is always an eigenvalue. We need to show that it is a simple eigenvalue, meaning $Uf = f$ implies f is constant almost everywhere. But f being T-invariant implies that it is constant almost everywhere by the equivalence of (1) $\iff$ (2).
(3) $\implies$ (1): Notice 1 is always an eigenvalue, since constant functions are invariant. For it to be a simple eigenvalue means that $Uf = f$ implies f is constant almost everywhere. So all $f \in L^2(X, \mathcal{M}, \mu)$ which are T-invariant are constant almost everywhere. Does this imply that every measurable function which is T-invariant is constant almost everywhere? Doesn't seem to be necessarily true.

(1) $\iff$ (4): The Birkhoff ergodic theorem doesn't necessarily say what we want for this to work.

(1) $\iff$ (5): The implication still works in infinite measure spaces. The converse doesn't necessarily hold (the containment doesn't hold true).

□

Problem 31 (Petersen 2.4.4)**.** Consider $(X, \mathcal{M}, \mu, T)$ a measure preserving system of a probability measure space.

(1) Prove that T is ergodic if and only if

$$\frac{1}{n} \sum_{0}^{n-1} \mu(T^k(A) \cap B) \to \mu(A)\mu(B) \text{ for all } A, B \in \mathcal{M}.$$

(2) Prove that T is ergodic if and only if (1) holds on a semialgebra which generates the σ-algebra.

Proof.

(1) ($\Longrightarrow$): Assume that T is ergodic. We can write the condition as

$$\lim_{n\to\infty}\frac{1}{n}\sum_{0}^{n-1}\mu(T^k(A)\cap B)=\lim_{n\to\infty}\frac{1}{n}\sum_{0}^{n-1}\int\chi_A(T^{-k}(x))\chi_B(x)d\mu.$$

Now using the fact that T is ergodic and the Birkhoff Ergodic Theorem, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{0}^{n-1}\chi_A(T^{-k}(x))=\chi_A^*(x).$$

Notice we have

$$\int\chi_A^*(x)d\mu=\chi_A^*(x)\mu(X)=\mu(A)\implies\chi_A^*(x)=\frac{\mu(A)}{\mu(X)}=\mu(A),$$

since χ_A^* is constant almost everywhere and since $\mu(X)=1$. Multiplying this by χ_B gives

$$\lim_{n\to\infty}\frac{1}{n}\sum_{0}^{n-1}\chi_A(T^{-k}(x))\chi_B(x)=\mu(A)\chi_B(x).$$

Now use the dominated convergence theorem to get

$$\int\left(\lim_{n\to\infty}\frac{1}{n}\sum_{0}^{n-1}\chi_A(T^{-k}(x))\right)d\mu=\lim_{n\to\infty}\frac{1}{n}\sum_{0}^{n-1}\mu(T^k(A)\cap B)=\mu(A)\mu(B).$$

($\Longleftarrow$): Let $A=B=E$, where E is a T-invariant set. Then

$$\frac{1}{n}\sum_{0}^{n-1}\mu(E)\to\mu(E)=\mu(E)^2.$$

This means $\mu(E)=0$ or $\mu(E)=1$.

(2) ($\Longrightarrow$): Clear, it holds on the entire algebra.
($\Longleftarrow$): We just need to show that the property on the semialgebra implies the property on the σ-algebra. See **Theorem 1.17** [4]. Elements in the algebra can be written as disjoint unions of elements in the semialgebra, so it's clear that it will hold on the algebra. Now assume that A, B are in the σ-algebra. Fix $\epsilon>0$ small (where the smallness will be chosen in the future). We can find A_0, B_0 in the algebra so that $\mu(A\triangle A_0),\mu(B\triangle B_0)<\epsilon$. Notice that

$$(T^{-k}(A)\cap B)\triangle(T^{-k}(A_0)\cap B)\subseteq(T^{-k}(A)\triangle T^{-k}(A_0))\cup(B\triangle B_0).$$

Taking the measure and using measure preserving, we have

$$\mu\left((T^{-k}(A)\cap B)\triangle(T^{-k}(A_0)\cap B)\right)<2\epsilon.$$

Thus

$$|\mu(T^{-k}(A)\cap B)-\mu(T^{-k}(A_0)\cap B_0)|<2\epsilon.$$

Now, notice

$$\begin{aligned}|\mu(T^{-k}(A)\cap B)-\mu(A)\mu(B)|\le|\mu(T^{-k}(A)\cap B)-\mu(T^{-k}(A_0)\cap B_0)|+|\mu(T^{-k}(A_0)\cap B_0)-\mu(A_0)\mu(B_0)|\\+|\mu(A_0)\mu(B_0)-\mu(A)\mu(B_0)|+|\mu(A)\mu(B_0)-\mu(A)\mu(B)|\\<4\epsilon+|\mu(T^{-k}(A_0)\cap B_0)-\mu(A_0)\mu(B_0)|.\end{aligned}$$

Now

$$\left|\frac{1}{n}\sum_0^{n-1}\mu(T^{-k}(A)\cap B)-\mu(A)\mu(B)\right| \leq \left|\frac{1}{n}\sum_0^{n-1}\mu(T^{-k}(A)\cap B)-\frac{1}{n}\sum_0^{n-1}\mu(T^{-k}(A_0)\cap B_0)\right|$$
$$+\left|\frac{1}{n}\sum_0^{n-1}\mu(T^{-k}(A_0)\cap B_0)-\mu(A_0)\mu(B_0)\right| + |\mu(A_0)\mu(B_0)-\mu(A)\mu(B_0)|$$
$$+|\mu(A)\mu(B_0)-\mu(A)\mu(B)|$$
$$< \frac{\epsilon}{n}+\left|\frac{1}{n}\sum_0^{n-1}\mu(T^{-k}(A_0)\cap B_0)-\mu(A_0)\mu(B_0)\right| + \mu(A_0)\epsilon+\mu(A)\epsilon.$$

We can choose $\epsilon > 0$ small enough so that $\epsilon < 1/\mu(A)$ and smaller than 1. Notice that $|\mu(A)-\mu(A_0)| < \epsilon$ so that $\mu(A_0) < \epsilon + \mu(A)$. Putting this all together, we get

$$\left|\frac{1}{n}\sum_0^{n-1}\mu(T^{-k}(A)\cap B)-\mu(A)\mu(B)\right| < 4\epsilon + \left|\frac{1}{n}\sum_0^{n-1}\mu(T^{-k}(A_0)\cap B_0)-\mu(A_0)\mu(B_0)\right|.$$

Now taking the limit as $n\to\infty$ of both sides to get

$$\lim_{n\to\infty}\left|\frac{1}{n}\sum_0^{n-1}\mu(T^{-k}(A)\cap B)-\mu(A)\mu(B)\right| < 4\epsilon.$$

We have $\epsilon > 0$ small arbitrary, so taking $\epsilon \to 0$ gives us the desired result. So this holds on the σ-algebra.

□

Problem 32 (Petersen 2.4.5)**.** Let T be ergodic and $\nu \ll \mu$ a measure on $(X, \mathcal{M})$ such that $\nu(T^{-1}) \leq \nu$. Show that $\nu(T^{-1}) = \nu$ and ν is a constant multiple of μ. *Note:* $(X, \mathcal{M}, \mu)$ *is a probability measure space.*

Remark. I followed this, but I don't actually think it is right as is. I hopefully cleaned it up.

Proof. Assume without loss of generality ν is a probability measure as well, so $\nu(X) = 1$. Let $g = \frac{\partial\nu}{\partial\mu}$ be the Radon-Nikodym derivative, so that for all $E \in \mathcal{M}$ we have

$$\nu(E) = \int_E g\,d\mu.$$

Consider $A = \{g > 1\}$. Notice that

$$\mu(A) = \int_A d\mu < \int_A g\,d\mu = \nu(A) \leq 1,$$

so we have $\mu(A) < 1$.

Notice as well that

$$T^{-1}(A) = (T^{-1}(A)\setminus A) \sqcup (T^{-1}(A)\cap A),$$

so

$$\nu(T^{-1}(A)) = \nu(T^{-1}(A)\setminus A) + \nu(T^{-1}(A)\cap A).$$

We now use the fact that $\nu(T^{-1}) \leq \nu$ to get

$$\nu(T^{-1}(A)\setminus A) + \nu(T^{-1}(A)\cap A) \leq \nu(A).$$

Use the fact that

$$A = (A\setminus T^{-1}(A)) \sqcup (A\cap T^{-1}(A))$$

to get
$$\nu(T^{-1}(A) \setminus A) + \nu(T^{-1}(A) \cap A) \leq \nu(A \setminus T^{-1}(A)) + \nu(A \cap T^{-1}(A)).$$
Simplifying, we have
$$\nu(T^{-1}(A) \setminus A) \leq \nu(A \setminus T^{-1}(A)).$$

Now assume that $\mu(A) > 0$. Since T is ergodic, we have
$$\mu(A \setminus T^{-1}(A)) \geq \mu(A \triangle T^{-1}(A)) > 0.$$
The first observation we have is
$$\mu(T^{-1}(A) \setminus A) = \int_{T^{-1}(A)\setminus A} d\mu.$$
Now
$$T^{-1}(A) \setminus A = \{x \in X : g(T(x)) > 1, g(x) \leq 1\}.$$
Substitute this in and use the fact that $\nu(T^{-1}(E)) \leq \nu(E)$ as well as the earlier inequality we derived to get
$$\mu(T^{-1}(A) \setminus A) < \int_{T^{-1}(A)\setminus A} g(T(x)) d\mu(x) = \nu(T^{-1}(T^{-1}(A) \setminus A)) \leq \nu(T^{-1}(A) \setminus A) \leq \nu(A \setminus T^{-1}(A)).$$
Now again use the inequality $\nu(T^{-1}(E)) \leq \nu(E)$ as well as the fact that T is measure preserving and invertible to get
$$\nu(A \setminus T^{-1}(A)) = \nu(T^{-1}(T(A) \setminus A)) \leq \nu(T(A) \setminus A) = \int_{T(A)\setminus A} g d\mu \leq \mu(T(A) \setminus A) = \mu(A \setminus T^{-1}(A)).$$
But we have
$$\mu(A \setminus T^{-1}(A)) + \mu(A \cap T^{-1}(A)) = \mu(A) = \mu(T^{-1}(A)) = \mu(T^{-1}(A) \setminus A) + \mu(T^{-1}(A) \cap A),$$
which implies
$$\mu(A \setminus T^{-1}(A)) = \mu(T^{-1}(A) \setminus A).$$
Thus we have
$$\mu(T^{-1}(A) \setminus A) < \mu(T^{-1}(A) \setminus A),$$
which is a contradiction. Thus $\mu(A) = 0$. Thus $g \leq 1$ almost everywhere. Notice that if $g < 1$ strictly almost everywhere, then
$$\nu(X) = \int_X g d\mu < \int_X d\mu = \mu(X) = 1,$$
which contradicts the fact that ν is a probability measure. This implies that $g = 1$ almost everywhere, but this means that $\mu = \nu$.

Now if we had normalized ν above, so that $\nu(E) = \kappa(E)/\kappa(X)$ for some finite measure κ, then we see that this adjusts g by a constant. However if g is a constant almost everywhere, then κ is a constant multiple of μ. □

Problem 33 (Petersen 2.4.6)**.** *Let $X = [0, 1]$ with Lebesgue measure m. Then T (preserving m) is ergodic on X if and only if for every continuous f*
$$\frac{1}{n} \sum_{0}^{n-1} f(T^k(x)) \to \int f dm \text{ almost everywhere.}$$

Proof. The forward direction is clear, since continuous functions are measurable. The backward direction is more interesting. By an earlier equivalence, we just need to show that this property holding for continuous functions, then it holds for all L^1 functions. Use Lusin's theorem for density of continuous functions in L^1. □

Problem 34 (Petersen 2.4.7)**.** Let T be an ergodic measure preserving transformation on a nonatomic probability space $(X, \mathcal{M}, \mu)$. Let $U : L^2(X, \mathcal{M}, \mu) \to L^2(X, \mathcal{M}, \mu)$ be the unitary operator associated to T.

(1) Show that every point of the unit circle is an **approximate eigenvalue** of U in the following sense: Given λ with $|\lambda| = 1$ there are $f_n \in L^2$ with $\|f_n\|_2 = 1$ for all n and $\|Uf_n - \lambda f_n\|_2 \to 0$.
(2) Deduce the spectrum of U is the entire unit circle.

We recall **Rokhlin Lemma** (see here).

Theorem (Rokhlin Lemma)**.** Let $T : X \to X$ be an invertible measure-preserving transformation on a probability measure space. If the collection of periodic points has measure zero, then for every $\epsilon > 0$ and n fixed there is a measurable set E such that the sets $\{T^j(E)\}_{j=0}^{n-1}$ are pairwise disjoint and such that

$$\mu(E \cup \cdots \cup T^{n-1}(E)) > 1 - \epsilon.$$

Remark. See here.

Proof.

(1) Take $\lambda \in \mathbb{C}$ with $|\lambda| = 1$ (so that it lies on the unit circle). Fix $\epsilon > 0$. Take the set in Rokhlin's lemma, $E_{n,\epsilon}$ so that $\{T^j(E_{n,\epsilon})\}_{j=0}^{n}$ almost covers X and they are pairwise disjoint. Define $f_{n,\epsilon}$ by

$$f_{n,\epsilon}(T^i(x)) = \lambda^i \text{ for } x \in E_{n,\epsilon},$$

and 1 everywhere else. Then

$$\|f_{n,\epsilon}\|_2 = \epsilon + (1-\epsilon)|\lambda| = 1.$$

Notice that if $x \in E_{n,\epsilon}$ and $0 \le j \le n-1$ we have

$$U(f_{n,\epsilon})(T^j(x)) = f_{n,\epsilon}(T^{j+1}(x)) = \lambda^{j+1} = \lambda f_{n,\epsilon}(T^j(x)).$$

Notice as well that

$$\mu(E_{n,\epsilon}) \le \frac{1}{n+1}.$$

Now on $F = \bigcup_0^{n-1} T^j(E_{n,\epsilon})$ we have $Uf_{n,\epsilon} = \lambda f_{n,\epsilon}$. So they are equal on a set of measure greater than or equal to $1 - \epsilon - 1/(n+1)$. Now

$$\|Uf_{n,\epsilon} - \lambda f_{n,\epsilon}\|_2 \le \left(\int_{F^c} |Uf_{n,\epsilon} - \lambda f_{n,\epsilon}| d\mu\right)^{1/2} \le (2\mu(F^c))^{1/2} \le \sqrt{2}\left(\epsilon + \frac{1}{n+1}\right)^{1/2}.$$

We can set $\epsilon = n^{-1}$ and take $n \to \infty$ to get the result.

(2) Notice the choice of $\lambda \in \mathbb{C}$ with $|\lambda| = 1$ didn't matter.

□

Problem 35 (Petersen 2.4.8)**.** Suppose that $(\Omega, \mathcal{M}, \mu, \sigma)$ is a Markov shift determined by a given stochastic matrix A and fixed probability vector p with all $p_i > 0$. Prove that if $(\Omega, \mathcal{M}, \mu, \sigma)$ is ergodic, then A is irreducible.

Proof. The goal is to show that for any i, j we have there is a k so that $a_{i,j}^k > 0$, where $a_{i,j}^k = (A^k)_{i,j}$. In other words, $a_{i,j}^k$ is the probability that $\omega_0 = i$ and the probability that $\omega_k = j$, where $\omega \in \Omega$. We define this as

$$a_{i,j}^k = \frac{\mu(\sigma^{-k}(C_j) \cap C_i)}{\mu(C_i)}, \quad C_k = \{\omega \in \Omega : \omega_0 = k\}.$$

Notice that

$$\frac{1}{n}\sum_0^{n-1} \mu(\sigma^{-k}(C_j) \cap C_i) \to \mu(C_j)\mu(C_i) > 0,$$

so there must be some n so that the term on the left hand side is positive, and for this n there must be some $k \in \{0, \dots, n-1\}$ so that $\mu(\sigma^{-k}(C_j) \cap C_i) > 0$. Since $\mu(C_i) > 0$, we are done. □

Problem 36 (Petersen 2.4.9)**.** Prove that for any ergodic measure preserving transformation on a nonatomic space there is a set A of positive measure so that the return time n_A is unbounded.

Remark. See here.

Proof. We invoke Rokhlin's lemma again. Fix $\epsilon > 0$ arbitrary. We see that for every n we can find an n so that

$$\mu(E_n \cup T(E_n) \cup \cdots \cup T^n(E_n)) > 1 - \epsilon$$

and so that these sets are disjoint. Notice that by disjointness and measure preserving, we have

$$1 \geq n\mu(E_n) > 1 - \epsilon \implies \frac{1}{n+1} \geq \mu(E_n) > \frac{1-\epsilon}{n+1}.$$

So take $x \in E_n$, then $n_{E_n}(x) \geq n$. Consider $E = \bigcap_{n=1}^{\infty} E_n$. If there is an $x \in E$, Then $x \in E_n$ for all n, so $n_{E_n}(x) \geq n$ for each n and hence $n_E(x)$ is unbounded. It might be, however, that E is empty. We see from here we need to be careful about our choices of E_n.

How do we fix this? We can first find E_1 so that $E_1, T(E_1)$ are pairwise disjoint and $\mu(E_1) < 1/2$. Set $X_1 = X \setminus T(E_1)$. Then for $x \in E_1$, $n_{X_1}(x) \geq 2$ and $\mu(X_1) > 1/2$. Now using the proof of Rokhlin, we can find $E_2 \subseteq E_1$ so that $T^2(E_2) \subseteq E_1$, $T(E_2), T^3(E_2) \subseteq T(E_1)$ and $\mu(E_2) < 1/4$. Set $X_2 = X_1 \setminus T^2(E_2)$. Then $\mu(X_2) > 1/4$ and $n_{X_2}(x) \geq 4$ for $x \in E_2$. Continue in this fashion. Then $\mu(X_n) > 1/2^n$ and $n_{X_n}(x) \geq 2^n$ for $x \in E_n$, where E_n is chosen appropriately. Set $X_\infty = \bigcap X_n$. Then $\mu(X_\infty) > 0$ and $n_{X_\infty}(x)$ is unbounded. □

Problem 37. Let G be a (first countable) compact topological group, and let $g \in G$. Show that there exists a sequence $n_k \nearrow \infty$ so that $g^{n_k} \to e$ as $k \to \infty$.

Proof. Notice that the map $L_g : G \to G$ defined by $L_g(h) = gh$ is a measure preserving homeomorphism. Fix $y \in G$. Take a neighborhood basis $\{U_n\}$ for y. Without loss of generality, assume that this is a decreasing sequence of sets, i.e. $U_{n+1} \subseteq U_n$. By Poincare recurrence, we know that

$$\Gamma_n = \{x \in U_n : x \text{ is infinitely recurrent}\}$$

is such that $\mu(\Gamma_n) = \mu(U_n) > 0$, so $\Gamma_n \neq \varnothing$. For each n, let $y_n \in \Gamma_n$. Then we have that there is a sequence (m_k^n) so that $g^{m_k^n} y_n \in U_n$ for all k (note here that we need to choose these sequences so that $m_k^k \leq m_n^n$ for $k \leq n$, but doing so isn't hard and it's just a matter of refining the sequences if needed). Taking $m_n = m_n^n$, we note that $g^{m_n} y_n \in U_n$ and we claim that this shows $g^{m_n} y_n \to y$. To see this, take any neighborhood U of y. Since $\{U_n\}$ is a decreasing basis, we have that there exists a N so that for all $n \geq N$ $U_n \subseteq U$. Now for all $n \geq N$, we have that $g^{m_n} y_n \in U_n \subseteq U$. This holds for all neighborhoods U, so this implies convergence.

Now we have $y_n \to y$, $g^{m_n} y_n \to y$, and we wish to show $g^{m_n} \to e$. To get this, we claim that $(g^{m_n} y_n) y_n^{-1} = g^{m_n} \to e$. Since multiplication is continuous, we have

$$\lim_{n\to\infty} (g^{m_n} y_n) y_n^{-1} = \left(\lim_{n\to\infty} g^{m_n} y_n\right)\left(\lim_{n\to\infty} y_n^{-1}\right).$$

By continuity of inversion, $y_n^{-1} \to y^{-1}$, so using this and the above we have

$$\lim_{n\to\infty} g^{m_n} = \lim_{n\to\infty} (g^{m_n} y_n) y_n^{-1} = \left(\lim_{n\to\infty} g^{m_n} y_n\right)\left(\lim_{n\to\infty} y_n^{-1}\right) = (y)(y^{-1}) = e.$$

□

Problem 38 (Petersen 2.4.10)**.** Show that if T has discrete spectrum, then there is a sequence of integers $n_k \nearrow \infty$ with $T^{n_k} \to I$ in the strong operator topology on L^2. In other words,

$$\|T^{n_k}(f) - f\|_2 \to 0 \text{ for all } f \in L^2.$$

Remark. See here.

As Thomas pointed out, I'm implicitly assuming the space is "nice" enough for things to work (i.e. L^2 is a separable Hilbert space).

Proof. Let $\{f_k\}$ be a sequence of orthonormal eigenfunctions for U on L^2. Then $U(f_i) = f_i(T) = \lambda_i f_i$ with $|\lambda_i| = 1$. For each i, we can find a sequence $\{n_i\}$ so that $\lambda^{n_i} \to 1$ (follows by the last problem). Now take any $f \in L^2$. We can approximate it with the span $\{f_k\}$, so there is some $g \in \text{span}\{f_k\}$ so that $\|f - g\|_2 < \epsilon$, $\epsilon > 0$ fixed. Notice that

$$\|T^{n_k} f - f\|_2 \leq \|T^{n_k} g - g\|_2 + \|T^{n_k} g - T^{n_k} f\|_2 + \|f - g\|_2 < 2\epsilon.$$

This holds for all $\epsilon > 0$, so we get the result. □

Recall that a measure preserving transformation has **Lebesgue spectrum of multiplicity** N (where N is a finite or infinite cardinal number) in the case that there is a set Λ of cardinality N and a set of functions

$$\{f_{\lambda,j} : \lambda \in \Lambda, j \in \mathbb{Z}\}$$

which together with 1 form an orthonormal basis for $L^2(X, \mathcal{M}, \mu)$ and such that

$$U_T f_{\lambda,j} = f_{\lambda,j+1} \text{ for all } \lambda \in \Lambda, j \in \mathbb{Z}.$$

A measure preserving transformation T on $(X, \mathcal{M}, \mu)$ is a K-automorphism if there is a subσ-algebra $\mathcal{A} \subseteq \mathcal{M}$ such that

(1) $T^{-1}(\mathcal{A}) \subseteq \mathcal{A}$.
(2) $\bigcup_{n=-\infty}^{\infty} T^n(\mathcal{A})$ generates $\mathcal{M}$.
(3) $\bigcap_{n=-\infty}^{0} T^n(\mathcal{A})$ is trivial.

Recall **Petersen Proposition 2.5.11**.

Proposition (Petersen Proposition 2.5.11)**.** Every K-automorphism has countable Lebesgue spectrum.

Problem 39. Show that for every N there exists $A_1, \ldots, A_n \in \mathcal{A}$ pairwise disjoint sets with positive measure.

Proof. We can use Rokhlin's lemma to find $B_1, \ldots, B_n \in \mathcal{M}$ pairwise disjoint with positive measure. By property (2) of K-automorphisms, we have that we canapproximate the B_i arbitrarily well with $A_i \in \mathcal{A}$. By (1), we can choose $A_i \in T^n(\mathcal{A})$ by choosing n large enough, so $L^2(X, T^n(\mathcal{A}), \mu)$ has dimension at least N, forcing $T^n(\mathcal{A})$ to contain at least N pairwise disjoint sets of positive measure. We can then take preimages to get them in $\mathcal{A}$. □

Problem 40 (Petersen 2.5.1)**.** Finish the proof of 2.5.11 by showing that the orthogonal complement of UM is countable.

Proof. Fix n. Use the last problem to find $A_1, \ldots, A_n \in \mathcal{A}$ with positive measure which are pairwise disjoint. Pick $f \in W \setminus \{0\}$. Note that such an f exists because otherwise $(X, \mathcal{M}, \mu)$ is atomic. Set $w_i = f\chi_{A_i} \circ T$. Note three things:

(1) The w_i are linearly independent (disjoint support).
(2) The $w_i \in V$.
(3) The $w_i \in (UM)^\perp$.

To check this last fact, take $E \in \mathcal{A}$ and examine $\chi_E \circ T \in UM$. Taking arbitrary w_i, we see that

$$(w_i, \chi_E \circ T) = \int w_i \chi_{T^{-1}(E)} = \int f(T(x))\chi_{T^{-1}(A_i)}\chi_{T^{-1}(E)} = \int f(T(x))\chi_{T^{-1}(A_i \cap E)} = 0.$$

So $\dim(W) \geq N$ for all N, and we get it's infinite. □

A map is **weakly mixing** if

$$\frac{1}{n}\sum_{0}^{n-1}|\mu(T^{-k}(A)\cap B)-\mu(A)\mu(B)|\to 0 \text{ for all } A,B\in\mathcal{M}.$$

A map is **strongly mixing** if

$$|\mu(T^{-k}(A)\cap B)-\mu(A)\mu(B)|\to 0 \text{ for all } A,B\in\mathcal{M}.$$

Problem 41 (Petersen 2.5.4, Ornsteins Criterion)**.** Prove that a weakly mixing Markov shift is strongly mixing.

Proof. We follow **Theorem 1.31** [4]. We will show that weak mixing implies the matrix A is irreducible and aperiodic (in this context, there exists some N so that $A^N > 0$). Then we will use that to establish strong mixing.

Let $C_j = \{\omega \in \Omega : \omega_0 = j\}$ be the cylinders. Weak mixing here says that

$$\frac{1}{n}\sum_{0}^{n-1}|\mu(T^{-l}(C_i)\cap C_j)-\mu(C_i)\mu(C_j)|\to 0 \text{ for all } i,j.$$

Notice

$$a_{i,j}^k = \frac{\mu(T^{-k}(C_i)\cap C_j)}{\mu(C_i)},$$

so substituting this in we get

$$\frac{\mu(C_i)}{n}\sum_{0}^{n-1}|a_{i,j}^l - p_j|\to 0 \text{ for all } i,j,$$

or

$$\frac{1}{n}\sum_{0}^{n-1}|a_{i,j}^l - p_j|\to 0 \text{ for all } i,j.$$

Using a real analysis exercise **Theorem 1.20** [4] we get that there is some subsequence (n_k) so that

$$a_{i,j}^{n_k}\to p_j.$$

This gives us irreducible and aperiodic. The renewal theroem tells us that

$$a_{i,j}^{n}\to p_j.$$

Notice that for any two cylinders now,

$$\mu(T^{-k}(C_i)\cap C_j) = a_{i,j}^k\mu(C_i)\to p_j\mu(C_i) = \mu(C_j)\mu(C_i).$$

Since the cylinders generate things, this is sufficient. ☐

Problem 42 (Petersen 2.5.5)**.** Prove that T is weakly mixing if and only if

$$\lim_{n\to\infty}\frac{1}{n}\sum_{0}^{n-1}|\mu(T^{-k}(A)\cap A)-\mu(A)^2| = 0 \text{ for all } A\in\mathcal{M}.$$

Proof. The forward direction is clear, so let's assume this condition and show that T is weakly mixing. We follow the proof of **Theorem 5.5** [3] (for a similar result on strong mixing). Fix $A\in\mathcal{M}$. Consider the subspace $H\subseteq L^2(X,\mathcal{M},\mu)$ generated by the constant functions along with $\{U^k\chi_A : k\in\mathbb{Z}\}$. Notice that

$$\langle U^k\chi_A, 1\rangle = \int \chi_{T^{-k}(A)}d\mu = \mu(T^{-k}(A)) = \mu(A),$$

so

$$\frac{1}{n}\sum_{0}^{n-1}\left|\langle U^k\chi_A,1\rangle - \langle\chi_A,1\rangle\langle 1,1,\rangle\right| = \frac{1}{n}\sum_{0}^{n-1}|\mu(A)-\mu(A)| = 0.$$

Notice as well that

$$\langle U^k\chi_A, U^j\chi_A\rangle = \langle U^{k-j}\chi_A,\chi_A\rangle = \mu(T^{k-j}(A)\cap A)$$

so for fixed j and varying k we have

$$\frac{1}{n}\sum_{k=0}^{n-1}\left|\langle U^k\chi_A, U^j\chi_A\rangle - \langle U^k\chi_A,1\rangle\langle 1,U^j\chi_A\rangle\right| = \frac{1}{n}\sum_{0}^{n-1}\left|\mu(T^{k-j}(A)\cap A)-\mu(A)^2\right| = 0.$$

Thus for all $f\in H$ we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\left|\langle U^k\chi_A,f\rangle - \langle\chi_A,1\rangle\langle 1,f\rangle\right| = 0.$$

Notice this is a closed subspace. Decompose $L^2(X,\mathcal{M},\mu) = H\oplus H^\perp$, so for all $f\in L^2(\mu)$ we can write $f = f_1+f_2$, $f\in H$ and $f_2\in H^\perp$. Then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\left|\langle U^k\chi_A,f\rangle - \langle\chi_A,1\rangle\langle 1,f\rangle\right| = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\left|\langle U^k\chi_A,f_1\rangle - \langle\chi_A,1\rangle\langle 1,f_1\rangle\right| = 0.$$

Note here we utilized the fact that the constant functions are in H so that $\langle 1,f_2\rangle = 0$, and then we used linearity. Consequently we have the result holds for all $f\in L^2(\mu)$. Now for $B\in\mathcal{M}$, we have $\chi_B\in L^2(\mu)$, so

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\left|\langle U^k\chi_A,\chi_B\rangle - \langle\chi_A,1\rangle\langle 1,\chi_B\rangle\right| = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\left|\mu(T^{-k}(A)\cap B)-\mu(A)\mu(B)\right| = 0.$$

Thus T is weak mixing. □

Recall that a system is **n-mixing** if for all choices of n sets $A_1,\dots,A_n\in\mathcal{M}$ we have

$$\lim_{\substack{\inf_i m_i\to\infty\\ \inf_{i\neq j}|m_i-m_j|\to\infty}} \mu(T^{m_1}A_1\cap\cdots\cap T^{m_n}A_n) = \mu(A_1)\cdots\mu(A_n).$$

Problem 43 (Petersen 2.5.6)**.** Prove that Bernoulli shifts, mixing Markov shifts, and ergodic automorphisms of a compact abelian group are n-mixing for all n.

Proof. Let's show Bernoulli shifts are n-mixing for all n.

Consider $(\Omega,\mathcal{B},p,\sigma)$ where $\Omega=\{0,1\}^{\mathbb{Z}}$, $\sigma(\omega)=\omega'$, where $\omega'(n)=\omega(n+1)$ and p and $\mathcal{B}$ are the usual things (p is generated by vector $p_0=1/2$ and $p_1=1/2$). If we show it on cylinders, we win. Consider the cylinders

$$C_j^n = \{\omega\in\Omega : \omega(n)=j\}.$$

We examine

$$\lim_{\substack{\inf_i m_i\to\infty\\ \inf_{i\neq j}|m_i-m_j|\to\infty}} p(\sigma^{m_1}(C_{j_1}^{t_1})\cap\cdots\cap\sigma^{m_n}(C_{j_n}^{t_n})).$$

Notice

$$\sigma^{m_1}(C_{j_1}^{t_1}) = \{\omega\in\Omega : \omega(t_1+m_1)=j_1\}.$$

Thus fixing $m_1,\dots,m_n$ distinct and far enough apart (which we can by construction of the limit) we have

$$\sigma^{m_1}(C_{j_1}^{t_1})\cap\cdots\cap\sigma^{m_n}(C_{j_n}^{t_n}) = \{\omega\in\Omega : \omega(t_1+m_1)=j_1,\dots,\omega(t_n+m_n)=j_n\}.$$

Since we assume things are far enough apart and distinct, we can use the definition of p to calculate

$$p(\sigma^{m_1}(C_{j_1}^{t_1}) \cap \cdots \cap \sigma^{m_n}(C_{j_n}^{t_n})) = p(C_{j_1}^{t_1}) \cdots p(C_{j_n}^{t_n}).$$

As we take things further and further apart, this doesn't change, so we get n-mixing.

Mixing Markov shifts are isomorphic to a Bernoulli shift, so I don't think we need to do any kind of argument there. I believe ergodic automorphisms of a compact abelian group also have this nice property, so I think this suffices. □

Problem 44 (Petersen 2.5.7)**.** Show that there is no concept of "uniform mixing" for measure preserving transformations. That is, if

$$\mu(T^{-n}(A) \cap B) \to \mu(A)\mu(B)$$

uniformly for all $A, B \in \mathcal{B}$ with $A \subseteq B$, then every set in $\mathcal{M}$ has measure 0 and 1, so $(X, \mathcal{M}, \mu)$ is isomorphic with the space consisting of a single point.

Remark. There is a paper by **Halmos** (see here) which discusses this problem. The hint Petersen gives is essentially the gist of Halmos' argument.

Proof. Let's first just assume that it converges uniformly for all $A, B \in \mathcal{M}$. That is, for all $A, B \in \mathcal{M}$ and $\epsilon > 0$, there exists an N so that for all $n \geq N$ we have

$$|\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| < \epsilon.$$

If we let $B = T^{-n}(A)$, then this says that

$$|\mu(A) - \mu(A)^2| < \epsilon.$$

We can do this for all $\epsilon > 0$, so this forces $\mu(A)$ to be 0 or 1. The choice of A was arbitrary, so the measure of all sets must be either 0 or 1.

Now we go back to the original condition. The goal is to show that with the condition of uniform convergence for all $A, B \in \mathcal{M}$ with $A \subseteq B$, we have uniform convergence for all $A, B \in \mathcal{M}$. So take $A, B \in \mathcal{M}$ arbitrary. The goal is to show that for all $\epsilon > 0$, there is an N so that for $n \geq N$ we have

$$|\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| < \epsilon.$$

Fix $\epsilon > 0$. Notice that $A \cap B \subseteq B$, so there is an N such that for $n \geq N$ we have

$$|\mu(T^{-n}(A \cap B) \cap B) - \mu(A \cap B)\mu(B)| < \epsilon/2.$$

Simultaneously, we have $A \cap B^c \subseteq B^c$, so for all $\epsilon > 0$ there is an N such that for $n \geq N$ we have

$$|\mu(T^{-n}(A \cap B^c) \cap B^c) - \mu(A \cap B^c)\mu(B^c)| < \epsilon/2.$$

Now, notice that

$$\mu(A \cap B^c) - \mu(A \cap B^c)\mu(B^c) = \mu(A \cap B^c)(\mu(X) - \mu(B^c)) = \mu(A \cap B^c)\mu(B).$$

So we may rewrite the above as

$$|\mu(T^{-n}(A \cap B^c) \cap B^c) - \mu(A \cap B^c) + \mu(A \cap B^c)\mu(B)| < \epsilon/2.$$

Notice

$$A \cap B^c = (A \cap B^c \cap T^n B) \sqcup (A \cap B^c \cap T^n B^c),$$

so that

$$\begin{aligned} T^{-n}(A \cap B^c) &= (T^{-n}(A \cap B^c) \cap B) \sqcup (T^{-n}(A \cap B^c) \cap B^c), \\ \mu(A \cap B^c) = \mu(T^{-n}(A \cap B^c)) &= \mu(T^{-n}(A \cap B^c) \cap B) + \mu(T^{-n}(A \cap B^c) \cap B^c), \\ \mu(T^{-n}(A \cap B^c) \cap B^c) &= \mu(A \cap B^c) - \mu(T^{-n}(A \cap B^c) \cap B). \end{aligned}$$

Use this to rewrite the above again as

$$|\mu(A \cap B^c)\mu(B) - \mu(T^{-n}(A \cap B^c) \cap B)| = |\mu(T^{-n}(A \cap B^c) \cap B) - \mu(A \cap B^c)\mu(B)| < \epsilon/2.$$

Now note

$$T^{-n}(A) = T^{-n}(A \cap B) \sqcup T^{-n}(A \cap B^c),$$
$$A = (A \cap B) \sqcup (A \cap B^c),$$

hence

$$\begin{aligned}
&|\mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)| \\
&= |\mu(T^{-n}(A \cap B) \cap B) + \mu(T^{-n}(A \cap B^c) \cap B) - \mu(A \cap B)\mu(B) - \mu(A \cap B^c)\mu(B)| \\
&\leq |\mu(T^{-n}(A \cap B) \cap B) - \mu(A \cap B)\mu(B)| + |\mu(T^{-n}(A \cap B^c) \cap B) - \mu(A \cap B^c)\mu(B)| < \epsilon.
\end{aligned}$$

□

References

[1] Karma Dajani and Sjoerd Dirksin. *A Simple Introduction to Ergodic Theory*. 2008. Link.
[2] Anatole Katok and Boris Hasselblatt. *Introduction to the Modern Theory of Dynamics Systems*. 1995.
[3] Karl Petersen. *Ergodic Theory*. 1983.
[4] Peter Walters. *An Introduction to Ergodic Theory*. 1982.

Remark. Thomas O'Hare was a collaborator.

Problem 45. Suppose T acts ergodically on a probability measure space $(X, \mathcal{M}, \mu)$. Let λ be a T-invariant probablility measure on X. Prove the following.

(1) If $\lambda \ll \mu$, then $\lambda = \mu$.
(2) If λ is T-ergodic and $\lambda \neq \mu$, then $\lambda \perp \mu$.

Proof.

(1) Let $g = \frac{d\lambda}{d\mu}$ be the Radon-Nikdoym derivative. The first step is to show that this is T-invariant, using the fact that μ and ν are T-invariant. So we need to show that
$$g \circ T = g.$$
Since μ is T-invariant, we have
$$\lambda(T^{-1}(A)) = \int_{T^{-1}(A)} g d\mu(x) = \int_A (g \circ T)(x) d(\mu \circ T^{-1})(x) = \int_A (g \circ T)(x) d\mu(x).$$
Since λ is T-invariant, we have
$$\lambda(T^{-1}(A)) = \lambda(A) = \int_A g(x) d\mu(x).$$
Putting these together, we see that for all $A \in \mathcal{M}$ we have
$$\int_A g d\mu = \int_A g \circ T d\mu.$$
By **Folland Proposition 2.23**, we see that this implies $g = g \circ T$ almost everywhere, so that g is T-invariant μ almost everywhere. This means that g is constant, so that $\lambda = c\mu$ for some constant $c \geq 0$. Notice that $\lambda(X) = 1 = c\mu(X) = c$, so $c = 1$. Thus $\lambda = \mu$.

(2) Now use the **Lebesgue-Radon-Nikdoym theorem** to write
$$\lambda = \lambda_1 + \lambda_2, \qquad \lambda_1 \ll \mu \text{ and } \lambda_2 \perp \mu.$$
We can also decompose $X = E_1 \sqcup E_2$ with $\lambda_2(E_1) = 0$ and $\lambda_1(E_2) = 0$. The next claim is that λ_1 and λ_2 are T-invariant. This follows, since
$$\lambda_1 + \lambda_2 = \lambda = \lambda \circ T^{-1} = \lambda_1 \circ T^{-1} + \lambda_2 \circ T^{-1}.$$
Notice that
$$\lambda_1 \circ T^{-1} \ll \mu \circ T^{-1} = \mu,$$
$$\lambda_2 \circ T^{-1} \perp \mu \circ T^{-1} \implies \lambda_2 \circ T^{-1} \perp \mu.$$
By the uniqueness of Lebesgue-Radon-Nikdoym, we have
$$\lambda_1 = \lambda_1 \circ T^{-1}, \qquad \lambda_2 = \lambda_2 \circ T^{-1}.$$
Thus λ_1 and λ_2 are T-invariant. Notice that
$$E_1 \sqcup E_2 = X = T^{-1}(X) = T^{-1}(E_1) \sqcup T^{-1}(E_2).$$
By the T-invariance of λ_1 and λ_2, we see that
$$\lambda_2(T^{-1}(E_1)) = \lambda_2(E_1) = 0, \qquad \lambda_1(T^{-1}(E_2)) = \lambda_1(E_2) = 0.$$
By the uniqueness of Lebesgue-Radon-Nikodym, we have that $E_1 = T^{-1}(E_1)$ and $E_2 = T^{-1}(E_2)$. Since T is ergodic, one of these must have measure zero and the other must have full measure with respect to λ. If $\lambda(E_1) = 1$ we have a contradiction, since this would imply $\lambda = \lambda_1 \ll \mu$ and (1) tells us that $\lambda = \mu$. Thus we must have $\lambda(E_2) = 1$ so that $\lambda = \lambda_2 \perp \mu$.

