Further Topics in Measure Theory

Jyunyi Liao

Contents

1 The Hardy-Littlewood Maximal Inequality and Differentiation

1.1 The Marcinkiewicz Interpolation Theorem

Distribution function and weak L^p spaces. Let $(X, \mathscr{F}, \mu)$ be a measure space. For a measurable function f on $(X, \mathscr{F}, \mu)$, define its *distribution function* $\lambda_f : (0, \infty) \to [0, \infty]$ by

$$\lambda_f(\alpha) = \mu\left(\{x \in X : |f(x)| > \alpha\}\right).$$

Some properties of the distribution function is clear:

- λ_f is decreasing on $(0, \infty)$.
- λ_f is right-continuous, since $\{|f| > \alpha\} = \bigcup_{n=1}^\infty \{|f| > \alpha + \epsilon_n\}$ for all $\epsilon_n > 0$ with $\epsilon_n \downarrow 0$.
- $\lambda_{f+g}(2\alpha) \leq \lambda_f(\alpha) + \lambda_g(\alpha)$. In addition, if $|f| \leq |g|$, then $\lambda_f \leq \lambda_g$.
- If the sequence (f_n) satisfies $|f_n| \uparrow |f|$, then $\lambda_{f_n} \to \lambda_f$ pointwise, since $\{|f| > \alpha\} = \bigcup_{n=1}^\infty \{|f_n| > \alpha\}$.

Let $1 \leq p < \infty$. For a measurable function f, define

$$[f]_p = \left(\sup_{\alpha > 0} \alpha^p \lambda_f(\alpha)\right)^{1/p}.$$

The *weak L^p space* is then defined to be the set of all measurable functions on $(X.\mathscr{F}, \mu)$ such that $[f]_p < \infty$. Note that $[\cdot]_p$ is not a norm, because it does not satisfy the triangle inequality.

By Chebyshev's inequality, we have $[f]_p \leq \|f\|_{L^p}$ for all $f \in L^p(X, \mathscr{F}, \mu)$. Therefore the classical L^p is contained in the weak L^p space.

Definition 1.1 (Sublinear operators of strong/weak type p)**.** Let T be an operator on the some vector space V of measurable functions from $(X, \mathscr{F}, \mu)$ to the space of all measurable functions on $(X, \mathscr{F}, \mu)$.

(i) The operator T is said to be *sublinear*, if $|T(f+g)| \leq |Tf| + |Tg|$ and $|T(cf)| = c|Tf|$ for all functions $f, g \in V$ and $c > 0$.

(ii) Let $1 \leq p \leq \infty$. The operator T is said to be *of strong type* p, if $L^p(X, \mathscr{F}, \mu) \subset V$ and there exists a constant $C_p > 0$ such that

$$\|Tf\|_{L^p} \leq C_p \|f\|_{L^p}$$

for all $f \in L^p(X, \mathscr{F}, \mu)$.

(iii) Let $1 \leq p < \infty$. The operator T is said to be *of weak type* p, if $L^p(X, \mathscr{F}, \mu) \subset V$ and there exists a constant $C_p > 0$ such that

$$[Tf]_p \leq C_p \|f\|_{L^p}$$

for all $f \in L^p(X, \mathscr{F}, \mu)$. Clearly, a sublinear operator T of strong type p is also of weak type p.

Theorem 1.2 (Marcinkiewicz interpolation theorem)**.** *If T is a sublinear operator of weak type p and q, where $1 \leq p < q < \infty$, then T is of strong type r for each $r \in (p, q)$.*

Remark. We take a glimpse at the idea of interpolation. If $f \in L^p(X, \mathscr{F}, \mu) \cap L^q(X, \mathscr{F}, \mu)$, then for each $r \in (p, q)$, the Hölder's inequality implies

$$\begin{aligned}\int_X |f|^r \, d\mu &= \int_X |f|^{\frac{p(q-r)}{q-p}} |f|^{\frac{q(r-p)}{q-p}} \, d\mu \\ &\leq \left(\int_X |f|^p \, d\mu\right)^{\frac{q-r}{q-p}} \left(\int_X |f|^q \, d\mu\right)^{\frac{r-p}{q-p}} = \|f\|_{L^p}^{\frac{p(q-r)}{q-p}} \|f\|_{L^q}^{\frac{q(r-p)}{q-p}}.\end{aligned}$$

Therefore $f \in L^r(X, \mathscr{F}, \mu)$, and $\|f\|_r$ can be recovered from L^p and L^q norms. If T of strong type p and q,

$$\|Tf\|_r = \|Tf\|_{L^p}^{\frac{p(q-r)}{r(q-p)}} \|Tf\|_{L^q}^{\frac{q(r-p)}{r(q-p)}} \leq A_r \|f\|_{L^p}^{\frac{p(q-r)}{r(q-p)}} \|f\|_{L^q}^{\frac{q(r-p)}{r(q-p)}} = A_r \|Tf\|_r.$$

The proof Marcinkiewicz interpolation theorem requires the following lemma.

Lemma 1.3. *If f is a measurable function on $(X, \mathscr{F}, \mu)$ and $0 < p < \infty$, then*

$$\int_X |f|^p \, d\mu = p \int_0^\infty \alpha^{p-1} \lambda_f(\alpha) \, d\alpha.$$

Proof. We may assume $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$, otherwise both integrals are infinite. We may also assume $f \geq 0$ by replacing f with $|f|$ if necessary. If f is simple, λ_f is a step function with jump discontinuities $0 < \alpha_1 < \cdots < \alpha_n$. We let $\alpha_0 = 0$. Then

$$\begin{aligned} p \int_0^\infty \alpha^{p-1} \lambda_f(\alpha) \, d\alpha &= \sum_{j=1}^n \int_{\alpha_{j-1}}^{\alpha_j} p\alpha^{p-1} \lambda_f(\alpha) \, d\alpha = \sum_{j=1}^n \left(|\alpha_j|^p - |\alpha_{j-1}|^p\right) \lambda_f(\alpha_{j-1}) \\ &= \sum_{j=1}^n |\alpha_j|^p \left(\lambda_f(\alpha_{j-1}) - \lambda_f(\alpha_j)\right) = \sum_{j=1}^n |\alpha_j|^p \mu(\{f = \alpha_j\}) = \int_X |f|^p \, d\mu. \end{aligned}$$

Since $f_n \uparrow f$ implies the pointwise convergence $\lambda_{f_n} \to \lambda_f$, the general result follows from simple function approximation and monotone convergence theorem. ☐

Lemma 1.4. *If f is a measurable function and $\alpha > 0$, define*

$$h_\alpha = f \chi_{\{|f| \leq \alpha\}} + \alpha(\operatorname{sgn} f) \chi_{\{|f| > \alpha\}}, \quad \textit{and} \quad g_\alpha = f - h_\alpha = (\operatorname{sgn} f)(|f| - \alpha) \chi_{\{|f| > \alpha\}}.$$

Then

$$\lambda_{g_\alpha}(t) = \lambda_f(t + \alpha), \quad \textit{and} \quad \lambda_{h_\alpha}(t) = \begin{cases} \lambda_f(t), & t < \alpha, \\ 0, & t \geq \alpha. \end{cases}$$

Proof. By definition, h_α is in fact the α-truncation of f, i.e. $h_\alpha = f$ when $|f| \leq \alpha$, and $h_\alpha = \alpha(\operatorname{sgn} f)$ when $|f| > \alpha$. Hence $\{|h| > t\} = \{|f| > t\}$ when $t < \alpha$, and $\{|h| > t\} = \{|f| > t\} = \emptyset$ when $t \geq \alpha$. On the other hand, note that $g_\alpha = 0$ on $\{|f| < \alpha\}$. For any $t > 0$, we have $\{|g_\alpha| > t\} = \{|f| - \alpha > t\} = \{|f| > t + \alpha\}$. ☐

Now we prove the Marcinkiewicz interpolation theorem.

Proof of Theorem 1.2. By assumption, there exists constants $C_p > 0$ and $C_q > 0$ such that

$$\lambda_{Tf}(\alpha) \leq \left(\frac{C_p}{\alpha}\right)^p \int_X |f|^p \, d\mu \quad \text{and} \quad \lambda_{Tf}(\alpha) \leq \left(\frac{C_q}{\alpha}\right)^q \int_X |f|^q \, d\mu.$$

For $f \in L^r(\mathbb{R}^d)$, we take g_α and h_α as in Lemma 1.4. By sublinearity of T,

$$\lambda_{Tf}(2\alpha) \leq \lambda_{Tg_\alpha}(\alpha) + \lambda_{Th_\alpha}(\alpha) \leq \left(\frac{C_p}{\alpha}\right)^p \int_X |g_\alpha|^p \, d\mu + \left(\frac{C_q}{\alpha}\right)^q \int_X |h_\alpha|^q \, d\mu. \tag{1.1}$$

Clearly $g_\alpha \in L^p(\mathbb{R}^d)$, and $h \in L^q(\mathbb{R}^d)$. By Lemma 1.3, we have

$$\begin{aligned} \int_X |g_\alpha|^p \, d\mu &= p \int_0^\infty t^{p-1} \lambda_{g_\alpha}(t) \, dt = p \int_0^\infty t^{p-1} \lambda_f(t + \alpha) \, dt \\ &= p \int_\alpha^\infty (t - \alpha)^{p-1} \lambda_f(t) \, dt \leq p \int_\alpha^\infty t^{p-1} \lambda_f(t) \, dt, \end{aligned} \tag{1.2}$$

and similarly,

$$\int_X |h_\alpha|^q \, d\mu = q \int_0^\infty t^{q-1} \lambda_{h_\alpha}(t) \, dt = q \int_0^\alpha t^{q-1} \lambda_f(t) \, dt. \tag{1.3}$$

We combine Lemma 1.3, the inequality (1.1) and the estimates (1.2)-(1.3):

$$\begin{aligned}\|Tf\|_{L^r}^r &= r\int_0^\infty (2\alpha)^{r-1}\lambda_{Tf}(2\alpha)\,d(2\alpha) = r2^r\int_0^\infty \alpha^{r-1}\lambda_{Tf}(2\alpha)\,d\alpha\\ &\le r2^r\int_0^\infty \left(C_p^p\alpha^{r-p-1}\int_X |g_\alpha|^p\,d\mu + C_q^q\alpha^{r-q-1}\int_X |h_\alpha|^q\,d\mu\right)d\alpha\\ &\le r2^r\left(C_p^p\int_0^\infty\int_\alpha^\infty \alpha^{r-p-1}t^{p-1}\lambda_f(t)\,dt\,d\alpha + C_q^q\int_0^\infty\int_0^\alpha \alpha^{r-q-1}t^{q-1}\lambda_f(t)\,dt\,d\alpha\right).\end{aligned}$$

Then apply Fubini's theorem to change the order of integration:

$$\begin{aligned}\|Tf\|_{L^r}^r &\le r2^r\left(C_p^p\int_0^\infty t^{p-1}\lambda_f(t)\int_0^t \alpha^{r-p-1}\,d\alpha\,dt + C_q^q\int_0^\infty t^{q-1}\lambda_f(t)\int_t^\infty \alpha^{r-q-1}\,d\alpha\,dt\right)\\ &= r2^r\left(\frac{C_p^p}{r-p}\int_0^\infty t^{r-1}\lambda_f(t)\,dt + \frac{C_q^q}{q-r}\int_0^\infty t^{r-1}\lambda_f(t)\,dt\right)\\ &= 2^r\left(\frac{C_p^p}{r-p}+\frac{C_q^q}{q-r}\right)\|f\|_{L^r}^r,\end{aligned}$$

where the last equality follows from Lemma 1.3. Hence T is of strong type r for each $r\in(p,q)$, as desired. □

Finally we deal with the exceptional case $q=\infty$.

Theorem 1.5 (Marcinkiewicz interpolation theorem)**.** *If T is a sublinear operator of weak type p and of strong type ∞, where $1\le p<\infty$, then T is of strong type r for each $r\in(p,\infty)$.*

Proof. We use the same notation as in the proof of Theorem 1.2. By assumption, there exists constants $C_p>0$ and $C_\infty>0$ such that for all $\phi\in L^p(\mathbb{R}^n)$,

$$\lambda_{T\phi}(\alpha)\le\left(\frac{C_p}{\alpha}\right)^p\int_X|\phi|^p\,d\mu \quad\text{and}\quad \|T\phi\|_{L^\infty}\le C_\infty\|\phi\|_{L^\infty}.$$

Take $f\in L^r(\mathbb{R}^d)$ and $\alpha>0$. Then $\|h_\alpha\|_{L^\infty}\le\alpha$ and $\|Th_\alpha\|_{L^\infty}\le C_\infty\alpha$, and consequently $\lambda_{Th_\alpha}(C_\infty\alpha)=0$. By sublinearity of T, we have

$$\lambda_{Tf}(2C_\infty t)\le\lambda_{Tg_\alpha}(C_\infty\alpha)+\lambda_{Th_\alpha}(C_\infty\alpha)=\lambda_{Tg_\alpha}(C_\infty\alpha)\le\left(\frac{C_p}{C_\infty\alpha}\right)^p\int_X|g_\alpha|^p\,d\mu. \tag{1.4}$$

Similar to the proof of Theorem 1.2, we apply the estimates (1.4), (1.2), and Fubini's theorem to get

$$\begin{aligned}\|Tf\|_{L^r}^r &= r\int_0^\infty (2C_\infty\alpha)^{r-1}\lambda_{Tf}(2C_\infty\alpha)\,d(2C_\infty\alpha) = r(2C_\infty)^r\int_0^\infty \alpha^{r-1}\lambda_{Tf}(2C_\infty\alpha)\,d\alpha\\ &\le r(2C_\infty)^r\int_0^\infty \alpha^{r-1}\left(\frac{C_p}{C_\infty\alpha}\right)^p\int_X|g_\alpha|^p\,d\mu\,d\alpha\\ &\le r2^rC_p^pC_\infty^{r-p}\int_0^\infty\int_\alpha^\infty \alpha^{r-p-1}t^{p-1}\lambda_f(t)\,dt\,d\alpha\\ &= r2^rC_p^pC_\infty^{r-p}\int_0^\infty t^{p-1}\lambda_f(t)\int_0^t\alpha^{r-p-1}\,d\alpha\,dt\\ &= r2^rC_p^pC_\infty^{r-p}\frac{1}{r-p}\int_0^\infty t^{r-1}\lambda_f(t)\,dt = \frac{2^rC_p^pC_\infty^{r-p}}{r-p}\|f\|_{L^r}^r.\end{aligned}$$

Therefore, T is of strong type r for each $r\in(p,\infty)$, as desired. □

Remark. We also say a sublinear operator T is of weak type ∞ if and only if T is of strong type ∞. Then q is allowed to be ∞ in Theorem 1.2, which is the case Theorem 1.5.

1.2 The Hardy-Littlewood Maximal Inequality

In this section, we work in Euclidean space $\mathbb{R}^d$ with the Lebesgue measure m. The measure (or volume) of an d-dimensional ball $B(x,r)$ of radius r is given by

$$m(B(x,r)) = \frac{\pi^{d/2}}{\Gamma\left(\frac{n}{2}+1\right)} r^d.$$

For a locally integrable function $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$, we define the local average

$$(A_r f)(x) = \frac{1}{m(B(x,r))} \int_{B(x,r)} f(y)\,dy, \quad x \in \mathbb{R}^d.$$

To obtain a uniform estimate for $A_r f$, we define the *Hardy-Littlewood maximal operator* by

$$(Mf)(x) = \sup_{r>0} (A_r|f|)(x) = \sup_{r>0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y)|\,dy, \quad x \in \mathbb{R}^d.$$

Clearly M is sublinear. The function Mf is also called the *Hardy-Littlewood maximal function of f.*

Theorem 1.6 (Hardy-Littlewood maximal inequality, weak type)**.** *The Hardy-Littlewood operator M is of weak type 1. In other words, there exists a constant $C_d > 0$ such that for all $f \in L^1(\mathbb{R}^d)$ and all $\lambda > 0$,*

$$m\left(\{Mf \geq \lambda\}\right) \leq \frac{C_d}{\lambda} \|f\|_{L^1}. \tag{1.5}$$

Remark. The inequality (1.5) may look a bit stronger than the condition $[Mf]_1 \leq C_d \|Mf\|_{L^1(\mathbb{R}^d)}$ of weak type 1. As we will see, the two assertions are indeed equivalent.

The proof of this inequality requires a finite version of Vitali covering theorem.

Lemma 1.7 (Vitali covering lemma)**.** *For any finite collection $\mathcal{F}$ of open balls $B_1, B_2, \cdots, B_N$ in an arbitrary metric space X, there exists a subcollection $\mathcal{G} \subset \mathcal{F}$ of disjoint balls such that*

$$\bigcup_{j=1}^N B_j \subset \bigcup_{B \in \mathcal{G}} 3B,$$

where $3B$ denotes the ball with the same center as B but with 3 times the radius.

Proof. We choose balls in $\mathcal{G}$ by the greedy algorithm. First take B_1' to be the largest ball among $\mathcal{F}$. Having chosen $\{B_1', B_2', \cdots, B_k'\}$, repeat the inductive step:

- if the remaining balls each have nonempty intersection with $\bigcup_{i=1}^k B_i'$, stop;
- otherwise, take B_{k+1}' to be the largest among $\mathcal{F} \backslash \{B_1', B_2', \cdots, B_k'\}$ that are disjoint from $\bigcup_{i=1}^k B_i'$.

