Further Topics in Measure Theory

Jyunyi Liao

Contents

1 Interpolation of L^p spaces

1.1 The Riesz-Thorin Interpolation Theorem

We begin from the interpolation of L^p norms. Let $1 \le p < r < q < \infty$. If $f \in L^p(X, \mathscr{F}, \mu) \cap L^q(X, \mathscr{F}, \mu)$, the Hölder's inequality implies

$$\begin{aligned}\int_X |f|^r \, d\mu &= \int_X |f|^{\frac{p(q-r)}{q-p}} |f|^{\frac{q(r-p)}{q-p}} \, d\mu \\ &\le \left(\int_X |f|^p \, d\mu \right)^{\frac{q-r}{q-p}} \left(\int_X |f|^q \, d\mu \right)^{\frac{r-p}{q-p}} = \|f\|_{L^p}^{1-t} \|f\|_{L^q}^t,\end{aligned}$$

where $t = \frac{q(r-p)}{q-p}$ satisfies $\frac{1}{r} = \frac{1-t}{p} + \frac{t}{q}$. This estimate holds even when $q = \infty$, since

$$\int_X |f|^r \, d\mu = \int_X |f|^p |f|^{r-p} \, d\mu \le \int_X |f|^p \|f\|_{L^\infty}^{r-p} \, d\mu = \|f\|_{L^p}^p \|f\|_{L^\infty}^{r-p}.$$

Therefore $f \in L^r(X, \mathscr{F}, \mu)$, and $\|f\|_{L^r}$ can be bounded by L^p and L^q norms. More generally, we have the following interpolation theorem for linear operators.

Theorem 1.1 (Riesz-Thorin interpolation theorem). *Let $p_0, p_1, q_0, q_1 \in [1, \infty]$. Let $(X, \mathscr{F}, \mu)$ and $(Y, \mathscr{G}, \nu)$ be measure spaces. If $q_0 = q_1 = \infty$, we further assume that ν is semifinite. Let T be a linear operator from $L^{p_0}(X, \mathscr{F}, \mu) + L^{p_1}(X, \mathscr{F}, \mu)$ into $L^{q_0}(Y, \mathscr{G}, \nu) + L^{q_1}(Y, \mathscr{G}, \nu)$ such that $\|Tf\|_{L^{q_0}} \le M_0 \|f\|_{L^{p_0}}$ for all $f \in L^{p_0}(X, \mathscr{F}, \mu)$, and $\|Tg\|_{L^{q_1}} \le M_1 \|g\|_{L^{p_1}}$ for all $g \in L^{p_1}(X, \mathscr{F}, \mu)$. For each $0 < t < 1$, define*

$$\frac{1}{p} = \frac{1-t}{p_0} + \frac{t}{p_1}, \qquad \frac{1}{q} = \frac{1-t}{q_0} + \frac{t}{q_1}.$$

Then $\|Tf\|_{L^q} \le M_0^t M_1^{1-t} \|f\|_{L^p}$ for $f \in L^p(X, \mathscr{F}, \mu)$.

We begin by introducing an estimate of L^p-norms using a dual space argument.

Lemma 1.2. *Let $(X, \mathscr{F}, \mu)$ be a measure space, and $p, q \in [1, \infty]$ conjugate exponents. If $q = \infty$, we further assume that μ is semifinite. For each $f \in L^q(X, \mathscr{F}, \mu)$,*

$$\|f\|_{L^q} = \sup \left\{ \left| \int_X fg \, d\mu \right| : \|g\|_{L^p} \le 1, \ g \text{ is simple} \right\} \tag{1.1}$$

Proof. Let M be the right-hand side of (1.1). By Hölder's inequality, we have $\|f\|_{L^q} \|g\|_{L^p} \ge \left| \int_X fg \, d\mu \right|$, and $\|f\|_{L^q} \ge M$. Then it suffices to show the other direction $\|f\|_{L^q} \le M$. We discuss two cases.

Case I: $1 < p, q < \infty$**.** Given $f \in L^q(X, \mathscr{F}, \mu)$, we take a sequence (f_n) of simple functions such that $|f_n| \uparrow |f|$ and $f_n \to f$ a.e., and define

$$g_n = \frac{|f_n|^{q-1} \cdot \overline{\mathrm{sgn} f_n}}{\|f_n\|_{L^q}^{q-1}}.$$

Then $\|g_n\|_{L^p}^p = 1$, and

$$|(f_n - f)g| \le 2|f| \cdot \frac{|f_n|^{q-1}}{\|f_1\|_{L^q}^{q-1}} \le \frac{2|f|^q}{\|f_1\|_{L^q}^{q-1}} \in L^1(X, \mathscr{F}, \mu).$$

By dominated convergence theorem,

$$\lim_{n \to \infty} \int_X (f_n - f) g_n \, d\mu = 0.$$

We then use Fatou's lemma to bound the L^q norm of f:

$$\|f\|_{L^q} = \int_X \frac{|f|^q}{\|f\|_{L^q}^{q-1}}\, d\mu \le \liminf_{n\to\infty} \int_X \frac{|f_n|^q}{\|f_n\|_{L^q}^{q-1}}\, d\mu = \liminf_{n\to\infty} \left| \int_X f_n g_n \, d\mu \right|. \tag{1.2}$$

Passing to a suitable subsequence and applying (1.2), we obtain

$$\|f\|_{L^q} \le \lim_{n\to\infty} \left| \int_X f_n g_n \, d\mu \right| = \lim_{n\to\infty} \left| \int_X f g_n \, d\mu \right| \le M.$$

Case II: $p = 1$ and $q = \infty$. Argue by contradiction. If $\|f\|_{L^\infty} > M$, we choose $\epsilon > 0$ such that the set $E = \{x \in X : |f(x)| > M + \epsilon\}$ has positive measure. Since μ is a semifinite measure, we can choose $F \subset E$ with $0 < \mu(F) < \infty$. Let $g = \chi_F \cdot \overline{\operatorname{sgn} f}/\mu(F)$, and take a sequence of simple functions $g_n \to g$ and $|g_n| \uparrow g$. Then $\|g\|_{L^1} = 1$, $\|g_n\|_{L^1} \le 1$, and

$$|f g_n| = \frac{1}{\mu(F)} \chi_F |f| \le \frac{\|f\|_{L^\infty}}{\mu(F)} \chi_F,$$

which is an integrable function. By dominated convergence theorem and definition of F,

$$M \ge \int_X f g \, d\mu = \lim_{n\to\infty} \int_X f g_n \, d\mu = \frac{1}{\mu(F)} \int_F |f| \, d\mu \ge M + \epsilon,$$

a contradiction! Hence $\|f\|_{L^\infty} \le M$. □

Lemma 1.3 (The three lines lemma)**.** *Let ϕ be a bounded continuous function on the strip $0 \le \operatorname{Re}(z) \le 1$ that is holomorphic in the interior of the strip. If $|\phi(z)| \le M_0$ on $\operatorname{Re}(z) = 0$ and $|\phi(z)| \le M_1$ on $\operatorname{Re}(z) = 1$, then $|\phi(z)| \le M_0^{1-t} M_1^t$ on $\operatorname{Re}(z) = t$, where $0 < t < 1$.*

Proof. We define the function

$$\phi_\epsilon(z) = \phi(z) M_0^{-z} M_1^{z-1} e^{-\epsilon z(1-z)},$$

which also satisfies the hypothesis of the lemma with M_0 and M_1 replaced by 1, and

$$|\phi_\epsilon(x+iy)| = |\phi(x+iy)| M_0^{-x} M_1^{x-1} e^{-\epsilon x(1-x) - \epsilon y^2} \le M_0^{1-x} M_1^{x-1} e^{-\epsilon y^2}.$$

Hence $\phi_\epsilon(z) \to 0$ as $|\operatorname{Im}(z)| \to \infty$. By our hypothesis, for sufficiently large $A > 0$, we have $|\phi_\epsilon| \le 1$ on the boundary of the region $D = \{z : 0 \le \operatorname{Re}(z) \le 1, -A \le \operatorname{Im}(z) \le A\}$. By the maximum modulus principle, $\max_{z \in \partial D} |\phi_\epsilon(z)| = \max_{z \in D} |\phi_\epsilon(z)|$. Hence $|\phi_\epsilon| \le 1$ on D, and hence on the strip $0 \le \operatorname{Re}(z) \le 1$. Letting $\epsilon \to 0$, we obtain $|\phi(z)| M_0^{-t} M_1^{t-1} \le \lim_{\epsilon\to 0} |\phi_\epsilon(z)| \le 1$, where $t = \operatorname{Re}(z)$. □

Proof of Theorem 1.1. The proof has three steps.

Step I: We begin with the case $p_0 = p_1 = p$. Since the case $q_0 = q_1$ is clear, we may assume $q_0 < q_1$. Then

$$\|Tf\|_{L^q} \le \|Tf\|_{L^{q_0}}^{\frac{q_0(q_1-q)}{q(q_1-q_0)}} \|Tf\|_{L^{q_1}}^{\frac{q_1(q-q_0)}{q(q_1-q_0)}} = \|Tf\|_{L^{q_0}}^t \|Tf\|_{L^{q_1}}^{1-t} \le M_0^t M_1^{1-t} \|f\|_{L^p}.$$

Step II. Now we assume $p_0 < p_1 \le \infty$, and in particular $p < \infty$ for all $0 < t < 1$. We begin by taking a simple function $f = \sum_{j=1}^n a_j \chi_{E_j} = \sum_{j=1}^n |a_j| e^{i\theta_j} \chi_{E_j}$ and show that $\|Tf\|_{L^q} \le M_0^{1-t} M_1^t \|f\|_{L^p}$.

By homogeneity of $\|\cdot\|$ and linearity of T, it suffices to show the case $\|f\|_{L^p} = 1$. We estimate $\|Tf\|_{L^q}$ by taking $g = \sum_{k=1}^m b_j \chi_{F_k} = \sum_{k=1}^m |b_j| e^{i\xi_k} \chi_{F_k}$ with $\|g\|_{L^{q'}} = 1$ in (1.1). Define functions α and β as follows:

$$\alpha(z) = \frac{1-z}{p_0} + \frac{z}{p_1}, \quad \beta(z) = \frac{1-z}{q_0} + \frac{z}{q_1}, \quad z \in \mathbb{C},\ 0 \le \operatorname{Re}(z) \le 1.$$

Then $\alpha(t) = 1/p$ and $\beta(t) = 1/q$. We let

$$f_z = \sum_{j=1}^n |a_j|^{\alpha(z)/\alpha(t)} e^{i\theta_j} \chi_{E_j}, \qquad g_z = \begin{cases} \sum_{k=1}^m |b_k|^{(1-\beta(z))/(1-\beta(t))} e^{i\xi_k} \chi_{F_k}, & \beta(t) \neq 1, \\ g, & \beta(t) = 1. \end{cases}$$

Finally, we define

$$\Phi(z) = \int_Y (Tf_z) g_z \, d\nu = \begin{cases} \sum_{j=1}^n \sum_{k=1}^m |a_j|^{\frac{\alpha(z)}{\alpha(t)}} |b_k|^{\frac{1-\beta(z)}{1-\beta(t)}} e^{i(\theta_j + \xi_k)} \int_Y (T\chi_{E_j}) \chi_{F_k} \, d\nu, & \beta(t) \neq 1 \\ \sum_{j=1}^n \sum_{k=1}^m |a_j|^{\frac{\alpha(z)}{\alpha(t)}} |b_k| e^{i(\theta_j + \xi_k)} \int_Y (T\chi_{E_j}) \chi_{F_k} \, d\nu, & \beta(t) = 1. \end{cases}$$

Then $\Phi(z)$ is a bounded and continuous function on the strip $0 \leq \mathrm{Re}(z) \leq 1$ that is holomorphic in the strip. We claim that $\Phi(z) \leq M_0$ on $\mathrm{Re}(z) = 0$ and $\Phi(z) \leq M_1$ on $\mathrm{Re}(z) = 1$. We let $z = i\omega$, where $\omega \in \mathbb{R}$. Since $E_1, \cdots, E_n$ are disjoint, at most one χ_{E_j} is nonzero, and

$$|f_{i\omega}| = \sum_{j=1}^n |a_j|^{p/p_0} \chi_{E_j} = |f|^{p/p_0}.$$

A similar calculation yields

$$|g_{i\omega}| = \sum_{k=1}^m |b_k|^{q'/q_0'} \chi_{F_k} = |g|^{q'/q_0'},$$

where q' and q_0' are the conjugate exponents of q and q_0, respectively, and we set $\frac{\infty}{\infty} = 1$. By Hölder's inequality,

$$|\Phi(i\omega)| \leq \|Tf_{i\omega}\|_{L^{q_0}} \|g_{i\omega}\|_{L^{q_0'}} \leq M_0 \|f_{i\omega}\|_{L^{p_0}} \|g_{i\omega}\|_{L^{q_0'}} = M_0 \|f\|_{L^p} \|g\|_{L^{q'}} = M_0.$$

Similarly, we can show $|\Phi(1 + i\omega)| \leq M_1$. By three lines lemma [Lemma 1.3] and Lemma 1.2, we have

$$\|Tf\|_{L^q} \leq |\Phi(t)| \leq M_0^{1-t} M_1^t.$$

Step III. We have shown that $\|Tf\|_{L^q} \leq M_0^{1-t} M_1^t \|f\|_{L^p}$ for all simple functions. For each $f \in L^p(X, \mathscr{F}, \mu)$ with $f \geq 0$, we choose a sequence of simple functions such that $|f_n| \uparrow |f|$ and $f_n \to f$ pointwise. We let $E = \{x \in X : |f(x)| > 1\}$, and define

$$g = f\chi_E, \quad g_n = f_n \chi_E, \quad h = f - g, \quad h_n = f_n - g_n.$$

Since $p_0 < p < p_1$, we have $g \in L^{p_0}(X, \mathscr{F}, \mu)$ and $h \in L^{p_1}(X, \mathscr{F}, \mu)$. By dominated convergence theorem, $\|f_n - f\|_{L^p} \to 0$, $\|g_n - g\|_{L^{p_0}} \to 0$ and $\|h - h_p\|_{L^{p_1}} \to 0$. Hence $\|Tg_n - Tg\|_{L^{q_0}} \leq M\|g_n - g\|_{L^{p_0}} \to 0$ and $\|Th_n - Th\|_{L^{q_1}} \leq M\|h_n - h\|_{L^{p_1}} \to 0$. By passing to a suitable subsequence we may also assume $Tg_n \to Tg$ a.e. and $Th_n \to Th$ a.e., and then $Tf_n \to Tf$. By Fatou's lemma,

$$\|Tf\|_{L^q} \leq \liminf_{n\to\infty} \|Tf_n\|_{L^q} \leq \liminf_{n\to\infty} M_0^{1-t} M_1^t \|f_n\|_{L^p} = M_0^{1-t} M_1^t \|f\|_{L^p}.$$

Then we finish the proof. □

1.2 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 $\epsilon_n \downarrow 0$.
- $\lambda_{f+g}(2\alpha) \le \lambda_f(\alpha) + \lambda_g(\alpha)$. In addition, if $|f| \le |g|$, then $\lambda_f \le \lambda_g$.
- 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 $1 \le p < \infty$. 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, $[f]_p \le \|f\|_{L^p}$ for all $f \in L^p(X, \mathscr{F}, \mu)$. Therefore the classical L^p is contained in the weak L^p space.

Lemma 1.4. *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 $f \ge 0$ 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. ☐

Definition 1.5 (Sublinear operators of strong and weak types)**.** 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 $(Y, \mathscr{G}, \nu)$.

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

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

$$\|Tf\|_{L^q} \le C_{p,q}\|f\|_{L^p}.$$

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

$$[Tf]_q \le C_{p,q}\|f\|_{L^p}.$$

Clearly, a sublinear operator T of strong type (p, q) is also of weak type (p, q). Also, we say T is of weak type (p, ∞) if and only if it is of strong type (p, ∞).

Theorem 1.6 (Marcinkiewicz interpolation theorem)**.** *Let $1 \leq p_0 \leq q_0 \leq \infty$, $1 \leq p_1 \leq q_1 \leq \infty$ and $q_0 \neq q_1$. Let T be a sublinear operator from $L^{p_0}(X, \mathscr{F}, \mu) + L^{p_1}(X, \mathscr{F}, \mu)$ into the space of measurable functions on $(Y, \mathscr{G}, \nu)$. For each $0 < \gamma < 1$, define*

$$\frac{1}{p} = \frac{1-\gamma}{p_0} + \frac{\gamma}{p_1}, \qquad \frac{1}{q} = \frac{1-\gamma}{q_0} + \frac{\gamma}{q_1}.$$

If T is a sublinear operator of weak types (p_0, q_0) and (p_1, q_1), then T is of strong type (p, q).

The proof Marcinkiewicz interpolation theorem requires the following lemma.

Lemma 1.7 (Minkowski's integral inequality)**.** *Let $(X, \mathscr{F}, \mu)$ and $(Y, \mathscr{G}, \nu)$ be two measure spaces and let $\Phi : X \times Y \to \mathbb{C}$ be a measurable function on the produce space. If $p \geq 1$, we have*

$$\left(\int_X \left| \int_Y \phi(x, y)\, d\nu(y) \right|^p d\mu(x) \right)^{\frac{1}{p}} \leq \int_X \left(\int_X |\phi(x, y)|^p\, d\mu(x) \right)^{\frac{1}{p}} d\mu(y).$$

Proof. Let $\Phi(x) = \int_Y \phi(x, y)\, d\nu(y)$. Similar to the proof of Minkowski's inequality, we estimate $\|\Phi\|_{L^p}^p$ by

$$\begin{aligned}
\int_X |\Phi|^p\, d\mu &\leq \int_X |\Phi(x)|^{p-1} \int_Y |\phi(x, y)|\, d\nu(y)\, d\mu(x) \\
&= \int_Y \int_X |\Phi(x)|^{p-1} |\phi(x, y)|\, d\mu(x)\, d\nu(y) \\
&\leq \int_Y \left(\int_X |\Phi(x)|^{(p-1)\cdot \frac{p}{p-1}}\, d\mu(x) \right)^{\frac{p-1}{p}} \left(\int_X |\phi(x, y)|^p\, d\mu(x) \right)^{\frac{1}{p}} d\mu(y) \\
&= \|\Phi\|_{L^p}^{p-1} \int_X \left(\int_X |\phi(x, y)|^p\, d\mu(x) \right)^{\frac{1}{p}} d\mu(y),
\end{aligned}$$

where we interchange the integrals by Fubini's theorem and use Hölder's inequality to the inner integral. ☐

Lemma 1.8. *If f is a measurable function and $\alpha > 0$, define*

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

Then

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

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 1.6. For notation simplicity we also write $[\cdot]_\infty = \|\cdot\|_{L^\infty}$. Since T is of weak types (p_0, q_0) and (p_1, q_1), there exist constants C_0 and C_1 such that

$$\lambda_f(\alpha) \leq \left(\frac{C_0}{\alpha} \right)^{q_0} \|f\|_{L^{p_0}}^{q_0} \quad \text{and} \quad \lambda_g(\alpha) \leq \left(\frac{C_1}{\alpha} \right)^{q_1} \|g\|_{L^{p_1}}^{q_1}$$

for all $f \in L^{p_0}(X, \mathscr{F}, \mu)$, $g \in L^{p_1}(X, \mathscr{F}, \mu)$ and all $\alpha > 0$. There are several cases to consider.

- Case I: $p_0 = p_1 = p$. We may assume $q_0 < q_1$ by switching subscripts 0 and 1 when necessary.
- Case II: $p_0 \neq p_1$. We may assume $p_0 < p_1$ by switching subscripts 0 and 1 when necessary.

Now we prove the theorem case by case.

Case I (1): $p_0 = p_1 = p$ and $q_0 < q_1 < \infty$. If $f \in L^p(X, \mathscr{F}, \mu)$,

$$\begin{aligned}
\|Tf\|_q^q &= q\int_0^\infty \alpha^{q-1}\lambda_{Tf}(\alpha)\,d\alpha \le q\int_0^{\|f\|_{L^p}} \alpha^{q-1}\lambda_{Tf}(\alpha)\,d\alpha + q\int_{\|f\|_{L^p}}^\infty \alpha^{q-1}\lambda_{Tf}(\alpha)\,d\alpha \\
&\le q\int_0^{\|f\|_{L^p}} \alpha^{q-q_0-1}C_0^{q_0}\|f\|_{L^p}^{q_0}\,d\alpha + q\int_{\|f\|_{L^p}}^\infty \alpha^{q-q_1-1}C_1^{q_1}\|f\|_{L^p}^{q_1}\,d\alpha \\
&= \left(\frac{qC_0^{q_0}}{q-q_0} + \frac{qC_1^{q_1}}{q_1-q}\right)\|f\|_{L^p}^q
\end{aligned}$$

Case I (2): $p_0 = p_1 = p$ and $q_0 < q_1 = \infty$. If $f \in L^p(X, \mathscr{F}, \mu)$, we have $\|Tf\|_{L^{q_1}} \le C_1\|f\|_{L^p}$. Then $\lambda_{Tf}(\alpha) = 0$ when $\alpha > C_1\|f\|_{L^p}$, and

$$\begin{aligned}
\|Tf\|_q^q &= q\int_0^{C_1\|f\|_{L^p}} \alpha^{q-1}\lambda_{Tf}(\alpha)\,d\alpha \le q\int_0^{C_1\|f\|_{L^p}} \alpha^{q-q_0-1}[Tf]_{q_0}^{q_0}\,d\alpha \\
&\le q\int_0^{C_1\|f\|_{L^p}} \alpha^{q-q_0-1}C_0^{q_0}\|f\|_{L^p}^{q_0}\,d\alpha = \frac{qC_0^{q_0}C_1^{q-q_0}}{q-q_0}\|f\|_{L^p}^q.
\end{aligned}$$

Case II (1): $p_0 < p_1 < \infty$ and $q_0, q_1 < \infty$. For $f \in L^p(X, \mathscr{F}, \mu)$, we take g_α and h_α as in Lemma 1.8, where $\alpha > 0$ is to be determined. Then

$$\lambda_{Tf}(2\beta) \le \lambda_{Tg_\alpha}(\beta) + \lambda_{Th_\alpha}(\beta) \le \left(\frac{C_0}{\beta}\right)^{q_0}\left(\int_X |g_\alpha|^{p_0}\,d\mu\right)^{\frac{q_0}{p_0}} + \left(\frac{C_1}{\beta}\right)^{q_1}\left(\int_X |h_\alpha|^{p_1}\,d\mu\right)^{\frac{q_1}{p_1}}. \tag{1.3}$$

Here we allow α to depend on β. By Lemma 1.4, we have

$$\begin{aligned}
\int_X |g_\alpha|^{p_0}\,d\mu &= p\int_0^\infty t^{p_0-1}\lambda_{g_\alpha}(t)\,dt = p_0\int_0^\infty t^{p_0-1}\lambda_f(t+\alpha)\,dt \\
&= p_0\int_\alpha^\infty (t-\alpha)^{p_0-1}\lambda_f(t)\,dt \le p_0\int_\alpha^\infty t^{p_0-1}\lambda_f(t)\,dt,
\end{aligned} \tag{1.4}$$

and similarly,

$$\int_X |h_\alpha|^{p_1}\,d\mu = p_1\int_0^\infty t^{p_1-1}\lambda_{h_\alpha}(t)\,dt = p_1\int_0^\alpha t^{p_1-1}\lambda_f(t)\,dt. \tag{1.5}$$