□

Recall that a map T is weakly mixing if, for all $A, B \in \mathcal{M}$, we have

$$\frac{1}{n}\sum_{j=0}^{n-1} |\mu(T^{-j}(A) \cap B) - \mu(A)\mu(B)| \to 0.$$

Problem 46 (Petersen 2.6.2)**.** Suppose T is weakly mixing.

(1) Show that $S = T^m$, $m \geq 1$, is weakly mixing.
(2) Show that S defined so that $S^m = T$, $m \geq 1$, is weakly mixing.

Proof. Recall one of the equivalences for weakly mixing. That is, T is weakly mixing iff there exists $J \subseteq \mathbb{Z}_{\geq 0}$ of density zero so that

$$\lim_{\substack{n\to\infty \\ n\notin J}} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B).$$

Recall as well that T is weak mixing iff T has no measurable eigenfunctions other than the constants.

(1) Fix m. We claim that $J_1 = \{n \in \mathbb{N} : mn \in J\}$ has density zero. To see this, notice that for fixed n we have

$$\frac{|J_1 \cap \{0,\dots,n-1\}|}{n} = \frac{|J \cap \{0,\dots,m(n-1)\}|}{n}$$

$$\frac{|J \cap \{0,\dots,m(n-1)\}|}{n} \cdot \frac{mn}{mn} = \frac{|J \cap \{0,\dots,m(n-1)\}|}{mn} \cdot \frac{mn}{n} = m \cdot \frac{|J \cap \{0,\dots,m(n-1)\}|}{mn}.$$

Take the limit as $n \to \infty$ to get

$$\overline{d}(J_1) = m \cdot \overline{d}(J).$$

Since $\overline{d}(J) = 0$, this tells us that $\overline{d}(J_1) = 0$. Moreover,

$$\lim_{\substack{n\to\infty \\ n\notin J_1}} \mu(T^{-mn}(A) \cap B) = \lim_{\substack{n\to\infty \\ n\notin J_1}} \mu(S^{-n}(A) \cap B) = \mu(A)\mu(B).$$

This gives us that $S = T^m$ is weak mixing for all $m \geq 1$.

(2) Now suppose that $S^m = T$, where $m \geq 1$ is fixed. Suppose that S had a measurable eigenfunction which is not constant, say f. Then $Sf = \lambda f$ for some $|\lambda| = 1$, $\lambda \neq 1$. Notice that $S^m f = Tf = \lambda^m f$, with $|\lambda^m| = 1$, $\lambda \neq 1$. Therefore we have that T is not weak mixing. The contrapositive gives us the result.

□

Problem 47 (Petersen 2.6.4)**.** There are examples of weakly mixing measure preserving transformations that are not strongly mixing. For now, consider some easier counterexamples.

(1) Find an example of a sequence $\{a_n\}$ so that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} |a_k| = 0, \qquad \lim_{n\to\infty} a_n \neq 0.$$

(2) A sequence $\{A_j\}$ of measurable sets, each having measure α, is called strongly mixing if

$$\lim_{n\to\infty} \mu(A_n \cap B) = \alpha\mu(B) \text{ for all } B \in \mathcal{M}.$$

It is called weakly mixing if

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} |\mu(A_j \cap B) - \alpha\mu(B)| = 0.$$

Find an example which is weakly mixing but not strongly mixing.

(3) A sequence $\{A_j\}$ as above is called mixing of order k if

$$\lim_{\substack{\inf_i n_i \to \infty \\ \inf_{i \neq j} |n_i - n_j| \to \infty}} \mu(A_{n_1} \cap \cdots \cap A_{n_k} \cap B) = \alpha^k \mu(B) \text{ for all } B \in \mathcal{M}.$$

Give an example of a sequence that is mixing of order 1 but not of order 2.

Proof.

(1) Let $J \subseteq \mathbb{Z}_{\geq 0}$ be an infinite set of density zero (for example, we could say J is the set of primes, see here). Define a sequence $\{a_k\}$ where

$$a_k = \begin{cases} 1 \text{ if } k \in J \\ 0 \text{ otherwise.} \end{cases}$$

Then we see that the limit of a_k does not exist, but the limit of the series tends to 0.

(2) Presumably the idea is to use the last part to prove this part. The idea (maybe) is to take A_j distributed around your space so that

$$\mu(A_k \cap B) = \begin{cases} \text{Something not } \alpha\mu(B) \text{ if } k \in J \\ \alpha\mu(B) \text{ otherwise.} \end{cases}$$

Then by the same argument as the last part, we get that this will come out to give us weak mixing but not strong mixing. The question is whether we can place these sets so that this is true.

(3) TODO

□

Remark. Thomas O'Hare was a collaborator.

Problem 48 (Petersen 5.2.4)**.** Show the following.

(1) We have
$$h(\alpha, T) = \lim_{n\to\infty} H\left(T^{-n}(\alpha)\middle|\bigvee_{k=0}^{n-1} T^{-k}(\alpha)\right).$$

(2) We have
$$h(\alpha, T) = \lim_{n\to\infty} H\left(\alpha\middle|\bigvee_{k=1}^{n-1} T^{k}(\alpha)\right) = H\left(\alpha\middle|\bigvee_{k=1}^{\infty} T^{k}(\alpha)\right).$$

(3) We have
$$h(\alpha, T) = h(\alpha, T^{-1})$$

Proof by Suxuan.

(1) This is the same kind of trick. Notice that
$$H\left(T^{-n}(\alpha)\middle|\bigvee_{k=0}^{n-1} T^{-k}(\alpha)\right) = H\left(\bigvee_{k=0}^{n} T^{-k}(\alpha)\right) - H\left(\bigvee_{k=0}^{n-1} T^{-k}(\alpha)\right).$$
Now $H(T^{-1}(\alpha)) = H(\alpha)$ for any α, so we have
$$H\left(\bigvee_{k=0}^{n-1} T^{-k}(\alpha)\right) = H\left(\bigvee_{k=1}^{n} T^{-k}(\alpha)\right),$$
thus
$$H\left(T^{-n}(\alpha)\middle|\bigvee_{k=0}^{n-1} T^{-k}(\alpha)\right) = H\left(\alpha\middle|\bigvee_{k=1}^{n} T^{-k}(\alpha)\right).$$
Take the limit.

(2) Same kind of trick as in (1). Notice
$$H\left(\alpha\middle|\bigvee_{k=1}^{n-1} T^{-k}(\alpha)\right) = H\left(T^{-(n-1)}(\alpha)\middle|\bigvee_{k=0}^{n-2} T^{-k}(\alpha)\right).$$
Take the limit.

(3) Notice by (2) we have
$$h(\alpha, T) = \lim_{n\to\infty} H\left(\alpha\middle|\bigvee_{k=1}^{n} T^{-k}(\alpha)\right) = \lim_{n\to\infty} H\left(\alpha\middle|\bigvee_{k=1}^{n} T^{k}(\alpha)\right) = h(\alpha, T^{-1}).$$

□

Problem 49 (Petersen 5.2.6)**.** Show that $\alpha \le \beta$ implies $h(\alpha, T) \le h(\beta, T)$.

Proof. Examine **Petersen Proposition 5.2.13**. This says that for any countable partitions, we have
$$h(\alpha, T) \le h(\beta, T) + H(\alpha|\beta).$$
Since $\alpha \le \beta$, we can use **Petersen Proposition 5.2.7** to get $H(\alpha|\beta) = 0$.

Alternatively it follows by definition. Since $\alpha \leq \beta$, we know that $H(\alpha) \leq H(\beta)$. Notice that $T^{-1}(\alpha) \leq T^{-1}(\beta)$ since the preimage plays nicely with unions, and by induction this keeps holding. Thus $\bigvee_{k=0}^{n-1} T^{-k}(\alpha) \leq \bigvee_{k=0}^{n-1} T^{-k}(\beta)$. Finally we see that

$$H\left(\bigvee_{k=0}^{n-1} T^{-k}(\alpha)\right) \leq H\left(\bigvee_{k=0}^{n-1} T^{-k}(\beta)\right).$$

This holds for all n, so dividing by n and taking limits gives the result. □

Problem 50 (Petersen 5.2.7)**.** Show that

$$h\left(\bigvee_{k=n}^{m} T^{-k}\alpha, T\right) = h(\alpha, T).$$

Proof. We follow **Walters Theorem 4.12**. Note

$$\begin{aligned} h\left(\bigvee_{k=n}^{m} T^{-k}\alpha, T\right) &= \lim_{r\to\infty} \frac{1}{r} H\left(\bigvee_{k=0}^{r-1} T^{-k}\left(\bigvee_{i=n}^{m} T^{-i}(\alpha)\right)\right) \\ = \lim_{r\to\infty} \frac{1}{r} H\left(\bigvee_{i=n}^{m+r-1} T^{-i}(\alpha)\right) &= \lim_{r\to\infty} \left(\frac{m+r-1}{r}\right) \frac{1}{m+r-1} H\left(\bigvee_{i=n}^{m+r-1} T^{-i}(\alpha)\right) \\ &= h(T^{-n}(\alpha), T). \end{aligned}$$

So this boils down to showing that $h(T^{-n}(\alpha), T) = h(\alpha, T)$. By an induction argument, it suffices to show that $h(T^{-1}(\alpha), T) = h(\alpha, T)$. This follows, since

$$\begin{aligned} h(T^{-1}(\alpha), T) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=1}^{n} T^{-k}(\alpha)\right) &= \lim_{n\to\infty} \frac{1}{n} H\left(T^{-1}\left(\bigvee_{k=0}^{n-1} T^{-k}(\alpha)\right)\right) \\ &= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^{-k}(\alpha)\right) = h(\alpha, T). \end{aligned}$$

□

The **entropy** of the transformation T is defined as

$$h(T) = \sup_{\alpha} h(\alpha, T).$$

This gives a numeric value to the average uncertainty of where T moves points with respect to a partition α.

Problem 51 (Petersen 5.2.8)**.** Show that

$$h(T^k) = |k| h(T).$$

Proof. Assume $k > 0$. Let α denote finite partitions. We see that

$$h(T^k) \geq \sup_{\alpha} h(\alpha, T^k).$$

In particular, by the last problem we see that

$$\begin{aligned} h(\alpha, T^k) = h\left(\bigvee_{m=0}^{k-1} T^{-m}(\alpha), T^k\right) &= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-ik}\left(\bigvee_{j=0}^{k-1} T^{-j}(\alpha)\right)\right) \\ = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{kn-1} T^{-j}(\alpha)\right) &= \lim_{n\to\infty} \frac{k}{kn} H\left(\bigvee_{j=0}^{kn-1} T^{-j}(\alpha)\right) = kh(\alpha, T). \end{aligned}$$

So for every finite partition we have the result, and thus

$$\sup_{\alpha} kh(\alpha, T) \leq h(T^k).$$

Now for any partition α we see that

$$h(\alpha, T^k) \leq h\left(\bigvee_{j=0}^{k-1} T^{-j}(\alpha), T^k\right) = kh(\alpha, T^k),$$

since $\alpha \leq \bigvee_{j=0}^{k-1} T^{-j}(\alpha)$. For negative k, we simply use $h(T^{-1}) = h(T)$. We know this holds since $h(\alpha, T^{-1}) = h(\alpha, T)$ for all α. □

Problem 52 (Petersen 5.2.9)**.** Show that $I(T^{-1}\alpha) = I(\alpha) \circ T$.

Proof. We recall

$$I(\alpha) = -\sum_{A\in\alpha} \log(\mu(A))\chi_A(x).$$

Notice that

$$I(\alpha) \circ T = -\sum_{A\in\alpha} \log(\mu(A))\chi_A(T(x)) = -\sum_{A\in\alpha} \log(\mu(A))\chi_{T^{-1}(A)}(x).$$

Assuming the transformation is measure preserving, we have

$$\log(\mu(T^{-1}(A))) = \log(\mu(A)),$$

so

$$I(\alpha) \circ T = -\sum_{A\in T^{-1}(\alpha)} \log(\mu(A))\chi_A(x).$$

□

Whenever not specified, assume that $T : X \to X$ is a measure preserving transformation of a Borel probability measure space $(X, \mathcal{B}, \mu)$.

Problem 53. Consider the two-to-one map

$$T(x) = \frac{1}{2}(x - 1/x), \quad T(0) = 0$$

on $\mathbb{R}$.

(1) Show that T preserves the measure $dx/(1+x^2)$.
(2) Show that the change of variables $x = \tan(t)$ carries T to a Lebesgue measure preserving map S of $(-\pi/2, \pi/2)$.
(3) Show that S is isomorphic to the one-sided Bernoulli shift $\mathcal{B}(1/2, 1/2)$; here we normalize the Lebesgue measure on $(-\pi/2, \pi/2)$.

Proof.

(1) The half-open intervals generate the Borel σ-algebra, so by **Walters Theorem 1.1** it suffices to check that T is measure preserving on them. Let $I := [a, b)$ be an interval. Let μ be the measure generated by $dx/(1+x^2)$; i.e.

$$\mu(E) = \int_{\mathbb{R}} \chi_E(x) \frac{dx}{1+x^2}.$$

The goal is to show

$$\arctan(b) - \arctan(a) = \int_a^b \frac{dx}{1+x^2} = \mu(I) = \mu(T^{-1}(I)).$$

Fix $z \neq 0$ in $\mathbb{R}$. We solve the equation

$$T(x) = z \implies x - \frac{1}{x} = 2z \implies x^2 - 2xz - 1 = 0.$$

Solving, we have solutions given by

$$x = z \pm \sqrt{z^2 + 1}.$$

Assume for now we have an interval I such that $0 \notin I$ and $b \neq 0$; we eliminate these cases since we need to deal with the fact that the preimage has three values at zero. For such an interval I, we have

$$\mu(T^{-1}(I)) = \int \chi_{T^{-1}(I)}(x) \frac{dx}{1+x^2}.$$

Using the above analysis, we see that $T^{-1}(I) = [a_1, b_1) \sqcup [a_2, b_2)$, where $[a_1, b_1) \subseteq (-\infty, 0)$ and $[a_2, b_2) \subseteq (0, \infty)$. Explicitly, we have $a_1 = a - \sqrt{a^2+1}$, $b_1 = b - \sqrt{b^2+1}$, $a_2 = a + \sqrt{a^2+1}$, $b_2 = b + \sqrt{b^2+1}$. We can rewrite the above integral as

$$\mu(T^{-1}(I)) = \int_{a-\sqrt{a^2+1}}^{b-\sqrt{b^2+1}} \frac{dx}{1+x^2} + \int_{a+\sqrt{a^2+1}}^{b+\sqrt{b^2+1}} \frac{dx}{1+x^2}$$

$$= \arctan(b - \sqrt{b^2+1}) - \arctan(a - \sqrt{a^2+1}) + \arctan(b + \sqrt{b^2+1}) - \arctan(a + \sqrt{a^2+1}).$$

Recall the following trig identity:

$$\arctan(x) + \arctan(y) = \arctan\left(\frac{x+y}{1-xy}\right) \text{ if } xy < 1.$$

Notice that for all $z \in \mathbb{R}$ we have

$$(z - \sqrt{z^2+1})(z + \sqrt{z^2+1}) = z^2 - z^2 - 1 = -1 < 1.$$

Hence we can apply the identity for all $z \in \mathbb{R}$. Doing so, we see

$$\arctan(z - \sqrt{z^2+1}) + \arctan(z + \sqrt{z^2+1}) = \arctan\left(\frac{z - \sqrt{z^2+1} + z + \sqrt{z^2+1}}{1 - (z - \sqrt{z^2+1})(z + \sqrt{z^2+1})}\right)$$
$$= \arctan\left(\frac{2z}{2}\right) = \arctan(z).$$

Thus substituting in a and b above for z, we get that

$$\mu(T^{-1}(I)) = \arctan(b) - \arctan(a) = \mu(I).$$

We now deal with the cases involving zero. Suppose $a = 0$. Then $T^{-1}(I) = [a_1, b_1) \sqcup [a_2, b_2) \sqcup \{0\}$, where a_i and b_i defined as before. We notice that 0 doesn't contribute anything, so the argument still works the same. The same kind of argument applies if $b = 0$, and if $0 \in I$ we can break up $I = [a, 0) \sqcup [0, b)$ and apply the prior arguments to each case there. Thus T is measure preserving on all half-open intervals and the result follows.

(2) Define $S : (-\pi/2, \pi/2) \to (-\pi/2, \pi/2)$ by

$$S(t) := \arctan\left(\frac{1}{2}\left(\tan(t) - \frac{1}{\tan(t)}\right)\right).$$

The goal is to show that this is a Lebesgue measure preserving transformation. If we can show that $\tan(x) : (-\pi/2, \pi/2) \to \mathbb{R}$ is measure preserving with respect the Lebesgue measure on $(-\pi/2, \pi/2)$ and the measure μ defined above on $\mathbb{R}$, and if we can show that $\arctan(x) : \mathbb{R} \to (-\pi/2, \pi/2)$ is measure preserving with respect to Lebesgue measure, then we get that S is measure preserving (see **Walters Remark (2)** on **Page 19**, though the result is an easy calculation). Take an interval $I := [a, b) \subseteq \mathbb{R}$. Then the first step is to show that

$$\lambda(\tan^{-1}(I)) = \mu(I).$$

In other words,

$$\int_{(-\pi/2,\pi/2)} \chi_I(\tan(x))dx = \int_{\mathbb{R}} \chi_I(y)\frac{dy}{1+y^2}.$$

This is just a substitution though – let $y = \tan(x)$, $dy = dx(1+y^2)$. Then

$$\int \chi_I(\tan(x))dx = \int \chi_I(y)\frac{dy}{1+y^2}$$

as desired.

To see that $\arctan : \mathbb{R} \to (-\pi/2, \pi/2)$ is measure preserving, let $I := [a, b) \subseteq (-\pi/2, \pi/2)$. Again, we just need to check

$$\int \chi_I(\arctan(x))\frac{dx}{1+x^2} = \int \chi_I(y)dy.$$

If we let $y = \arctan(x)$ then $dy = dx/(1+x^2)$ and we get the above result. This shows that S is a composition of measure preserving transformations, hence measure preserving.

(3) We consider normalized Lebesgue measure on S (that is, if λ denotes Lebesgue measure, we take the measure ν defined by $\nu(E) := \lambda(E)/\pi$). Consider $X = (0, 1) \subseteq [0, 1)$ – this set has full measure since we've just removed a point. Take

$$\varphi : X \to (-\pi/2, \pi/2), \qquad \varphi(x) := \pi x - \frac{\pi}{2}.$$

This map is invertible with inverse given by

$$\varphi^{-1} : (-\pi/2, \pi/2) \to X, \qquad \varphi^{-1}(x) := \frac{x}{\pi} + \frac{1}{2}.$$

We check that the map is measure preserving. Take $I \subseteq (-\pi/2, \pi/2)$. Then

$$\lambda(\varphi^{-1}(I)) = \int_0^1 \chi_I(\varphi(x))dx = \int_{-\pi/2}^{\pi/2} \chi_I(y)\frac{dy}{\pi} = \frac{\lambda(I)}{\pi} = \nu(I).$$

We see φ^{-1} is also measure preserving; taking $I \subseteq (0, 1)$, we have

$$\nu(\varphi(I)) = \int_{-\pi/2}^{\pi/2} \chi_I(\varphi^{-1}(x))\frac{dx}{\pi} = \int_0^1 \chi_I(y)dy = \lambda(I).$$

Now if $T_2 : [0, 1) \to [0, 1)$ is the doubling map defined by

$$T_2(x) \equiv 2x \pmod 1,$$

then we claim that $((-\pi/2, \pi/2), \nu, S)$ is isomorphic to $([0, 1), \lambda, T_2)$. This amounts to showing that $\varphi^{-1} \circ S = T_2 \circ \varphi^{-1}$ almost everywhere. Notice that (using a little bit of precalculus)

$$\begin{aligned}
\varphi^{-1}(S(x)) &= \frac{S(x)}{\pi} + \frac{1}{2} \\
&= \frac{\arctan\left(\frac{\tan(x)^2-1}{2\tan(x)}\right)}{\pi} + \frac{1}{2} \\
&= \frac{\arctan(-\cot(2x))}{\pi} + \frac{1}{2} \\
&= \begin{cases} 2x/\pi + 1 \text{ if } -\pi/2 < x < 0 \\ 2x/\pi \text{ if } 0 < x < \pi/2, \end{cases}
\end{aligned}$$

and

$$\begin{aligned}
T_2(\varphi^{-1}(x)) &= \begin{cases} 2\varphi^{-1}(x) \text{ if } 0 < \varphi^{-1}(x) < 1/2 \\ 2\varphi^{-1}(x) - 1 \text{ if } 1/2 \le \varphi^{-1}(x) < 1 \end{cases} \\
&= \begin{cases} 2x/\pi + 1 \text{ if } -\pi/2 < x < 0 \\ 2x/\pi \text{ if } 0 < x < \pi/2. \end{cases}
\end{aligned}$$

Thus these are equal almost everywhere, and so this is an isomorphism between T_2 and S. We can then use the isomorphism between σ the one-sided left shift on $\mathcal{B}(1/2, 1/2)$ and T_2 (see **Walters (2)** near the top of **Page 58**) to get the isomorphism from S to σ (see **Walters Remark (1)** on **Page 58** – here we use the transitivity of an equivalence relation).

□

Problem 54. Let $X = [0,1]$ with Lebesgue measure m. Then T (measure preserving) is ergodic on X iff

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) \to \int f dm \text{ a.e.}$$

for each continuous function f.

Proof. ($\Longrightarrow$): This follows by **Petersen Theorem 4.4**, since continuous functions are measurable. The argument is as follows: assume T is ergodic. Then from **Petersen Theorem 2.2.3 (1)** we know that for continuous f we have

$$\overline{f}(x) := \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)$$

exists almost everywhere. Since $\overline{f}$ is T-invariant by **Petersen Theorem 2.2.3 (2)**, we get that $\overline{f}$ is constant almost everywhere by **Petersen Proposition 2.4.1**. Set $\overline{f}(x) = C \in \mathbb{R}$ almost everywhere. Using **Petersen Theorem 2.2.3 (4)**, we have

$$\int_X f d\lambda = \int_X \overline{f} d\lambda = C\lambda([0,1]) = C.$$

This gives us the result.
($\Longleftarrow$): Assume we have it for all f continuous. Let $g \in L^1$. Notice that for arbitrary continuous f we have

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=0}^{n-1} g(T^k(x)) - \int g dm\right| &\le \left|\frac{1}{n}\sum_{k=0}^{n-1}[g(T^k(x)) - f(T^k(x))]\right| + \left|\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)) - \int fm\right| + \left|\int (f-g)dm\right| \\ &\le \left|\frac{1}{n}\sum_{k=0}^{n-1}[g(T^k(x)) - f(T^k(x))]\right| + \left|\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)) - \int f dm\right| + \|f-g\|_1.\end{aligned}$$

Fix $\epsilon > 0$. By **Folland Proposition 7.9** (i.e. density of continuous functions in L^1), we can choose f so that $\|f-g\|_1 < \epsilon/3$. Integrating both sides of the inequality, this leaves us with

$$\begin{aligned}\int \left|\frac{1}{n}\sum_{k=0}^{n-1} g(T^k(x)) - \int g dm\right| dm \le& \int \left|\frac{1}{n}\sum_{k=0}^{n-1}[g(T^k(x)) - f(T^k(x))]\right| dm \\ &+ \int \left|\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)) - \int f dm\right| dm + \epsilon/3.\end{aligned}$$

Since f continuous on a compact domain, we get that it's bounded. Defining

$$f_n = \frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)),$$

we see that each $f_n \in L^1$ since $f \in L^1$, and $|f_n| \le \sup_{x\in[0,1]} |f(x)|$. These observations tell us that the dominated convergence theorem (**Folland Theorem 2.24**) applies. Using the assumption we get

$$\int \left|\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)) - \int f dm\right| dm \to 0.$$

Thus we can choose N sufficiently large so that for $n \geq N$ we have

$$\int \left| \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) - \int f dm \right| dm < \epsilon/3.$$

Now using linearity of the integral and the triangle inequality, we get an upper bound

$$\int \left| \frac{1}{n} \sum_{k=0}^{n-1} g(T^k(x)) - \int g dm \right| dm \leq \frac{1}{n} \sum_{k=0}^{n-1} \int |g(T^k(x)) - f(T^k(x))| dm + 2\epsilon/3.$$

Applying a change of variables and using the fact that T is measure preserving, we then get

$$\begin{aligned} \int \left| \frac{1}{n} \sum_{k=0}^{n-1} g(T^k(x)) - \int g dm \right| dm &\leq \frac{1}{n} \sum_{k=0}^{n-1} \|g - f\|_1 + 2\epsilon/3 \\ &< \frac{1}{n} \sum_{k=0}^{n-1} (\epsilon/3) + 2\epsilon/3 = \epsilon. \end{aligned}$$

The choice of $\epsilon > 0$ was arbitrary, so this implies that $\frac{1}{n} \sum_{k=0}^{n-1} g(T^k(x)) \to \int g dm$ in L^1. Recall convergence in L^1 implies there is a subsequence along which it converges almost everywhere. Since we know that $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(T^k(x)) =: \overline{g}(x)$ exists almost everywhere by **Petersen Theorem 2.2.3 (1)**, we get that we must have $\overline{g} = \int g dm$ almost everywhere for every L^1 function g. We can now apply **Petersen Theorem 2.4.4** to finish the result. $\square$

Problem 55. Suppose that T is ergodic and $f \geq 0$ is measurable. Prove that if

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) < \infty \text{ a.e.}$$

then $f \in L^1$.

Proof. Let

$$f^*(x) := \limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)).$$

We claim that f^* is T-invariant. This follows from the first paragraph of the proof of **Walters Theorem 1.14**. The idea is to let

$$a_n(x) := \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)).$$

Then we have that

$$\left(\frac{n+1}{n}\right) a_{n+1}(x) - a_n(T(x)) = \frac{f(x)}{n},$$

$$\limsup_{n\to\infty} \frac{f(x)}{n} = 0,$$

and

$$\limsup_{n\to\infty} \left(\left(\frac{n+1}{n}\right) a_{n+1}(x) - a_n(T(x)) \right) = f^*(x) - f^*(T(x)).$$

Since T is ergodic, this implies that f^* is constant almost everywhere. Since $f \geq 0$, this implies that $0 \leq f^* = C < \infty$. Let

$$f_k := f\chi_{f\leq k} + k\chi_{f>k}.$$

This is a bounded function, so $f_k \in L^1$. Moreover $f_k \nearrow f$, so we get that

$$\frac{1}{n} \sum_{j=0}^{n-1} f_k(T^j(x)) \leq \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) \text{ for all } n.$$

Let

$$f_k^*(x) := \limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f_k(T^j(x)).$$

The above observation tells us that $f_k^* \leq f^* = C$ almost everywhere. **Petersen Theorem 2.2.3 (4)** tells us that

$$\int_X f_k d\mu = \int_X f_k^* d\mu \leq C\mu(X) \text{ for all } k.$$

We can now apply the monotone convergence theorem (**Folland Theorem 2.14**) to get

$$\int f d\mu = \lim_{k\to\infty} \int f_k d\mu \leq C\mu(X) < \infty.$$

Thus we have $f \in L^1$. □

Problem 56. For $m = 1, 2, \ldots$, prove that T is weakly mixing iff T^m is weakly mixing.

Remark. We may assume T is nice enough so that the criteria applies.

Proof. Recall **Petersen Theorem 2.6.1**. We have two equivalences for weakly mixing. First, T is weakly mixing iff for every measurable A, B there exists $J \subseteq \mathbb{Z}_{\geq 0}$ of density zero so that

$$\lim_{\substack{n\to\infty \\ n\notin J}} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B).$$

Second, T is weakly mixing iff T has no measurable eigenfunctions other than the constants.
($\implies$): Fix m and let $J \subseteq \mathbb{Z}_{\geq 0}$ be a set of density zero. We claim that $J_1 = \{n \in \mathbb{N} : mn \in J\}$ has density zero. To see this, notice that for fixed n we have

$$\frac{|J_1 \cap \{0, \ldots, n-1\}|}{n} = \frac{|J \cap \{0, \ldots, m(n-1)\}|}{n}$$

$$\frac{|J \cap \{0, \ldots, m(n-1)\}|}{n} \cdot \frac{mn}{mn} = \frac{|J \cap \{0, \ldots, m(n-1)\}|}{mn} \cdot \frac{mn}{n} = m \cdot \frac{|J \cap \{0, \ldots, m(n-1)\}|}{mn}.$$

Take the limit as $n \to \infty$ to get

$$\overline{d}(J_1) = m \cdot \overline{d}(J).$$

Since $\overline{d}(J) = 0$, this tells us that $\overline{d}(J_1) = 0$. Now, take A, B measurable. Since T is weakly mixing, we have that there exists a J with density zero so that

$$\lim_{\substack{n\to\infty \\ n\notin J}} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B).$$

By the prior observation, we have J_1 has density zero, and if we set $S := T^m$ we see that

$$\lim_{\substack{n\to\infty \\ n\notin J_1}} \mu(T^{-mn}(A) \cap B) = \lim_{\substack{n\to\infty \\ n\notin J_1}} \mu(S^{-n}(A) \cap B) = \mu(A)\mu(B).$$

This gives us that $S := T^m$ is weak mixing for any $m \geq 1$.
($\impliedby$): The goal is to prove that for fixed $m \geq 1$, T^m weakly mixing implies T is weakly mixing. We proceed by contrapositive; namely we will show that T not weakly mixing implies that T^m is not weakly mixing. Since T is not weakly mixing, we have a measurable eigenfunction f which is nonconstant, so $U_T(f) = \lambda f$, $\lambda \neq 0$. But then we see that $U_{T^m}(f) = (U_T)^m(f) = \lambda^m f$, so f is also a non-constant eigenfunction for T^m. By our equivalence, this forces T^m to not be weakly mixing, thus proving the contrapositive. $\square$

Problem 57. Show that

$$h(T) = \sup\{h(\alpha, T) : \alpha \text{ is a countable measurable partition with } H(\alpha) < \infty\}.$$

Remark. Recall that

$$h(T) := \sup\{h(\alpha, T) : \alpha \text{ is a finite partition}\}.$$

Proof. If α is a finite partition, say $\alpha = \{A_1, \ldots, A_n\}$, then

$$H(\alpha) = -\sum_{i=1}^{n} \mu(A_i)\log_2(\mu(A_i)).$$

It's a finite sum of finite things, so we get that $H(\alpha) < \infty$. We can make a finite partition countable by adding empty sets. Let $\beta = \{A_1, \ldots, A_n, B_1, \ldots\}$ be a countable partition, where $B_i = \varnothing$ for all $i \geq 1$. Then

$$H(\beta) = -\sum_{i=1}^{n} \mu(A_i)\log_2(\mu(A_i)) - \sum_{i=1}^{\infty} \mu(B_i)\log_2(\mu(B_i)) = -\sum_{i=1}^{n} \mu(A_i)\log_2(\mu(A_i)) = H(\alpha) < \infty.$$

We now claim that $h(\alpha, T) = h(\beta, T)$. Recall

$$h(\alpha, T) := \lim_{n\to\infty} H(\alpha \vee T^{-1}\alpha \vee \cdots \vee T^{-n+1}\alpha)/n.$$

It follows readily from the above calculation and the definition of join that for each n we have

$$H(\alpha \vee T^{-1}\alpha \vee \cdots \vee T^{-n+1}\alpha) = H(\beta \vee T^{-1}\beta \vee \cdots \vee T^{-n+1}\beta).$$

Hence the result follows. So we can view the set of finite partitions as a subset of the set of countable partitions with finite entropy, and this tells us that

$$\begin{aligned} h(T) &:= \sup\{h(\alpha, T) : \alpha \text{ is a finite partition}\} \\ &\leq \sup\{h(\alpha, T) : \alpha \text{ is a countable measurable partition with } H(\alpha) < \infty\}. \end{aligned}$$

We now need to show the other inequality. Let α be a countable partition such that $H(\alpha) < \infty$. Write $\alpha = \{A_1, \ldots\}$. Let $\alpha_n = \{A_1, \ldots, A_{n-1}, B_n\}$, where

$$B_n = \bigcup_{j=n}^{\infty} A_j.$$

Now by **Petersen Proposition 5.2.13** we have

$$h(\alpha, T) \leq h(\alpha_n, T) + H(\alpha|\alpha_n) \text{ for all n.}$$

By construction, $\alpha_1 \leq \alpha_2 \leq \cdots$ and $\alpha_\infty = \bigvee_{n=1}^{\infty} \alpha_n = \alpha$. Using **Petersen Proposition 5.2.7** and **Petersen Poposition 5.2.11**[1] we get that

$$h(\alpha, T) \leq \lim_{n\to\infty} h(\alpha_n, T) + H(\alpha|\alpha) = \lim_{n\to\infty} h(\alpha_n, T) \leq \sup\{h(\alpha, T) : \alpha \text{ is a finite partition}\}.$$

This holds for all α countable measurable partitions with $H(\alpha) < \infty$, so by supremum properties we get

$$\begin{aligned} &\sup\{h(\alpha, T) : \alpha \text{ is a countable measurable partition with } H(\alpha) < \infty\} \\ &\qquad \leq h(T) := \sup\{h(\alpha, T) : \alpha \text{ is a finite partition}\}. \end{aligned}$$

Hence we have equality. □

[1] Slight caveat here – technically the proposition only works for finite partitions, but the remark right after **Petersen Proposition 5.2.12** points out that it still works if we have a countable partition with finite entropy using **Petersen Corollary 6.2.2**.

Problem 58. Use the Shannon-McMillan-Breiman Theorem to compute the entropy of an ergodic Markov shift.

Proof. Let $A = (a_{ij})$ be an $n \times n$ stochastic matrix with fixed row probability vector p (i.e. we have $pA = p$). Assume our elements are given by $\{1, \ldots, n\}$ without loss of generality. Let α be the partition given by the time 0 cylinders, so

$$\alpha := \{\{(x_n) : x_0 = i\} : 1 \leq i \leq n\}.$$

This partition is a generator (see **Petersen Example 5.3.4**, although the calculation is easy). Thus the Kolmogorov-Sinai theorem (**Petersen Theorem 5.3.1**) tells us that $h(\alpha, \sigma) = h(\sigma)$. We now use the Shannon-McMillan-Breiman theorem (**Petersen Theorem 6.2.3**) to calculate $h(\alpha, \sigma)$.

For $1 \leq i, j \leq n$ let

$$C_{i,j} := \{(x_n) : x_0 = i, x_1 = j\}.$$

For a fixed sequence (x_n) let $k_{i,j,m}((x_n))$ be the number of occurrences of a_i followed by a_j in the sequence $\{x_0, \ldots, x_m\}$. That is, we have

$$k_{i,j,m}((x_n)) := \sum_{k=0}^{m-1} \chi_{C_{i,j}}(\sigma^{-k}((x_n))).$$

Notice that we can write the information function of the sequence (x_n) as

$$I_{\alpha_0^m}((x_n)) = -\log_2 \left[\prod_{i=1}^{n} p_i^{\chi_{C_i}((x_n))} \cdot \prod_{i,j=1}^{n} a_{ij}^{k_{i,j,m}((x_n))} \right]$$

$$= -\left[\sum_{i=1}^{n} \chi_{C_i}((x_n)) \log_2(p_i) + \sum_{i,j=1}^{n} k_{i,j,m}((x_n)) \log_2(a_{ij}) \right].$$

Dividing by $m+1$ and taking the limit, we see that the sum on the left vanishes. It suffices to then look at

$$\lim_{m\to\infty} \frac{k_{i,j,m}((x_n))}{m+1}.$$

Since σ is ergodic, **Petersen Theorem 2.4.4** tells us that

$$\lim_{m\to\infty} \frac{1}{m} k_{i,j,m} = \lim_{m\to\infty} \frac{1}{m} \sum_{k=0}^{m-1} \chi_{C_{i,j}} \circ \sigma^{-k} = \int_X \chi_{C_{i,j}} d\mu = \mu(C_{i,j}) = p_i a_{i,j} \text{ almost everywhere.}$$

Notice now that

$$\lim_{m\to\infty} \frac{k_{i,j,m}((x_n))}{m+1} = \lim_{m\to\infty} \left(\frac{m}{m+1}\right) \left(\frac{k_{i,j,m}((x_n))}{m}\right) = p_i a_{i,j}.$$

Therefore

$$\frac{1}{m+1} I_{\alpha_0^m} \to -\sum_{i,j=1}^{n} p_i a_{i,j} \log_2(a_{i,j}) \text{ almost everywhere as } m \to \infty.$$

Putting it all together, we see that

$$h(\sigma) = -\sum_{i,j=1}^{n} p_i a_{i,j} \log_2(a_{i,j}).$$

Notice this matches the calculation given in **Petersen Example 5.3.5**. ☐

Problem 59. Let $T : X \to X$ is a homeomorphism of a compact metric space, as is $S : Y \to Y$. Let $\varphi : X \to Y$ be a continuous surjective map such that $S \circ \varphi = \varphi \circ T$. Prove that $h_{\text{Top}}(T) \geq h_{\text{Top}}(S)$ by using

(1) the Adler-Konheim-McAndrew definition;
(2) the Bowen definition.

Remark. In the above scenario, we say that the system (X, T) is a **continuous extension** of the system (Y, S).

Proof. First notice that the compatibility condition (i.e. the condition that $S \circ \varphi = \varphi \circ T$) applies for iterates, meaning

$$S^k \circ \varphi = S^{k-1} \circ S \circ \varphi = S^{k-1} \circ \varphi \circ T = \cdots = \varphi \circ T^k.$$

We use this fact in both proofs.

(1) We follow **Walters Theorem 7.2**. Recall that

$$h_{\text{Top}}(T) := \sup\{h(\mathcal{U}, T) : \mathcal{U} \text{ an open cover of } X\}.$$

Let $\mathcal{U}$ be an open cover of Y. We first claim that

$$H(\varphi^{-1}(\mathcal{U})) = H(\mathcal{U}),$$

where we recall that

$$H(\mathcal{U}) = \log(N(\mathcal{U})),$$

$$N(\mathcal{U}) = \text{ smallest cardinality of finite subcover of } \mathcal{U},$$

$$\text{and } \varphi^{-1}(\mathcal{U}) = \{\varphi^{-1}(U) : U \in \mathcal{U}\}.$$

Since φ is continuous we have $\varphi^{-1}(\mathcal{U})$ consists of open sets. Since $Y \subseteq \bigcup_{U \in \mathcal{U}} U$ we have that $\varphi^{-1}(Y) = X \subseteq \bigcup_{U \in \mathcal{U}} \varphi^{-1}(U)$. Thus for every open cover $\mathcal{U}$ of Y we have a corresponding open cover $\mathcal{W}$ for X. Let $\mathcal{A} = \{A_1, \ldots, A_n\} \subseteq \mathcal{U}$ be the finite subcover with smallest cardinality. The above observations tell us that $\varphi^{-1}(\mathcal{A}) \subseteq \mathcal{W}$ is a finite subcover, and therefore $H(\varphi^{-1}(\mathcal{U})) \leq H(\mathcal{U})$. If we can refine it further, then we have $\mathcal{B} = \{\varphi^{-1}(A_1), \ldots, \varphi^{-1}(A_m)\} \subseteq \mathcal{W}$ for $m \leq n$. We have $\varphi(\mathcal{B}) = \{A_1, \ldots, A_m\} \subseteq \mathcal{A}$ is an open subcover by surjectivity. Since $\mathcal{A}$ was chosen to have smallest cardinality, we must have $m = n$. Therefore $H(\varphi^{-1}(\mathcal{U})) = H(\mathcal{U})$.

