§10. Measure + Topology.

This section will combine ideas from topological dynamics and measure dynamics. Recall that any topological space X determines a σ-algebra $\mathcal{B}_X$ of *Borel subsets* of X , and any continuous map $f : X \to Y$ is automatically measurable as a function from $(X, \mathcal{B}_X)$ into $(Y, \mathcal{B}_Y)$. (Compare §8D.) By a *measure* on X we will always mean a measure on this σ-algebra $\mathcal{B}_X$. The following will help us to relate measure and topology.

Definition. The *support* $\mathrm{supp}(\mu) \subset X$ of a measure μ on X is defined as the set of all $x \in X$ such that every neighborhood of x has strictly positive measure. Alternatively, the complement $X \smallsetminus \mathrm{supp}(\mu)$ is the union of all open sets of measure zero. In making use of this definition, we will always assume that there is a countable basis for the topology of X , so that this union of sets of measure zero will itself have measure zero, or in other words so that $\mathrm{supp}(\mu)$ is a set of full measure.

> **Lemma 10.1.** *If $f : X \to X$ is a continuous map, and $f_*(\mu) = \mu$ is an invariant measure, then f carries the closed set $\mathrm{supp}(\mu)$ into itself, so that $f : \mathrm{supp}(\mu) \to \mathrm{supp}(\mu)$ can be considered as a dynamical system in its own right. Furthermore, if $\mu(X) < \infty$, then $\mathrm{supp}(\mu)$ is necessarily contained in the non-wandering set $\Omega(f)$. (See §4B.)*

As an example, if μ has the property that *every* non-vacuous open subset has strictly positive measure, then it follows that $\mathrm{supp}(\mu)$ is the whole space X . If $\mu(X) < \infty$, it then follows that every point of X is non-wandering. (Compare the Poincaré Recurrence Theorem in §8A.)

Proof of 10.1. Suppose $x \in \mathrm{supp}(\mu)$, so that every neighborhood of x has positive measure. Then every neighborhood U of $f(x)$ has measure $\mu(U) = \mu(f^{-1}(U)) > 0$, since $f^{-1}(U)$ is a neighborhood of x . Therefore, $f(x) \in \mathrm{supp}(\mu)$. Similarly, if $x \in \mathrm{supp}(\mu)$ and $\mu(X) < \infty$, then no neighborhood U of x can have the property that the iterated pre-images $f^{-n}(U)$ are pairwise disjoint. For these sets all have measure equal to $\mu(U) > 0$. Therefore, x must be a non-wandering point. $\square$

---

§10A. Ergodic Theory on Compact Metric Spaces. This section will characterize ergodic mappings, and also ergodic flows, on a compact metric space. Recall from §9 that the *time average* $A_\varphi(x_0)$ of a real valued function φ over an orbit $x_0 \mapsto x_1 \mapsto \cdots$ is the limit as $n \to \infty$ of $\frac{1}{n}(\varphi(x_0) + \varphi(x_1) + \cdots + \varphi(x_{n-1}))$, when this limit exists.

> **Theorem 10.2.** *Let X be compact metric and let μ be a probability measure on $(X, \mathcal{B}_X)$ Then a measurable map $f : X \to X$ is both ergodic and measure preserving if and only if, for every* continuous *function $\varphi : X \to \mathbb{R}$, the time average $A_\varphi(x)$ exists for $[\mu]$-almost every $x \in X$, and is equal to the space average $\int \varphi \, d\mu$.*

In particular, in order to check that f is ergodic, we only need to check this condition for *continuous* real valued functions, rather than for the much larger class of integrable (or bounded measurable functions), as in 9.3. The proof will be based on the following well known facts.