This algorithm must stop after less than N rounds, with the chosen balls $B_1', B_2', \cdots, B_n'$ disjoint. Then it remains to show that $B_i \subset E := \bigcup_{j=1}^n 3B_j'$ for every $i = 1, \cdots, N$. We claim $B_i \cap E \neq \emptyset$, otherwise the algorithm would not have stopped at $B_1', B_2', \cdots, B_n'$. We let j_0 be the minimal j such that $B_j' \cap B_i \neq \emptyset$. Then B_i does not intersect $\bigcup_{i=1}^{j_0-1} B_j'$, and the radius of B_i is no greater than B_{j_0}, since B_{j_0} is maximal at step j_0. Recalling that $B_{j_0}' \cap B_i \neq \emptyset$, by triangle inequality, $3B_{j_0}' \supset B_i$. $\square$

Proof of Hardy-Littlewood maximal inequality [Theorem 1.6]. We will show that for all $f \in L^1(\mathbb{R}^d)$,

$$m(\{Mf > \lambda\}) \leq \frac{3^d}{\lambda} \|f\|_{L^1}, \quad \lambda > 0.$$

Noticing that $m(\{Mf \geq \lambda\}) \leq m(\{Mf > \lambda - \epsilon\}) \leq 3^d (\lambda - \epsilon)^{-1} \|f\|_{L^1}$ for sufficiently small $\epsilon > 0$, the desired inequality (1.5) follows by perturbing $\epsilon \downarrow 0$.

Using the inner regularity of the Lebesgue measure, it suffices to show that $m(K) \le 3^d\lambda^{-1}\|f\|_{L^1}$ for each compact subset $K \subset \{Mf > \lambda\}$. For each $x \in K$, we take $r_x > 0$ such that

$$\frac{1}{m(B(x,r_x))}\int_{B(x,r_x)} |f|\,dm > \lambda.$$

The collection of balls $B(x, r_x)$ forms an open cover of K, and we may take by compactness a finite subcollection that covers K. By Vitali covering lemma [Lemma 1.7], we take a further collection of disjoint balls $B_1, B_2, \cdots, B_n$ such that $K \subset \bigcup_{j=1}^n B_j$. Consequently,

$$m(K) \le 3^d \sum_{j=1}^n m(B_j) \le \frac{3^d}{\lambda}\sum_{j=1}^n \int_{B_j} |f|\,dm = \frac{3^d}{\lambda}\int_{\bigcup_{j=1}^n B_j} |f|\,dm \le \frac{3^d}{\lambda}\|f\|_{L^1}. \qquad \square$$

Using the Marcinkiewicz interpolation theorem, we immediately obtain the following result.

Theorem 1.8 (Hardy-Littlewood maximal inequality, strong type). *Let $1 < p \le \infty$. The Hardy-Littlewood operator M is of strong type p. That is, there exists a constant $C_{p,d} > 0$ such that for all $f \in L^p(\mathbb{R}^d)$,*

$$\|Mf\|_{L^p} \le C_{p,d}\|f\|_{L^p}.$$

Proof. The Hardy-Littlewood operator M is sublinear and of weak type 1. By definition of Mf, we also have $\|Mf\|_{L^\infty} \le \|f\|_{L^\infty}$ when f is a.e. bounded. Hence M is of strong type ∞, and is of strong type p for each $1 < p \le \infty$ by Marcinkiewicz interpolation theorem [Theorem 1.5] $\square$

1.3 The Differentiation Theorems

In this section we apply the Hardy-Littlewood maximal inequality to prove some differentiation theorems.

Theorem 1.9 (Lebesgue differentiation theorem). *Let $f \in L^1_{\text{loc}}(\mathbb{R}^d)$. For almost every $x \in \mathbb{R}^n$,*

$$\lim_{r\to 0^+} \frac{1}{m(B(x,r))}\int_{B(x,r)} |f(y) - f(x)|\,dy = 0. \tag{1.6}$$

Consequently, the local average function $A_r f$ converges almost everywhere to f, i.e.

$$\lim_{r\to 0^+}(A_r f)(x) = \lim_{r\to 0^+}\frac{1}{m(B(x,r))}\int_{B(x,r)} f(y)\,dy = f(x) \tag{1.7}$$

for almost every $x \in \mathbb{R}^n$.

Remark. Let f be a measurable function on $\mathbb{R}^d$. A point $x \in \mathbb{R}^d$ is said to be a *Lebesgue point of f* if the identity (1.6) holds. The Lebesgue differentiation theorem implies that if $f \in L^1_{\text{loc}}(\mathbb{R}^d)$, then almost every point in $\mathbb{R}^n$ is a Lebesgue point of f.

Proof. We first prove the result for $g \in C_c(\mathbb{R}^d)$. If $x \in \mathbb{R}^d$ and $\epsilon > 0$, by uniform continuity of g, there exists $\delta > 0$ such that $|g(y) - g(x)| < \epsilon$ for all $y \in B(x,\delta)$. Then for all $r < \delta$,

$$\frac{1}{m(B(x,r))}\int_{B(x,r)} |g(y) - g(x)|\,dy < \epsilon.$$

Hence (1.6) holds for all continuous functions with compact support.

Now we prove the general case. Since differentiation is a local property, we may assume that $f \in L^1(\mathbb{R}^d)$. For $\epsilon > 0$, choose $g \in C_c(\mathbb{R}^d)$ such that $\|f - g\|_{L^1} \le \epsilon$. We put $h = f - g$. By the triangle inequality,

$$|(A_r f)(x) - f(x)| \le |(A_r g)(x) - g(x)| + |(A_r h)(x) - h(x)| \le |(A_r g)(x) - g(x)| + (A_r|h|)(x) + |h(x)|.$$

Let $\lambda > 0$. Then

$$
\begin{aligned}
&m\left(\left\{x \in \mathbb{R}^d : \limsup_{r\to 0^+} |A_r f - f|(x) \geq \lambda\right\}\right) \\
&\quad \leq m\left(\left\{x \in \mathbb{R}^d : \limsup_{r\to 0^+} |A_r g - g|(x) \geq \frac{\lambda}{3}\right\}\right) + m\left(\left\{x \in \mathbb{R}^d : \limsup_{r\to 0^+} (A_r|h|)(x) \geq \frac{\lambda}{3}\right\}\right) + m\left(|h| \geq \frac{\lambda}{3}\right) \\
&\quad \leq m\left(\left\{x \in \mathbb{R}^d : \sup_{r>0} (A_r|h|)(x) \geq \frac{\lambda}{3}\right\}\right) + m\left(|h| \geq \frac{\lambda}{3}\right).
\end{aligned}
$$

By weak L^1 Hardy-Littlewood maximal inequality [Theorem 1.5] and Markov inequality,

$$
m\left(\left\{x \in \mathbb{R}^d : \limsup_{r\to 0^+} |A_r f - f|(x) \geq \lambda\right\}\right) \leq \frac{3C_d}{\lambda}\|h\|_{L^1} + \frac{3}{\lambda}\|h\|_{L^1} \leq \frac{3(C_d+1)\epsilon}{\lambda}.
$$

Since $\epsilon > 0$ is arbitrary, the left-hand side of the last display is zero. The result then follows by taking the union on the sequence $\lambda_n = \frac{1}{n} \downarrow 0$. □

Following is a particular case of Lebesgue differentiation theorem.

Theorem 1.10 (Lebesgue density theorem). *Let $E \subset \mathbb{R}^d$ be a Lebesgue measurable set. For almost every point $x \in E$, the density*

$$
\lim_{r\to 0^+} \frac{m(E \cap B(x,r))}{m(B(x,r))} = 1, \tag{1.8}
$$

and for almost every point $y \notin E$,

$$
\lim_{r\to 0^+} \frac{m(E \cap B(y,r))}{m(B(y,r))} = 0. \tag{1.9}
$$

Remark. Let $E \subset \mathbb{R}^d$. A point $x \in \mathbb{R}^d$ is said to be a *density point of E* if the identity (1.8) holds. The Lebesgue density theorem implies that almost every point of a measurable set is a density point, and almost every point outside the measurable set is not a density point.

Proof. The identities (1.8) and (1.9) are special cases of (1.7) when $f = \chi_E$. □

We can employ the Lebesgue differentiation theorem to prove the Fundamental theorem of calculus.

Theorem 1.11 (Fundamental Theorem of Calculus). *Let $F : \mathbb{R} \to \mathbb{C}$ be an absolutely continuous function. Then F is differentiable almost everywhere, and the derivative $f = F'$ satisfies*

$$
F(x) = F(a) + \int_a^x f(t)\,dt, \quad -\infty < a < x < \infty.
$$

Proof. Since the differentiability is a local property, it suffices to deal with the restriction of F on a compact interval $[a,b]$. Let μ_F be the Lebesgue-Stieltjes measure generated by f on $[a,b]$.

Step I. We claim that μ_F is absolutely continuous with respect to the Lebesgue measure m.

We fix $\epsilon > 0$, and choose $\delta > 0$ such that $\sum_{j=1}^N |F(b_j) - F(a_j)| < \epsilon$ for all disjoint intervals $\{(a_j,b_j)\}_{j=1}^N$ with total length less than δ. If E is a Borel set with $m(E) = 0$, by outer regularity of m, we take an open $U \supset E$ with $m(U) < \delta$. Then U is a disjoint union of at most countably many intervals $\{(a_j,b_j)\}_{j=1}^\infty$, and

$$
\sum_{j=1}^N \mu_F((a_j,b_j)) \leq \sum_{j=1}^N (F(b_j) - F(a_j)) \leq \epsilon.
$$

Letting $N \to \infty$, we have $\mu_F(U) < \epsilon$, and $\mu_F(E) < \epsilon$. Since $\epsilon > 0$ is arbitrary, $\mu_F(E) = 0$.

Step II. By Radon-Nikodym theorem, we take $f \in L^1([a,b])$ such that $\mu_F(E) = \int_E f\,dm$. We may further globalize this result and assert that there exists a locally integrable function $f \in L^1_{\text{loc}}(\mathbb{R})$ such that

$$\mu_F((x,y]) = F(y) - F(x) = \int_x^y f(t)\,dt \quad \text{for all } -\infty < x < y < \infty.$$

Step III. If $x \in \mathbb{R}$ is a Lebesgue point of f, by Lebesgue differentiation theorem,

$$\lim_{r\to 0^+} \frac{1}{2r} \int_{x-r}^{x+r} |f(y) - f(x)|\,dy = 0.$$

We split the integral to $[x-r, x]$ and $[x, x+r]$ to get

$$\lim_{r\to 0^+} \frac{1}{r} \int_{x}^{x+r} |f(y) - f(x)|\,dy = \lim_{r\to 0^+} \frac{1}{r} \int_{x-r}^{x} |f(y) - f(x)|\,dy = 0.$$

Hence the right derivative of F at x is

$$\lim_{r\to 0^+} \frac{F(x+r) - F(x)}{r} = \lim_{r\to 0^+} \frac{1}{r} \int_{x}^{x+r} |f(y) - f(x)|\,dy = f(x),$$

and the same for the left derivative. Therefore F is differentiable almost everywhere, and $F' = f$. $\square$

Remark. A special case of this theorem is the one-dimensional Rademacher's theorem. If we further assume that $F : \mathbb{R} \to \mathbb{C}$ is Lipschitz continuous, then F is almost everywhere differentiable and $F' \in L^\infty(\mathbb{R})$. Indeed, the essential supremum of F' is bounded by the Lipschitz constant.

Theorem 1.12 (Rademacher's Theorem)**.** *If $f : \mathbb{R}^d \to \mathbb{C}$ is a Lipschitz continuous function, then f is almost everywhere differentiable.*

2 Radon Measures

2.1 Locally Compact Hausdorff (LCH) Spaces

Topology review. Throughout this section, we are mainly concerned with the *Locally Compact Hausdorff (LCH)* space. To be specific, the topological space X of our interest has the following topological properties:

- X is *Hausdorff*, i.e. for each pair of distinct points x and y in X, there exists a neighborhood U_x of x and a neighborhood U_y of y such that U and V are disjoint.
- X is *locally compact*, i.e. every point in X has a compact neighborhood.

The following proposition describes that, for any set K compactly included in an open set U, we can always find a set V between them in sense of compact inclusion.

Proposition 2.1. *If X is an LCH space and $K \subset U \subset X$, where K is compact and U is open, there exists a precompact open set V such that $K \subset V \subset \overline{V} \subset U$.*

Proof. Our proof are divided into three steps.

• *Step I:* We first show that, in a Hausdorff space X, we can separate a compact set K and a single point $x \notin K$ outside the set with disjoint neighborhoods. Formally, we find two disjoint open sets $U \supset K$ and $V \ni x$.

For each $y \in K$, by Hausdorff property, we can find two disjoint neighborhoods U_y of y and V_y of x. By compactness of K, it is possible to cover K by finitely many such neighborhoods $U_{y_1}, \cdots, U_{y_n}$. We then set $U = \bigcap_{j=1}^n U_{y_j}$ and $V = \bigcap_{j=1}^n V_{y_j}$, which has the desired properties.

• *Step II:* Next, we assume X is LCH and show that any open neighborhood U of a point x contains a compact neighborhood of x. We may assume that $\overline{U}$ is compact, otherwise we may replace U by its intersection with the interior of a compact neighborhood of x. Then ∂U is also a compact set, and we can separate x and ∂U by two disjoint open sets $V \ni x$ and $W \supset \partial U$ in U. Hence V satisfies $\overline{V} \subset (W^c \cap \overline{U}) \subset U$, and since U is precompact, $\overline{V}$ is a compact subset of U. Therefore $\overline{V}$ is a compact neighborhood of x.

• *Step III:* Finally we come to the original proposition. By Step II, we find a precompact open neighborhood V_x for each $x \in K$ such that $x \subset V_x \subset \overline{V}_x \subset U$. By compactness of K, we take finitely many such neighborhoods $V_{x_1}, \cdots, V_{x_n}$ to cover K. Setting $V = \bigcup_{j=1}^n V_{x_j}$, we have $K \subset V \subset \overline{V} \subset U$, and $\overline{V}$ is compact. □

Now we discuss the generalized version of Urysohn's lemma and Tietze extension theorem in LCH spaces. Recall that *every compact Hausdorff is normal*, to which the original version of these theorems applies.

Theorem 2.2 (Urysohn's lemma in LCH spaces)**.** *Let X be an LCH space and $K \subset U \subset X$, where K is compact and U is open. There exists $f \in C(X, [0,1])$ such that $f = 1$ on K and $f = 0$ outside a compact subset of U.*

Proof. We take a precompact open set V such that $K \subset V \subset \overline{V} \subset U$, as in Proposition 2.1, so $\overline{V}$ is normal. By Urysohn's lemma for normal spaces, there exists $f \in C(\overline{V}, [0,1])$ such that $f = 1$ on K and $f = 0$ on ∂V. We extend f to X by setting $f = 0$ on $\overline{V}^c$. It remains to show that $f \in C(X)$.

Let E be a closed subset of $[0,1]$. If $0 \notin E$, we have $f^{-1}(E) = (f|_{\overline{V}})^{-1}(E)$, and if $0 \in E$, we have $f^{-1}(E) = (f|_{\overline{V}})^{-1}(E) \cup \overline{V}^c = (f|_{\overline{V}})^{-1}(E) \cup V^c$ since $(f|_{\overline{V}})^{-1}(E) \supset \partial V$. In either case, $f^{-1}(E)$ is closed. Therefore f is continuous. □

The following theorem can be proved in a similar approach.

Theorem 2.3 (Tietze extension theorem in LCH spaces)**.** *Let X be an LCH space and $K \subset X$, where K is compact. If $f \in C(K)$, there exists $F \in C(X)$ such that $F|_K = f$. Moreover, F may be taken to vanish outside a compact set, i.e. $F \in C_c(X)$.*

Proof. We take a precompact set V such that $K \subset V \subset \overline{V} \subset X$, so $\overline{V}$ is normal. By Tietze extension theorem for normal spaces, we can extend f to a function $g \in C(\overline{V})$ with $g|_K = f$. We also take a function $\phi \in C(\overline{V}, [0,1])$ such that $\phi = 1$ on K and $\phi = 0$ on ∂V by Urysohn's lemma. Then $g\phi \in C(\overline{V})$ agrees with f on K. We take $F = g\phi$ on $\overline{V}$ and $F = 0$ in $\overline{V}^c$. Then $F \in C_c(X)$ and $F|_K = f$. □

Alexandroff compactification. If X is a noncompact LCH space, it is possible to make X into a compact Hausdorff space by adding a single point at the "infinity". Let us take some object that is not a point of X, denoted by the symbol ∞ for convenience, and adjoin it to X, forming the set $X^* = X \cup \{\infty\}$. We topologize X^* by defining the collection $\mathscr{T}^*$ of open sets of X^* to consist of

(i) all sets U that are open in X, and

(ii) all sets of the form $X^* \backslash K$, where K is a compact subset of X.

We first check that such collection is indeed a topology on X^*.

- The empty set $\emptyset$ and X^* are open sets of type (i) and (ii), respectively.
- Let U_1 and U_2 be open sets in X, and let K_1 and K_2 be compact sets in X. Then
  - $U_1 \cap U_2$ is of type (i),
  - $(X^* \backslash K_1) \cap (X^* \backslash K_2) = X^* \backslash (K_1 \cup K_2)$ is of type (ii), and
  - $U \cap (X^* \backslash K) = U \cap (X \backslash K)$ is of type (i).

  Hence $\mathscr{T}^*$ is closed under the finite intersection operation.
- Let $\{U_\alpha\}$ be a collection of open sets of X, and let $\{K_\beta\}$ be a collection of compact sets in X. Then
  - $\bigcup_\alpha U_\alpha = U$ is of type (i),
  - $\bigcup_\beta (X^* \backslash K_\beta) = X^* \backslash \bigcap_\beta K_\beta = X^* \backslash K$ is of type (ii), and
  - $U \cup (X^* \backslash K) = X^* \backslash (K \backslash U)$ is of type (ii) since $K \backslash U$ is a compact subset of X.

  Hence $\mathscr{T}^*$ is closed under the union operation.

Then we need to verify that X is a subspace of X^*:

- Given any open set in X^*, its intersection with X is open in X. If the open set is of type (i), it is clearly open in X. If it is of type (ii), then $(X^* \backslash K) \cap X = X \backslash K$ is open in Hausdorff space X.
- Conversely, given any open set in X, it is a type (i) open set in X^*.