Recall that for two open covers $\mathcal{U}$ and $\mathcal{W}$, we define $\mathcal{U} \vee \mathcal{W} = \{U \cap W : U \in \mathcal{U}, W \in \mathcal{W}\}$. We claim that $\varphi^{-1}(\mathcal{U} \vee \mathcal{W}) = \varphi^{-1}(\mathcal{U}) \vee \varphi^{-1}(\mathcal{W})$. Notice that

$$\varphi^{-1}(\mathcal{U} \vee \mathcal{W}) = \{\varphi^{-1}(U \cap W) : U \in U, W \in \mathcal{W}\}$$
$$= \{\varphi^{-1}(U) \cap \varphi^{-1}(W) : U \in U, W \in \mathcal{W}\} = \varphi^{-1}(\mathcal{U}) \vee \varphi^{-1}(\mathcal{W}).$$

Now let $\mathcal{U}$ be an open cover. By definition (**Petersen Proposition 6.3.2**) we have

$$h(S, \mathcal{U}) := \lim_{n \to \infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} S^{-i}(\mathcal{U})\right).$$

By what we've just shown and the compatibility condition for φ, we have

$$h(S,\mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} H\left(\varphi^{-1}\left(\bigvee_{i=0}^{n-1} S^{-i}(\mathcal{U})\right)\right)$$

$$= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} \varphi^{-1} S^{-i}(\mathcal{U})\right)$$

$$= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}\varphi^{-1}(\mathcal{U})\right) = h(T,\varphi^{-1}(\mathcal{U})).$$

Hence $h_{\text{Top}}(S) \leq h_{\text{Top}}(T)$.

(2) **Petersen Proposition 6.3.7** tells us that

$$h_{\text{Top}}(X) = \lim_{\epsilon\to 0^+} \limsup_{n\to\infty} \frac{1}{n} \log_2(r(n,\epsilon)),$$

where

$$r(n,\epsilon) := \min\{|E| : E \subseteq X \text{ is } (n,\epsilon) - \textit{spanning}\},$$

and we recall that an (n,ϵ)-spanning set is a set $F \subseteq X$ so that for all $x \in X$ there is a $y \in F$ so that

$$d(T^k(x), T^k(y)) \leq \epsilon \text{ for } 0 \leq k \leq n-1.$$

Since φ is continuous (hence uniformly continuous) we have that for all $\epsilon > 0$ there is a $\delta(\epsilon) > 0$ such that

$$d_X(x,y) < \delta(\epsilon) \implies d_Y(\varphi(x),\varphi(y)) < \epsilon.$$

Let $F \subseteq X$ be an (n,ϵ)-spanning set. We can define the Bowen-Dinaburg metric as

$$d_{X,n}^T(x,y) := \max\{d_X(T^i(x),T^i(y)) : 0 \leq i \leq n\}.$$

Notice that this measures the distance of orbits. Using the observation above, we have the for all $\epsilon > 0$ fixed and each $0 \leq i \leq n$ there is a $\delta(\epsilon, i) > 0$ so that

$$d_X(T^i(x),T^i(y)) < \delta(\epsilon,i) \implies d_Y(\varphi(T^i(x)),\varphi(T^i(y))) < \epsilon.$$

Using the compatibility condition for iterates, we have

$$d_X(T^i(x),T^i(y)) < \delta(\epsilon,i) \implies d_Y(S^i(\varphi(x)),S^i(\varphi(y))) < \epsilon.$$

Take $\delta(\epsilon) = \min\{\delta(\epsilon,i) : 0 \leq i \leq n\}$. In terms of the Bowen-Dinaburg metric, we have

$$d_{X,n}^T(x,y) < \delta(\epsilon) \implies d_{Y,n}^S(\varphi(x),\varphi(y)) < \epsilon.$$

Let

$$B_\epsilon^X(x) := \{y \in X : d_{X,n}^T(x,y) < \epsilon\}$$

denote the balls with respect to this new metric. The above tells us that

$$\varphi(B_{\delta(\epsilon)}^X(x)) \subseteq B_\epsilon^Y(\varphi(x)).$$

Consider an $(n,\delta(\epsilon))$-spanning set for X. The above observation coupled with surjectivity says that the image of this set gives an (n,ϵ)-spanning set for Y. Thus the minimal cardinality for an $(n,\delta(\epsilon))$-spanning set for X will be at least the minimal cardinality for an (n,ϵ)-spanning set for Y. Hence

$$r_Y(n,\epsilon) \leq r_X(n,\delta(\epsilon)).$$

Monotonicity of logarithms and the independence of n tells us

$$\limsup_{n\to\infty} \frac{1}{n} \log(r_Y(n,\epsilon)) \leq \limsup_{n\to\infty} \frac{1}{n} \log(r_X(n,\delta(\epsilon))).$$

Taking $\epsilon \to 0^+$ gives

$$h_{\text{Top}}(S) \leq h_{\text{Top}}(T),$$

as desired.

$\square$

Let G be a locally compact abelian group. Let

$$\widehat{G} = \{\chi : G \to S^1 : \chi \text{ is a continuous homomorphism}\}.$$

We call $\widehat{G}$ the collection of characters of G.

Problem 60. Show that $\widehat{G}$ is an abelian group under pointwise multiplication.

Proof. Let $\chi_1, \chi_2 \in \widehat{G}$. Then for any $x \in G$ we have

$$(\chi_1\chi_2)(x) = \chi_1(x)\chi_2(x) = \chi_2(x)\chi_1(x) = (\chi_2\chi_1)(x).$$

Thus $\chi_1\chi_2 = \chi_2\chi_1$ and we have that the group is abelian. □

We now recall the compact open topology. This is the topology generated by sets of the form

$$B(K, U) = \{\chi \in \widehat{G} : \chi(K) \subseteq U, K \subseteq G \text{ compact and } U \subseteq S^1 \text{ open}\}.$$

It is a difficult exercise to prove that $\widehat{G}$ equipped with the open compact topology is a LCA group (see here). Here's a list of facts.

(1) G has a countable topological basis iff $\widehat{G}$ has a countable topological basis.
(2) G is compact iff $\widehat{G}$ is discrete.
(3) $\widehat{(\widehat{G})}$ is naturally isomorphic to G, with isomorphism given by $\alpha \mapsto a$ where $\alpha(\gamma) = \gamma(a)$ for all $\gamma \in \widehat{G}$.
(4) If G is compact then G is connected iff $\widehat{G}$ is torsion free.
(5) If G_1, G_2 are locally compact abelian groups, then

$$\widehat{G_1 \times G_2} = \widehat{G_1} \times \widehat{G_2}.$$

(6) If Γ a subgroup of $\widehat{G}$, then

$$H = \{g \in G : \gamma(g) = 1 \forall \gamma \in \Gamma\}$$

is a closed subgroup of G, and $\widehat{G/H} = \Gamma$.
(7) If H is a closed subgroup of G and $H \neq G$, then there exists a $\gamma \in \widehat{G}$ with $\gamma \not\equiv 1$ such that $\gamma(h) = 1$ for all $h \in G$.
(8) Let G be compact. The members of $\widehat{G}$ are all mutually orthogonal members of $L^2(m)$, where m is Haar measure.
(9) If G is compact, the members of $\widehat{G}$ form an orthonormal basis for $L^2(m)$.
(10) If $A : G \to G$ is an endomorphism, we can define the dual endomorphism $\widehat{A} : \widehat{G} \to \widehat{G}$ by $\widehat{A}\gamma = \gamma \circ A$ for $\gamma \in \widehat{G}$. Note $\widehat{A}$ is an automorphism iff A is an automorphism.

Problem 61. Prove that the only homomorphisms of $\mathbb{T}^n$ to $\mathbb{T}^1 = S^1$ are maps of the form

$$(z_1, \ldots, z_n) \mapsto z_1^{m_1} \cdots z_n^{m_n} \text{ where } m_1, \ldots, m_n \in \mathbb{Z}.$$

To prove this, follow these steps.

(1) Show that every closed subgroup of $\mathbb{T}^1$ is either $\mathbb{T}^1$ or a finite cyclic subgroup consisting of all pth roots of unity for some $p > 0$.
(2) Show the only automorphisms of K are the map $z \mapsto -z$ and the identity.
(3) Show that the only homomorphisms of K are maps of the form

$$\varphi_n(z) = nz \pmod 1, \qquad n \in \mathbb{Z}.$$

(4) Deduce the result.

Proof. We follow the steps.

(1) Let's view $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z} = [0,1]/\sim$ with $0 \sim 1$ so that things are in terms of addition. Consider H a closed subgroup of $\mathbb{T}^1$. Suppose it had an infinite number of elements. This implies that there is some limit point z_0. So for all $\epsilon > 0$, there is an $a \neq z_0$ which satisfies $d(a, z_0) < \epsilon$. The metric is invariant under the group operation, so $d(a - z_0, 0) < \epsilon$. So $b_\epsilon = a - z_0 \in H$ are ϵ-dense around 0. By adding these b_ϵ to themselves over and over, we get that H is ϵ dense in $\mathbb{T}^1$, so $H = \mathbb{T}^1$.

If H is finite, it has order say p, so for all $a \in H$ we have $pa \equiv 0 \pmod 1$. For pa to be an integer implies that it is a rational number with p as the denominator. The only rational numbers with p as a denominator are

$$\left\{0, \frac{1}{p}, \dots, \frac{p-1}{p}\right\} = \langle 1/p \rangle.$$

This is a cyclic group with p elements, and we see that $H \subseteq \langle 1/p \rangle$ and has the same size, so $H = \langle 1/p \rangle$.

(2) Consider $\theta : \mathbb{T}^1 \to \mathbb{T}^1$ an automorphism. Notice $\theta(0) = 0$. Notice $1/2$ is the only element of order 2, and so we must have $\theta(1/2) = 1/2$. Now $1/4$ and $3/4$ are the only elements of order 4, so either $\theta(1/4) = 1/4$ or $\theta(1/4) = 3/4$ and vice versa for $3/4$. Consider $\theta(1/4) = 1/4$. We need to have intervals are mapped to intervals, so consider the interval $[0, 1/4] \subseteq \mathbb{T}^1$. We have $\theta([0, 1/4])$ is an interval, and the endpoints are fixed, so it is either going to be $[1/4, 1]$ (where we flip the order and note that $1 = 0$) or $[0, 1/4]$ (we keep the order the same). There is no element of order 2 in $[0, 1/4]$, so we cannot have it mapped to $[1/4, 1]$, and thus we must have $[0, 1/4]$.

Now suppose we have that $\theta([0, 1/2^n]) = [0, 1/2^n]$ for $0 \leq n \leq k-1$. The goal is to show it holds for k. Take $1/2^k \in [0, 1/2^{k-1}]$, and notice it is an element with order 2^k so must be mapped to an element of order 2^k. The only elements in $\mathbb{T}^1$ with order 2^k are

$$\left\{\frac{1}{2^k}, \frac{3}{2^k}, \dots \frac{2^k - 1}{2^k}\right\}.$$

We see that none of these are in $[0, 1/2^{k-1}]$ except for $1/2^k$, so this must be fixed. Thus we have that the interval $[0, 1/2^k]$ is fixed by the same argument as above.

Note that there was nothing special about $[0, 1/2^n]$; we can apply this argument to all subintervals. This gives us that all elements of order 2^n are fixed. But by continuity this gives us that it is the identity.

The same kind of argument works if we set $\theta(1/4) = 3/4$, except $\theta(x) = -x$.

(3) We now check that the only homomorphisms of $\mathbb{T}^1$ are of the form $\theta_n(x) = nx$, where $n \in \mathbb{Z}$. Suppose $\theta : \mathbb{T}^1 \to \mathbb{T}^1$ is an endomorphism. If it is non-trivial, its image is a closed connected subgroup of $\mathbb{T}^1$, so using (1) we have that it must be the whole group (i.e. it is surjective). The kernel is a closed subgroup, so either $\ker(\theta) = \mathbb{T}^1$ (trivial) or $\ker(\theta) = H_p$, where

$$H_p = \{x \in [0,1) : px \equiv 0 \pmod 1\} = \langle 1/p \rangle.$$

Notice we have an isomorphism

$$\theta_1 : \mathbb{T}^1 / H_p \to \mathbb{T}^1, \qquad \theta_1(x + H_p) = px \pmod 1.$$

Examine the induced isomorphism

$$\overline{\theta} : \mathbb{T}_1 / H_p \to \mathbb{T}^1, \qquad \overline{\theta}(x + H_p) = \theta(x).$$

Then we have that $\overline{\theta} \circ \theta_1^{-1}$ is an automorphism of K. By (2), we know that this must be either the identity (so $\overline{\theta} = \theta_1$) or the inverse map (so $\overline{\theta} = \theta_1^{-1}$). This forces $\theta(x) = px \pmod 1$ or $\theta(x) = (-p)x \pmod 1$.

(4) Embed and check on generators.

□

Problem 62. Suppose

$$T : (X, \mathcal{M}, \mu) \to (X, \mathcal{N}, \nu)$$

is a measurable map between probability spaces, and suppose we have $\mathcal{A}$ is a semialgebra that generates $\mathcal{N}$ such that for all $A \in \mathcal{A}$,

$$\mu(T^{-1}(A)) = \nu(A).$$

Show that

$$\mathcal{E} = \left\{C \in \mathcal{N} : \mu(T^{-1}(C)) = \nu(C)\right\}$$

is a σ-algebra equal to $\mathcal{N}$. Use this to deduce (5) from (4) in the prior exercise.

[In other words, state/prove **Theorem 1.1** from Walters and use it to establish (5) in the last problem.]

Proof. Note that we have

$$\mathcal{A} \subseteq \mathcal{E}.$$

Let $\alpha(\mathcal{A})$ be the algebra generated by the semialgebra. The first remark is that

$$\alpha(\mathcal{A}) \subseteq \mathcal{E}.$$

To see this, recall that the algbera is created via finite disjoint unions (see **Walters Theorem 0.1**), so all $E \in \alpha(A)$ are of the form $\bigsqcup_{i=1}^n E_i$, where $E_i \in \mathcal{A}$. We have that

$$\mu\left(T^{-1}\left(\bigsqcup_{i=1}^n E_i\right)\right) = \mu\left(\bigsqcup_{i=1}^n T^{-1}(E_i)\right) = \sum_{i=1}^n \mu(T^{-1}(E_i)) = \sum_{i=1}^n \nu(E_i) = \nu\left(\bigsqcup_{i=1}^n E_i\right),$$

hence $E \in \mathcal{E}$. The next thing to note is that $\mathcal{E}$ is a monotone class. This follows from the continuity of measures and the fact that we're dealing with a probability measure space. By the monotone class theorem, this implies that

$$\sigma(\alpha(A)) = \mathcal{N} \subseteq \mathcal{E} \subseteq \mathcal{N},$$

so $\mathcal{E} = \mathcal{N}$. This implies measure preserving by definition.

The conditions here are such that $\mu = \nu = \lambda$ (Lebesgue measure) and the σ-algebras are the Borel σ-algebras. The collection of all intervals forms a semialgebra which generates the Borel σ-algebra, and as we've shown before the measures agree on all intervals, so invoking the theorem we have that they agree on all Borel sets, telling us that our map is measurable. □

Remark. The actual statement of Walters is as follows:

Theorem (Walters, Theorem 1.1)**.** Suppose $T : (X, \mathcal{M}, \mu) \to (Y, \mathcal{N}, \nu)$ is a measurable transformation of probability spaces. Let $\mathcal{C}$ be a semi-algebra that generates $\mathcal{N}$. If for each $A \in \mathcal{C}$ we have $T^{-1}(A) \in \mathcal{M}$ and $\mu(T^{-1}(A)) = \nu(A)$, then T is measure-preserving.

I tried to modify the above to fit the original spirit of the problem.

Problem 63. Suppose $T : (X, \mathcal{M}, \mu) \to (X, \mathcal{M}, \mu)$ is a measure preserving transformation of probability spaces. Let $E \subseteq X$ be a measurable set with $0 < \mu(E)$. Then almost every $x \in E$ returns to E infinitely often.

Proof. Consider

$$E^{**} := E \setminus \left(\bigcap_{n=1}^{\infty} \bigcup_{j \geq n} T^{-j} E \right).$$

Notice that if $x \in E^{**}$, then x does not return to E infinitely often. The goal is to show that the measure of E^{**} is 0. By DeMorgan's, we have

$$\begin{aligned} E^{**} &= E \cap \left(\bigcup_{n=1}^{\infty} \bigcap_{j \geq n} (T^{-j}(E))^c \right) \\ &= \bigcup_{n=1}^{\infty} \left(E \cap \bigcap_{j \geq n} (T^{-j}(E))^c \right). \end{aligned}$$

Taking the measure then gives

$$\begin{aligned} \mu(E^{**}) &\leq \sum_{n=1}^{\infty} \mu \left(E \cap \bigcap_{j \geq n} (T^{-j}(E))^c \right) \\ &\leq \sum_{n=1}^{\infty} \mu \left(E \cap \bigcap_{j \geq 1} (T^{-j}(E))^c \right) \\ &= \sum_{n=1}^{\infty} \mu(E^*) = 0, \end{aligned}$$

where

$$E^* = E \setminus \left(\bigcup_{j=1}^{\infty} T^{-j} E \right).$$

Poincare's theorem tells us that $\mu(E^*) = 0$. $\square$

Problem 64. Let $(X, \mathcal{M}, \mu)$ be a probability space, $T : X \to X$ invertible, injective, measure preserving transformation. Let $A \subseteq X$ be a measurable set with $\mu(A) \neq 0$. Define

$$n_A : X \to \mathbb{N}, \qquad n_A(x) := \inf\{n \geq 1 : T^n x \in A\},$$

$$\mu_A := \frac{\mu}{\mu(A)},$$

$$\mathcal{N} := \{E \cap A : E \in \mathcal{M}\},$$

$$T_A : A \to A, \qquad T_A(x) := T^{n_A(x)} x,$$

$$A_n = \{x \in A : n_A(x) = n\}.$$

We call T_A a **derivative transformation** (or **induced transformation**).

(1) Express why n_A should be finite and defined for almost every $x \in A$.
(2) Show that n_A is measurable.
(3) Show that $\mathcal{N}$ is a σ-algebra.
(4) Show that μ_A is a measure on $\mathcal{N}$.
(5) Show that $(A, \mathcal{N}, \mu_A)$ is a probability space.
(6) Show that T_A is measurable.
(7) Show that T_A is a measure preserving transformation.

Proof.

(1) This follows by Poincare's theorem.
(2) The assumed σ-algebra on $\mathbb{N}$ is the trivial one, $\mathcal{P}(\mathbb{N})$. So it suffices to show that it is measurable on each $k \in \mathbb{N}$. Notice that

$$n_A^{-1}(\{k\}) = \{x \in X : T^k x \in A, T^j x \notin A \text{ for } 1 \leq j < k\} = T^{-k}(A) \setminus \left(\bigcup_{j=1}^{k-1} T^{-j}(A) \right).$$

These are all measurable sets (since T is measurable) so $n_A^{-1}(\{k\})$ is measurable for all $k \in \mathbb{N}$.
(3) We see that $A \in \mathbb{N}$, since $X \cap A = A$. Let $\{E_i \cap A\}_{i=1}^{\infty} \subseteq \mathcal{N}$. Then

$$\bigcup_{i=1}^{\infty}(E_i \cap A) = A \cap \bigcup_{i=1}^{\infty} E_i \in \mathcal{N}.$$

If $E \in \mathcal{N}$, then $E = A \cap F$, $F \in \mathcal{M}$. We need to show $A \setminus E \in \mathcal{N}$. To do so, notice

$$A \setminus E = A \cap E^c = A \cap (A^c \cup F^c) = A \cap F^c \in \mathcal{N}.$$

(4) Notice

$$\mu_A(\varnothing) = \frac{\mu(\varnothing)}{\mu(A)} = 0.$$

Notice

$$\mu_A\left(\bigsqcup_{i=1}^{\infty}(E_i \cap A)\right) = \frac{1}{\mu(A)}\mu\left(\bigsqcup_{i=1}^{\infty}(E_i \cap A)\right) = \frac{1}{\mu(A)}\left(\sum_{i=1}^{\infty}\mu(E_i \cap A)\right) = \sum_{i=1}^{\infty}\mu_A(E_i \cap A).$$

This is indeed a measure. Notice

$$\mu_A(A) = 1,$$

so it is a probability measure.
(5) Follows by (4).
(6) Let $E \subseteq A$. Then

$$T_A^{-1}(E) = \{x \in A : T^{n_A(x)}x \in E\} = \bigcup_{n \geq 1} A_n \cap \{x \in A : T^n x \in E\} = \bigcup_{n \geq 1}(A_n \cap T^{-n}(E)).$$

Moreover, notice that the A_n are disjoint (useful for next part).
(7) We define a bunch of sets which have convenient properties and hope things work out. Let $E \subseteq A$ be measurable.

First, notice that

$$T_A^{-1}(E) = \bigsqcup_{n \geq 1}(A_n \cap T^{-n}(E)),$$

so

$$\mu_A(T_A^{-1}(E)) = \frac{1}{\mu(A)}\sum_{n \geq 1}\mu(A_n \cap T^{-n}(E)).$$

Next, let

$$F_0 = A, \quad F_k = \{x \in X : T^k x \in A, T^j x \notin A \text{ for } 0 \leq j < k\} \text{ for } k \geq 1.$$

Notice that

$$\begin{aligned} T^{-1}(F_k) = \{x \in X : T^{k+1}x \in A, T^j \notin A \text{ for } 1 \leq j < k+1\} \\ = A_{k+1} \sqcup F_{k+1}. \end{aligned}$$

Now we see that

$$\mu(E) = \mu(E \cap A) = \mu(E \cap F_0).$$

Since T is measure preserving, we have

$$\mu(E \cap F_0) = \mu(T^{-1}(E \cap F_0)) = \mu(T^{-1}(E) \cap T^{-1}(F_0)) = \mu(T^{-1}(E) \cap A_1) + \mu(T^{-1}(E) \cap F_1).$$

We can continue this inductively; that is, we have

$$\mu(T^{-n}(E) \cap F_n) = \mu(T^{-n-1}(E) \cap T^{-1}(F_n)) = \mu(T^{-(n+1)}(E) \cap F_{n+1}) + \mu(T^{-(n+1)}(E) \cap E_{n+1}).$$

Letting this go to infinity gives

$$\mu(E) = \sum_{n \geq 1} \mu(T^{-n}(E) \cap A_n).$$

Thus

$$\begin{aligned} \mu_A(T_A^{-1}(E)) &= \frac{1}{\mu(A)} \sum_{n \geq 1} \mu(A_n \cap T^{-n}(E)) \\ &= \frac{1}{\mu(A)} \mu(E) = \mu_A(E). \end{aligned}$$

So T_A is measure preserving.

□

We now go through the construction of a primitive transformation on a superset.

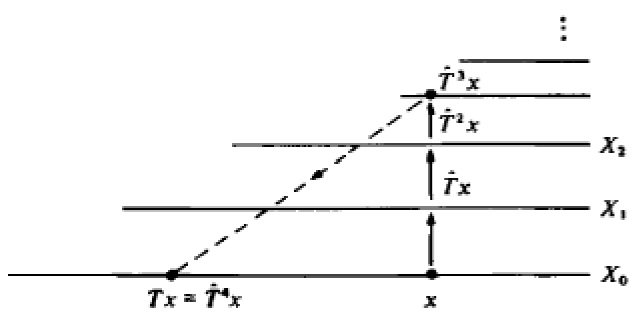
Consider $\cdots \subseteq Y_3 \subseteq Y_2 \subseteq Y_1 \subseteq Y_0 = X$ to be a decreasing sequence of measurable sets. We can take $\{X_i\}$ to be the copies of Y_i such that they are all disjoint. Notice each X_i is a measure space, via the induced σ-algebra $\mathcal{M}_i = \{E \cap X_i : E \in \mathcal{M}\}$ and induced measure $\mu_i = \frac{\mu}{\mu(X_i)}$. Let $\widehat{X} = \bigsqcup_{i \geq 0} X_i$. We equip $\widehat{X}$ with the appropriate σ-algebra and measure, labeled $\widehat{\mathcal{M}}$ and $\widehat{\mu}$. Let $E \subseteq \widehat{X}$. Then we have that

$$\widehat{\mu}(E) = \sum_{n=0}^{\infty} \mu(E \cap X_n).$$

We have a picture of $\widehat{X}$ being a tower built over $X_0 = X$. Suppose $T : X \to X$ is an invertible measure preserving transformation of a probability space. Define $\varphi_{j,i} : X_j \to X_i$ to be the map defined by the inclusion $Y_j \subseteq Y_i$. If $T : X \to X$ is a measure preserving transformation of a probability space, we get an induced map

$$\widehat{T} : \widehat{X} \to \widehat{X}, \qquad \widehat{T} = \begin{cases} \varphi_{i+1,i}^{-1}(\widehat{x}) \text{ if } \widehat{x} \in X_i \text{ and } \varphi_{i+1,i}^{-1}(\widehat{x}) \neq \varnothing, \\ T(\varphi_{i,0}(\widehat{x})) \text{ otherwise.} \end{cases}.$$

While notationally cumbersome, this can also be described via a picture (sometimes called Kakutani's skyscraper):



For simplicity, define $\pi : \widehat{X} \to X$ to be the projection map (so that we can view $\pi(\widehat{x}) \in X$).

Problem 65.

(1) Show that $\widehat{T}$ is measurable.
(2) Show that $\widehat{T}$ preserves $\widehat{\mu}$.

Proof.

(1) Let $E \subseteq \widehat{X}$ be a measurable set. The goal is to show $\widehat{T}^{-1}(E)$ is also measurable. Let $E_n = E \cap X_n$, $n \geq 0$. Note that each $E_n \subseteq X_n$ is a measurable set. Then we get $E = \bigsqcup_{n \geq 0} E_n$, so if we show each $\widehat{T}^{-1}(E_n)$ is measurable then we have the result. Now, for $n \geq 1$, $\widehat{T}^{-1}(E_n)$ can be thought of as $\varphi_{n,n-1}(E_n)$. The claim then is that for $n \geq 1$, $\varphi_{n,n-1}(E_n)$ is a measurable set. But this follows by viewing this as the inclusion $Y_n \subseteq Y_{n-1}$ and then using the fact that measurable subsets of Y_n are measurable subsets of Y_{n-1}. Now we consider E_0. We see

$$\widehat{T}^{-1}(E_0) = \{\widehat{x} \in \widehat{X} : \text{ For some i, } \widehat{x} \in X_i \text{ and } \varphi_{i+1,i}^{-1}(\widehat{x}) = \varnothing \text{ and } T(\pi(\widehat{x})) \in E_0\}.$$

In other words, if we view $E_0 \subseteq X$ as well, then

$$\widehat{T}^{-1}(E_0) = \bigsqcup_{n \geq 0} F_n, \text{ where } F_n = T^{-1}(E_0) \cap Y_n \cap Y_{n+1}^c \subseteq X_n.$$

Each of these are measurable sets, so we get that $\widehat{T}^{-1}(E_0)$ is measurable as well.

(2) We now need to show that for $E \subseteq \widehat{X}$, we have

$$\widehat{\mu}(\widehat{T}^{-1}(E)) = \widehat{\mu}(E).$$

Notice by the description in (1), we have

$$\begin{aligned}\widehat{T}^{-1}(E) &= \bigsqcup_{n \geq 0} \widehat{T}^{-1}(E_n) = \widehat{T}^{-1}(E_0) \sqcup \bigsqcup_{n \geq 1} \widehat{T}^{-1}(E_n) \\ &= \bigsqcup_{n \geq 0} (T^{-1}(E_0) \cap X_n \cap X_{n+1}^c) \sqcup \bigsqcup_{n \geq 1} E_n',\end{aligned}$$

where E_n' is $E_n = E \cap X_n$ but viewed in X_{n-1}. Taking the measure of this, we have

$$\begin{aligned}\widehat{\mu}(\widehat{T}^{-1}(E)) &= \sum_{n \geq 0} \mu(T^{-1}(E_0) \cap Y_n \cap Y_{n+1}^c) + \sum_{n \geq 1} \widehat{\mu}(E_n') \\ &= \sum_{n \geq 0} \mu(T^{-1}(E_0) \cap Y_n \cap Y_{n+1}^c) + \sum_{n \geq 1} \mu(E_n)\end{aligned}$$

$$= \mu(T^{-1}(E_0)) + \sum_{n \geq 1} \mu(E_n) = \mu(E_0) + \sum_{n \geq 1} \mu(E_n) = \sum_{n \geq 0} \mu(E_n) = \widehat{\mu}(E).$$

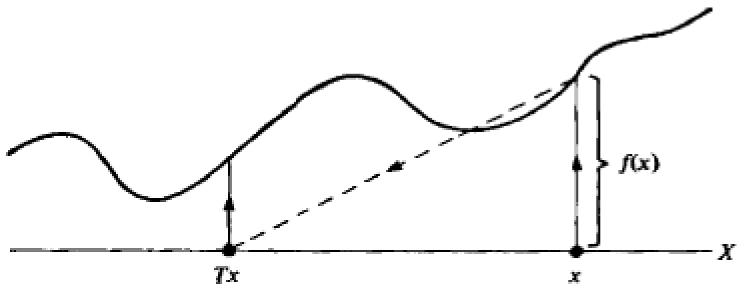
So this transformation is measure preserving.

□

Now let $(X, \mathcal{M}, \mu, T)$ be a measure-preserving system (meaning $(X, \mathcal{M}, \mu)$ is a probability space, $T : X \to X$ is an invertible measure preserving transformation). For $f : X \to (0, \infty)$, consider

$$\Gamma_f = \{(x, t) : 0 \leq t < f(x)\}.$$

This is the collection of points "under f." We identify $(x, f(x))$ and $(Tx, 0)$. We have the following picture:



For $n \in \mathbb{Z}$, define

$$S_n(x) = \begin{cases} \sum_{k=0}^{n-1} f(T^k(x)) \text{ if } n > 0 \\ 0 \text{ if } n = 0 \\ -\sum_{k=1}^{-n} f(T^{-k}(x)) \text{ if } n < 0. \end{cases}$$

Problem 66. Use Poincare recurrence to show that $S_n(x) \to \infty$ as $n \to \infty$. The same kind of argument can be used to show $S_n(x) \to -\infty$ as $n \to -\infty$.

Proof. Notice that $f^{-1}((0, \infty)) = X$. By continuity of measures, there must be some $a > 0$ so that $\mu(f^{-1}((a, \infty))) \neq 0$. If we let $E_a = f^{-1}((a, \infty))$, then by Poincare recurrence almost every $x \in E_a$ returns to E_a infinitely often, so $f(T^k(x)) > a$ infinitely often for almost every $x \in E_a$. Consequently, for almost every $x \in E_a$, we have $S_n(x) \to \infty$ as $n \to \infty$. Now this holds for each E_a (technically, even if the set E_a has measure zero it will still hold), and we can write

$$X = \bigcup_{a > 0} E_a.$$

The union of sets of measure zero will be measure zero, so it holds for almost every $x \in X$. □

For $x \in X$, $0 \leq t < f(x)$, $s \in \mathbb{R}$, define

$$n(x, t, s) := \min \{k \in \mathbb{Z}_{\geq 0} : s + t < S_{k+1}(x)\}.$$

This is called the **hitting number**.

Problem 67. Show that $n(x, t, s)$ is well-defined, and satisfies the property that

$$S_{n(x,t,s)} \leq s + t < S_{n(x,t,s)+1}.$$

Proof. The fact that $n(x, t, s)$ is well-defined follows from the fact that $S_n(x) \to \infty$, so there must be some n so that $s + t < S_n(x)$, and the minimum will be unique. The fact that it's a minimum tells us that we have the above identity. □

For $t \geq 0$, define

$$T_s^f(x, t) = T^f(x, t, s) := (T^{n(x,t,s)}(x), s + t - S_{n(x,t,s)}(x)).$$

Problem 68. Show that S_n satisfies the cocycle relation; i.e.,

$$S_{n+m} = S_n + S_m \circ T^n.$$

Proof. We see that

$$S_m \circ T^n(x) = \sum_{j=0}^{m-1} f(T^{j+n}(x)) = \sum_{j=n}^{m+n-1} f(T^j(x)) = S_{m+n}(x) - S_n(x).$$

□

Problem 69. Show that $n_s = n(\cdot, s)$ satisfies the coycle relation; i.e.,

$$n(x,t,s+q) = n(x,t,s) + n(T^{n(x,t,s)}(x), s+t-S_{n(x,t,s)}(x), q).$$

Proof. Notice that

$$n(T^{n(x,t,s)}(x), s+t-S_{n(x,t,s)}(x), q) = \min\left\{k \in \mathbb{Z} : s+t+q-S_{n(x,t,s)}(x) < S_{k+1}(T^{n(x,t,s)}(x))\right\}.$$

By the cocyle relation for S_n, this is the same as

$$n(T^{n(x,t,s)}(x), s+t-S_{n(x,t,s)}(x), q) = \min\left\{k \in \mathbb{Z} : s+t+q < S_{n(x,t,q)+k+1}(x)\right\}.$$

After changing variables appropriately, we see

$$n(T^{n(x,t,s)}(x), s+t-S_{n(x,t,s)}(x), q) = \min\{\alpha \in \mathbb{Z} : s+t+q < S_{\alpha+1}(x)\} - n(x,t,q) = n(x,t,s+q) - n(x,t,q).$$

This gives us the cocycle property. □

Problem 70. Show that T_s^f is a flow.

Proof. There are two things we need to show.

(1) We see that

$$n(x,t,0) = \min\{k \in \mathbb{Z} : t \le S_{k+1}(x)\} = \min\left\{k \in \mathbb{Z} : t \le \sum_{j=0}^{k} f(T^j(x))\right\}.$$

Since $0 \le t < f(x)$, this implies that

$$t \le S_1(x) = f(x),$$

so $n(x,t,0) = 0$. Therefore

$$T_0^f(x,t) = (T^0(x), t - S_0(x)) = (x,t).$$

So T_0^f is the identity.

(2) We next need to check that the $\mathbb{R}$ action is satisfied, meaning

$$T_{s+q}^f(x,t) = T_s^f \circ T_q^f(x,t).$$

Notice

$$\begin{aligned} T_s^f\left(T_q^f(x,t)\right) &= T_s^f(T^{n(x,t,q)}(x), q+t-S_{n(x,t,q)}(x)) \\ &= \Big(T^{n(T^{n(x,t,q)}(x), s+t-S_{n(x,t,q)}(x), s)}(T^{n(x,t,q)}(x)), \\ & s+q+t-S_{n(x,t,q)}(x) - S_{n(T^{n(x,t,q)}(x), s+t-S_{n(x,t,q)}(x), s)}(T^{n(x,t,q)}(x))\Big). \end{aligned}$$

Use the cocycle property for n to get

$$n(T^{n(x,t,q)}(x), s+t-S_{n(x,t,q)}(x), s) = n(x,t,s+q) - n(x,t,q),$$

so

$$T^{n(T^{n(x,t,q)}(x), s+t-S_{n(x,t,q)}(x), s)}(T^{n(x,t,q)}(x)) = T^{n(x,t,s+q)}(x).$$

Now

$$S_{n(T^{n(x,t,q)}(x), s+t-S_{n(x,t,q)}(x), s)}(T^{n(x,t,q)}(x)) = S_{n(x,t,s+q)-n(x,t,q)}(T^{n(x,t,q)}(x)).$$

Plugging in the definition, we get

$$S_{n(x,t,s+q)-n(x,t,q)}(T^{n(x,t,q)}(x)) = \sum_{j=0}^{n(x,t,s+q)-n(x,t,q)-1} f(T^{j+n(x,t,q)}(x)) = \sum_{j=n(x,t,q)}^{n(x,t,s+q)-1} f(T^j(x))$$
$$= S_{n(x,t,s+q)}(x) - S_{n(x,t,q)}.$$

Substituting this in, we have

$$S_{n(x,t,q)}(x) + S_{n(T^{n(x,t,q)}(x),s+t-S_{n(x,t,q)}(x),s)}(T^{n(x,t,q)}(x)) = S_{n(x,t,s+q)}(x).$$

So

$$T_s^f\left(T_q^f(x,t)\right) = \left(T^{n(x,t,s+q)}(x), s+q+t-S_{n(x,t,s+q)}(x)\right) = T_{s+q}^f(x,t).$$

Thus this is actually a flow.

□

We call $\{T_s^f\}_{s\in\mathbb{R}}$ the **induced flow**. Note this is the flow going upward with unit speed.

Let $n \in \mathbb{Z}$ and define

$$\Gamma_{n,s} = \{(x,t) \in \Gamma : n(x,t,s) = n\}.$$

If we fix $s \in \mathbb{R}$, we denote the above as just Γ_n. Another way to view these points is as the following:

$$\{(x,t) \in \Gamma : T_s^f(x,t) = (T^n(x), s+t-S_n(x))\}.$$

That is,

$$T_s^f(\Gamma_{n,s}) = \{(T^n(x), t+s-S_n(x)) : (x,t) \in \Gamma_{n,s}\}.$$

Set

$$X_n = \pi(\Gamma_{n,s}).$$

Define

$$t_1 : X_n \to [0,\infty), \qquad t_1(x) = \inf\{t \in [0, f(x)) : n(x,t,s) = n\},$$
$$t_2 : X_n \to [0,\infty), \qquad t_2(x) = \sup\{t \in [0, f(x)) : n(x,t,s) = n\}.$$

We note that

$$n(x, t_1(x), s) = n.$$

This says that for $t_1(x)$ and $t_2(x)$ we have the property that

$$0 \le s + t_1(x) \le s + t_2(x) < S_{n+1}(x), \qquad S_{n+1}(x) = S_n(x) + f(T^n(x)),$$
$$0 \le s + t_1(x) - S_n(x) \le s + t_2(x) - S_n(x) < f(T^n(x)).$$

We also have

$$0 \le t_1(x) \le t_2(x) < f(x).$$

Notice these are measurable functions by construction. Finally, we can write

$$\Gamma_{n,s} = \{(x,t) : x \in X_n, t_1(x) \le t \le t_2(x)\},$$
$$T_s^f(\Gamma_{n,s}) = \{(T^n x, t+s-S_n(x)) : (x,t) \in \Gamma_n\}.$$

Let μ_f denote $\mu \times \lambda|_\Gamma$ (μ times Lebesgue measure restricted to Γ). For any $E \subseteq \Gamma$, we can write

$$\mu_f(E) = \int_{x\in X} \left(\int_0^{f(x)} \chi_E(x,t)dt\right) d\mu(x).$$

Problem 71. Using the above, verify that

$$\mu_f(T_s^f(E)) = \mu(E).$$

Recall the **Mean Ergodic Theorem**.

Theorem (Mean Ergodic Theorem)**.** Let $U : H \to H$ be an isometry of a complex Hilbert space. Let $\mathcal{M} = \{x \in H : Ux = x\}$ be the space of vectors invariant under U. Let $P : H \to \mathcal{M}$ be the projection map. Then we have

$$A_n(x) = \frac{1}{n}\sum_{j=0}^{n-1} U^j(x) \to Px \text{ for all } x \in H.$$

Problem 72. Proof the Mean Ergodic Theorem by following these steps.

(1) Show that it holds true for the space $\mathcal{M}$.
(2) Let
$$\mathcal{N} = \{x - Ux : x \in H\}.$$
Show that the theorem holds true for this subspace.
(3) Show that it holds for $\overline{\mathcal{N}}$, the norm closure of $\mathcal{N}$.
(4) Show that it holds for $\overline{\mathcal{N}}^{\perp}$ (the orthogonal closure of $\overline{\mathcal{N}}$).
(5) Deduce that it holds for all of H.

Proof.

(1) Notice that for all $x \in \mathcal{M}$, we have
$$U^j x = x \text{ for all } j \geq 0,$$
so
$$\frac{1}{n}\sum_{j=0}^{n-1} U^j(x) = \frac{nx}{n} = x = Px \text{ for all } n \geq 1.$$
(2) Take $y = x - Ux \in \mathcal{N}$. Then
$$\frac{1}{n}\sum_{j=0}^{n-1} U^j(x - Ux) = \frac{1}{n}\sum_{j=0}^{n-1}\left[U^j(x) - U^{j+1}(x)\right] = \frac{1}{n}\left[x - U^n(x)\right].$$
Taking the norm of both sides, we have
$$\left\|\frac{1}{n}\sum_{j=0}^{n-1} U^j(y)\right\| \leq \frac{2}{n}\|x\|,$$
since U is an isometry. Taking the limit as $n \to \infty$, we get that the norm tends to zero, which is $P(y)$.
(3) Take $\overline{\mathcal{N}}$ the norm closure of $\mathcal{N}$. Let $(y_k) \subseteq \mathcal{N}$ be a sequence with $y_k \to y \in \overline{\mathcal{N}}$. Then
$$\|A_n(y)\| \leq \|A_n(y - y_k)\| + \|A_n(y_k)\|.$$
Now $y \to y_k$, so choose k sufficiently large so that $\|y - y_k\| < \epsilon/2$. Then
$$\|A_n(y - y_k)\| = \left\|\frac{1}{n}\sum_{j=0}^{n-1} U^j(y - y_k)\right\| \leq \frac{1}{n}\sum_{j=0}^{n-1}\|U^j(y - y_k)\| < \frac{1}{n}\sum_{j=0}^{n-1}\epsilon/2 = \epsilon/2.$$
This holds for arbitrary n, so choose n sufficiently large so that $\|A_n(y_k)\| < \epsilon/2$. Then we have that for n large,
$$\|A_n(y)\| < \epsilon.$$
We can find such an n for all $\epsilon > 0$, so $\|A_n(y)\| \to 0$.