Lemma 10.3. *If X is compact metric and μ is a finite measure on $(X, \mathcal{B}_X)$, then for any $S \in \mathcal{B}_X$ the measure $\mu(S)$ can be described as the supremum of $\mu(K)$ as K varies over compact subsets of S, or as the infimum of $\mu(U)$ as U varies over open neighborhoods of S.*

Proof. First suppose that S itself is compact. Then the first statement is trivially true, and the second is true since S is the countable intersection of its $1/n$-neighborhoods. But it is easy to check that the statement is preserved when we pass to countable unions, or to complements. Hence it is true for all Borel sets. $\square$

Lemma 10.4. *If $\mu \neq \nu$ are two distinct measures on the compact metric space X, then there exists a continuous real valued function φ with $\int \varphi \, d\mu \neq \int \varphi \, d\nu$.*

Proof. Choose a set $S \in \mathcal{B}_X$ with say $\mu(S) < \nu(S)$, and choose a compact subset K and an open neighborhood U so that $\mu(U) < \nu(K)$. Let $L = X \smallsetminus U$, so that $K \cap L = \emptyset$. Then the function

$$\varphi(x) = \mathrm{dist}(x, L)/(\mathrm{dist}(x, L) + \mathrm{dist}(x, K)) \qquad (10:1)$$

satisfies $0 \leq \varphi(x) \leq 1$, with $\varphi(K) = 1$ and $\varphi(L) = 0$. It follows easily that

$$\int \varphi \, d\mu \leq \mu(U) < \nu(K) \leq \int \varphi \, d\nu \ ,$$

as required. $\square$

Proof of 10.2. If f is ergodic with $f_*(\mu) = \mu$, then for any continuous $\varphi : X \to \mathbb{R}$ the time average $A_\varphi(x)$ takes the constant value $C = \int \varphi$ for almost all x by 9.3. Conversely, if time averages equal space averages, then given f and the measure class $[\mu]$, we can compute $\int \varphi \, d\mu$ for any continuous φ simply as the common value $A_\varphi(x)$ for $[\mu]$-almost all x. Since $A_\varphi = A_{\varphi \circ f}$, this implies that $\int \varphi \, d\mu = \int (\varphi \circ f) \, d\mu = \int \varphi \, d(f_* \mu)$. By 10.4, this implies that $\mu = f_* \mu$, as required.

We must also show that f is ergodic. Otherwise, we could find a fully invariant set $S = f^{-1}(S)$ with $0 < \mu(S) < 1$. Let T be the complementary set $X \smallsetminus S$, and let $\epsilon = \mu(S)\mu(T)/2 > 0$. By 10.3, we could choose an open set $U \supset S$ and a compact set $K \subset S$ so that

$$\mu(U \smallsetminus K) \ < \ \epsilon \ . \qquad (10:2)$$

As in (10 : 1), we could then choose a continuous function $\varphi : X \to [0, 1]$ which takes the value 1 on K and the value 0 outside U. But we have assumed that the time average $A_\varphi(x)$ takes the constant value $C = \int \varphi$ for $[\mu]$-almost all x. This leads to a contradiction as follows. Since $\varphi(x) = 1$ for $x \in K$, this integral C must satisfy $C \geq \mu(K) \geq \mu(S) - \epsilon$. By the Birkhoff Theorem, applied to the restriction of φ to the invariant set $T = X \smallsetminus S$, we have $\int_T \varphi = \int_T A_\varphi$. But $\int_T \varphi \leq \mu(U \cap T) < \epsilon$ since φ is identically zero outside U, while $\int_T A_\varphi = C \, \mu(T) > \mu(S)\mu(T) - \epsilon = \epsilon$. This contradiction completes the proof of 10.2. $\square$

Now consider a flow $\{f_t : X \to X\}$ on the compact metric space X. We assume that $f_t(x)$ is defined for all $(t, x) \in \mathbb{R} \times X$, and that it is continuous in both variables. For any bounded real valued function φ on X we can try to form the forward and backward time averages,

$$\varphi^+(x) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \varphi(f_t(x)) \, dt \ , \qquad \varphi^-(x) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \varphi(f_{-t}(x)) \, dt \ .$$

Theorem 10.5. *If $\varphi : X \to \mathbb{R}$ is bounded and measurable, and if $\mu = f_{t*}(\mu)$ is an invariant probability measure on X, then these forward and backward time averages $\varphi^+(x)$ and $\varphi^-(x)$ are defined and equal to each other for $[\mu]$-almost all x, and furthermore,*

$$\int_X \varphi^{\pm}\, d\mu \;=\; \int_X \varphi\, d\mu \;.$$

Proof. The function $(t, x) \mapsto \varphi(f_t(x))$ is measurable as a function of two variables, hence by Fubini's Theorem it is measurable as a function of t for almost all x. (See for example [Rudin].) Furthermore, the integral $\psi(x) = \int_0^1 \varphi(f_t(x))\, dt$ is bounded and measurable as a function of x. Now it is easy to check that the forward or backward time average $\varphi^{\pm}(x)$ for the flow $\{f_t\}$ is the same as the forward or backward time average of the function ψ with respect to the iterated map f_1. Hence 10.5 follows immediately from 9.1 and 9.7. $\square$