Next we verify that X^* is a compact topological space.

- If $\mathscr{A}$ is an open cover of X^*, it must contain at least one open set $X^* \backslash K$ of type (ii), to contain ∞.
- Taking all members in $\mathscr{A}$ but $X^* \backslash K$ and intersect them with X, we obtain a cover of X. Since K is a compact subset of X, finitely many of them cover K. Then the corresponding finite collection of elements of $\mathscr{A}$ along with $X^* \backslash K$ form a cover of X^*.

Finally we verify that X^* is a Hausdorff space. Let x and y be two distinct points of X^*:

- The case that both x and y lies in X is clear since X is Hausdorff.
- If $y = \infty$, we choose a compact set K in X that contains a neighborhood U of x, then U and $X^* \backslash K$ are disjoint neighborhoods of x and ∞, respectively, in X^*.

The comapact Hausdorff space X^* is called the *one point compactification/Alexandroff compactification of* X.

Functions vanishing at infinity. Let X be a topological space. A continuous function $f \in C(X)$ is said to *vanish at infinity* if the set $\{x \in X : |f(x)| \geq \epsilon\}$ is compact for every $\epsilon > 0$. We define $C_0(X)$ to be the space of functions vanishing at infinity.

Proposition 2.4. *Let X be an LCH space, and $f \in C(X)$. The function f extends continuously to the Alexandroff compactification X^* of X if and only if there exists function $g \in C_0(X)$ and $z \in \mathbb{C}$ such that $f = g + c$, in which case the continuous extension is given by $f(\infty) = c$.*

Proof. Assume $f = g + c$, where $g \in C_0(X)$ and $c \in \mathbb{C}$. Replacing f by $f - c$, we may further assume $c = 0$. We extend f to X^* by setting $f(\infty) = 0$, and show that f is continuous. Let U be an open subset of $\mathbb{C}$.

- If $0 \notin U$, then $f^{-1}(U) = (f|_X)^{-1}(U)$, which is open by continuity of $f|_X$.
- If $0 \in U$, there exists $\epsilon > 0$ such that $|z| \geq \epsilon$ for all $z \in U^c$. Since $f|_X \in C_0(X)$, $(f|_X)^{-1}(U^c)$ is a closed subset of the compact set $\{x \in X : |f(x)| \geq \epsilon\}$ in X. Hence $f^{-1}(U) = X^* \backslash (f|_X)^{-1}(U^c)$ is open.

Conversely, if $f \in C(X)$ extends continuously to X^*, we let $c = f(\infty)$ and $g = f - c$. For each $\epsilon > 0$, the set $g^{-1}(B(0, \epsilon)) = \{x \in X^* : |g(x)| < \epsilon\}$ is open in X^* and contains ∞. Consequently, the complement $\{x \in X^* : |g(x)| \geq \epsilon\}$ is a compact set in X. Therefore $g \in C_0(X)$. □

Topologies on $\mathbb{C}^X$. Let X be a topological space. There are various ways to topologize the space $\mathbb{C}^X$ of all complex-valued functions on X:

- *The topology of pointwise convergence/the product topology* is generated by the sets
$$U^{\epsilon}_{x_1,\cdots,x_m}(f) = \left\{g \in \mathbb{C}^X : |f(x_j) - g(x_j)| < \epsilon,\ j = 1, 2, \cdots, m\right\},$$
where $f \in \mathbb{C}^X$, $\epsilon > 0$ and $x_1, \cdots, x_m \in X$. In this topology, a sequence (f_n) of functions converges to f when $f_n \to f$ pointwise.
- *The topology of compact convergence* is generated by the sets
$$U^{\epsilon}_{K}(f) = \left\{g \in \mathbb{C}^X : \sup_{x \in K} |f(x) - g(x)| < \epsilon\right\},$$
where $f \in \mathbb{C}^X$, $\epsilon > 0$ and K is a compact subset of X. In this topology, a sequence (f_n) of functions converges to f when $f_n \to f$ uniformly on every compact subset K of X.
- *The topology of uniform convergence* is generated by the sets
$$U^{\epsilon}_{\infty}(f) = \left\{g \in \mathbb{C}^X : \sup_{x \in X} |f(x) - g(x)| < \epsilon\right\},$$
where $f \in \mathbb{C}^X$ and $\epsilon > 0$. In this topology, a sequence (f_n) of functions converges to f when $f_n \to f$ uniformly on X.

Basic analysis shows that the space $C(X)$ of continuous functions on X is not a closed subspace of $\mathbb{C}^X$ in the topology of pointwise convergence, but when we switch to the uniform topology, it is. The following theorem asserts that $C(X)$ is also closed in the topology of compact convergence when X is an LCH space.

Proposition 2.5. *If X is an LCH space, $C(X)$ is closed in $\mathbb{C}^X$ in the topology of compact convergence.*

Proof. We claim that, *a subset E of X is closed if and only if $E \cap K$ is closed for each compact set $K \subset X$.* In fact, if E is closed, $E \cap K$ must be closed since it is the intersection of two closed sets. On the other hand, if E is not closed, we choose a point $x \in \overline{E} \backslash E$ and let K be a compact neighborhood of x. Then x is a limit point of $E \cap K$, however it is not in $E \cap K$.

Now we prove the desired result. If f is in the closure of $C(X)$, then for each compact subset K of X, the restriction $f|_K$, being a uniform limit of continuous functions on K, is continuous. Then for any closed set $E \subset X$, the intersection $f^{-1}(E) \cap K = (f|_K)^{-1}(E)$ is closed for all compact subset K of X, and hence $f^{-1}(E)$ is closed. Therefore f is also in $C(X)$. □

Proposition 2.6. *If X is an LCH space, $C_0(X) = \overline{C_c(X)}$ in the uniform topology.*

Proof. If f is in the closure of $C_c(X)$, for every $\epsilon > 0$, we can take some $g \in C_c(X)$ such that $\|f - g\|_\infty < \epsilon$. Then $\{x \in X : |f(x)| \geq \epsilon\} \subset \operatorname{supp} g$, which are compact sets.

Conversely, if $f \in C_0(X)$, we show how to find a function $g \in C_c(X)$ with $\|f - g\|_\infty < \epsilon$ for any $\epsilon > 0$. We take the compact set $K = \{x \in X : |f(x)| > \epsilon\}$, and take $\phi \in C_c(X, [0, 1])$ such that $\phi = 1$ on K by Urysohn's lemma [Theorem 2.2]. Setting $g = f\phi$ completes the proof. □

Proposition 2.7 (Partition of unity)**.** *Let X be an LCH space, K a compact subset of X, and $(U_j)_{j=1}^n$ an open cover of K. There exists a family of functions $\phi_j \in C_c(U_j, [0, 1])$ such that $\sum_{j=1}^n \phi_j(x) = 1$ for all $x \in K$.*

Proof. By Proposition 2.1, for each $x \in X$, we take a precompact open neighborhood V_x of x contained in some U_j. Then by compactness of K, there exist finitely many $V_{x_1}, \cdots, V_{x_m}$ that form a cover of K. We denote by K_j the union of neighborhoods V_{x_k} contained in U_j. By Urysohn's lemma, for each $j = 1, 2, \cdots, n$ we can find a function $g_j \in C_c(U_j, [0,1])$ such that $g_j = 1$ on K_j. Furthermore, there also exists a function $f \in C_c(X, [0,1])$ such that $f = 1$ on K and $\text{supp}(f) \subset \{x \in X : \sum_{j=1}^n g_k(x) > 0\}$. Let $g_{n+1} = 1 - f$, so that $\sum_{j=1}^{n+1} g_j > 0$ everywhere. Taking $\phi_j = g_j / \sum_{k=1}^{n+1} g_k$, we have $\phi_j \in C_c(U_j, [0,1])$ and $\sum_{j=1}^n \phi_j = 1$ on K. ☐

σ-compactness. A topological space is said to be *σ-compact* if it is a countable union of compact sets. Formally, if X is σ-compact, there exists compact subsets $K_n \subset X$ such that $X = \bigcup_{n=1}^\infty K_n$. Replacing K_n by the union of itself and all preceding members, we may assume that (K_n) is an increasing sequence.

A second countable LCH space is σ-compact. To see this, we take a precompact open neighborhood U_x for each $x \in X$. Consequently, we can find a base set $B_x \in \mathscr{B}$ such that $x \in B_x \subset U_x$, and $\overline{B}_x$ is compact. We choose $\mathscr{B}_c \subset \mathscr{B}$ to be the collection of all precompact base sets. Then $B_x \in \mathscr{B}_c$ for all $x \in X$, and $X = \bigcup_{B \in \mathscr{B}_c} \overline{B}$ is a countable union of compact sets. Therefore, X is a σ-compact topological space.

Proposition 2.8. *Let X be a σ-compact LCH space. There exists a sequence $(U_n)_{n=1}^\infty$ of precompact open sets such that $U_1 \subset \overline{U}_1 \subset U_2 \subset \overline{U}_2 \subset U_3 \subset \cdots U_n \subset \overline{U}_n \subset U_{n+1} \subset \cdots$ and $X = \bigcup_{n=1}^\infty U_n$. Furthermore, for all compact set $K \subset X$, there exists $n \in \mathbb{N}$ such that $U_n \supset K$.*

Proof. By σ-compactness of X, there exists a sequence $(K_n)_{n=1}^\infty$ of compact sets increasing to X. We start by taking a precompact open neighborhood U_x for each $x \in X$ and setting $U_0 = \emptyset$. With U_{n-1} constructed, the union $\overline{U}_{n-1} \cup K_n$ is compact, and there exists finitely many $x_1, \cdots, x_k \in X$ such that $(\overline{U}_{n-1} \cup K_n) \subset \bigcup_{j=1}^k U_{x_j}$. We construct $U_n = \bigcup_{j=1}^k U_{x_j}$, which is also precompact open. Then we have $\overline{U}_{n-1} \subset U_n$. Moreover,

$$\bigcup_{n=1}^\infty U_n \supset \bigcup_{n=1}^\infty K_n = X.$$

Hence the sequence (U_n) has the desired property. Moreover, for any compact subset K of X, $\{U_n\}_{n=1}^\infty$ is an open cover of K, hence there exists U_n such that $K \subset U_n$. ☐

Proposition 2.9. *Let X be a σ-compact LCH space, and let $(U_n)_{n=1}^\infty$ be a sequence of precompact sets as in Proposition 2.8. Then for each $f \in \mathbb{C}^X$, the sets*

$$\left\{ g \in \mathbb{C}^X : \sup_{x \in \overline{U}_n} |g(x) - f(x)| < \frac{1}{m} \right\}, \quad m, n \in \mathbb{N} \tag{2.1}$$

form a neighborhood base for f in the topology of compact convergence. Hence this topology is first countable, and $f_k \to f$ uniformly on compact sets if and only if $f_n \to f$ uniformly on each $\overline{U}_n$.

Proof. For $f \in \mathbb{C}^X$, any neighborhood of f in the topology of compact convergence contains a set of the form

$$U_K^\epsilon(f) = \left\{ g \in \mathbb{C}^X : \sup_{x \in K} |g(x) - f(x)| < \epsilon \right\},$$

where K is a compact subset of X and $\epsilon > 0$. We choose $n, m \in \mathbb{N}$ such that $K \subset U_n$ and $\frac{1}{m} < \epsilon$. Then

$$U_K^\epsilon(f) \supset \left\{ g \in \mathbb{C}^X : \sup_{x \in \overline{U}_n} |g(x) - f(x)| < \frac{1}{m} \right\}.$$

Therefore the sets of the form (2.1) form a neighborhood base for f. ☐

2.2 Positive Linear Functionals on $C_c(X)$ and Radon Measures

Throughout this section, we assume that X is an LCH space. One of the vector spaces we are interested in is the space $C_c(X)$ of continuous functions on X with compact support.

Definition 2.10 (Positive linear functionals)**.** Let X be an LCH space. A *positive linear functional on* $C_c(X)$ is a linear functional $T : C_c(X) \to \mathbb{C}$ such that $Tf \geq 0$ for all $f \in C_c(X)$ with $f \geq 0$.

Though there is no mention of continuity in the definition above, the positivity condition itself implies a rather strong continuity property.

Proposition 2.11. *If T is a positive linear functional on $C_c(X)$, for each compact set $K \subset X$, there exists a constant $C_K > 0$ such that $|Tf| \leq C_K \|f\|_\infty$ for all $f \in C_c(X)$ with* $\operatorname{supp}(f) \subset K$.

Proof. By dividing $f \in C_c(X)$ into real and imaginary parts, it suffices to consider real-valued functions f. By Urysohn's lemma, for any compact $K \subset X$, there is a function $\phi \in C_c(U, [0, 1])$ such that $\phi = 1$ on K. Then if $\operatorname{supp}(f) \subset K$, we have $|f| \leq \|f\|_\infty \phi$. Hence both $\|f\|_\infty \phi - f$ and $\|f\|_\infty \phi + f$ are nonnegative, and

$$T\phi \|f\|_\infty - Tf \geq 0, \quad T\phi \|f\|_\infty + Tf \geq 0$$

Therefore $|Tf| \leq T\phi \|f\|_\infty$, which concludes the proof by setting $C_K = T\phi$. □

The positive linear functionals on $C_c(X)$ is closely related to a family of Borel measures on X with some regular properties. Intuitively, we let μ be a Borel measure on X such that $\mu(K) < \infty$ for all compact $K \subset U$. Then the map $f \mapsto \int_X f \, d\mu$ is a positive linear functional on $C_c(X)$, since $f \in C_c(X) \subset L^1(\mu)$.

Definition 2.12 (Radon measures)**.** Let X be a topological space, $\mathscr{B}$ the Borel σ-algebra on X, and μ a measure on $(X, \mathscr{B})$. Let E be a Borel subset of X.

(i) μ is said to be *outer regular on* E, if

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

(ii) μ is said to be *inner regular on* E, if

$$\mu(E) = \inf \{\mu(K) : K \subset E,\ K \text{ is compact}\}.$$

(iii) μ is said to be *regular*, if it is outer and inner regular on all Borel sets.

(iv) μ is called a *Radon measure*, if it is finite on all compact sets, outer regular on all Borel sets, and inner regular on all open sets.

The following theorem relates every positive linear functional on $C_c(X)$ with a Radon measure on X.

Theorem 2.13 (Riesz representation theorem)**.** *Let X be a LCH space. If T is a positive linear $C_c(X)$, there exists a unique Radon measure μ on X such that*

$$Tf = \int_X f \, d\mu, \quad \forall f \in C_c(X).$$

Furthermore, for all open sets $U \subset X$, μ satisfies

$$\mu(U) = \sup \{Tf : f \in C_c(U),\ 0 \leq f \leq 1\},$$

and for all compact sets $K \subset X$,

$$\mu(K) = \inf \{Tf : f \in C_c(X),\ f \geq \chi_K\}.$$

We begin by constructing a Radon measure from a positive linear functional on $C_c(X)$.

Lemma 2.14. *Let T be a positive linear functional on $C_c(X)$. For each open $U \subset X$, define*

$$\mu(U) = \sup \{Tf : f \in C_c(U, [0,1])\},$$

and for each subset $E \in 2^X$, define

$$\mu^*(E) = \inf \{\mu(U) : U \supset E,\ U \text{ is open}\}. \tag{2.2}$$

Then μ^* is an outer measure on X, and every open set $U \subset X$ is μ^*-measurable, i.e.

$$\mu^*(E) = \mu^*(E \cap U) + \mu^*(E \backslash U) \quad \text{for all } E \in 2^X. \tag{2.3}$$

Proof. By definition of μ, we have $\mu(\emptyset) = 0$, and $\mu(U) \leq \mu(V)$ for any open sets $U \subset V$. Hence $\mu^*(E) \leq \mu^*(F)$ for all $E \subset F \subset X$, and $\mu^*(U) = \mu(U)$ for all open U. We then show that for a sequence of open sets $(U_n)_{n=1}^\infty$ and $U = \bigcup_{n=1}^\infty U_n$, it holds $\mu(U) \leq \sum_{n=1}^\infty \mu(U_n)$. For any $f \in C_c(U, [0,1])$, let $K = \operatorname{supp}(f)$. By compactness of K, we have $K \subset \bigcup_{j=1}^n \mu(U_j)$ for some finite $n \in \mathbb{N}$. By Proposition 2.7, there exists a family of functions $g_j \in C_c(U_j, [0,1])$ such that $\sum_{j=1}^n g_n = 1$ on K. Then $f = \sum_{j=1}^n fg_j$, and

$$Tf = \sum_{j=1}^n T(fg_j) \leq \sum_{j=1}^n \mu(U_j) \leq \sum_{n=1}^\infty \mu(U_n).$$

By taking the supremum over $f \in C_c(U, [0,1])$, we have $\mu(U) \leq \sum_{n=1}^\infty \mu(U_n)$. More generally, if $(E_n)_{n=1}^\infty$ is a sequence of subsets of X and $E = \bigcup_{n=1}^\infty E_n$, we take an open set $U_n \supset E_n$ for each E_n and get

$$\sum_{n=1}^\infty \mu(U_n) \geq \mu\left(\bigcup_{n=1}^\infty U_n\right) \geq \mu(E).$$

By taking the infimum over $(U_n)_{n=1}^\infty$, we have $\sum_{n=1}^\infty \mu(E_n) \geq \mu(E)$. Hence μ^* is an outer measure on X.