(4) First, we remark that $\overline{\mathcal{N}}^{\perp} = \mathcal{N}^{\perp}$. So take $y \in N^{\perp}$. That is, y is such that $\langle x, y\rangle = 0$ for all $x \in \mathcal{N}$. Since $x \in \mathcal{N}$, we can write it as $x = z - Uz$ for $z \in H$. Thus we have $\langle z - Uz, y\rangle = 0$ for all $z \in H$. Using linearity, we have

$$\langle z - Uz, y\rangle = \langle z, y\rangle - \langle Uz, y\rangle = 0.$$

So

$$\langle z, y\rangle = \langle Uz, y\rangle.$$

Now U is an isometry, so we can take its adjoint to get

$$\langle z, y\rangle = \langle z, U^* y\rangle.$$

Subtract again to get

$$\langle z, y - U^* y\rangle = 0.$$

This holds for all $z \in H$, so $y - U^* y = 0$, or $y = U^* y$. The goal now is to use this to show that $y = Uy$. Notice

$$\begin{aligned}\|Uy - y\|^2 = \langle Uy - y, Uy - y\rangle &= \langle Uy, Uy\rangle - \langle y, Uy\rangle - \langle Uy, y\rangle + \langle y, y\rangle \\ &= 2\langle y, y\rangle - \langle y, Uy\rangle - \langle Uy, y\rangle.\end{aligned}$$

Notice

$$\begin{aligned}\langle y, Uy\rangle &= \langle U^* y, y\rangle = \langle y, y\rangle, \\ \langle Uy, y\rangle &= \langle y, U^* y\rangle = \langle y, y\rangle.\end{aligned}$$

Substituting this in gives

$$\|Uy - y\|^2 = 0 \implies Uy = y.$$

Thus, we have that $\mathcal{N}^{\perp} = \mathcal{M}$. Now apply (1).

(5) We can write

$$H = \overline{\mathcal{N}} \oplus \mathcal{M}.$$

So every $x \in H$ can be written uniquely as $x = x_1 + x_2$, where $x_1 \in \overline{\mathcal{N}}$ and $x_2 \in \mathcal{M}$. Apply A_n to this and use the linearity to get the desired result.

□

We now move on to the Birkhoff Ergodic theorem. Suppose we have a sequence of real numbers $(a_j)_{j=1}^n$. Let m be a positive integer such that $m \le n$. A term a_k of the sequence we be called a **m-leader** if there exists a positive integer p with $1 \le p \le m$ and such that

$$a_k + \cdots + a_{k+p-1} \ge 0.$$

For example, the 1-leaders are the non-negative terms of the sequence.

Problem 73. Show that the sum of m-leaders is non-negative.

Proof. First, notice that if a_k is an m-leader, then we have a $1 \le p \le m$ so that

$$a_k + \cdots + a_{k+p-1} \ge 0.$$

Let p be the smallest such, and suppose k is the smallest numbers to that a_k is an m-leader (that is, a_k is the first m-leader). The claim then is that each of the a_t in this are m-leaders themselves. If a_t in this is not an m-leader, then we see there is no $1 \le p \le m$ so that

$$a_t + \cdots + a_{t+p-1} \ge 0.$$

Consequently

$$a_t + \cdots + a_{k+p-1} < 0,$$

so we can omit it to get a smaller p. Thus

$$a_k + \cdots + a_{t-1} \ge 0.$$

This however contradicts minimality of p. So a_t must be an m-leader.

Thus we have $(a_j)_{j=k}^{k+p-1}$ gives us some of the m-leaders, and we have that the sum of these is greater than or equal to 0. For $a_{k+p}, \ldots, a_n$ we repeat this process with all of the other m-leaders. Thus each of these have their associated sequences, which are greater than or equal to 0, and adding them all up gives us a number greater than or equal to zero. □

We will use this to prove the **Maximal Ergodic Theorem**.

Theorem (Maximal Ergodic Theorem)**.** Let $(X, \mathcal{M}, \mu, T)$ be a measure-preserving system. Let

$$f_j = f(T^j(x)).$$

If

$$E = \left\{ x \in X : \sum_{j=0}^{n-1} f_j(x) \geq 0 \text{ for some n} \right\},$$

then

$$\int_E f(x)dx \geq 0.$$

Problem 74. Follow the proof of Halmos for the Maximal Ergodic theorem. That is, prove the Maximal Ergodic theorem following these steps.

(1) Let

$$E_m = \left\{ x \in X : \sum_{j=0}^{p} f_j(x) \geq 0 \text{ for some } p \leq m \right\}.$$

Show that $E_m \nearrow E$.

(2) Deduce that it is sufficient to show

$$\int_{E_m} f(x)dx \geq 0 \text{ for each } m.$$

(3) Let n be an arbitrary positive integer. Consider for each point x the m-leaders in the sequence $f_0(x), \ldots, f_{n+m-1}(x)$. Let $s(x)$ be the sum of the m-leaders. Let

$$D_k = \{x \in X : f_k(x) \text{ is an m-leader of the sequence } f_0(x), \ldots, f_{n+m-1}(x)\}.$$

Show D_k is measurable.

(4) Let

$$g_k = \chi_{D_k}.$$

Show that

$$s = \sum_{k=0}^{n+m-1} f_k g_k.$$

(5) Deduce that s is measurable and integrable.

(6) Deduce that

$$\sum_{k=0}^{n+m-1} \int_{D_k} f_k(x)dx \geq 0.$$

(7) Observe that if $k = 1, \ldots, n-1$, then the following conditions are equivalent:

(a) $T(x) \in D_{k-1}$.

(b) $f_{k-1}(Tx) + \cdots + f_{k-1+p-1}(Tx) \geq 0$ for some $p \leq m$.

(c) $f_k(x) + \cdots + f_{k+p-1}(x) \geq 0$ for some $p \leq m$.

(d) $x \in D_k$.

(8) Use the prior part to establish $D_k = T^{-1}D_{k-1}$.

(9) Deduce $D_k = T^{-k} D_0$.
(10) Show that $D_0 = E_m$.
(11) Use the prior part to calculate

$$\int_{D_k} f_k(x)dx.$$

(12) Use (6), (10), and (11) to conclude

$$n \int_{E_m} f(x)dx + m \int |f(x)|dx \geq 0.$$

(13) Finish the Maximal Ergodic theorem.

Proof.

(1) If $n \leq m$, $x \in E_n$ implies that $\sum_{j=0}^{p} f_j(x) \geq 0$ for some $p \leq n \leq m$, so $x \in E_m$. This implies $E_n \subseteq E_m$, which gives us increasing. Next, let

$$F = \bigcup_{m \geq 0} E_m.$$

Take $x \in F$. Then $x \in E_m$ for some m, so $\sum_{j=0}^{p} f_j(x) \geq 0$ for some $p \leq m$. But this implies that $x \in E$, taking $n = p+1$. So $F \subseteq E$. For the other direction, we see that $x \in E$ implies $\sum_{j=0}^{n-1} f_j(x) \geq 0$, so $x \in E_{n-1}$ for some n. Thus $E = F$.

(2) If we can show

$$\int_{E_m} f(x)dx \geq 0$$

for each m, then

$$\int_E f(x)dx = \lim_{m \to \infty} \int_{E_m} f(x)dx \geq 0.$$

(3) If $x \in D_k$, then there is a $1 \leq p \leq m$ so that

$$f_k(x) + \cdots + f_{k+p-1}(x) \geq 0.$$

Let $G_{k,p}(x) = f_k(x) + \cdots + f_{k+p-1}(x)$ for each such p. Then $G_{k,p}$ is a measurable function, and

$$F_{k,p} = G_{k,p}^{-1}([0, \infty))$$

is a measurable set. We can then express

$$D_k = \bigcup_{1 \leq p \leq m} F_{k,p}.$$

So D_k is measurable.

(4) Examine $x \in X$ fixed. Then $s(x)$ is the sum of the m-leaders. Notice that $f_k(x)$ is an m-leader if and only if $x \in D_k$. Thus

$$\sum_{k=0}^{n+m-1} f_k(x) g_k(x)$$

gives us the sum of all of the m-leaders in $(f_j(x))_{j=0}^{n+m-1}$. So

$$s(x) = \sum_{k=0}^{n+m-1} f_k(x) g_k(x).$$

This works for each $x \in X$.

(5) Since D_k is measurable, g_k is measurable. The product of measurable functions is measurable, and the sum of measurable functions is measurable, so s is measurable. Integrability follows since $s(x) \geq 0$ (using **Problem 16**).

(6) We use the linearity of integration to get

$$\int s(x)dx = \int \sum_{k=0}^{n+m-1} f_k(x)g_k(x)dx = \sum_{k=0}^{n+m-1} \int_{D_k} f_k(x)dx \geq 0.$$

(7) We observe the equivalence.

$(a) \implies (b)$: If $T(x) \in D_{k-1}$, then this says that $f_{k-1}(Tx)$ is an m-leader, giving us (b).
$(b) \implies (c)$: This follows by definition of $f_j(x)$.
$(c) \implies (d)$: This says $f_k(x)$ is an m-leader, so $x \in D_k$.
$(d) \implies (a)$: If $x \in D_k$, then $f_k(x)$ is an m-leader, meaning there is some $p \leq m$ so that

$$f_k(x) + \cdots + f_{k+p-1}(x) \geq 0,$$

and using the definition of f_j we have

$$f_{k-1}(Tx) + \cdots + f_{k+p-2}(Tx) \geq 0,$$

implying that $f_{k-1}(Tx)$ is an m-leader as well, or $T(x) \in D_{k-1}$.

(8) We have $x \in D_k$ if and only if $T(x) \in D_{k-1}$ by the equivalence in (7). Thus

$$T^{-1}(D_{k-1}) = \{x \in X : T(x) \in D_{k-1}\} = \{x \in X : x \in D_k\} = D_k.$$

(9) Proceed by induction to get $T^{-k}(D_0) = D_k$.

(10) We see

$$D_0 = \{x \in X : f_0(x) \text{ is an m-leader of the sequence } f_0(x), \dots, f_{n+m-1}(x)\}.$$

If $f_0(x)$ is an m-leader of the sequence, we get that there is some $1 \leq p \leq m$ so that

$$\sum_{j=0}^{p-1} f_j(x) \geq 0.$$

This says that $x \in E_m$. The same argument backwards works.

(11) We see

$$\int_{D_k} f_k(x)dx = \int_{D_0} f_k(T^{-k}(x))dx = \int_{D_0} f_0(x)dx = \int_{D_0} f(x)dx.$$

(12) We see that

$$\begin{aligned} 0 \leq \sum_{k=0}^{n+m-1} \int_{D_k} f_k(x)dx &= \sum_{k=0}^{n-1} \int_{D_k} f_k(x) + \sum_{k=n}^{n+m-1} \int_{D_k} f_k(x) \\ &\leq n \int_{E_m} f(x)dx + m\|f\|_1 \geq 0. \end{aligned}$$

(13) Dividing by n in (12), we have

$$\int_{E_m} f(x)dx + \frac{m}{n}\|f\|_1 \geq 0.$$

Taking the limit as $n \to \infty$, we get

$$\int_{E_m} f(x)dx \geq 0.$$

This holds for all $m \geq 0$, so we get the result in (2).

□

We now use this for **Birkhoff's Ergodic theorem**.

Theorem (Individual Ergodic Theorem)**.** If $(X, \mathcal{M}, \mu, T)$ is a measure-preserving system (X may have possibly infinite measure) and if $f \in L^1$, then

$$A_n(f) = \frac{1}{n}\sum_{j=0}^{n-1} f(T^j(x))$$

converges almost everywhere. Let $f'(x)$ be the value to which it converges. The function f' is integrable and invariant. Furthermore, if $\mu(X) < \infty$, then

$$\int f' = \int f.$$

Problem 75. Follow Halmos' proof for the Individual Ergodic theorem. That is, prove the theorem using these steps.

(1) Let
$$f^*(x) = \limsup_{n\to\infty} A_n(f)(x), \qquad f_*(x) = \liminf_{n\to\infty} A_n(f)(x).$$
Show that these functions are T-invariant.

(2) Let $a < b$ be real numbers. Let
$$Y_{a,b} = \{x \in X : f_*(x) < a < b < f^*(x)\}.$$
Show that $Y = Y_{a,b}$ is measurable and T-invariant.

(3) Show that Y can be assumed to be σ-finite.

(4) Show that $\mu(Y) < \infty$. To do so, take $C \subseteq Y$ a set with finite measure and show it is uniformly bounded. (*Hint: Use the Maximal Ergodic theorem with* $h = f - b\chi_C$*.*)

(5) Show that $\mu(Y) = 0$. (*Hint: Use the Maximal Ergodic theorem with the functions* $h = f - b$*,* $g = a - f$*)*

(6) Applying the result to all (a, b) rational points, we get that the limit converges almost everywhere. Show that f' is integrable and measurable.

(7) Show that f' is invariant.

(8) Show that if $\mu(X) < \infty$, then f and f' have the same integral.

Proof.

(1) The first thing to remark is that these functions are measurable, since limsups and liminfs are measurable. Next, to see they are T-invariant, notice
$$\begin{aligned} f^*(T(x)) = \limsup_{n\to\infty} A_n(f)(Tx) = \limsup_{n\to\infty} \left(\frac{1}{n}\sum_{j=1}^{n} f(T^j(x))\right) \\ = \limsup_{n\to\infty}\left(A_{n+1}(f)(x) - \frac{1}{n}f(x)\right) \\ = \limsup_{n\to\infty} A_{n+1}(f)(x) = f^*(x). \end{aligned}$$
The same argument applies with liminfs.

(2) Notice that
$$E_a = (f_*)^{-1}((-\infty, a))$$
is a measurable set, and
$$F_b = (f^*)^{-1}((b, \infty))$$

is a measurable set. We then have

$$Y = Y_{a,b} = E_a \cap F_b$$

is a measurable set. To get that it is T-invariant, notice that

$$\begin{aligned} T^{-1}(Y) = \{x \in X : T(x) \in Y\} &= \{x \in X : f_*(T(x)) < a < b < f^*(T(x))\} \\ &= \{x \in X : f_*(x) < a < b < f^*(x)\} = Y. \end{aligned}$$

(3) Since $f \in L^1(\mu)$, we have that $\sigma(f)$ (the support of f) is σ-finite. Consequently $Y \subseteq \sigma(f)$, so Y must be σ-finite as well.

(4) Note that we can assume $b > 0$; if $b \leq 0$, then we can repeat the same kind of argument with $-f$ and $-a$ in place of f and b. Let $g = \chi_C$. Consider the function

$$h = f - bg.$$

Let

$$F = \{x \in X : A_n(h)(x) \geq 0 \text{ for some } n\}.$$

We note that $Y \subseteq F$. Let $x \in Y$. Since $b > 0$, we have that $f^*(x) > b$, which implies that for each $x \in Y$ there is some n so that $A_n(f)(x) > b$. In particular, there is some n so that $A_n(h)(x) \geq 0$.

Applying the Maximal Ergodic theorem, we have

$$\int_F h(x)dx \geq 0 \implies \int_F f(x)dx \geq b\mu(C).$$

Notice that

$$\int_F f(x)dx \leq \|f\|_1,$$

so we have

$$\mu(C) \leq \frac{\|f\|_1}{b}.$$

So every set of finite measure is uniformly bounded above by $\|f\|_1/b$, and hence $\mu(Y) < \infty$.

(5) Now apply the maximal Ergodic theorem to X with the function $h = f - b$. We see that the set E in this context will be Y. So

$$\int_Y (f(x) - b) \geq 0.$$

We can also apply it to $g = a - f$ to get

$$\int_Y (a - f(x)) \geq 0.$$

Putting these facts together, we have

$$\mu(Y)(a - b) \geq 0 \implies \mu(Y) = 0.$$

(6) Apply the result to all rational points to get $\mu(Y_{a,b}) = 0$ for all rational endpoints. Use limits to get that the average limits actually do converge. We get integrability and measurability by using Fatou's lemma;

$$\int \left| \frac{1}{n} \sum_{j=0}^{n-1} f_j(x) \right| dx \leq \frac{1}{n} \int \sum_{j=0}^{n-1} |f_j(x)| dx = \int |f(x)| dx < \infty.$$

(7) T-invariance follows from the fact that f_* and f^* are T-invariant.

(8) Finally we need to check the integrals are equal. If $f'(x) \geq a$, then there is at least one n so that for all $\epsilon > 0$,

$$\sum_{j=0}^{n-1}(f_j(x) - a + \epsilon) \geq 0.$$

Hence

$$\int f(x)dx \geq (a - \epsilon)\mu(X)$$

for each $\epsilon > 0$, and so

$$\int f(x)dx \geq a\mu(X).$$

A similar argument applies for $f'(x) \leq b$, giving us

$$\int f(x)dx \leq b\mu(X).$$

Write

$$X_{k,n} = \left\{x \in X : \frac{k}{2^n} \leq f'(x) \leq \frac{k+1}{2^n}\right\}.$$

Note that $X_{k,n}$ is T-invariant (since f' is). Thus we see

$$\frac{k}{2^n}\mu(X_{k,n}) \leq \int_{X_{k,n}} f(x)dx \leq \frac{k+1}{2^n}\mu(X_{k,n}),$$

$$\frac{k}{2^n}\mu(X_{k,n}) \leq \int_{X_{k,n}} f'(x)dx \leq \frac{k+1}{2^n}\mu(X_{k,n}).$$

Taking the difference and summing over k, we get

$$\left|\int f(x)dx - \int f'(x)dx\right| \leq \frac{1}{2^n}\mu(X).$$

The choice of n was arbitrary, so let it go to infinity.

□

Problem 76. Assume the setting of the individual ergodic theorem. If $f_* = f^*$ almost everywhere, show that the limit function f^* is integrable.

Proof. Let

$$E = \{x \in X : f_*(x) = f^*(x)\}.$$

Then on E we have

$$f^*(x) = \lim_{n\to\infty} A_n f(x).$$

Now Fatou gives us

$$\int_X |f^*|d\mu = \int_E |f^*|d\mu = \int_E \lim_{n\to\infty} |A_n f|d\mu \leq \liminf_{n\to\infty} \int_E |A_n f|d\mu \leq \liminf_{n\to\infty} \int_X |A_n f|d\mu.$$

Notice

$$|A_n f| = \left|\frac{1}{n}\sum_{j=0}^{n-1} f(T^j(x))\right| \leq \frac{1}{n}\sum_{j=0}^{n-1} |f(T^j(x))|,$$

so

$$\liminf_{n\to\infty} \int_X |A_n f|d\mu \leq \liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \int_X |f(T^j(x))|d\mu.$$

Since $(X, \mathcal{M}, \mu, T)$ is a measure preserving system, we have

$$\int_X |f(T(x))| d\mu = \int_{T^{-1}(X)} |f| d\mu = \int_X |f| d\mu = \|f\|_1.$$

Inducting gives us that

$$\int_X |f(T^j(x))| d\mu = \|f\|_1 \text{ for } j \geq 0.$$

Substituting this in, we have

$$\int_X |f^*| d\mu \leq \liminf_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \|f\|_1 = \|f\|_1 < \infty.$$

So f^* is integrable. □

Problem 77. Prove the following corollary of the Individual Ergodic Theorem: If $(X, \mathcal{M}, \mu, T)$ is a measure-preserving system of a probability measure space and if $f \in L^1(\mu)$, then

$$\int |A_n(f) - f^*| \, d\mu \to 0,$$

where

$$f^* = \lim_{n\to\infty} A_n(f).$$

Hint: First prove it for bounded f. Then approximate.

Proof. Recall the dominated convergence theorem:

Theorem (Dominated Convergence Theorem)**.** If (f_n) is a sequence of measurable functions which converges pointwise to a function f and is dominated by a function $g \in L^1(\mu)$, meaning

$$|f_n| \leq g \text{ for all } n,$$

then

$$\lim_{n\to\infty} \int |f_n - f| d\mu = 0.$$

We have $A_n(f) \to f^*$ pointwise. If f is bounded, say $|f| \leq M$, then

$$|A_n(f)| = \left| \frac{1}{n} \sum_0^{n-1} f \circ T^j \right| \leq \frac{1}{n} \sum_0^{n-1} |f \circ T^j| \leq \frac{1}{n} \sum_0^{n-1} M = M.$$

So $A_n(f)$ is dominated by M, and we apply the dominated convergence theorem (which applies since we are over a probability space), and we get the desired result.

Now assume f is not bounded. Using simple functions, we can approximate f by bounded functions. Consequently, we have

$$\|A_n(f) - f^*\|_1 \leq \|A_n(f-g)\|_1 + \|A_n(g) - g^*\|_1 + \|g^* - f^*\|_1.$$

As we've noted earlier,

$$\|A_n(f-g)\| \leq \|f - g\|_1.$$

On the other hand, we examine the L^1 norm on the far right. Using the last part of the Individual Ergodic theorem, we recall

$$\int g^* = \int g, \qquad \int f^* = \int f,$$

so

$$\int |g^* - f^*| = \int |g - f| = \|g - f\|_1.$$

So we can make these terms as small as we want, say $\epsilon/3$ for some $\epsilon > 0$ chosen arbitrarily. We may then take n as large as we want so the middle term is smaller than $\epsilon/3$ by the first part. This gives us

$$\|A_n(f) - f^*\|_1 < \epsilon$$

for all $\epsilon > 0$. Thus it goes to 0. □

Let $(X, \mathcal{M}, \mu, T)$ be a measure-preserving system on a probability space. We say that T is **ergodic** if for all $E \in \mathcal{M}$, we have

$$T^{-1}(E) = E \implies \mu(E) = 0 \text{ or } 1.$$

Problem 78. Let $(X, \mathcal{M}, \mu, T)$ be a measure-preserving system of a probability space. Show the following are equivalent.

(1) T is ergodic.
(2) The only members $E \in \mathcal{M}$ with $\mu(T^{-1}(E) \triangle E) = 0$ are those with $\mu(E) = 0$ or 1.
(3) For every $E \in \mathcal{M}$ with $\mu(E) > 0$, we have
$$\mu\left(\bigcup_{n=1}^{\infty} T^{-n}(E)\right) = 1.$$
(4) For every $A, B \in \mathcal{M}$ with $\mu(A) > 0$, $\mu(B) > 0$, there exists $n > 0$ with $\mu(T^{-n}(A) \cap B) > 0$.
(5) Whenever f is measurable and $U_T(f) = f \circ T = f$ for all $x \in X$, then f is constant almost everywhere.
(6) Whenever f is measurable and $U_T(f) = f$ almost everywhere, then f is constant almost everywhere.
(7) Whenever $f \in L^2(\mu)$ and $U_T(f) = f$ for all $x \in X$, then f is constant almost everywhere.
(8) Whenever $f \in L^2(\mu)$ and $U_T(f) = f$ almost everywhere, then f is constant almost everywhere.

Proof. (1) $\implies$ (2): Assume T is ergodic. Then if $T^{-1}(E) = E$ implies $\mu(E) = 0$ or 1. Assume now that $\mu(T^{-1}(E) \triangle E) = 0$. Notice for $n \geq 0$ we have

$$\mu(T^{-n}(E) \triangle E) = 0.$$

To see this, first see that

$$T^{-n}(E) \triangle E \subseteq \bigcup_{0}^{n-1} T^{-(j+1)}(E) \triangle T^{-j}(E).$$

Let $x \in T^{-n}(E) \triangle E$. This says that

$$x \in T^{-n}(E) \setminus E \text{ or } x \in E \setminus T^{-n}(E).$$

In other words, x is such that either x is in E but $T^n(x) \notin E$ or $T^n(x) \in E$ but x is not in E. This says that there is some point where $T^{-k}(x) \in E$ but $T^{-(k+1)}(x) \notin E$ or vice versa, which implies it is in the set on the right.

Now, recall

$$T^{-j}(E) \triangle T^{-j-k}(E) = T^{-j}\left(E \triangle T^{-k}(E)\right).$$

This is because preimages play nicely with unions and intersections. So using this fact, we have

$$T^{-n}(E) \triangle E \subseteq \bigcup_{0}^{n-1} T^{-(j+1)}(E) \triangle T^{-j}(E) = \bigcup_{0}^{n-1} T^{-j}\left(T^{-1}(E) \triangle E\right).$$

Now taking the measure, we have

$$\mu(T^{-n}(E) \triangle E) \leq n\mu(T^{-1}(E) \triangle E)$$

since the system is measure preserving, and since we assumed $\mu(T^{-1}(E)\triangle E) = 0$, this tells us that

$$\mu(T^{-n}(E)\triangle E) = 0.$$

Let

$$E_\infty = \bigcap_{n=0}^{\infty} \bigcup_{j=n}^{\infty} T^{-j}(E).$$

We have from the work above that

$$\mu(E_\infty \triangle E) \leq \mu\left(E\triangle \bigcup_{j=0}^{\infty} T^{-j}(E)\right) \leq \sum_{j=0}^{\infty} \mu(E\triangle T^{-j}(E)) = 0.$$

So $\mu(E_\infty) = \mu(E)$, and we see $T^{-1}(E_\infty) = E_\infty$ by construction, so $\mu(E_\infty) = 0$ or 1. Thus $\mu(E) = 0$ or 1.
(2) $\implies$ (3): Let

$$F = \bigcup_{n=1}^{\infty} T^{-n}(E).$$

Notice

$$T^{-1}(F) = \bigcup_{n=1}^{\infty} T^{-n-1}(E) \subseteq F.$$

Since $\mu(T^{-1}(F)) = \mu(F)$, we have $\mu(T^{-1}(F)\triangle F) = 0$. Now $T^{-1}(E) \subseteq F$, so

$$0 < \mu(E) = \mu(T^{-1}(E)) \leq \mu(F),$$

and (2) forces $\mu(F) = 1$.
(3) $\implies$ (4): Since $\mu(A) > 0$, (3) tells us that

$$F = \bigcup_{n=1}^{\infty} T^{-n}(A)$$

is such that $\mu(F) = 1$. Now $\mu(F \cap B) = \mu(B) > 0$, so

$$0 < \mu\left(\bigcup_{n=1}^{\infty} (T^{-n}(A) \cap B)\right) \leq \sum_{n=1}^{\infty} \mu(T^{-n}(A) \cap B),$$

so there must be some n with

$$0 < \mu(T^{-n}(A) \cap B).$$

(4) $\implies$ (1): If $E \in \mathcal{M}$ with $T^{-1}(E) = E$, $0 < \mu(E) < 1$, then

$$0 = \mu(E \cap E^c) = \mu(T^{-n}(E) \cap E^c) \text{ for } n \geq 1,$$

a contradiction.
(1) $\implies$ (6): Consider

$$E_a = \{x \in X : f(x) > a\}.$$

We see that

$$T^{-1}(E_a) = \{x \in X : f(T(x)) > a\}.$$

Since $U_T(f) = f$ almost everywhere, this gives us that (up to a set of measure zero which we take out)

$$T^{-1}(E_a) = \{x \in X : f(T(x)) = f(x) > a\} = E_a.$$

By ergodicity, we have $\mu(E_a) = 0$ or 1. This applies for all real $a \in \mathbb{R}$. If f non-constant, there would be some E_a with $0 < \mu(E_a) < 1$, a contradiction.

(6) $\implies$ (5): This is clear.
(5) $\implies$ (7): Functions which are L^2 are measurable, so this follows.
(6) $\implies$ (8): Same kind of argument as above.
(8) $\implies$ (7): This is clear.
(7) $\implies$ (1): Suppose $T^{-1}(E) = E$ with $E \in \mathcal{M}$. Then $\chi_E \in L^2(\mu)$, $\chi_E \circ T = \chi_E$ for all x, so χ_E is constant. Thus $\chi_E = 0$ or 1 almost everywhere, and

$$\mu(E) = \int \chi_E d\mu = 0 \text{ or } 1.$$

□

Problem 79. Prove that if T is an ergodic invertible transformation of a probability measure space, then T_A (the induced or derivative transformation) is ergodic.

Proof. Suppose $E \subseteq A$ a measurable set. We need to show that if $T_A^{-1}(E) = E$, then $\mu_A(E) = 0$ or 1. Recall

$$T_A^{-1}(E) = \bigcup_{n\geq 1} (A_n \cap T^{-n}(E)).$$

We will prove on induction that for $i \geq 0$ we have

$$T^i(E) \cap A \subseteq E.$$

In other words, if E under T comes back to A, it must also come back to E. For $i = 0$, we have

$$T^0(E) \cap A = E \cap A = E \subseteq E.$$

Now suppose we know this for $0 \leq i \leq n-1$. The goal is to show it holds for n. Fix some $k \in \mathbb{N}$. Recall

$$A_k = \{x \in A : n_A(x) = k\}.$$

Then

$$T^i(E \cap A_k) \cap A = \varnothing \text{ if } i < k.$$

Notice that

$$T^k(E \cap A_k) = T_A(E \cap A_k),$$

since by definition this is the set of elements in E which return to A at time k. E is invariant under T_A, so this tells us

$$T_A(E \cap A_k) \subseteq E.$$

Now take $i > k$. We see

$$T^i(E \cap A_k) \cap A = T^{i-k}(T^k(E \cap A_k)) \cap A = T^{i-k}(T_A(E \cap A_k)) \cap A.$$

From our prior calculation this tells us

$$T^{i-k}(T_A(E \cap A_k)) \cap A \subseteq T^{i-k}(E) \cap A.$$

This holds for all k. Since $k \geq 1$, $i - k \leq n - 1$. So the induction hypothesis applies to tell us

$$T^{i-k}(E) \cap A \subseteq E.$$

Now

$$X_E = \bigcup_{i\geq 0} T^i E$$

is T-invariant, so X_E is null or conull by ergodicity. If it is null, we win (since $E \subseteq X_E$). If it is conull, then $A \setminus (X_E \cap A)$ is null, and $X_E \cap A = E$ by our claim above. So $A \setminus E$ is null. □

Problem 80. Suppose $(X, \mathcal{M}, \mu, T)$ is an invertible ergodic system on a probability measure space. If $A \in \mathcal{M}$ with $\mu(A) > 0$, then

$$\int_A n_A d\mu = 1.$$

Proof. Recall

$$A_n = \{x \in A : n_A(x) = n\}.$$

Let

$$F_n = A_n \cup TA_n \cup \cdots \cup T^{n-1} A_n.$$

Using the Kakutani diagram, we see

$$X = \bigcup_{n \geq 1} F_n.$$

Now, notice that

$$X = \bigcup_{n \geq 1} F_n \text{ almost everywhere.}$$

Since T is measure preserving and invertible, we have

$$\mu(F_n) = n\mu(A_n).$$

So

$$\sum_{n \geq 1} \mu(F_n) = \sum_{n \geq 1} n\mu(A_n) = \mu(X) = 1.$$

Finally, notice that we can decompose A into

$$A = \bigcup_{n \geq 1} A_n,$$

and

$$\int_A n_A(x) d\mu(x) = \sum_{n \geq 1} \int_{A_n} n_A(x) d\mu(x).$$

On A_n, we see $n_A(x) = n$, so this is equal to

$$\int_A n_A(x) d\mu(x) = \sum_{n \geq 1} n \int_{A_n} d\mu(x) = \sum_{n \geq 1} n\mu(A_n) = 1.$$

□

Problem 81. Prove that

$$T : [0, 1) \to [0, 1), \qquad T(x) \equiv 2x \pmod 1$$

is an ergodic transformation.

Proof. Use the L^2 equivalence. Let $f \in L^2(\mu)$ be such that $f \circ T = f$ almost everywhere. Then if we write

$$f(x) = \sum a_n e^{-2\pi i n x}$$

as the Fourier series, we have

$$f(T(x)) = \sum a_n e^{-2\pi i n 2x}.$$

These series are equal almost everywhere, so this says that $|a_n| = |a_{2n}| = |a_{4n}| = \cdots$, so by Riemann Lebesgue we get that this is 0 unless $n = 0$. □

Problem 82. Let X be a compact metric space, $\mathcal{B}$ the Borel σ-algebra, and let μ be a probability measure with the property that $\mu(U) > 0$ for all non-empty open U. Suppose $T : X \to X$ is a continuous measure preserving transformation which is also ergodic. Then almost all points of x have dense orbit under T.

Proof. Suppose $x \in X$ does not have a dense orbit $\mathcal{O}_T(x) = \{T^n(x) : n \geq 0\}$. The base for the topology of X is countable, so we have $\{U_n\}_{n=1}^{\infty}$ is a base. Since the orbit is not dense, there is a U_j with the property that $\mathcal{O}_T(x) \cap U_j = \varnothing$. Let $f = \chi_{U_j}$. Then

$$\frac{1}{n}\sum_{0}^{n-1} \chi_{U_j} \circ T^k(x) = 0.$$

In the proof of Birkhoff ergodic theorem, we have that for almost every x this series converges to the measure of U_j which should be greater than 0. So the collection of all points $x \in X$ whose orbit does not intersect U_J has measure 0. Let E_j be the set of these points. Now

$$E = \bigcup_{j=1}^{\infty} E_j = \{x \in X : x \text{ does not have a dense orbit}\}.$$

Notice $\mu(E) = 0$. So almost every point has a dense orbit. □

Problem 83. Prove the following corollary of the Individual Ergodic theorem:

Theorem (Borel's Normal Number theorem)**.** Almost all numbers in $[0, 1)$ are normal to base 2. That is, the frequency of 1's in the binary expansion of almost every number in $[0, 1)$ is $1/2$.

Hint: Construct a system and a function which measures the frequency.

Proof. Let $\mathcal{B}$ be the Borel σ-algbera on $[0, 1)$. Then we consider the measure-preserving system $([0, 1), \mathcal{B}, \lambda, T)$, where $T : [0, 1) \to [0, 1)$ is

$$T(x) \equiv 2x \pmod 1.$$

□

Problem 84. Let $(X, \mathcal{M}, \mu, T)$ be a measure preserving probability space. Show the following are equivalent.

(1) T is ergodic.
(2) For all $A, B \in \mathcal{M}$, we have

$$\frac{1}{n}\sum_{0}^{n-1} \mu(T^{-j}(A) \cap B) \to \mu(A)\mu(B).$$

Proof. (1) $\implies$ (2): Assume T is ergodic. Consider χ_A, χ_B. These are measurable functions, and moreover

$$\chi_A \circ T^j = \chi_{T^{-j}(A)}.$$

Notice that

$$(\chi_A \circ T^j) \cdot \chi_B = \chi_{T^{-j}(A) \cap B}.$$

We use the Birkhoff ergodic theorem and the fact that T is ergodic to get

$$\frac{1}{n}\sum_{0}^{n-1} \chi_A \circ T^j \to \int \chi_A d\mu = \mu(A).$$

Notice the Birkhoff ergodic theroem only tells us that the thing on the left converges to some function f^*. We know this f^* is T-invariant, so f^* is constant almost everywhere by ergodicity. Furthermore, since we have a probability measure space, we know that

$$\int f^* d\mu = \int f d\mu.$$

So $f^* = C$ for some C constant, and

$$C \int d\mu = C\mu(X) = C = \int f d\mu.$$

Thus

$$f^* = \int f d\mu \text{ almost everyhwere,}$$

and we have the desired above result.

Now we can mulitply χ_B to get

$$\frac{1}{n} \sum_{0}^{n-1} (\chi_A \circ T^j)\chi_B \to \mu(A)\chi_B \text{ almost everywhere.}$$

Integrating both sides grants us

$$\int \frac{1}{n} \sum_{0}^{n-1} (\chi_A \circ T^j)\chi_B d\mu \to \mu(A)\mu(B).$$

For each n, notice that linearity gives

$$\int \frac{1}{n} \sum_{0}^{n-1} (\chi_A \circ T^j)\chi_B d\mu = \frac{1}{n} \sum_{0}^{n-1} \int \left((\chi_A \circ T^j)\chi_B d\mu \right) = \frac{1}{n} \sum_{0}^{n-1} \mu(T^{-j}(A) \cap B).$$

Thus we have the desired result.

(2) $\implies$ (1): Assume now $A = B = E$, and E is T-invariant (meaning $T^{-1}(E) = E$). Then

$$\frac{1}{n} \sum_{0}^{n-1} \mu(T^{-j}(E) \cap E) = \frac{1}{n} \sum_{0}^{n-1} \mu(E) = \mu(E) = \mu(E)^2.$$

This forces either $\mu(E) = 0$ or $\mu(E) = 1$. □

Remark. This same kind of result can be proven for just a generating algebra.

Problem 85. Show that the two sided shift $(p_0, \dots, p_{k-1})$ shift is ergodic.

Proof. Recall that our space is $X = A^{\mathbb{Z}}$, and our measure is defined by the property that if $F \in C(X)$ depends on coordinates $-N$ to N, then

$$\int_X F d\mu := \int_{A^{2N+1}} \left(F|_{\prod_{-N}^{N} A} \right) dp^{2N+1}.$$

Let

$$\sigma((x_k)) = (y_k), \qquad \text{where } y_k = x_{k+1}.$$

The goal is to show that σ is ergodic. Let $E, F \subseteq X$ be subsets which depend only on finitely many coordinates, say $-N$ to N. Then

$$\sigma^{-i}(E) = \{(y_k) : y_{j+i} \in E \text{ for } j \in \{-N, \dots, N\}\}.$$

Now

$$\chi_E \circ \sigma^i = \chi_{\sigma^{-i}(E)}$$

depends on $N-i$ to $N+i$ coordinates. Now if $|i| \geq 2N+1$, then χ_E and χ_F depend on different coordinates entirely, so

$$\mu(\sigma^{-i}(E) \cap F) = \int \chi_{\sigma^{-i}(E)} \chi_F d\mu = \mu(\sigma^{-i}(E))\mu(F) = \mu(E)\mu(F).$$

Now use the prior problem. □

Problem 86. Let $(X, \mathcal{M}, \mu, T)$ be a measure-preserving system of a probability space. Prove the following are equivalent:

(1) T is ergodic.
(2) For all $f, g \in L^2(\mu)$,

$$\frac{1}{n} \sum_{0}^{n-1} \langle U_T^i f, g \rangle \to \langle f, 1 \rangle \langle 1, g \rangle.$$

(3) For all $f \in L^2(\mu)$,

$$\frac{1}{n} \sum_{0}^{n-1} \langle U_T^i f, f \rangle \to \langle f, 1 \rangle \langle 1, f \rangle.$$

Proof. (1) $\implies$ (2): Assume T is ergodic. By the Birkhoff ergodic theorem, we know

$$\frac{1}{n} \sum_{0}^{n-1} U_T^i f \to \int f d\mu = \langle f, 1 \rangle.$$

Notice that, since we are on a Hilbert space, $\langle \cdot, g \rangle = g^*(\cdot)$ is a bounded linear functional. By continuity, we get

$$g^* \left(\frac{1}{n} \sum_{0}^{n-1} U_T^i f \right) \to g^*(\langle f, 1 \rangle) = \langle f, 1 \rangle \langle 1, g \rangle.$$

Using the fact that it's linear, we have

$$g^* \left(\frac{1}{n} \sum_{0}^{n-1} U_T^i f \right) = \frac{1}{n} \sum_{0}^{n-1} \langle U_T^i f, g \rangle.$$

(2) $\implies$ (3): Set $f = g$.
(3) $\implies$ (1): Take $f = \chi_E$ where E is T-invariant. We know

$$\langle U_T^i \chi_E, \chi_E \rangle = \int U_T^i \chi_E \cdot \chi_E d\mu = \mu(T^{-i}(E) \cap E) = \mu(E).$$

By (3), this is equal to $\mu(E)^2$. This forces $\mu(E) = 0$ or 1. □

Let $\sigma(T)$ denote the **spectrum** of a bounded operator T. That is, we define

$$\sigma(T) = \{\lambda \in \mathbb{C} : T - \lambda I \text{ does not have an inverse}\}.$$

Problem 87 (Bounded inverse theorem)**.** Show that if $T - \lambda I$ does have an inverse, it must be bounded.