We combine Lemma 1.4, the inequality (1.3) and the estimates (1.4)-(1.5):

$$\begin{aligned}
\|Tf\|_{L^q}^q &= q\int_0^\infty (2\beta)^{q-1}\lambda_{Tf}(2\beta)\,d(2\beta) = q2^q\int_0^\infty \beta^{q-1}\lambda_{Tf}(2\beta)\,d\beta \\
&\le q2^q\int_0^\infty \left(C_0^{p_0}\beta^{q-q_0-1}\left(\int_X |g_\alpha|^{p_0}\,d\mu\right)^{\frac{q_0}{p_0}} + C_1^{p_1}\beta^{q-q_1-1}\left(\int_X |h_\alpha|^{p_1}\,d\mu\right)^{\frac{q_1}{p_1}}\right)d\beta \\
&\le q2^qC_0^{p_0}p_0^{\frac{q_0}{p_0}}\int_0^\infty \beta^{q-q_0-1}\left(\int_0^\infty \chi_{\{t>\alpha\}}t^{p_0-1}\lambda_f(t)\,dt\right)^{\frac{q_0}{p_0}}d\beta \\
&\quad + q2^qC_1^{p_1}p_1^{\frac{q_1}{p_1}}\int_0^\infty \beta^{q-q_1-1}\left(\int_0^\infty \chi_{\{t\le\alpha\}}t^{p_1-1}\lambda_f(t)\,dt\right)^{\frac{q_1}{p_1}}d\beta.
\end{aligned} \tag{1.6}$$

Since this estimate holds for any $\alpha > 0$, we choose $\alpha = \beta^\sigma$, where

$$\sigma = \frac{p_0(q-q_0)}{q_0(p-p_0)} = \frac{(1-\gamma)\left(\frac{q}{q_0}-1\right)}{(1-\gamma)\left(\frac{p}{p_0}-1\right)} = \frac{\gamma\left(1-\frac{q}{q_1}\right)}{\gamma\left(1-\frac{p}{p_1}\right)} = \frac{p_1(q_1-q)}{q_1(p_1-p)}.$$

We also write $\chi_0 = \chi_{\{t>\alpha\}}$, $\chi_1 = \chi_{\{t\le\alpha\}}$, and

$$\phi_i(t,\beta) = \beta^{\frac{p_i}{q_i}(q-q_i-1)}\chi_i t^{p_i-1}\lambda_f(t),$$

where $i = 0, 1$. Then (1.6) becomes

$$\|Tf\|_{L^q}^q \le \sum_{i=0}^{1} q2^q C_i^{p_i} p_i^{\frac{q_i}{p_i}} \int_0^\infty \left(\int_0^\infty \phi_i(t,\beta)\,dt\right)^{\frac{q_i}{p_i}} d\beta. \tag{1.7}$$

We write $\Phi(\beta) = \int_0^\infty \phi_i(t,\beta)\,dt$. Since $\frac{q_i}{p_i} \ge 1$, by Minkowski's inequality, in either case $i = 0, 1$,

$$\begin{aligned}\int_0^\infty \left(\int_0^\infty \phi_i(t,\beta)\,dt\right)^{\frac{q_i}{p_i}} d\beta &\le \left(\int_0^\infty \left(\int_0^\infty |\phi_i(t,\beta)|^{\frac{q_i}{p_i}}\,d\beta\right)^{\frac{p_i}{q_i}} dt\right)^{\frac{q_i}{p_i}} \\ &= \left(\int_0^\infty \left(\int_0^\infty \beta^{q-q_i-1}\chi_i\,d\beta\right)^{\frac{p_i}{q_i}} t^{p_i-1}\lambda_f(t)dt\right)^{\frac{q_i}{p_i}}.\end{aligned} \tag{1.8}$$

If $q_1 > q_0$, the exponents $q - q_0$ and σ are positive, and $\{t > \beta^\sigma\} = \{\beta < t^{1/\sigma}\}$. Then (1.8) becomes

$$\begin{aligned}\int_0^\infty \left(\int_0^\infty \phi_0(t,\beta)\,dt\right)^{\frac{q_0}{p_0}} d\beta &\le \left(\int_0^\infty \left(\int_0^{t^{1/\sigma}} \beta^{q-q_0-1}\,d\beta\right)^{\frac{p_0}{q_0}} t^{p_0-1}\lambda_f(t)dt\right)^{\frac{q_0}{p_0}} \\ &= \frac{1}{q-q_0}\left(\int_0^\infty t^{p-1}\lambda_f(t)dt\right)^{\frac{q_0}{p_0}} = \frac{1}{q-q_0}\left(\frac{1}{p}\right)^{\frac{q_0}{p_0}} \|f\|_{L^p}^{\frac{pq_0}{p_0}}.\end{aligned}$$

On the other hand, $q - q_1 < 0$, and $\{t \le \beta^\sigma\} = \{\beta \ge t^{1/\sigma}\}$. Then (1.8) becomes

$$\begin{aligned}\int_0^\infty \left(\int_0^\infty \phi_1(t,\beta)\,dt\right)^{\frac{q_1}{p_1}} d\beta &\le \left(\int_0^\infty \left(\int_{t^{1/\sigma}}^\infty \beta^{q-q_1-1}\,d\beta\right)^{\frac{p_1}{q_1}} t^{p_1-1}\lambda_f(t)dt\right)^{\frac{q_1}{p_1}} \\ &= \frac{1}{q_1-q}\left(\int_0^\infty t^{p-1}\lambda_f(t)dt\right)^{\frac{q_1}{p_1}} = \frac{1}{q_1-q}\left(\frac{1}{p}\right)^{\frac{q_1}{p_1}} \|f\|_{L^p}^{\frac{pq_1}{p_1}}.\end{aligned}$$

If $q_1 < q_0$, the exponents $q - q_0$ and σ are negative, and $\{t > \beta^\sigma\} = \{\beta > t^{1/\sigma}\}$. Then (1.8) becomes

$$\begin{aligned}\int_0^\infty \left(\int_0^\infty \phi_0(t,\beta)\,dt\right)^{\frac{q_0}{p_0}} d\beta &\le \left(\int_0^\infty \left(\int_{t^{1/\sigma}}^\infty \beta^{q-q_0-1}\,d\beta\right)^{\frac{p_0}{q_0}} t^{p_0-1}\lambda_f(t)dt\right)^{\frac{q_0}{p_0}} \\ &= \frac{1}{q_0-q}\left(\int_0^\infty t^{p-1}\lambda_f(t)dt\right)^{\frac{q_0}{p_0}} = \frac{1}{q_0-q}\left(\frac{1}{p}\right)^{\frac{q_0}{p_0}} \|f\|_{L^p}^{\frac{pq_0}{p_0}}.\end{aligned}$$

A similar calculation gives

$$\int_0^\infty \left(\int_0^\infty \phi_1(t,\beta)\,dt\right)^{\frac{q_1}{p_1}} d\beta \le \frac{1}{q-q_1}\left(\frac{1}{p}\right)^{\frac{q_1}{p_1}} \|f\|_{L^p}^{\frac{pq_1}{p_1}}.$$

In either case, we plug in (1.8) to (1.7) to get

$$\|Tf\|_{L_q}^q \le \sum_{i=0}^{1} \frac{q2^q C_i^{p_i}}{|q-q_i|}\left(\frac{p_i}{p}\right)^{\frac{q_i}{p_i}} \|f\|_{L^p}^{\frac{pq_i}{p_i}}$$

Therefore

$$\sup\{\|Tf\|_{L^q} : \|f\|_{L^p} = 1\} \le B_{p,q} := 2q^{1/q}\left[\sum_{i=0}^{1} \frac{C_i^{p_i}}{|q-q_i|}\left(\frac{p_i}{p}\right)^{\frac{q_i}{p_i}}\right]^{1/q}.$$

By homogeneity of norms and sublinearity of T, we have $\|Tf\|_{L^q}\| \le B_{p,q}\|f\|_{L^p}$ for all $f \in L^p(X, \mathscr{F}, \mu)$. The remaining cases follow by modifying this procedure.

Case II (2): $p_0 < p_1 = \infty$ **and** $q_0 < q_1 = \infty$**.** We have $p_0 = pt$ and $q_0 = qt$. We take $\alpha = \beta/C_1$, so $\|Th_\alpha\| \le C_1\|h_\alpha\| \le \beta$, and $\lambda_{Th_\alpha}(\beta) = 0$. Then the second term in the estimate (1.3) vanishes, and $\phi_1 = 0$ in the estimate (1.7). We then apply an analogue of (1.8) to get

$$\begin{aligned}\|Tf\|_{L^q}^q &\lesssim \int_0^\infty \left(\int_0^\infty \phi_0(t,\beta)\,dt\right)^{\frac{q_0}{p_0}} d\beta \le \left(\int_0^\infty \left(\int_0^\infty |\phi_0(t,\beta)|^{\frac{q_0}{p_0}}\,d\beta\right)^{\frac{p_0}{q_0}} dt\right)^{\frac{q_0}{p_0}} \\ &= \left(\int_0^\infty \left(\int_0^{C_1 t} \beta^{q-q_0-1}\,d\beta\right)^{\frac{p_0}{q_0}} t^{p_0-1}\lambda_f(t)dt\right)^{\frac{q_0}{p_0}} = \frac{1}{|q-q_0|}\left(C_1^{q-q_0}\int_0^\infty t^{p-1}\lambda_f(t)dt\right)^{\frac{q_0}{p_0}} = \frac{C_1^{q-q_0}}{|q-q_0|}\|f\|_{L^p}^q.\end{aligned}$$

Case II (3): $p_0 < p_1 < \infty$ **and** $q_0 < q_1 = \infty$**.** Since $\|Th_\alpha\|_{L^\infty} \le C_1\|h_\alpha\|_{L^{p_1}}$,

$$\begin{aligned}\|Th_\alpha\|_{L^\infty}^{p_1} &\le C_1^{p_1}\|h_\alpha\|_{L^{p_1}}^{p_1} = C_1^{p_1}p_1\int_0^\alpha t^{p_1-1}\lambda_f(t)\,dt \\ &\le C_1^{p_1}p_1\alpha^{p_1-p}\int_0^\alpha t^{p-1}\lambda_f(t)\,dt \le C_1^{p_1}\frac{p_1}{p}\alpha^{p_1-p}\|f\|_{L^p}^p.\end{aligned} \tag{1.9}$$

We take $\alpha = (\frac{\beta}{\kappa})^\sigma$, where $\kappa = C_1(\frac{p_1}{p}\|f\|_{L^p}^p)^{1/p_1}$ and $\sigma = \frac{p_1}{p_1-p} = \frac{p_0(q-q_0)}{q_0(p-p_0)} > 0$. The the estimate (1.9) is β^{p_1}. Since $\|Th_\alpha\|_{L^\infty} \le \beta$, the second term in the estimate (1.3) vanishes, and $\phi_1 = 0$ in the estimate (1.7). Then $\chi_0 = \chi_{\{t>(\beta/\kappa)^\sigma\}} = \chi_{\{\beta<\kappa t^{1/\sigma}\}}$, and we apply an analogue of (1.8) to get

$$\begin{aligned}\|Tf\|_{L^q}^q &\lesssim \int_0^\infty \left(\int_0^\infty \phi_0(t,\beta)\,dt\right)^{\frac{q_0}{p_0}} d\beta \le \left(\int_0^\infty \left(\int_0^\infty |\phi_0(t,\beta)|^{\frac{q_0}{p_0}}\,d\beta\right)^{\frac{p_0}{q_0}} dt\right)^{\frac{q_0}{p_0}} \\ &= \left(\int_0^\infty \left(\int_0^{\kappa t^{1/\sigma}} \beta^{q-q_0-1}\,d\beta\right)^{\frac{p_0}{q_0}} t^{p_0-1}\lambda_f(t)dt\right)^{\frac{q_0}{p_0}} = \frac{1}{q-q_0}\left(\kappa^{q-q_0}\int_0^\infty t^{p-1}\lambda_f(t)dt\right)^{\frac{q_0}{p_0}} = B_{p,q}\|f\|^{\frac{pq_0}{p_0}}.\end{aligned}$$

Case II (4): $p_0 < p_1 < \infty$ **and** $q_1 < q_0 = \infty$**.** Since $\|Tg_\alpha\|_{L^\infty} \le C_0\|g_\alpha\|_{L^{p_0}}$,

$$\|Tg_\alpha\|_{L^\infty}^{p_0} \le C_0^{p_0}\|g_\alpha\|_{L^{p_0}}^{p_0} \le C_0^{p_0}p_0\alpha^{p_0-p}\int_0^\alpha t^{p-1}\lambda_f(t)\,dt \le C_0^{p_0}\frac{p_0}{p}\alpha^{p_0-p}\|f\|_{L^p}^p. \tag{1.10}$$

We take $\alpha = (\frac{\beta}{\kappa})^\sigma$, where $\kappa = C_0(\frac{p_0}{p}\|f\|_{L^p}^p)^{1/p_0}$ and $\sigma = \frac{p_0}{p_0-p} = \frac{p_1(q_1-q)}{q_1(p_1-p)} < 0$, so the estimate (1.10) is β^{p_0}. Since $\|Tg_\alpha\|_{L^\infty} \le \beta$, the first term in the estimate (1.3) vanishes, and $\phi_0 = 0$ in the estimate (1.7). Then $\chi_1 = \chi_{\{t\le(\beta/\kappa)^\sigma\}} = \chi_{\{\beta\le\kappa t^{1/\sigma}\}}$, and we apply an analogue of (1.8) to get

$$\|Tf\|_{L^q}^q \lesssim \frac{1}{q-q_1}\left(\kappa^{q-q_1}\int_0^\infty t^{p-1}\lambda_f(t)dt\right)^{\frac{q_1}{p_1}} = B_{p,q}\|f\|^{\frac{pq_0}{p_0}}.$$

Then we complete the whole proof. □

Corollary 1.9 (Marcinkiewicz interpolation theorem)**.** *Let $1 \le p_0 < p_1 \le \infty$. If T is a sublinear operator of weak types (p_0, p_0) and (p_1, p_1), then T is of strong type (p, p) for each $p \in (p_0, p_1)$.*

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 <u>any open neighborhood U of a point x contains a compact neighborhood of x</u>. 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 $g \in C_0(X)$. $\square$

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 $C(X)$. $\square$

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. $\square$

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 $\operatorname{supp}(f) \subset \{x \in X : \sum_{j=1}^n g_k(x) > 0\}$. 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. $\square$

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 $K \subset U_n$. $\square$

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. $\square$

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 $f \geq 0$.

The positivity condition implies a continuity property of T.

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 $C_K = T\phi$. □

Remark. If we replace $C_c(X)$ by $C^\infty(X)$, this proposition still holds, because we can make $\phi \in C_c^\infty(U, [0, 1])$ in our proof by C^∞-Urysohn Lemma.

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 functional on $C_c(X)$, 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 $f = \sum_{j=1}^n f g_j$, and

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

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 $K = \mathrm{supp}(f)$. 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 $K_0 = \mathrm{supp}(f)$. 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

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

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

Remark. For any $f \in C_c(X)$ supported on K, we can take a mollification sequence $f^\epsilon = \phi_\epsilon * f \in C_c^\infty(X)$ that converges to f uniformly. Therefore, if T is a positive linear functional on $C_c^\infty(X)$, by the remark under Proposition 2.11, we can extend T to a linear functional on $C_c(X)$. Through this procedure, we can also relate each positive linear functional on a subspace of $C_c(X)$ containing $C_c^\infty(X)$ to a unique Radon measure on X.

2.3 Regularity and Approximation 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. □

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. □

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 $F \subset E$ 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. □

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 $1 \le p < \infty$.*

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 $\chi_K \le f \le \chi_U$. Then $\|\chi_E - f\|_p^p \le \mu(U \backslash K) \le \epsilon$, and we are done. □

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 $\|\phi\|_\infty \le \|f\|_\infty$.*

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 $A_n = \{0 \le |f| \le n\}$, 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 $\mu(\{\phi \neq f\}) < \epsilon$. □

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 $\phi \ge 0$, 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, $\nu(K) = \int_K \phi \, d\mu \le \int_X \phi \, d\mu < \infty$.
- 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. □

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 $\mu \mapsto \|\mu\|$ 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 μ_R^+, μ_R^-, μ_I^+, μ_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 $\nu_f = \nu$. ☐

2.5 Lebesgue Decomposition for Radon Measures on $\mathbb{R}^n$

In this section, we work in the Euclidean space $(\mathbb{R}^n, m)$, which is a locally compact, Hausdorff and second countable space. According to Theorem 2.19, if a Borel measure μ on $\mathbb{R}^n$ is finite on compact sets, it is a Radon measure. By Lebesgue decomposition theorem, μ has a unique decomposition

$$\mu = \rho + \nu,$$

where

- ρ is *absolutely continuous* with respect to the Lebesgue measure m, written $\rho \ll m$, i.e. $\rho(E) = 0$ for all Borel sets E with $m(E) = 0$.
- ν and the Lebesgue measure m are *mutually singular*, written $\nu \perp m$, i.e. there is a Borel set A such that $m(\mathbb{R}^n \backslash A) = \nu(A) = 0$.

Clearly, both ρ and ν are Radon measures on $\mathbb{R}^n$. The following theorem gives a further decomposition of μ.

Theorem 2.29 (Lebesgue decomposition for Radon measure on $\mathbb{R}^n$)**.** *If μ is a Radon measure on $\mathbb{R}^n$, there exists a locally integrable function $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ and a Radon measure $\nu \perp m$ such that*

$$\mu(E) = \int_E f\, dm + \nu(E), \quad E \in \mathscr{B}(\mathbb{R}^n). \tag{2.5}$$

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

$$\lim_{r \to 0^+} \frac{\mu(B(x,r))}{m(B(x,r))} = f(x). \tag{2.6}$$

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

Lemma 2.30 (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, $3B_{j_0}' \supset B_i$. $\square$

Lemma 2.31. *If ν is a Radon measure on $\mathbb{R}^n$, and $\nu \perp m$, then for almost every $x \in \mathbb{R}^n$,*

$$\lim_{r \to 0^+} \frac{\nu(B(x,r))}{m(B(x,r))} = 0. \tag{2.7}$$

Proof. We take the Borel set A such that $m(\mathbb{R}^n \backslash A) = \nu(A) = 0$, and define

$$E_k = \left\{ x \in A : \limsup_{r \to 0^+} \frac{\nu(B(x,r))}{m(B(x,r))} > \frac{1}{k} \right\}, \quad k = 1, 2, \cdots.$$

By outer regularity of ν, for any $\epsilon > 0$, we can find an open set $U \supset A$ with $\nu(U) < \epsilon$. By definition of E_k, for each $x \in E_k$, we can take a ball $B(x, r_x) \subset$ such that

$$\frac{\nu(B(x, r_x))}{m(B(x, r_x))} > \frac{1}{k}. \tag{2.8}$$

We take a compact subset $K \subset E_k$, then K is covered by finitely many such balls. By Vitali covering lemma [Lemma 2.30], we can further take finitely many disjoint balls $B(x_1, r_1), \cdots, B(x_N, r_N)$ such that

$$K \subset \bigcup_{j=1}^{N} B(x_j, 3r_j).$$

Applying the estimate (2.8), we have

$$m(K) \leq \sum_{j=1}^{N} m(B(x_j, 3r_j)) = 3^n \sum_{j=1}^{N} m(B(x_j, r_j)) < 3^n k \sum_{j=1}^{N} \nu(B(x_j, r_j)) \leq 3^n k \nu(U) \leq 3^n k\epsilon.$$

Since the compact $K \subset E_k$ is arbitrary, by inner regularity of the Lebesgue measure, $m(F_k) \leq 3^n k\epsilon$. Also, since $\epsilon > 0$ is arbitrary, $m(F_k) = 0$ for all $k \in \mathbb{N}$. Hence

$$\left\{ x \in A : \limsup_{r \to 0^+} \frac{\nu(B(x,r))}{m(B(x,r))} > 0 \right\} = \bigcup_{k=1}^{\infty} E_k$$

has Lebesgue measure zero. Since $m(\mathbb{R}^n \backslash A) = 0$, the limit (2.7) holds for m-a.e. $x \in \mathbb{R}^n$. □

Proof of Theorem 2.29. We take the Lebesgue decomposition $\mu = \rho + \nu$, where $\rho \ll m$ and $\nu \perp m$. By Radon Nikodym theorem, there exists $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ such that

$$\rho(E) = \int_E f \, dm, \quad \forall E \in \mathscr{B}(\mathbb{R}^n).$$

Then f satisfies the identity (2.5). The second result (2.6) is an immediate consequence of the Lebesgue differentiation theorem [Theorem 3.3] and Lemma 2.31. □

Remark. The locally integrable function f satisfying (2.6) is also called the derivative or *density* of the Radon measure μ with respect to the Lebesgue measure m.

3 The Hardy-Littlewood Maximal Inequality and Differentiation

3.1 The Hardy-Littlewood Maximal Inequality

In this section, we work in the Euclidean space $\mathbb{R}^n$ with the Lebesgue measure m. For a locally integrable function $f \in L^1_{\text{loc}}(\mathbb{R}^n)$, we define the local average

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

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

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

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

Theorem 3.1 (Hardy-Littlewood maximal inequality, weak type)**.** *The Hardy-Littlewood operator M is of weak type $(1,1)$. In other words, there exists a constant $C_n > 0$ such that for all $f \in L^1(\mathbb{R}^n)$ and all $\lambda > 0$,*

$$m(\{Mf \geq \lambda\}) \leq \frac{C_n}{\lambda} \|f\|_{L^1}. \tag{3.1}$$

Remark. The inequality (3.1) may look a bit stricter than the condition $[Mf]_1 \leq C_n \|Mf\|_{L^1(\mathbb{R}^n)}$ of weak type $(1,1)$. But, as we will see, the two assertions are indeed equivalent.

Proof of Hardy-Littlewood maximal inequality [Theorem 3.1]. We will show that for all $f \in L^1(\mathbb{R}^n)$,

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

Noticing that $m(\{Mf \geq \lambda\}) \leq m(\{Mf > \lambda - \epsilon\}) \leq 3^n(\lambda-\epsilon)^{-1}\|f\|_{L^1}$ for sufficiently small $\epsilon > 0$, the desired inequality (3.1) follows by perturbing $\epsilon \downarrow 0$.