Now we verify the condition 2.3. We first assume that E is open, so that $E \cap U$ is open. For any $\epsilon > 0$, we can find $f \in C_c(E \cap U, [0,1])$ such that $Tf > \mu(E \cap U) - \epsilon$. Similarly, we can find $g \in C_c(E \backslash \operatorname{supp}(f), [0,1])$ such that $Tg > \mu(E \backslash \operatorname{supp}(f)) - \epsilon$. Then $f + g \in C_c(E, [0,1])$, and

$$\mu(E) \geq Tf + Tg \geq \mu(E \cap U) + \mu(E \backslash \operatorname{supp}(f)) - 2\epsilon \geq \mu^*(E \cap U) + \mu^*(E \backslash U) - 2\epsilon.$$

Letting $\epsilon \to 0$, we obtain the desired inequality. For the general case $E \in 2^X$, we may assume $\mu^*(E) < \infty$ and find an open $V \supset E$ such that $\mu^*(V) < \mu^*(E) + \epsilon$, and hence

$$\mu^*(E) + \epsilon > \mu^*(V) \geq \mu^*(V \cap U) + \mu^*(V \backslash U) \geq \mu^*(E \cap U) + \mu^*(E \backslash U).$$

Letting $\epsilon \to 0$, we are done. $\square$

Remark. By Carathéodory's extension theorem, the family of μ^*-measurable sets is a σ-algebra on X, which contains the Borel σ-algebra $\mathscr{B}$. By taking the restriction $\mu = \mu^*|_{\mathscr{B}}$, we obtain a Borel measure on X.

Lemma 2.15. *The restriction $\mu = \mu^*|_{\mathscr{B}}$ of the outer measure μ^* in Lemma 2.14 on the Borel algebra $\mathscr{B}$ defines a Radon measure on X. Furthermore, for each compact set $K \subset X$,*

$$\mu(K) = \inf \{Tf : f \in C_c(X),\ f \geq \chi_K\}. \tag{2.4}$$

Proof. By (2.2), the Borel measure μ is outer regular on all Borel sets in X. If K is compact, $f \in C_c(X)$ and $f \geq \chi_K$, we define $U_\epsilon = \{x \in X : f(x) \geq 1 - \epsilon\}$, which is an open set. If $g \in C_c(U_\epsilon, [0,1])$, we have

$f-(1-\epsilon)g \geq 0$, and $Tf \geq (1-\epsilon)Tg$. Hence

$$\mu(K) \leq \mu(U_\epsilon) = \inf\{Tg : g \in C_c(U_\epsilon,[0,1])\} \leq \frac{Tf}{1-\epsilon}.$$

Letting $\epsilon \to 0$, we have $\mu(K) \leq Tf$, and hence $\mu(K) < \infty$. On the other hand, for any open $U \supset K$, by Urysohn's lemma, there exists $f \in C_c(U,[0,1])$ such that $f \geq \chi_K$, and we have $Tf \leq \mu(U)$ by definition of μ in Lemma 2.14. Since μ is outer regular, the result (2.4) follows.

To verify that μ is a Radon measure, it remains to show that it is inner regular on all open sets. If U is open and $\epsilon > 0$, we choose $f \in C_c(U,[0,1])$ such that $Tf > \mu(U) - \epsilon$ and let $K = \operatorname{supp}(f)$. If $g \in C_c(X)$ and $g \geq \chi_K$, we have $g - f \geq 0$ and $Tg \geq Tf > \mu(U) - \epsilon$. Then $\mu(K) > \mu(U) - \epsilon$, and μ is inner regular on U. □

Proof of Theorem 2.13. We start by establishing the uniqueness. Assume μ is a Radon measure such that $\int_X f\,d\mu = Tf$ for all $f \in C_c(X)$. If $U \subset X$ is open, we have $Tf = \int_X f\,d\mu \leq \mu(U)$ for all $f \in C_c(U,[0,1])$. On the other hand, if $K \subset U$ is a compact set, we take $f \in C_c(U,[0,1])$ such that $f = 1$ on K by Urysohn's lemma, so that $\mu(K) \leq \int_X f\,d\mu = Tf$. Since μ is inner regular on U, we have

$$\mu(U) = \sup\left\{Tf : f \in C_c(U,[0,1])\right\}.$$

Thus μ is determined by T on all open sets, hence on all Borel sets by outer regularity.

To prove the existence, we take the Radon measure constructed in Lemmata 2.14 and 2.15. It remains to show that $Tf = \int_X f\,d\mu$ for all $f \in C_c(X)$. We may assume $0 \leq f \leq 1$, since f is a linear combination of functions in $C_c(X,[0,1])$. Fix $N \in \mathbb{N}$. We define $K_j = \left\{x \in X : f(x) \geq \frac{j}{N}\right\}$ for each $j = 1,2,\cdots,N$ and $K_0 = \operatorname{supp}(f)$. Also, we divide f by $f = \sum_{j=1}^N f_j$, where $f_1,\cdots,f_N \in C_c(X)$ are defined as the truncation of f on the interval $\left[\frac{j-1}{N},\frac{j}{N}\right]$:

$$f_j = \min\left\{\max\left\{f - \frac{j-1}{N}, 0\right\}, \frac{1}{N}\right\}.$$

Then $N^{-1}\chi_{K_j} \leq f_j \leq N^{-1}\chi_{j-1}$, and

$$\frac{\mu(K_j)}{N} \leq \int_X f_j\,d\mu \leq \frac{\mu(K_{j-1})}{N}.$$

If $U \supset K_{j-1}$ is an open set, we have $Nf_j \in C_c(U,[0,1])$, and $Tf_j \leq \frac{\mu(U)}{N}$. Hence by (2.4) and outer regularity,

$$\frac{\mu(K_j)}{N} \overset{(2.4)}{\leq} Tf_j \leq \frac{1}{N}\inf\left\{\mu(U) : U \supset K_{j-1},\ U \text{ is open}\right\} = \frac{\mu(K_{j-1})}{N}.$$

Using $f = \sum_{j=1}^N f_j$, we have

$$\frac{1}{N}\sum_{j=1}^N \mu(K_j) \leq \int_X f\,d\mu \leq \frac{1}{N}\sum_{j=0}^{N-1}\mu(K_j), \quad \text{and} \quad \frac{1}{N}\sum_{j=1}^N \mu(K_j) \leq Tf \leq \frac{1}{N}\sum_{j=0}^{N-1}\mu(K_j).$$

Hence

$$\left|Tf - \int_X f\,d\mu\right| \leq \frac{\mu(K_0) - \mu(K_N)}{N} \leq \frac{\mu(\operatorname{supp}(f))}{N}.$$

Since $\mu(\operatorname{supp}(f)) < \infty$, we let $N \to \infty$ and conclude that $Tf = \int_X f\,d\mu$. □

Remark. For every Radon measure μ on X, define $T_\mu f = \int_X f\,d\mu$ for $f \in C_c(X)$. Then $\mu \mapsto T_\mu$ is a one-to-one map from the positive linear functionals on $C_c(X)$ to the Radon measures on X.

2.3 Regularity of Radon Measures

In this section we discuss more properties of Radon measures.

Proposition 2.16. *Every Radon measure is inner regular on all of its σ-finite sets.*

Proof. Let μ be a Radon measure on X and $E \subset X$ a σ-finite set. If $\mu(E) < \infty$, for any $\epsilon > 0$, we take an open set $U \supset E$ with $\mu(U) < \mu(E) + \epsilon$ and a compact set $F \subset U$ such that $\mu(F) > \mu(U) - \epsilon$. Since $\mu(U \backslash E) < \epsilon$, we can also take an open set $V \supset U \backslash E$ such that $\mu(V) < \epsilon$. Let $K = F \backslash V$, which is compact. Then $K \subset U \backslash V \subset E$, and

$$\mu(K) = \mu(F) - \mu(F \cap V) > \mu(U) - \epsilon - \mu(V) > \mu(E) - 2\epsilon.$$

Hence μ is inner regular on E. On the other hand, if $\mu(E) = \infty$, E is the limit of an increasing sequence $(E_n)_{n=1}^{\infty}$ of μ-finite sets such that $\mu(E_n) \to \infty$. Hence for any $N > 0$ there exists $n \in \mathbb{N}$ such that $\mu(E_n) > N$. By the preceding argument, one can take a compact $K \subset E_n$ with $\mu(K) > N$. Hence the supremum of $\mu(K)$ over compact $K \subset E$ is ∞, and μ is inner regular on E. $\square$

We have some immediate corollaries of this proposition.

Corollary 2.17. *Every σ-finite Radon measure is regular. Particularly, if X is a σ-compact space, every Radon measure on X is regular.*

Proposition 2.18. *Let μ be a σ-finite Radon measure on X and E a Borel set in X.*

(i) For every $\epsilon > 0$, there exists an open U and a closed F with $F \subset E \subset U$ and $\mu(U \backslash F) < \epsilon$.

(ii) There exists an F_σ set A and a G_δ set B such that $A \subset E \subset B$ and $\mu(B \backslash A) = 0$.

Proof. We write $E = \bigcup_{n=1}^{\infty} E_n$ where the E_j's are disjoint and have finite measure. For each E_n, choose an open $U_n \supset E_n$ with $\mu(U_n) < \mu(E_n) + 2^{-1-n}\epsilon$ and let $U = \bigcup_{n=1}^{\infty} U_n$. Then U is an open set containing E and $\mu(U \backslash E) \leq \sum_{n=1}^{\infty} \mu(U_n \backslash E_n) < \epsilon/2$. Applying the same approach to E^c, we get an open $V \supset E^c$ with $\mu(V \backslash E^c) < \epsilon/2$. Let $F = V^c$. Then F is a closed set contained in E, and

$$\mu(U \backslash F) = \mu(U \backslash E) + \mu(E \backslash F) = \mu(U \backslash E) + \mu(V \backslash E^c) < \epsilon.$$

Now for each $k \in \mathbb{N}$, by the preceding argument, we choose an open U_k and a closed F_k with $F_k \subset E \subset U_k$ and $\mu(U_k \backslash F_k) < 1/k$. We may also assume $U_k \subset U_{k-1}$ by taking $U_k \cap U_{k-1}$ if necessary. Similarly we assume $F_k \supset F_{k-1}$. Let $B = \bigcap_{k=1}^{\infty} U_k$, which is a G_δ set, and $A = \bigcup_{k=1}^{\infty} F_k$, which is an F_σ set. Then

$$\mu(B \backslash A) = \mu\left(\bigcap_{k=1}^{\infty} (U_k \backslash F_k)\right) = \lim_{k \to \infty} \mu(U_k \backslash F_k) = 0,$$

and $A \subset E \subset B$, which concludes the proof. $\square$

The following theorem discusses the regularity of Borel measures in LCH spaces.

Theorem 2.19. *Let μ be a Borel measure on an LCH space X in which every open set is σ-compact (which is the case, for example, if X is second countable). If μ is finite on compact sets, it is regular.*

Proof. Since μ is finite on compact sets, we have $\int_X f \, d\mu < \infty$ for all $f \in C_c(X)$, and $T_\mu f = \int_X f \, d\mu$ defines a positive linear functional T_μ on $C_c(X)$. Let ν be the associated Radon measure according to Theorem 2.13. If $U \subset X$ is open, let $(K_n)_{n=1}^{\infty}$ be a sequence of compact sets increasing to U. We take $f_1 \in C_c(U, [0,1])$ such that $f = 1$ on K_1, and inductively take $f_n \in C_c(U, [0,1])$ such that $f = 1$ on $K_n \cup \operatorname{supp}(f_{n-1})$. Then $f_n \uparrow \chi_U$ pointwise, and by monotone convergence theorem,

$$\mu(U) = \lim_{n \to \infty} \int_X f_n \, d\mu = \lim_{n \to \infty} \int_X f_n \, d\nu = \nu(U).$$

Next, if E is any Borel set and $\epsilon > 0$, by Proposition 2.18, there exists open $U \supset E$ and closed $V \subset E$ with $\nu(U\backslash F) < \epsilon$. Since $U\backslash F$ is open, $\mu(U\backslash F) = \nu(U\backslash F) < \epsilon$. In particular, $\mu(U) < \mu(E) + \epsilon$, and μ is outer regular. Also, we have $\mu(F) > \mu(E) - \epsilon$. Since X is σ-compact, there exist compact sets $K_n \subset F$ with $\mu(K_n) \to \mu(F)$, and μ is inner regular. Therefore μ is regular on X. $\square$

Remark. By the uniqueness part of Theorem 2.13, since μ is Radon, we have $\mu = \nu$.

Proposition 2.20. *If μ a Radon measure on an LCH space X, $C_c(X)$ is dense in $L^p(\mu)$ for $1 \leq p < \infty$.*

Proof. Since the simple functions are dense in $L^p(\mu)$, it suffices to approximate each simple function χ_E in L^p-norm, where $E \subset X$ is a Borel set with $\mu(E) < \infty$. For any $\epsilon > 0$, we pick an open set U and a compact set K such that $K \subset E \subset U$ and $\mu(U\backslash K) < \epsilon$. By Urysohn's lemma, there exists $f \in C_c(X)$ such that $\chi_K \leq f \leq \chi_U$. Then $\|\chi_E - f\|_p^p \leq \mu(U\backslash K) \leq \epsilon$, and we are done. $\square$

Theorem 2.21 (Lusin). *Let μ be a Radon measure on an LCH space X, and $f : X \to \mathbb{C}$ a measurable function that vanishes outside a μ-finite set. Then for any $\epsilon > 0$, there exists $\phi \in C_c(X)$ such that $\mu(\{\phi \neq f\}) < \epsilon$. Moreover, if f is bounded, we may take $\|\phi\|_\infty \leq \|f\|_\infty$.*

Proof. We assume first that f is bounded, so $f \in L^1(\mu)$. Let $E = \{x \in X : f(x) \neq 0\}$. By Proposition 2.20, there exists a sequence (g_n) in C_c that converges to f in L^1. We take a subsequence that converges to f a.e. and still denote it by (g_n) for simplicity. By Egoroff's theorem, there exists $A \subset E$ with $\mu(E\backslash A) < \epsilon/3$ and $g_n \to f$ uniformly on A, and there exists a compact $B \subset A$ and an open $U \supset E$ such that $\mu(A\backslash B) < \epsilon/3$ and $\mu(U\backslash E) < \epsilon/3$. Since $g_n \to f$ uniformly on B, $f|_B$ is continuous, and by Tietze extension theorem, there exists $\phi \in C_c(U)$ such that $\phi = f$ on B. Since $\{\phi \neq f\} \subset U\backslash B$ and $\mu(U\backslash B) < \epsilon$, we have $\mu(\{\phi \neq f\}) < \epsilon$. Furthermore, if $|\phi(x)| > \|f\|_\infty$, we may truncate $\phi(x)$ to $\|f\|_\infty \frac{\phi(x)}{|\phi(x)|}$, which does not change $\phi|_B$ and does not impact the continuity of ϕ. Therefore we mat take $\|\phi\|_\infty < \|f\|_\infty$.

On the other hand, if f is unbounded, we make $A_n = \{0 \leq |f| \leq n\}$, which increases to $E = \{f \neq 0\}$ as $n \to \infty$. Then there exists sufficient large n such that $\mu(E\backslash A_n) < \epsilon/2$. By the preceding argument, there exists $\phi \in C_c(X)$ such that $\phi = f\chi_{A_n}$ except on a set of measure less than $\epsilon/2$. Hence $\mu(\{\phi \neq f\}) < \epsilon$. $\square$

Finally we discuss how to construct a Radon measure from another one.

Proposition 2.22. *Let μ be a Radon measure on a topological space X. If $\phi \in L^1(\mu)$ and $\phi \geq 0$, we define*

$$\nu(E) = \int_E \phi \, d\mu, \quad E \in \mathscr{B}.$$

Then ν is also a Radon measure on X.

Proof. One can easily verify that ν is a Borel measure on X, and $\nu \ll \mu$. Then for each $\epsilon > 0$, there exists $\delta > 0$ such that $\nu(E) < \epsilon$ for all $\mu(E) < \delta$. Now we verify that ν is a Radon measure on X.

- If $K \subset X$ is a compact set, $\nu(K) = \int_K \phi \, d\mu \leq \int_X \phi \, d\mu < \infty$.
- For any Borel set $E \subset X$ and any $\epsilon > 0$, there exists an open $U \supset E$ such that $\mu(U\backslash E) < \delta$, and $\nu(U\backslash E) < \epsilon$. Hence ν is outer regular on E.
- For any open set $U \subset X$ and any $\epsilon > 0$, there exists a compact $K \subset U$ such that $\mu(U\backslash K) < \delta$, and $\nu(U\backslash K) < \epsilon$. Hence ν is inner regular on U.

To summarize, ν is a Radon measure on X. $\square$

2.4 Riesz-Markov-Kakutani Representation of $C_0(X)^*$

Positive bounded linear functionals on $C_0(X)$. Let X be an LCH space. Proposition 2.6 states that $C_0(X)$ is the uniform closure of $C_c(X)$. If μ is a Radon measure on X, the functional $T_\mu f = \int_X f\,d\mu$ extends continuously to $C_0(X)$ if and only if it is bounded with respect to the uniform norm $\|\cdot\|_\infty$, i.e. there exists a constant $\gamma > 0$ such that $|T_\mu f| \le \gamma \|f\|_\infty$ for all $f \in C_c(X)$. In view of the equality

$$\mu(X) = \sup\left\{\int_X f\,d\mu : f \in C_c(X),\ 0 \le f \le 1\right\} = \sup\left\{T_\mu f : f \in C_c(X),\ 0 \le f \le 1\right\},$$

we know that $T_\mu : C_c(X) \to \mathbb{C}$ is bounded with respect to $\|\cdot\|_\infty$ if and only if $\mu(X) < \infty$, in which case $\mu(X)$ is the operator norm of T_μ. Therefore, we have identified the positive bounded linear functionals on $C_0(X)$, which are given by integration against finite Radon measures.

In this section, we identify the dual space of $C_0(X)$, denoted by $C_0(X)^*$, which consists of all bounded linear functionals on $C_0(X)$.

Definition 2.23 (Signed Radon measures and complex Radon measures)**.** Let X be a topological space.

(i) A *signed Radon measure on* X is a signed Borel measure on X whose positive and negative variations are Radon measures.

(ii) A *complex Radon measure on* X is a complex Borel measure on X whose real and imaginary parts are signed Radon measures. We denote the space of complex Radon measures on X by $M(X)$, and define $\|\mu\| = |\mu|(X)$, where $|\mu|$ is the total variation of μ.