By definition, a set $S \in \mathcal{B}_X$ is *invariant* under the flow $\{f_t\}$ if $f_t(S) = S$ for all t. The flow $\{f_t\}$ is *ergodic* with respect to the measure class $[\mu]$ if every invariant Borel set must have either measure zero or full measure. It follows easily from 10.2 and 10.5 that $\{f_t\}$ is measure preserving and ergodic for the probability measure μ if and only if, for every continuous $\varphi : X \to \mathbb{R}$, the time average $\varphi^{\pm}(x)$ is constant almost everywhere, and hence equal to the space average $\int \varphi\, d\mu$ almost everywhere.

We can sharpen these statements slightly, using the following. Let $\mathrm{C}(X, \mathbb{R})$ be the vector space consisting of all continuous real valued functions $\varphi : X \to \mathbb{R}$. Note that $\mathrm{C}(X, \mathbb{R})$ is a real *Banach space*, with respect to the *max norm*

$$\|\varphi\| \;=\; \max_{x \in X} |\varphi(x)| \;\geq\; 0 \;.$$

That is, $\mathrm{C}(X, \mathbb{R})$ is a real vector space, and this norm satisfies

$$\begin{aligned} &\|\varphi + \psi\| \leq \|\varphi\| + \|\psi\| \;, \\ &\|\lambda\, \varphi\| = |\lambda|\; \|\varphi\| \quad \text{for} \quad \lambda \in \mathbb{R} \;, \quad \text{and} \\ &\|\varphi\| = 0 \iff \varphi = 0 \;, \end{aligned}$$

and furthermore the associated metric $\mathrm{dist}(\varphi, \psi) = \|\varphi - \psi\|$ is complete. The topology for $\mathrm{C}(X, \mathbb{R})$ which is determined by this max norm is called the *topology of uniform convergence* on the space of real valued functions.

Lemma 10.6. *If X is compact metric, then this Banach space $\mathrm{C}(X, \mathbb{R})$ admits a countable dense subset.*

Proof. For each integer $n > 0$, choose a covering of X by finitely many open sets U_{nj} of diameter less than $1/n$. Let $\rho_{nj}(x) = \mathrm{dist}(x\,,\, X \smallsetminus U_{nj})$, and define functions $\psi_{nj} : X \to [0, 1]$ by

$$\psi_{nj}(x) \;=\; \rho_{nj}(x) \,/\, \sum_i \rho_{ni}(x) \;.$$

Then for each fixed n the ψ_{nj} form a partition of unity. That is, $0 \leq \psi_{nj}(x) \leq 1$, with $\sum_j \psi_{nj}(x) = 1$, and $\psi_{nj}(x) = 0$ for $x \notin U_{nj}$. Now for each n let $\Psi_n \subset \mathrm{C}(X, \mathbb{R})$ be the set of all finite linear combinations $\sum_j r_j\, \psi_{nj}$ with rational coefficients. Then $\Psi_1 \cup \Psi_2 \cup \cdots$ will be the required countable dense subset. In fact, for any

$\varphi \in \mathrm{C}(X, \mathbb{R})$ and any $\epsilon > 0$, we can choose n so that $|\varphi(x) - \varphi(y)| < \epsilon$ whenever $\mathrm{dist}(x, y) < 1/n$. If $\hat{x}_{nj} \in U_{nj}$ are representative points, then it follows easily that

$$\| \varphi - \sum_j \varphi(\hat{x}_{nj})\, \psi_{nj} \| \;<\; \epsilon\ .$$

Now approximate the real coefficients $\varphi(\hat{x}_{nj})$ by rational coefficients r_j, and the conclusion follows. □