Using the inner regularity of the Lebesgue measure, it suffices to show that $m(K) \leq 3^n \lambda^{-1} \|f\|_{L^1}$ 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 of K a finite subcollection that covers K. By Vitali covering lemma [Lemma 2.30], we take a further collection of disjoint balls $B_1, B_2, \cdots, B_k$ such that $K \subset \bigcup_{j=1}^k B_j$. Consequently,

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

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

Theorem 3.2 (Hardy-Littlewood maximal inequality, strong type)**.** *Let $1 < p \leq \infty$. The Hardy-Littlewood operator M is of strong type p. That is, there exists a constant $C_{n,n} > 0$ such that for all $f \in L^p(\mathbb{R}^n)$,*

$$\|Mf\|_{L^p} \leq C_{p,n} \|f\|_{L^p}.$$

Proof. The Hardy-Littlewood operator M is sublinear and of weak type 1. By definition of Mf, we also have $\|Mf\|_{L^\infty} \leq \|f\|_{L^\infty}$ when f is a.e. bounded. Hence M is of strong type ∞, and is of strong type (p,p) for each $1 < p \leq \infty$ by Marcinkiewicz interpolation theorem [Corollary 1.9] $\square$

3.2 The Lebesgue Differentiation Theorem and a.e. Differentiability

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

Theorem 3.3 (Lebesgue differentiation theorem)**.** *Let $f \in L^1_{\text{loc}}(\mathbb{R}^n)$. For almost every $x \in \mathbb{R}^n$,*

$$\lim_{r\to 0^+} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y) - f(x)|\, dy = 0. \tag{3.2}$$

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

$$\lim_{r\to 0^+} (A_r f)(x) = \lim_{r\to 0^+} \frac{1}{m(B(x,r))} \int_{B(x,r)} f(y)\, dy = f(x) \tag{3.3}$$

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

Remark. Let f be a measurable function on $\mathbb{R}^n$. A point $x \in \mathbb{R}^n$ is said to be a *Lebesgue point of f* if the identity (3.2) holds. The Lebesgue differentiation theorem implies that if $f \in L^1_{\text{loc}}(\mathbb{R}^n)$, then almost every point in $\mathbb{R}^n$ is a Lebesgue point of f.

Proof. We first prove the result for $g \in C_c(\mathbb{R}^n)$. If $x \in \mathbb{R}^n$ 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 (3.2) holds for all continuous functions with compact support.

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

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

Let $\lambda > 0$. Then

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

By weak L^1 Hardy-Littlewood maximal inequality [Theorem 3.1] and Markov inequality,

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

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

$$\lim_{r\to 0^+} \frac{m(E \cap B(x,r))}{m(B(x,r))} = \begin{cases} 1, & \text{if } x \in E, \\ 0, & \text{if } x \notin E. \end{cases}. \tag{3.4}$$

Remark. Let $E \subset \mathbb{R}^n$. A point $x \in \mathbb{R}^n$ is said to be a *density point of E* if the

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

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 identity (3.4) is a special case of (3.3) when $f = \chi_E$. □

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

Theorem 3.5 (Fundamental Theorem of Calculus)**.** *Let $F : \mathbb{R} \to \mathbb{C}$ be an absolutely continuous function. Then F is almost everywhere differentiable, and the derivative $f = F'$ satisfies $f \in L^1_{\text{loc}}(\mathbb{R})$, and*

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

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

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

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 $F' = f$. □

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

Proof. Since the differentiability is a local property, we may assume that f is Lipschitz continuous on $\mathbb{R}^n$. From the one-dimensional case, we know that for each unit vector $|v| = 1$, the directional derivative $f_v = \frac{\partial f}{\partial v}$ exists almost everywhere. In particular, the partial derivatives $(\frac{\partial f}{\partial x_j})_{j=1}^n$ exist almost everywhere. We write $G = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\right)$. We show that for almost every $x \in \mathbb{R}^n$,

$$\lim_{|h|\to 0} \frac{f(x+h) - f(x) - G \cdot h}{|h|} = 0.$$

This implies that f is almost everywhere differentiable, and the gradient $\nabla f = G$.

Step I. For each $|v| = 1$, we claim that $f_v = G \cdot v$ almost everywhere. We take a test function $\phi \in C_c^\infty(\mathbb{R}^n)$. Since f is Lipschitz, by Lebesgue dominated convergence theorem,

$$\lim_{t\to 0} \int_{\mathbb{R}^n} \frac{f(x+tv) - f(x)}{t} \phi(x)\, dx = \int_{\mathbb{R}^n} f_v(x)\phi(x)\, dx.$$

Since ϕ is smooth, $\phi_v = \nabla\phi \cdot v$. Applying integration by parts, we have

$$\begin{aligned}\int_{\mathbb{R}^n} f_v(x)\phi(x)\, dx &= \lim_{t\to 0} \int_{\mathbb{R}^n} \frac{f(x+tv) - f(x)}{t} \phi(x)\, dx = \lim_{t\to 0} \int_{\mathbb{R}^n} f(x) \frac{\phi(x-tv) - \phi(x)}{t}\, dx \\ &= -\int_{\mathbb{R}^n} f(x)\phi_v(x)\, dx = -\int_{\mathbb{R}^n} \sum_{j=1}^n f(x) v_j \frac{\partial \phi}{\partial x_j}(x)\, dx = \int_{\mathbb{R}^n} \sum_{j=1}^n v_j \frac{\partial f}{\partial x_j}(x)\phi(x)\, dx.\end{aligned}$$

Therefore $\int_{\mathbb{R}^n} (f_v - G \cdot v)\phi\, dm = 0$ for all $\phi \in C_c^\infty(\mathbb{R}^n)$, and $f_v = G \cdot v$ a.e.. For each $|v| = 1$, we write

$$A_v = \{x \in \mathbb{R}^n : G(x) \text{ and } f_v(x) \text{ exists, and } f_v(x) = G(x) \cdot v\}.$$

Step II. We take a countable dense subset (v_k) of the unit sphere $\{|v| = 1\}$. By Step I, $\mu(A_{v_k}^c) = 0$ for all $k \in \mathbb{N}$, and $\mu(A^c) = 0$, where we take $A = \bigcap_{k=1}^\infty A_{v_k}$. We claim that f is differentiable at all $x \in A$. Since f is Lipschitz continuous, there exists a constant $K > 0$ such that $|f(x) - f(y)| \le K|x - y|$, and all partial derivatives are bounded by K. We take $h \neq 0$ in $\mathbb{R}^n$. Then for all $k \in \mathbb{N}$,

$$\begin{aligned}&\frac{|f(x+h) - f(x) - G \cdot h|}{|h|} \\ &= \frac{|f(x+|h|v) - f(x+|h|v_k)|}{|h|} + \frac{|f(x+|h|v_k) - f(x) - G \cdot |h|v_k|}{|h|} + \frac{|G \cdot |h|(v_k - v)|}{|h|} \\ &= \frac{|f(x+|h|v) - f(x+|h|v_k)|}{|h|} + \left|\frac{f(x+|h|v_k) - f(x)}{|h|} - G \cdot v_k\right| + |G \cdot (v_k - v)| \\ &\le K|v - v_k| + \left|\frac{f(x+|h|v_k) - f(x)}{|h|} - G \cdot v_k\right| + K\sqrt{n} \cdot |v_k - v|\end{aligned}$$

where $v = h/|h|$ is a unit vector. By density of (v_k) in the unit sphere, for each $\epsilon > 0$, we take v_k such that $|v_k - v| < \epsilon$. Then for all $x \in A$,

$$\lim_{|h|\to 0} \frac{|f(x+h) - f(x) - G \cdot h|}{|h|} \le K|v - v_k| + |f_v - G \cdot v_k| + K\sqrt{n} \cdot |v_k - v| \le K(1 + \sqrt{n})\epsilon,$$

where the last inequality follows from $x \in A_{v_k}$. Since $\epsilon > 0$ is arbitrary, the above limit is zero. Therefore f is differentiable at x, and the gradient $\nabla f = G$. $\square$

Remark. Since each component of ∇f is the limit of bounded measurable functions, $|\nabla f| \in L^1_{\text{loc}}(\mathbb{R}^n)$.

Finally we record some technical facts we will use later.

Theorem 3.7 (Differentiability on Level Sets)**.** *The following statements hold:*

(i) Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz continuous function, and

$$Z = \{x \in \mathbb{R}^n : f(x) = 0\}.$$

Then $Df(x) = 0$ for m-a.e. $x \in Z$.

(ii) Let $f, g : \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitz continuous functions, and

$$Y = \{x \in \mathbb{R}^n : g(f(x)) = x\},$$

Then $Dg(f(x))Df(x) = \mathrm{Id}$ for m-a.e. $x \in Y$.

Proof. (i) We may assume $m = 1$. Choose $x \in Z$ so that $\nabla f(x)$ exists, and

$$\lim_{r \downarrow 0} \frac{m(Z \cap B(x,r))}{m(B(x,r))} = 1 \quad \text{for } m\text{-a.e. } x \in Z. \tag{3.5}$$

If $\nabla f(x) \neq 0$, we put $V = \{\xi \in \partial B(0,1) : \nabla f(x)^\top \xi \geq \frac{1}{2}|\nabla f(x)|\}$. Then for all $\xi \in V$,

$$0 = \lim_{t \downarrow 0} \frac{f(x + t\xi) - t\nabla f(x)^\top \xi}{t} = \lim_{t \downarrow 0} \left(\frac{f(x + t\xi)}{t} - \frac{1}{2}|\nabla f(x)| \right).$$

Since $|f(x)| > 0$, there exists $t_0 > 0$ such that $f(x + t\xi) > 0$ for all $t \in (0, t_0)$ and $\xi \in V$. Then

$$\lim_{r \downarrow 0} \frac{m(Z \cap B(x,r))}{m(B(x,r))} < 1,$$

which contradicts (3.5). Hence $m\{x \in Z : |\nabla f(x)| \neq 0\} = 0$.

(ii) We define $A = \{x \in \mathbb{R}^n : Df(x) \text{ exists}\}$, $B = \{x \in \mathbb{R}^n : Dg(x) \text{ exists}\}$, and $X = Y \cap A \cap f^{-1}(B)$. Since $x \in Y \backslash f^{-1}(B)$ implies $f(x) \in \mathbb{R}^n \backslash B$, and $x = g(f(x)) \in g(\mathbb{R}^n \backslash B)$, we have $Y \backslash X \subset (\mathbb{R}^n \backslash A) \cup g(\mathbb{R}^n \backslash B)$. By Rademacher's theorem, we have $m(Y \backslash X) = 0$.

Finally, for each $x \in X$, both $Df(x)$ and $Dg(x)$ exist, and

$$D(g \circ f)(x) = Dg(f(x))Df(x)$$

exists. Since $(g \circ f)(x) - x = 0$ a.e. on Y, and assertion (i) implies $D(g \circ f)(x) = \mathrm{Id}$ a.e. on Y. $\square$

3.3 Second Differentiability of Convex Functions

In this section, we discuss the differentiation of convex functions on Euclidean spaces. Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is called *convex* if for all $x, y \in \mathbb{R}^n$ and $0 \le \lambda \le 1$,

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y).$$

Proposition 3.8. *Every convex function on $\mathbb{R}^n$ is locally Lipschitz continuous.*

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function.

Step I. We first prove that f is locally bounded. We consider the compact hypercube $Q = [-N, N]^n$, with vertices $(x_k)_{k=1}^{2^n}$. Then every $x \in Q$ is a convex combination $x = \sum_{k=1}^{2^n} \lambda_k x_k$ of the vertices, and

$$f(x) \le \sum_{k=1}^{2^n} \lambda_k f(x_k) \le M := \max_{1 \le k \le 2^n} f(x_k) < \infty.$$

Then $\sup_{x \in Q} f(x) \le M$. To derive a lower bound, note that for every $x \in Q$,

$$f(0) \le \frac{1}{2} f(x) + \frac{1}{2} f(-x) \le \frac{1}{2} f(x) + \frac{1}{2} M.$$

Hence $\inf_{x \in Q} f(x) \ge 2f(0) - M$.

Step II. Now we prove the local Lipschitz continuity of f. Fix $x, y \in \overline{B(0,N)}$ with $x \ne y$, where $N > 0$ and $\overline{B(0,N)}$ is the closed ball of radius N centered at 0. We choose $\mu > 0$ such that $z = x + \mu(y - x)$ satisfies $|z| = 3N$. Then $\mu = \frac{|z-x|}{|y-x|} \ge 1$, and

$$f(y) = f\left(\frac{1}{\mu} z + \left(1 - \frac{1}{\mu}\right) x\right) \le f(x) + \frac{f(z) - f(x)}{\mu} \le f(x) + \frac{2}{\mu} \sup_{\xi \in B(0,3N)} |f(\xi)|.$$

Since $|z - x| \ge 2N$, we obtain

$$f(y) - f(x) \le \frac{2}{\mu} \sup_{\xi \in B(0,3N)} |f(\xi)| = \frac{|y-x|}{N} \sup_{\xi \in B(0,3N)} |f(\xi)|.$$

Interchanging x and y, the same estimate holds for $f(x) - f(y)$. Hence f is locally Lipschitz continuous. $\square$

Remark. According to Proposition 3.8 and Rademacher's theorem [Theorem 3.6], every convex function is almost everywhere differentiable. In this section, we step further and deal with the second differentiability. We begin by discussing some properties of derivatives of convex functions.

Lemma 3.9. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function.*

(i) If f is differentiable at x, then

$$f(y) \ge f(x) + \nabla f(x) \cdot (y - x). \tag{3.6}$$

(ii) If, in addition, $f \in C^2(\mathbb{R}^n)$, then $\nabla^2 f \succeq 0$ on $\mathbb{R}^n$.

Proof. (i) For each $y \in \mathbb{R}^n$ and $0 < \lambda < 1$, by convexity of f, we have

$$\frac{f(x + \lambda(y-x)) - f(x)}{\lambda} \le f(y) - f(x).$$

Letting $\lambda \to 0$, we have

$$f(y) \ge f(x) + \nabla f(x) \cdot (y - x).$$

(ii) By Taylor's theorem,

$$f(y) = f(x) + \nabla f(x) \cdot (y - x) + (y - x)^\top \int_0^1 (1 - t)\nabla^2 f(x + t(y - x))\, dt \cdot (y - x)$$

Then the estimate 3.6 implies

$$(y - x)^\top \int_0^1 (1 - t)\nabla^2 f(x + t(y - x))\, dt \cdot (y - x) \geq 0.$$

Hence given any $\xi \in \mathbb{R}^n$, we set $y = x + s\xi$ with $s > 0$. Then the above inequality becomes

$$\xi^\top \int_0^1 (1 - t)\nabla^2 f(x + st\xi)\, dt \cdot \xi \geq 0.$$

Letting $s \to 0$, we have

$$\xi^\top \nabla^2 f(x) \cdot \xi \geq 0.$$

This proves assertion (ii). $\square$

Indeed, for any convex function, we can find its second derivatives in the distributional sense.

Theorem 3.10 (Second derivatives of convex functions as measures)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ to a convex function. Then there exist signed Radon measures $\mu^{ij} = \mu^{ji}$ such that for all functions $\phi \in C_c^2(\mathbb{R}^n)$,*

$$\int_{\mathbb{R}^n} f\phi_{x_i x_j}\, dx = \int_{\mathbb{R}^n} \phi\, d\mu^{ij}, \quad i, j = 1, 2, \cdots, n.$$

Proof. We define $f^\epsilon = \eta_\epsilon \cdot f \in C^\infty(\mathbb{R}^n)$, where η_ϵ is the standard mollifier. For any $v \in \mathbb{R}^n$ with $|v| = 1$,

$$\sum_{i,j=1}^n \int_{\mathbb{R}^n} f^\epsilon \phi_{x_i x_j} v_i v_j\, dx = \int_{\mathbb{R}^n} \phi \sum_{i,j=1}^n f^\epsilon_{x_i x_j} v_i v_j\, dx \geq 0, \quad \phi \in C_c^2(\mathbb{R}^n) \text{ and } \phi \geq 0.$$

Letting $\epsilon \downarrow 0$, we have

$$T_v \phi := \sum_{i,j=1}^n \int_{\mathbb{R}^n} f\phi_{x_i x_j} v_i v_j\, dx.$$

Thus we define a positive linear functional T_v on $C_c^2(\mathbb{R}^n)$. According to the Remark under the Theorem 2.13, there exists a Radon measure μ^v on $\mathbb{R}^n$ such that

$$T_v \phi = \int_{\mathbb{R}^n} \phi\, d\mu^v, \quad \text{for all } \phi \in C^2(\mathbb{R}^n).$$

For each $i = 1, 2, \cdots, n$, we define $\mu^{ii} = \mu^{e_i}$. If $i \neq j$, we set $v = \frac{e_i + e_j}{\sqrt{2}}$. Then

$$\begin{aligned}
\int_{\mathbb{R}^n} f\phi_{x_i x_j}\, dx &= \int_{\mathbb{R}^n} f \left(\frac{\phi_{ii} + \phi_{jj}}{2} + \phi_{ij} \right) dx - \frac{1}{2} \int_{\mathbb{R}^n} f\phi_{x_i x_i}\, dx - \frac{1}{2} \int_{\mathbb{R}^n} f\phi_{x_j x_j}\, dx \\
&= \sum_{k,l=1}^n \int_{\mathbb{R}^n} f\phi_{kl} v_k v_l\, dx - \frac{1}{2} \int_{\mathbb{R}^n} f\phi_{x_i x_i}\, dx - \frac{1}{2} \int_{\mathbb{R}^n} f\phi_{x_j x_j}\, dx \\
&= \int_{\mathbb{R}^n} \phi\, d\mu^v - \frac{1}{2} \int_{\mathbb{R}^n} \phi\, d\mu^{ii} - \frac{1}{2} \int_{\mathbb{R}^n} \phi\, d\mu^{jj} = \int_{\mathbb{R}^n} \phi\, d\mu^{ij},
\end{aligned}$$

where we set $\mu^{ij} = \mu^v - \frac{1}{2}\mu^{ii} - \frac{1}{2}\mu^{ij}$. Then we complete the proof. $\square$

Remark. According to the Lebesgue decomposition theorem, every signed Radon measure μ^{ij} has a unique decomposition $\mu^{ij} = \mu^{ij}_{\mathrm{ac}} + \mu^{ij}_{\mathrm{s}}$, where $\mu^{ij}_{\mathrm{ac}} \ll m$ and $\mu^{ij}_{\mathrm{s}} \perp m$. We write

$$M = (\mu^{ij})^n_{i,j=1}, \quad M_{\mathrm{ac}} = (\mu^{ij}_{\mathrm{ac}})^n_{i,j=1}, \quad M_{\mathrm{s}} = (\mu^{ij}_{\mathrm{s}})^n_{i,j=1}.$$

By Radon-Nikodym theorem, we define $f^{ij} = d\mu^{ij}_{\mathrm{ac}}/dm$ to be the density of the absolute continuous part of μ^{ij} with respect to the Lebesgue measure m. Then

$$D^2 f = \begin{pmatrix} f^{11} & f^{12} & \dots & f^{1n} \\ f^{21} & f^{22} & \dots & f^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f^{n1} & f^{n2} & \dots & f^{nn} \end{pmatrix}$$

is a matrix valued function, and every element f^{ij} is locally integrable. According to the Theorem 2.29, we have the decomposition

$$M(E) = \int_E D^2 f \, dm + M_{\mathrm{s}}(E).$$

In fact, a convex function has not only distributional second derivatives as Radon measures, but also classical derivatives almost everywhere. The main result of this section is presented below.

Theorem 3.11 (Alexandrov Theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ to a convex function. Then f is almost everywhere twice differentiable. More precisely, there exists $\nabla^2 f : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ such that for almost every $x \in \mathbb{R}^n$,*

$$\lim_{y\to x} \frac{1}{|y-x|^2} \left(f(y) - f(x) - \nabla f(x) \cdot (y-x) - \frac{1}{2}(y-x)^\top \nabla^2 f(x)(y-x) \right) = 0. \tag{3.7}$$

To prove this theorem, we need some maximal inequalities concerning convex functions in a ball.

Lemma 3.12. *If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, there exists a constant $C_n > 0$ such that for each ball $B(x,r) \subset \mathbb{R}^n$,*

$$\sup_{y\in B(x,\frac{r}{2})} |f(y)| \le \frac{C_n}{m(B(x,r))} \int_{B(x,r)} |f(y)| \, dy, \tag{3.8}$$

and

$$\operatorname*{ess\,sup}_{y\in B(x,\frac{r}{2})} |\nabla f(y)| \le \frac{C_n}{r \cdot m(B(x,r))} \int_{B(x,r)} |f(y)| \, dy. \tag{3.9}$$

Proof. **Step I.** We first prove (3.8) for $f \in C^2(\mathbb{R}^n)$. Given $B(x,r) \subset \mathbb{R}^n$, we fix $z \in B(x, \frac{r}{2})$. Then

$$f(y) \ge f(z) + \nabla f(z) \cdot (y-z).$$

We integrate this inequality with respect to y over $B(z, \frac{r}{2})$ to obtain

$$f(z) \le \frac{1}{m(B(z,\frac{r}{2}))} \int_{B(z,\frac{r}{2})} f(y) \, dy \le \frac{2^n}{m(B(x,r))} \int_{B(x,r)} |f(y)| \, dy. \tag{3.10}$$

Next, we choose $\phi \in C^\infty_c(\mathbb{R}^n)$ such that $0 \le \phi \le 1$ on $\mathbb{R}^n$, $\phi = 1$ on $B(0, \frac{1}{2})$ and $\phi = 0$ outside $B(0,1)$, and write $M_1 = \sup_{|y|\le 1} |\nabla\phi(y)|$. Then the function $\phi_{x,r}(y) = \phi(\frac{y-x}{r})$ satisfies

$$\begin{cases} 0 \le \phi_{x,r} \le 1, \quad |\nabla\phi_{x,r}| \le \frac{M_1}{r}, \\ \phi = 1 \text{ on } B(x, \frac{r}{2}), \quad \phi = 0 \text{ on } \mathbb{R}^n \backslash B(x,r). \end{cases}$$

We multiply by $\phi_{x,r}(y)$ the estimate $f(z) \geq f(y) + \nabla f(y) \cdot (z-y)$ and integrate with respect to y on $B(x,r)$:

$$\begin{aligned} f(z) \int_{B(x,r)} \phi_{x,r}(y)\, dy &\geq \int_{B(x,r)} f(y)\phi_{x,r}(y)\, dy + \int_{B(x,r)} \phi_{x,r}(y) \nabla f(y) \cdot (z-y)\, dy \\ &= \int_{B(x,r)} f(y) \left(\phi_{x,r}(y) - \nabla \cdot \phi_{x,r}(y)(z-y) \right) dy \geq -M_1 \int_{B(x,r)} |f(y)|\, dy. \end{aligned}$$

This inequality implies

$$f(z) \geq -\frac{M_1}{m(B(x,r))} \int_{B(x,r)} |f(y)|\, dy. \tag{3.11}$$

Since $z \in B(x, \frac{r}{2})$ is arbitrary, we combine (3.10) and (3.11) to obtain the estimate (3.8).

Step II. More generally, for a convex function $f : \mathbb{R}^n \to \mathbb{R}$, we define $f^\epsilon = \eta_\epsilon * f$, where $\epsilon > 0$ and $\eta_\epsilon(y) = \epsilon^{-n}\eta(\epsilon^{-1}y)$ is the standard mollifier. Clearly $f^\epsilon \in C^\infty(\mathbb{R}^n)$, and f^ϵ is convex, since

$$\begin{aligned} f^\epsilon(\lambda x + (1-\lambda)y) &\leq \int_{\mathbb{R}^n} f(\lambda x + (1-\lambda)y - z)\eta_\epsilon(z)\, dz \\ &\leq \lambda \int_{\mathbb{R}^n} f(x-z)\eta_\epsilon(z)\, dz + (1-\lambda) \int_{\mathbb{R}^n} f(y-z)\eta_\epsilon(z)\, dz \leq \lambda f^\epsilon(x) + (1-\lambda) f^\epsilon(y). \end{aligned}$$

Applying the assertion (i) for C^2 functions, we have

$$\sup_{z \in B(x,\frac{r}{2})} |f^\epsilon(z)| \leq \frac{C_n}{m(B(x,r))} \int_{B(x,r)} |f^\epsilon(y)|\, dy$$

for each ball $B(x,r) \subset \mathbb{R}^n$. Since f is locally Lipschitz continuous, $f^\epsilon \to f$ uniformly on $B(x,r)$ as $\epsilon \to 0$, which gives the same estimate (3.8) for f.