Remark. Since a complex measure is always finite, every complex Radon measure is regular. Furthermore, every complex Borel measure is Radon in an LCH space in which every open set is σ-compact (for example, a second countable LCH space).

Theorem 2.24. *If μ is a complex Borel measure on X, then μ is Radon if and only in $|\mu|$ is Radon. Furthermore, $M(X)$ is a vector space and $\mu \to \|\mu\|$ is a norm on it.*

Proof. By Proposition 2.16, we note that a finite positive Borel measure μ is Radon if and only if for every Borel set E and every $\epsilon > 0$, there exists compact $K \subset E$ and open $U \supset E$ such that $\mu(U\backslash K) < \epsilon$.

If $\mu = (\mu_1 - \mu_2) + i(\mu_3 - \mu_4)$ and $|\mu|(U\backslash K) < \epsilon$, we have $\mu_j(U\backslash K) < \epsilon$ for $j = 1, 2, 3, 4$. Conversely, if $\mu_j(U_j\backslash K_j) < \epsilon/4$ for all j, we have $|\mu|(U\backslash K) < \epsilon$ for $U = \bigcap_{j=1}^4 U_j$ and $K_j = \bigcup_{j=1}^4 K_j$. Hence μ is Radon if and only if its total variation $|\mu|$ is Radon.

For the second assertion, a similar argument shows that $M(X)$ is closed under addition and scalar multiplication. Finally, to show $\mu \to \|\mu\|$ is a norm on X, let $\mu_1, \mu_2 \in M(X)$ and $\nu = |\mu_1 + \mu_2|$, and take the Radon Nikodym derivative $f_1 = d\mu_1/d\nu$ and $f_2 = \mu_2/d\nu$. Then

$$|\mu + \nu|(X) \le \int_X |f_1 + f_2| d\nu \le \int_X |f_1| d\nu + \int_X |f_2| d\nu \le |\mu_1|(X) + |\mu_2|(X).$$

Hence the triangle inequality holds, and $\|\mu\| = |\mu|(X)$ is a norm. □

We now discuss how to identify each $T \in C(X)^*$ with a complex Radon measure on X.

Theorem 2.25 (Riesz-Markov-Kakutani)**.** *Let X be an LCH space. For each $\mu \in M(X)$, define*

$$T_\mu f = \int_X f\,d\mu, \quad f \in C_0(X).$$

Then the map $\mu \mapsto T_\mu$ defines an isometric isomorphism of $M(X)$ onto the dual space $C_0(X)^*$.

We begin from the real case. While studying a possibly non-positive linear functional on $C_0(X, \mathbb{R})$, the following decomposition is extremely useful.

Theorem 2.26 (Jordan decomposition)**.** *If $T \in C_0(X,\mathbb{R})^*$, there exists positive bounded linear functionals $T^\pm \in C_0(X,\mathbb{R})^*$ such that $T = T^+ - T^-$.*

Proof. For $f \in C_0(X,\mathbb{R})$ with $f \geq 0$, we define

$$T^+ f = \sup\{Tg : g \in C_0(X,\mathbb{R}),\ 0 \leq g \leq f\}$$

We claim that T^+ is the restriction to $C_0(X,[0,\infty))$ of a positive bounded linear functional on $C_0(X,\mathbb{R})$.

- For $\lambda \geq 0$, we have

$$T(\lambda f) = \sup\{Th : h \in C_0(X,\mathbb{R}),\ 0 \leq h \leq \lambda f\} = \sup\{\lambda Tg : g \in C_0(X,\mathbb{R}),\ 0 \leq g \leq f\} = \lambda Tf.$$

- If $0 \leq g_1 \leq f_1$ and $0 \leq g_2 \leq f_2$, we have $0 \leq g_1 + g_2 \leq f_1 + f_2$, so that $T^+(f_1+f_2) \geq Tg_1 + Tg_2$, and hence $T^+(f_1+f_2) \geq T^+f_1 + T^+f_2$. On the other hand, if $0 \leq g \leq f_1 + f_2$, let $g_1 = \min\{f_1, g\}$ and $g_2 = g - g_1 = \max\{0, g - f_1\}$, so that $0 \leq g_1 \leq f_1$ and $0 \leq g_2 \leq f_2$. Then

$$Tg = Tg_1 + Tg_2 \leq T^+f_1 + T^+f_2,$$

  and $T^+(f_1+f_2) \leq T^+f_1 + T^+f_2$. Therefore $T^+(f_1+f_2) = T^+f_1 + T^+f_2$.
- Since $|Tg| \leq \|T\|\,\|g\|_\infty \leq \|T\|\,\|f\|_\infty$ for $0 \leq g \leq f$ and $T0 = 0$, we have $0 \leq T^+f \leq \|T\|\,\|f\|_\infty$

Now for any $f \in C_0(X,\mathbb{R})$, both its positive $f^+ = \max\{f,0\}$ and negative parts $f^- = \max\{-f,0\}$ are in $C_0(X,[0,\infty))$, and we define $T^+f = T^+f^+ - T^+f^-$. If $f = g - h$, where $g,h \geq 0$, we have $f^+ + h = g + f^-$, and $Tf = Tf^+ - Tf^- = Tg - Th$. It follows easily that T^+ is a linear functional in $C_0(X,\mathbb{R})$, and

$$|T^+f| \leq \max\left\{T^+f^+, T^+f^-\right\} \leq \|T\| \max\left\{\|f^+\|_\infty, \|f^-\|_\infty\right\} = \|T\|\,\|f\|_\infty.$$

Hence T^+ is bounded, and $\|T^+\| \leq \|T\|$.

Finally, we define $T^- = T^+ - T \in C_0(X,\mathbb{R})^*$. By definition of T^+, we have $T^+f \geq Tf$ for $f \in C_0(X,\mathbb{R})$ with $f \geq 0$, hence T^- is a positive linear functional. Thus we concludes the proof. ☐

Remark. For any $T \in C_0(X)^*$, consider its restriction $T_R = U + iV$ to $C_0(X,\mathbb{R})$, where $U, V \in C_0(X,\mathbb{R})^*$. If $f = u + iv \in C_0(X)$, where $u, v \in C_0(X,\mathbb{R})$, by $\mathbb{C}$-linearity,

$$Tf = Tu + iTv = T_Ru + iT_Rv = (U+iV)u + i(U+iV)v = (Uu - Vv) + i(Uv + Vu).$$

It is seen T is uniquely determined by T_R. We then decompose $U = U^+ - U^-$ and $V = V^+ - V^-$, where $U^\pm, V^\pm \in C_0(X,\mathbb{R})^*$ are positive. By Riesz representation theorem, we can find finite positive Radon measures $\mu_R^\pm$ and $\mu_I^\pm$ associated with $U^\pm$ and $V^\pm$, respectively. We define the complex Radon measure

$$\mu = \left(\mu_R^+ - \mu_R^-\right) + i\left(\mu_I^+ - \mu_I^-\right).$$

Then

$$\begin{aligned}
\int_X f\,d\mu &= \left(\int_X f\,d\mu_R^+ - \int_X f\,d\mu_R^-\right) + i\left(\int_X f\,d\mu_I^+ - \int_X f\,d\mu_I^-\right) \\
&= \left(\int_X u\,d\mu_R^+ - \int_X u\,d\mu_R^- - \int_X v\,d\mu_I^+ + \int_X v\,d\mu_I^-\right) + i\left(\int_X v\,d\mu_R^+ - \int_X v\,d\mu_R^- + \int_X u\,d\mu_I^+ - \int_X u\,d\mu_I^-\right) \\
&= \left(U^+u - U^-u - V^+v + V^-v\right) + i\left(U^+v - U^-v + V^+u - V^-u\right) \\
&= (Uu - Vv) + i(Uv + Vu) = Tf.
\end{aligned}$$

Therefore, every $T \in C_0(X)^*$ is associated with a complex Radon measure $\mu \in M(X)$ such that $Tf = \int_X f\,d\mu$. Furthermore, since $\mu_R^+, \mu_R^-, \mu_I^+, \mu_I^-$ are unique determined by T, the complex Radon measure μ is unique.

Proof of Theorem 2.25. We have already shown that every $T \in C_0(X)^*$ is of the form T_μ. On the other hand, if $\mu \in M(X)$, we have

$$\left| \int_X f\, d\mu \right| \leq \int_X |f|\, d|\mu| \leq \|f\|_\infty \|\mu\|, \quad f \in C_0(X).$$

Hence $T_\mu \in C_0(X)^*$, and $\|T_\mu\| \leq \|\mu\|$. Furthermore, we take $h = d\mu/d|\mu|$, so that $|h| = 1$ $|\mu|$-a.e.. By Lusin's theorem [Theorem 2.21], for each $\epsilon > 0$, there exists $\phi \in C_c(X)$ such that $\|\phi\|_\infty = 1$ and $\phi = \overline{h}$ except on a set E with $|\mu|(E) < \epsilon/2$. Then

$$\begin{aligned} \|\mu\| &= \int_X |h|^2\, d|\mu| = \int_X \overline{h}\, d\mu \leq \left| \int_X \phi\, d\mu \right| + \left| \int_X (\phi - \overline{h})\, d\mu \right| \\ &= \left| \int_X \phi\, d\mu \right| + \left| \int_E (\phi - \overline{h})\, d\mu \right| \leq \|T_\mu\|\, \|\phi\|_\infty + \|\phi - \overline{h}\|_\infty |\mu|(E) \leq \|T_\mu\| + \epsilon. \end{aligned}$$

Letting $\epsilon \to 0$, we have $\|\mu\| \leq \|T_\mu\|$. Hence $\|\mu\| = \|T_\mu\|$, and the proof is complete. □

Remark. If we consider the real case, the mapping $\mu \mapsto T_\mu$ is an isometric isomorphism from the space of finite signed Radon measures to $C_0(X, \mathbb{R})^*$.

Corollary 2.27. *Let X be a compact Hausdorff space, $C(X)^*$ is isometrically isomorphic to the space $M(X)$ of complex Radon measures on X.*

Remark. If in addition, X is metrizable, then X is second countable, and we know that every finite Borel measure on X is Radon by Theorem 2.19. Since complex measures are always finite, $M(X)$ is indeed the space of complex Borel measures on X, and $C(X)^* \simeq M(X)$.

Corollary 2.28. *Let μ be a Radon measure on an LCH space X. For each $f \in L^1(\mu)$, define*

$$\nu_f(E) = \int_E f\, d\mu, \quad E \in \mathscr{B}.$$

The mapping $f \mapsto \nu_f$ is an isometric embedding of $L^1(\mu)$ into $M(X)$ whose range consists precisely of those $\nu \in M(X)$ such that $\nu \ll \mu$.

Proof. By Proposition 2.22, the complex measure ν_f on X is Radon and satisfies $\nu_f \ll \mu$. Moreover,

$$\|\nu_f\| = |\nu|(X) = \int_X |f|\, d\mu = \|f\|_{L^1}.$$

Finally, if $\nu \in M(X)$ and $\nu \ll \mu$, taking f to be the Radon-Nikodym derivative $d\nu/d\mu$ yields $\nu_f = \nu$. □

3 Ergodic Theory

Setting. Let $(X, \mathscr{F}, \mu)$ be a σ-finite measure space. We consider a mapping $T : X \to X$ such that

(i) T is *measurable*, i.e. $T^{-1}(E) \in \mathscr{F}$ for each $E \in \mathscr{F}$;

(ii) T is *measure-preserving*, i.e. $\mu(T^{-1}(E)) = \mu(E)$ for each $E \in \mathscr{F}$.

(iii) We call the quadruple $(X, \mathscr{F}, \mu, T)$ a *measure-preserving system.*

If in addition for such a transformation T we have that T is a bijection and T^{-1} is also a measure-preserving transformation, then T is called a *measure-preserving isomorphism.*

If $f : X \to \mathbb{C}$ is a measurable function and T is a measure-preserving transformation, the composition $f \circ T$ is measurable. Furthermore, if f is integrable, so is $f \circ T$, and

$$\int_X f \, d\mu = \int_X f \circ T \, d\mu.$$

The setting described above is of interest, in part, because it abstracts the idea of a dynamical system, one whose totality of states is represented by the space X, with each point $x \in X$ giving a particular state of the system. The mapping $T : X \to X$ describes the transformation of the system after a unit of time has elapsed. The iterates, $T^n = T \circ T \circ \cdots \circ T$ (n times) describe the evolution of the system after n units of time. In many scenarios, we are interested in the average behavior of the system as the time $n \to \infty$. To be specific, given a measurable function f on $(X, \mathscr{F}, \mu)$, we aim to study the *ergodic averages*

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

and their limit as $n \to \infty$.

3.1 The Mean Ergodic Theorem

We first discuss a general ergodic result for Banach spaces.

Theorem 3.1 (Mean ergodic theorem)**.** *Let $T : X \to X$ be a bounded linear operator on a Banach space X, and assume that $\sup_{n\in\mathbb{N}} \|T^n\| < \infty$. For $n \in \mathbb{N}$, define the ergodic average*

$$A_n = \frac{1}{n} \sum_{k=0}^{n-1} T^k.$$

(i) If $x \in X$, the sequence $(A_n x)_{n=1}^{\infty}$ converges if and only if it has a weakly convergent subsequence;

(ii) The set

$$L = \{x \in X : \textit{the sequence } (A_n x)_{n=1}^{\infty} \textit{ converges}\}$$

is a closed T-invariant subspace of X, and $L = \ker(\mathrm{Id} - T) \oplus \overline{\mathfrak{R}(\mathrm{Id} - T)}$.

(iii) If X reflexive, then $L = X$.

(iv) Define the operator $A : L \to L$ by $A(x_0 + x_1) = x_0$ for $x_0 \in \ker(\mathrm{Id} - T)$ and $x_1 \in \overline{\mathfrak{R}(\mathrm{Id} - T)}$. Then

$$\lim_{n\to\infty} A_n x = Ax$$

for all $x \in L$, and A satisfies

$$AT = TA = A^2 = A, \quad \textit{and} \quad \|A\| \leq \sup_{n\in\mathbb{N}} \|T^n\|.$$

The proof of this theorem requires some lemmata.

Lemma 3.2. *Assume $c = \sup_{n\in\mathbb{N}} \|T^n\| < \infty$.*

(i) For each $n \in \mathbb{N}$, $\|A_n\| \le c$ and $\|A_n(\mathrm{Id} - T)\| \le \frac{1+c}{n}$.

(ii) If $x \in \ker(\mathrm{Id} - T)$, then for each $n \in \mathbb{N}$, we have $A_n x = x$ and $\|x\| \le c\|x + (\mathrm{Id} - T)\xi\|$ for all $\xi \in X$.

(iii) If $x \in \ker(\mathrm{Id} - T)$ and $y \in \overline{\mathfrak{R}(\mathrm{Id} - T)}$, then $\|x\| \le c\|x + y\|$.

(iv) $\ker(\mathrm{Id} - T) \cap \overline{\mathfrak{R}(\mathrm{Id} - T)} = 0$, and $L := \ker(\mathrm{Id} - T) \oplus \overline{\mathfrak{R}(\mathrm{Id} - T)}$ is a closed subspace of X.

(v) $T(L) \subset L$.

(vi) If $y \in \overline{\mathfrak{R}(\mathrm{Id} - T)}$, then $\lim_{n\to\infty} A_n y = 0$.

Proof. (i) Since $A_n = \frac{1}{n}(\mathrm{Id} + T + T^2 + \cdots + T^n)$, we have

$$\|A_n\| \le \frac{1}{n}\sum_{k=0}^{n-1} \|T^k\| \le \sup_{n\in\mathbb{N}} \|T^n\| = c, \quad \text{and} \quad \|A_n(\mathrm{Id} - T)\| = \frac{1}{n}\|\mathrm{Id} - T^n\| \le \frac{1 + \|T^n\|}{n} \le \frac{1+c}{n}.$$

(ii) If $x \in \ker(\mathrm{Id} - T)$, we have $Tx = x$ and by induction $T^n x = x$ for all $n \in \mathbb{N}$ and hence $A_n x = x$. Moreover, by (i) we have $A_n(\mathrm{Id} - T)\xi \to 0$ as $n \to \infty$ for all $\xi \in X$, and

$$\|x\| = \lim_{n\to\infty} \|x + A_n(\mathrm{Id} - T)\xi\| = \lim_{n\to\infty} \|A_n(x + (\mathrm{Id} - T)\xi)\| \le c\|x + (\mathrm{Id} - T)\xi\|.$$

(iii) If $y \in \overline{\mathfrak{R}(\mathrm{Id} - T)}$, there exists a sequence $\xi_n \in X$ such that $(\mathrm{Id} - T)\xi_n \to y$. We take $\xi = \xi_n$ in (ii) and take the limit $n \to \infty$ to obtain $\|x\| \le c\|x + y\|$.

(iv) We let $x \in \ker(\mathrm{Id} - T) \cap \overline{\mathfrak{R}(\mathrm{Id} - T)}$. Then $-x \in \overline{\mathfrak{R}(\mathrm{Id} - T)}$, and by (iii) we have $\|x\| \le c\|x + (-x)\| = 0$. Next we show that $\ker(\mathrm{Id} - T) \oplus \overline{\mathfrak{R}(\mathrm{Id} - T)}$ is closed. Let $x_n \in \ker(\mathrm{Id} - T)$ and $y_n \in \overline{\mathfrak{R}(\mathrm{Id} - T)}$ be sequences whose sum $z_n = y_n + z_n$ converges to some element $z \in X$. Then (z_n) is a Cauchy sequence in X, and by (iii) the sequence (x_n) is also Cauchy, and hence $y_n = z_n - x_n$ is also Cauchy. Since $\ker(\mathrm{Id} - T)$ and $\overline{\mathfrak{R}(\mathrm{Id} - T)}$ are closed subspaces of X, the Cauchy sequences (x_n) and (y_n) converge to $x \in \ker(\mathrm{Id} - T)$ and $y \in \overline{\mathfrak{R}(\mathrm{Id} - T)}$, respectively, and $z = x + y \in \ker(\mathrm{Id} - T) \oplus \overline{\mathfrak{R}(\mathrm{Id} - T)}$.