Recall from §3B that a sequence of points $x_0, x_1, x_2, \ldots$ in X is said to be *evenly distributed* with respect to the probability measure μ, if the following condition is satisfied: *For every* $\varphi \in \mathrm{C}(X, \mathbb{R})$*, the limit*

$$\lim_{n\to\infty} \frac{1}{n}\left(\varphi(x_0) + \varphi(x_1) + \cdots + \varphi(x_{n-1})\right)$$

must exist and be equal to the space average $\int_X \varphi\, d\mu$. Similarly, we say that a continuous curve $t \mapsto x(t) \in X$ for $t \geq 0$ is *evenly distributed* with repect to μ if

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \varphi(x(t))\, dt$$

exists and is equal to $\int_X \varphi\, d\mu$ for every $\varphi \in \mathrm{C}(X, \mathbb{R})$.

Theorem 10.7. Topological Ergodic Theorem. *A map $f : X \to X$ is ergodic and measure preserving for the probability measure μ if and only if the forward orbit of x is evenly distributed for $[\mu]$-almost every x. Similarly, the flow $\{f_t : X \to X\}$ is ergodic and measure preserving if and only if the forward trajectory $t \mapsto f_t(x)$, $t \geq 0$, is evenly distributed for $[\mu]$-almost every x.*

Proof. Suppose that f or $\{f_t\}$ is ergodic and measure preserving. According to 10.2 or 10.5, we know that for each $\varphi \in \mathrm{C}(X, \mathbb{R})$ there is a set of measure zero, say N_φ, so that the required equality is satisfied for all $x \notin N_\varphi$. But now we require something sharper: We must find a single set of measure zero which works simultaneously for every $\varphi \in \mathrm{C}(X, \mathbb{R})$. To do this, we choose a countable dense $\{\varphi_i\}$ by 10.6, and then let N be the union of the N_{φ_i}. Evidently this union is itself a set of measure zero. Now for any $\varphi \in \mathrm{C}(X, \mathbb{R})$ and any $\epsilon > 0$ we can choose some φ_i with $\|\varphi - \varphi_i\| < \epsilon$. If $x \notin N$, then the upper and lower time averages for φ_i along the forward orbit of x are equal, hence the upper and lower time averages for φ differ by at most 2ϵ. Since ϵ can be arbitrarily small, this proves that the time average for φ is well defined. Further details of the proof are straightforward. □

---

§10B. Existence of Invariant and Ergodic Measures. We will need to make use of the following fundamental result. Let X be a compact metric space, and again let $\mathrm{C}(X, \mathbb{R})$ be the Banach space of continuous real valued functions. Evidiently any finite measure μ on the Borel sets of X gives rise to a linear mapping $L = L_\mu$ from $\mathrm{C}(X, \mathbb{R})$ to $\mathbb{R}$, where

$$L_\mu(\varphi) \;=\; \int_X \varphi\, d\mu\ .$$

If $\varphi \geq 0$ (that is, if $\varphi(x) \geq 0$ for all x), then evidently $L_\mu(\varphi) \geq 0$.

Theorem 10.8. The Riesz Representation Theorem. *Conversely, given any linear map* $L : \mathrm{C}(X,\mathbb{R}) \to \mathbb{R}$ *such that* $\varphi \geq 0$ *implies* $L(\varphi) \geq 0$ *, there is one and only one finite measure* μ *on* $(X, \mathcal{B}_X)$ *such that* $L(\varphi)$ *is equal to* $\int \varphi\, d\mu$ *for every* $\varphi \in \mathrm{C}(X,\mathbb{R})$ *.*

The uniqueness of μ is just Lemma 10.4 above. A proof of existence may be found in [Rudin] or [Parthasarathy]. □

It will often be convenient to identify each $\mu \in \mathcal{M}$ with the corresponding linear functional $L : \mathrm{C}(X,\mathbb{R}) \to \mathbb{R}$. Note that each such linear functional $L(\varphi) = \int \varphi\, d\mu$ is continuous, or equivalently *bounded*, in the sense that

$$|L(\varphi)| \ \leq \ c\,\|\varphi\| \ ,$$

where $c = \mu(X)$ and $\|\varphi\| = \max|\varphi(x)|$. In particular, if we restrict attention to probability measures, with $\mu(X) = 1$ (or equivalently $L(\mathbf{1}_X) = 1$), then $|L(\varphi)| \leq \|\varphi\|$.