Step III. For each $z \in B(x, \frac{r}{2})$ such that $\nabla f(z)$ exists, define

$$S_z = \left\{ y \in \mathbb{R}^n : \frac{r}{4} \leq |y-z| \leq \frac{r}{2},\ \nabla f(z) \cdot (y-z) \geq \frac{1}{2} |\nabla f(z)|\, |y-z| \right\}.$$

Then $m(S_z) \geq Ar^n$, where $A > 0$ is a constant only depending on n. Using the estimate (3.6), we have

$$f(y) \geq f(z) + \frac{r}{8} |\nabla f(z)|.$$

Integrating with respect to y over $S(z)$ gives

$$|\nabla f(z)| \leq \frac{8}{r} \cdot \frac{1}{m(S_z)} \int_{S_z} |f(y) - f(z)|\, dy \leq \frac{8}{Ar^{n+1}} \int_{B(x,r)} |f(y) - f(z)|\, dy.$$

This estimate and (3.8) complete the proof of assertion (i) for convex functions f. □

Proof of Theorem 3.11. The proof has four steps.

Step I. According to the Lebesgue differentiation theorem [Theorem 3.3] and Rademacher's theorem [Theorem 3.6], for almost every $x \in \mathbb{R}^n$, the gradient $\nabla f(x)$ exists and satisfies

$$\lim_{r \to 0^+} \frac{1}{m(B(x,r))} \int_{B(x,r)} |\nabla f(y) - \nabla f(x)|\, dy = 0, \tag{3.12}$$

and

$$\lim_{r \to 0^+} \frac{1}{m(B(x,r))} \int_{B(x,r)} \|D^2 f(y) - D^2 f(x)\|_{\mathrm{F}}\, dy = 0. \tag{3.13}$$

By singularity of μ_{s}^{ij} and Lemma 2.31, for almost every $x \in \mathbb{R}^n$, the measures (μ_s^{ij}) satisfy

$$\lim_{r\to 0^+} \frac{|\mu_{\mathrm{s}}^{ij}|(B(x,r))}{m(B(x,r))} = 0, \quad i,j = 1,2,\cdots,n. \tag{3.14}$$

We fix such a point x, and we also assume $x = 0$ for simplicity, since our proof is adapted to the convex function $(\tau_x f)(y) = f(y-x)$. We choose $r > 0$ and let $f^\epsilon = \eta_\epsilon * f$. For $y \in B(0,r)$, by Taylor's theorem,

$$\begin{aligned} f^\epsilon(y) &= f^\epsilon(0) + \nabla f^\epsilon(0)\cdot y + \int_0^1 (1-t) y^\top \nabla^2 f^\epsilon(ty) y\, dt \\ &= f^\epsilon(0) + \nabla f^\epsilon(0)\cdot y + \frac{1}{2} y^\top D^2 f(0) y + \int_0^1 (1-t) y^\top \left[\nabla^2 f^\epsilon(ty) - D^2 f(0)\right] y\, dt. \end{aligned}$$

Step II. For any function $\phi \in C_c^2(B(0,r))$ with $|\phi| \le 1$, we multiply the equation above by ϕ and take the average over $B(0,r)$. Then

$$\begin{aligned} &\frac{1}{m(B(0,r))} \int_{B(0,r)} \phi(y) \left(f^\epsilon(y) - f^\epsilon(0) - \nabla f^\epsilon(0)\cdot y - \frac{1}{2} y^\top D^2 f(0) y \right) dy \\ &= \frac{1}{m(B(0,r))} \int_{B(0,r)} \phi(y) \left(\int_0^1 (1-t) y^\top \left[\nabla^2 f^\epsilon(ty) - D^2 f(0)\right] y\, dt \right) dy \qquad (3.15) \\ &= \int_0^1 (1-t) \left(\frac{1}{m(B(0,r))} \int_{B(0,r)} \phi(y) y^\top \left[\nabla^2 f^\epsilon(ty) - D^2 f(0)\right] y\, dy \right) dt \qquad \text{(By Fubini's theorem)} \\ &= \int_0^1 \frac{1-t}{t^2} \left(\frac{1}{m(B(0,tr))} \int_{B(0,tr)} \phi\left(\frac{z}{t}\right) z^\top \left[\nabla^2 f^\epsilon(z) - D^2 f(0)\right] z\, dz \right) dt. \qquad \text{(Change the vairable } z = ty) \end{aligned}$$

To estimate the inner integral, we use integration by parts:

$$g_\epsilon(t) = \int_{B(0,tr)} \phi\left(\frac{z}{t}\right) z^\top \nabla^2 f^\epsilon(z) z\, dz = \int_{B(0,tr)} f^\epsilon(z) \sum_{i,j=1}^n \frac{\partial^2}{\partial z_i \partial z_j} \left(\phi\left(\frac{z}{t}\right) z_i z_j \right) dz.$$

Letting $\epsilon \to 0^+$, we obtain

$$\begin{aligned} \lim_{\epsilon\to 0^+} g_\epsilon(t) &= \int_{B(0,tr)} f(z) \sum_{i,j=1}^n \frac{\partial^2}{\partial z_i \partial z_j} \left(\phi\left(\frac{z}{t}\right) z_i z_j \right) dz \\ &= \sum_{i,j=1}^n \int_{B(0,tr)} \phi\left(\frac{z}{t}\right) z_i z_j\, d\mu^{ij} \qquad (3.16) \\ &= \sum_{i,j=1}^n \int_{B(0,tr)} f^{ij} \phi\left(\frac{z}{t}\right) z_i z_j\, dz + \sum_{i,j=1}^n \int_{B(0,tr)} \phi\left(\frac{z}{t}\right) z_i z_j\, d\mu_{\mathrm{s}}^{ij}. \end{aligned}$$

Furthermore, we have the following estimate:

$$\begin{aligned} \frac{g_\epsilon(t)}{t^2} &\le r^2 \int_{B(0,tr)} \|\nabla^2 f(z)\|_{\mathrm{F}}\, dz = r^2 \int_{B(0,tr)} \left\| \int_{\mathbb{R}^n} \nabla^2 \eta_\epsilon(z-y) f(y)\, dy \right\|_{\mathrm{F}} dz \\ &= r^2 \int_{B(0,tr)} \left\| \int_{B(z,\epsilon)} \eta_\epsilon(z-y)\, dM(y) \right\|_{\mathrm{F}} dz \\ &\le r^2 \int_{B(0,tr)} \sum_{i,j=1}^n \int_{B(z,\epsilon)} |\eta_\epsilon(z-y)|\, d|\mu^{ij}|(y)\, dz. \end{aligned}$$

By Fubini's theorem, we have

$$\begin{aligned}\frac{g_\epsilon(t)}{t^2} &\le r^2\sum_{i,j=1}^n\int_{B(0,tr+\epsilon)}\left(\int_{B(y,\epsilon)\cap B(0,tr)}|\eta_\epsilon(z-y)|\,dz\right)d|\mu^{ij}|(y)\\ &\le \frac{r^2}{\epsilon^n}\sum_{i,j=1}^n\int_{B(0,tr+\epsilon)}\left(\int_{B(y,\epsilon)\cap B(0,tr)}dz\right)d|\mu^{ij}|(y).\end{aligned}$$

Since (μ^{ij}) are Radon measures, we have $\sum_{i,j=1}^n|\mu^{ij}|(B(0,r+1))<\infty$

$$\frac{g_\epsilon(t)}{t^2}\le\frac{r^2\min\{\epsilon^n,t^nr^n\}}{\epsilon^n}\sum_{i,j=1}^n|\mu^{ij}|(B(0,tr+\epsilon))\le Cr^2\min\left\{1,\frac{t^nr^n}{\epsilon^n}\right\}\le Cr^2,$$

where C is a constant only depending on f, r and n. Hence we applying dominated convergence theorem to let $\epsilon\to0^+$ in (3.15), and plug-in (3.16):

$$\begin{aligned}&\frac{1}{m(B(0,r))}\int_{B(0,r)}\phi(y)\left(f(y)-f(0)-\nabla f(0)\cdot y-\frac12y^\top D^2f(0)y\right)dy\\ &=\int_0^1\frac{1-t}{t^2}\left(\frac{1}{m(B(0,tr))}\int_{B(0,tr)}\phi\left(\frac zt\right)z^\top[D^2f(z)-D^2f(0)]\,z\,dz\right)dt\\ &\quad+\sum_{i,j=1}^n\int_0^1\frac{1-t}{t^2}\left(\frac{1}{m(B(0,tr))}\int_{B(0,tr)}\phi\left(\frac zt\right)z_iz_j\,d\mu_{\mathrm s}^{ij}\right)dt\\ &\le Cr^2\int_0^1\left(\frac{1}{m(B(0,tr))}\int_{B(0,tr)}\|D^2f(z)-D^2f(0)\|_{\mathrm F}\,dz\right)dt+Cr^2\max_{1\le i,j\le n}\int_0^1\frac{|M_{\mathrm s}|(B(0,tr))}{(tr)^n}\,dt,\end{aligned}$$

where C is a constant depending on f,r and n only. According to the properties (3.13) and (3.14), and taking the supremum over all $\phi\in C_c^2(B(0,r))$ with $|\phi|\le1$, we have

$$\frac{1}{m(B(0,r))}\int_{B(0,r)}|h(y)|dy=o(r^2),\tag{3.17}$$

where

$$h(y)=f(y)-f(0)-\nabla f(0)\cdot y-\frac12y^\top D^2f(0)y.\tag{3.18}$$

Step III. We claim that there exists a constant C depending on f and n only, such that

$$\operatorname*{ess\,sup}_{x\in B(0,\frac r2)}|\nabla h(x)|\le\frac{C}{r\cdot m(B(0,r))}\int_{B(0,r)}|h(y)|\,dy+Cr\tag{3.19}$$

for all $r>0$. This estimate follows by applying (3.9) on the convex function $g(y)=h(y)+\frac12\|D^2f(0)\|_2|y|^2$.

Step IV. We fix $0<\epsilon,\eta<1$. By (3.17), for some $r_0>0$ depending on η and ϵ, we have

$$m(\{z\in B(0,r):|h(z)|\ge\epsilon r^2\})\le\frac{1}{\epsilon r^2}\int_{B(0,r)}|h(z)|\,dz\le\eta\cdot m(B(0,r))$$

for all $0<r<r_0$. Thus for each point $y\in B(0,\frac r2)$, there exists $z\in B(0,r)$ with $|y-z|\le\eta^{1/n}r$ such that $|h(z)|<\epsilon r^2$. If not, a contradiction arises from

$$m(\{z\in B(0,r):|h(z)|\ge\epsilon r^2\})\ge m(B(y,\eta^{1/n}r))=\eta\cdot m(B(0,r)).$$

By the estimate (3.19), for all $0 < r < r_0$ and all $y \in B(0, \frac{r}{2})$,

$$\begin{aligned}\frac{|h(y)|}{r^2} \leq \frac{|h(z)| + |h(y) - h(z)|}{r^2} &\leq \epsilon + \frac{\eta^{1/n}}{r} \operatorname*{ess\,sup}_{B(0,r)} |\nabla h| \\ &\leq \epsilon + C\eta^{1/n} \left(\frac{1}{r^2 \cdot m(B(0,r))} \int_{B(0,r)} |h(\xi)| \, d\xi + 1 \right) \leq \epsilon + C\eta^{1/n} \left(\eta\epsilon + 1 \right)\end{aligned}$$

Since $0 < \epsilon, \eta < 1$ are arbitrary, we have

$$\lim_{r \to 0^+} \frac{1}{r^2} \sup_{y \in B(0, \frac{r}{2})} |h(y)| = 0.$$

Recalling the definition (3.18) of h, we have

$$\lim_{r \to 0^+} \frac{1}{r^2} \sup_{y \in B(0, \frac{r}{2})} \left| f(y) - f(0) - \nabla f(0) \cdot y - \frac{1}{2} y^\top D^2 f(0) y \right| = 0,$$

which implies (3.7) for $x = 0$. The same estimate holds for every $x \in \mathbb{R}^n$ satisfying (3.12), (3.13) and (3.14), which concludes the proof. $\square$

4 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 $n \to \infty$.

4.1 The Mean Ergodic Theorem

We first discuss a general ergodic result for Banach spaces.

Theorem 4.1 (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

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

The proof of this theorem requires some lemmata.

Lemma 4.2. *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 4.3. *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 4.2 (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 4.2 (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 4.1. By Lemma 4.3, 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 4.2 (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 4.2 (iii), we have $\|A\| \le c$, and by Lemma 4.2 (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 4.4 (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

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

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 4.5 (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

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

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 4.6. *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 4.5,

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

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

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

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

4.2 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 4.7 (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

$$\mu(E_\alpha^f) \leq \frac{1}{\alpha} \int_{E_\alpha} f \, d\mu \leq \frac{1}{\alpha} \|f\|_{L^1}. \tag{4.1}$$

This result is a corollary of the following maximal inequality.

Lemma 4.8 (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 4.7. Define $g = f - \alpha$ and $Ug = g \circ T$ in Proposition 4.8. 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 4.8 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 (4.1), 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.

4.3 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 4.9 (Ergodic transformation)**.** Let $T : X \to X$ be a measure-preserving transformation on a measure space $(X, \mathscr{F}, \mu)$. Define the *invariant σ-algebra* of T by

$$\mathscr{I}_T = \left\{E \in \mathscr{F} : T^{-1}(E) = E\right\}.$$

The mapping T is said to be *μ-ergodic* if $\mathscr{I}_T$ is trivial, i.e. for each $E \in \mathscr{I}_T$, either $\mu(E) = 0$ or $\mu(X \backslash E) = 0$.

Following are some alternate characterizations of ergodicity.

Proposition 4.10. *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 $E_t = \{f \le t\}$ and $T^{-1}(E_t) = \{f \circ T \le t\}$ only differ by a μ-null set. By (ii), we have $\mu(\{f \le t\}) = 0$ or $\mu(\{f > t\}) = 0$. We take

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

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 $\mu(X\backslash E) = 0$. $\square$

Proposition 4.11. *A measurable function f on $(X, \mathscr{F})$ is T-invatiant if and only if f is $\mathscr{I}_T$-measurable.*

Proof. If f is T-invariant, then for all $\alpha \in \mathbb{R}$,

$$\{x \in X : f(x) > \alpha\} = \{x \in X : f(Tx) > \alpha\} = T^{-1}\{x \in X : f(x) > \alpha\} \in \mathscr{I}_T.$$

If f is $\mathscr{I}_T$-measurable, then for each $x \in X$,

$$x \in \{y \in X : f(y) = f(x)\} = T^{-1}\{y \in X : f(y) = f(x)\} = \{y \in X : f(Ty) = f(x)\}.$$

Hence $f(x) = f(Tx)$, and we complete the proof. $\square$

Now we are ready to introduce the main result.

Theorem 4.12 (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 $\bar{f} = \mathbb{E}[f|\mathscr{I}_T]$.

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

Proof. (i) We may assume $\bar{f} = 0$, otherwise we replace f by $f - \bar{f}$. Consider $g = \limsup_{n\to\infty} A_n f$. Note that

$$\frac{n+1}{n}(A_{n+1}f)(x) = (A_n f)(Tx) + \frac{f(x)}{n}.$$

Letting $n \to \infty$ and take the supremum, we see that $g(Tx) = g(x)$, and g is $\mathscr{I}_T$-measureable. Then we take $D = \{g > \epsilon\} \in \mathscr{I}_T$, define $f^* = (f - \epsilon)\mathbb{1}_D$, and choose $F_n = \{\sup_{1\le k\le n} A_k f^* > 0\}$. Then

$$F = \bigcup_{n=1}^{\infty} F_n = \left\{\sup_{n\ge 1} A_n f^* > 0\right\}.$$

Since D is T-invariant, $f^* \circ T = (f \circ T - \epsilon)\mathbb{1}_D$, and $A_n f^* = (A_n f - \epsilon)\mathbb{1}_D$. Hence

$$F = D \cap \left\{\sup_{k\ge 1} A_k f > \epsilon\right\} = D.$$

Since $f \mapsto f \circ T$ is norm-preserving in L^1, we apply maximal ergodic lemma [Lemma 4.8] and the dominated convergence theorem on $\mathbb{1}_{F_n} \uparrow \mathbb{1}_F = \mathbb{1}_D$ to obtain

$$0 \le \int_{F_n} f^*\, d\mu \to \int_D f^*\, d\mu.$$

To proceed, note that

$$0 \le \int_D f^*\, d\mu \le \int_D (f - \epsilon)\, d\mu = \int_D (\bar{f} - \epsilon)\, d\mu = -\epsilon\mu(D).$$

Hence $\mu(D) = 0$. Since $\epsilon > 0$ is arbitrary, we know that $g \le 0$ a.s., and $\limsup_{n\to\infty} A_n f \le 0$. We apply the same result on $-f$ to conclude $\liminf_{n\to\infty} A_n f \ge 0$. Therefore $\lim_{n\to\infty} A_n f = 0$ a.s..

(ii) If T is μ-ergodic, then $\mathscr{I}_T$ is trivial, and $\bar{f} = \mathbb{E}[f|\mathscr{I}_T] = \mathbb{E}f = \int_X f\, d\mu$. $\square$

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

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

Proof. Let $g_M = f\chi_{|f|\le M}$ and $h_M = f - g_M = f\chi_{|f|>M}$. Then

$$\frac{1}{n}\sum_{k=0}^{n-1} g_M \circ T^k \to \mathbb{E}[g_M|\mathscr{I}_T] \quad a.e.,$$

and the convergence also holds in L^p by dominated convergence theorem. Meanwhile,

$$\left\| \frac{1}{n}\sum_{k=0}^{n-1} h_M \circ T^k \right\|_{L^p} \le \frac{1}{n}\sum_{k=0}^{n-1} \left\| h_M \circ T^k \right\|_{L^p} = \|h_M\|_{L^p}.$$

Also,

$$\int_X |\mathbb{E}[h_M|\mathscr{I}_T]|^p \, d\mu \le \int_X \mathbb{E}[|h_M|^p|\mathscr{I}_T] \, d\mu = \|h_M\|_{L^p}^p.$$

By triangle inequality,

$$\limsup_{n\to\infty} \left\| \frac{1}{n}\sum_{k=0}^{n-1} h_M \circ T^k - \mathbb{E}[h_M|\mathscr{I}_T] \right\|_{L^p} \le 2\|h_M\|_{L^p}.$$

As $M \to \infty$, we have $\|h_M\|_{L^p}^p \to 0$ by dominated convergence theorem, which completes the proof. ☐

Remark. To summarize, if $1 \le p < \infty$ 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.

4.4 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 4.14 (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 $\operatorname{ext}(K)$ 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 4.15. *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 4.16. *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 4.17 (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 4.15, $(K, \preceq)$ 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 4.16, 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 4.18 (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 4.17, 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 4.17, there exists a continuous linear functional ℓ such that $\ell(x) > \sup_{y \in \overline{\mathrm{conv}}(\mathrm{ext}(K))} \ell(y)$. By Lemma 4.16, 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 4.17, the compact convex set F_ℓ has an extreme point x, which is also an extreme point of K by Lemma 4.15. This contradicts the fact that $F_\ell \cap \mathrm{ext}(K) = \emptyset$. Thus we complete the proof. $\square$

4.5 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 4.19. *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 $\mathscr{B}$. □

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 4.20 (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 4.6, 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. $\square$

Corollary 4.21. *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 4.19, it has an extreme point μ by Krein-Milman theorem. According to Theorem 4.20, μ is a T-ergodic measure. $\square$

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

Definition 4.22 (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 4.23 (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 $f \in C(X)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \int_X f \, d\mu \quad \text{for all } x \in X. \tag{4.2}$$

(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 (4.2). 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

$$\left| \frac{1}{n_m} \sum_{k=0}^{n_m - 1} f(T^k x_m) - \int f \, d\mu \right| \geq \epsilon. \tag{4.3}$$

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 (4.3) 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,

$$\int_X f \, d\nu = \lim_{n\to\infty} \int_X A_n f \, d\nu = \int_X \lim_{n\to\infty} A_n f \, d\nu = \int_X \left(\int_X f \, d\mu \right) d\nu = \int_X f \, d\mu. \tag{4.4}$$

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

4.6 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 4.24 (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 4.25 (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 4.26 (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.

5 Hausdorff Measures, Area and Coarea Formulae

In this section, we study measures that estimate the "low-dimensional" volume of "very small" subset of $\mathbb{R}^n$, which are obtained by considering the size of coverings.

5.1 Definitions and Fundamental Properties

Definition 5.1 (Hausdorff measures)**.** Let X be a metric space, $E \subset X$, $0 \le s < \infty$ and $0 < \delta \le \infty$. Define

$$\mathcal{H}^s_\delta(E) = \inf\left\{\sum_{j=1}^\infty \alpha_s \left(\frac{\operatorname{diam} E_j}{2}\right)^s \,\middle|\, E \subset \bigcup_{j=1}^\infty E_j,\ \operatorname{diam} E_j \le \delta\right\}, \quad \text{where } \alpha_s = \frac{\pi^{s/2}}{\Gamma\left(\frac{s}{2}+1\right)}.$$

Based on this notation, we define the *s-dimensional Hausdorff measure on X* by

$$\mathcal{H}^s(E) := \lim_{\delta\downarrow 0} \mathcal{H}^s_\delta(E) = \sup_{\delta>0} \mathcal{H}^s_\delta(E). \tag{5.1}$$

Remark. If s is an integer, the constant α_s is the volume of the unit ball in s-dimensional Euclidean space. Also, the set on which we take infimum is decreasing as $\delta \downarrow 0$, hence the limit (5.1) is well-defined. We can view this definition as evaluating the "volume" of a set by convering it with finer and finer balls.

To establish some fundamental results about Hausdorff criterions, we first introduce a criterion to verify that an outer measure in $\mathbb{R}^n$ is Borel.