Proof. Suppose T is a bounded linear operator (on a Hilbert space). The goal is to show that if T has an inverse, it's inverse is bounded. We will do this with the closed graph theorem. Let $(x_n, T^{-1}(x_n)) \to (x, y)$. The goal is to show that $T(x) = y$. Let $y_n = T^{-1}(x_n)$. Then

$$(y_n, T(y_n)) \to (y, T(y))$$

since T is bounded. But this says that $T(T^{-1}(x_n)) = x_n \to T(y)$ and $x_n \to x$. So we have $x = T(y)$, and taking the inverse this forces $T^{-1}(x) = y$. Thus we have that T^{-1} is bounded. □

Using the last problem, we can equivalently write

$$\sigma(T) = \{\lambda \in \mathbb{C} : T - \lambda I \text{ does not have a bounded inverse}\}.$$

Problem 88. Show the following.

(1) If the operator T is invertible, then
$$\sigma(T^{-1}) = (\sigma(T))^{-1} = \{\lambda^{-1} : \lambda \in \sigma(T)\}.$$

(2) If T is an operator, then
$$\sigma(T^*) = (\sigma(T))^* = \{\lambda^* : \lambda \in \sigma(T)\}.$$

Proof.

(1) First note that T invertible implies that $0 \notin \sigma(T)$. So inverting the set makes sense. Next, let $\lambda \in \sigma(T^{-1})$. Then $T^{-1} - \lambda I$ is invertible. But we can write this as
$$T^{-1} - \lambda I = (\lambda^{-1} I - T)(\lambda T^{-1}).$$
So $(\lambda^{-1} I - T)$ is invertible, and hence $\lambda^{-1} \in \sigma(T)$. This tells us that $\sigma(T^{-1}) \subseteq \sigma(T)^{-1}$. Now apply what we've done to T^{-1} to get $\sigma(T) \subseteq \sigma(T^{-1})^{-1}$ or $\sigma(T)^{-1} \subseteq \sigma(T^{-1})$.

(2) Suppose $\lambda \in \sigma(T)$ so that $T - \lambda I$ is invertible. Then $(T - \lambda I)^* = T^* - \lambda^* I^* = T^* - \lambda^* I$ is invertible as well, so that $\lambda^* \in \sigma(T^*)$. This tells us $\sigma(T)^* \subseteq \sigma(T^*)$. Now use the same argument on T^* to get $\sigma(T^*)^* \subseteq \sigma(T)$. Applying $*$ to both sides gives $\sigma(T^*) \subseteq \sigma(T)^*$ and we're done.

□

A **Hermitian operator** is one on a Hilbert space which satisfies the condition that
$$\langle f, Tg \rangle = \langle Tf, g \rangle.$$
A **normal operator** is an operator T on a Hilbert space which satisfies the condition that
$$TT^* = T^*T.$$

Problem 89. Show that a Hermitian operator is normal.

Proof. We see that
$$\langle f, Tg \rangle = \langle Tf, g \rangle = \langle f, T^*g \rangle,$$
so for all $g \in H$ we have
$$Tg = T^*g \implies T = T^*.$$
Now $TT^* = T^2 = T^*T$. □

An operator T is **bounded from below** if $\|Tx\| \geq C\|x\|$ for all $x \in H$.

Problem 90. Show that if T is a bounded operator on a Hilbert space, then it is injective if and only if it is bounded from below.

Proof. ($\implies$) Assume it is injective. Notice that $T : H \to \operatorname{Im}(T)$ is a bijective bounded continuous map. Thus T is open by the open mapping theorem. Now being an open mapping implies that $T^{-1} : \operatorname{Im}(T) \to H$ is continuous as well, so $\|T^{-1}x\| \leq C\|x\|$ for some $C > 0$ and all $x \in \operatorname{Im}(T)$. This means that $\|Tx\| \geq \frac{1}{C}\|x\|$ for all $x \in H$. This implies T is bounded from below.
($\impliedby$): Assume that T is bounded from below. Then if $Tx = 0$ if and only if $\|Tx\| = 0$, and since for all non-trivial x we have $\|Tx\| > 0$ this forces $x = 0$. □

Problem 91. Show that if T is an injective bounded operator, then $\operatorname{Im}(T) \subseteq H$ is a Banach space. Deduce it is closed.

Proof. By the prior problem, we know that T injective iff T is bounded from below. Let $T : H \to \text{Im}(T)$. Let $(y_n) \subseteq \text{Im}(T)$ be a Cauchy sequence. We know that $y_n \to y \in H$. The goal is to show that $y \in \text{Im}(T)$. Notice we can define $T^{-1} : \text{Im}(T) \to T$. Thus

$$\|y_n - y_m\| = \|TT^{-1}(y_n) - TT^{-1}(y_m)\| \geq C\|T^{-1}(y_n) - T^{-1}(y_m)\|.$$

Since (y_n) Cauchy, we have $(T^{-1}(y_m))$ Cauchy in H, so it converges to some $x \in H$. Notice that

$$T(x) = \lim T(T^{-1}(y_n)) = \lim y_n = y.$$

□

For a subspace $K \subseteq H$, we define

$$K^{\perp} = \{x \in H : \langle x, y\rangle = 0 \text{ for all } y \in K\}.$$

Problem 92. Show that $K^{\perp}$ is a closed subspace.

Proof. The fact that it is a subspace is clear. Let $(y_n) \subseteq K^{\perp}$, and suppose $y_n \to y$. The goal is to show that $y \in K^{\perp}$. Let $z^* = \langle \cdot, z\rangle$. This is a bounded linear operator, and furthermore $z^*(y_n) = 0$ for all n. Taking the limit, we have $z^*(y) = 0$. So $y \in K^{\perp}$. □

Problem 93 (Parallelogram identity)**.** Prove that for all $x, y \in H$, we have

$$\|x - y\|^2 + \|x + y\| = 2\left[\|x\|^2 + \|y\|^2\right]$$

Proof. Notice

$$\|x - y\|^2 = \langle x - y, x - y\rangle = \|x\|^2 - \langle x, y\rangle - \langle y, x\rangle + \|y\|^2,$$
$$\|x + y\|^2 = \langle x + y, x + y\rangle = \|x\|^2 + \langle x, y\rangle + \langle y, x\rangle + \|y\|^2.$$

Adding these together, we have

$$\langle x - y, x - y\rangle + \langle x + y, x + y\rangle = 2\|x\|^2 + 2\|y\|^2.$$

□

Problem 94. Let $C \subseteq H$ a closed convex subset and $z \notin C$. Then there is a unique $x \in C$ so that

$$\|x - z\| = \inf_{y \in C} \|y - z\|.$$

In other words, this infimum is achieved.

Proof. Let's first prove it for $z = 0 \notin C$. If $(x_n) \subseteq C$ is such that

$$\|x_n\| \to r = \inf_{y \in C} \|y\|,$$

then we can apply the parallelogram identity to $x_n/2, x_m/2$. This gives us

$$\left\|\frac{x_n - x_m}{2}\right\|^2 + \left\|\frac{x_n - x_m}{2}\right\| = 2\left[\left\|\frac{x_n}{2}\right\|^2 + \left\|\frac{x_m}{2}\right\|^2\right].$$

Simplifying, we have

$$\|x_n - x_m\|^2 = 2\|x_m\|^2 + 2\|x_n\|^2 - 4\|\frac{x_m + x_n}{2}\|^2.$$

We see $\|x_m\| \to r$, $\|x_n\| \to r$, and $\|(x_m + x_n)/2\| \geq r$. So therefore

$$\limsup_{m,n} \|x_n - x_m\|^2 = 0.$$

This tells us (x_n) is Cauchy, so $x_n \to x$. Since C closed, $x \in C$. For uniqueness, suppose $z \in C$ satisfies $\|z\| = r$. Then the sequence (y_k) where $y_k = x$ if k even and $y_k = z$ if k is odd is Cauchy. The only thing it could converge to is 0, giving us the result. □

Problem 95. Prove that $H = \overline{M} \oplus M^{\perp}$.

Proof. First let's show that $M \cap M^{\perp} = \{0\}$. Notice $x \in M \cap M^{\perp}$ satisfies $\langle x, x\rangle = 0$ which means $\|x\| = 0$, or $x = 0$.

Let $\delta = \inf_{y \in \overline{M}} \|x - y\|$. Let y be the point minimizing this distance, so $\|x - y\| = \delta$. Take $u \in \overline{M}$, let $z = x - y$. The goal is to show that $\langle z, u\rangle = 0$. Scale u so that $\langle z, u\rangle$ is real. Now

$$f(t) = \|z + tu\|^2 = \|z\|^2 + 2t\langle z, u\rangle + t^2\|u\|^2.$$

This is a real function, and it has a minimum at $t = 0$. This follows since

$$z + tu = x - (y - tu),\ y - tu \in \overline{M},\ f(t) = \|x - (y - tu)\|^2 \geq \delta^2.$$

We see that $f(0) = \delta^2$ as well. Now

$$f'(t) = 2\langle z, u\rangle + 2t\|u\|^2.$$

If we plug in $t = 0$, we have

$$f'(0) = 2\langle z, u\rangle = 0,$$

since $t = 0$ was a critical point. This tells us the result. If we scaled u by say $\alpha \neq 0$, then we see that

$$2\alpha\langle z, u\rangle = 0,$$

so this doesn't change anything. □

Problem 96. Prove that

$$(K^{\perp})^{\perp} = \overline{K}.$$

Proof. It's a matter of checking definitions to see $K \subseteq (K^{\perp})^{\perp}$, so $\overline{K} \subseteq (K^{\perp})^{\perp}$. For the other direction, it follows easily from the decomposition. We see $K^{\perp} \cap (K^{\perp})^{\perp} = \{0\}$, and so $(K^{\perp})^{\perp} \subseteq H$ implies $(K^{\perp})^{\perp} \subseteq \overline{K}$. □

Problem 97. Show that if T is an operator on a Hilbert space, then

(1) $\ker(T^*) = (\mathrm{Im}(T))^{\perp}$;
(2) $\overline{\mathrm{Im}(T)} = (\ker(T^*))^{\perp}$.

Proof.

(1) Let $y \in \ker(T^*)$. Then for all $z \in \mathrm{Im}(T)$, we have $T(x) = z$ for some x and

$$\langle y, z\rangle = \langle y, T(x)\rangle = \langle T^*(y), x\rangle = \langle 0, x\rangle = 0.$$

So $y \in (\mathrm{Im}(T))^{\perp}$, giving $\ker(T^*) \subseteq \mathrm{Im}(T)$. Now let $z \in \mathrm{Im}(T)^{\perp}$. Then for all $x \in H$, we see $T(x) \in \mathrm{Im}(T)$ and we have

$$\langle T(x), z\rangle = \langle x, T^*(z)\rangle = 0.$$

This forces $T * (z) = 0$, so $z \in \ker(T^*)$.

(2) Apply (1) to T^*. We see $(T^*)^* = T$, so $\ker(T) = (\mathrm{Im}(T^*)^{\perp}$. Applying $\perp$ to both sides gives $\ker(T)^{\perp} = ((\mathrm{Im}(T^*)^{\perp})^{\perp} = \overline{\mathrm{Im}(T^*)}$.

□

Problem 98. Show the following are equivalent:

(1) T is invertible.
(2) There is a constant $\alpha > 0$ so that $T^*T \geq \alpha I$ and $TT^* \geq \alpha I$.

Proof. (1) $\implies$ (2): Assume T is invertible. Then T^* is also invertible. Define

$$\|T^{-1}\|^{-2} = \alpha = \|(T^*)^{-1}\|^{-2}.$$

Notice that

$$\|x\| = \|T^{-1}(T(x))\| \leq \|T^{-1}\|\|Tx\|,$$

so

$$\|Tx\| \geq \|T^{-1}\|^{-1}\|x\|.$$

Now

$$\langle T^*Tx, x\rangle = \langle Tx, Tx\rangle = \|Tx\|^2 \geq \|T^{-1}\|^{-2}\|x\|^2 = \alpha\langle x, x\rangle = \langle \alpha x, x\rangle.$$

So

$$\langle (T^*T - \alpha I)x, x\rangle \geq 0.$$

We have the result.
(2) $\implies$ (1): Notice that

$$\|Tx\|^2 = \langle Tx, Tx\rangle = \langle T^*Tx, x\rangle \geq \alpha\langle x, x\rangle = \alpha\|x\|^2,$$

so

$$\|Tx\| \geq \sqrt{\alpha}\|x\|.$$

This shows that T is injective. This shows that $\text{Im}(T)$ is closed, and we use the prior problem to deduce that $\text{Im}(T) = (\ker(T^*))^{\perp}$. Now from the exact same argument,

$$\|T^*x\| \geq \sqrt{\alpha}\|x\|,$$

so T^* is injective and $\ker(T^*) = \{0\}$. Thus T is surjective. □

Problem 99. Show that if T is a bounded operator, then TT^* and T^*T are both positive.

Proof. This follows from the following observations:

$$\langle TT^*x, x\rangle = \langle T^*x, T^*x\rangle = \|T^*x\|^2 \geq 0,$$
$$\langle T^*Tx, x\rangle = \langle Tx, Tx\rangle = \|Tx\|^2 \geq 0.$$

□

Problem 100. Show that every Hermitian has real spectrum (i.e. $\sigma(T) \subseteq \mathbb{R}$).

Proof. Suppose T is normal. Let $\lambda \in \mathbb{C} \setminus \mathbb{R}$ so $\lambda = a + bi$ with $b \neq 0$. Notice that if $X = T - \lambda I$, we have

$$\begin{aligned} X^*X = XX^* = (T - \lambda I)(T - \lambda^* I) &= |\lambda|^2 I - 2(\text{Re}(\lambda))T + T^2 \\ &= (a^2 + b^2)I - 2aT + T^2 = b^2 I + (aI - T)^2. \end{aligned}$$

Now

$$(aI - T)^2 = (aI - T)(aI - T)^*$$

is a positive operator, so we have that $XX^* \geq b^2 I$. Hence X is invertible. □

Problem 101. Show that if T is a Hermitian operator, then $\|T\| = \sup\{|\lambda| : \lambda \in \sigma(T)\}$.

Proof. TODO □

Let T be an operator, f a comple-valued function on $\sigma(T)$. We define

$$N_T(f) = \sup\{|f(\lambda)| : \lambda \in \sigma(T)\}.$$

Problem 102. Show that if T is a Hermitian operator and p is a real polynomial, then

$$\|p(T)\| = N_T(p).$$

Proof. TODO □

Suppose M is a closed subspace of H. From **Problem 38** we know we can decompose the space H into a direct sum $H = M \oplus M^\perp$. We define a projection of M to be a mapping $P : H \to H$ where it takes each vector $v = v_1 + v_2 \in H$ and maps it to v_1. Note that P is necessarily an operator.

Problem 103.

(1) Show that an operator P is a projection if and only if it is Hermitian and idempotent (meaning $P^2 = P$).

(2) Show that for a projection operator we have

$$\|Pv\|^2 = \langle Pv, v\rangle.$$

Proof.

(1) ($\implies$): If P is a projection, then $P : H \to H$ is defined by some closed subspace M. We first note that it is idempotent, since $P(v) = P(v_1 + v_2) = v_1$ and $P(v_1) = P(v_1 + 0) = v_1$, so $P^2(v) = P(v)$ for all $v \in H$.

Next, we claim that it is Hermitian. Let $v = v_1 + v_2, w = w_1 + w_2 \in H$. Then

$$\langle Pv, w\rangle = \langle v_1, w_1 + w_2\rangle = \langle v_1, w_1\rangle + \langle v_1, w_2\rangle = \langle v_1, w_1\rangle = \langle v_1, w_1\rangle + \langle v_2, w_1\rangle = \langle v, Pw\rangle.$$

($\impliedby$) : We now show that if P is idempotent and Hermitian, then you are a projection mapping for some closed subspace. Let

$$M = \{w \in H : P(w) = w\}.$$

We claim that $H = M \oplus M^\perp$ (so that M is a closed subspace). Take $v \in H$, then this is equivalent to showing

$$\langle Pv, v - Pv\rangle = 0.$$

This is just a calcuation:

$$\langle Pv, v - Pv\rangle = \langle Pv, v\rangle - \langle Pv, Pv\rangle = \langle Pv, v\rangle - \langle Pv, v\rangle = 0.$$

(2) Notice

$$\|Pv\|^2 = \langle Pv, Pv\rangle = \langle Pv, v\rangle.$$

□

Denote by $\mathcal{P}(H)$ the collection of all projection operators on H. We can give this a partial ordering by $P_i \le P_j$ if $M_i \subseteq M_j$ (where we note from this last problem that a projection operator is uniquely characterized by it's fixed points). We can define

$$\sum_{i\in I} P_i \in \mathcal{P}(H) \text{ is the operator with fixed points } \bigcup_{i\in I} M_i.$$

Let X be a set and $\mathcal{M}$ a σ-algebra on it. A **spectral measure** is a function $E : \mathcal{M} \to \mathcal{P}(H)$ satisfying the following properties:

(1) $E(\varnothing) = 0$ and $E(H) = \mathrm{Id}$ (here, these are the 0 operator and the identity operator).

(2) If $\{U_n\} \subseteq \mathcal{M}$ are disjoint, then

$$E\left(\bigcup U_n\right) = \sum E(U_n).$$

Problem 104. Show that condition (1) is slightly superfluous in the sense that we only require $E(H) = 1$ and condition (2). That is, deduce $E(\varnothing) = 0$ from $E(H) = 1$ and condition (2).

Proof. Notice that for any $M \subseteq H$ we have $E(H - M) = E(H) - E(M)$. This follows from the fact that $E(H - M) + E(M) = E(H) = 1$. Now if we take $M = H$, we have $E(H - H) = E(\varnothing) = E(H) - E(H) = \mathrm{Id} - \mathrm{Id} = 0$. □

Problem 105.

(1) Prove that E is modular, meaning

$$E(M \cup N) + E(M \cap N) = E(M) + E(N).$$

(2) Prove that

$$E(M)E(M \cap N) = E(M \cap N) \text{ and } E(M)E(M \cup N) = E(M).$$

(3) Prove that E is multiplicative, meaning

$$E(M \cap N) = E(M)E(N).$$

Proof.

(1) We can write

$$M \cup N = (M - N) \sqcup (M \cap N) \sqcup (N - M).$$

So

$$E(M \cup N) = E(M - N) + E(M \cap N) + E(N - M).$$

Now add $E(M \cap N)$ to both sides.

$$E(M \cup N) + E(M \cap N) = (E(M - N) + E(M \cap N)) + (E(N - M) + E(M \cap N)) = E(M) + E(N).$$

(2) We have

$$E(M \cap N) \leq E(M) \leq E(M \cup N),$$

where this is with respect to the partial ordering on fixed spaces. The claim then follows from observing that if $P_0 \leq P_1$, then $P_1 P_0 = P_0$. Examine the fixed space of $P_1 P_0$. This is going to be

$$M = \{v \in H : (P_1 P_0)(v) = v\} = \{v \in H : P_1(P_0(v)) = v\}.$$

Write $v = v_1 + v_2$, where $v_1 \in M_1$ which is the fixed space of P_1. Then using the fact that $P_0 \leq P_1$ so $M_1^{\perp} \subseteq M_2^{\perp}$, where M_2 is the fixed space of P_0, we have

$$P_1(P_0(v_1 + v_2)) = P_1(P_0(v_1) + P_0(v_2)) = P_1(P_0(v_1)) + P_1(P_0(v_2)) = P_1(P_0(v_1)) = v.$$

Now take $v_1 = w_1 + w_2$, $w_1 \in M_2$. Applying the same reasoning gives

$$P_1(P_0(v)) = w_1 = v.$$

So its the collection of all v contained in M_2. This means that $P_1 P_0$ is the projection operator with fixed space M_2. It is uniquely characterized by this, so $P_1 P_0 = P_0$. The same kind of argument also gives $P_0 P_1 = P_0$. Using this, we have

$$E(M)E(M \cap N) = E(M \cap N) \text{ and } E(M)E(M \cup N) = E(M).$$

(3) Now multiply both sides of the modular equation by $E(M)$ to get

$$E(M) + E(M \cap N) = E(M) + E(M)E(N).$$

Note here we used the property that projection operators are idempotent. Subtract to get the result.

□

Problem 106. Show the following are equivalent for $E : \mathcal{M} \to \mathcal{P}(H)$.

(1) E is a spectral measure.
(2) $E(H) = 1$ and for any pairs $x, y \in H$ the complex set-valued function μ defined by

$$\mu_{x,y}(M) = \langle E(M)x, y \rangle$$

is countably additive.

Proof. (1) $\implies$ (2): The first part follows by definition of a spectral measure. Inner products with an infinite sum can be formed term by term, so that gives us the second part.
(2) $\implies$ (1): We see for arbitrary $x, y \in H$ we have

$$\mu_{x,y}(M \sqcup N) = \langle E(M \sqcup N)x, y\rangle = \mu_{x,y}(M) + \mu_{x,y}(N) = \langle E(M)x, y\rangle + \langle E(N)x, y\rangle = \langle (E(M) + E(N))x, y\rangle.$$

This holds for arbitrary choice of x and y, so we get $E(M \sqcup N) = E(M) + E(N)$. Notice we also get multiplicativity by the argument in the prior problem.

We now wish to extend this to countable unions. Here we need to be careful, as the sum doesn't quite make sense without some argument. Suppose $\{M_n\} \subseteq \mathcal{M}$ is a disjoint sequence of sets with

$$\bigsqcup_n M_n = M.$$

Then by the prior paragraph, $\{E(M_n)\}$ is a sequence of orthogonal projections. So $\{E(M_n)x\rangle$ is a sequence of orthogonal vectors for any choice of $x \in H$. Now

$$\sum_n \|E(M_n)x\|^2 = \sum_n \langle E(M_n)x, x\rangle = \langle E(M)x, x\rangle = \|E(M)x\|^2.$$

The sequence $\{E(M_n)x\}$ is thus summable. If we set

$$\sum_n E(M_n)x = Ax,$$

then $A : H] \to H$ is a bounded linear operator. Thus the sum makes sense, and we can make the calculation

$$\langle E(M)x, y\rangle = \sum_n \langle E(M_n)x, y\rangle = \langle \sum_n E(M_n)x, y\rangle.$$

□

We will use the symbol $\mathcal{B}$ to denote the class of complex-valued bounded measurable functions on X. We will write

$$N(f) = \sup\{|f(\lambda)| : \lambda \in X\}$$

whenever $f \in \mathcal{B}$. *Warning: do not confuse $\mathcal{B}$ with the Borel σ-algebra!*

Problem 107. Show that if E is a spectral measure and if $f \in \mathcal{B}$, then there exists a unique operator A such that

$$\langle Ax, y\rangle = \int f(\lambda) d(\langle E(\lambda)x, y\rangle)$$

for every pair of vectors x, y. The dependence of A on f and E will be denoted by writing

$$A = \int f dE.$$

Proof. The function

$$\varphi(x, y) = \int f(\lambda) d(\langle E(\lambda)x, y\rangle)$$

makes sense for ever yx and y by the boundedness of f. We see that φ is a bilinear functional, and furthermore it is bounded by

$$|\varphi(x, y)| \leq \int |f(\lambda) d(\|E(\lambda)x\|^2) \leq N(f)\|x\|^2.$$

This gives us a unique operator. □

This gives us some nice properties.

Problem 108. Show the following properties for a spectral measure E, $f, g \in \mathcal{B}$, and $\alpha \in \mathbb{C}$:

(1) We have

$$\int (\alpha f) dE = \alpha \int f dE.$$

(2) We have

$$\int (f+g) dE = \int f dE + \int g dE.$$

(3) We have

$$\int f^* dE = \left(\int f dE\right)^*.$$

(4) We have

$$\left(\int f dE\right)\left(\int g dE\right) = \int f g dE.$$

(5) If B is an operator that commutes with E, then B commutes with A on the level of

$$\langle ABx, y\rangle = \langle BAx, y\rangle.$$

Proof. The proof of all of these is essentially the same. I'll prove (1), (3), and (4). We follow Halmos.

(1) By the prior problem, we have operators A, A_α such that

$$A = \int f dE, \qquad A_\alpha = \int (\alpha f) dE.$$

The goal is to show $\alpha A = A_\alpha$. This is shown by taking arbitrary x, y, and noting that

$$\langle \alpha A x, y\rangle = \alpha \int f(\lambda) d(\langle E(\lambda x), y\rangle) = \int \alpha f(\lambda) d(\langle E(\lambda x), y\rangle) = \langle A_\alpha x, y\rangle.$$

(3) Let

$$A = \int f dE, \qquad B = \int f^* dE.$$

The goal is to show that $A^* = B$. That is, for all x, y we have

$$\langle Ax, y\rangle = \langle x, By\rangle.$$

Notice

$$\langle x, By\rangle = \langle By, x\rangle^* = \left(\int f^*(\lambda) d(\langle E(\lambda)x, y\rangle)\right)^* = \int f(\lambda) d(\langle x, E(\lambda)y\rangle)$$
$$= \int f(\lambda) d(\langle E(\lambda)x, y\rangle) = \langle Ax, y\rangle.$$

Here we really use the fact that $E(\lambda)$ is a projective operator.

(4) Let

$$A = \int f dE, \qquad B = \int g dE.$$

Define the (complex) measure μ in X by $\mu(M) = \langle E(M)Bx, y\rangle$, where x, y are any fixed vectors. Since $E(M)$ is a projective operator, we see

$$\mu(M) = \langle E(M)Bx, y\rangle = \langle Bx, E(M)y\rangle = \int g(\lambda) d(\langle E(\lambda)x, E(M)y\rangle) = \int g(\lambda) d(\langle E(M)E(\lambda)x, y\rangle)$$
$$= \int g(\lambda) d(\langle E(M\cap\lambda)x, y\rangle) = \int_M g(\lambda) d(\langle E(\lambda)x, y\rangle).$$

Now

$$\langle ABx, y\rangle = (\langle A^*x, Bx\rangle)^* = \left(\int f^*(\lambda) d(\langle E(\lambda)y, Bx\rangle)\right)^* = \left(\int f^*(\lambda) d(\langle y, E(\lambda)Bx\rangle)\right)^*$$
$$= \int f(\lambda) d(\langle E(\lambda)Bx, y\rangle) = \int f(\lambda) d\mu(\lambda) = \int f(\lambda)g(\lambda) d(\langle (E\lambda)x, y\rangle).$$

This gives the result. □

If E is a spectral measure, we say that E is **regular** if for every Borel set M_0 we have

$$E(M_0) = \sup_{\substack{M \subseteq M_0 \\ M \text{ is compact}}} E(M)$$

Let E be a spectral measure. Define

$$\gamma(E) = \{U \subseteq X : U \text{ is open and } E(U) = 0\}.$$

We can set

$$\Gamma(E) = \bigcup_{U \in \gamma(E)} U.$$

We define the **spectrum** of a spectral measure E, denoted $\sigma(E)$, by

$$\sigma(E) = X \setminus \Gamma(E).$$

A spectral measure is **compact** if its spectrum is compact. Note this doesn't even make sense if X is not a topological space, so we assume that from now on. To make things doable, we assume X is a locally compact Hausdorff space.

Problem 109. Show that if E is a regular spectral measure, then $\sigma(E)$ is a closed set such that $E(\sigma(E)) = 1$.

Proof. We note $\Gamma(E)$ is, by definition, a union of open sets, so open. Hence $\sigma(E)$ is closed. By regularity, it suffices to prove that if $M \subseteq X \setminus \sigma(E)$ is a compact subset, then $E(M) = 0$. Since $M \subseteq X \setminus \sigma(E) = \Gamma(E)$, we have that for all $x \in M$ there exists $U_x \in \gamma(E)$ so that U open and $E(U) = 0$. Thus $M \subseteq \bigcup_{x \in M} U_x$. Since M is compact, we can take a finite refinement so that $M \subseteq \bigcup_0^n U_i$, $E(U_i) = 0$. Now by monotonicity, we get $E(M) \leq \sum_0^n E(U_i) = 0$. □

Problem 110 (Spectral Theorem for Hermitian Operators)**.** Show that if A is a Hermitian operator, then there exists a (necessarily real and necessarily unique) compact, complex spectral measure E, called the **spectral measure of** A, such that $A = \int \lambda d(E(\lambda))$.

Proof. Let $p \in \mathbb{R}[x]$, and recall that for a Hermitian operator A we can associate to it $p(A)$ a matrix. Define $L_{x,y}(p) = \langle p(A)x, y\rangle$ for some vectors x, y. Notice that

$$|L_{x,y}(p)| \leq N_A(p) \cdot \|x\| \cdot \|y\|.$$

Thus L is a bounded linear functional, $L_{x,y} : \mathbb{R}[x] \to \mathbb{R}$. Hence we can find a unique complex measure μ in the compact set $\sigma(A)$ so that

$$L_{x,y}(p) = \int p(\lambda) d\mu(\lambda).$$

Moreover, we see that for every Borel set M we have

$$|\mu(M)| \leq \|x\| \cdot \|y\|.$$

We will denote this measure by $\mu_M(x, y)$ to indicate the dependence on x and y. Notice that $\mu_M : H \times H \to \mathbb{R}$ is a symmetric bilinear functional. Moreover, the bilinear functionals are bounded.

For each M, there exists a unique Hermitian operator $E(M)$ such that $\mu_M(x, y) = \langle E(M)x, y\rangle$. To get that E is projection valued, we show it is multiplicative. Define

$$\nu(M) = \int_M q(\lambda)d(\langle E(\lambda)x, y\rangle).$$

If p is any real polynomial, then we see that

$$\int p(\lambda)d\nu(\lambda) = \int p(\lambda)d(\langle E(\lambda)x, q(A)y\rangle),$$

so

$$\nu(M) = \int q(\lambda)d(\langle E(\lambda)E(M)x, y\rangle).$$

Thus

$$\langle E(M \cap N)x, y\rangle = \langle E(M)E(N)x, y\rangle.$$

This completes it. □

A similar kind of trick gives us the **Spectral Theorem for Normal Operators**.

Theorem. If A is a normal operator, then there exists a (necessarily unique) compact, complex spectral measure E, called the **spectral measure of** A, such that $A = \int \lambda dE(\lambda)$.

Recall that T is **strongly mixing** if

$$|\mu(T^{-k}(A) \cap B) - \mu(A)\mu(B)| \to 0 \text{ for all } A, B \in \mathcal{M}.$$

Problem 111 (Petersen 2.5.2)**.** Show that T is strongly mixing if and only if

$$\langle U^n f, g\rangle \to \langle f, 1\rangle\langle g, 1\rangle \text{ for all } f, g \in L^2.$$

Proof. Notice that if we let $f = \chi_A$ and $g = \chi_B$ for $A, B \in \mathcal{M}$ then the result follows. Fix $B \in \mathcal{M}$ and suppose f is a simple function. That is, write

$$f = \sum_{i=1}^{n} a_i \chi_{A_i}.$$

Plugging things in, we see

$$\langle U^k f, \chi_B\rangle = \int \sum_{i=1}^{n} a_i \chi_{T^{-k}(A_i)}\chi_B = \sum_{i=1}^{n} a_i \mu(T^{-k}(A_i) \cap B).$$

Taking the limit as $k \to \infty$ and using the result on measurable sets, we see that

$$\begin{aligned} |\langle U^k f, \chi_B\rangle - \langle f, 1\rangle\langle 1, g\rangle| &= \left|\sum_{i=1}^{n} a_i \left(\mu(T^{-k}(A_i) \cap B) - \mu(A_i)\mu(B)\right)\right| \\ &\leq |a_i| \sum_{i=1}^{n} |\mu(T^{-k}(A_i) \cap B) - \mu(A_i)\mu(B)| \to 0. \end{aligned}$$

This gives the result for simple functions. Now let f be in L^2. Since simple functions are dense in L^2, we can find a sequence of simple function so that for all $\epsilon > 0$ there is an N so that for $n \geq N$ we have $\|f_n - f\|_2 < \epsilon$. Still fixing $g = \chi_B$, we see that for arbitrary $n \in \mathbb{N}$ we get the estimate

$$\begin{aligned} &|\langle U^k f, \chi_B\rangle - \langle f, 1\rangle\langle 1, \chi_B\rangle| \\ &= |\langle U^k f, \chi_B\rangle - \langle U^k f_n, \chi_B\rangle + \langle U^k f_n, \chi_B\rangle - \langle f_n, 1\rangle\langle 1, \chi_B\rangle + \langle f_n, 1\rangle\langle 1, \chi_B\rangle - \langle f, 1\rangle\langle 1, \chi_B\rangle| \\ &\leq |\langle U^k f, \chi_B\rangle - \langle U^k f_n, \chi_B\rangle| + |\langle U^k f_n, \chi_B\rangle - \langle f_n, 1\rangle\langle 1, \chi_B\rangle| + |\langle f_n, 1\rangle\langle 1, \chi_B\rangle - \langle f, 1\rangle\langle 1, \chi_B\rangle|. \end{aligned}$$

Using a change of variables and the fact that T is measure preserving, we get

$$|\langle U^k f, \chi_B\rangle - \langle U^k f_n, \chi_B\rangle| \leq \int_B |f - f_n|(x) d\mu(x).$$

If we assume a probability measure space (which I don't see why we're not, see **Walters, Theorem 1.23**) we can use **Folland, Proposition 6.12** to get

$$\int_B |f - f_n|(x) d\mu(x) \leq \|f - f_n\|_2 C, \qquad C > 0 \text{ is some constant.}$$

Note that the Folland proposition actually gives the constant.

For all $\epsilon > 0$, we can choose n independent of k so that this is less than ϵ. A similar argument applies for the last term in the sum, so we have

$$|\langle U^k f, \chi_B\rangle - \langle f, 1\rangle\langle 1, \chi_B\rangle| \leq 2\epsilon + |\langle U^k f_n, \chi_B\rangle - \langle f_n, 1\rangle\langle 1, \chi_B\rangle|.$$

Since ϵ is independent of k and we know the result for simple functions, we can take $k \to \infty$ to get

$$\lim_{k\to\infty} |\langle U^k f, \chi_B\rangle - \langle f, 1\rangle\langle 1, \chi_B\rangle| \leq 2\epsilon.$$

Now $\epsilon > 0$ was arbitrary, so take $\epsilon \to 0$ to get the result. This gives us that for all $f \in L^2$, $g = \chi_B$ with $B \in \mathcal{M}$, we have the result. Now fix $f \in L^2$ and do the same kind of argument for g, first establishing it for simple functions then using density to establish it for all $g \in L^2$. ☐

Problem 112 (Petersen 2.5.3)**.** Suppose $\mathcal{J}$ is a semialgebra generating $\mathcal{M}$. Show that T is strongly mixing iff

$$|\mu(T^{-k}(A) \cap B) - \mu(A)\mu(B)| \to 0 \text{ for all } A, B \in \mathcal{J}.$$

Proof. This follows **Walters, Theorem 1.17**. The implication is trivial, so we prove the converse. The result holds trivially for the algebra $\mathcal{A}$ generated by $\mathcal{J}$, so let's assume that it holds for all $A, B \in \mathcal{A}$ which generates $\mathcal{M}$. We use the fact that for $\epsilon > 0$ and $A, B \in \mathcal{M}$ there are $A_n, B_n \in \mathcal{A}$ so that $\mu(A \triangle A_n), \mu(B \triangle B_n) < \epsilon$. Now note that

$$(T^{-k}(A) \cap B) \triangle (T^{-k}(A_n) \cap B_n) \subseteq (T^{-k}(A) \triangle T^{-k}(A_n)) \cup (B \triangle B_n).$$

The argument from here is the usual one. Notice

$$\begin{aligned} &|\mu(T^{-k}(A) \cap B) - \mu(A)\mu(B)| \\ \leq\ &|\mu(T^{-k}(A) \cap B) - \mu(T^{-k}(A_n) \cap B_n)| + |\mu(T^{-k}(A_n) \cap B_n) - \mu(A_n)\mu(B_n)| \\ &+ |\mu(A_n)\mu(B_n) - \mu(A)\mu(B_n)| + |\mu(A)\mu(B_n) - \mu(A)\mu(B)|. \end{aligned}$$

Fix $\epsilon > 0$ small (the size will be chosen later). By the above estimate and the fact that T is measure preserving, we get

$$|\mu(T^{-k}(A) \cap B) - \mu(T^{-k}(A_n) \cap B_n)| < 2\epsilon.$$

For the others, note that

$$\mu(B_n)|\mu(A_n) - \mu(A)| < \mu(B_n)\epsilon,$$
$$\mu(A)|\mu(B) - \mu(B_n)| < \mu(A)\epsilon.$$

Since $|\mu(B) - \mu(B_n)| < \epsilon$, we have $\mu(B_n) < \epsilon + \mu(B)$, so

$$\mu(B_n)|\mu(A_n) - \mu(A)| < \epsilon^2 + \epsilon\mu(B).$$

As long as $\epsilon < 1$, we have

$$\mu(B_n)|\mu(A_n) - \mu(A)| < \epsilon + \epsilon\mu(B).$$

So for $0 < \epsilon' < 1$, choose

$$\epsilon < \min\left\{\epsilon', \frac{\epsilon'}{\mu(A)}, \frac{\epsilon'}{\mu(B)}\right\}.$$

Then
$$|\mu(T^{-k}(A)\cap B)-\mu(A)\mu(B)|<|\mu(T^{-k}(A_n)\cap B)-\mu(A_n)\mu(B_n)|+5\epsilon'.$$
Since ϵ' and k are independent, we can take $k\to\infty$ to get
$$\lim_{k\to\infty}|\mu(T^{-k}(A)\cap B)-\mu(A)\mu(B)|\leq 5\epsilon'.$$
Now take $\epsilon'\to 0$ to get the result. □

Recall a set $E\subseteq\mathbb{N}$ has density zero if
$$\frac{|E\cap[0,n-1]|}{n}=\frac{1}{n}\sum_{0}^{n}\chi_E(k)\to 0.$$

Problem 113 (Koopman-von Neumann, Petersen Lemma 6.2)**.** Let $f:\mathbb{N}\to\mathbb{R}$ be a nonnegative bounded function. Show that
$$\lim_{n\to\infty}\frac{1}{n}\sum_{0}^{n-1}f(k)=0$$
iff there is a subset $E\subseteq\mathbb{N}$ of density zero such that
$$\lim_{n\to\infty,n\notin E}f(n)=0.$$

Proof. ($\implies$): Assume that
$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}f(k)=0.$$
Let $E=\{k\in\mathbb{N}:f(k)=0\}$. We can rewrite this as
$$0=\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}f(k)=\lim_{n\to\infty}\frac{1}{n}\sum_{0}^{n-1}((f\chi_E)(k)+(f\chi_{E^c})(k))=\lim_{n\to\infty}\frac{1}{n}\sum_{0}^{n-1}(f\chi_{E^c})(k)$$
$$\geq\lim_{n\to\infty}\frac{1}{n}\sum_{0}^{n-1}(\chi_{E^c})(k)\geq 0.$$
Thus the set E^c has density zero and we see that
$$\lim_{n\to\infty,n\notin E^c}f(n)=0.$$

($\impliedby$): Let $E\subseteq\mathbb{N}$ be the set of density zero such that we have the property. Then again we examine
$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}f(k)=\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}((f\chi_E)(k)+(f\chi_{E^c})(k)).$$
Examine the left part first. We have
$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}(f\chi_E)(k).$$
Fixing n, we have by Cauchy Schwarz that
$$\frac{1}{n}\sum_{k=0}^{n-1}(f\chi_E)(k)\leq\sqrt{\frac{1}{n}\sum_{k=0}^{n-1}f(k)^2}\sqrt{\frac{1}{n}\sum_{k=0}^{n-1}\chi_E(k)}.$$

Taking the limit as $n \to \infty$ of both sides and using the fact that E has density zero, this gives us that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (f\chi_E)(k) = 0.$$

Thus

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(k) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(k)\chi_{E^c}(k).$$

Now

$$\lim_{n\to\infty} f(n)\chi_{E^c}(n) = 0$$

implies for all $\epsilon > 0$ there is an N so that for $n \geq N$ we have

$$f(n)\chi_{E^c}(n) < \epsilon.$$

Fix $\epsilon > 0$ and N. For $n \geq N + 1$ we can rewrite this sum as

$$\begin{aligned}\frac{1}{n} \sum_{k=0}^{n-1} f(k)\chi_{E^c}(k) &= \frac{1}{n} \sum_{k=0}^{N} f(k)\chi_{E^c}(k) + \frac{1}{n} \sum_{k=N+1}^{n} f(k)\chi_{E^c}(k) \\ &< \frac{1}{n} \sum_{k=0}^{N} f(k)\chi_{E^c}(k) + \frac{\epsilon(n-N)}{n}.\end{aligned}$$

Taking $n \to \infty$ of both sides gives us

$$0 \leq \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(k) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(k)\chi_{E^c}(k) \leq \epsilon.$$

The choice of $\epsilon > 0$ was arbitrary, so we let $\epsilon \to 0$ and this gives us the result. □

Problem 114. Assume $(X, \mathcal{M}, \mu, T)$ is a measure preserving system of a probability measure space. Show that if $T \times T$ is weakly mixing, then so is T.