(v) We take $x \in \ker(\mathrm{Id} - T)$ and $y \in \overline{\mathfrak{R}(\mathrm{Id} - T)}$, and take a sequence $\xi_n \in X$ such that $(\mathrm{Id} - T)\xi_n \to y$. Then

$$T(x + y) = x + Ty = x + \lim_{n\to\infty} T(\mathrm{Id} - T)\xi_n = x + \lim_{n\to\infty} (\mathrm{Id} - T)(T\xi_n) \in \ker(\mathrm{Id} - T) \oplus \overline{\mathfrak{R}(\mathrm{Id} - T)}.$$

(vi) For any $\epsilon > 0$, we take $\xi \in X$ such that $c\|y - (\mathrm{Id} - T)\xi\| < \frac{\epsilon}{3}$. By (i), we have $\|A_n(\mathrm{Id} - T)\xi\| \le \frac{1+c}{n}\|\xi\|$, which tends to 0 as $n \to \infty$. Then there exists N such that $\|(A_n - A_m)(\mathrm{Id} - T)\xi\| \le \frac{\epsilon}{3}$ for all $n, m \ge N$, and

$$\begin{aligned}
\|A_n y - A_m y\| &\le \|A_n y - A_n(\mathrm{Id} - T)\xi\| + \|(A_n - A_m)(\mathrm{Id} - T)\xi\| + \|A_m(\mathrm{Id} - T)\xi - A_m y\| \\
&\le \|A_n\| \, \|y - (\mathrm{Id} - T)\xi\| + \|(A_n - A_m)(\mathrm{Id} - T)\xi\| + \|A_m\| \, \|(\mathrm{Id} - T)\xi - y\| \\
&\le 2c\|y - (\mathrm{Id} - T)\xi\| + \frac{\epsilon}{3} \le \epsilon.
\end{aligned}$$

Hence $(A_n y)$ is a Cauchy sequence, and

$$\lim_{n\to\infty} \|A_n y\| = \lim_{n\to\infty} \|A_n(y - (\mathrm{Id} - T)\xi)\| + \lim_{n\to\infty} \|A_n(\mathrm{Id} - T)\xi\| < \frac{\epsilon}{3},$$

which implies $A_n y \to 0$ as $n \to \infty$. □

Lemma 3.3. *Let $x, x_0 \in X$. The following are equivalent:*

(a) $x_0 \in \ker(\mathrm{Id} - T)$ and $x - x_0 \in \overline{\mathfrak{R}(\mathrm{Id} - T)}$.

(b) $\lim_{n\to\infty} \|A_n x - x_0\| = 0$.

(c) There exists a subsequence n_k such that for all $f \in X^*$,

$$\lim_{k\to\infty} f(A_{n_k} x) = f(x_0).$$

Proof. The Lemma 3.2 (vi) implies (a) $\Rightarrow$ (b), and obviously (a) $\Rightarrow$ (c). Then it remains to prove (c) $\Rightarrow$ (a). If (c) holds, we take $f \in X^*$. Then $T^* f = f \circ T : X \to \mathbb{C}$ is a bounded linear functional, and

$$f(x_0 - Tx_0) = (f - T^* f)(x_0) = \lim_{k\to\infty} (f - T^* f)(A_{n_k} x) = \lim_{k\to\infty} f((\mathrm{Id} - T)A_{n_k} x) = 0,$$

where the last equality follows from Lemma 3.2 (i), and we have $Tx_0 = x_0$ by Hahn-Banach theorem.

Now we assume that $x - x_0 \in \overline{\mathfrak{R}(\mathrm{Id} - T)}$. By Hahn-Banach theorem, there exists $f \in X^*$ such that $f(x - x_0) = 1$ and $f(\xi - T\xi) = 0$ for all $\xi \in X$. This implies that $f(T^{k+1}\xi - T^k\xi) = 0$ for all $\xi \in X$ and all $k \in \mathbb{N}_0$. By induction, we have $f(T^k \xi) = f(\xi)$. Hence

$$f(A_n x) = \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x_0) = f(x_0)$$

for all $n \in \mathbb{N}_0$. According to (c), we have $f(x) = f(x_0)$, and $f(x - x_0) = 0$, which is a contradiction. Therefore $x - x_0 \in \overline{\mathfrak{R}(\mathrm{Id} - T)}$, and we complete the proof. $\square$

Now we prove the main theorem.

Proof of Theorem 3.1. By Lemma 3.3, the sequence $(A_n x)_{n=1}^\infty$ converges in norm if and only if it has a weakly convergent subsequence, if and only if $x \in L = \ker(\mathrm{Id} - T) \oplus \overline{\mathfrak{R}(\mathrm{Id} - T)}$. By Lemma 3.2 (iv) and (v), the subspace L is closed and T-invariant. Furthermore, since $\|A_n\| \le c$ for all $n \in \mathbb{N}$, for every $x \in X$, the sequence $(A_n x)$ is bounded. If X is reflexive, by Banach-Alaoglu theorem, every $(A_n x)$ has a weakly convergent subsequence $(A_{n_k} x)$, which implies $x \in L$, and hence $L = X$.

Finally we consider the operator A defined in (iv). Then $A^2 = A$ by definition. By Lemma 3.2 (iii), we have $\|A\| \le c$, and by Lemma 3.2 (vi), $\lim_{n\to\infty} A(x_0 + x_1) = Ax_0$. Since A commutes with $T|_L$, and A vanishes on the range of operator $\mathrm{Id} - T$, we have $TA = AT = A$. $\square$

Since Hilbert spaces are reflexive, we have the following mean ergodic theorem for Hilbert spaces.

Corollary 3.4 (Mean ergodic theorem)**.** *Let T be a bounded linear operator on the Hilbert space H such that $\sup_{n\in\mathbb{N}} \|T^n\| < \infty$, and let P_T be the projection operator onto the subspace*

$$\ker(\mathrm{Id} - T) = \{x \in H : Tx = x\}.$$

Then for every $x \in H$, the ergodic average

$$A_n x := \frac{1}{n} \sum_{k=0}^{n-1} T^k x \to P_T x \quad \textit{in norm as } n \to \infty.$$

In particular, we take the Hilbert space to be $L^2(X, \mathscr{F}, \mu)$. If T is a measure-preserving operator on X, we regard T as a linear operator on $L^2(X, \mathscr{F}, \mu)$ by writing $Tf = f \circ T$. Then T is an isometry on $L^2(X, \mathscr{F}, \mu)$, i.e. $\|Tf\|_{L^2} = \|f\|_{L^2}$ for all $f \in L^2(X, \mathscr{F}, \mu)$, and $\|T\| = 1$. Consequently, we have $\|T^n\| \le 1$ for all $n \in \mathbb{N}$, and we can apply the mean ergodic theorem on this system.

Corollary 3.5 (Mean ergodic theorem)**.** *Let $(X, \mathscr{F}, \mu, T)$ be a measure-preserving system, and let P_T be the projection operator onto the subspace*

$$G = \left\{g \in L^2(X, \mathscr{F}, \mu) : g \circ T = g\right\}$$

Then for every $f \in L^2(X, \mathscr{F}, \mu)$, the ergodic average

$$A_n f := \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \to P_T f \quad \textit{in } L^2(X, \mathscr{F}, \mu) \textit{ as } n \to \infty.$$

In finite measure spaces, the ergodic average $A_n f$ also converges in L^1. This conclusion follows from the convergence result in L^2 and the density of L^2 in L^1.

Corollary 3.6. *Let $(X, \mathscr{F}, \mu, T)$ be a measure-preserving system such that μ is finite. For each $f \in L^1(X, \mathscr{F}, \mu)$, the ergodic average $A_n f = \sum_{k=0}^{n-1} f \circ T^k$ converges in L^1 to a T-invariant function $\bar{f} \in L^1(X, \mathscr{F}, \mu)$.*

Proof. Since μ is finite, we know by Cauchy-Schwartz inequality that $L^2(X, \mathscr{F}, \mu) \subset L^1(X, \mathscr{F}, \mu)$. For any $g \in L^2(X, \mathscr{F}, \mu)$, by Cauchy's inequality and Corollary 3.5,

$$\|A_n g - P_T g\|_{L^1} \leq \sqrt{\|A_n g - P_T g\|_{L^2} \, \|\mathbf{1}\|_{L^2}} = \sqrt{\mu(X)} \|A_n g - P_T g\|_{L^2} \to 0.$$

Hence $(A_n g)$ is a Cauchy sequence in L^1. If $f \in L^1(X, \mathscr{F}, \mu)$ and $\epsilon > 0$, we choose $g \in L^2(X, \mathscr{F}, \mu)$ such that $\|f - g\|_{L^1} < \epsilon/3$. Since $\|T(f-g)\|_{L^1} = \|f - g\|_{L^1}$, we have $\|A_n(f-g)\| \leq \|f-g\|_{L^1} < \epsilon/3$ for all $n \in \mathbb{N}$. Furthermore, there exists N such that $\|A_n g - A_m g\|_{L^1} < \epsilon/3$ for all $n, m > N$, and

$$\|A_n f - A_m f\|_{L^1} \leq \|A_n f - A_n g\|_{L^1} + \|A_n g - A_m g\|_{L^1} + \|A_m g - A_m f\|_{L^1} < \epsilon.$$

Hence $(A_n f)$ is also a Cauchy sequence in L^1, which converges to a function $\bar{f} \in L^1(X, \mathscr{F}, \mu)$ by L^1-completeness. To show that $\bar{f}$ is T-invariant, note that

$$\|A_n f \circ T - A_n f\|_{L^1} = \left\| \frac{1}{n} \left(f \circ T^n - f \right) \right\|_{L^1} \leq \frac{2}{n} \|f\|_{L^1},$$

which converges to 0 as $n \to \infty$. Hence $\bar{f} \circ T = \bar{f}$ a.e., and $\bar{f}$ is T-invariant. $\square$

3.2 The Maximal Ergodic Theorem

We now turn to the question of almost everywhere convergence of the ergodic averages. As in the case of the averages that occur in the Lebesgue differentiation theorem, the key to dealing with such pointwise limits lies in estimate for the corresponding maximal function:

$$f^* = \sup_{1 \leq n < \infty} A_n f = \sup_{1 \leq n < \infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k.$$

We first state our main result below.

Theorem 3.7 (Maximal ergodic theorem)**.** *Let $(X, \mathscr{F}, \mu, T)$ be a measure-preserving system, and fix $\alpha \in \mathbb{R}$. For each $f \in L^1(X, \mathscr{F}, \mu)$, define*

$$E_\alpha^f = \left\{ x \in X : \sup_{1 \leq n < \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) > \alpha \right\}.$$

Then

$$\alpha \mu(E_\alpha^f) \leq \int_{E_\alpha} f \, d\mu \leq \|f\|_{L^1}.$$

Remark. If $\alpha > 0$, the result can be written as

$$\mu(E_\alpha^f) \leq \frac{1}{\alpha} \int_{E_\alpha} f \, d\mu \leq \frac{1}{\alpha} \|f\|_{L^1}. \tag{3.1}$$

This result is a corollary of the following maximal inequality.

Proposition 3.8 (Maximal inequality). *Let $U : L^1(X, \mathscr{F}, \mu) \to L^1(X, \mathscr{F}, \mu)$ be a positive linear operator such that $\|U\| \leq 1$. For $g \in L^1(X, \mathscr{F}, \mu)$, define the functions*

$$g_n = g + Ug + U^2 g + \cdots U^{n-1} g$$

for $n \in \mathbb{N}$, with $g_0 = 0$. Let $G_N(x) = \max_{0 \leq n \leq N} g_n(x)$ for all $x \in X$. Then for every $N \geq 1$,

$$\int_{\{G_N > 0\}} g \, d\mu \geq 0.$$

Proof. Since U is a positive linear operator, for $0 \leq n \leq N$, we have $UG_N + g \geq Ug_n + g = g_{n+1}$. Hence

$$UG_N + g \geq \max_{1 \leq n \leq N+1} g_n \geq \max_{1 \leq n \leq N} g_n.$$

Since $g_0 = 0$, on the set $E = \{G_N > 0\}$, we have

$$UG_N + g \geq \max_{1 \leq n \leq N} g_n = \max_{0 \leq n \leq N} g_n = G_N.$$

Therefore $g \geq G_N - UG_N$ on E. Since $G_N \geq 0$, we have $UG_N \geq 0$, and

$$\begin{aligned}
\int_E g \, d\mu &\geq \int_E G_N \, d\mu - \int_E UG_N \, d\mu = \int_X G_N \, d\mu - \int_E UG_N \, d\mu \\
&\geq \int_X G_N \, d\mu - \int_X UG_N \, d\mu = \|G_N\|_{L^1} - \|UG_N\|_{L^1} \geq 0,
\end{aligned}$$

where the last inequality follows from $\|U\| \leq 1$. Then we complete the proof. □

Now we prove the main theorem.

Proof of Theorem 3.7. Define $g = f - \alpha$ and $Ug = g \circ T$ in Proposition 3.8. Then

$$E_\alpha^f = \left\{ x \in X : \sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) > \alpha \right\} = \bigcup_{N=0}^{\infty} \{x \in X : G_N(x) > 0\}.$$

By Proposition 3.8 and Lebesgue dominated convergence theorem,

$$\int_{E_\alpha^f} f \, d\mu - \alpha \mu(E_\alpha^f) = \int_{E_\alpha^f} g \, d\mu \geq 0.$$

Thus we complete the proof. □

Remark. When $\alpha > 0$, we apply the same result on the negation $-f \in L^1(X, \mathscr{F}, \mu)$, we have

$$\mu \left(\inf_{1 \leq n < \infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k < -\alpha \right) \leq \frac{1}{\alpha} \|f\|_{L^1}.$$

Combining this with (3.1), we get the two-sided bound:

$$\mu \left(\sup_{1 \leq n < \infty} |A_n f| > \alpha \right) = \mu \left(\sup_{1 \leq n < \infty} \left| \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \right| > \alpha \right) \leq \frac{2}{\alpha} \|f\|_{L^1}.$$

We later use this conclusion in the proof of pointwise convergence result.

3.3 The Birkhoff Ergodic Theorem

In this section, we focus on the pointwise convergence theorem of ergodic averages. Our result is established on finite measure spaces, and it is convenient to assume that the measure-preserving system $(X, \mathscr{F}, \mu, T)$ is on a probability space. Before we proceed, we first introduce the definition of ergodicity.

Definition 3.9 (Ergodic transformation)**.** A measure-preserving transformation $T : X \to X$ on a measure space $(X, \mathscr{F}, \mu)$ is said to be *μ-ergodic* if every T-invariant subset of X is trivial, i.e. for any $E \in \mathscr{F}$,

$$T^{-1}(E) = E \quad \Rightarrow \quad \mu(E) = 0 \ \text{ or } \ \mu(X\backslash E) = 0.$$

Following are some alternate characterizations of ergodicity.

Proposition 3.10. *Let $(X, \mathscr{F}, \mu, T)$ be a measure-preserving system. The following are equivalent:*

(i) T is μ-ergodic;

(ii) For any $E \in \mathscr{F}$, if $T^{-1}(E)$ and E only differ by a μ-null set, i.e. $\mu(T^{-1}(E)\backslash E) + \mu(E\backslash T^{-1}(E)) = 0$, then $\mu(E) = 0$ or $\mu(X\backslash E) = 0$;

(iii) For any measurable function $f : X \to \mathbb{C}$, if $f \circ T = f$ a.e., then f is constant a.e..

Proof. (i) $\Rightarrow$ (ii). Let E be a set such that $\mu(T^{-1}(E)\backslash E) + \mu(E\backslash T^{-1}(E)) = 0$. Then

$$\begin{aligned} T^{-n}(E)\backslash E &\subset \bigcup_{k=0}^{n-1} T^{-k-1}(E)\backslash T^{-k}(E) = \bigcup_{k=0}^{n-1} T^{-k}(T^{-1}(E)\backslash E), \\ E\backslash T^{-n}(E) &\subset \bigcup_{k=0}^{n-1} T^{-k}(E)\backslash T^{-k-1}(E) = \bigcup_{k=0}^{n-1} T^{-k}(E\backslash T^{-1}(E)). \end{aligned}$$

Since T is measure-preserving, we have $\mu(T^{-n}(E)\backslash E) + \mu(E\backslash T^{-n}(E)) = 0$. We define

$$F = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} T^{-n}(E).$$

Then $T^{-1}(F) = F$, and we have either $\mu(F) = 0$ or $\mu(X\backslash F) = 0$ by ergodicity of μ. Moreover,

$$\begin{aligned} F\backslash E &= \bigcap_{N=1}^{\infty} \left(\bigcup_{n=N}^{\infty} T^{-n}(E) \right) \backslash E = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} T^{-n}(E)\backslash E, \\ E\backslash F &= \bigcup_{N=1}^{\infty} E\backslash \left(\bigcup_{n=N}^{\infty} T^{-n}(E) \right) = \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} E\backslash T^{-n}(E). \end{aligned}$$

Hence $\mu(E\backslash F) = \mu(F\backslash E) = 0$, and we have either $\mu(E) = 0$ or $\mu(X\backslash E) = 0$.