Theorem 5.2 (Caratheodory's criterion)**.** *Let μ be an outer measure on a metric space X. If for all sets $A, B \subset X$ with $d(A,B) > 0$, we have $\mu(A \cup B) = \mu(A) + \mu(B)$, then μ is a Borel measure.*

Proof. Step I. Assume $E, F \subset X$ and F is closed. We claim that

$$\mu(E) \ge \mu(E \backslash F) + \mu(E \cap F). \tag{5.2}$$

Without loss of generality we assume $\mu(E) < \infty$, otherwise the inequality (5.2) is obvious. Define

$$F_m = \left\{x \in \mathbb{R}^n : d(x,F) \le \frac{1}{m}\right\}, \quad m = 1, 2, \cdots.$$

Then $d(E\backslash F_m, E \cap F) \ge \frac{1}{m} > 0$, and

$$\mu(E\backslash F_m) + \mu(E \cap F) = \mu(E \cap (F \cup F_m^c)) \le \mu(E). \tag{5.3}$$

Step II. We then set

$$A_k = \left\{x \in E : \frac{1}{k+1} < d(x,F) \le \frac{1}{k}\right\}, \quad k = 1, 2, \cdots.$$

Since F is closed, $E\backslash F = (E\backslash F_m) \cup (\bigcup_{k=m}^\infty A_k)$, and

$$\mu(E\backslash F_m) \le \mu(E\backslash F) \le \mu(E\backslash F_m) + \sum_{k=m}^\infty \mu(A_k). \tag{5.4}$$

If $j \ge i+2$, we have $d(A_i, A_j) > 0$. By induction, for all $m \in \mathbb{N}$,

$$\sum_{k=1}^m \mu(A_{2k-1}) = \mu\left(\bigcup_{k=1}^m A_{2k-1}\right) \le \mu(E), \quad \text{and} \quad \sum_{k=1}^m \mu(A_{2k}) = \mu\left(\bigcup_{k=1}^m A_{2k}\right) \le \mu(E).$$

Combining these results and letting $m \to \infty$, we have $\sum_{k=1}^{\infty} \mu(A_k) \le 2\mu(E) < \infty$. Recalling (5.4), we have $\mu(E \backslash F_m) \to \mu(E \backslash F)$ as $m \to \infty$. We plug in this limit to (5.3) to get (5.2).

Step III. Since the outer measure is subadditive, the opposite of (5.2) holds, and we have for all closed sets $F \subset X$ the Carathéodory condition

$$\mu(E) = \mu(E \backslash F) + \mu(E \cap F), \quad \text{for all } E \subset X.$$

Hence μ is a measure on the σ-algebra generated by all closed subsets of X, which is the Borel σ-algebra. □

Theorem 5.3. *Let $n \in \mathbb{N}$. For each $0 \le s < \infty$, the Hausdorff measure $\mathcal{H}^s$ on a metric space X satisfies:*
(i) $\mathcal{H}^s$ is a Borel measure on X;
(ii) (Borel Regularity). For each $E \subset X$, there exists a Borel set $B \supset E$ such that $\mathcal{H}^s(B) = \mathcal{H}^s(E)$.

Proof. Step I. We claim that $\mathcal{H}^s$ is an outer measure. Let $A_1, A_2, \cdots \subset X$. For any $\delta > 0$, let $A_k \subset \bigcup_{j=1}^{\infty} E_j^k$, with $\operatorname{diam} E_j^k \le \delta$. Then $\{E_j^k\}_{j,k=1}^{\infty}$ covers $\bigcup_{k=1}^{\infty} A_k$, and

$$\mathcal{H}_\delta^s \left(\bigcup_{k=1}^{\infty} A_k \right) \le \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \alpha_s \left(\frac{\operatorname{diam} E_j^k}{2} \right)^s$$

Taking the infima over $\{E_j^k\}_{j,k=1}^{\infty}$, we have

$$\mathcal{H}_\delta^s \left(\bigcup_{k=1}^{\infty} A_k \right) \le \sum_{k=1}^{\infty} \mathcal{H}_\delta^s(A_k) \le \sum_{k=1}^{\infty} \mathcal{H}^s(A_k).$$

Finally, we take $\delta \downarrow 0$ to conclude the proof of our claim.

Step II. Let $A, B \subset X$ be two sets with $d(A, B) > 0$. By Caratheodory's criterion, we can conclude that $\mathcal{H}^s$ is a Borel measure if $\mathcal{H}^s(A \cup B) = \mathcal{H}^s(A) + \mathcal{H}^s(B)$. Below we prove this identity.

We take $0 < \delta < \frac{1}{4} d(A, B)$, and assume that $A \cup B \subset \bigcup_{j=1}^{\infty} E_j$ with $\operatorname{diam} E_j < \delta$ for all $j \in \mathbb{N}$. We define $\mathcal{A} = \{E_j : A \cap E_j \neq \emptyset\}$ and $\mathcal{B} = \{E_j : B \cap E_j \neq \emptyset\}$. Then $A \subset \bigcup_{E_j \in \mathcal{A}} E_j$ and $B \subset \bigcup_{E_j \in \mathcal{B}} E_j$. Furthermore, if $E_i \in \mathcal{A}$ and $E_j \in \mathcal{B}$, we have $d(E_i, E_j) > 2\delta$ by the fact $\operatorname{diam} E_i, \operatorname{diam} E_j < \delta$. Hence

$$\sum_{k=1}^{\infty} \alpha_s \left(\frac{\operatorname{diam} E_k}{2} \right)^s = \sum_{E_j \in \mathcal{A}} \alpha_s \left(\frac{\operatorname{diam} E_j}{2} \right)^s + \sum_{E_j \in \mathcal{B}} \alpha_s \left(\frac{\operatorname{diam} E_j}{2} \right)^s \ge \mathcal{H}_\delta^s(A) + \mathcal{H}_\delta^s(B).$$

Taking the infimum over all such sets $(E_j)_{j=1}^{\infty}$, we conclude that $\mathcal{H}_\delta^s(A \cup B) \ge \mathcal{H}_\delta^s(A) + \mathcal{H}_\delta^s(B)$ for $0 < \delta < \frac{1}{4} d(A, B)$. Letting $\delta \downarrow 0$, we have $\mathcal{H}^s(A \cup B) \ge \mathcal{H}^s(A) + \mathcal{H}^s(B)$. The opposite holds by subadditivity of μ.

Step III. To verify the assertion (ii), we assume $E \subset X$ and $\mathcal{H}^s(E) < \infty$ (otherwise just take $B = X$). For each $k \in \mathbb{N}$, take $E_1^k, E_2^k, \cdots \subset X$ such that $E \subset \bigcup_{j=1}^{\infty} X$, $\operatorname{diam} E_j^k \le \frac{1}{k}$ for all $j \in \mathbb{N}$, and

$$\sum_{j=1}^{\infty} \alpha_s \left(\frac{\operatorname{diam} E_j^k}{2} \right)^s \le \mathcal{H}_{1/k}^s(E) + \frac{1}{k}.$$

We can further assume E_j^k's are closed, since the cover property and set diameters do not change when we take closures. We pick $B = \bigcap_{k=1}^{\infty} \bigcup_{j=1}^{\infty} E_j^k$, which is a Borel set that contains E. Then for each $k \in \mathbb{N}$,

$$\mathcal{H}_{1/k}^s(E) \le \mathcal{H}_{1/k}^s(B) \le \mathcal{H}^s \left(\bigcup_{j=1}^{\infty} E_j^k \right) \le \sum_{j=1}^{\infty} \alpha_s \left(\frac{\operatorname{diam} E_j^k}{2} \right)^s \le \mathcal{H}_{1/k}^s(E) + \frac{1}{k}.$$

Letting $k \to \infty$ concludes the proof. □

We next study some properties of Hausdorff measures on Euclidean spaces.

Proposition 5.4. *Let $0 \le s < \infty$ and let $\mathcal{H}^s$ be the s-diemnsional Hausdorff measure on $\mathbb{R}^n$.*

(i) (Scaling). Let $\lambda \in \mathbb{R}$ and $E \subset \mathbb{R}^n$. Then $\mathcal{H}^s(\lambda E) = |\lambda|^s \mathcal{H}^s(E)$.

(ii) (Affine Invariance). Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an affine isometry, and $E \subset \mathbb{R}^n$. Then $\mathcal{H}^s(TE) = \mathcal{H}^s(E)$.

(iii) Let $E \subset \mathbb{R}^n$. If $\mathcal{H}^s_\delta(E) = 0$ for some $0 < \delta < \infty$, then $\mathcal{H}^s(E) = 0$.

(iv) If $s > n$, then $\mathcal{H}^s \equiv 0$ on $\mathbb{R}^n$.

Proof. The properties (i) and (ii) is clear by definition.

(iii) The conclusion is clear for $s = 0$, so we may assume $s > 0$. For any $\epsilon > 0$, there exists $E_1, E_2, \cdots \subset \mathbb{R}^n$ such that $E \subset \bigcup_{j=1}^\infty E_j$ and

$$\sum_{j=1}^\infty \alpha_s \left(\frac{\operatorname{diam} E_j}{2} \right)^s < \epsilon.$$

Then for each j, we have $\operatorname{diam} E_j \le \delta_\epsilon := 2 \left(\frac{\epsilon}{\alpha_s} \right)^{1/s}$, and $\mathcal{H}^s_{\delta_\epsilon}(E) < \epsilon$. Since $\delta_\epsilon \downarrow 0$ as $\epsilon \downarrow 0$, we have $\mathcal{H}^s(E) = 0$.

(iv) For each $m \in \mathbb{N}$, the unit cube $Q \in \mathbb{R}^n$ can be divided into m^n cubes with side $\frac{1}{m}$ and diameter $\frac{\sqrt{n}}{m}$. Then

$$\mathcal{H}^s_{\frac{\sqrt{n}}{m}}(Q) \le \sum_{j=1}^{m^n} \alpha_s \left(\frac{\sqrt{n}}{2m} \right)^s = \alpha_s \left(\frac{\sqrt{n}}{2} \right)^s m^{n-s}.$$

When $s > n$, the last term goes to zero as $m \to \infty$. Then $\mathcal{H}^s(Q) = 0$, and $\mathcal{H}^s(\mathbb{R}^n) = 0$. □

Now we focus on Hausdorff measures of low dimensions.

Proposition 5.5. *(i) $\mathcal{H}^0$ is the counting measure. (ii) $\mathcal{H}^1$ on $\mathbb{R}$ coincides the Lebesgue measure m.*

Proof. (i) By definition, for each $x \in \mathbb{R}^n$, we have $\mathcal{H}^0_\delta(\{x\}) = 1$ for all $\delta > 0$, and $\mathcal{H}^0(\{x\}) = 1$.

(ii) Note that $\alpha_1 = 2$. Let $E \subset \mathbb{R}$ and $\delta > 0$. Since the Lebesgue measure as an outer measure on $\mathbb{R}$,

$$m(E) = \inf \left\{ \sum_{j=1}^\infty \operatorname{diam} E_j \,\middle|\, \bigcup_{j=1}^\infty E_j \supset E \right\} \le \inf \left\{ \sum_{j=1}^\infty \operatorname{diam} E_j \,\middle|\, \bigcup_{j=1}^\infty E_j \supset E,\ \operatorname{diam} E_j \le \delta \right\} = \mathcal{H}^1_\delta(E).$$

On the other hand, set $I_k = [(k-1)\delta, k\delta]$ for $k \in \mathbb{Z}$. Then

$$m(E) = \inf \left\{ \sum_{j=1}^\infty \operatorname{diam} E_j \,\middle|\, \bigcup_{j=1}^\infty E_j \supset E \right\} \ge \inf \left\{ \sum_{j=1}^\infty \sum_{k \in \mathbb{Z}} \operatorname{diam}(I_k \cap E_j) \,\middle|\, \bigcup_{j=1}^\infty E_j \supset E \right\} \ge \mathcal{H}^1_\delta(E).$$

Hence $\mathcal{H}^1_\delta = m$ for all $\delta > 0$, and $\mathcal{H}^1 = m$ on $\mathbb{R}$. □

We next study what happens to the Hausdorff measure $\mathcal{H}^s$ of a $E \subset \mathbb{R}^n$ when we change the dimension s, and introduce the Hausdorff dimension.

Proposition 5.6. *Let X be a metric space, $E \subset X$ and $0 \le s < t < \infty$.*

(i) If $\mathcal{H}^s(E) < \infty$, then $\mathcal{H}^t(E) = 0$.

(ii) If $\mathcal{H}^t(E) > 0$, then $\mathcal{H}^s(E) = \infty$.

Proof. We take $\delta > 0$, and $\bigcup_{j=1}^\infty E_j \supset E$ with $\operatorname{diam} E_j \le \delta$. Then

$$\sum_{j=1}^\infty \alpha_s \left(\frac{\operatorname{diam} E_j}{2} \right)^s \le \mathcal{H}^s_\delta(E) + 1 \le \mathcal{H}^s(E) + 1.$$

Note that $\operatorname{diam} E_j \le \delta$. Hence

$$\mathcal{H}^t_\delta(E) = \sum_{j=1}^{\infty} \alpha_t \left(\frac{\operatorname{diam} E_j}{2}\right)^t \le \frac{\alpha_t \delta^{t-s}}{\alpha_s 2^{t-s}} \sum_{j=1}^{\infty} \alpha_s \left(\frac{\operatorname{diam} E_j}{2}\right)^s \le \frac{\alpha_t \delta^{t-s}}{\alpha_s 2^{t-s}} \left(\mathcal{H}^s(E) + 1\right).$$

If $\mathcal{H}^s(E) < \infty$, we send $\delta \downarrow 0$ to conclude $\mathcal{H}^t(E) = 0$. Note that (ii) is the contrapositive of (i). $\square$

The above proposition justifies the following definition of Hausdorff dimension.

Definition 5.7 (Hausdorff dimension)**.** Let X be a metric space. The *Hausdorff dimension* of a set $E \subset X$ is

$$\dim_{\mathcal{H}}(E) = \inf\{0 \le s < \infty : \mathcal{H}^s(E) = 0\}.$$

Remark. By Proposition 5.6,

$$\mathcal{H}^s(E) = \begin{cases} \infty, & 0 \le s < \dim_{\mathcal{H}}(E), \\ 0, & s > \dim_{\mathcal{H}}(E). \end{cases}$$

Hence we have following equivalent characterization of Hausdorff dimension:

$$\begin{aligned} \dim_{\mathcal{H}}(E) &= \inf\{0 \le s < \infty : \mathcal{H}^s(E) < \infty\} = \sup\{0 \le s < \infty : \mathcal{H}^s(E) > 0\} \\ &= \sup\{0 \le s < \infty : \mathcal{H}^s(E) = \infty\} = \inf\{0 \le s < \infty : \mathcal{H}^s(E) = 0\}. \end{aligned}$$

The Hausdorff dimension of a set in a general metric space can be infinity. But in the Euclidean space $\mathbb{R}^n$, it is clear that $\dim_{\mathcal{H}}(E) \le n$ for all $E \subset \mathbb{R}^n$.

Theorem 5.8 (Countable stability)**.** *Let X be a metric space and $E_1, E_2, \cdots \subset X$. Then*

$$\dim_{\mathcal{H}}\left(\bigcup_{j=1}^{\infty} E_j\right) = \sup_{j\in\mathbb{N}} \dim_{\mathcal{H}}(E_j).$$

Proof. If $s < \sup_{j\in\mathbb{N}} \dim_{\mathcal{H}}(E_j)$, there exists $j \in \mathbb{N}$ such that $s < \dim_{\mathcal{H}}(E_j)$, and $\mathcal{H}^s(E_j) = \infty$. Then

$$\mathcal{H}^s\left(\bigcup_{j=1}^{\infty} E_j\right) \ge \mathcal{H}^s(E_j) = \infty. \tag{5.5}$$

Hence $s \le \dim_{\mathcal{H}}(\bigcup_{j=1}^{\infty} E_j)$. Since $s < \sup_{j\in\mathbb{N}} \dim_{\mathcal{H}}(E_j)$ is arbitrary, we let $s \uparrow \dim_{\mathcal{H}}(E_j)$ to obtain

$$\sup_{j\in\mathbb{N}} \dim_{\mathcal{H}}(E_j) \le \dim_{\mathcal{H}}\left(\bigcup_{j=1}^{\infty} E_j\right).$$

Also, if $s > \sup_{j\in\mathbb{N}} \dim_{\mathcal{H}}(E_j)$, then $\mathcal{H}^s(E_j) = 0$ for all $j \in \mathbb{N}$, and

$$\mathcal{H}^s\left(\bigcup_{j=1}^{\infty} E_j\right) \le \sum_{j=1}^{n} \mathcal{H}^s(E_j) = 0.$$

Hence $s \ge \dim_{\mathcal{H}}(\bigcup_{j=1}^{\infty} E_j)$, and we can similarly obtain

$$\sup_{j\in\mathbb{N}} \dim_{\mathcal{H}}(E_j) \ge \dim_{\mathcal{H}}\left(\bigcup_{j=1}^{\infty} E_j\right),$$

which is the opposite of (5.5). $\square$

Example 5.9 (Hausdorff dimension of Cantor sets)**.** Let $0 < \gamma < \frac{1}{2}$. Beginning from the unit interval $[0, 1]$, the *γ-ary Cantor set* is obtained by repeatedly removing the middle open $1 - 2\gamma$ from the interval. To elaborate, we first remove the open interval $(\gamma, 1 - \gamma)$ from $I_{0,1} = [0, 1]$ and get

$$I_{1,1} = [0, \gamma], \quad I_{1,2} = [1 - \gamma, \gamma].$$

Next, we remove the middle open interval of length $\gamma(1 - 2\gamma)$ of each of the above intervals and get

$$I_{2,1} = [0, \gamma^2], \quad I_{2,2} = [\gamma - \gamma^2, \gamma], \quad I_{2,3} = [1 - \gamma, 1 - \gamma + \gamma^2], \quad I_{2,4} = [1 - \gamma^2, 1].$$

Continuing this process, after the n^{th} step there are 2^n closed intervals $I_{n,1}, I_{n,2}, \cdots, I_{n,2^n}$ of length γ^n, and we remove the middle open interval of length $\gamma^n(1 - 2\gamma)$ from each intervals at the next step. We define the *γ-ary Cantor set* as the decreasing intersection

$$C_\gamma = \bigcap_{n=1}^{\infty} \bigcup_{j=1}^{2^n} I_{n,j}.$$

Since for each $n \in \mathbb{N}$, we can cover C_γ by 2^n intervals of length γ^n. Hence

$$\mathcal{H}^s_{\delta_n}(C_\gamma) \le \frac{\alpha_s}{2^s} (2\gamma^s)^n, \quad \delta_n = \frac{\gamma^n}{2}.$$

If we choose $s = \frac{\log 2}{\log(1/\gamma)}$, we have

$$\mathcal{H}^s(C_\gamma) = \lim_{n\to\infty} \mathcal{H}^s_{\delta_n}(C_\gamma) \le \frac{\alpha_s}{2^s} < \infty.$$

We also give a lower bound of the s-dimensional Hausdorff measure of C_γ:

$$\mathcal{H}^s(C_\gamma) \ge \frac{\alpha_s}{4 \cdot 2^s} > 0. \tag{5.6}$$

According to the two estimates, we conclude that the Hausdorff dimension of C_γ is $s = \frac{\log 2}{\log(1/\gamma)}$.

Proof of (5.6). Let $\mathscr{J}$ be a collection of open intervals that cover C_γ. We prove that

$$\sum_{J_i \in \mathscr{J}} m(J_i)^s \ge \frac{1}{4}. \tag{5.7}$$

We assume that J_i contains some interval I_{l,j_l} from the l^{th} stage, and let k be the smallest integer such that J_i contains some interval I_{k,j_k} from the k^{th} stage. Clearly $k \le l$. Furthermore, following our construction of C_γ we know that no more than 4 intervals from the k^{th} stage can intersect J_i for otherwise J_i must contain some interval $I_{k-1,j_{k-1}}$ from the $(k-1)^{\text{th}}$ stage. Thus

$$4m(J_i)^s \ge \sum_{I_{k,j} \cap J_i \ne \emptyset} m(J_i)^s \ge \sum_{I_{k,j} \cap J_i \ne \emptyset} m(I_{k,j})^s = \sum_{I_{k,j} \cap J_i \ne \emptyset} \sum_{I_{l,j'} \subset I_{k,j}} m(I_{l,j'})^s \ge \sum_{I_{l,j} \subset J_i} m(I_{l,j})^s.$$

By compactness of C_γ and the Lebesgue number lemma, for sufficiently large l, every interval $I_{l,j}$ is contained in some $J_i \in \mathscr{J}$. Hence

$$4 \sum_{J_i \in \mathscr{J}} m(J_i)^s \ge \sum_{J_i \in \mathscr{J}} \sum_{I_{l,j} \subset J_i} m(I_{l,j})^s \ge \sum_{j=1}^{2^l} m(I_{l,j})^s = 2^l(\gamma^l)^s = 1,$$

which establishes (5.7). Now if the estimate (5.6) did not hold, there would exist $E_1, E_2, \cdots \subset \mathbb{R}$ such that

$\bigcup_{i=1}^\infty E_i \supset C_\gamma$, and for some $\epsilon > 0$,

$$\sum_{i=1}^\infty (\operatorname{diam} E_i)^s < \frac{1}{4(1+\epsilon)^s}.$$

We then replace each E_i by an open interval J_i of length $(1+\epsilon)\operatorname{diam} E_i$ with $J_i \supset E_i$. This implies

$$\sum_{i=1}^\infty m(J_i)^s < \frac{1}{4},$$

which contradicts (5.7). Hence we complete the proof. $\square$

5.2 Vitali's Covering Theorem

In this subsection, we study how to fill an open set in $\mathbb{R}^n$ with countably many balls and introduce Vitali's covering theorem. We first discuss an analogue of the finite result discussed in Lemma 2.30.

Theorem 5.10 (Vitali's covering theorem). *For any collection $\mathcal{F}$ of non-degenerate closed balls in $\mathbb{R}^n$ with*

$$\sup_{B\in\mathcal{F}} \operatorname{diam} B < \infty.$$

Then there exists a countable subcollection $\mathcal{G} \subset \mathcal{F}$ of disjoint balls such that

$$\bigcup_{B\in\mathcal{F}} B \subset \bigcup_{B\in\mathcal{G}} 5B,$$

where $5B$ denotes the closed ball with the same center as B but with 5 times the radius.

Proof. We write $M = \sup_{B\in\mathcal{F}} \operatorname{diam} B$ and set

$$\mathcal{F}_k = \left\{ B \in \mathcal{F} : \frac{M}{2^k} < \operatorname{diam} B \le \frac{M}{2^{k-1}} \right\}, \quad k = 1, 2, \cdots .$$

We define $\mathcal{G}_k \subset \mathcal{F}_k$ as follows:

- Let $\mathcal{G}_1$ be any maximal disjoint subcollection of balls in $\mathcal{F}_1$, which is clearly countable;
- With $\mathcal{G}_1, \cdots, \mathcal{G}_{k-1}$ selected, we choose $\mathcal{G}_k$ to be any maximal disjoint subcollection of

$$\left\{ B \in \mathcal{F}_k : B \cap B' = \emptyset \text{ for all } B' \in \bigcup_{j=1}^{k-1} \mathcal{G}_j \right\}.$$

Finally, we define $\mathcal{G} = \bigcup_{k=1}^\infty \mathcal{G}_k$, which is clearly a subcollection of disjoint balls in $\mathcal{F}$. To conclude the proof, we claim that, for each ball $B \in \mathcal{F}$, there exists $B' \in \mathcal{G}$ such that $B \cap B' \neq \emptyset$ and $B \subset 5B'$.

Fix $B \in \mathcal{F}$. There then exists an index k such that $B \in \mathcal{F}_k$. By maximality of $\mathcal{G}_k$, there exists a ball $B' \in \bigcup_{j=1}^k \mathcal{G}_j$ with $B \cap B' \neq \emptyset$. Since $\operatorname{diam} B' > 2^{-k}M$ and $\operatorname{diam} B \le 2^{1-k}M$, we have $\operatorname{diam} B \le 2 \operatorname{diam} B'$ and so $B \subset 5B'$ by the triangle inequality. $\square$

Theorem 5.11 (Filling open sets with balls). *Let $U \subset \mathbb{R}^n$ be open, and $\delta > 0$. Then there exists a countable collection $\mathcal{G}$ of disjoint closed balls contained in U such that* $\operatorname{diam} B < \delta$ *for all $B \in \mathcal{G}$ and*

$$m\left(U \backslash \bigcup_{B\in\mathcal{G}} B \right) = 0,$$

where m is the Lebesgue measure on $\mathbb{R}^n$.