Proof. Let $A, B \in \mathcal{M}$. Examine $A \times X$, $B \times X \in \mathcal{M} \otimes \mathcal{M}$. Since $T \times T$ is weakly mixing, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \left| (\mu \otimes \mu)(T^k(A \times X) \cap (X \times B)) - (\mu \otimes \mu)(A \times X)(\mu \otimes \mu)(X \times B) \right| = 0.$$

Using the fact that these are measurable cylinders, we can rewrite this as

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \left| \mu(T^k(A) \cap B)\mu(T^k(X) \cap X) - \mu(A)\mu(B)\mu(X)^2 \right| = 0.$$

Since we're assuming that $(X, \mathcal{M}, \mu)$ is a probability measure space, this is equivalent to

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \left| \mu(T^k(A) \cap B) - \mu(A)\mu(B) \right| = 0.$$

This implies T is weakly mixing. □

Problem 115 (Petersen Theorem 2.6.1)**.** Let $(X, \mathcal{M}, \mu, T)$ be a measure preserving system on a probability measure space. Show the following are equivalent.

(1) T is weakly mixing.

(2) For all $f, g \in L^2(\mu)$, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} |\langle U^k f, g\rangle - \langle f, 1\rangle\langle 1, g\rangle| = 0.$$

(3) Given $A, B \in \mathcal{M}$, there is a set $J \subseteq \mathbb{N}$ of density zero so that

$$\lim_{n\to\infty, n\notin J} \mu(T^n(A) \cap B) = \mu(A)\mu(B).$$

(4) $T \times T$ is weakly mixing.
(5) $T \times S$ is ergodic on $X \times Y$ for each ergodic system $(Y, \mathcal{N}, \nu, S)$.
(6) $T \times T$ is ergodic.

Proof. We note that (1) $\iff$ (2) is clear by an earlier problem. We note (5) $\implies$ (6) is clear.
(1) $\iff$ (3): Let

$$f(n) = |\mu(T^n(A) \cap B) - \mu(A)\mu(B)|.$$

Assuming (1), we get

$$\frac{1}{n} \sum_{0}^{n-1} f(k) = 0,$$

and by an earlier problem that implies the existence of J with density zero so that

$$\lim_{n\to 0, n\notin J} f(n) = \lim_{n\to 0, n\notin J} |\mu(T^n(A) \cap B) - \mu(A)\mu(B)| = 0.$$

Assuming (3), we do the prior argument backwards, again using the an earlier problem.
Note that we have (1) $\iff$ (2) $\iff$ (3).
(3) $\implies$ (4): Consider $X \times X$ with measure space $\mathcal{M} \otimes \mathcal{M}$ and measure $\mu \otimes \mu$. We can consider the cylindrical sets. Take $A, B, C, D \in \mathcal{M}$. By (3), there exists a J_1 and J_2 of density zero so that

$$\lim_{n\to\infty, n\notin J_1} \mu(T^n(A) \cap B) = \mu(A)\mu(B),$$

$$\lim_{n\to\infty, n\notin J_2} \mu(T^n(C) \cap D) = \mu(C)\mu(D).$$

Notice that $J_1 \cup J_2$ has density zero, since

$$0 \le \frac{1}{n} \sum_{k=0}^{n-1} \chi_{J_1 \cup J_2}(k) \le \frac{1}{n} \sum_{k=0}^{n-1} [\chi_{J_1}(k) + \chi_{J_2}(k)] \to 0.$$

Now consider

$$\begin{aligned}
&\lim_{n\to\infty, n\notin J_1\cup J_2} |(\mu\otimes\mu)(T^n(A\times C) \cap (B\times D)) - (\mu\otimes\mu)(A\times C)(\mu\otimes\mu)(B\times D)| \\
&= \lim_{n\to\infty, n\notin J_1\cup J_2} |\mu(T^n(A)\cap B)\mu(T^n(C)\cap D) - \mu(A)\mu(B)\mu(C)\mu(D)| \\
&= \lim_{n\to\infty, n\notin J_1\cup J_2} |\mu(T^n(A)\cap B)\mu(T^n(C)\cap D) - \mu(T^n(A)\cap B)\mu(C)\mu(D) \\
&\qquad + \mu(T^n(A)\cap B)\mu(C)\mu(D) - \mu(A)\mu(B)\mu(C)\mu(D)| \\
&\le \lim_{n\to\infty, n\notin J_1\cup J_2} |\mu(T^n(A)\cap B)\mu(T^n(C)\cap D) - \mu(T^n(A)\cap B)\mu(C)\mu(D)| \\
&\qquad + \lim_{n\to\infty, n\notin J_1\cup J_2} |\mu(T^n(A)\cap B)\mu(C)\mu(D) - \mu(A)\mu(B)\mu(C)\mu(D)| \\
&= \mu(T^n(A)\cap B) \lim_{n\to\infty, n\notin J_1\cup J_2} |\mu(T^n(C)\cap D) - \mu(C)\mu(D)| \\
&\qquad + \mu(C)\mu(D) \lim_{n\to\infty, n\notin J_1\cup J_2} |\mu(T^n(A)\cap B) - \mu(A)\mu(B)| = 0.
\end{aligned}$$

We thus have (3) for $T\times T$ on the semialgebra $\mathcal{J}=\mathcal{M}\times\mathcal{M}$ which generates $\mathcal{M}\otimes\mathcal{M}$. By a density argument, this tells us that it applies for $\mathcal{M}\otimes\mathcal{M}$.
(4) $\implies$ (5): Notice that $T\times T$ weakly mixing implies T is weakly mixing by the last problem. Consider an ergodic system $(Y,\mathcal{N},\nu,S)$. We will show that $T\times S$ is weakly mixing, which implies that it is ergodic. Take $A,B\in\mathcal{M}$, $C,D\in\mathcal{N}$, and consider

$$\left|\frac{1}{n}\sum_{k=0}^{n-1}(\mu\otimes\nu)((T\times S)^k(A\times C)\cap(B\times D))-(\mu\otimes\nu)(A\times C)(\mu\otimes\nu)(B\times D)\right|.$$

We may rewrite this as

$$\left|\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^k(A)\cap B)\nu(S^k(C)\cap D)-\mu(A)\nu(C)\mu(B)\nu(D)\right|.$$

We can now add and subtract

$$\frac{1}{n}\sum_{k=0}^{n-1}\nu(S^k(C)\cap D)\mu(A)\mu(B)$$

on the inside to get

$$\left|\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^k(A)\cap B)\nu(S^k(C)\cap D)-\frac{1}{n}\sum_{k=0}^{n-1}\nu(S^k(C)\cap D)\mu(A)\mu(B)\right.$$
$$\left.+\frac{1}{n}\sum_{k=0}^{n-1}\nu(S^k(C)\cap D)\mu(A)\mu(B)-\mu(A)\nu(C)\mu(B)\nu(D)\right|.$$

We then get an upper bound

$$\nu(S^k(C)\cap D)\frac{1}{n}\sum_{k=0}^{n-1}|\mu(T^k(A)\cap B)-\mu(A)\mu(B)|+\mu(A)\mu(B)\left|\frac{1}{n}\sum_{k=0}^{n-1}\nu(S^k(C)\cap D)-\nu(C)\nu(D)\right|.$$

Taking $n\to\infty$ of both sides gives us that the first term tends to 0, since T is weakly mixing, and the second term tends to 0 since S is ergodic. Thus we have it is weakly mixing on the cylinders, so it is weakly mixing.
(5) $\implies$ (6): Clear, since weakly mixing implies ergodic.
(6) $\implies$ (3): Assume $T\times T$ is ergodic. Take $A,B\in\mathcal{M}$. We examine

$$\frac{1}{n}\sum_{k=0}^{n-1}\Big[\mu(T^k(A)\cap B)-\mu(A)\mu(B)\Big]^2$$
$$=\frac{1}{n}\sum_{k=0}^{n-1}\Big[\mu(T^k(A)\cap B)^2-2\mu(T^k(A)\cap B)\mu(A)\mu(B)+\mu(A)^2\mu(B)^2\Big]$$
$$=\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^k(A)\cap B)^2-2\mu(A)\mu(B)\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^k(A)\cap B)+(\mu(A)\mu(B))^2.$$

Now we rewrite this in terms of $T\times T$. We have

$$\frac{1}{n}\sum_{k=0}^{n-1}(\mu\otimes\mu)(T^k(A\times A)\cap(B\times B))-2\mu(A)\mu(B)\frac{1}{n}\sum_{k=0}^{n-1}(\mu\otimes\mu)(T^k(A\times X)\cap(B\times X))$$
$$+(\mu\otimes\mu)(A\times A)(\mu\otimes\mu)(B\times B).$$

Taking $n \to \infty$, we use ergodicity. The above is then equal to

$(\mu\otimes\mu)(A\times A)(\mu\otimes\mu)(B\times B)-2\mu(A)\mu(B)(\mu\otimes\mu)(A\times X)(\mu\otimes\mu)(B\times X)+(\mu\otimes\mu)(A\times A)(\mu\otimes\mu)(B\times B).$

These are cylinders, so we can evaluate this to get

$$\mu(A)^2\mu(B)^2 - 2\mu(A)^2\mu(B)^2 + \mu(A)^2\mu(B)^2 = 0.$$

Therefore

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1}\Big[\mu(T^k(A)\cap B) - \mu(A)\mu(B)\Big]^2 = 0.$$

Define

$$f(k) = \Big[\mu(T^k(A)\cap B) - \mu(A)\mu(B)\Big]^2.$$

By an earlier problem, we get that there is a density zero subset $J \subseteq \mathbb{N}$ with

$$\lim_{n\to\infty, n\notin J} f(k) = \lim_{n\to\infty, n\notin J}\Big[\mu(T^k(A)\cap B) - \mu(A)\mu(B)\Big]^2 = 0.$$

This forces

$$\lim_{n\to\infty, n\notin J}\Big|\mu(T^k(A)\cap B) - \mu(A)\mu(B)\Big| = 0,$$

which forces T to be weakly mixing.

Thus we've shown (1) $\iff$ (2) $\iff$ (3) $\implies$ (4) $\implies$ (5) $\implies$ (6) $\implies$ (3). This gives us all of the equivalences. □

Recall that a unitary operator $U : H \to H$ is said to have continuous spectrum if 1 is the only eigenvalue and the only eigenfunctions are the constants.

Problem 116. Let $(X, \mathcal{M}, \mu, T)$ be an invertible measure preserving system of a probability measure space. Let $U : L^2(\mu) \to L^2(\mu)$ be the corresponding unitary operator. Show that if λ is a non-trivial eigenvalue of U, then $|\lambda| = 1$. Thus all eigenvalues of the unitary operator lie on the circle.

Proof. Let λ be an eigenvalue, f an associated eigenfunction. We have

$$\|f\|^2\langle f, f\rangle = \langle Uf, Uf\rangle = \langle U^2 f, f\rangle = \lambda^2\langle f, f\rangle = \lambda^2\|f\|^2.$$

Taking square roots, we have $|\lambda| = 1$ or $\|f\| = 0$. Since λ was assumed to be non-trivial, f assumed to be non-trivial, this gives us the result. □

We now include Walters version of the spectral theorem for unitary operators.

Theorem. Suppose U is a unitary operator on a Hilbert space H. Then for each $f \in H$ there exists a unique finite Borel measure μ_f on K so that

$$\langle U^n f, f\rangle = \int_K \lambda^n d\mu_f(\lambda) \text{ for all } n \in \mathbb{Z}.$$

If T is an invertible measure preserving transformation, then U_T is unitary. If T has continuous spectrum and $\langle f, 1\rangle = 0$, then μ_f has no atoms.

Let $(X, \mathcal{M}, \mu, T)$ be an ergodic measure-preserving system on a probability measure space. We say that $(X, \mathcal{M}, \mu)$ has **discrete spectrum** if there exists an orthonormal basis for $L^2(\mu)$ which consists of eigenfunctions of T.

Problem 117. Show that if T has discrete spectrum and $(X, \mathcal{M}, \mu)$ is a Lebesgue space, then T is (measure theoretically) invertible.

Proof. Consider $U_T : L^2(\mu) \to L^2(\mu)$. This is going to be surjective, since every element can be expressed as a linear combination of eigenfunctions. Since it's an eigenfunction, we have that $\widehat{T}^{-1} : (\mathcal{M}, \mu) \to (\mathcal{M}, \mu)$ (modulo sets of measure zero) is an automorphism of measure spaces. If it is a Lebesgue space, we recall this induces a point measure map. □

Problem 118. Let $(X, \mathcal{M}, \mu, T)$ be an ergodic system of a probability measure space. Prove the following:

(1) If $U_T f = \lambda f$, $f \neq 0$, then $|\lambda| = 1$ and $|f|$ is constant almost everywhere.
(2) Eigenfunctions corresponding to different eigenvalues are orthogonal.
(3) If f and g are both eigenfunctions corresponding to the eigenvalue λ, then $f = cg$ almost everywhere for some c.
(4) The eigenvalues of T form a subgroup of the unit circle.

Proof.

(1) Assume $f \neq 0$ and
$$U_T f = \lambda f.$$
By definition,
$$\|U_T f\|_2^2 = \int |f \circ T|^2 d\mu.$$
Do the change of variables formula and use the fact that T is measure preserving to get
$$\|U_T f\|_2^2 = \int |f|^2 d(\mu \circ T^{-1}) = \int |f|^2 d\mu = \|f\|_2^2.$$
Thus we have $\|f\|_2 = \|U_T f\|_2 = \|\lambda f\|_2 = |\lambda| \|f\|_2$. Subtracting from both sides, we see
$$(1 - |\lambda|)\|f\|_2 = 0.$$
This implies either $|\lambda| = 1$ or $\|f\|_2 = 0$. By assumption $f \neq 0$, so this forces $|\lambda| = 1$.
To see $|f|$ is constant almost everywhere, notice
$$|f| \circ T = |f \circ T| = |\lambda f| = |\lambda||f| = |f|.$$
Thus $|f|$ is T-invariant. Since T is ergodic, this forces $|f|$ to be constant almost everywhere.

(2) Suppose $U_T f = \lambda_1 f$, $U_T g = \lambda_2 g$, $\lambda_1 \neq \lambda_2$. Recalling that U_T is an isometry, we have
$$\langle f, g\rangle = \langle U_T f, U_T g\rangle = \lambda_1 \overline{\lambda_2} \langle f, g\rangle.$$
Notice this forces either $\langle f, g\rangle = 0$ or $\lambda\overline{\lambda_2} = 1$. Since $|\lambda_2| = 1$, notice that $\lambda_2\overline{\lambda_2} = 1$ so that $\overline{\lambda_2} = \lambda_2^{-1}$. Thus the second condition can be rewritten as $\lambda_1 = \lambda_2$. Since we assumed $\lambda_1 \neq \lambda_2$, this forces $\langle f, g\rangle = 0$. This gives us orthogonality.

(3) Without loss of generality, $g \neq 0$. Notice f/g is T-invariant, since
$$U_T(f/g) = U_T(f)/U_T(g) = \lambda f/(\lambda g) = f/g.$$
Thus f/g is constant almost everywhere.

(4) Let $\sigma(T)$ denote the eigenvalues. We've seen that if $\lambda \in \sigma(T)$, then $\lambda^{-1} \in \sigma(T)$. If f, g are eigenfunctions for λ_1, λ_2, then
$$U_T(fg) = U_T(f)U_T(g) = \lambda_1\lambda_2 fg.$$
So $\lambda_1\lambda_2 \in \sigma(T)$. This concludes that it's a subgroup.

□

Let $(X_1, \mathcal{M}_1, \mu_1, T_1)$, $(X_2, \mathcal{M}_2, \mu_2, T_2)$ be two systems. We say T_1 and T_2 are **spectrally isomorphic** if there exists a $W : L^2(\mu_2) \to L^2(\mu_1)$ so that

(1) W is invertible,

(2) $\langle Wf, Wg\rangle = \langle f, g\rangle$ for all $f, g \in L^2(\mu_2)$,
(3) $U_{T_1}W = WU_{T_2}$.

Problem 119. Show that if $\sigma(T_1)$ and $\sigma(T_2)$ denote the set of eigenvalues for T_1 and T_2 and T_1 and T_2 are spectrally isomorphic, then $\sigma(T_1) = \sigma(T_2)$.

Proof. Let $\lambda \in \sigma(T_1)$. There is some $f \in L^2(\mu_1)$ so that $U_{T_1}(f) = \lambda f$, $f \neq 0$. Since W invertible, there is some $\widehat{f} \in L^2(\mu_2)$ so that $W(\widehat{f}) = f$ (in L^2). Notice that

$$\lambda f = U_{T_1}(f) = U_{T_1}W(\widehat{f}) = WU_{T_2}(\widehat{f}).$$

Applying W^{-1} to both sides yields

$$\lambda W^{-1}(f) = \lambda\widehat{f} = U_{T_2}(\widehat{f}).$$

Notice as well

$$\|\widehat{f}\|_2^2 = \langle \widehat{f}, \widehat{f}\rangle = \langle W(\widehat{f}), W(\widehat{f})\rangle = \langle f, f\rangle = \|f\|_2^2,$$

so $\widehat{f} \neq 0$ almost everywhere. Thus $\lambda \in \sigma(T_2)$. If $\lambda \in \sigma(T_2)$, there is some g non-zero so that $U_{T_2}(g) = \lambda g$. Consider $\widehat{g} = W(g)$. This is non-zero (again using the isometry condition) and we see that

$$U_{T_1}(\widehat{g})U_{T_1}W(g) = WU_{T_2}(g) = \lambda W(g) = \lambda\widehat{g}.$$

Thus $\lambda \in \sigma(T_1)$. This gives equality. □

Problem 120. Let $(X, \mathcal{M}, \mu)$ be a probability space, $h \in L^2(\mu)$. Show the following are equivalent.
(1) We have h is bounded.
(2) We have $h \cdot f \in L^2(\mu)$ for all $f \in L^2(\mu)$.

Proof. (1) $\implies$ (2): If h is bounded, then $h \in L^\infty(\mu) \cap L^2(\mu)$ so that

$$\|hf\|_2^2 = \int |hf|^2 d\mu \leq \int \|h\|_\infty |f|^2 d\mu = \|h\|_\infty \|f\|_2 < \infty.$$

(2) $\implies$ (1): Consider

$$X_n = \{x \in X : n - 1 \leq |h| < n\} \qquad n \geq 1.$$

Then

$$X = \bigsqcup_{n=1}^\infty X_n.$$

Write

$$f = \sum_{i=1}^M i^{-1}\mu(X_i)^{-1/2}\chi_{X_i}(x).$$

Notice that we have

$$\|f\|_2^2 = \int |f|^2 d\mu \leq \sum_{i=1}^\infty i^{-2} < \infty.$$

Let $F = \{i : \mu(X_i) \neq 0\}$. Then

$$\|hf\|_2^2 = \int |hf|^2 d\mu \geq \sum_{i \in F}\left(\frac{i-1}{i}\right)^2.$$

Thus F must be finite, so h is bounded. □

Recall this fact from algebra (a sort of Baer's criterion argument, see Problem 26 in my notes).

Theorem (Walters, Lemma 3.3)**.** Let H be a discrete abelian group and K a divisible subgroup of H. Then there exists a homomorphism $\varphi : H \to K$ such that $\varphi|_K = \mathrm{Id}_K$.

Theorem. Let T_i be an ergodic measure preserving transformation of a probability space $(X_i, \mathcal{B}_i, m_i)$ and suppose T_i has discrete spectrum for $i = 1, 2$. Then the following are equivalent:

(1) T_1 and T_2 are spectrally isomorphic.
(2) T_1 and T_2 have the same eigenvalues.
(3) T_1 and T_2 are conjugate.

Let $(X, \mathcal{M}, \mu, T)$ and $(Y, \mathcal{N}, \nu, S)$ be measure preserving transformations of probability measure spaces. We say that T is **isomorphic** to S if there exists $M \in \mathcal{M}$ and $N \in \mathcal{N}$ with $\mu(M) = 1$ and $\nu(N) = 1$ so that

(1) $T(M) \subseteq M$, $S(N) \subseteq N$ (in other words M is T-invariant and N is S-invariant),
(2) there is an invertible measure-preserving transformation

$$\varphi : M \to N \text{ with } \varphi T(x) = S\varphi(x) \text{ for all } x \in M.$$

Write this as $T \cong S$.

Problem 121. Prove that isomorphism is an equivalence relation.

Proof. There are three properties to check.

(1) We see that $T \cong T$ by just taking $M = X$, $N = X$, and φ the identity.
(2) If $T \cong S$, then $S \cong T$ by taking $\varphi = \varphi^{-1}$.
(3) This is the more interesting thing to check. Suppose $T \cong S$ and $S \cong Q$. Write the systems as $(X, \mathcal{M}, \mu, T)$, $(Y, \mathcal{N}, \nu, S)$, and $(Z, \mathcal{B}, \rho, Q)$. Since $T \cong S$, we have that there exists a T-invariant set M and an S-invariant set N, both of full measure, so that there is an invertible measure-preserving transformation

$$\varphi : M \to N \text{ with } \varphi T = S\varphi \text{ on } M.$$

Since $S \cong Q$, we have that there is an S-invariant set of full measure K and a Q-invariant set of full measure O so that there is an invertible measure-preserving transformation

$$\psi : K \to O \text{ with } \psi S = Q\psi \text{ on } K.$$

Consider $K \cap N$. Notice that

$$\nu(K \cap N) = 1 - \nu(K^c \cup N^c) \geq 1 - (\nu(K^c) + \nu(N^c)) = 1,$$

so this is still a set of full measure. If we take $M \cap \varphi^{-1}(K \cap N)$, this is still a set of full measure by the same argument, and φ restricted to this set is going to be an invertible measure-preserving transformation. Consider $O \cap \psi(K \cap N)$. The same argument says this is a set of full measure. Thus if we relabel $K \cap N$ as N, relabel $M \cap \varphi^{-1}(K \cap N)$ as M, and relabel $O \cap \psi(K \cap N)$ as O, and relabel the transformations restricted to these sets, we have $\varphi : M \to N$ is an invertible measure preserving transformation so that $\varphi T = S\varphi$ on M and $\psi : N \to O$ is an invertible measure preserving transformation so that $\psi S = Q\psi$ on N. Thus $\kappa = \psi \circ \varphi : M \to O$ is measure preserving (as the composition of measure preserving transformations is measure preserving) and invertible, and moreover on M we have

$$\kappa T = (\psi \circ \varphi)T = \psi \circ S \circ \varphi = Q(\psi \circ \varphi) = Q\kappa.$$

This gives an isomorphism. Thus $T \cong Q$.

□

Problem 122. Show that if T and S in the above definition are invertible, then we can take M and N so that $TM = M$ and $SN = N$.

Proof. If T is invertible, then $\mu(T(M)) = \mu(M) = 1$ and $\mu(T^{-1}(M)) = \mu(M) = 1$. The intersection of countably many sets of full measure has full measure, so consider

$$M' = \bigcap_{n=-\infty}^{\infty} T^n M.$$

Since T invertible, $T(M') = M'$. Notice the same trick applies for S. Moreover, these sets work for isomorphisms. $\square$

A **partition** of a space $(X, \mathcal{M}, \mu)$ is a disjoint collection of elements in $\mathcal{M}$ whose union is X. We will mostly be focused on finite partitions, denoted with a Greek letter (for example, $\xi = \{A_1, \dots, A_n\}$.

Problem 123. Let $(X, \mathcal{M}, \mu)$ be a measure space and let $\xi = \{A_1, \dots, A_n\}$ be a partition. Consider the collection $\mathcal{A}(\xi)$ which consists of all unions of elements of ξ. Show this is a sub-σ-algebra.

Proof. We need to show three things.

(1) We see that $X \in \mathcal{A}(\xi)$, since $\bigcup_{i=1}^n A_i = X \in \mathcal{A}(\xi)$.
(2) Let $\{C_i\} \in \mathcal{A}(\xi)$. Then $\bigcup_i C_i \in \mathcal{A}(\xi)$, since a union of C_i is just a union of A_i.
(3) Let $C \in \mathcal{A}(\xi)$. Then by definition C is a union of A_i. After relabeling, we can assume that C is a union of the first k. That is,

$$C = \bigcup_{i=1}^{k} A_i.$$

Now

$$C^c = X \setminus C = X \setminus \left(\bigcup_{i=1}^{k} A_i \right) = \bigcup_{i=k+1}^{n} A_i.$$

To see this last equality, let $x \in X \setminus \left(\bigcup_{i=1}^{k} A_i \right)$. Since $X = \bigcup_{i=1}^n A_i$, we see that this forces $x \in \bigcup_{i=k+1}^n A_i$, so it is a subset. The other direction is clear; $x \in \bigcup_{i=k+1}^n A_i$ implies $x \notin \bigcup_{i=1}^k A_i$ and $x \in X$.

$\square$

Problem 124. Assume that $\mathbb{C}$ is a finite sub-σ-algebra of $\mathcal{M}$. Then $\mathcal{C} = \{C_i : i = 1, \dots, n\}$. Let B_i be of the form C_i or C_i^c. Then show that the non-empty sets of the form $B_1 \cap \cdots \cap B_n$ form a finite partition of $(X, \mathcal{M}, \mu)$. Denote this partition by $\xi(\mathcal{C})$.

Proof. We need to show that they are disjoint and that they union to the whole set. One can write

$$B_1 \cap \cdots \cap B_n = B_{i_1 \cdots i_n},$$

where $i_j \in \{0, 1\}$ is 1 if $B_i = C_i$ and is 0 if $B_i = C_i^c$. Then if $i_1 \cdots i_n \neq j_1 \cdots j_n$, there is some $k \in \{1, \dots, n\}$ so that $i_k \neq j_k$. Without loss of generality, assume that $i_k = 1$ and $j_k = 0$. Then we see that

$$B_{i_1 \cdots i_n} \cap B_{j_1 \cdots j_n} \subseteq C_k \cap C_k^c = \varnothing,$$

so these sets are disjoint.

We have that $\bigcup_{i=1}^n C_i = X$, since $\mathcal{C}$ is a sub-σ-algebra. So $x \in X$ implies $x \in C_i$ for some i. But unioning over $B_{i_1 \cdots i_n}$, we get $C_i \subseteq \bigcup_{i_1 \cdots i_n} B_{i_1 \cdots i_n}$, so we must have $X = \bigcup_{i_1 \cdots i_n} B_{i_1 \cdots i_n}$. $\square$

Problem 125. Deduce there is a one-to-one correspondence between finite sub-σ-algebras and finite partitions.

Proof. Let $\mathcal{C} = \{C_i : i = 1, \ldots, n\}$. We see that $\xi(\mathcal{C})$ is going to be

$$\xi(\mathcal{C}) = \{B_{i_1 \cdots i_n} : i_1, \ldots, i_n \in \{0, 1\}\}.$$

Now we see that we can get C_k by fixing $i_k = 1$ and union over all of the other elements varying. Thus $C_k \in \mathcal{A}(\xi(\mathcal{C}))$ for $k = 1, \ldots, n$, and we see that $\mathcal{C} = \{C_1, \ldots, C_n\} \subseteq \mathcal{A}(\xi(\mathcal{C}))$. Since $\xi(\mathcal{C}) \subseteq \mathcal{C}$, we get equality here by minimality. So $\mathcal{A}(\xi(\mathcal{C})) = \mathcal{C}$. The other direction is similar. □

Consider the space of finite partitions, call it Γ. We can introduce a partial ordering on Γ, denoted $\leq$, by saying $\xi \leq \eta$ if each element if ξ is a union of elements of η. We call η a **refinement** of ξ.

Problem 126. Show

$$\xi \leq \eta \iff \mathcal{A}(\xi) \subseteq \mathcal{A}(\eta)$$

Proof. ($\Longrightarrow$): Notice $\xi \leq \eta$ if every element of ξ is a union of elements from η. Since $\mathcal{A}(\xi)$ is a union of elements from ξ, this implies that $\mathcal{A}(\xi) \subseteq \mathcal{A}(\eta)$.
($\Longleftarrow$): Same kind of idea. □

The same kind of idea shows that $\mathcal{A} \subseteq \mathcal{C}$ iff $\xi(\mathcal{A}) \leq \eta(\mathcal{C})$. Essentially what we've done is induced a partial ordering via the one-to-one correspondence.

Let $\xi = \{A_1, \ldots, A_n\}$ and $\eta = \{C_1, \ldots, C_k\}$ be two finite partitions of $(X, \mathcal{M}, \mu)$. Their **join** is the partition

$$\xi \vee \eta = \{A_i \cap C_j : 1 \leq i \leq n, 1 \leq j \leq k\}.$$

Similarly the **join** of $\mathcal{A}$ and $\mathcal{C}$, which are two finite sub-σ-algebras, is the smallest sub-σ-algebra of $\mathcal{M}$ containing $\mathcal{A}$ and $\mathcal{C}$.

Problem 127.

(1) Show that $\mathcal{A} \vee \mathcal{C}$ is a finite sub-σ-algebra. Deduce that the space of finite sub-σ-algebras are closed under the join operation.
(2) Show that
$$\mathcal{A}(\xi \vee \eta) = \mathcal{A}(\xi) \vee \mathcal{A}(\eta).$$
(3) Show that
$$\xi(\mathcal{A} \vee \mathcal{C}) = \xi(\mathcal{A}) \vee \xi(\mathcal{C}).$$

Proof.

(1) Let $\mathcal{N} = \mathcal{A} \vee \mathcal{C}$. This is the smallest sub-σ-algebra containing $\mathcal{A}$ and $\mathcal{C}$. Let
$$\mathcal{B} = \{A \cap C : A \in \mathcal{A}, C \in \mathcal{C}\},$$
then $\sigma(\mathcal{B})$, which is the union of all elements in $\mathcal{B}$, is a finite sub-σ-algebra. Notice that $\mathcal{A} \subseteq \sigma(\mathcal{B})$, $\mathcal{C} \subseteq \sigma(\mathcal{B})$, so $\mathcal{N} \subseteq \sigma(\mathcal{B})$ by minimality. Finally if $K \in \sigma(\mathcal{B})$, then it can be written as
$$K = \bigcup_{i,j} A_i \cap C_j.$$
If τ is any σ algebra containing $\mathcal{A}$ and $\mathcal{C}$, then we see that we must have $K \in \tau$ (since τ contains $\{A_i\}$ and $\{C_j\}$) implying that $K \in \mathcal{N}$. Thus $\mathcal{N} = \sigma(\mathcal{B})$, and $\mathcal{A} \vee \mathbb{C}$ is a finite sub-σ-algebra.
(2) We can view $\mathcal{A}(\xi) = \sigma(\{A_i : 1 \leq i \leq n\})$, $\mathcal{A}(\eta) = \sigma(\{C_j : 1 \leq j \leq k\})$ by (1). By definition
$$\xi \vee \eta = \{A_i \cap C_j : 1 \leq i \leq n, 1 \leq j \leq k\}.$$
Thus
$$\mathcal{A}(\xi \vee \eta) = \sigma(\{A_i \cap C_j : 1 \leq i \leq n, 1 \leq j \leq k\}).$$

So this is unions of elements of this form. Meanwhile $\mathcal{A}(\xi) \vee \mathcal{A}(\eta)$ is the smallest sub-σ-algebra containing $\mathcal{A}(\xi)$ and $\mathcal{A}(\eta)$. Again by (1) this is a finite sub-σ-algebra, and we explicitly calculated in (1) that it will be $\mathcal{A}(\xi \vee \eta)$.

(3) Write $\mathcal{A} = \{A_1, \ldots, A_n\}$, $\mathcal{C} = \{C_1, \ldots, C_k\}$, where elements in $\mathcal{A}$ are really unions of A_i and same for $\mathcal{C}$. We can write

$$\xi(\mathcal{A}) = \{B_{i_1 \cdots i_n} : i_j \in \{0,1\} \text{ for } 1 \leq j \leq n\}.$$

Similarly

$$\xi(\mathcal{C}) = \{K_{j_1 \cdots j_k} : j_s \in \{0,1\} \text{ for } 1 \leq s \leq k\}.$$

We have an explicit calculation for $\mathcal{A} \vee \mathcal{C}$ from (1) which says it looks like unions of elements of the form $A_i \cap C_j$. Notice that

$$\xi(\mathcal{A} \vee \mathcal{C}) = \{T_{i_1 \cdots i_n j_1 \cdots j_k} : i_s, j_r \in \{0,1\} \text{ for } 1 \leq s \leq n, 1 \leq r \leq k\},$$

where $T_{i_1 \cdots i_n j_1 \cdots j_k} = B_{i_1 \cdots i_n} \cap K_{j_1 \cdots j_k}$. Notice that

$$\xi(\mathcal{A}) \vee \xi(\mathcal{C}) = \{B_{i_1 \cdots i_n} \cap K_{j_1 \cdots j_k}\} = \{T_{i_1 \cdots i_n j_1 \cdots j_k}\}.$$

Thus these sets are the same.

□

Problem 128. Let $\xi = \{A_1, \ldots, A_n\}$ be a partition of $(X, \mathcal{M}, \mu, T)$ a (invertible) measure preserving system of a probability measure space (so the partition is of the space). Show that $T^{-1}(\xi)$ is a partition of X.

Proof. We note $T^{-1}(X) = X$, so

$$\bigcup_{i=1}^{n} T^{-1}(A_i) = T^{-1}\left(\bigcup_{i=1}^{n} A_i\right) = X.$$

Moreover these $\{T^{-1}(A_i)\}$ are disjoint. Thus it is a partition. □

If $\mathcal{C}$ and $\mathcal{D}$ are sub-σ-algebras of $\mathcal{M}$, we write $\mathcal{C} \overset{\circ}{\subset} \mathcal{D}$ if for every $C \in \mathcal{C}$ there is a $D \in \mathcal{D}$ with $\mu(D \triangle C) = 0$. If ξ and η are finite partitions, then $\xi \overset{\circ}{=} \eta$ means $\mathcal{A}(\xi) \overset{\circ}{=} \mathcal{A}(\eta)$.

Suppose $\mathcal{C}$ and $\mathcal{D}$ are finite and $\mathcal{C} \overset{\circ}{=} \mathcal{D}$. Then if we can write $\xi(\mathcal{C}) = \{C_1, \ldots, C_p, C_{p+1}, \ldots, C_q\}$ where $\mu(C_i) > 0$ for $1 \leq i \leq p$ and $\mu(C_i) = 0$ for $p < i \leq q$, we are then able to write $\xi(\mathcal{D}) = \{D_1, \ldots, D_p, D_{p+1}, \ldots, D_s\}$ with $\mu(C_i \triangle D_i) = 0$ for $1 \leq i \leq p$ and $\mu(D_i) = 0$ for $p+1 \leq i \leq s$.

Recall the **Radon-Nikodym** theorem. If μ and ν are two measures on the same space $(X, \mathcal{M})$, we say $\mu \ll \nu$ if for every $E \in \mathcal{M}$ with $\nu(E) = 0$ we have $\mu(E) = 0$. We say $\mu \perp \nu$ if there exists $E \in \mathcal{M}$ with $\mu(E) = 0$ and $\nu(E^c) = 0$.

Theorem (Radon-Nikodym Theorem)**.** Let μ, ν be two probability measures on the space $(X, \mathcal{M})$. Then $\mu \ll \nu$ iff there exists $f \in L^1(\nu)$ with $f \geq 0$ and $\int f d\nu = 1$ so that $\mu(B) = \int_B f d\nu$ for all $B \in \mathcal{M}$. The function f is unique almost everywhere.

Theorem (Lebesgue Decomposition Theorem)**.** Let μ and ν be two probability measures on $(X, \mathcal{M})$. There exists $p \in [0,1]$ and probability measures μ_1, μ_2 so that $\mu = p\mu_1 + (1-p)\mu_2$ and $\mu_1 \ll \nu$, $\mu_2 \perp \nu$. The number p and the measures μ_1, μ_2 are uniquely determined.

We use these theorems to define conditional expectation. Let $(X, \mathcal{M}, \mu)$ be a measure space and $\mathcal{C}$ a sub-σ-algebra of $\mathcal{M}$. The goal is to define the operator $E(\cdot|\mathcal{C}) : L^1(X, \mathcal{M}, \mu) \to L^1(X, \mathcal{C}, \mu)$. If $f \in L^1(X, \mathcal{M}, \mu) \cap L^+(X, \mathcal{M}, \mu)$, then $\mu_f(E) = a^{-1} \int_C f d\mu$ (where $a = \int f d\mu$) defines a probability measure μ_f on $(X, \mathcal{C}, \mu)$ with $\mu_f \ll \mu$.

Problem 129. Prove this fact. That is, prove that μ_f defines a probability measure so that $\mu_f \ll \mu$.

Proof. To be a measure it needs to satisfy two conditions. These conditions are satisfied trivially by the definition of integrals, though. To see it's a probability measure, we have $\mu_f(X) = a^{-1}\int_X f d\mu = aa^{-1} = 1$. The second condition also follows trivially by how we define integration; $\mu(E) = 0$ implies $\int_E f d\mu = 0$. □

So by Radon-Nikodym there is a function $g \in L^1(X, \mathcal{C}, \mu)$ so that $\mu_f(E) = \int_E g d\mu$. Thus define $E(f|\mathcal{C}) = g$. By uniqueness, this operator is well-defined. Note that

$$\int_E f d\mu = \int_E g d\mu = \int_E E(f|\mathcal{C}) d\mu \text{ for } C \in \mathcal{C}.$$

Problem 130. Show that $E(\cdot|\mathcal{C})$ is additive on the positive function and $E(cf|\mathcal{C}) = cE(f|\mathcal{C})$ for $c > 0$.

Proof. Notice that

$$\mu_{f+g}(E) = \int_E (f+g) d\mu = \mu_f(E) + \mu_g(E)$$

for all E. Thus $\mu_{f+g} = \mu_f + \mu_g$. By this relation, we have

$$\int_E (E(f|\mathcal{C}) + E(g|\mathcal{C})) d\mu = \int_E E(f+g|\mathcal{C}) d\mu.$$

Since these are positive functions and this equality holds for all $E \in \mathcal{C}$, we get that they are almost everywhere the same. Thus $E(f|\mathcal{C}) + E(g|\mathcal{C}) = E(f+g|\mathcal{C})$. The scaling property follows by the same argument. □

For arbitrary f, write $f = f^+ - f^-$ and define $E(f|\mathcal{C}) = E(f^+|\mathcal{C}) - E(f^-|\mathcal{C})$. Similarly works for complex valued functions. This gives us that $E(\cdot|\mathcal{C})$ is an operator.

Problem 131. Prove the remaining properties.

(1) If $f \in L^1(X, \mathcal{M}, \mu)$ and g is $\mathcal{C}$-measurable and bounded, then

$$E(fg|\mathcal{C}) = gE(f|\mathcal{C}).$$

(2) Show

$$|E(f|\mathcal{C})| \leq E(|f||\mathcal{C}).$$

(3) Show that if $\mathcal{C}_2 \subseteq \mathcal{C}_1$, then

$$E(E(f|\mathcal{C}_1)|\mathcal{C}_2) = E(f|\mathcal{C}_2)$$

for $f \in L^1(X, \mathcal{M}, \mu)$.

Proof.