Given any real Banach space B , we can form the *dual vector space* B^* , consisting of all continuous linear maps $L : \mathrm{B} \to \mathbb{R}$. In particular, taking $\mathrm{B} = \mathrm{C}(X,\mathbb{R})$, we can form the dual space $\mathrm{B}^* = \mathrm{C}(X,\mathbb{R})^*$. As in 8.7, let $\mathcal{M} = \mathcal{M}(X, \mathcal{B}_X)$ be the convex set consisting of all probability measures on $(X, \mathcal{B}_X)$. *The Riesz Representation Theorem asserts that the space* $\mathcal{M}(X, \mathcal{B}_X)$ *of all Borel probability measures on* X *can be naturally embedded in this dual space* $\mathrm{C}(X,\mathbb{R})^*$ *.*

Thus, in order to put a topology on $\mathcal{M}(X, \mathcal{B}_X)$, it suffices to put a topology on such dual spaces B^* . In fact there are two standard methods of putting a topology on such a dual space B^* . One is the *norm topology*. For any Banach space B , it is not difficult to check that B^* is itself a Banach space, where the norm of an element $L \in \mathrm{B}^*$ is defined by the formula

$$\|L\| \ = \ \sup\{\, \|L(\varphi)\| \ ; \ \|\varphi\| = 1\} \ .$$

Equivalently, $\|L\|$ is the smallest constant k such that $\|L(\varphi)\| \leq k\,\|\varphi\|$ for every $\varphi \in \mathrm{B}$.

However, this topology is much too large[1] for our purposes. That is, it has too many open sets, so that it is too difficult for a sequence $\{L_n\}$ to converge to a limit in B^* . Instead, we will make use of the *weak* topology*, which can be characterized intuitively by the following property:

> *A sequence of elements* $L_i \in \mathrm{B}^*$ *converges to the limit* L *in this topology if and only if, for every* $b \in \mathrm{B}$ *, the sequence* $L_i(b)$ *of real numbers converges to* $L(b)$ *. In particular, applying this to the subset* $\mathcal{M} = \mathcal{M}(X, \mathcal{B}_X) \subset \mathrm{C}(X,\mathbb{R})^*$ *, a sequence of measures* $\mu_1, \mu_2, \ldots$ *in* $\mathcal{M}$ *converges to the limit* $\mu \in \mathcal{M}$ *if and only if, for every continuous function* $\varphi \in \mathrm{C}(X,\mathbb{R})$ *the sequence of real numbers* $\int \varphi\, d\mu_i$ *converges to the limit* $\int \varphi\, d\mu$ *.*

A precise definition can be given as follows.

Definition. The *weak* topology* in B^* is defined as the smallest topology (the topology with the fewest open sets) having the following property: *For every fixed element* $\psi \in \mathrm{B}$,

[1] A larger topology (one with more open sets) can also be called a *finer* topology. Analysts call it a *stronger* topology since it has fewer convergent sequences. (Unfortunately, many topologists have confused the issue by calling it a *weaker* topology.)

the function

$$L \ \mapsto \ L(\psi)$$

from B^* *to* $\mathbb{R}$ *is continuous; or equivalently, for every* ψ *and every open interval* $U \subset \mathbb{R}$ *, the set*

$$N(\psi, U) \ = \ \{\, L \in \mathrm{B}^* \ : \ L(\psi) \in U\}$$

is an open subset of B^* . More explicitly, it follows that the finite intersections of the form

$$N(\psi_1\,,\,U_1) \cap \cdots \cap N(\psi_k\,,\,U_k)\ ,$$

where the U_i are open intervals in $\mathbb{R}$ and $\psi_i \in \mathrm{B}$, form a basis for the required weak* topology. We will use the notation $\mathrm{B}^*_{\mathrm{weak}}$ to emphasize that we are using this topology. One fundamental property is the following:

> *A sequence of linear maps* $L_j \in \mathrm{B}^*_{\mathrm{weak}}$ *converges to a limit* $L \in \mathrm{B}^*_{\mathrm{weak}}$ *if and only if, for each fixed* $\psi \in \mathrm{B}$ *, the sequence of real numbers* $L_j(\psi)$ *converges to* $L(\psi)$ *.*

The proof is easily supplied.