Proof. Step I. We first assume $m(U) < \infty$, and fix $1 - \frac{1}{5^n} < \theta < 1$. We claim that there exists finitely many disjoint balls $B_1, \cdots, B_{M_1} \subset U$ such that

$$m\left(U \backslash \bigcup_{j=1}^{M_1} B_j\right) \leq \theta m(U). \tag{5.8}$$

To prove this, we let

$$\mathcal{F}_1 = \{B \subset U : \ B \text{ is a closed ball, } \operatorname{diam} B < \delta\}.$$

By Vitali's covering theorem [Theorem 5.10], there exists a countable subcollection $\mathcal{G}_1 \subset \mathcal{F}_1$ of disjoint balls contained in U such that

$$\bigcup_{B \in \mathcal{G}_1} 5B \supset U.$$

Hence

$$m(U) \leq m\left(\bigcup_{B \in \mathcal{G}_1} 5B\right) \leq \sum_{B \in \mathcal{G}_1} m(5B) = 5^n \sum_{B \in \mathcal{G}_1} m(B) = 5^n m\left(\bigcup_{B \in \mathcal{G}_1} B\right).$$

As a result,

$$m\left(U \backslash \bigcup_{B \in \mathcal{G}_1} B\right) \leq \left(1 - \frac{1}{5^n}\right) m(U).$$

Since $m(U) < \infty$ and $\theta > 1 - \frac{1}{5^n}$, there exists finitely many balls $B_1, \cdots, B_{M_1} \in \mathcal{G}_1$ satisfying (5.8).

Step II. Define

$$U_2 = U \backslash \bigcup_{B \in \mathcal{G}_1} B,$$

and

$$\mathcal{F}_2 = \{B \subset U_2 : \ B \text{ is a closed ball, } \operatorname{diam} B < \delta\}.$$

As above, we may find finitely many disjoint balls $B_{M_1+1}, \cdots, B_{M_2} \in \mathcal{F}_2$ such that

$$m\left(U \backslash \bigcup_{j=1}^{M_2} B\right) = m\left(U_2 \backslash \bigcup_{j=M_1+1}^{M_2} B\right) \leq \theta m(U_2) \leq \theta^2 m(U).$$

Step III. Continuing this procedure, for each k, there exist finitely many balls $B_1, \cdots, B_{M_k} \subset U$ with diameter less than δ such that

$$m\left(U \backslash \bigcup_{j=1}^{M_k} B\right) \leq \theta^k m(U), \quad k = 1, 2, \cdots.$$

Since $\theta^k \to 0$ as $k \to \infty$, we complete the proof for the case $m(U) < \infty$.

Step IV. If $m(U) = \infty$, we apply the above construction to each of the open sets

$$U_m = \{x \in U : m < |x| < m+1\}, \quad m = 1, 2, \cdots.$$

Then we conclude the proof. $\square$

5.3 The Isodiametric Inequality

From now on, to distinguish between Lebesgue measures on Euclidean spaces of different dimensions, we use $\mathcal{L}^n$ to denote the Lebesgue measure on $\mathbb{R}^n$. This is obtained by applying Carathéodory's extension on $\mathbb{R}^n$ to a pre-measure on the semi-ring comprised of cells of the form $\prod_{i=1}^n (a_i, b_i]$. To be specific, one define

$$\mathcal{L}^n\left(\prod_{i=1}^n (a_i, b_i]\right) = \prod_{i=1}^n (b_i - a_i) \quad \text{for all } a_i \leq b_i,\ i = 1, 2, \cdots, n,$$

and for $E \subset \mathbb{R}^n$, define

$$\mathcal{L}^n(E) = \inf\left\{\sum_{j=1}^\infty \mathcal{L}^n(Q_j) \,\middle|\, (Q_j)_{j=1}^\infty \text{ are cells of the form } \prod_{i=1}^n (a_i, b_i], \text{ and } \bigcup_{j=1}^\infty Q_j \supset E\right\}. \tag{5.9}$$

In this subsection, we are going to establish that $\mathcal{L}^n = \mathcal{H}^n$ on $\mathbb{R}^n$ by the isodiametric inequality. To proceed, we first introduce the Steiner symmetrization.

Definition 5.12 (Steiner symmetrization)**.** Let $\xi \in \mathbb{R}^n$ with $|\xi| = 1$, and $b \in \mathbb{R}^n$. Denote by

$$L_\xi(b) = \{b + t\xi : t \in \mathbb{R}\}$$

the line through b in the direction ξ, and

$$P_\xi = \{x \in \mathbb{R}^n : \xi^\top x = 0\}$$

by the plane through the origin perpendicular to ξ. Given $E \in \mathbb{R}^n$, we define the *Steiner symmetrization* of E with respect to the plane P_ξ to be the set

$$S_\xi(E) = \bigcup_{b \in P_\xi,\, E \cap L_\xi(b) \neq \emptyset} \left\{b + ta : |t| \leq \frac{1}{2}\mathcal{H}^1\big(E \cap L_\xi(b)\big)\right\}$$

Proposition 5.13 (Properties of the Steiner symmetrization)**.** *Let $\xi \in \mathbb{R}^n$ with $|\xi| = 1$, and $E \subset \mathbb{R}^n$. Then*

(i) $\operatorname{diam} S_\xi(E) \leq \operatorname{diam} E$.

(ii) If E is $\mathcal{L}^n$-measurable, so is $S_\xi(E)$, and $\mathcal{L}^n(S_\xi(E)) = \mathcal{L}^n(E)$.

Proof. (i) We may assume that $\operatorname{diam} E < \infty$ and E is closed. We fix $\epsilon > 0$ and choose $x, y \in S_\xi(E)$ with $\operatorname{diam} S_\xi(E) \leq |x - y| + \epsilon$. Set $b = x - (\xi^\top x)\xi \in P_\xi$ and $c = y - (\xi^\top y)\xi \in P_\xi$, and

$$r = \inf\{t : b + t\xi \in E\}, \quad s = \sup\{t : b + t\xi \in E\}, \quad u = \inf\{t : c + t\xi \in E\}, \quad v = \sup\{t : c + t\xi \in E\}.$$

Without loss of generality, we may assume $v - r \geq s - u$. Then

$$v - r \geq \frac{1}{2}(v - r) + \frac{1}{2}(s - u) = \frac{1}{2}(s - r) + \frac{1}{2}(v - u) \geq \frac{1}{2}\mathcal{H}^1(E \cap L_\xi(b)) + \frac{1}{2}\mathcal{H}^1(E \cap L_\xi(c)).$$

Also, by definition of $S_\xi(E)$, we have $|x^\top \xi| \leq \frac{1}{2}\mathcal{H}^1(E \cap L_\xi(b))$ and $|y^\top \xi| \leq \frac{1}{2}\mathcal{H}^1(E \cap L_\xi(c))$. Hence

$$v - r \geq |x^\top \xi| + |y^\top \xi| = |(x - y)^\top \xi|.$$

Therefore,

$$\begin{aligned}(\operatorname{diam} S_\xi(E) - \epsilon)^2 &\leq |x - y|^2 = |b - c|^2 + |(x - y)^\top \xi|^2 \\ &\leq |b - c|^2 + (v - r)^2 = |(b + r\xi) - (c + v\xi)|^2 \leq (\operatorname{diam} E)^2,\end{aligned}$$

where the last inequality holds because E is closed and so $b+r\xi, c+v\xi \in E$. Since $\epsilon > 0$ is arbitrary, we have

$$\operatorname{diam} S_\xi(E) \leq \operatorname{diam} E.$$

(ii) Since $\mathcal{L}^n$ is rotation invariant, we may assume $\xi = e_n = (0, \cdots, 0, 1)^\top$. Note that $\mathcal{L}^1 = \mathcal{H}^1$ on $\mathbb{R}$. By Tonelli-Fubini theorem, the map $f : \mathbb{R}^{n-1} \to \mathbb{R}$ defined by

$$f(b) = \mathcal{H}^1(E \cap L_\xi(b,0)) = \mathcal{L}^1(E \cap L_\xi(b,0)) = \int_{\mathbb{R}} \mathbb{1}_E(b,t)\, dt$$

is $\mathcal{L}^{n-1}$ measurable, and $\mathcal{L}^n(E) = \int_{\mathbb{R}^{n-1}} f(b)\, db$. We use the following lemma.

Lemma 5.14. *Let $f : \mathbb{R}^{n-1} \to [0,\infty]$ be an $\mathcal{L}^{n-1}$-measurable function. Then the* ***hypograph*** *of f, defined by*

$$H_f = \left\{(x,y) : x \in \mathbb{R}^{n-1}, y \in \mathbb{R} \quad \text{and} \quad y \leq f(x)\right\},$$

is $\mathcal{L}^n$-measurable.

Proof. Note that H_f is the level set $\{g \geq 0\}$ of the $\mathcal{L}^{n+1}$-measurable function $g(x,y) = f(x) - y$. □

Proof of Proposition 5.13 (Continued). We note that

$$S_\xi(E) = \left\{(b,t) \in \mathbb{R}^n : -\frac{f(b)}{2} \leq t \leq \frac{f(b)}{2}\right\} \setminus \left\{(b,0) : b \in \mathbb{R}^{n-1}, E \cap L_\xi(b,0) = \emptyset\right\}.$$

By Lemma 5.14, the symmetrized set $S_\xi(E)$ is $\mathcal{L}^{n-1}$ measurable, and

$$\mathcal{L}^n(S_\xi(E)) = \int_{\mathbb{R}^n} f(b)\, db = \mathcal{L}^n(E).$$ □

Remark. In fact, throughout our subsequent proof of $\mathcal{L}^n = \mathcal{H}^n$, we only use the statement (ii) above in the special case that ξ is a standard coordinate vector e_i, $i = 1, 2, \cdots, n$. Since $\mathcal{L}^n$ is obviously rotation invariant, we therefore indeed prove that $\mathcal{L}_n$ is rotation invariant.

We next introduce the isodiametric inequality, which gives an estimate of $\mathcal{L}^n(E)$ in terms of $\operatorname{diam} E$.

Theorem 5.15 (Isodiametric inequality)**.** *For all sets $E \subset \mathbb{R}^n$,*

$$\mathcal{L}^n(E) \leq \alpha_n \left(\frac{\operatorname{diam} E}{2}\right)^n.$$

Proof. We may assume $\operatorname{diam} E < \infty$. We take the standard basis $e_i = (0, \cdots, 0, \underset{i\text{-th}}{1}, 0, \cdots, 0)$ of $\mathbb{R}^n$, and consider the Steiner symmetrizations $E_1 = S_{e_1}(E), E_2 = S_{e_2}(E_1), \cdots, E_n = S_{e_n}(E_{n-1})$. Write $E^* = E_n$.

Step I. We claim that E^* is symmetric with respect to the origin, i.e. for every $x \in E$, we have $-x \in E$.

To prove this claim, we note that E_1 is clearly symmetric with respect to P_{e_1}. We hence assume $1 \leq k < n$ and E_k is symmetric with respect to $P_{e_1}, \cdots, P_{e_k}$. Then $E_{k+1} = S_{e_{k+1}}(E_k)$ is symmetric with respect to P_{e_k}. We fix $1 \leq j \leq k$, and let $R_j : \mathbb{R}^n \to \mathbb{R}^n$ be the reflection through P_{e_j}. For any $b \in P_{e_{k+1}}$, since $R_j(E_k) = E_k$,

$$\mathcal{H}^1(E_k \cap L_{e_{k+1}}(b)) = \mathcal{H}^1(R_j(E_k) \cap L_{e_{k+1}}(R_j b)) = \mathcal{H}^1(E_k \cap L_{e_{k+1}}(R_j b)).$$

As a result,

$$\{t \in \mathbb{R} : b + te_{k+1} \in E_{k+1}\} = \{t \in \mathbb{R} : R_j b + te_{k+1} \in E_{k+1}\}, \quad b \in P_{e_{k+1}},$$

and $R_j(E_{k+1}) = E_{k+1}$. Therefore E_{k+1} is symmetric with respect to P_{e_j}, $j = 1, 2, \cdots, k$. By induction, E^* is symmetric with respect to $P_{e_1}, \cdots, P_{e_n}$, and hence with respect to the origin.

Step II. If $x \in E^*$, we have $-x \in E^*$ by Step I, and so $\operatorname{diam} E^* \geq 2|x|$. Therefore $E^* \subset B(0, \frac{1}{2} \operatorname{diam} E^*)$, and

$$\mathcal{L}^n(E^*) \leq \mathcal{L}^n\left(B\left(0, \frac{\operatorname{diam} E^*}{2}\right)\right) = \alpha_n \left(\frac{\operatorname{diam} E^*}{2}\right)^n. \tag{5.10}$$

Since $\overline{E}$ is $\mathcal{L}^n$-measurable, by Proposition 5.13,

$$\mathcal{L}^n((\overline{E})^*) = \mathcal{L}^n(E), \quad \text{and} \quad \operatorname{diam}(\overline{E})^* \leq \operatorname{diam} \overline{E}.$$

We then apply (5.10) on $(\overline{E})^*$ to obtain

$$\mathcal{L}^n(E) \leq \mathcal{L}^n(\overline{E}) = \mathcal{L}^n((\overline{E})^*) \leq \alpha_n \left(\frac{\operatorname{diam}(\overline{E})^*}{2}\right)^n \leq \alpha_n \left(\frac{\operatorname{diam} \overline{E}}{2}\right)^n = \alpha_n \left(\frac{\operatorname{diam} E}{2}\right)^n.$$

Thus we complete the proof. $\square$

Remark. Generally, when the dimension $n \geq 2$, we cannot contain a set $E \subset \mathbb{R}^n$ in a ball with diameter $\operatorname{diam} E$. For example, consider an equilateral triangle in $\mathbb{R}^2$.

Using the isodiametric inequality, we can establish the equivalence of Lebesgue and Hausdorff measures in multidimensional Euclidean spaces.

Theorem 5.16 (n-dimensional Hausdorff and Lebesgue measures)**.** *We have*

$$\mathcal{H}^n = \mathcal{L}^n \quad \textit{on } \mathbb{R}^n.$$

Proof. Note that we can write each cell of the form $\prod_{i=1}^n (a_i, b_i]$ as the union of countably many (equilateral) cubes. According to (5.9), for each $\delta > 0$, we have

$$\mathcal{L}^n(E) = \inf\left\{\sum_{j=1}^\infty \mathcal{L}^n(Q_j) \,\middle|\, (Q_j)_{j=1}^\infty \text{ are cubes, } \operatorname{diam} Q_j \leq \delta, \text{ and } \bigcup_{j=1}^\infty Q_j \supset E\right\}, \quad E \subset \mathbb{R}^n.$$

Step I. We claim that $\mathcal{H}^n$ is absolutely continuous with respect to $\mathcal{L}^n$. For each cube $Q_j \subset \mathbb{R}^n$,

$$\mathcal{L}^n(Q_j) = \left(\frac{\operatorname{diam} Q_j}{\sqrt{n}}\right)^n = \frac{\alpha_n}{C_n}\left(\frac{\operatorname{diam} Q_j}{2}\right)^n, \quad \text{where } C_n = \alpha_n \left(\frac{\sqrt{n}}{2}\right)^n.$$

Therefore we have

$$\mathcal{H}^n_\delta(E) \leq \inf\left\{\sum_{j=1}^\infty \alpha_n \left(\frac{\operatorname{diam} Q_j}{2}\right)^n \,\middle|\, (Q_j)_{j=1}^\infty \text{ are cubes, } \operatorname{diam} Q_j \leq \delta, \text{ and } \bigcup_{j=1}^\infty Q_j \supset E\right\} \leq C_n \mathcal{L}^n(Q_j).$$

Letting $\delta \downarrow 0$, we conclude that $\mathcal{H}^n$ is absolutely continuous with respect to $\mathcal{L}^n$.

Step II. We fix $\epsilon, \delta > 0$ and take cubes $(Q_j)_{j=1}^\infty$ such that $E \subset \bigcup_{j=1}^\infty Q_j$, $\operatorname{diam} Q_j < \delta$ for all $j \in \mathbb{N}$, and

$$\sum_{j=1}^\infty \mathcal{L}^n(Q_j) \leq \mathcal{L}^n(E) + \epsilon.$$

By Theorem 5.11, for each $j \in \mathbb{N}$, there exist disjoint closed balls $(B_k^j)_{k=1}^\infty$ contained in the interior of Q_j such that $\operatorname{diam} B_k^j < \delta$ for all k and

$$\mathcal{L}^n\left(Q_j \setminus \bigcup_{k=1}^\infty B_k^j\right) = \mathcal{L}^n\left(\mathring{Q}_j \setminus \bigcup_{k=1}^\infty B_k^j\right) = 0.$$

By Step I, we have

$$\mathcal{H}^n\left(Q_j \backslash \bigcup_{k=1}^{\infty} B_k^j\right) = 0.$$

Therefore

$$\begin{aligned}
\mathcal{H}^n_\delta(E) &\le \sum_{j=1}^{\infty} \mathcal{H}^n_\delta(Q_j) = \sum_{j=1}^{\infty} \mathcal{H}^n_\delta\left(\bigcup_{k=1}^{\infty} B_k^j\right) \le \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \mathcal{H}^n_\delta(B_k^j) \le \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \alpha_n \left(\frac{\operatorname{diam} B_k^j}{2}\right)^n \\
&= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \mathcal{L}^n(B_k^j) = \sum_{j=1}^{\infty} \mathcal{L}^n\left(\bigcup_{k=1}^{\infty} B_k^j\right) = \sum_{j=1}^{\infty} \mathcal{L}^n(Q_j) \le \mathcal{L}^n(E) + \epsilon.
\end{aligned}$$

Letting $\delta, \epsilon \downarrow 0$, we conclude that $\mathcal{H}^n(E) \le \mathcal{L}^n(E)$.

Step III. We fix $\delta > 0$, and choose $(E_j)_{j=1}^{\infty}$ such that $\bigcup_{j=1}^{\infty} E_j \supset E$ and $\operatorname{diam} E_j \le \delta$ for all j. By the isodiametric inequality [Theorem 5.15],

$$\mathcal{L}^n(E) \le \sum_{j=1}^{\infty} \mathcal{L}^n(E_j) \le \sum_{j=1}^{\infty} \alpha_n \left(\frac{\operatorname{diam} E_j}{2}\right)^n.$$

Taking the infima, we have $\mathcal{L}^n(E) \le \mathcal{H}^n_\delta(E)$. Letting $\delta \downarrow 0$, we conclude that $\mathcal{L}^n(E) \le \mathcal{H}^n(E)$. $\square$

5.4 Hausdorff Measure under Hölder and Lipschitz Continuous Mappings

In this subsection, we study Hölder and Lipschitz continuous mappings. Generally, we fix $0 < \gamma \le 1$, and let $A \subset \mathbb{R}^n$. A function $f : A \to \mathbb{R}^m$ is called *Hölder continuous with exponent* γ, provided

$$[f]_{C^{0,\gamma}(A)} := \sup_{x,y \in A,\, x \ne y} \frac{|f(x) - f(y)|^\gamma}{|x - y|} < \infty.$$

This is a seminorm on the vector space $C^{0,\gamma}(A)$ of γ-Hölder continuous functions on A.

Theorem 5.17 (Hausdorff measure under Hölder mappings)**.** *Let $\gamma \in (0, 1]$, $A \subset \mathbb{R}^n$, and $f : A \to \mathbb{R}^m$ be a γ-Hölder continuous function. Then for every $0 \le s < \infty$,*

$$\mathcal{H}^s(f(A)) \le \frac{\alpha_s [f]^s_{C^{0,\gamma}(A)}}{\alpha_{\gamma s} 2^{(1-\gamma)s}} \mathcal{H}^{\gamma s}(A). \tag{5.11}$$

Proof. We fix $\delta > 0$ and choose subsets $(E_j)_{j=1}^{\infty}$ of $\mathbb{R}^n$ with $A \subset \bigcup_{j=1}^{\infty} E_j$ and $\operatorname{diam}(E_j) \le \delta$ for each j. Then

$$\operatorname{diam} f(E_j) = \sup_{x,y \in E_j} |f(x) - f(y)| \le [f]_{C^{0,1}(A)} \sup_{x,y \in E_j} |x - y|^\gamma \le [f]_{C^{0,1}(A)} (\operatorname{diam} E_j)^\gamma \le \delta^\gamma [f]_{C^{0,1}(A)}.$$

Hence

$$\mathcal{H}^s_{\delta^\gamma [f]_{C^{0,1}(A)}}(f(A)) \le \sum_{j=1}^{\infty} \alpha_s \left(\frac{\operatorname{diam} f(E_j)}{2}\right)^s \le [f]^s_{C^{0,1}(A)} \sum_{j=1}^{\infty} \frac{\alpha_s}{2^{(1-\gamma)s}} \left(\frac{\operatorname{diam} E_j}{2}\right)^{\gamma s}.$$

Taking the infima over all $(E_j)_{j=1}^{\infty}$, we have $\mathcal{H}^s_{\delta^\gamma [f]_{C^{0,1}(A)}}(f(A)) \le \frac{\alpha_s [f]^s_{C^{0,\gamma}(A)}}{\alpha_{\gamma s} 2^{(1-\gamma)s}} \mathcal{H}^{\gamma s}_\delta(A)$. Send $\delta \downarrow 0$. $\square$

Remark. (I) By estimate (5.11), for all $s > \frac{\dim_{\mathcal{H}}(A)}{\gamma}$, we have $\mathcal{H}^s(f(A)) = 0$. Then

$$\dim_{\mathcal{H}}(f(A)) \le \frac{\dim_{\mathcal{H}}(A)}{\gamma}. \tag{5.12}$$

By countable stability of Hausdorff dimension, (5.12) also holds for locally γ-Hölder continuous functions.

(II) In particular, for a Lipschitz continuous function $f : A \to \mathbb{R}^m$, we have

$$\mathcal{H}^s(f(A)) \leq [f]^s_{C^{0,1}(A)} \mathcal{H}^s(A).$$

One typical example is the projection map. If $P : \mathbb{R}^n \to \mathbb{R}^n$ is a projection, i.e. $Px \perp x - Px$ for all $x \in \mathbb{R}^n$, then $[P]_{C^{0,1}(\mathbb{R}^n)} = 1$, and $\mathcal{H}^s(P(A)) \leq \mathcal{H}^s(A)$ for all $A \subset \mathbb{R}^n$.