(1) Notice that for all $K \in \mathcal{C}$ we have

$$\int_K E(f|\mathcal{C}) d\mu = \int_K f d\mu.$$

Fixing $K \in \mathcal{C}$ and taking $E \in \mathcal{M}$ arbitrary we get that

$$\int_E E(f|\mathcal{C}) \chi_K d\mu = \int_E f \chi_K d\mu.$$

Thus we have $E(f|\mathcal{C})\chi_K = f\chi_K$ almost everywhere for every $K \in \mathcal{C}$. Thus

$$\int_K E(fg|\mathcal{C}) d\mu = \int_K fg d\mu = \int f \chi_K g d\mu = \int_K E(f|\mathcal{C}) g d\mu.$$

This holds for every $K \in \mathcal{C}$. Finally notice that since g is $\mathcal{C}$ measurable we have that the support of g is in $\mathcal{C}$, so taking arbitrary $K \in \mathcal{M}$ we can use the above equality to get

$$\int_K E(fg|\mathcal{C})d\mu = \int_K E(f|\mathcal{C})gd\mu.$$

Thus $E(fg|\mathcal{C}) = E(f|\mathcal{C})g$ almost everywhere.

(2) Taking $f = f^+ - f^-$, we see that $E(f|\mathcal{C}) = E(f^+|\mathcal{C}) - E(f^-|\mathcal{C})$, so

$$|E(f|\mathcal{C})| = |E(f^+|\mathcal{C}) - E(f^-|\mathcal{C})| \leq E(f^+|\mathcal{C}) + E(f^-|\mathcal{C}) = E(|f||\mathcal{C}).$$

(3) We see that $E(E(f|\mathcal{C}_1)|\mathcal{C}_2)$ is the unique function in $\mathcal{C}_2$ so that for every $K \in \mathcal{C}_2$ we have

$$\int_K E(E(f|\mathcal{C}_1)|\mathcal{C}_2)d\mu = \int_K E(f|\mathcal{C}_1)d\mu = \int_K fd\mu = \int_K E(f|\mathcal{C}_2)d\mu.$$

Since the support of $E(f|\mathcal{C}_2)$ and $E(E(f|\mathcal{C}_1)|\mathcal{C}_2)$ lies in $\mathcal{C}_2$, we get that this holds for all K measurable, so these things are equal almost everywhere (and thus equal). The choice of $f \in L^1(X, \mathcal{M}, \mu)$ was arbitrary.

□

The goal is to capture the amount of randomness or uncertainty a transformation T generates on a probability measure space $(X, \mathcal{M}, \mu)$. This will be some quantitative value $h(T)$ which represents the **entropy** of the transformation. We want $h(T)$ to have two properties:

(1) The amount of information gained by an application of T is proportional to the amount of uncertainty removed.
(2) We have $h(T)$ is an isomorphism invariant.

Let α be a partition of our space X. We define the **entropy of the partition** by

$$H(\alpha) := -\sum_{i=1}^{n} \mu(A_i)\log(\mu(A_i)).$$

Problem 132. Let

$$f(t) = \begin{cases} -t\log(t) \text{ if } 0 < t \leq 1 \\ 0 \text{ if } t = 0. \end{cases}$$

Show that f is continuous, nonnegative, and concave downward. Moreover, show that for $\lambda_1, \dots, \lambda_n$, we have

$$\frac{1}{n}\sum_{i=1}^{n} f(\lambda_i) \leq f\left(\frac{1}{n}\sum_{i=1}^{n}\lambda_i\right).$$

Proof. Continuity follows if we can show that $\lim_{t\to 0^+} f(t) = 0$. Notice that

$$\lim_{t\to 0^+} f(t) = \lim_{t\to 0^+} -t\log(t) = \lim_{t\to 0^+} -\frac{\log(t)}{1/t}.$$

We apply L'Hospital to get that this is 0, as desired. For nonnegative, we have that $\log(t) \leq 0$ for $0 < t \leq 1$, so $-t\log(t) \geq 0$ for $0 < t \leq 1$. Concave downward follows from taking derivatives. Use definitions for the moreover part (see here). □

Problem 133.

(1) Show that

$$0 \leq H(\alpha) < \infty.$$

(2) Establish $H(\alpha) \leq \log(n)$, where $\alpha = \{A_1, \dots, A_n\}$. Thus we really have $0 \leq H(\alpha) \leq \log(n)$.

Proof.

(1) Notice that $0 \leq \mu(A_i) < 1$, so $\log(\mu(A_i)) < 0$ (modulo the case $\mu(A_i) = 0$, to be addressed), hence $-\log(\mu(A_i)) > 0$, and $-\mu(A_i)\log(\mu(A_i)) > 0$. In the case $\mu(A_i) = 0$, we have by convention $-\mu(A_i)\log(\mu(A_i)) = 0$. This shows that $H(\alpha) \geq 0$. Wlog, assume $\mu(A_i) > 0$ (just throw out the 0 ones since they won't help with an upper bound). Let A_j be such that $\mu(A_j) \leq \mu(A_i)$ for all i. Then
$$H(\alpha) \leq -n\log(\mu(A_j)) < \infty.$$
(2) We can improve the bound in (1). Let $\lambda_i = \mu(A_i)$. Let $f(t) = -t\log(t)$. Then
$$\frac{1}{n}\sum_{i=1}^{n} f(\lambda_i) \leq f\left(\frac{1}{n}\sum_{i=1}^{n}\lambda_i\right) = f(1/n) = \frac{1}{n}\log(n).$$

□

We define the information content of a set to be $I(A) = -\log(\mu(A))$. The information function of a (countable even) measurable partition α is given by
$$I(\alpha)(x) = \sum_{A\in\alpha} I(A)\chi_A(x).$$

Problem 134. Show that
$$\int I(\alpha)d\mu = H(\alpha).$$

Proof. We see
$$E(I(\alpha)) = \int I(\alpha)d\mu = \sum_{i=1}^{n} -\log(\mu(A_i))\int \chi_{A_i}d\mu = \sum_{i=1}^{n} -\log(\mu(A_i))\mu(A_i) = H(\alpha).$$

□

Let α, β be two partitions of (X,μ). Define the **conditional entropy of** α **given** β by
$$H(\alpha|\beta) = -\sum_{A\in\alpha}\sum_{B\in\beta}\log\left(\frac{\mu(A\cap B)}{\mu(B)}\right)\mu(A\cap B)$$
using the convention $0\log(0) = 0$. This is interpreted as the average uncertainty about which element of the partition α the point x will enter if we already know which element of β the point will enter.

A useful trick will be **Jensen's inequality.**

Problem 135 (Jensen's Inequality)**.** Suppose $g : (X,\mathcal{M},\mu) \to \mathbb{R}$ is integrable, $(X,\mathcal{M},\mu)$ a probability measure space, and φ is a convex function on the real line. Show
$$\varphi(E(g|\mathcal{F})) \leq E(\varphi(g)|\mathcal{F}).$$
Deduce from this the usual Jensen's inequality
$$\varphi\left(\int_X g d\mu\right) \leq \int_X \varphi\circ g d\mu.$$

Proof. Since φ is convex, we get that
$$\varphi(x) = \sup_{\substack{h\leq\varphi \\ h \text{ is linear}}} h(x).$$
Notice that if $E(\varphi(g)|\mathcal{F}) = \infty$, then the result clearly follows, so assume it is finite. Then for $h \leq \varphi$ linear we have
$$E(\varphi(g)|\mathcal{F}) \geq E(h(g)|\mathcal{F}) = h(E(g|\mathcal{F})).$$

Notice this was arbitrary, so

$$E(\varphi(g)|\mathcal{F}) \geq \sup_{\substack{h \leq \varphi \\ h \text{ is linear}}} h(E(g|\mathcal{F})) = \varphi(E(g|\mathcal{F})).$$

Now let $\mathcal{F} = \{\varnothing, X\}$ be the trivial σ-algebra. Then $E(g|\mathcal{F})$ defined on this must be a constant function (say C), and we have

$$\int_X E(g|\mathcal{F})(x) d\mu(x) = C = \int_X g d\mu.$$

So $E(g|\mathcal{F})(x) = \int_X g d\mu$. Applying what we have, this tells us that

$$E(\varphi(g)|\mathcal{F}) = \int_X \varphi(g) d\mu \geq \varphi(E(g|\mathcal{F})) = \varphi\left(\int_X g d\mu\right).$$

□

Problem 136. Prove the following for α, β, γ partitions of X.

(1) $H(T^{-1}(\alpha)) = H(\alpha)$.
(2) $H(\alpha \vee \beta) = H(\alpha) + H(\beta|\alpha)$.
(3) $H(\beta|\alpha) \leq H(\beta)$.
(4) $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$.
(5) If $\alpha \leq \beta$ (i.e. a refinement) then $H(\alpha) \leq H(\beta)$.

Proof.

(1) We see that

$$H(T^{-1}(\alpha)) = -\sum_{i=1}^{n} \mu(T^{-1}(A_i)) \log(\mu(T^{-1}(A_i))) = -\sum_{i=1}^{n} \mu(A_i) \log(\mu(A_i)) = H(\alpha)$$

since T is measure preserving.

(2) Let $\alpha = \{A_1, \ldots, A_n\}$, $\beta = \{B_1, \ldots, B_k\}$. Then

$$\alpha \vee \beta = \{A_i \cap B_j : 1 \leq i \leq n, 1 \leq j \leq k\}.$$

Notice

$$\begin{aligned}
H(\alpha \vee \beta) = -\sum_{i,j=1}^{i=n,j=k} \mu(A_i \cap B_j) \log(\mu(A_i \cap B_j)) &= \sum_{i=1}^{n} \left(-\sum_{j=1}^{k} \mu(A_i \cap B_j) \log(\mu(A_i \cap B_j))\right) \\
&= \sum_{i=1}^{n} \left(-\sum_{j=1}^{k} \mu(A_i \cap B_j)[\log(\mu(A_i \cap B_j)) - \log(\mu(A_i)) + \log(\mu(A_i))]\right) \\
&= -\sum_{i=1}^{n} \left(\sum_{j=1}^{k} \mu(A_i \cap B_j) \left[\log\left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right) + \log(\mu(A_i))\right]\right) \\
&= -\sum_{i=1}^{n} \sum_{j=1}^{k} \mu(A_i \cap B_j) \log\left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right) - \sum_{i=1}^{n} \sum_{j=1}^{k} \mu(A_i \cap B_j) \log(\mu(A_i)) \\
&= -\sum_{i=1}^{n} \sum_{j=1}^{k} \mu(A_i \cap B_j) \log\left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right) - \sum_{i=1}^{n} \mu(A_i) \log(A_i) \\
&= H(\alpha|\beta) + H(\alpha).
\end{aligned}$$

(3) Define $f(t) = -t\log(t)$. We see that

$$H(\beta|\alpha) = -\sum_{i=1}^{n}\sum_{j=1}^{k}\mu(A_i\cap B_j)\log\left(\frac{\mu(A_i\cap B_j)}{\mu(A_i)}\right)$$
$$= -\sum_{i=1}^{n}\sum_{j=1}^{k}\mu(A_i)\frac{\mu(A_i\cap B_j)}{\mu(A_i)}\log\left(\frac{\mu(A_i\cap B_j)}{\mu(A_i)}\right)$$
$$= \sum_{j=1}^{k}\sum_{i=1}^{n}\mu(A_i)f\left(\frac{\mu(A_i\cap B_j)}{\mu(A_i)}\right)$$
$$\leq \sum_{j=1}^{k} f\left(\sum_{i=1}^{n}\mu(A_i)\frac{\mu(A_i\cap B_j)}{\mu(A_i)}\right)$$
$$= \sum_{j=1}^{k} f(\mu(B_j)) = -\sum_{j=1}^{k}\mu(B_j)\log(\mu(B_j)) = H(\beta).$$

(4) Combine (2) and (3) to get

$$H(\alpha\vee\beta) = H(\alpha) + H(\beta|\alpha) \leq H(\alpha) + H(\beta).$$

(5) If α is a refinement, then elements of α are unions of elements in β, so $\alpha\vee\beta = \beta$. Thus

$$H(\alpha\vee\beta) = H(\beta) = H(\alpha) + H(\beta|\alpha).$$

Thus

$$H(\alpha) \leq H(\beta).$$

□

We define the **conditional information function** of a countable partition α given a sub-σ-algebra $\mathcal{F}\subseteq\mathcal{M}$ to be

$$I(\alpha|\mathcal{F})(x) = -\sum_{A\in\alpha}\log(\mu(A|\mathcal{F})(x))\chi_A(x),$$

where

$$\mu(A|\mathcal{F}) = E(\chi_A|\mathcal{F})$$

is a function.

Note that if α a partition and $\mathcal{F}$ a σ-algebra, we write $\alpha\vee\mathcal{F}$ to mean $\sigma(\alpha)\vee\mathcal{F}$; i.e. it is the join of the σ-algebra generated by α and the σ-algebra $\mathcal{F}$.

Problem 137. Show that

$$\mu(B|\alpha\vee\mathcal{F}) = \sum_{A\in\alpha}\frac{\mu(A\cap B|\mathcal{F})}{\mu(A|\mathcal{F})}\chi_A.$$

Of course this is almost everywhere.

Proof. Take $A'\cap F$, $A'\in\sigma(\alpha)$ and $F\in\mathcal{F}$. Notice that

$$\int_{A'\cap F}\mu(B|\alpha\vee\mathcal{F})d\mu = \int_{A'\cap F}E(\chi_B|\alpha\vee\mathcal{F})d\mu = \int_{A'\cap F}\chi_B d\mu = \int_F \chi_{A'\cap B}d\mu.$$

Notice that this is equivalent to

$$\int_F\chi_{A'\cap B}d\mu = \int_F E(\chi_{A'\cap B}|\mathcal{F})d\mu = \int_F\mu(A'\cap B|\mathcal{F})d\mu = \int_F E(\chi_{A'}|\mathcal{F})\frac{\mu(A'\cap B|\mathcal{F})}{\mu(A'|\mathcal{F})}d\mu.$$

This last equality comes from simply multiplying and dividing (note that $E(\chi_{A'}|\mathcal{F}) = \mu(A'|\mathcal{F})$). We now wish to claim that

$$\int_F E(\chi_{A'}|\mathcal{F})\frac{\mu(A'\cap B|\mathcal{F})}{\mu(A'|\mathcal{F})}d\mu = \int_F E\left(\chi_{A'}\frac{\mu(A'\cap B|\mathcal{F})}{\mu(A'|\mathcal{F})}\middle|\mathcal{F}\right)d\mu.$$

This, however, follows by one of the earlier problems involving conditional expectation. Thus

$$\int_F E(\chi_{A'}|\mathcal{F})\frac{\mu(A'\cap B|\mathcal{F})}{\mu(A'|\mathcal{F})}d\mu = \int_F E\left(\chi_{A'}\frac{\mu(A'\cap B|\mathcal{F})}{\mu(A'|\mathcal{F})}\middle|\mathcal{F}\right)d\mu = \int_F \chi_{A'}\frac{\mu(A'\cap B|\mathcal{F})}{\mu(A'|\mathcal{F})}d\mu.$$

Now recall that all of the $A \in \alpha$ are disjoint, so

$$\int_F \chi_{A'}\frac{\mu(A'\cap B|\mathcal{F})}{\mu(A'|\mathcal{F})}d\mu = \int_{A'\cap F}\sum_{\alpha\in A}\frac{\mu(A\cap B|\mathcal{F})}{\mu(A|\mathcal{F})}\chi_A d\mu.$$

Since this holds for all $A'\cap F \in \alpha\vee\mathcal{F}$ we have the result. □

Problem 138. Prove the following.

(1) $I(\alpha\vee\beta) = I(\alpha) + I(\alpha|\beta)$.
(2) $I(\alpha\vee\beta|\mathcal{F}) = I(\alpha|\mathcal{F}) + I(\beta|\alpha\vee\mathcal{F})$.

Proof.

(1) This will follow from (2) using $\mathcal{F} = \{\varnothing, X\}$. To see this, notice that

$$I(\alpha\vee\beta|\mathcal{F}) = -\sum_{A_i\cap B_j\in\alpha\vee\beta}\log\left(\mu(A_i\cap B_j|\mathcal{F})(x)\right)\chi_{A_i\cap B_j}(x).$$

Recall that for the trivial σ-algebra we have

$$\mu(A_i\cap B_j|\mathcal{F})(x) = \mu(A_i\cap B_j),$$

so

$$I(\alpha\vee\beta|\mathcal{F}) = -\sum_{A_i\cap B_j\in\alpha\vee\beta}\log(\mu(A_i\cap B_j))\chi_{A_i\cap B_j}(x) = I(\alpha\vee\beta).$$

The same kind of argument applies to the other two functions. Thus, assuming (2) we have (1).

(2) Note that since the A are disjoint in α we can move the sum in and out of the log. Thus

$$\begin{aligned}
I(\beta|\alpha\vee\mathcal{F}) &= -\sum_{B\in\beta}\log(\mu(B|\alpha\vee F))\chi_B(x)\\
&= -\sum_{B\in\beta}\log\left(\sum_{A\in\alpha}\frac{\mu(A\cap B|\mathcal{F})}{\mu(A|\mathcal{F})}\chi_A\right)\chi_B\\
&= -\sum_{B\in\beta}\sum_{A\in\alpha}\log\left(\frac{\mu(A\cap B|\mathcal{F})}{\mu(A|\mathcal{F})}\right)\chi_A\chi_B\\
&= -\sum_{B\in\beta}\sum_{A\in\alpha}[\log\left(\mu(A\cap B|\mathcal{F})\right) - \log(\mu(A|\mathcal{F}))]\chi_A\chi_B\\
&= -\sum_{\substack{A\in\alpha\\ B\in\beta}}\log\left(\mu(A\cap B|\mathcal{F})\right)\chi_{A\cap B} + \sum_{A\in\alpha}\log(\mu(A|\mathcal{F}))\chi_A\\
&= I(\alpha\vee\beta|\mathcal{F}) - I(\alpha|\mathcal{F}).
\end{aligned}$$

This gives the result.

□

The **conditional entropy of α given $\mathcal{F}$** is defined by

$$H(\alpha|\mathcal{F}) = \int I(\alpha|\mathcal{F})(x)d\mu(x).$$

Problem 139. Prove the following.

(1) $H(\alpha \vee \beta|\mathcal{F}) = H(\alpha) + H(\beta|\alpha \vee \mathcal{F})$.
(2) If $\mathcal{F}_2 \subseteq \mathcal{F}_1$, then $H(\alpha|\mathcal{F}_1) \leq H(\alpha|\mathcal{F}_2)$.
Hint: Jensen's inequality.
(3) $\mu(T^{-1}(A)|T^{-1}(\mathcal{F}))(x) = \mu(A|\mathcal{F})(Tx)$.
(4) $H(T^{-1}(\alpha)|T^{-1}(\mathcal{F})) = H(\alpha|\mathcal{F})$.

Proof.

(1) Integrate (2) from the last problem.
(2) Let $f = -t\log(t)$. Then
$$E(f \circ \mu(A|\mathcal{F}_1)|\mathcal{F}_2) \leq f \circ E(\mu(A|\mathcal{F}_1)|\mathcal{F}_2) = f \circ E(A|\mathcal{F}_2) = f \circ \mu(A|\mathcal{F}_2).$$
Integrate both sides of the inequality to get
$$\int E(f \circ \mu(A|\mathcal{F}_1)|\mathcal{F}_2)d\mu \leq \int f \circ \mu(A|\mathcal{F}_2)d\mu.$$
Notice that over all $F \in \mathcal{F}_2$, we have that the left hand side is such that
$$\int_F E(f \circ \mu(A|\mathcal{F}_1)|\mathcal{F}_2)d\mu = \int_F f \circ \mu(A|\mathcal{F}_1)d\mu.$$
Thus integrating over $F \in \mathcal{F}_2$ we have
$$\int_F f \circ \mu(A|\mathcal{F}_1)d\mu \leq \int_F f \circ \mu(A|\mathcal{F}_2)d\mu.$$
Writing out the definition grants us
$$\int_F -\log(\mu(A|\mathcal{F}_1))\mu(A|\mathcal{F}_1)d\mu \leq \int_F -\log(\mu(A|\mathcal{F}_2))\mu(A|\mathcal{F}_2)d\mu.$$
Since F_2 a σ-algebra, we have in particular
$$\int_X -\log(\mu(A|\mathcal{F}_1))\mu(A|\mathcal{F}_1)d\mu \leq \int_X -\log(\mu(A|\mathcal{F}_2))\mu(A|\mathcal{F}_2)d\mu.$$
Now sum over all $A \in \alpha$ to get
$$H(\alpha|\mathcal{F}_1) \leq H(\alpha|\mathcal{F}_2).$$
(3) Notice
$$\mu(T^{-1}(A)|T^{-1}(\mathcal{F}))(x) = E(\chi_{T^{-1}(A)}|T^{-1}(\mathcal{F}))(x).$$
Integrate over $F \in T^{-1}(\mathcal{F})$ to get
$$\int_F E(\chi_{T^{-1}(A)}|T^{-1}(\mathcal{F}))(x)d\mu(x) = \int_F \chi_{T^{-1}(A)}(x)d\mu(x) = \int_F \chi_A(Tx)d\mu(x).$$
Notice that for $K \in \mathcal{F}$ we have (using the fact T measure preserving)
$$\int_K E(\chi_A|\mathcal{F})(x)d\mu(x) = \int_K \chi_A(x)d\mu(x) = \int_{T^{-1}(K)} \chi_A(Tx)d\mu(x).$$
Choosing $T^{-1}(K) = F$ gives us the desired result.
(4) Use (3).

□

Two partitions α and β are **independent** if $\mu(A \cap B) = \mu(A)\mu(B)$ for all $A \in \alpha$, $B \in \beta$. We denote it by $\alpha \perp \beta$.

For a measure ν on $(X, \mathcal{M})$ and a partition α define

$$H(\nu, \alpha) = -\sum_{A \in \alpha} \log(\nu(A))\nu(A).$$

Problem 140.

(1) Show that $H(\cdot, \alpha)$ is concave.
(2) Show that it is strictly concave.

Proof. We need to show that for $0 < \gamma < 1$ and μ, ν probability measures we have

$$H(\gamma\mu + (1-\gamma)\nu, \alpha) \geq \gamma H(\mu, \alpha) + (1-\gamma)H(\nu, \alpha).$$

This, however, follows from a simple calculation:

$$H(\gamma\mu + (1-\gamma)\nu, \alpha) = \sum_{A \in \alpha} f(\gamma\mu(A) + (1-\gamma)\nu(A))$$

$$\geq \sum_{A \in \alpha} (\gamma f(\mu(A)) + (1-\gamma) f(\nu(A))) = \gamma H(\mu, \alpha) + (1-\gamma)H(\nu, \alpha).$$

Since f is strictly concave, we have equality iff $\mu = \nu$ on α. □

Problem 141. Show that $H(\alpha|\mathcal{F}) = 0$ iff $\alpha \subseteq \mathcal{F}$ (up to sets of measure 0). Conclude $H(\alpha|\beta) = 0$ iff $\alpha \leq \beta$.

Proof. ($\Longrightarrow$): Assume $H(\alpha|\mathcal{F}) = 0$. This is saying

$$\sum_{A \in \alpha} \int -\log(\mu(A|\mathcal{F}))\chi_A d\mu = 0.$$

Since this is a sum of positive things, each component must be 0. Since they are disjoint, we have

$$-\int \log(\mu(A|\mathcal{F}))\chi_A d\mu = 0.$$

Again, the integral of a positive function is 0, so this implies (almost everywhere) that

$$-\log(\mu(A|\mathcal{F}))\mu(A|\mathcal{F}) = 0.$$

This implies either $\mu(A|\mathcal{F}) = 0$ or 1 for all $x \in X$, so it is a characteristic function. Since $\mu(A|\mathcal{F}) = E(\chi_A|\mathcal{F})$, this implies that for some $F \in \mathcal{F}$ we have $\mu(A|\mathcal{F}) = \chi_F$. Finally, integrating over this characteristic function, we have

$$\mu(F) = \int_F \mu(A|\mathcal{F}) d\mu = \mu(A \cap F),$$

so $A \subseteq F$. On the other hand,

$$\mu(F) = \int \mu(A|\mathcal{F}) d\mu = \mu(A),$$

so $A = F$ (up to a set of measure zero). This holds for all $A \in \alpha$, so $\alpha \subseteq \mathcal{F}$.
($\Longleftarrow$): If $\alpha \subseteq \mathcal{F}$, then $\mu(A|\mathcal{F}) = \chi_A$ almost everywhere (by almost the same argument as above), which gives the result.
If $H(\alpha|\beta) = 0$, then $\alpha \subseteq \sigma(\beta)$ by the above, but this then forces $\alpha \leq \beta$, since elements in α can be written as unions of elements in β. Vice versa is the same. □

Problem 142. Show that $H(\alpha|\beta) = H(\alpha)$ iff $\alpha \perp \beta$.

Throughout, define

$$\mu_B(A) = \frac{\mu(A \cap B)}{\mu(B)}.$$

This is the **conditional probability measure.**

Proof. ($\Longrightarrow$): We see

$$H(\alpha|\beta) = H(\mu, \alpha|\beta) = \sum_{B \in \beta} \mu(B) H(\mu_B, \alpha)$$

$$\leq H\left(\sum_{B \in \beta} \mu(B)\mu_B, \alpha\right) = H(\mu, \alpha) = H(\alpha).$$

By strict concavity, equality only occurs when we have

$$\frac{\mu(A \cap B_i)}{\mu(B_i)} = \frac{\mu(A \cap B_k)}{\mu(B_k)}$$

for all $B_i, B_k \in \beta$. Thus

$$\mu(A) = \sum_{B \in \beta} \mu(A \cap B_i).$$

Fix B_1 wlog. Notice $\mu(A \cap B) = \frac{\mu(B)\mu(A \cap B_1)}{\mu(B_1)}$ for all $B \in \beta$, so

$$\mu(A) = \sum_{B \in \beta} \frac{\mu(B)\mu(A \cap B_1)}{\mu(B_1)} = \frac{\mu(A \cap B_1)}{\mu(B_1)}.$$

Thus $\mu(A)\mu(B_1) = \mu(A \cap B_1)$. Notice that the choices B_1 and A were arbitrary, so in fact $\mu(A \cap B) = \mu(A)\mu(B)$ for all $B \in \beta$ and $A \in \alpha$. This give $\alpha \perp \beta$.
($\Longleftarrow$): Take $\mathcal{F} = \{\varnothing, X\}$. By an earlier problem, we see that

$$\mu(A|\beta) = \mu(A|\beta \vee \mathcal{F}) = \mu(A|\mathcal{F}) = \mu(A).$$

Thus

$$H(\alpha|\beta) = -\sum_{A \in \alpha} \int \log(\mu(A|\beta))\chi_A d\mu = -\sum_{A \in \alpha} \int \log(\mu(A))\chi_A d\mu = H(\alpha).$$

□

For a measure preserving transformation T and a partition α define

$$h(\alpha, T) = \lim_{n \to \infty} \frac{1}{n} H(\alpha \vee T^{-1}(\alpha) \vee \cdots \vee T^{-n+1}(\alpha)).$$

The heuristic for this is that it measures the entropy of the transformation T with respect to the partition α. In other words, it measures the average uncertainty per time on which element of α the point x will enter under T given it's history. An important question is whether this exists.

Problem 143. For each *countable* partition α, $h(\alpha, T)$ exists (it may be ∞).

Proof. The gist is to apply Problem 15. To do so, we define

$$H_n = H(\alpha \vee T^{-1}(\alpha) \vee \cdots \vee T^{-n+1}(\alpha)).$$

The claim is that this is an increasing, subadditive sequence of non-negative real numbers. By earlier problems, we get H_n is increasing and non-negative. Thus we just need to show subadditivity. However, we see

$$\begin{aligned} H_{n+m} &= H(\alpha \vee T^{-1}(\alpha) \vee \cdots \vee T^{-n+1}(\alpha) \vee T^{-n}(\alpha) \vee \cdots \vee T^{-m+n+1}(\alpha)) \\ &\leq H(\alpha \vee T^{-1}(\alpha) \vee \cdots \vee T^{-n+1}(\alpha)) + H(T^{-n}(\alpha) \vee \cdots \vee T^{-m+n+1}(\alpha)) \\ &= H(\alpha \vee T^{-1}(\alpha) \vee \cdots \vee T^{-n+1}(\alpha)) + H(\alpha \vee \cdots \vee T^{-m+1}(\alpha)) = H_n + H_m. \end{aligned}$$

Thus we have subadditivity. We can then apply the result ☐

Let $(X, \mathcal{M}, \mu)$ be a probability measure space. Let $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots$ be an increasing sequence of sub-σ-algebras of $\mathcal{F}$. A sequence $X_1, \dots$ of functions in $L^1(\mu)$ such that X_n is measurable with respect to $\mathcal{F}_n$ for $n = 1, 2, \dots$ is called

(1) a **submartingale** if $E(X_{n+1}|\mathcal{F}_n) \geq X_n$ a.e.,
(2) a **martingale** if $E(X_{n+1}|\mathcal{F}_n) = X_n$ a.e.,
(3) a **supermartingale** if $E(X_{n+1}|\mathcal{F}_n) \leq X_n$ a.e.

Theorem (Doob's Martingale Convergence Theorem)**.** If $\{X_n\}$ is an L^1 submartingale which is bounded in the sense of $\sup_n E(|X_n|) < \infty$, then it converges a.e.

Problem 144. Let $\mathcal{M}_1 \subseteq \mathcal{M}_2 \subseteq \cdots$ be a sequence of sub-σ-algebras on $\mathcal{M}$, and let $\bigvee_{i=1}^\infty \mathcal{M}_i = \mathcal{M}_\infty$. If α is a finite partition, then

$$\lim_{n\to\infty} H(\alpha|\mathcal{M}_n) = H(\alpha|\mathcal{M}_\infty).$$

Proof. Fix $A \in \alpha$ arbitrary. Our submartingale will be the family $\{X_n = \mu(A|\mathcal{M}_n)\}$. Notice

$$E(X_{n+1}|\mathcal{M}_n) = E(\mu(A|\mathcal{M}_{n+1})|\mathcal{M}_n) = E(E(\chi_A|\mathcal{M}_{n+1})|\mathcal{M}_n) = E(\chi_A|\mathcal{M}_n) = \mu(A|\mathcal{M}_n) = X_n.$$

So this is actually a submartingale (in fact, a martingale). We see that

$$E(|X_n|) = E(\mu(A|\mathcal{M}_n)) = \int \mu(A|\mathcal{M}_n)(x) d\mu = \int \chi_A d\mu = \mu(A).$$

Thus it is bounded. By Doob, it converges almost everywhere. Moreover we see

$$\mu(A|\mathcal{M}_n) \to \mu(A|\mathcal{M}_\infty) \text{ a.e.}$$

Now $f(t) = -t\log(t)$ is a bounded continuous function, so

$$f(\mu(A|\mathcal{M}_n)) \to f(\mu(A|\mathcal{M}_\infty)).$$

Use the bounded convergence theorem to get

$$\int f(\mu(A|\mathcal{M}_n)) d\mu \to \int f(\mu(A|\mathcal{M}_\infty)) d\mu.$$

Sum over all $A \in \alpha$ to get the result. ☐

Problem 145. Assume α is a finite partition. Show that

$$h(\alpha, T) := \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^{-k}(\alpha)\right) = \lim_{n\to\infty} H\left(\alpha \middle| \bigvee_{k=1}^{n} T^{-k}(\alpha)\right) = H\left(\alpha \middle| \bigvee_{k=1}^{\infty} T^{-k}(\alpha)\right).$$

Proof. Let $\beta = \bigvee_{k=1}^n T^{-k}(\alpha)$. Recall that

$$H(\alpha|\beta) = H(\alpha \vee \beta) - H(\beta).$$

Using the definition of β, we have

$$H\left(\alpha\middle|\bigvee_{k=1}^{n} T^{-k}(\alpha)\right) = H\left(\bigvee_{k=0}^{n} T^{-k}(\alpha)\right) - H\left(\bigvee_{k=1}^{n} T^{-k}(\alpha)\right).$$

If we sum over n we have a telescoping series. Thus

$$\sum_{n=1}^{j} H\left(\alpha\middle|\bigvee_{k=1}^{n} T^{-k}(\alpha)\right) = H\left(\bigvee_{k=0}^{j} T^{-k}(\alpha)\right) - H(\alpha).$$

From earlier problems, we know that $H\left(\alpha\middle|\bigvee_{k=1}^{n} T^{-k}(\alpha)\right)$ is nonnegative and decreasing, so its limit exists as $n \to \infty$. Thus the Cesaro averages converge to the same point. This gives us

$$\lim_{j\to\infty} \frac{1}{j+1} \sum_{n=1}^{j} H\left(\alpha\middle|\bigvee_{k=1}^{n} T^{-k}(\alpha)\right) = \lim_{j\to\infty} H\left(\alpha\middle|\bigvee_{k=1}^{j} T^{-k}(\alpha)\right).$$

Moreover,

$$\lim_{j\to\infty} H\left(\alpha\middle|\bigvee_{k=1}^{j} T^{-k}(\alpha)\right) = \lim_{j\to\infty} \frac{1}{j+1} H\left(\bigvee_{k=0}^{j} T^{-k}(\alpha)\right) = h(\alpha, T).$$

Finally the last equality follows from the previous problem. □

Problem 146 (Petersen 5.2.1)**.** Show that $\alpha \leq \beta$ implies $I(\alpha|\mathcal{F}) \leq I(\beta|\mathcal{F})$.

Proof. Like before, notice $\alpha \leq \beta$ implies $\alpha \vee \beta = \beta$, so

$$I(\beta|\mathcal{F}) = I(\alpha|\mathcal{F}) + I(\beta|\alpha \vee \mathcal{F}).$$

Since $I \geq 0$, this gives us the result. □

Problem 147. For any countable measurable partitions α and β, show

$$h(\beta, T) \leq h(\alpha, T) + H(\beta|\alpha).$$

Proof. Write

$$\beta_0^{m-1} = \bigvee_{k=0}^{m-1} T^{-k}(\beta), \qquad \alpha_0^{n-1} = \bigvee_{k=0}^{n-1} T^{-k}(\alpha).$$

Notice

$$H(\beta \vee T^{-1}(\beta)|\alpha_0^{n-1}) = H(\beta) + H(T^{-1}(\beta)|\beta \vee \alpha_0^{n-1}).$$

We have

$$H(\beta) \leq H(\beta|\alpha_0^{n-1}),$$

and

$$H(T^{-1}(\beta)|\beta \vee \alpha_0^{n-1}) \leq H(T^{-1}(\beta)|\alpha_0^{n-1}),$$

so

$$H(\beta_0^2|\alpha_0^{n-1}) \leq H(\beta|\alpha_0^{n-1}) + H(T^{-1}(\beta)|\alpha_0^{n-1}).$$

An induction argument establishes

$$H(\beta_0^{n-1}|\alpha_0^{n-1}) \leq \sum_{k=0}^{n-1} H(T^{-k}(\beta)|\alpha_0^{n-1}).$$

Since $T^{-k}(\alpha) \leq \alpha_0^{n-1}$ for $0 \leq k \leq n-1$, we have

$$H(\beta_0^{n-1}|\alpha_0^{n-1}) \leq \sum_{k=0}^{n-1} H(T^{-k}(\beta)|T^{-k}(\alpha)) = nH(\beta|\alpha).$$

Finally $\beta_0^{n-1} \leq \alpha_0^{n-1}$, so

$$H(\beta_0^{n-1}) \leq H(\beta_0^{n-1} \vee \alpha_0^{n-1}) = H(\alpha_0^{n-1}) + H(\beta_0^{n-1}|\alpha_0^{n-1}) \leq H(\alpha_0^{n-1}) + nH(\beta|\alpha).$$

Dividing by n and taking the limit gives the result. □

We call a finite partition α a **generator** with respect to T in the case that $\alpha_{-\infty}^{\infty} = \mathcal{M}$ (i.e. the countable join generates the σ-algebra.

Problem 148 (Kolmogorov-Sinai)**.** Show that if α is a generator with respect to T, then

$$h(T) = h(\alpha, T).$$

Proof. Let β be a finite partition. It suffices (by earlier arguments) to show that $h(\beta, T) \leq h(\alpha, T)$ for all possible β. Notice that

$$h(\beta, T) \leq h(\alpha_{-n}^{n}, T) + H(\beta|\alpha_{-n}^{n}) \leq h(\alpha, T) + H(\beta|\alpha_{-n}^{n}).$$

Notice $\beta \leq \alpha_{-\infty}^{\infty}$, so

$$\lim_{n\to\infty} H(\beta|\alpha_{-n}^{n}) = H(\beta|\alpha_{-\infty}^{\infty}) = 0.$$

Taking the limit as $n \to \infty$ in the first inequality gives the result. □

Problem 149. Consider the Bernoulli scheme $\mathcal{B}(p_1, \ldots, p_n)$ on the alphabet $\mathcal{A} = \{a_1, \ldots, a_n\}$. Let

$$A_i = \{x = (\ldots, x_{-1}, x_0, x_1, \ldots) : x_0 = a_i\}, \qquad i = 1, \ldots, n$$

be the time-zero cylinder sets.

(1) Show that $\alpha = \{A_i\}_{i=1}^{n}$ forms a measurable partition.
(2) Show that α is a generator.
(3) Calculate the entropy.

Proof.

(1) These sets are measurable by construction. Take $x \in \mathcal{A}^{\mathbb{Z}}$. Define $[\cdot]_0 : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}$ by $[x]_j = x_j$, where $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$. Then $[x]_0 = a_i$ for some i, so $x \in A_i$. Now $A_i \cap A_j = \{x : [x]_0 = a_i, [x]_0 = a_j\}$. For $i \neq j$, we have that this must be empty, so α indeed partitions the space.
(2) Define $\sigma : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ by $\sigma(x) = y$, with $y_i = x_{i+1}$ (i.e. the right shift). Examine $\bigvee_{-\infty}^{\infty} \mathcal{M}(\sigma^{-n}(\alpha))$. Notice that this is a σ-algebra containing every measurable cylinder, so this is the entire σ-algebra. Therefore α is a generator.
(3) By Kolmogorov-Sinai, it suffices to examine the entropy of a generator. Thus we need to calculate

$$h(\alpha, \sigma) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} \sigma^{-j}(\alpha)\right).$$

Notice elements in $\bigvee_{j=0}^{n-1} \sigma^{-j}(\alpha)$ are of the form

$$A_{i_1} \cap \sigma^{-1}(A_{i_2}) \cap \cdots \cap \sigma^{n-1}(A_{i_n}) = \{x \in \mathcal{A}^{\mathbb{Z}} : [x]_0 = a_{i_1}, \ldots, [x]_n = a_{i_n}\}.$$

We claim $\alpha \vee \sigma^{-1}(\alpha)$ are independent. This is clear by the observation above and the fact that the measure is the product measure. Hence by an induction argument we see that $\bigvee_{j=0}^{n-1} \sigma^{-j}(\alpha)$ are independent. By **Petersen 5.2.9** and **Petersen 5.2.3'** this implies that

$$H\left(\bigvee_{j=0}^{n-1} \sigma^{-j}(\alpha)\right) = \sum_{j=0}^{n-1} H(\sigma^{-j}(\alpha)) = \sum_{j=0}^{n-1} H(\alpha) = nH(\alpha).$$

Therefore

$$h(\sigma) = h(\alpha, \sigma) = H(\alpha) = -\sum_{i=1}^{n} p_i \log_2(p_i).$$

□

Problem 150 (Petersen 5.3.4)**.** Show that for any $r > 0$ there is a Bernoulli shift of entropy r.