Example. Let X be the unit interval $[0,1]$. Let $\mathcal{I} \in \mathrm{C}(X,\mathbb{R})^*$ be the linear map $\varphi \mapsto \int_0^1 \varphi(x)\,dx$ (using either the Riemann integral or the Lebesgue integral), and let

$$L_n(\varphi) \ = \ \frac{1}{n}\sum_{i=1}^{n} \varphi(i/n)$$

be the linear map corresponding to the average $(\delta_{1/n} + \cdots + \delta_{n/n})/n$ of Dirac measures. For any $\varphi \in \mathrm{C}(X,\mathbb{R})$, the limit $\lim_n L_n(\varphi)$ is, almost by definition, the Riemann integral $\mathcal{I}(\varphi)$. Hence the sequence $\{L_n\}$ converges to $\mathcal{I}$ in the weak* topology. On the other hand, it is not difficult to check that $\|L_n - \mathcal{I}\| = \|L_n - L_{n+1}\| = 2$, so this sequence does not converge in the norm topology.

> **Theorem 10.9.** *For any compact metric* X *, the space* $\mathcal{M} = \mathcal{M}(X, \mathcal{B}_X)$ *of Borel probability measures on* X *is a compact convex metrizable subset of the topological vector space* $\mathrm{C}(X,\mathbb{R})^*_{\mathrm{weak}}$ *. The space* X *itself is canonically embedded in* $\mathcal{M}$ *by the correspondence* $x \mapsto \delta_x$ *.*

Proof. More generally, let B be any real Banach space which has a countable dense subset $\{\varphi_i\}$. We will first prove that the dual space B^* , with the weak* topology, is metrizable. For example, one suitable metric is given by the formula

$$\mathrm{dist}(L, L') \ = \ \sum_{i=1}^{\infty} \min\left(\,|L(\varphi_i) - L'(\varphi_i)|\,,\ 1/2^i\right)\ , \qquad (10\!:3)$$

taking the smaller of the two numbers on the right. It is not difficult to check that the topology associated with this metric coincides precisely with the weak* topology, as defined above.

Now let $U^* \subset B^*$ be the dual closed unit ball, consisting of all linear maps $L : B \to \mathbb{R}$ which satisfy $|L(\varphi)| \le \|\varphi\|$ for all $\varphi \in B$. We will prove that this dual unit ball, with the weak* topology, is compact. Since B^*_{weak} is a metric space, it suffices to show that every sequence of points $L_i \in U^*_{\mathrm{weak}}$ has a convergent subsequence. To do this, let us start with the countable dense subset $\{\varphi_i\} \subset B$, and set $\psi_i = \varphi_i/\|\varphi_i\|$, so that $\|\psi_i\| = 1$. First fix the test function ψ_1 and note that $L(\psi_1) \in [-1,1]$ for every $L \in U^*$. Hence we can choose an infinite set N_1 of natural numbers so that the infinite sequence of points

$$\left\{ L_i(\psi_1) \right\}_{i\in N_1} ,$$

all belonging to the compact interval $[-1,1]$, converges to a limit within this interval. Then choose an infinite set $N_2 \subset N_1$ so that $\{ L_i(\psi_2) \}_{i\in N_2}$ converges to a limit, and continue inductively. Finally, using a diagonal construction, we take the first element of N_1 , the second element of N_2 , and so on, yielding a new set N' of natural numbers with the property that for every fixed j the points $L_i(\psi_j)$ with $i \in N'$ converge to a limit as $i \to \infty$ within N' . Call this limit $L(\psi_j)$. Then it is easy to check that L extends uniquely to a linear map $L : B \to \mathbb{R}$ which belongs to U^* and is equal to the weak* limit of L_i as i tends to infinity through N' . Thus $U^*_{\rm weak}$ is compact.