(ii) $\Rightarrow$ (iii). For f given in (iii), by considering $\mathrm{Re}f$ and $\mathrm{Im}f$ separately, we may assume $f : X \to \mathbb{R}$ and $f \circ T = f$ a.e.. For any $t \in \mathbb{R}$, the sets $E_t = \{f \leq t\}$ and $T^{-1}(E_t) = \{f \circ T \leq t\}$ only differ by a μ-null set. By (ii), we have $\mu(\{f \leq t\}) = 0$ or $\mu(\{f > t\}) = 0$. We take

$$c = \sup\{t \in \mathbb{R} : \mu(\{f \leq t\}) = 0\} = \inf\{t \in \mathbb{R} : \mu(\{f > t\}) = 0\}.$$

Since $\{f > c\} = \bigcup_{n\in\mathbb{N}}\{f > c+n^{-1}\}$ and $\{f < c\} = \bigcup_{n\in\mathbb{N}}\{f < c+n^{-1}\}$, we have $\mu(\{f > c\}) = \mu(\{f < c\}) = 0$. Hence $f = c$ a.e..

(iii) $\Rightarrow$ (i). If E is a T-invariant set, i.e. $E = T^{-1}(E)$, we take $f = \chi_E$ in (iii). Then $\chi_E \circ T = \chi_{T^{-1}(E)}$, which equals χ_E a.e.. By (iii), $\chi_E = c$ a.e., where $c \in \{0, 1\}$. Hence either $\mu(E) = 0$ or $\mu(X\backslash E) = 0$. ☐

Now we are ready to introduce the main result.

Theorem 3.11 (Birkhoff's theorem)**.** *Let $(X, \mathscr{F}, \mu, T)$ be a measure-preserving system on a probability space.*

(i) For each $f \in L^1(X, \mathscr{F}, \mu)$, the ergodic average

$$A_n f = \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k$$

converges almost everywhere to a T-invariant function $\bar{f} \in L^1(X, \mathscr{F}, \mu)$, where $\int_X f \, d\mu = \int_X \bar{f} \, d\mu$.

(ii) In addition, if T is μ-ergodic, $\bar{f} = \int_X f \, d\mu$.

Proof. We first assume that $g \in L^\infty(X, \mathscr{F}, \mu)$. By Corollary 3.6, there is a T-invariant function $\bar{g} \in L^1(X, \mathscr{F}, \mu)$ such that $A_n g \to \bar{g}$ in L^1. For any $\epsilon > 0$, we choose n sufficiently large so that $\|\bar{g} - A_n g\|_{L^1} < \epsilon^2$. Applying maximal ergodic theorem [Theorem 3.7] to the function $h = \bar{g} - A_n g$, we have

$$\mu\left(\left\{x \in X : \sup_{m \in \mathbb{N}} |A_m(\bar{g} - A_n g)| > \epsilon\right\}\right) \leq \frac{2}{\epsilon} \|\bar{g} - A_n g\|_{L^1} < 2\epsilon.$$

Since $\bar{g}$ is T-invariant, $A_m \bar{g} = \bar{g}$. Also,

$$\begin{aligned} A_m(A_n g) &= \frac{1}{mn} \sum_{j=0}^{m-1} \sum_{k=0}^{n-1} g \circ T^{j+k} = \frac{1}{mn} \sum_{j=0}^{m-1} \sum_{k=0}^{n-1} \left(g \circ T^j + g \circ T^{j+k} - g \circ T^j\right) \\ &= A_m g + \frac{1}{mn} \sum_{k=0}^{n-1} \sum_{j=0}^{m-1} \left(g \circ T^{j+k} - g \circ T^j\right) = A_m g + \frac{1}{mn} \sum_{k=1}^{n-1} \sum_{j=0}^{k-1} \left(g \circ T^{m+j} - g \circ T^j\right), \end{aligned}$$

and

$$|A_m(A_n g) - A_m g| \leq \frac{1}{mn} \sum_{k=1}^{n-1} \sum_{j=0}^{k-1} \left|g \circ T^{m+j} - g \circ T^j\right| \leq \frac{1}{mn} \frac{n(n-1)}{2} 2\|g\|_\infty = \frac{n-1}{m} \|g\|_\infty,$$

which converges to 0 as $m \to \infty$. Then $\limsup_{m \to \infty} |\bar{g} - A_m g| = \limsup_{m \to \infty} |\bar{g} - A_m(A_n g)|$, and

$$\mu\left(\limsup_{m \to \infty} |\bar{g} - A_m g| > \epsilon\right) = \mu\left(\limsup_{m \to \infty} |\bar{g} - A_m(A_n g)| > \epsilon\right) = \mu\left(\limsup_{m \to \infty} |A_m(\bar{g} - A_n g)| > \epsilon\right) < 2\epsilon.$$

Thus $\limsup_{m \to \infty} |\bar{g} - A_m g| = 0$, and $A_m g \to \bar{g}$ a.e. as $m \to \infty$. To generalize the result to $f \in L^1(X, \mathscr{F}, \mu)$, we take $g \in L^\infty(X, \mathscr{F}, \mu)$ such that $\|f - g\|_{L^1} < \epsilon^2$. By Corollary 3.6, we take a T-invariant $\bar{f} \in L^1(X, \mathscr{F}, \mu)$ such that $\|A_n f - \bar{f}\|_{L^1} \to 0$. Since $\|A_n f - A_n g\|_{L^1} \leq \|f - g\|_{L^1} < \epsilon^2$, we have $\|\bar{f} - \bar{g}\|_{L^1} < \epsilon^2$. Thus

$$\begin{aligned} \mu\left(\limsup_{n \to \infty} |\bar{f} - A_n f| > 2\epsilon\right) &\leq \mu\left(|\bar{f} - \bar{g}| + \limsup_{n \to \infty} |\bar{g} - A_n g| + \sup_{n \in \mathbb{N}} |A_n f - A_n g| > 2\epsilon\right) \\ &\leq \mu\left(|\bar{f} - \bar{g}| > \epsilon\right) + \mu\left(\sup_{n \in \mathbb{N}} |A_n(f - g)| > \epsilon\right) \\ &\leq \frac{\|\bar{f} - \bar{g}\|_{L^1}}{\epsilon} + \frac{2\|f - g\|_{L^1}}{\epsilon} < 3\epsilon, \end{aligned}$$

where the third inequality follows from Markov's inequality and the maximal ergodic theorem. Hence $A_n f \to \bar{f}$ a.e. and in L^1. Since T is measure preserving,

$$\int_X \bar{f} \, d\mu = \int_X A_n f \, d\mu = \int_X f \, d\mu.$$

In addition, if T is a μ-ergodic transformation, by Proposition 3.10 the T-invariant function $\bar{f}$ is constant a.e.. Since $\mu(X) = 1$, we have $\bar{f} = \int_X f \, d\mu$ a.e.. $\square$

Remark. For a measure-preserving system $(X, \mathscr{F}, \mu, T)$ on a probability space and a function $f \in L^1(X, \mathscr{F}, \mu)$, we define the *time average at* $x \in X$ to be $(A_n f)(x) = \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)$ and the *space average* $\int_X f\, d\mu$.

A brief interpretation for Birkhoff theorem is that, if T is μ-ergodic, then for almost every $x \in X$, the time average converges to the space average as the time n goes to infinity.

We can obtain a stronger mean ergodic theorem as a consequence of Birkhoff's theorem.

Theorem 3.12 (Mean ergodic theorem)**.** *Let $(X, \mathscr{F}, \mu, T)$ be a measure-preserving system on a probability space, and let $1 \leq p < \infty$. If $f \in L^p(X, \mathscr{F}, \mu)$, the ergodic average $A_n f = \sum_{k=0}^{n-1} f \circ T^k$ converges in L^p to a T-invariant function $\bar{f} \in L^p(X, \mathscr{F}, \mu)$, i.e.*

$$\lim_{n\to\infty} \left\| \sum_{k=0}^{n-1} f \circ T^k - \bar{f} \right\|_{L^p} = 0.$$

Proof. We first take a bounded function $g \in L^\infty(X, \mathscr{F}, \mu)$, so that $g \in L^1(X, \mathscr{F}, \mu)$. By Birkhoff's theorem, there exists a T-invariant $\bar{g} \in L^1(X, \mathscr{F}, \mu)$ such that $A_n g \to \bar{g}$ a.e.. Then for almost all $x \in X$,

$$\lim_{n\to\infty} |(A_n g)(x) - \bar{g}(x)|^p = 0.$$

Since g is bounded, so is $A_n g$ and $\bar{g}$, and $\bar{g} \in L^\infty(X, \mathscr{F}, \mu)$. By Lebesgue dominated convergence theorem,

$$\lim_{n\to\infty} \|A_n g - \bar{g}\|_{L^p}^p = 0.$$

We now consider the general case $f \in L^p(X, \mathscr{F}, \mu)$. Since T is measure-preserving, we have $\|f \circ T^n\|_{L^p} = \|f\|_{L^p}$ for all $n \in \mathbb{N}$, and $\|A_n f\|_{L^p} \leq \|f\|_{L^p}$ by the triangle inequality. Given any $\epsilon > 0$, we take a bounded $g \in L^\infty(X, \mathscr{F}, \mu)$ such that $\|f - g\|_{L^p} < \epsilon/3$. Since $(A_n g)$ is a Cauchy sequence in L^p, we take $N > 0$ such that $\|A_n g - A_m g\|_{L^p} < \epsilon/3$ for all $m, n \geq N$, and hence

$$\begin{aligned}\|A_n f - A_m f\|_{L^p} &\leq \|A_n f - A_n g\|_{L^p} + \|A_n g - A_m g\|_{L^p} + \|A_m g - A_m f\|_{L^p} \\ &\leq 2\|f - g\|_{L^p} + \|A_n g - A_m g\|_{L^p} < \epsilon.\end{aligned}$$

Thus $(A_n f)$ is Cauchy in L^p, and there exists $\bar{f} \in L^p(X, \mathscr{F}, \mu)$ such that $\|A_n f - \bar{f}\|_{L^p} \to 0$. By Birkhoff's theorem, $(A_n f)$ also admits a pointwise limit, which coincides the L^p limit $\bar{f}$. Hence $\bar{f}$ is T-invariant. $\square$

Remark. To summarize, if $1 \leq p < \infty$ and $f \in L^p(X, \mathscr{F}, \mu)$, the ergodic average sequence $(A_n f)$ admits a limit $\bar{f} \in L^p(X, \mathscr{F}, \mu)$ such that

$$\lim_{n\to\infty} \|A_n f - \bar{f}\|_{L^p} = 0, \quad \text{and} \quad \lim_{n\to\infty} (A_n f)(x) = \bar{f}(x) \;\; \text{for a.e. } x \in X.$$

In a nutshell, $A_n f \to \bar{f}$ both a.e. and in L^p.

3.4 The Krein-Milman Theorem

In this section we introduce a general result about compact convex subsets of a locally convex Hausdorff topological vector space, which is used in the proof of unique ergodicity.

Definition 3.13 (Extreme point and face)**.** Let X be a vector space and $K \subset X$ a nonempty convex subset.

(i) A point x of K is called an *extreme point* of K if there do not exist $y, z \in K$ and $0 < \lambda < 1$ such that $\lambda y + (1-\lambda)z = x$. We denote by $\mathrm{ext}(K)$ the set of extreme points of K.

(ii) A nonempty convex subset $F \subset K$ is called a *face* of K if for all $x, y \in K$ and $0 < \lambda < 1$ such that $\lambda x + (1-\lambda)y \in F$, we have $x, y \in F$.

Remark. A point $x \in K$ is an extreme point of K if and only if the singleton $\{x\}$ is a face of K.

Lemma 3.14. *Let X be a vector space, and let A, B, C be convex subsets of K. If B is a face of A and C is a face of B, then C is a face of A.*

Proof. Let $x, y \in A$ and $0 < \lambda < 1$. If $\lambda x + (1-\lambda)y \in C$, since $C \subset B$ and B is a face of A, we have $x, y \in B$. Again, since C is a face of B, we have $x, y \in C$. Therefore C is a face of A. $\square$

Lemma 3.15. *Let X be a locally convex Hausdorff topological space. If $K \in \mathscr{K}$ is a compact convex set and $\ell : X \to \mathbb{R}$ is a continuous linear functional, the set*

$$F_\ell := \left\{ x \in K : \ell(x) = \sup_{y \in K} \ell(y) \right\}$$

is a nonempty compact convex subset of K, and F_ℓ is a face of K.

Proof. We abbreviate $c = \sup_{y \in K} \ell(y)$.

- Since K is compact and ℓ is continuous, there exists $x \in K$ such that $\ell(x) = c$, and F is nonempty.
- Since X is Hausdorff and ℓ is continuous, both K and $\ell^{-1}(\{c\})$ is closed. Hence F_ℓ is closed and compact.
- Since K is convex and f is linear, $\ell^{-1}(\{c\})$ is convex, and so is F.

To summarize, F is nonempty, compact and convex. To prove that F is a face of K, we fix $x, y \in K$ and $0 < \lambda < 1$ such that $\lambda x + (1-\lambda)y \in F$. Then $\lambda\ell(x) + (1-\lambda)\ell(y) = \ell(\lambda x + (1-\lambda)y) = c$. Since both $\ell(x)$ and $\ell(y)$ are no greater than c, we have $\ell(x) = \ell(y) = c$, and $x, y \in F$. Hence F_ℓ is a face of K. $\square$

Lemma 3.16 (Existence)**.** *Let X be a locally convex Hausdorff topological vector space, and let $K \subset X$ be a nonempty compact convex set. Then the set of extreme points of K is nonempty.*

Proof. The proof is divided to three steps.

Step I. Let $\mathscr{K}$ be the set of all nonempty compact convex subset of X, and define the relation $\preceq$ on $\mathscr{K}$ by $F \preceq K$ if and only if F is a face of K. By Lemma 3.14, $(K, \preceq)$ is a partially ordered set. Since X is Hausdorff, every nonempty chain $\mathscr{C} \subset \mathscr{K}$ has a infimum $C_0 = \bigcap_{C \in \mathscr{C}} C \in \mathscr{K}$.

Step II. We claim that every minimal element of $\mathscr{K}$ is a singleton.

If $K \subset \mathscr{K}$ is not a singleton, we take $x, y \in K$ such that $x \neq y$ and take a convex open neighborhood U of x that does not contain y. Using the hyperplane separation theorem, there exists a continuous linear functional $\ell : X \to \mathbb{R}$ such that $\ell(y) < \ell(z)$ for all $z \in U$. By Lemma 3.15, the set $F_\ell \in \mathscr{K}$ is a face of K and $y \in K \backslash F$. Hence K is not a minimal element of $\mathscr{K}$.

Step IV. By Step I and Zorn's lemma, there exists a minimal element $E \subset \mathscr{K}$. By Step III, the minimal element E is a singleton $\{x\}$. Then $x \in \mathrm{ext}(K)$. $\square$

Now we introduce the Krein-Milman theorem.

Theorem 3.17 (Krein-Milman theorem)**.** *Let X be a locally convex Hausdorff topological vector space, and let $K \subset X$ be a nonempty compact convex set. Then K is the closed convex hull of its extreme points, i.e.*

$$K = \overline{\mathrm{conv}}(\mathrm{ext}(K)).$$

Proof. Following the proof of Lemma 3.16, we have $K \in \mathscr{K}$. To prove the desired result, it suffices to show $K \subset \overline{\mathrm{conv}}(\mathrm{ext}(K))$. We argue by contradiction. If $x \in K \backslash \overline{\mathrm{conv}}(\mathrm{ext}(K))$, there exists an open convex neighborhood $U \subset X$ of x such that $U \cap \overline{\mathrm{conv}}(\mathrm{ext}(K)) = \emptyset$. Since $\mathrm{ext}(K)$ is nonempty by Lemma 3.16, there exists a continuous linear functional ℓ such that $\ell(x) > \sup_{y \in \overline{\mathrm{conv}}(\mathrm{ext}(K))} \ell(y)$. By Lemma 3.15, the set $F_\ell = \{x \in K : f(x) = \sup f(K)\}$ is a face of K and $F_\ell \cap \mathrm{ext}(K) = \emptyset$. On the other hand, by Lemma 3.16, the compact convex set F_ℓ has an extreme point x, which is also an extreme point of K by Lemma 3.14. This contradicts the fact that $F_\ell \cap \mathrm{ext}(K) = \emptyset$. Thus we complete the proof. $\square$

3.5 Ergodic Measures and Unique Ergodicity

Invariant measures. For convenience, we focus on a compact metrizable space X equipped with the Borel σ-algebra $\mathscr{B}$. Then X is a second countable space, and the space $C(X)$ of all continuous functions $f: X \to \mathbb{C}$ with the supremum norm $\|\cdot\|_\infty$ is a separable Banach space. Furthermore, by Corollary 2.27, the dual space of $C(X)$ is isomorphic to the space $M(X)$ of complex Borel measures on X. Let $T: X \to X$ is a homeomorphism on X. A Borel probability measure μ on X is said to be T-*invariant* if

$$\int_X f \circ T\, d\mu = \int_X f\, d\mu, \quad \text{for all } f \in C(X).$$

We denote by $M_T(X)$ the set of all T-invariant Borel probability measures on X.

Lemma 3.18. *Let X be a compact metrizable space, and $T: X \to X$ a homeomorphism.*

(i) $M_T(X)$ is a weak compact convex subset of the unit sphere in $M(X)$.*

(ii) $M_T(X)$ is nonempty.

(iii) If $\mu \in M_T(X)$, then $(X, \mathscr{B}, \mu, T)$ is a measure-preserving system, i.e. $\mu(E) = \mu(T^{-1}(E))$ for all $E \in \mathscr{B}$.

Proof. (i) By definition $M_T(X)$ is a convex subset of the unit sphere in $M(X)$. By Banach-Alaoglu theorem, the closed unit ball is compact in the weak* topology on $M(X)$. Then it suffices to show that $M_T(X)$ is weak* closed. We note that a sequence of complex Borel measures $\mu_n \to \mu$ in the weak* topology on $M(X)$ if and only if $\int_X f\, d\mu_n \to \int_X f\, d\mu$ for all $f \in C(X)$. If $\mu_n \in M_T(X)$, by setting $f = 1$ we know that $\mu(X) = \lim_{n\to\infty} \mu_n(X) = 1$. Furthermore,

$$\int_X f \circ T\, d\mu = \lim_{n\to\infty} f \circ T\, d\mu_n = \lim_{n\to\infty} f\, d\mu_n = \int_X f\, d\mu, \quad \forall f \in C(X).$$

Therefore $\mu \in M_T(X)$, and $M_T(X)$ is closed and hence compact in the weak* topology on $M(X)$.