Next we study the Hausdorff dimension of function graphs. Given $A \subset \mathbb{R}^n$ and a function $f : \mathbb{R}^n \to \mathbb{R}^m$, we write the *graph of f over A* by

$$\mathrm{Gr}_f(A) = \{(x, f(x)) : x \in A\} \subset \mathbb{R}^{n+m}.$$

Theorem 5.18 (Hausdorff dimension of graphs)**.** *Let $A \subset \mathbb{R}^n$ with $\mathcal{L}^n(A) > 0$, and $f : \mathbb{R}^n \to \mathbb{R}^m$.*
(i) $\dim_{\mathcal{H}}(\mathrm{Gr}_f(A)) \geq n$*;*
(ii) If f is locally γ-Hölder continuous for some exponent $\gamma > 0$, then

$$\dim_{\mathcal{H}}(\mathrm{Gr}_f(A)) \leq n + (1-\gamma)\left(m \wedge \frac{1}{\gamma}\right).$$

Proof. (i) Consider the standard projection $P : \mathbb{R}^{n+m} \to \mathbb{R}^n, (x_1, \cdots, x_n, x_{n+1}, \cdots, x_{n+m}) \mapsto (x_1, \cdots, x_n)$. Then $[P]_{C^{0,1}(\mathbb{R}^{n+m})} = 1$, and $\mathcal{H}^n(\mathrm{Gr}_f(A)) \geq \mathcal{H}^n(P(\mathrm{Gr}_f(A))) = \mathcal{H}^n(A) > 0$. Hence $\dim_{\mathcal{H}} \mathrm{Gr}_f(A) \geq n$.

(ii) We let Q denote any cube in $\mathbb{R}^n$ of side length 1, and divide Q into k^n subcubes $Q_1, Q_2, \cdots, Q_{k^n}$ of side length $\frac{1}{k}$. Then $\operatorname{diam} Q_j = \frac{\sqrt{n}}{k}$ for each j. Define

$$a^i_j = \inf_{x \in A \cap Q_j} f^i(x), \quad b^i_j = \sup_{x \in A \cap Q_j} f^i(x), \quad i = 1, \cdots, m, \quad j = 1, \cdots, k^n.$$

By γ-Hölder continuity,

$$b^i_j - a^i_j \leq [f]_{C^{0,\gamma}(Q)} \operatorname{diam}(Q_j \cap A)^\gamma \leq [f]_{C^{0,\gamma}(Q)} \left(\frac{\sqrt{n}}{k}\right)^\gamma.$$

Then the image of f over $A \cap Q_j$ satisfies $f(A \cap Q_j) = \prod_{i=1}^m [a^i_j, b^i_j]$.

(ii.1) By our estiamte of $b^i_j - a^i_j$, each the image $f(A \cap Q_j)$ can be covered by $Ck^{m(1-\gamma)}$ cubes in $\mathbb{R}^m$ of side length $\frac{1}{k}$, where C is a constant depending on f, n, m and γ. Consequently, the graph

$$\mathrm{Gr}_f(A \cap Q) \subset \bigcup_{j=1}^{k^n} Q_j \times f(A \cap Q_j)$$

is covered by a constant multiple of $k^{n+(1-\gamma)m}$ cubes in $\mathbb{R}^{n+m}$ of side length $\frac{1}{k}$. Then

$$\begin{aligned}\mathcal{H}^{n+(1-\gamma)m}_{\frac{\sqrt{n+m}}{k}}(\mathrm{Gr}_f(A \cap Q)) &\leq Ck^{n+(1-\gamma)m} \alpha_{n+(1-\gamma)m} \left(\frac{\sqrt{n+m}}{2k}\right)^{n+(1-\gamma)m} \\ &\leq C\alpha_{n+(1-\gamma)m} \left(\frac{\sqrt{n+m}}{2}\right)^{n+(1-\gamma)m}.\end{aligned}$$

Letting $k \uparrow \infty$, we obtain $\mathcal{H}^s(\mathrm{Gr}_f(A \cap Q)) < \infty$, and

$$\dim_{\mathcal{H}}(\mathrm{Gr}_f(A \cap Q)) \leq n + (1-\gamma)m,$$

which holds for all cubes $Q \subset \mathbb{R}^n$. By countable stability of Hausdorff dimension, we can subdivide $\mathbb{R}^n$ into countably many cubes and conclude that $\dim_{\mathcal{H}}(\mathrm{Gr}_f(A)) \leq n + (1-\gamma)m$.

(ii.2) We let $E_j = Q_j \times f(A \cap Q_j)$, $j = 1, 2, \cdots, k^n$, which together cover $\mathrm{Gr}_f(A \cap Q)$. Then

$$\operatorname{diam} E_j \le \sqrt{\frac{n}{k^2} + \sum_{i=1}^m (b_j^i - a_j^i)^2} \le \frac{1}{k^\gamma} \sqrt{n + mn^\gamma [f]_{C^{0,\gamma}(Q)}^2} =: \frac{C}{k^\gamma}.$$

Consequently,

$$\mathcal{H}^{n/\gamma}_{k^{-\gamma}C}(A \cap Q) \le \sum_{j=1}^{k^n} \alpha_{n/\gamma} \left(\frac{\operatorname{diam} E_j}{2}\right)^{n/\gamma} \le \alpha_{n/\gamma} \left(\frac{C}{2}\right)^{n/\gamma}.$$

Letting $k \uparrow \infty$, we obtain $\mathcal{H}^{n+(1-\gamma)m}(\mathrm{Gr}_f(A \cap Q)) < \infty$, and $\dim_{\mathcal{H}}(\mathrm{Gr}_f(A \cap Q)) \le n/\gamma$. Again by countable stability of Hausdorff dimension, we have $\dim_{\mathcal{H}}(\mathrm{Gr}_f(A)) \le \frac{n}{\gamma}$, which complete the proof. $\square$

Remark. In particular, if $f : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous on A, then $\dim_{\mathcal{H}}(\mathrm{Gr}_f(A)) = n$.

Aside: Kirszbraun's Extension Theorem. In some scenarios, we may concerns if a Lipschitz continuous function on a subset $A \subsetneq \mathbb{R}^n$ can be extended to the whole space. We can indeed prove the possibility of extension in a more general setting. Let X, Y be two Hilbert spaces, $A \subset X$, and $f : A \to Y$ be a Lipschitz continuous function. We write the global Lipschitz constant by

$$[f]_{C^{0,1}(A)} = \sup_{x,y \in X,\, x \neq y} \frac{\|f(x) - f(y)\|}{\|x - y\|}.$$

The formal statament is presented below.

Theorem 5.19 (Kirszbraun)**.** *Let X, Y be two Hilbert spaces, $A \subset X$, and $f : A \to Y$ a Lipschitz continuous function. Then there exists an extension $\bar{f} : X \to Y$ such that*
(i) $f = \bar{f}$ on A, and
(ii) $[\bar{f}]_{C^{0,1}(X)} = [f]_{C^{0,1}(A)}$.

To begin with, we first prove a weaker result.

Lemma 5.20. *Let $I, J \subset X$ be two finite sets, $K \ge 0$, and let $f : I \to Y$ be a function such that*

$$\|f(x) - f(y)\| \le K\|x - y\|$$

for all $x, y \in I$. Then there exists a function $g : I \cup J \to Y$ such that $g(x) = f(x)$ for all $x \in I$, and $\|g(x) - g(y)\| \le K\|x - y\|$ for all $x, y \in I \cup J$.

Proof. By induction on the number of points in J, it suffices to show the case $J = \{a\}$ with $a \in X \backslash I$. We claim that there exists $b \in Y$ such that $\|f(x) - b\| \le K\|x - a\|$ for each $x \in I$. If $K = 0$, then f is constant on I, and we take b to be the constant value of f. If $K > 0$, we may assume $K = 1$ by replacing f with f/K.

Step I. We write $I = \{x_1, \cdots, x_m\}$ and set $D = \mathrm{Conv}(f(I))$. Then D is a compact set for the norm topology since it is the image of the continuous mapping

$$(\lambda_1, \cdots, \lambda_m) \mapsto \lambda_1 f(x_1) + \cdots + \lambda_m f(x_m)$$

over the compact set $\{\lambda \in [0,1]^m : \lambda_1, \cdots, \lambda_m \ge 0,\, \mathbf{1}^\top \lambda = 1\}$.

Step II. For each $x \in I$, the function $z \mapsto \frac{\|z - f(x)\|}{\|a - x\|}$ from K to $[0, \infty)$ is continuous. Since I is finite,

$$h(z) = \max_{x \in I} \frac{\|z - f(x)\|}{\|a - x\|}$$

is also a continuous function, which attains its infimum over D.

Step III. We let $b \in K$ be such that $h(b) \leq h(z)$ for all $z \in D$. Define

$$I^* = \left\{ x \in I : \frac{\|b - f(x)\|}{\|a - x\|} = h(b) \right\},$$

which is a nonempty subset of I. We claim that $b \in \mathrm{Conv}(f(I^*))$ and argue this by contradiction.

Note that $\mathrm{Conv}(f(I^*))$ is a compact convex set. If $b \notin \mathrm{Conv}(f(I^*))$, we take $z_0 \in \mathrm{Conv}(f(I^*))$ such that $\|b - z_0\| = \inf_{z \in \mathrm{Conv}(f(I^*))} \|b - z\|$. Then for all $z \in \mathrm{Conv}(f(I^*))$,

$$\langle b - z_0, z - z_0 \rangle \leq -\frac{1}{2}\|z - z_0\|^2 \leq 0.$$

In particular, $\langle b - z_0, f(x) - z_0 \rangle \leq 0$ for every $x \in I^*$. Then we fix $0 < \epsilon \leq 1$ and take $b_\epsilon = b + \epsilon(z_0 - b) \in D$. For every $x \in I^*$,

$$\langle b_\epsilon - b, f(x) - b \rangle = \langle b_\epsilon - b, f(x) - z_0 \rangle + \langle b_\epsilon - b, z_0 - b \rangle \geq \epsilon \|z_0 - b\|^2.$$

Consequently,

$$\begin{aligned} \|f(x) - b_\epsilon\|^2 &= \|f(x) - b\|^2 - 2\langle b_\epsilon - b, f(x) - b \rangle + \|b_\epsilon - b\|^2 \\ &\leq \|f(x) - b\|^2 - (2\epsilon - \epsilon^2)\|z_0 - b\|^2 < \|f(x) - b\|^2. \end{aligned}$$

Recalling the definition of I^*, we have

$$\frac{\|f(x) - b_\epsilon\|}{\|x - a\|} < \frac{\|f(x) - b\|}{\|x - a\|} = h(b), \quad x \in I^*.$$

On the other hand,

$$\lim_{\epsilon \downarrow 0} \frac{\|f(x) - b_\epsilon\|}{\|x - a\|} = \frac{\|f(x) - b\|}{\|x - a\|} < h(b), \quad x \in I \backslash I^*.$$

Since $I \backslash I^*$ is a finite set, we can find $\epsilon_1 > 0$ such that for each $\epsilon \in (0, \epsilon_1)$,

$$\frac{\|f(x) - b_\epsilon\|}{\|x - a\|} < h(b), \quad \text{for all } x \in I.$$

This implies $h(b_\epsilon) < h(b)$, contradicting the fact that $h(b) \leq h(z)$ for all $z \in D$.

Step IV. We assume $h(b) > 1$, which implies that $\|f(x) - b\| = h(b)\|x - a\| > \|x - a\|$ for all $x \in I^*$. Note that $\|f(x) - f(y)\| \leq \|x - y\|$ for all $x, y \in I^*$. Then we have

$$\begin{aligned} \langle x - a, y - a \rangle &= \frac{1}{2}\left(\|x - a\|^2 + \|y - a\|^2 - \|x - y\|^2\right) \\ &< \frac{1}{2}\left(\|f(x) - b\|^2 + \|f(y) - b\|^2 - \|f(x) - f(y)\|^2\right) = \langle f(x) - b, f(y) - b \rangle. \end{aligned}$$

Since $b \in \mathrm{Conv}(f(I^*))$, we write $b = \sum_{x \in I^*} \lambda_x f(x)$, where $\lambda_x \geq 0$ for all $x \in I^*$ and $\sum_{x \in I^*} \lambda_x = 1$. Then

$$\begin{aligned} 0 = \left\langle \sum_{x \in I^*} \lambda_x f(x) - b, \sum_{y \in I^*} \lambda_y f(y) - b \right\rangle &= \sum_{x, y \in I^*} \lambda_x \lambda_y \langle f(x) - b, f(y) - b \rangle \\ &> \sum_{x, y \in I^*} \lambda_x \lambda_y \langle x - a, y - a \rangle = \left\langle \sum_{x \in I^*} \lambda_x x - a, \sum_{x \in I^*} \lambda_y y - a \right\rangle \geq 0, \end{aligned}$$

again a contradiction. Therefore $h(b) \leq 1$, that is, $\|f(x) - b\| \leq \|x - a\|$ for all $x \in I$.

Step VI. Finally, we set $g(x) = f(x)$ for all $x \in I$, and $g(a) = b$. Then g has the desired property. $\square$

To prove Kirszbraun's Theorem, we need extend the above result from finite to infinite sets. A key tool is the product topology and Tychonoff's theorem.

Proof of Theorem 5.19. We may assume A is nonempty, otherwise we just make $f \equiv 0$. Fix $a \in A$.

Step I. For each $x \in X$, let

$$B_x = \Big\{ y \in Y : \|y - f(a)\| \le [f]_{C^{0,1}(A)} \|x - a\| \Big\},$$

which is a compact subset Y under the weak topology. By Tychonoff's Theorem, $F = \prod_{x \in X} B_x$ is also compact in the product topology. Next, for any finite set $I \subset X$, set

$$F_I = \Big\{ g \in F : g(x) = f(x) \text{ for all } x \in I \cap A,\ \|g(x) - g(y)\| \le [f]_{C^{0,1}(A)} \|x - y\| \text{ for all } x, y \in I \Big\}.$$

Clearly $F = F_{\{a\}} \supset F_I$ for any finite subset $I \subset X$. Furthermore, by Lemma 5.20, we can find an extension g_0 of f on $I \cup \{a\}$ such that $g_0(x) = f(x)$ for all $x \in (I \cap A) \cup \{a\}$ and $\|g(x) - g(y)\| \le [f]_{C^{0,1}(A)} \|x - y\|$ for all $x, y \in I \cup \{a\}$. We let $g(x) = g_0(x)$ for $x \in I \cup \{a\}$ and $g(x) = f(a)$ for $x \in X \backslash (I \cup \{a\})$. Then

$$\|g(x) - f(a)\| \le \|g_0(x) - g_0(a)\| = \begin{cases} \|x - a\|, & x \in I \cup \{a\}, \\ 0, & x \notin I \cup \{a\}. \end{cases}$$

Therefore $g \in F$, and consequently F_I is nonempty.

Step II. Now we check that each set F_I is closed. To show this, we note that:

- For each $x \in (I \cap A) \cup \{a\}$, the projection map $g \mapsto g(x) : X \to B_x$ is continuous, and the set $\{f(x)\} \subset Y$ is closed in the weak topology. Then $\{g \in F : g(x) = f(x)\}$ is closed.
- For all $x, y \in I \cup \{a\}$, the maps $g \mapsto g(x)$ and $g \mapsto g(y)$ from F to Y given its weak topology are continuous. Also, the function $z \mapsto \langle z, w \rangle : Y \to \mathbb{R}$ is continuous for each $w \in Y$. Then the functions $g \mapsto \langle g(x) - g(y), w \rangle$ from F to $\mathbb{R}$ are continuous. Therefore

  $$\{g \in F : \langle g(x) - g(y), w \rangle \le [f]_{C^{0,1}(A)} \|x - y\|\}$$

  is closed in F given the product weak topology. Also,

  $$\{g \in F : \|g(x) - g(y)\| \le [f]_{C^{0,1}(A)} \|x - y\|\} = \bigcap_{\|w\| \le 1} \{g \in F : \langle g(x) - g(y), w \rangle \le [f]_{C^{0,1}(A)} \|x - y\|\}$$

  is an intersection of closed subsets of F, hence is also closed.
- Finally, the intersection

  $$F_I = \left(\bigcap_{x \in I \cap A} \{g \in F : g(x) = f(x)\} \right) \cap \left(\bigcap_{x, y \in I} \{g \in F : \|g(x) - g(y)\| \le [f]_{C^{0,1}(A)} \|x - y\|\} \right)$$

  is also closed.

Step III. We define $\mathscr{F} = \{F_I : I \subset X \text{ is finite}\}$, which is a collection of closed subsets of F. Also, $\mathscr{F}$ has the finite intersection property. For finite subset $I_1, \cdots, I_m$ of X, then $I = \bigcup_{j=1}^m I_j$ is also a finite subset of X, and

$$\bigcap_{j=1}^m F_{I_j} = \bigcap_{j=1}^m F_{I_j \cap \{a\}} \supset F_{I \cap \{a\}} = F_I \neq \emptyset.$$

Step IV. By compactness of F, the intersection of all members of $\mathscr{F}$ is nonempty, and we take an element $\bar{f} \in \bigcap_{I \subset X \text{ is finite}} F_I$. In particular, for all $x \in A$, we have $\bar{f} \in F_{\{x,a\}}$ and $\bar{f}(x) = f(x)$. Also, for all $x, y \in X$, we have $\bar{f} \in F_{\{x,y,a\}}$ and $\|\bar{f}(x) - \bar{f}(y)\| \le [f]_{C^{0,1}(A)} \|x - y\|$. Hence $\bar{f}$ is the desired extension. $\square$

5.5 Hausdorff Measure under Linear Transformations

Review: Linear maps. Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. If we equip both $\mathbb{R}^n$ and $\mathbb{R}^m$ the standard orthonormal bases, we can identify L with a $m \times n$ matrix $L = (L_{ij})_{1\le i\le m, 1\le j\le n}$. We can establish the equivalence of a series of concepts and properties between linear maps and matrices:

- (Adjoint). L^* is the conjugate transpose matrix of $L \;\Leftrightarrow\; \langle Lx, y\rangle = \langle x, L^* y\rangle$.
- (Symmetry). $S^* = S \;\Leftrightarrow\; \langle Sx, y\rangle = \langle x, Sy\rangle$.
- (Orthogonality). $Q^*Q = \mathrm{Id} \;\Leftrightarrow\; \langle Qx, Qy\rangle = \langle x, y\rangle$.

We then introduce the *polar decomposition* of linear maps. Given a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$, we take its singular value decomposition

$$L = \sum_{j=1}^k \sigma_i u_i v_j^\top,$$

where $\mathrm{rank}(L) = \dim \mathfrak{R}(L) = k$, the singular values $\sigma_1, \cdots, \sigma_k > 0$, $\{u_1, \cdots, u_m\} \subset \mathbb{R}^m$ is some orthonormal basis of $\mathbb{R}^m$, and $\{v_1, \cdots, v_n\} \subset \mathbb{R}^n$ is some orthonormal basis of $\mathbb{R}^n$.

(i) If $m \ge n$, there exists a symmetric map $S : \mathbb{R}^n \to \mathbb{R}^n$ and an orthogonal map $Q : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$L = QS.$$

To see this, we take the singular value decomposition $Q = \sum_{i=1}^n u_i v_i^\top$ and $S = \sum_{i=1}^k \sigma_i v_i v_i^\top$.

(ii) If $m \le n$, there exists a symmetric map $S : \mathbb{R}^m \to \mathbb{R}^m$ and an orthogonal map $Q : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$L = SQ^*.$$

To see this, we take the singular value decomposition $Q = \sum_{i=1}^m v_i u_i^\top$ and $S = \sum_{i=1}^k \sigma_i u_i u_i^\top$.

We define the *Jacobian of* L as the determinant of the symmetric matrix S in the polar decomposition:

$$[\![L]\!] = [\![L^*]\!] = \det S = \begin{cases} \sqrt{\det(L^*L)}, & m \ge n, \\ \sqrt{\det(LL^*)}, & m \le n. \end{cases}$$

Lemma 5.21. *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map, and $A \subset \mathbb{R}^n$.*

(i) If $n \le m$, then

$$\mathcal{H}^n(L(A)) = [\![L]\!]\, \mathcal{L}^n(A).$$

(ii) If $n \ge m$ and $A \subset \mathbb{R}^n$ is $\mathcal{L}^n$-measurable, then $y \mapsto \mathcal{H}^{n-m}(A \cap L^{-1}\{y\})$ is $\mathcal{L}^m$-measurable, and

$$\int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap L^{-1}\{y\})\, dy = [\![L]\!]\, \mathcal{L}^n(A).$$

Proof. (i) We take the polar decomposition $L = QS$ as above.

Case i.1. If $[\![L]\!] = \det S = 0$, then $\dim S(\mathbb{R}^n) \le n-1$, and $\dim L(\mathbb{R}^n) \le n-1$, hence $\mathcal{H}^n(L(A)) = 0$.

Case i.2. If $[\![L]\!] = \det S > 0$, by Theorem 5.17, for any ball $B(x, r) \subset \mathbb{R}^n$, we have

$$\frac{\mathcal{H}^n(L(B(x,r)))}{\mathcal{L}^n(B(x,r))} = \frac{\mathcal{H}^n(Q^*L(B(x,r)))}{\mathcal{L}^n(B(x,r))} = \frac{\mathcal{L}^n(S(B(x,r)))}{\mathcal{L}^n(B(x,r))} = \frac{\mathcal{L}^n(S(B(0,1)))}{\mathcal{L}^n(B(0,1))} = \det S = [\![L]\!].$$

Next, we define $\nu(A) = \mathcal{H}^n(L(A))$ for all $A \subset \mathbb{R}^n$. Then ν is a Radon measure on $\mathbb{R}^n$ with $\nu \ll \mathcal{L}^n$, and by Theorem 2.29,

$$\frac{d\nu}{d\mathcal{L}^n}(x) = \lim_{r\downarrow 0} \frac{\nu(B(x,r))}{\mathcal{L}^n(B(x,r))} = [\![L]\!].$$

Hence for all Borel sets $B \subset \mathbb{R}^n$, we have $\mathcal{H}^n(L(B)) = \nu(B) = [\![L]\!]\, \mathcal{L}^n(B)$. the same formula holds for all subsets $A \subset \mathbb{R}^n$ by Borel regularity of Hausdorff measures.

(ii) We take the polar decomposition $L = SQ^*$ as above.

Case ii.1. If $\dim(L(\mathbb{R}^n)) < m$, we have $A \cap L^{-1}\{y\} = \emptyset$ for $\mathcal{L}^n$-a.e. $y \in \mathbb{R}^m$, and $\mathcal{L}^{n-m}(A \cap L^{-1}\{y\}) = 0$. Also, since $S = LQ$, we have $S(\mathbb{R}^m) = L(\mathbb{R}^n)$, which implies $\dim(S(\mathbb{R}^m)) < m$ and $[\![L]\!] = \det S = 0$.