Proof. Fix $r > 0$. The goal is to find a vector $(p_1, \ldots, p_n)$ and n so that

$$\sum_{i=1}^{n} -p_i \log(p_i) = r, \qquad \sum_{i=1}^{n} p_i = 1.$$

Notice

$$\sum_{i=1}^{n} -p_i \log(p_i) = r \Leftrightarrow \sum_{i=1}^{n} \log(p_i^{-p_i}) = r \Leftrightarrow \exp\left(\sum_{i=1}^{n} \log(p_i^{-p_i})\right) = \exp(r)$$
$$\Leftrightarrow \prod_{i=1}^{n} p_i^{-p_i} = \exp(r).$$

Without loss of generality just assume that we have

$$\prod_{i=1}^{n} p_i^{-p_i} = r, \qquad \sum_{i=1}^{n} p_i = 1.$$

Replace $p_n = 1 - \sum_{i=1}^{n-1} p_i$, then

$$\prod_{i=1}^{n-1} p_i^{-p_i} \cdot \left(1 - \sum_{i=1}^{n-1} p_i\right)^{-1+\sum_{i=1}^{n-1} p_i} = r.$$

Consider $p_1 = \cdots = p_{n-1} = p$. Then rewrite the left hand side of the above as

$$f(x) = (x^{-nx})(1 - (n-1)x)^{-1+(n-1)x}.$$

Notice $f(x)$ is continuous on the interval $0 \le x < \frac{1}{n-1}$ (assuming $n > 1$). Moreover, it is differentiable on this interval, with derivative

$$f'(x) = (1 - (n-1)x)^{(n-1)x-1} x^{-nx}(-n\log(x) + (n-1)\log(-nx + x + 1) - 1).$$

We see this is well-defined and continuous on the interval $0 \le x < \frac{1}{n-1}$. Noting that $f(0) = 1$ always and it increases until it hits a maximum (by looking at $f'(x)$), we just need to determine what the maximum is with respect to n. Noting

$$f(1/n) = n\left(1 - \frac{n-1}{n}\right)^{\frac{n-1}{n}-1}$$

and $1/n < 1/n - 1$, we see that for $r \ge 1$ we can find an n so that $f(x) = r$. Recall we replaced r with $\exp(r)$, so translating we can find an n and x so that $f(x) = \exp(r)$. But by taking logarithms, we have $\log(f(x)) = \sum_{i=1}^{n} -p_i \log(p_i) = r$, as desired. □

We say that T has a **one-sided generator** if there is a finite partition α so that $\alpha_1^\infty = \mathcal{M}$ up to sets of measure 0. The existence of this partition means that, in some sense, the present and future of the system $(X, \mathcal{M}, \mu, T)$ are completely determined by its past.

Problem 151. Show that if T has a one-sided generator then the entropy is 0.

Proof. We have

$$h(T) = h(\alpha, T) = h(\alpha|\alpha_1^\infty) = h(\alpha|\mathcal{M}) = 0.$$

□

Problem 152. Show that every discrete spectrum system has entropy zero.

Proof. We use **Petersen 2.4.10**. There exists a sequence of integers $\{n_k\}$ so that $T^{n_k} f \to f$ in L^2 for each $f \in L^2$. Let $\alpha = \{A_1, \ldots, A_n\}$ and define $f(x) = i$ if $x \in A_i$. We see that $T^{n_k} f \to f$ implies that $\alpha \subseteq \alpha_1^\infty = \bigvee_{k=1}^\infty \sigma(T^{-k}(\alpha))$, at least up to sets of measure 0. Thus

$$h(\alpha, T) = H(\alpha|\alpha_1^\infty) = 0.$$

This holds for all finite partitions. □

Problem 153. Let G be a compact abelian group. Show that $T : G \to G$ defined by $T_g(h) = gh$ has zero entropy (i.e. show $h(T_g) = 0$).

Proof. We note that T_g has discrete spectrum since $L^2(m)$ is spanned by $\widehat{G}$ (this follows by Stone-Weierstrass). □

Problem 154. Let T be the left shift on the set $X = \{1, 2, \ldots, n\}^{\mathbb{Z}}$ endowed with the σ-algebra $\mathcal{M}$ generated by the cylinder sets (i.e. the usual Borel one) and the Markov measure μ given by the stochastic matrix $P = (p_{ij})$ and the probability vector $\pi = (\pi_1, \ldots, \pi_n)$ so that $\pi P = \pi$. Show that

$$h(T) = -\sum_{j=1}^n \sum_{i=1}^n \pi_i p_{ij} \log(\pi_{ij}).$$

Proof. Let's recall how the stochastic matrix works in this context. We define the cylinders by

$$C_{x_1,\ldots,x_n}^{i_1,\ldots,i_n} = \{x \in X : x_{i_j} = x_j\}.$$

We then have the measure is defined by

$$\mu(C_{x_1,\ldots,x_n}^{i_1,\ldots,i_n}) = \pi_{i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}.$$

For notational simplicity, denote by $C_z := C_z^0$ a time-zero cylinder. We follow the Bernoulli scheme argument now. A generator for our σ-algebra is the time-zero partition,

$$\alpha := \{C_1, \ldots, C_n\}.$$

We now observe

$$\begin{aligned}\bigvee_{i=0}^m T^{-i}\alpha &= \left\{C_{i_0} \cap T^{-1}(C_{i_1}) \cap \cdots \cap T^{-m}(C_{i_m}) : i_k \in X, 1 \le k \le m\right\} \\ &= \{C_{i_0,\ldots,i_m}^{0,1,\ldots,m} : i_k \in X, 1 \le k \le n\}.\end{aligned}$$

Notice that $\sigma(\bigvee_{i=0}^\infty T^{-i}(\alpha)) = \mathcal{M}$, so α is indeed a generator. Next, we notice that

$$H\left(\bigvee_{i=0}^m T^{-i}\alpha\right) = \sum_{i=0}^m H(T^{-i}(\alpha)).$$

We calculate $H\left(\bigvee_{i=0}^m T^{-i}(\alpha)\right)$. Notice

$$\begin{aligned}H(\alpha) &= -\sum_{i=1}^n \mu(C_i) \log(\mu(C_i)) \\ &= -\sum_{i=1}^n \pi_i \log(\pi_i),\end{aligned}$$

$$H(T^{-1}(\alpha)) = -\sum_{i=1}^{n} \mu(C_i^2)\log(\mu(C_i^2))$$
$$= -\sum_{i=1}^{n} \pi_i \log(\pi_i),$$
$$H(\alpha \vee T^{-1}(\alpha)) = -\sum_{i_1,i_2=1}^{n} \pi_{i_1} p_{i_1 i_2} \log(\pi_{i_1} p_{i_1 i_2}).$$

We try to continue this idea. Let $f(t) = -t\log(t)$. Then

$$\begin{aligned}
H\left(\bigvee_{i=0}^{m} T^{-i}(\alpha)\right) &= \sum_{i_0,\dots,i_m=1}^{n} f(\pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m}) \\
&= -\sum_{i_0,\dots,i_m=1}^{n} \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m} \log(\pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m}) \\
&= -\sum_{i_0,\dots,i_m=1}^{n} \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-2} i_{m-1}} \left[p_{i_{m-1} i_m} \log(\pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-2} i_{m-1}}) + p_{i_{m-1} i_m} \log(p_{i_{m_1} i_m})\right] \\
&= \sum_{i_0,\dots,i_m=1}^{n} p_{i_{m-1} i_m} f(\pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-2} i_{m-1}}) + \sum_{i_0,\dots,i_m=1}^{n} \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-2} i_{m-1}} f(p_{i_{m-1} i_m}) \\
&= \sum_{i_0,\dots,i_{m-1}=1}^{n} \left(\sum_{i_m=1}^{n} p_{i_{m-1} i_m}\right) f(\pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-2} i_{m-1}}) \\
&\quad + \sum_{i_{m-1},i_m=1}^{n} \left(\sum_{i_0,\dots,i_{m-2}=1}^{n} \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-2} i_{m-1}}\right) f(p_{i_{m-1} i_m}) \\
&= \sum_{i_0,\dots,i_{m-1}=1}^{n} f(\pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-2} i_{m-1}}) + \sum_{i_{m-1},i_m=1}^{n} \pi_{i_{m-1}} f(p_{i_{m-1} i_m}).
\end{aligned}$$

So to conclude, we have

$$\sum_{i_0,\dots,i_m=1}^{n} f(\pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m}) = \sum_{i_0,\dots,i_{m-1}=1}^{n} f(\pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-2} i_{m-1}}) + \sum_{i_{m-1},i_m=1}^{n} \pi_{i_{m-1}} f(p_{i_{m-1} i_m}).$$

By a recursion argument, we get

$$H\left(\bigvee_{i=0}^{m} T^{-i}(\alpha)\right) = \sum_{i_0,\dots,i_m=1}^{n} f(\pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m}) = \sum_{i=1}^{n} f(\pi_i) + m \sum_{i,j=1}^{n} \pi_i f(p_{ij}).$$

In other words,

$$H\left(\bigvee_{i=0}^{m} T^{-i}(\alpha)\right) = -\sum_{i=1}^{n} \pi_i \log(\pi_i) - m \sum_{i,j=1}^{n} \pi_i p_{ij} \log(p_{ij}).$$

Using this identity, we have

$$h(T) = \lim_{m\to\infty} \frac{1}{m} H\left(\bigvee_{i=0}^{m} T^{-i}(\alpha)\right) = -\sum_{i,j=1}^{n} \pi_i p_{ij} \log(p_{ij}).$$

□

Problem 155. Suppose $(X_1, \mathcal{M}_1, \mu_1, T_1)$ and $(X_2, \mathcal{M}_2, \mu_2, T_2)$ are two dynamical systems. Show that

$$h(T_1 \times T_1) = h(T_1) + h(T_2).$$

Proof. We first show $h(T_1) \le h(T_1 \times T_2)$. Let α_1 be a partition for X_1, and let $\overline{\alpha_1} = \{A \times X_2 : A \in \alpha_1\}$. Then this is a partition for $X_1 \times X_2$. Moreover,

$$h(\alpha_1, T_1) = h(\overline{\alpha_1}, T_1 \times T_2).$$

This gives us the inequality, since these are defined as supremums. The same inequality applies for T_2. Therefore if at least one of T_1, T_2 are infinite the equality above holds. Assume now that both are finite. Let α_1 be a partition for X_1 and α_2 a partition for X_2. Define $\beta_1 = \overline{\alpha_1}$ and $\beta_2 \overline{\alpha_2}$ as before. Note these are independent of each other. Let $\beta = \beta_1 \vee \beta_2$. Then

$$H\left(\bigvee_{i=0}^{n} T^{-i}(\beta)\right) = H\left(\bigvee_{i=0}^{n} T^{-i}(\beta_1)\right) + H\left(\bigvee_{i=0}^{n} T^{-i}(\beta_2)\right).$$

Taking $1/n$ and the limit gives

$$h(\beta, T_1 \times T_2) = h(\beta_1, T_1) + h(\beta_2, T_2).$$

Now take increasing partitions α_1^n for X_1 and α_2^n for X_2 so that $\bigvee_{n=1}^{\infty} \alpha_1^n = \mathcal{M}_1$ and $\bigvee_{n=1}^{\infty} \alpha_2^n = \mathcal{M}_2$. Then $\bigvee_{n=1}^{\infty} \alpha_1^n \vee \alpha_2^n = \mathcal{M}_1 \times \mathcal{M}_2$, and invoking **Petersen Proposition 3.6** and the last remark we get the desired result. □

We now consider topological entropy. Throughout this next part, consider X a compact topological space.

An open cover α is a collection of open subsets $U \subseteq X$ so that

$$X \subseteq \bigcup_{U \in \alpha} U.$$

Problem 156 (Lebesgue Number Lemma)**.** Let X be a compact metric space. Let α be an open cover. Show that there exists an ϵ so that for all $x \in X$, we have $B(x, \epsilon) \subseteq U$ for some $U \in \alpha$. Such an ϵ is called the **Lebesgue number**.

Remark. The following proof is from Munkres.

Proof. Since α an open cover and X compact, we have that there exists $\{A_1, \ldots, A_n\} \subseteq \alpha$ so that

$$X \subseteq \bigcup_{i=1}^{n} A_i.$$

Notice redundant information doesn't help, so suppose all of the A_i are distinct. If $n = 1$ in this case, we can just take ϵ to be anything, so suppose $n > 1$ so that we have at least 2 distinct sets. Let $C_i = A_i^c$. Define a function

$$f : X \to \mathbb{R}, \qquad f(x) = \frac{1}{n} \sum_{i=1}^{n} d(x, C_i).$$

Recall

$$d(x, C_i) = \inf\{d(x, y) : y \in C_i\}.$$

We claim this is a continuous function. Since we are in a metric space, the function $d(x, \cdot) : X \to \mathbb{R}$ is continuous. The infimum of continuous functions is continuous, so $d(x, C_i)$ is continuous. The sum of continuous functions is continuous, so f is continuous. Since $\{A_1, \ldots, A_n\}$ is an open cover, we have that $x \in A_i$ for some i, so $f(x) > 0$ for all x. Since X a compact metric space, we have that f achieves it's minimum, call it $\epsilon > 0$. The goal now is to show that ϵ satisfies the desired criteria. Examine

$$B(x,\epsilon) = \{y \in X : d(x,y) < \epsilon\}.$$

Since ϵ is a minimum for f, we get that $f(x) \geq \epsilon$, meaning that $d(x, C_i) \geq \epsilon$ for some i. Therefore for each $y \in B(x,\epsilon)$, we have $y \in C_i^c = A_i$, hence $B(x,\epsilon) \subseteq A_i$. This finishes the proof. □

If α, β are two open covers of X, their **join**, denoted $\alpha \vee \beta$, is the collection of all sets of the form $A \cap B$ for all $A \in \alpha$, $B \in \beta$.

An open cover β is said to be an **refinement** of α, written $\alpha < \beta$, if every member of β is a subset of a member of α. In other words, for all $B \in \beta$, there is an $A \in \alpha$ so that $B \subseteq A$.

Recall that a cover β is a **subcover** of α if for all $B \in \beta$ we have that $B \in \alpha$.

Problem 157.

(1) Show that $\alpha < \alpha \vee \beta$ for any open covers α, β.
(2) Show that if β is a subcover of α then $\alpha < \beta$.

Proof.

(1) We have
$$\alpha \vee \beta = \{A \cap B : A \in \alpha, B \in \beta\}.$$
The goal is to show that every member $C \in \alpha \vee \beta$ is a subset of a member $A \in \alpha$. But this follows, since every member $C = A \cap B \in \alpha \vee \beta$ is naturally a subset of $A \in \alpha$.
(2) This is also easy. Let $B \in \beta$. Then $B \subseteq B \in \alpha$.

□

Problem 158.

(1) Suppose α is an open cover of X and $T : X \to X$ is continuous. Show that
$$T^{-1}(\alpha) = \{T^{-1}(A) : A \in \alpha\}$$
is also an open cover.
(2) Show that T^{-1} behaves well with our operations. That is, show
$$T^{-1}(\alpha \vee \beta) = T^{-1}(\alpha) \vee T^{-1}(\beta), \qquad \alpha < \beta \implies T^{-1}(\alpha) < T^{-1}(\beta).$$

Proof.

(1) Since T is continuous, $T^{-1}(A)$ is open for all $A \in \alpha$. Hence $T^{-1}(\alpha)$ is a collection of open sets. Next we note it is a cover, since
$$X = T^{-1}(X) \subseteq T^{-1}\left(\bigcup_{A\in\alpha} A\right) = \bigcup_{A\in\alpha} T^{-1}(A) = \bigcup_{A\in T^{-1}(\alpha)} A.$$
(2) Let $A \cap B \in \alpha \vee \beta$. Then $T^{-1}(A\cap B) = T^{-1}(A) \cap T^{-1}(B)$, so $T^{-1}(\alpha \vee \beta) \subseteq T^{-1}(\alpha) \vee T^{-1}(\beta)$. The same idea goes the other way. The same idea also applies to refinements.

□

If α is an open cover of X, let $N(\alpha)$ denote the minimum number of sets so that there exists $A_1, \ldots, A_{N(\alpha)} \in \alpha$ with $X \subseteq \bigcup_{i=1}^{N(\alpha)} A_i$. We define the **entropy** of the open cover α to be

$$H(\alpha) := \log(N(\alpha)).$$

Problem 159. Show the following properties for entropy.

(1) $H(\alpha) \geq 0$.
(2) $H(\alpha) = 0 \iff N(\alpha) = 1 \iff X \in \alpha$.
(3) $\alpha < \beta \iff H(\alpha) \leq H(\beta)$.

(4) $N(\alpha \vee \beta) \leq N(\alpha) \cdot N(\beta)$.
(5) $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$.
(6) If $T : X \to X$ is continuous, then $H(T^{-1}(\alpha)) \leq H(\alpha)$.
(7) If $T : X \to X$ is continuous and surjective, then $H(T^{-1}(\alpha)) = H(\alpha)$.

Proof.

(1) Consider $\theta \subseteq \mathcal{P}(\alpha)$, where for all $\gamma \in \theta$ we have
$$X \subseteq \bigcup_{A \in \gamma} A.$$
Define an equivalence relation $\sim$ on θ so that $\gamma_1 \sim \gamma_2$ iff $|\gamma_1| = |\gamma_2|$. This is easily checked to be an equivalence relation since this is just cardinality. Now quotient θ by this equivalence relation, and look at only the finite sets. This will be in bijection with $\mathbb{N}$, since it is impossible for an empty set to be an open cover for X (so there must be at least one set). Taking the image under the bijection with $\mathbb{N}$, we can find a minimum and label that $N(\alpha)$. This satisfies the condition $N(\alpha) \geq 1$, and so we have $H(\alpha) \geq 0$.

(2) Now $N(\alpha) = 1 \iff X \in \alpha$, since if $\gamma \subseteq \mathcal{P}(\alpha)$ satisfies $|\gamma| = 1$ and
$$A \subseteq X \subseteq \bigcup_{A \in \gamma} A = A,$$
then $A = X$, and if $X \in \alpha$ then we have that $N(\alpha) = 1$, the smallest possible value. It's clear $N(\alpha) = 1 \iff H(\alpha) = 0$.

(3) Let $n = N(\beta)$, take $\{B_1, \ldots, B_n\} \subseteq \beta$ so that $X \subseteq \bigcup_{i=1}^n B_i$. Since $\alpha < \beta$, we have that we can find $A_i \in \alpha$ so that $B_i \subseteq A_i$. Therefore $X \subseteq \bigcup_{i=1}^n A_i$, and hence $N(\alpha) \leq N(\beta)$ by definition of minimality. Taking logs, we get $H(\alpha) \leq H(\beta)$.

(4) This argument is the same as in the last one. Let $n = N(\alpha)$, $m = N(\beta)$, take $\{A_1, \ldots, A_n\} \subseteq \alpha$ so that $X \subseteq \bigcup_{i=1}^n A_i$, $\{B_1, \ldots, B_m\} \subseteq \beta$ so that $X \subseteq \bigcup_{i=1}^m$. Then taking
$$\gamma = \{A_i \cap B_j : 1 \leq i \leq n, 1 \leq j \leq m\},$$
we have $X \subseteq \bigcup_{C \in \gamma} C$, $|\gamma| = N(\alpha) \cdot N(\beta)$, and $\gamma \subseteq \alpha \vee \beta$, so by minimality we get $N(\alpha \vee \beta) \leq N(\alpha) \cdot N(\beta)$.

(5) Taking logs from the last part gives us the result.

(6) Let $n = N(\alpha)$, $\{A_1, \ldots, A_n\}$ as before. Then $\{T^{-1}(A_1), \ldots, T^{-1}(A_n)\}$ covers X, and so by definition of minimality we get $N(T^{-1}(\alpha)) \leq N(\alpha)$. Taking logs gives the result.

(7) Let $n = N(T^{-1}(\alpha))$, $\{B_1, \ldots, B_n\}$ such that it covers and $B_i = T^{-1}(A_i)$ for some i. We claim that $\{A_1, \ldots, A_n\}$ also covers X. This follows from the fact that T has a right inverse, so
$$X = T(X) \subseteq T\left(\bigcup_{i=1}^n B_i\right) = \bigcup_{i=1}^n T(B_i) = \bigcup_{i=1}^n A_i.$$
□

We define the **topological entropy of α with respect to T**, denoted $h(T, \alpha)$, to be
$$h(T, \alpha) = \lim_{n \to \infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\alpha)\right).$$
We need to check that this exists first.

Problem 160. Define
$$\alpha_0^n := \bigvee_{i=0}^{n-1} T^{-i}(\alpha), \qquad a_n := H(\alpha_0^n).$$

Show that the above exists by showing $a_{n+m} \leq a_n + a_m$ and then invoking Fekete's lemma.

Proof. We see

$$a_{n+m} = H\left(\bigvee_{i=0}^{n+m-1} T^{-i}(\alpha)\right) = H\left(\bigvee_{i=0}^{n-1} T^{-i}(\alpha) \vee \bigvee_{i=n}^{n+m-1} T^{-i}(\alpha)\right)$$
$$= H\left(\bigvee_{i=0}^{n-1} T^{-i}(\alpha) \vee T^{-n}\left(\bigvee_{i=0}^{m-1} T^{-i}(\alpha)\right)\right)$$
$$\leq H\left(\bigvee_{i=0}^{n-1} T^{-i}(\alpha)\right) + H\left(T^{-n}\left(\bigvee_{i=0}^{m-1} T^{-i}(\alpha)\right)\right)$$
$$\leq H\left(\bigvee_{i=0}^{n-1} T^{-i}(\alpha)\right) + H\left(\bigvee_{i=0}^{m-1} T^{-i}(\alpha)\right)$$
$$= a_n + a_m.$$

□

Problem 161. Prove the following properties.

(1) We have $h(T, \alpha) \geq 0$.
(2) If $\alpha < \beta$, then $h(T, \alpha) \leq h(T, \beta)$.
(3) We have $h(T, \alpha) \leq H(\alpha)$.
(4) We have $h(\mathrm{Id}, \alpha) = 0$ for all α open covers.

Proof.

(1) This follows since $H(\alpha) \geq 0$ for all open covers α.
(2) This follows since $\alpha < \beta$ implies $H(\alpha) \leq H(\beta)$, and $\alpha < \beta$ implies $T^{-1}(\alpha) < T^{-1}(\beta)$.
(3) This is the trickier calculation. Here use the fact that $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$. Then

$$H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) \leq \sum_{i=0}^{n-1} H(T^{-i}(\alpha)).$$

Now use the fact that $H(T^{-i}(\alpha)) \leq H(\alpha)$ to get

$$H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) \leq \sum_{i=0}^{n-1} H(T^{-i}(\alpha)) \leq nH(\alpha).$$

Hence

$$h(T, \alpha) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\alpha)\right) \leq \lim_{n\to\infty} \frac{1}{n}(nH(\alpha) = H(\alpha).$$

(4) This is a matter of showing that $\alpha < \alpha \vee \alpha < \alpha$. We know from prior that for any open cover β (and therefore for $\beta = \alpha$) we have $\alpha < \alpha \vee \beta$. Now to show that $\alpha \vee \alpha < \alpha$, we need to show that for all $A \in \alpha$ is a subset of a member of $\alpha \vee \alpha$. But this follows, since $A \in \alpha$ is a subset of $A \cap A = A \in \alpha \vee \alpha$. Therefore $H(\alpha \vee \alpha) = H(\alpha \vee \mathrm{Id}^{-1}(\alpha)) = H(\alpha)$. By induction, we have

$$H\left(\bigvee_{i=0}^{n-1} \mathrm{Id}^{-i}(\alpha)\right) = H(\alpha),$$

so

$$h(\mathrm{Id}, \alpha) = \lim_{n\to\infty} \frac{1}{n} H(\alpha) = 0.$$

□

If $T : X \to X$ is continuous, we define the **topological entropy** of T to be

$$h(T) = \sup_\alpha h(T, \alpha),$$

where α ranges over all open covers of X.

Problem 162. Prove the following properties.

(1) $h(T) \geq 0$.
(2) We have

$$h(T) = \sup_\beta h(T, \beta),$$

where β rangers over all finite open covers of X.
(3) $h(\text{Id}) = 0$.
(4) If Y is a closed subset of X and $TY = Y$, then $h(T|_Y) \leq h(T)$.

Proof.

(1) This follows since $h(T, \alpha) \geq 0$ for all α.
(2) Apriori we have

$$\sup_\beta h(T, \beta) \leq h(T).$$

We need to show the other direction. But this follows since for all open covers α there always exists an open subcover β, so that $\alpha < \beta$. Therefore $h(T, \alpha) \leq h(T, \beta)$, and this holds for all open covers α, so

$$h(T) \leq \sup_\beta h(T, \beta).$$

(3) This follows since $h(\text{Id}, \alpha) = 0$ for all α.
(4) Consider all open covers of Y in the subspace topology, call this space $\mathcal{Y}$. Consider all open covers of X with respect to its topology, call this space $\mathcal{X}$. We see there is a natural embedding $\mathcal{Y} \hookrightarrow \mathcal{X}$ via $\alpha \in \mathcal{Y}$ gets mapped to $\widehat{\alpha} = \alpha \cup \{Y^c\}$. We can then view $\mathcal{Y} \subseteq \mathcal{X}$, and therefore $h(T|_Y)$ is defined via a supremum over a subset, hence $h(T|_Y) \leq h(T)$.

□

Let $\{\alpha_n\}$ be a sequence of covers. We call this sequence **refining** if $\alpha_1 < \alpha_2 < \cdots$ and for each open cover β of X we have $\beta < \alpha_n$ for some n.

Problem 163. Show that if $\{\alpha_n\}$ is a refining sequence of covers then

$$h(T) = \lim_{n\to\infty} h(\alpha_n, T).$$

Proof. Let β be any cover. Since it is refining, we know that there exists an n so that $\beta < \alpha_n$. Hence

$$h(T, \beta) \leq \sup_n h(T, \alpha_n) = \lim_{n\to\infty} h(T, \alpha_n).$$

This holds for all possible covers β, so

$$\sup_\beta h(T, \beta) = h(T) \leq \lim_{n\to\infty} h(T, \alpha_n).$$

Since α_n is a collection of covers, we note that

$$\sup_n h(T, \alpha_n) = \lim_{n\to\infty} h(T, \alpha_n) \leq h(T).$$

Hence we have equality. □

Problem 164. Let σ be the shift transformation on $\{0,1\}^{\mathbb{Z}}$. Use the prior problem to prove the following:

(1) $h(\sigma) = 1$.
(2) If $X \subseteq \{0,1\}^{\mathbb{Z}}$ consists of a single periodic orbit, then $h(\sigma|_X) = 0$.
(3) If $X \subseteq \{0,1\}^{\mathbb{Z}}$ consists of all sequences containing only even length maximal strings of 0's and $1's$, then $h(\sigma|_X) = 1/2$.

Proof.

(1) From the prior problem, we know that
$$h(\sigma) = \lim_{n\to\infty} \frac{1}{n} H(\alpha_0^{n-1}),$$
where here $\alpha = \{C_0, C_1\}$ and
$$C_i = \{(x_n) \in \{0,1\}^{\mathbb{Z}} : x_0 = i\}.$$
So in other words, α_0^{n-1} is going to be the collection of all cylinders of length n of the form
$$C = \{(x_n) \in \{0,1\}^{\mathbb{Z}} : x_{-n+1} = i_{-n+1}, \dots, x_0 = i_0\}, \quad i_j \in \{0,1\}, -n+1 \le j \le 0.$$
Now for the set $\{0,1\}^{\mathbb{Z}}$, all of these are needed to cover the entire set, and there are 2^n of them. Hence
$$h(\sigma) = \lim_{n\to\infty} \frac{1}{n} \log_2(2^n) = 1.$$

(2) Consider $(x_n) \in \{0,1\}^{\mathbb{Z}}$ periodic, say of period m. Then there are exactly m-distinct words of length m in $\overline{X}$. So again using the prior problem, we have
$$h(\sigma|_X) = \lim_{n\to\infty} \frac{1}{n} H(\alpha_0^{n-1}),$$
where here α_0^{n-1} consists of the cylinders
$$C = \{(x_n) \in \{0,1\}^{\mathbb{Z}} : x_{-n+1} = i_{-n+1}, \dots, x_0 = i_0\} \cap X, \quad i_j \in \{0,1\}, -n+1 \le j \le 0.$$
Taking $n \ge m$, we see that only 2^m of these are needed to cover the entire set, so we have
$$h(\sigma|_X) = \lim_{n\to\infty} \frac{1}{n} \log_2(2^m) = \lim_{n\to\infty} \frac{m}{n} = 0.$$

(3) In essence, we have that X represents all possible orbits with an even number of 0s and 1s occurring. We consider then the total number of words of length n. Building such a word, we fix the first letter, and we know that the next letter must be the same one. So we have 2 options for the first letter, and then only one option for the next letter, 2 options for the following letter, and so on. This gives us a total of $2^{n/2}$ possible words if the length was even and $2^{(n+1)/2}$ possible words if the length was odd. Since this encompasses the total number of words, we take log and then the limit. Both of these limits agree, so we get $h(\sigma|_X) = 1/2$.

□

Problem 165 (Petersen 6.4.3)**.** Consider the Thue-Morse sequence. Let
$$A_0 = \{0\},$$
For a sequence $A = \{x_0, \dots, x_{n-1}\}$ write
$$\overline{A} = \{1 - x_0, \dots, 1 - x_{n-1}\},$$
and for two sequences $A = \{x_0, \dots, x_{n-1}\}$, $B = \{x_n, \dots, x_m\}$, we write
$$AB = \{x_0, \dots, x_m\}.$$

Then the Thue-Morse sequence is generated by the recursion

$$A_n = A_{n-1}\overline{A_{n-1}}.$$

Let A_∞ be the forward orbit. Then the full orbit is generated by setting

$$x = (x_n) = A = \overline{A_\infty}A_\infty.$$

Consider $X = \overline{\mathcal{O}_\sigma(x)}$. This is a closed σ-invariant set. Calculate $h(\sigma|_X)$.

Proof. Credit to Dr. Nimish Shah. As before, we wish to calculate the number of words of length $n = 2^k$ (we choose powers of 2 since that's how the sequence is generated). Notice that the set up for our shifts are

$$\cdots A\overline{A}\overline{A}A\overline{A}AA\overline{A}\cdots,$$

where here A represents blocks of length n. We claim that the only options for occurrences will be $A\overline{A}$, $\overline{A}A$, AA, or $\overline{A}\overline{A}$. This follow since we are enumerating all possible options without caring about order. Now, if we shift n times we see we are getting at most n new words in each combination, and this enumerates all possible words of length n. This gives us an upper bound of $4n$ total words of length n. Calculating topological entropy with this, we have

$$h(\sigma|_X) \leq \lim_{n\to\infty} \frac{1}{n} H(\alpha_0^n) = \lim_{k\to\infty} \frac{1}{2^k}\log_2(4\cdot 2^k) = 0.$$

Therefore the topological entropy is 0. □

We recall a few definitions.

Recall that a map $T : X \to X$ is **ergodic** if there exists a measure μ so that if B measurable satisfies $\mu(T^{-1}(B)\triangle B) = 0$, then $\mu(B) = 0$ or 1. It is called **uniquely ergodic** if there is a unique such measure. For X a compact metric space, $T : X \to X$ a homeomorphism, we have a **dynamical system** (X, T). The **orbit** for a point is the set

$$\mathcal{O}_T(x) = \{T^n(x) : n \in \mathbb{Z}\}.$$

We say that (X, T) is **minimal** if for all $x \in X$ we have $\mathcal{O}_T(x)$ is dense.

Problem 166. Suppose $Y \subseteq X$ is a T-invariant subset. Show that $\overline{Y}$ is also a T-invariant subset.

Proof. Let $y \in \overline{Y}$, then we have $(x_n) \subseteq Y$ so that $x_n \to y$. Since Y is T-invariant, we have that $(T(x_n)) \subseteq Y$, and by continuity $T(x_n) \to T(y)$, so $T(y) \in \overline{Y}$. Therefore $T(\overline{Y}) \subseteq \overline{Y}$, giving us that $\overline{Y}$ is T-invariant. □

Problem 167. Show that this is equivalent to X having no proper closed T-invariant subsets.

Proof. ($\implies$): Assume that (X, T) is minimal. Suppose $Y \subseteq X$ is a closed T-invariant subset. Then we have $y \in Y$, and we see $\mathcal{O}_T(y) \subseteq Y$. Taking the closure, we have $X = \overline{\mathcal{O}_T(y)} \subseteq Y \subseteq X$. Hence $X = Y$, and so we see that it is impossible for Y to be proper.
($\impliedby$): Assume that X has no proper closed T-invariant subsets. First we note that for all $x \in X$, $\overline{\mathcal{O}_T(x)}$ is a closed T-invariant subset. Since it is non-empty, it cannot be proper, so we must have $\overline{\mathcal{O}_T(x)} = X$. This shows that the orbit of every point is dense; i.e. the system is minimal. □

We say that a system (X, T) is **strictly ergodic** if it is minimal and uniquely ergodic.

We now discuss the Bowen definition for entropy. Let X be a compact metric space, $T : X \to X$ a homeomorphism. The goal is to count the number of different orbit-blocks of length n that can be observed, where we fail to distinguish points closer together than some positive error term, ϵ (in other words, our measurement system fails at this point).

We say that x and y are **(n, ϵ)-separated** if their initial blocks of length n can be distinguished by our measurement system. In other words, if

$$d(T^k x, T^k y) > \epsilon \text{ for some } 0, 1, \ldots, n-1.$$

A set $E \subseteq X$ is said to be **(n, ϵ)-separated** if x and y are (n, ϵ)-separated for all $x, y \in E$, $x \neq y$. The maximum number of distinguishable orbit n-blocks will be

$$s(n, \epsilon) = \max\{|E| : E \subseteq X \text{ is } (n, \epsilon)\text{-separated}\}.$$

Define

$$h(T, \epsilon) = \limsup_{n \to \infty} \frac{1}{n} \log_2(s(n, \epsilon)).$$

Set

$$h_{\text{Top}}(T) = \lim_{\epsilon \to 0^+} h(T, \epsilon).$$

Problem 168. Determine why should a limit should exist.

Proof. We need to show that $h(T, \epsilon)$ decreases as ϵ decreases. This will follow if we can show $s(n, \epsilon)$ decreases with ϵ. But this follows since we're taking a max over a smaller and smaller set. $\square$

We now wish to generalize this for arbitrary metric spaces. Assume that X is not compact. For each compact $K \subseteq X$, let

$$s_K(n, \epsilon) = \max\{|E| : E \subseteq K \text{ is } (n, \epsilon)\text{-separated}\}.$$

In other words, we view $K \subseteq X$ as a compact metric space. Define all of the other terms analogously;

$$h_K(T, \epsilon) = \limsup_{n \to \infty} \frac{1}{n} \log_2(s_K(n, \epsilon)),$$

$$h_K(T) = \lim_{\epsilon \to 0^+} h_K(T, \epsilon).$$

We can then define

$$h_{\text{Top}}(T) = \sup_K h_K(T).$$

We can go the opposite direction as well. For $\epsilon > 0$, $n = 1, 2, \ldots$, call a set $F \subseteq X$ **(n, ϵ)-spanning** if for each $x \in X$ there is a $y \in F$ so that

$$d(T^k x, T^k y) \leq \epsilon \text{ for all } k = 0, 1, \ldots, n-1.$$

Problem 169. Show that a maximal (n, ϵ)-separated set is (n, ϵ)-spanning. If we set

$$r(n, \epsilon) = \min\{|F| : F \subseteq X \text{ is } (n, \epsilon)\text{-spanning}\},$$

deduce that $r(n, \epsilon) \leq s(n, \epsilon)$.

Proof. The idea here is to note that if we add on a point, it will no longer be $(n, \epsilon$-separated by maximality. Hence by the above definition and minimality, we get $r(n, \epsilon) \leq s(n, \epsilon)$. $\square$

One can also show $s(n, \epsilon) \leq r(n, \epsilon/2)$ (see K&H), and this will give us the following proposition.

Proposition. We have

$$h_{\text{Top}}(T) = \lim_{\epsilon \to 0^+} \limsup_{n \to \infty} \frac{1}{n} \log_2(r(n, \epsilon)).$$

Recall Karamata's inequality.

Theorem (Karamata's Inequality)**.** Let $I = [0,1]$, and suppose we have to sets of numbers $\{x_i\}_{i=1}^n, \{y_i\}_{i=1}^n \subseteq I$ such that

$$x_1 \geq \cdots \geq x_n, \qquad y_1 \geq \cdots \geq y_n,$$

and for each $1 \leq k \leq n-1$ we have

$$x_1 + \cdots + x_k \leq y_1 + \cdots + y_k,$$

with equality in the case that $k = n$. If f is a real-valued convex function on I, then we have

$$f(x_1) + \cdots + f(x_n) \leq f(y_1) + \cdots + f(y_n).$$

Proof. If $x_i = y_i$ for all i, then the inequality holds clearly. Suppose then $x_i \neq y_i$ for some $1 \leq i \leq n$. If $x_i = y_i$ for some $1 \leq i \leq n-1$, then removing x_i and y_i from their sequences does not affect the assumptions nor conclusion, so iterating we may assume $x_i \neq y_i$ for all $1 \leq i \leq n-1$. For $1 \leq i \leq n$ let

$$A_0 = 0, \quad A_i = x_1 + \cdots + x_i,$$
$$B_0 = 0, \quad B_i = y_1 + \cdots + y_i.$$

By assumption, $A_i \leq B_i$ for $1 \leq i \leq n-1$ and $A_n = B_n$. Observe as well that

$$A_i - A_{i-1} = (x_1 + \cdots + x_i) - (x_1 + \cdots + x_{i-1}) = x_i,$$

and similarly $B_i - B_{i-1} = y_i$. Now for $1 \leq i \leq n-1$ let

$$c_i = \frac{f(y_i) - f(x_i)}{y_i - x_i}.$$

(Note here that we are using the fact that $x_i \neq y_i$ for all i). Observe that $c_{i+1} \leq c_i$ since it is a convex function. Hence we have

$$\begin{aligned}\sum_{i=1}^n (f(y_i) - f(x_i)) = \sum_{i=1}^n c_i(y_i - x_i) = \sum_{i=1}^n c_i(B_i - B_{i-1} - (A_i - A_{i-1})) \\ = \sum_{i=1}^n c_i(B_i - A_i) - \sum_{i=1}^n c_i(B_{i-1} - A_{i-1}) \\ = c_n(B_n - A_n) + \sum_{i=1}^{n-1}(c_i - c_{i+1})(B_i - A_i) - c_1(B_0 - A_0) \geq 0.\end{aligned}$$

This gives the result. □

Remark. We get the reverse equality for f concave. Simply use the fact that $-f$ is convex.

Problem 170. Show that if $p_i \leq q_j$ for all i and j, then the entropy of $\mathcal{B}(p_1, \ldots, p_n)$ is no less than that of $\mathcal{B}(q_1, \ldots, q_m)$.

Proof. Assume $n = m$. Then we are in a position to apply Karamata's inequality. After rearranging the p_i and q_j, we may assume that $p_1 \leq p_2 \leq \cdots \leq p_n$, $q_1 \leq q_2 \leq \cdots \leq q_n$, and we have the property that for each $1 \leq k \leq n$

$$p_1 + \cdots + p_k \leq q_1 + \cdots + q_k.$$

This follows since $p_1 \leq q_1, p_2 \leq q_2$, etc. Since it is a probability vector, we also have

$$p_1 + \cdots + p_n = q_1 + \cdots + q_n = 1.$$

Hence by Karamata we get that for $f(t) = -t\log(t)$ we have

$$\sum_{i=1}^n f(p_i) \geq \sum_{j=1}^n f(q_j).$$

Now suppose $n \neq m$. Consider $m < n$. Appending 0s at the end of the vector q to get $(q_1, \dots, q_m, 0, \dots, 0)$ so that there are n terms, we have that the assumptions for Karamata still hold and so the inequality still holds. Assuming $p_i, q_j \neq 0$ for all i and j, we claim that it is impossible for $m > n$. If this were the case, then we have

$$1 = p_1 + \cdots + p_n \leq q_1 + \cdots + q_n < q_1 + \cdots + q_m = 1,$$

a contradiction. □

We do some Einsiedler and Ward exercises here (separated since I don't want to try to organize these in with the other solutions).

Problem 171 (2.1.7). Let $(X, \mathcal{M}, \mu, T)$ be any measure-preserving system. A sub-σ-algebra $\mathcal{A} \subseteq \mathcal{M}$ with $T^{-1}(\mathcal{A}) = \mathcal{A}$ modulo μ is called a *T-invariant sub-σ-algebra*. Show that the system $(\widehat{X}, \widehat{B}, \widehat{\mu}, \widehat{T})$ defined by

- $\widehat{X} = \{x \in X^{\mathbb{Z}} : x_{k+1} = T(x_k) \text{ for all } k \in \mathbb{Z}\}$;
- $(\widehat{T}(x))_k = x_{k+1}$ for all $k \in \mathbb{Z}$ and $x \in \widehat{X}$;
- $\widehat{\mu}(\{x \in X : x_0 \in A\}) = \mu(A)$ for any $A \in \mathcal{M}$ and $\widehat{\mu}$ is invariant under $\widehat{T}$;
- $\widehat{B}$ is the smallest $\widehat{T}$-invariant σ-algebra for which the map $\pi : x \mapsto x_0$ from $\widehat{X}$ to X is measurable;

is an invertible measure-preserving system, and that the map $\pi : x \mapsto x_0$ is a factor map. The system $\widehat{X}$ is called the **invertible extension** of X.

Proof. Let's first check measure-preserving. □