(ii) Fix $x_0 \in X$. For each $n \in \mathbb{N}$, define the Borel probability measure $\mu_n: \mathscr{B} \to [0,1]$ by

$$\int_X f\, d\mu = \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x_0), \quad f \in C(X).$$

By Banach-Alaoglu theorem, the sequence has a weak* convergent subsequence (μ_{n_j}). We denote by μ its weak* limit in $M(X)$. Then $\mu(X) = \int_X 1\, d\mu = \lim_{n\to\infty} \int_X 1\, d\mu_n = 1$, and for all $f \in C(X)$,

$$\int_X f \circ T\, d\mu = \lim_{j\to\infty} \frac{1}{n_j} \sum_{k=1}^{n_j} f(T^k x_0) = \lim_{j\to\infty} \frac{1}{n_j} \sum_{k=0}^{n_j-1} f(T^k x_0) = \int_X f\, d\mu.$$

Therefore $\mu \in M_T(X)$, and $M_T(X)$ is nonempty.

(iii) We defined by $\nu(E) = (T_*\mu)(E) = \mu(T^{-1}(E))$ the pushforward of μ, which is also a measure on $\mathscr{B}$ by continuity of T. By the change-of-variable formula, it suffices to show that $\nu = \mu$ on $\mathscr{B}$.

For a closed subset $F \subset X$, define $f_n(x) = \max\{1 - nd(x, F), 0\}$. Then $f_n \in C(X)$ and $f_n \downarrow \chi_F$ as $n \to \infty$. By monotone convergence theorem,

$$\nu(F) = \lim_{n\to\infty} \int_X f_n\, d\nu = \lim_{n\to\infty} \int_X f_n \circ T\, d\mu = \lim_{n\to\infty} \int_X f_n\, d\mu = \mu(F).$$

Thus $\nu(F) = \mu(F)$ for all closed subset $F \subset X$, and $\mu(U) = \nu(U)$ for all open subset $U \subset X$. By outer-regularity of μ, we have $\mu = \nu$ everywhere on $\mathscr{B}$. $\square$

Remark. Since $T: X \to X$ is an homeomorphism, both T and T^{-1} are measurable. For all $E \in \mathscr{B}$, we have $\mu(E) = \mu(T^{-1}(T(E))) = \mu(T(E))$. Hence the inverse T^{-1} is also a measure-preserving transformation.

Ergodic measures. A T-invariant probability measure μ is said to be T-*ergodic* if T is μ-ergodic, i.e.

$$T^{-1}(E) = E \quad \Rightarrow \quad \mu(E) \in \{0, 1\}.$$

We have the following characterization of T-ergodic measures.

Theorem 3.19 (Ergodicity and Extremity). *Let X be a compact metrizable space, and let $T : X \to X$ be a homeomorphism. If $\mu \in M_T(X)$, the following are equivalent:*

(i) μ is T-ergodic;

(ii) μ is an extreme point of $M_T(X)$.

Proof. The proof has three steps.

Step I. Let μ_1, μ_2 be T-ergodic measures such that $\mu_1(E) = \mu_2(E)$ for every T-invariant Borel set $E \subset X$. We claim that $\int_X f\, d\mu_1 = \int_X f\, d\mu_2$ for each $f \in C(X)$, hence $\mu_1 = \mu_2$ by Riesz representation theorem.

By Corollary 3.6, the sequence $A_n f$ converges to $\int_X f\, d\mu_j$ in L^1, and hence a subsequence $A_{n_i} f$ converges a.e. to $\int_X f\, d\mu_j$, where $j = 1, 2$. Hence there exists $A_j \subset X$ such that $\mu(A_j) = 1$ and

$$\int_X f\, d\mu_j = \lim_{i\to\infty} \frac{1}{n_i} \sum_{k=0}^{n_i - 1} f(T^k x) \quad \text{for all } x \in A_j.$$

For $j = 1, 2$, define $E_j = \bigcap_{n\in\mathbb{Z}} T^n(A_j)$, so that E_j is a T-invariant set with $\mu_j(E_j) = 1$. By assumption, $\mu_1(E_1) = \mu_2(E_1) = \mu_1(E_2) = \mu_2(E_2) = 1$. Then the T-invariant set $E := E_1 \cap E_2$ is nonempty, because $\mu(E) = \mu(E_1) + \mu(E_2) - \mu(E_1 \cup E_2) = 1$. Since $E \subset A_1 \cap A_2$, we fix $x \in E$ and obtain

$$\int_X f\, d\mu_1 = \lim_{i\to\infty} \frac{1}{n_i} \sum_{k=0}^{n_i - 1} f(T^k x) = \int_X f\, d\mu_2.$$

Step II. If $\mu \in M_T(X)$ is ergodic, we claim that μ is an extreme point of $M_T(X)$. Take $\mu_1, \mu_2 \in M_T(X)$ and $0 < \lambda < 1$ such that $\mu = (1-\lambda)\mu_1 + \lambda\mu_2$. If $E \in \mathscr{B}$ is a T-invariant set, we have $\mu(E) \in \{0, 1\}$. Then

- If $\mu(E) = 0$, we have $(1-\lambda)\mu_1(E) + \lambda\mu_2(E) = 0$ and $\mu_1(E) = \mu_2(E) = 0$.
- Similarly, if $\mu(E) = 1$, we have $\mu_1(F) = \mu_2(F) = 1$.

In either case, we have $\mu_1(E) = \mu_2(E) = \mu(E) \in \{0, 1\}$. Hence μ_1 and μ_2 are T-ergodic measures that agree on all T-invariant Borel sets. By Step I, we have $\mu_1 = \mu_2 = \mu$, and hence μ is an extreme point of $M_T(X)$.

Step III. Conversely, if $\mu \in M_T(X)$ is not ergodic, we can find two probability measures $\mu_1, \mu_2 \in M_T(X)$ with $\mu_1 \neq \mu_2$ and $0 < \lambda < 1$ such that $(1-\lambda)\mu_1 + \lambda\mu_2 = \mu$, and hence μ is not an extreme point of $M_Y(X)$.

By non-ergodicity of (μ, T), there exists a Borel set $B \subset X$ such that $T^{-1}(B) = B$ and $0 < \mu(B) < 1$. We then define Borel probability measures

$$\mu_1(E) := \frac{\mu(E\backslash B)}{\mu(X\backslash B)} \quad \text{and} \quad \mu_2(E) := \frac{\mu(E \cap B)}{\mu(B)}, \quad E \in \mathscr{B},$$

and take $\lambda = \mu(B)$. For each $E \in \mathscr{B}$,

$$\mu_2(T^{-1}(E)) = \frac{\mu(T^{-1}(E) \cap B)}{\mu(B)} = \frac{\mu(T^{-1}(E \cap B))}{\mu(B)} = \frac{\mu(E \cap B)}{\mu(B)} = \mu_2(E).$$

Hence μ_2 is T-invariant, and similarly μ_1 is T-invariant. Furthermore, $(1-\lambda)\mu_1 + \lambda\mu_2 = \mu$, as desired. ☐

Corollary 3.20. *Every homeomorphism of a compact metrizable space admits an ergodic measure.*

Proof. Since $M_T(X)$ is a nonempty compact convex subset of $M(X)$ by Lemma 3.18, it has an extreme point μ by Krein-Milman theorem. According to Theorem 3.19, μ is a T-ergodic measure. ☐

Aside from existence, we also wonder whether the ergodic measure of a homeomorphism T is unique.

Definition 3.21 (Unique ergodicity)**.** A homeomorphism T of a compact metrizable space X is said to be *uniquely ergodic*, if there is only one Borel probability measure μ that is T-invariant, i.e. $|M_T(X)| = 1$.

Remark. Since $M_T(X)$ is the closed convex hull of the T-ergodic measures (extreme points), T is uniquely ergodic if and only if there is only one Borel probability measure μ that is T-ergodic.

Theorem 3.22 (Birkhoff's theorem)**.** *Let $T : X \to X$ be a homeomorphism of a compact metrizable space X. The following are equivalent.*

(i) T is uniquely ergodic.

(ii) There exists $\mu \in M_T(X)$ such that for all $f \in C(X)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \int_X f\, d\mu \quad \textit{for all } x \in X. \tag{3.2}$$

(iii) For all $f \in C(X)$, the sequence of functions $A_n f = \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$ converges pointwise to a constant.

(iv) For all $f \in C(X)$, the sequence of functions $A_n f = \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$ converges uniformly to a constant.

Proof. (i) $\Rightarrow$ (iv): If T is uniquely ergodic, we take for each $x \in X$ the sequence

$$\mu_n = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k x}, \quad n = 1, 2, \cdots.$$

By Banach-Alaoglu theorem, and since $|M_T(X)| = 1$, every subsequence of (μ_n) has a further subsequence converging in the weak* topology to the unique element $\mu \in M_T(X)$, which is ergodic. We claim that (μ_n) converges to μ in the weak* topology. If there exists a neighborhood U of μ in the weak* topology such that for each $k \in \mathbb{N}$, there exists $n_k > k$ with $\mu_{n_k} \notin U$, which gives a subsequence (μ_{n_k}) outside U. Therefore,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k x} \overset{w^*}{=} \mu.$$

Integrating both sides with $f \in C(X)$ gives (3.2). Argue (iv) by contradiction. If $(A_n f)$ does not converge uniformly to $\int_X f\, d\mu$, there exists $\epsilon > 0$ such that for each $m \geq 1$, there exists $n_m \geq m$ and $x_m \in X$ such that

$$\left| \frac{1}{n_m} \sum_{k=0}^{n_m - 1} f(T^k x_m) - \int f\, d\mu \right| \geq \epsilon. \tag{3.3}$$

We consider the sequence $\nu_m = \frac{1}{n_m}\sum_{k=0}^{n_m-1} \delta_{T^k x_m}$, which also converges to $\mu \in M_T(X)$ in weak* topology, by passing to a subsequence if necessary. Then the left-hand side of (3.3) goes to 0 as $m \to \infty$, a contradiction.

(iv) $\Rightarrow$ (iii) is clear.

(iii) $\Rightarrow$ (ii): Define the positive linear functional $Af = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$. Then $|Af| \leq \|f\|_\infty$, and $A : C(X) \to \mathbb{C}$ is continuous. By Riesz representation theorem, there is a Borel measure $\mu \in M(X)$ such that $Af = \int_X f\, d\mu$. Since $\mu(X) = A1 = 1$ and $A(f \circ T) = Af$, the measure $\mu \in M_T(X)$.

(ii) $\Rightarrow$ (i): Let $\mu, \nu \in M_T(X)$, where μ is the measure such that the hypothesis holds. For any $f \in C(X)$, by T-invariance of ν and dominated convergence theorem,

$$\int_X f\, d\nu = \lim_{n\to\infty} \int_X A_n f\, d\nu = \int_X \lim_{n\to\infty} A_n f\, d\nu = \int_X \left(\int_X f\, d\mu \right) d\nu = \int_X f\, d\mu. \tag{3.4}$$

By Riesz representation theorem, we have $\mu = \nu$, and $|M_T(X)| = 1$. $\square$

3.6 The Recurrence Theorems

In many scenarios, we are also interested in the recurrence property of a dynamical system $(X, \mathscr{F}, \mu, T)$. Beginning from a state $x_0 \in X$, we wonder if the system will return to a state arbitrarily closed to, or exactly the same as, the initial state x_0.

Definition 3.23 (Recurrence)**.** Let $(X, \mathscr{F}, \mu, T)$ be a measure-preserving system. For a subset $A \subset X$, the *first return time* of A is the map defined for almost every $x \in A$ by

$$n_A(x) = \inf \{n \geq 1 : T^n x \in A\}.$$

We write $n_A = n_A^1$. For each integer $k \geq 2$, we define the k^{th} *return time* by

$$n_A^k(x) = \inf\{n > n_A^{k-1}(x) : T^n x \in A\}.$$

We say that a point $x \in A$ is *infinitely recurrent to* A, or *returns infinitely to* A, if $(T^n x)_{n=1}^{\infty}$ contains a subsequence $(T^{n_k} x)_{k=1}^{\infty} \subset A$, or equivalently, $n_A^k(x) < \infty$ for every $k \in \mathbb{N}$.

Theorem 3.24 (Poincaré recurrence theorem)**.** *Let $(X, \mathscr{F}, \mu, T)$ be a measure-preserving system where μ is a probability measure. For each set $A \subset \mathscr{F}$, almost every $x \in A$ is infinitely recurrent to A. That is,*

$$\mu\left(\{x \in A : T^n x \in A \textit{ for infinitely many } n \in \mathbb{N}\}\right) = \mu(A).$$

Proof. We let $B = \{x \in A : T^n x \in A \text{ for infinitely many } n \in \mathbb{N}\}$. Then

$$\begin{aligned} B &= \{x \in A : T^n x \in A \text{ for infinitely many } n \in \mathbb{N}\} \\ &= \left\{x \in A : \text{for every } n \in \mathbb{N}, \text{ there exists } k \geq n \text{ such that } T^k x \in A\right\} \\ &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A \cap T^{-k}(A) = A \cap \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} T^{-k}(A). \end{aligned}$$

For every $n \in \mathbb{N}_0$, let $A_n = \bigcup_{k=n}^{\infty} T^{-k}(A)$. Then $T^{-n}(A_0) = A_n \subset A_0$. Since $A \backslash A_n \subset A_0 \backslash A_n = A_0 \backslash T^{-n}(A_0)$,

$$0 \leq \mu(A \backslash A_n) \leq \mu(A_0 \backslash T^{-n}(A_0)) = \mu(A_0) - \mu(T^{-n}(A_0)) = 0,$$

where the last inequality follows from the facts that T is measure-preserving and μ is finite. Then

$$\mu(B) = \mu\left(A \cap \bigcap_{n=1}^{\infty} A_n\right) = \mu\left(A \backslash \bigcup_{n=1}^{\infty} (A \backslash A_n)\right) = \mu(A) - \mu\left(\bigcup_{n=1}^{\infty} (A \backslash A_n)\right) = \mu(A).$$

Then we complete the proof. □

Asymptotic relative frequency. The Poincaré recurrence theorem implies that, for almost every $x \in A$, the trajectory $(T^n x)_{n=0}^{\infty}$ hits A infinitely many times. However, it does not predict the frequency of the visits that x makes to the set A. The relative number of elements of $\{x, Tx, T^2 x, \cdots, T^{n-1} x\}$ in A is

$$\frac{1}{n} \left|\left\{T^k x \in A : k = 0, 1, \cdots, n-1\right\}\right| = \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k x).$$

By Birkhoff's theorem, if T is μ-ergodic, for almost all $x \in X$, the asymptotic relative frequency is

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

The Poincaré recurrence theorem asserts that almost every point in a positive measure set returns to the set after a sufficiently long but finite time, but does not give an estimate of the return time. The Kac's lemma states that, in an ergodic system, the points in a positive measure set return to the set within an average time inversely proportional to the measure of the set.

Theorem 3.25 (Kac's lemma)**.** *Let $(X, \mathscr{F}, \mu, T)$ be an ergodic system on a probability space. For each set $A \in \mathscr{F}$ with $\mu(A) > 0$, the first return time n_A satisfies*

$$\int_A n_A \, d\mu = 1.$$

Proof. Let $A_n = \{x \in A : n_A(x) = n\}$ be the set of points in A that return to A after exactly n times. Then

$$A_n = \left\{x \in A : T^n x \in A \text{ and } Tx \notin A,\ T^2 x \notin A, \cdots, T^{n-1}x \notin A\right\} = A \cap T^{-n}(A) \cap \bigcap_{k=1}^{n-1} T^{-k}(X \backslash A).$$

Similarly, we define

$$B_n = \{x \notin A : x \text{ enters } A \text{ at time } n\} = T^{-n}(A) \cap \bigcap_{k=0}^{n-1} T^{-k}(X \backslash A).$$

Since T is μ-ergodic, and $\mu(A) > 0$, almost every $x \in X$ enters A after a sufficiently long time, and the set $\bigcap_{n=0}^{\infty} T^{-n}(X \backslash A)$ has measure zero. Hence both $\mu(A_n)$ and $\mu(B_n)$ goes to zero as $n \to \infty$. Furthermore, $(A_n, B_n)_{n=1}^{\infty}$ are disjoint sets that almost cover X. Also note that

$$T^{-1}(B_n) = T^{-n-1}(A) \cap \bigcap_{k=1}^{n} T^{-k}(X \backslash A) = A_{n+1} \cup B_{n+1}.$$

Since T is measure preserving, $\mu(B_n) = \mu(T^{-1}(B_n)) = \mu(A_{n+1}) + \mu(B_{n+1})$, and by induction we have

$$\mu(B_n) = \sum_{k=n+1}^{\infty} \mu(A_k) + \lim_{k \to \infty} \mu(B_k) = \sum_{k=n+1}^{\infty} \mu(A_k), \quad n \in \mathbb{N}.$$

By Poincaré recurrence theorem, $A = \bigcup_{n=1}^{\infty} A_n$. Therefore

$$\begin{aligned} 1 = \mu(X) &= \sum_{n=1}^{\infty} [\mu(A_n) + \mu(B_n)] = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mu(A_k) = \sum_{k=1}^{\infty} \sum_{n=1}^{k} \mu(A_k) \\ &= \sum_{k=1}^{\infty} k\mu(A_k) = \sum_{k=1}^{\infty} \int_{A_k} n_A \, d\mu = \int_A n_A \, d\mu. \end{aligned}$$

Thus we complete the proof. □

Remark. The Kac's lemma can also be stated as

$$\frac{1}{\mu(A)} \int_A n_A \, d\mu = \frac{1}{\mu(A)},$$

where the left-hand side of the equation is the *mean return time* to A.