Now, taking $B = \mathrm{C}(X,\mathbb{R})$, it is easy to check that $\mathcal{M}$ is a closed subset of $U^*_{\rm weak}$, so $\mathcal{M}$ is also compact. Finally, since X is compact, and since the embedding $x \mapsto \delta_x \in \mathcal{M}$ is clearly continuous and one-to-one, it must be a homeomorphism onto its image. $\square$

Next consider a continuous map $f : X \to X$. Let $f_* : \mathcal{M} \to \mathcal{M}$ be the induced linear map $(f_*\mu)(S) = \mu(f^{-1}(S))$ on the convex set $\mathcal{M} = \mathcal{M}(X,\mathcal{B}_X)$ of Borel probability measures. By definition, the fixed points $f_*(\mu) = \mu$ are the *invariant* probability measures for the dynamical system (X,f) .

Theorem 10.10 (Krylov and Bogoliubov). *Every continuous map $f : X \to X$ from a compact metric space to itself possesses at least one invariant probability measure. In fact, the space $\mathcal{M}_f$ consisting of all invariant probability measures for (X,f) forms a compact convex non-vacuous subset of $\mathcal{M} = \mathcal{M}(X,\mathcal{B}_X)$.*

Note that the compactness of X is essential. For example, the continuous tranformation $x \mapsto x+1$ on the real numbers has no invariant *probability* measure.

Proof of 10.10. Start with any $\mu_0 \in \mathcal{M}$ and let $\mu_n = f_*^{\circ n}(\mu_0)$. Consider the sequence of averages

$$\nu_n \;=\; \frac{\mu_0 + \mu_1 + \cdots + \mu_{n-1}}{n} \tag{10:4}$$

in the space $\mathcal{M} \subset \mathrm{C}(X,\mathbb{R})^*_{\rm weak}$. Since $\mathcal{M}$ is compact, some subsequence $\{\nu_{n_i}\}$ converges to a limit $\nu \in \mathcal{M}$. Now note that the difference

$$f_*(\nu_n) - \nu_n \;=\; (\mu_n - \mu_0)/n \;\in\; \mathrm{C}(X,\mathbb{R})^* ,$$

evaluated on $\psi \in \mathbb{C}(X,\mathbb{R})$, yields a real number

$$\Big(\int \psi\, d\mu_n \;-\; \int \psi\, d\mu_0\Big)/n$$

of absolute value at most $2\,\|\psi\|/n$. For any fixed ψ , this tends to zero as $n \to \infty$. It follows that the difference $f_*(\nu_n) - \nu_n \in \mathrm{C}(X,\mathbb{R})^*$ tends to zero as $n \to \infty$. Substituting $n = n_i$ and passing to the limit as $i \to \infty$, we see that $f_*(\nu) = \nu$. Further details of the proof are straightforward. $\square$

Remark. Recall from §3B that a measure $\mu \in \mathcal{M}$ is called an *asymptotic measure* for μ_0 under the action of f if for $[\mu_0]$-almost every $x_0 \in X$ the orbit $x_0 \mapsto x_1 \mapsto \cdots$ is *evenly distributed* with respect to μ. By definition, this means that the sequence

$$(\varphi(x_0) + \cdots + \varphi(x_{n-1}))/n \ ,$$

must converge to $\int \varphi \, d\mu$ for every continuous test function $\varphi : X \to \mathbb{R}$, or in other words that the sequence of measures

$$(\delta_{x_0} + \cdots + \delta_{x_{n-1}})/n$$

must converge to μ in the weak* topology. (Note that $\delta_{x_k} = f_*^{\circ k}(\delta_{x_0})$.) If such an asymptotic measure μ exists, then it follows that this sequence of measures $\nu_n \in \mathcal{M}$ defined by (10: 4) must also converge to μ. In fact we have

$$\int \varphi \, d\mu_k \ = \ \int \varphi \, d(f_*^{\circ k}\mu_0) \ = \ \int (\varphi \circ f^{\circ k}) \, d\mu_0 \ = \ \int \varphi(x_k) \, d\mu_0(x_0) \ .$$

Averaging from $k = 0$ to $n - 1$, we see that

$$\int \varphi \, d\nu_n \ = \ \int \tfrac{1}{n}(\varphi(x_0) + \cdots + \varphi(x_{n-1})) \, d\mu_0(x_0) \ .$$

Since the integrand on the right converges to the constant function $x_0 \mapsto \int \varphi \, d\mu$ for $[\mu_0]$-almost every x_0, it follows from the Lebesgue Dominated Convergence Theorem, that the sequnce $\int \varphi \, d\nu_n$ converges to $\int \varphi \, d\mu$, as asserted.

Recall from 8.7 that an ergodic probability measure for f can be characterized as an extreme point of the compact convex set $\mathcal{M}_f = \mathcal{M}_f(X, \mathcal{B}_X)$ consisting of all invariant probability measures on $(X, \mathcal{B}_X)$.

Theorem 10.11. *If X is compact metric and f is continuous, then the compact convex set $\mathcal{M}_f(X, \mathcal{B}_X)$ has at least one extreme point, hence there exists at least one invariant ergodic probability measure for f. In fact, every $\mu \in \mathcal{M}_f$ can be approximated arbitrarily closely by some weighted average $w_1 \mu_1 + \cdots + w_k \mu_k$ where the μ_i are ergodic.*

Remarks. The corresponding statement for a compact convex set K in an arbitrary locally convex topological vector space is known as the *Krein-Milman Theorem*. A more precise statement, due to *Choquet*, asserts that every $\mu_0 \in K$ can be represented as an integral

$$\mu_0 \ = \ \int_K \mu \, d\omega$$

where ω is a probability measure on K such that the set of extreme points has full measure. In the case $K = \mathcal{M}_f(X, \mathcal{B}_X)$ this "ergodic decomposition" of μ_0 is essentially unique. (Compare [Phelps], [Rochlin].)

Proof of Theorem 10.11. Let K_0 be an arbitrary non-vacuous compact convex subset of B^*_{weak}, where $B = \mathrm{C}(X, \mathbb{R})$. We will first show that K_0 has an extreme point. Choose a dense set $\{\varphi_i\} \subset B$ We will construct compact convex subsets $K_0 \supset K_1 \supset \cdots$ inductively as follows. For each $j > 0$ map the set K_{j-1} linearly into the real numbers by the correspondence $L \mapsto L(\varphi_j)$. The image of this map will be some closed interval of real numbers $[a_j \, , \, b_j]$, with $a_j \le b_j$. Define K_j inductively as the pre-image of b_j within

K_{j-1} . Then we claim that the intersection $\bigcap K_j$ consists of a single extreme point of K_0 . This intersection is non-vacuous, since the K_j are compact, non-vacuous and nested. It consists of a single point, since each projection $L \mapsto L(\varphi_j)$ carries it to a single point b_j . Suppose that this single intersection point $\hat{L}$ were not extreme, say $\hat{L} = (L_0 + L_1)/2$ with $L_0\,, L_1 \in K_0\,,\ L_0 \neq L_1$. Choosing the smallest j with $L_0(\varphi_j) \neq L_1(\varphi_j)$, we see that both L_0 and L_1 must belong to K_{j-1} . The projection from K_{j-1} to $[a_j\,,\,b_j] \subset \mathbb{R}$ is a linear map which carries L_0 and L_1 to different points of this interval, and yet carries their average $\hat{L}$ to the endpoint b_j . This is impossible.

To approximate some given $L \in K_0$ by a weighted average of extreme points, we modify this argument as follows. For any $n > 0$, note that the projection

$$L \ \mapsto \big(L(\varphi_1)\,,\,\ldots\,,\,L(\varphi_n)\big)$$

from B^* to the Euclidean space $\mathbb{R}^n$ carries K_0 onto some compact convex subset $K(n) \subset \mathbb{R}^n$. Now fix some $\epsilon > 0$. If n is sufficiently large, then we see from the expression (10: 3) that the preimage in K_0 of any point of $K(n)$ has diameter less than ϵ . Now express the image of L in $K(n)$ as a weighted average of extreme points $\mathbf{e}_i \in K(n)$, and then use the argument above to show that that each $\mathbf{e}_i$ is the image of an extreme point in K_0 . Hence L can be ϵ-approximated by a finite linear combination of extreme points.

The proof of 10.11 is now completed by choosing K_0 to be the set $\mathcal{M}_f$ of invarant probabilty measures, and applying 8.7. $\square$