Case ii.2. If $L = P$ is the projection map $(x_1, \cdots, x_m, x_{m+1}, \cdots, x_n) \mapsto (x_1, \cdots, x_m)$, then for each $y \in \mathbb{R}^m$, $P^{-1}\{y\} = P^{-1}\{0\} + P^*y$ is an $(n-m)$-dimensional affine subspace of $\mathbb{R}^n$ and a translation of $P^{-1}\{0\}$. By Fubini's theorem, $y \mapsto \mathcal{H}^{n-m}(A \cap P^{-1}\{y\})$ is $\mathcal{L}^m$-measurable, and

$$\begin{aligned}\int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap P^{-1}\{y\})\, dy &= \int_{\mathbb{R}^m} \int_{P^{-1}\{y\}} \mathbb{1}_A(x)\, d\mathcal{H}^{n-m}(x)\, dy = \int_{\mathbb{R}^m} \int_{P^{-1}\{0\}} \mathbb{1}_A(x + P^*y)\, d\mathcal{H}^{n-m}(x)\, dy \\ &= \int_{\mathbb{R}^m} \int_{\mathbb{R}^{n-m}} \mathbb{1}_A \begin{pmatrix} y \\ x \end{pmatrix} d\mathcal{L}^{n-m}(x)\, d\mathcal{L}^m(y) = \int_{\mathbb{R}^n} \mathbb{1}_A\, d\mathcal{L}^n = \mathcal{L}^n(A).\end{aligned}$$

Case ii.3. For the general case of $\dim(L(\mathbb{R}^n)) = m$, we write $Q^* = PU$, where $U : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal map. Then $L = SPU$. Similar to the Case ii.2, $L^{-1}\{0\}$ is an $(n-m)$-dimensional subspace of $\mathbb{R}^n$ and $L^{-1}\{y\} = L^{-1}\{0\} + QS^{-1}y$ is a translation of $L^{-1}\{0\}$ for each $y \in \mathbb{R}^m$. By Fubini's theorem, $y \mapsto \mathcal{H}^{n-m}(A \cap L^{-1}\{y\})$ is $\mathcal{L}^m$-measurable. Using the conclusion in Case ii.2, we have

$$\mathcal{L}^n(A) = \mathcal{L}^n(U(A)) = \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(U(A) \cap P^{-1}\{y\})\, dy = \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap (U^{-1} \circ P^{-1}\{y\}))\, dy$$

We set $z = Sy$. By the change of variables formula,

$$\mathcal{L}^n(A) = \frac{1}{|\det S|} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap (U^{-1} \circ P^{-1} \circ S^{-1}\{z\}))\, dz = \frac{1}{[\![L]\!]} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap (L^{-1}\{z\}))\, dz$$

Then we finish the proof. □

Remark. In the remain of this section, we will apply (i) to establish the area formula and (ii) to establish the coarea formula. Note that in the proof of (ii), we use the change of variables formula, which is an immediate corollary of the area formula.

5.6 The Area Formula

Throughout this subsection, we assume $n \le m$ and study the area formula. For a locally Lipschitz map $f = (f^1, f^2, \cdots f^m) : \mathbb{R}^n \to \mathbb{R}^m$, we define its Jacobian

$$Jf(x) = [\![Df(x)]\!], \quad \text{where } Df(x) = \begin{bmatrix} f^1_{x_1} & f^1_{x_2} & \cdots & f^1_{x_n} \\ f^2_{x_1} & f^2_{x_2} & \cdots & f^2_{x_n} \\ \vdots & \vdots & \ddots & \vdots \\ f^m_{x_1} & f^m_{x_2} & \cdots & f^m_{x_n} \end{bmatrix}$$

The area formula points out that, for every $A \subset \mathbb{R}^n$, the n-dimensional measure of the image $f(A) \subset \mathbb{R}^m$, counting multiplicity, can be computed by integrating the Jacobian Jf over A.

Definition 5.22. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a Lipschitz continuous function, $n \le m$, and $A \subset \mathbb{R}^n$. The mapping $y \mapsto \mathcal{H}^0(A \cap f^{-1}(y))$ is the *multiplicity function* of f.

Remark. The multiplicity function counts the number of points of A that are taken to the point $y \in \mathbb{R}^m$ by function f. We shall see that $f^{-1}\{y\}$ is at most countable for $\mathcal{H}^n$-a.e. $y \in \mathbb{R}^m$.

Theorem 5.23 (Area formula)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a Lipschitz continuous function, $n \le m$. Then for each $\mathcal{L}^n$-measurable subset $A \subset \mathbb{R}^n$,*

$$\int_A Jf(x)\, dx = \int_{\mathbb{R}^m} \mathcal{H}^0(A \cap f^{-1}\{y\})\, d\mathcal{H}^n(y).$$

From now on, we fix a Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}^m$, $n \le m$. To prove the are formula, we need to introduce some technical lemmata.

Lemma 5.24. *Let $A \subset \mathbb{R}^n$ be $\mathcal{L}^n$-measurable. Then*
(i) $f(A)$ is $\mathcal{H}^n$-measurable,
(ii) the multiplicity function $y \mapsto \mathcal{H}^0(A \cap f^{-1}(y))$ is $\mathcal{H}^n$-measurable, and

$$\int_{\mathbb{R}^m} \mathcal{H}^0(A \cap f^{-1}(y))\, d\mathcal{H}^n \le [f]_{C^{0,1}}^n \mathcal{L}^n(A).$$

Proof. We may assume that A is bounded by taking the intersections with countably many cubes covering $\mathbb{R}^n$.

Step I. By the inner regularity of $\mathcal{L}^n$, there exist compact sets $K_m \subset A$ such that $\mathcal{L}^n(A \backslash K_m) \le \frac{1}{m}$, $m \in \mathbb{N}$. By compactness of K_m and continuity of f, the images $f(K_m)$ are compact and $\mathcal{H}^n$-measurable. Also, the image $f(\bigcup_{m=1}^\infty K_m) = \bigcup_{m=1}^\infty f(K_m)$ is $\mathcal{H}^n$-measurable, and by Theorem 5.17,

$$\mathcal{H}^n\left(f(A)\backslash f\left(\bigcup_{m=1}^\infty K_m\right)\right) = \mathcal{H}^n\left(f\left(A\backslash \bigcup_{m=1}^\infty K_m\right)\right) \le [f]_{C^{0,1}}^n \mathcal{L}^n\left(A\backslash \bigcup_{m=1}^\infty K_m\right) = 0.$$

Therefore $f(A)$ is $\mathcal{H}^n$-measurable, which proves (i).

Step II. We subdivide $\mathbb{R}^n$ into cubes of side length 2^{-k}, and write

$$\mathbb{R}^n = \bigcup_{Q \in \mathscr{B}_k} Q, \quad \mathscr{B}_k = \left\{ \prod_{i=1}^n \left(\frac{a_i}{2^k}, \frac{b_i}{2^k} \right] : a_i, b_i \in \mathbb{Z},\ i = 1, \cdots, n \right\}.$$

Then the function

$$g_k = \sum_{Q \in \mathscr{B}_k} \mathbb{1}_{f(A \cap Q)}$$

is $\mathcal{H}^n$-measurable by (i). Furthermore, for each point $y \in \mathbb{R}^m$, $g_k(y)$ is the number of cubes $Q \in \mathscr{B}_k$ such that $f^{-1}\{y\} \cap (A \cap Q) \ne \emptyset$. As $k \to \infty$, we have $g_k(y) \uparrow \mathcal{H}^0(A \cap f^{-1}\{y\})$ for each $y \in \mathbb{R}^m$. Therefore $y \mapsto \mathcal{H}^0(A \cap f^{-1}\{y\})$ is $\mathcal{H}^n$-measurable. By monoton convergence theorem,

$$\begin{aligned}\int_{\mathbb{R}^m} \mathcal{H}^0(A \cap f^{-1}(y))\, d\mathcal{H}^n &= \lim_{k\to\infty} \int_{\mathbb{R}^m} \sum_{Q \in \mathscr{B}_k} \mathbb{1}_{f(A\cap Q)}\, d\mathcal{H}^n = \lim_{k\to\infty} \sum_{Q \in \mathscr{B}_k} \mathcal{H}^n\left(f(A \cap Q)\right) \\ &\le \limsup_{k\to\infty} \sum_{Q \in \mathscr{B}_k} [f]_{C^{0,1}}^n \mathcal{L}^n\left(A \cap Q\right) \le [f]_{C^{0,1}}^n \mathcal{L}^n\left(A\right)\end{aligned}$$

Then we complete the proof. □

Lemma 5.25. *Let $t > 1$, and define*

$$P = \{x \in \mathbb{R}^n : Df(x) \text{ exists, } Jf(x) > 0\}.$$

Then there is a countable collection $(E_k)_{k=1}^\infty$ of Borel subsets of $\mathbb{R}^n$ such that
(i) $\bigcup_{k=1}^\infty E_k = P$,
(ii) $f|_{E_k}$ is one-to-one ($k = 1, 2, \cdots$); and
(iii) For each $k = 1, 2, \cdots$, there exists a nonsingular symmetric linear map $T_k : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$[(f|_{E_k}) \circ T_k^{-1}]_{C^{0,1}} \le t, \quad [T_k \circ (f|_{E_k})^{-1}]_{C^{0,1}} \le t, \quad \text{and} \quad t^{-n}|\det T_k| \le Jf|_{E_k} \le t^n |\det T_k|.$$

Proof. We fix $\epsilon > 0$ such that $t^{-1} + \epsilon < 1 < t - \epsilon$. Let Q be a countable dense subset of P, and let $\mathcal{S}$ be a countable dense subset of nonsingular symmetric linear map on $\mathbb{R}^n$.

Step I. For each $q \in Q$, $T \in \mathcal{S}$ and $l \in \mathbb{N}$, define

$$E(q,T,l) = \Big\{x \in Q \cap B\left(q, l^{-1}\right) : \left(t^{-1}+\epsilon\right)|Tz| \le |Df(x)z| \le (t-\epsilon)|Tz| \text{ for all } z \in \mathbb{R}^n,$$
$$\text{and } |f(y)-f(x)-Df(x)(y-x)| \le \epsilon|T(x-y)| \text{ for all } y \in B\left(x, 2l^{-1}\right)\Big\}$$

Since Df is Borel measurable, $E(q,T,l)$ is a Borel set. Also, for all $x \in E(q,T,l)$ and $y \in B(x, \frac{2}{m})$,

$$t^{-1}|T(y-x)| \le |f(y)-f(x)| \le t|T(y-x)|. \tag{5.13}$$

We claim that for each $x \in E(q,T,l)$,

$$\left(t^{-1}+\epsilon\right)^n |\det T| \le Jf(x) \le (t-\epsilon)^n |\det T|. \tag{5.14}$$

Consider the polar decomposition $Df(x) = L(x) = Q(x)S(x)$, so $Jf(x) = [\![L(x)]\!] = |\det S(x)|$. For all $z \in \mathbb{R}^n$,

$$\left(t^{-1}+\epsilon\right)^n |Tz| \le |Q(x)S(x)z| = |S(x)z| \le (t-\epsilon)|Tz|,$$

and hence

$$\left(t^{-1}+\epsilon\right)^n |z| \le |S(x)T^{-1}z| \le (t-\epsilon)|z|.$$

Consequently, we have $(S(x)\circ T^{-1})B(0,1) \subset B(0, t-\epsilon)$, and

$$\alpha_n |\det(S(x)T^{-1})| \le \alpha_n (t-\epsilon)^n.$$

Hence $Jf(x) = \det|S(x)| \le (t-\epsilon)^n |\det T|$. The proof of the other inequality in (5.14) is similar.

Step II. We relabel the countable collection $\{E(q,T,l) : q \in Q, T \in \mathcal{S}, l \in \mathbb{N}\}$ as $\{E_k\}_{k=1}^{\infty}$. For each $x \in P$, we write the polar decomposition $Df(x) = Q(x)S(x)$ as above. Choose $T \in \mathcal{S}$ with $\|T - S(x)\|$ so small that

$$\|TS(x)^{-1}\| = \|(T-S(x))S(x)^{-1} + \mathrm{Id}\,\| \le \frac{1}{t^{-1}+\epsilon}, \quad \text{and} \quad \|S(x)T^{-1}\| = \|(S(x)-T)T^{-1} + \mathrm{Id}\,\| \le t-\epsilon.$$

Then $|Df(x)z| = |Q(x)S(x)z| = |S(x)z|$ satisfies

$$\left(t^{-1}+\epsilon\right)|Tz| \le \frac{|Tz|}{\|TS(x)^{-1}\|} \le |S(x)z| \le \|S(x)T^{-1}\| |Tz| \le (t-\epsilon)\,|Tz|\,, \quad \forall z \in \mathbb{R}^n.$$

Next, we choose $m \in \mathbb{N}$ and $q \in Q$ such that $q \in B(x, \frac{1}{l})$, and

$$|f(y)-f(x)-Df(x)(y-x)| \le \frac{\epsilon}{\|T^{-1}\|}|y-x| \le \epsilon|T(y-x)|$$

for all $y \in B(x, \frac{2}{l})$. Then $x \in E(q,T,l)$, which proves (i).

Step III. We fix any $E_k = E(q,T,l)$, and let $T_k = T$. By (5.13), for all $x \in E_k$ and $y \in B(x, \frac{2}{l})$,

$$t^{-1}|T_k(y-x)| \le |f(y)-f(x)| \le t|T_k(y-x)|.$$

Since $E_k \subset B(q, \frac{1}{l}) \subset B(x, \frac{2}{l})$, the above estimate in fact holds for all $x, y \in E_k$. Hence $f|_{E_k}$ is one-to-one, and

$$[(f|_{E_k})\circ T_k^{-1}]_{C^{0,1}} \le t, \quad [T_k \circ (f|_{E_k})^{-1}]_{C^{0,1}} \le t.$$

The estimate (5.14) implies

$$t^{-n}|\det T_k| \le Jf|_{E_k} \le t^n |\det T_k|.$$

Thus we complete the proof of (ii) and (iii). $\square$

Proof of Theorem 5.23. By Rademacher's Theorem, we may assume $Df(x)$ and $Jf(x)$ exist for all $x \in A$. We may also assume $\mathcal{L}^n(A) < \infty$.

Case I: $A \subset \{Jf > 0\}$**.** We fix $t > 1$ and choose $(E_k)_{k=1}^{\infty}$ as in Lemma 5.25. We may assume that $(E_k)_{k=1}^{\infty}$ are disjoint by sequentially removing the intersection. Also, we take the disjoint cubes $\mathscr{B}_l$ as in Lemma 5.24. We then set

$$F_{j,k}^l = E_k \cap Q_j \cap A, \quad Q_j \in \mathscr{B}_k, \quad j,k = 1,2,\cdots.$$

Then $(F_{j,k}^l)_{j,k=1}^{\infty}$ are disjoint, and $A = \bigcup_{j,k=1}^{\infty} F_{j,k}^l$. We define

$$g_l = \sum_{j,k=1}^{\infty} \mathbb{1}_{f(F_{j,k}^l)}.$$

Then for each $y \in \mathbb{R}^n$, $g_l(y)$ is the number of sets $F_{j,k}^l$ such that $F_{j,k}^l \cap f^{-1}\{y\} \neq \emptyset$. Consequently, $g_l(y) \uparrow \mathcal{H}^0(A \cap f^{-1}\{y\})$, and we use the monotone convergence theorem to conclude that

$$\int_{\mathbb{R}^m} \mathcal{H}^0(A \cap f^{-1}\{y\})\, d\mathcal{H}^n(y) = \lim_{l\to\infty} \int_{\mathbb{R}^m} g_l(y)\, d\mathcal{H}^n(y) = \lim_{l\to\infty} \sum_{j,k=1}^{\infty} \mathcal{H}^n(f(F_{j,k}^l)). \tag{5.15}$$

By estimate (5.13) and Theorem 5.17,

$$\mathcal{H}^n(f(F_{j,k}^l)) = \mathcal{H}^n(f|_{E_k} \circ T_k^{-1} \circ T_k(F_{j,k}^l)) \le t^n \mathcal{L}^n(T_k(F_{j,k}^l)),$$

and

$$\mathcal{L}^n(T_k(F_{j,k}^l)) = \mathcal{H}^n(T^k \circ (f|_{E_k})^{-1} \circ f|_{E_k}(F_{j,k}^l)) \le t^n \mathcal{H}^n(f(F_{j,k}^l)).$$

Next, we repeatedly use estimate (5.13) and Lemma 5.21 to obtain

$$\begin{aligned} t^{-2n}\mathcal{H}^n(f(F_{j,k}^l)) &\le t^{-n}\mathcal{L}^n(T_k(F_{j,k}^l)) = t^{-n}|\det T_k|\,\mathcal{L}^n(F_{j,k}^l) \\ &\le \int_{F_{j,k}^l} Jf\, d\mathcal{L}^n \\ &\le t^n|\det T_k|\,\mathcal{L}^n(F_{j,k}^l) = t^n\mathcal{L}^n(T_k(F_{j,k}^l)) = t^{2n}\mathcal{H}^n(f(F_{j,k}^l)) \end{aligned}$$

We sum on j and k to conclude

$$t^{-2n}\sum_{j,k=1}^{\infty} \mathcal{H}^n(f(F_{j,k}^l)) \le \int_A Jf\, d\mathcal{L}^n \le t^{2n}\sum_{j,k=1}^{\infty} \mathcal{H}^n(f(F_{j,k}^l))$$

Recalling (5.15), we can send $l \uparrow \infty$ and $t \downarrow 1$in the above estimate to obtain

$$\int_A Jf\, d\mathcal{L}^n = \int_{\mathbb{R}^m} \mathcal{H}^0(A \cap f^{-1}\{y\})\, d\mathcal{H}^n(y).$$

Case II: $A \subset \{Jf = 0\}$**.** We fix $0 < \epsilon < 1$, and define $g : \mathbb{R}^n \to \mathbb{R}^{m+n}$ as $g(x) = (f(x), \epsilon x)$. Then $f = P \circ g$, where $P : \mathbb{R}^{m+n} \to \mathbb{R}^m$ is the standard projection $P(y_1, \cdots, y_m, y_{m+1}, \cdots, y_n) = (y_1, \cdots, y_m)$. By definition,

$$Dg(x) = \begin{pmatrix} Df(x) \\ \epsilon\, \mathrm{Id}_{n\times n} \end{pmatrix},$$

and

$$Jg(x) = \sqrt{\det\left(Df(x)Df(x)^{\top} + \epsilon^2\, \mathrm{Id}_{n\times n}\right)} = \sqrt{\prod_{j=1}^{n}(\sigma_j(x)^2 + \epsilon^2)},$$

where $\sigma^1(x) \geq \cdots \geq \sigma_n^2(x) = 0$ are singular values of $Df(x)$ in decreasing order. By Lipschitz continuity of f over $\mathbb{R}^n$, we have $\sigma_1(x) = \|Df(x)\|_1 \leq [f]_{C^{0,1}}$. Also, since $Jf = 0$, we have $\sigma_n(x) = 0$. Therefore

$$0 < \epsilon^n \leq Jg(x) \leq \left(1 + [f]_{C^{0,1}}^2\right)^{\frac{n-1}{2}} \epsilon, \quad x \in A.$$

Since $P : \mathbb{R}^{m+n} \to \mathbb{R}^n$ is the standard projection, we apply Theorem 5.17 and the Case I above to obtain

$$\begin{aligned}\mathcal{H}^n(f(A)) \leq \mathcal{H}^n(g(A)) &= \int_{\mathbb{R}^{n+m}} \mathbb{1}_{g(A)}(y,z)\, d\mathcal{H}^n(y,z) \\ &\leq \int_{\mathbb{R}^{n+m}} \mathcal{H}^0(A \cap g^{-1}\{(y,z)\})\, d\mathcal{H}^n(y,z) = \int_A Jg\, d\mathcal{L}^n \leq \left(1 + [f]_{C^{0,1}}^2\right)^{\frac{n-1}{2}} \epsilon \mathcal{L}^n(A).\end{aligned}$$

Sending $\epsilon \downarrow 0$, we have $\mathcal{H}^n(f(A)) = 0$. Since $y \mapsto \mathcal{H}^0(A \cap f^{-1}\{y\})$ is supported on $f(A)$, we have

$$\int_{\mathbb{R}^m} \mathcal{H}^0(A \cap f^{-1}\{y\})\, d\mathcal{H}^n = 0 = \int_A Jf\, d\mathcal{L}^n.$$

Finally, for the general case $A \subset \mathbb{R}^n$, we just apply Cases I and II above to $A \cap \{Jf > 0\}$ and $A \cap \{Jf = 0\}$, respectively. Then we conclude the proof. $\square$

Remark. The area formula also implies that $f^{-1}\{y\}$ *is at most countable for* $\mathcal{H}^n$*-a.e.* $y \in \mathbb{R}^m$. Since f is a Lipschitz function, the Jacobian Jf is bounded on $\mathbb{R}^n$. Then for each cube $Q \subset \mathbb{R}$,

$$\int_{\mathbb{R}^m} \mathcal{H}^0(Q \cap f^{-1}\{y\})\, d\mathcal{H}^n = \int_Q Jf\, d\mathcal{L}^m < \infty,$$

and $\mathcal{H}^n\{y \in \mathbb{R}^m : \mathcal{H}^0(Q \cap f^{-1}\{y\}) = \infty\} = 0$. We take $\mathbb{R}^n = \bigcup_{k=1}^\infty Q_k$, where Q_k's are lattice cubes. Then $\mathcal{H}^0(Q_k \cap f^{-1}\{y\}) < \infty$ on each cube Q_k for $\mathcal{H}^n$-a.e. $y \in \mathbb{R}^m$, and $f^{-1}\{y\}$ is at most countable.

One most important corollary of the area formula is the change of variables formula.

Theorem 5.26 (Change of variables). *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz continuous, $n \leq m$. Then for each $\mathcal{L}^n$-integrable function $g : \mathbb{R}^n \to \mathbb{R}$,*

$$\int_{\mathbb{R}^n} g(x) Jf(x)\, dx = \int_{\mathbb{R}^m} \left[\sum_{x \in f^{-1}\{y\}} g(x)\right] d\mathcal{H}^n(y).$$

Proof. If $g \geq 0$ is measurable, we may write $g = \sum_{k=1}^\infty \frac{1}{k}\mathbb{1}_{A_k}$, where we define $A_1 = \{x \in \mathbb{R}^n : f(x) \geq 1\}$, and

$$A_k = \left\{x \in \mathbb{R}^n : f(x) \geq \frac{1}{k} + \sum_{j=1}^{k-1} \frac{1}{j}\mathbb{1}_{A_j}\right\}, \quad k = 2, 3, \cdots.$$

Then

$$\begin{aligned}\int_{\mathbb{R}^n} g(x) Jf(x)\, dx &= \sum_{k=1}^\infty \frac{1}{k} \int_{A_k} Jf(x)\, dx = \sum_{k=1}^\infty \frac{1}{k} \int_{\mathbb{R}^m} \mathcal{H}^0(A_k \cap f^{-1}\{y\})\, d\mathcal{H}^n(y) \\ &= \sum_{k=1}^\infty \frac{1}{k} \int_{\mathbb{R}^m} \sum_{x \in f^{-1}\{y\}} \mathbb{1}_{A_k}(x)\, d\mathcal{H}^n(y) = \int_{\mathbb{R}^m} \sum_{x \in f^{-1}\{y\}} \left[\sum_{k=1}^\infty \frac{1}{k}\mathbb{1}_{A_k}(x)\right] d\mathcal{H}^n(y) \\ &= \int_{\mathbb{R}^m} \left[\sum_{x \in f^{-1}\{y\}} g(x)\right] d\mathcal{H}^n(y).\end{aligned}$$

For the general case that g is an $\mathcal{L}^n$-integrable function, write $g = g^+ - g^-$ and apply the above conclusion. $\square$

Remark. In particular, if $f : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous and one-to-one, then for each $\mathcal{L}^n$-integrable function $g : \mathbb{R}^n \to \mathbb{R}$,

$$\int_{\mathbb{R}^n} g(x) Jf(x)\, dx = \int_{\mathbb{R}^n} g(f^{-1}(y))\, dy.$$

5.7 The Coarea Formula

Throughout this subsection, we assume $n \geq m$ and study the coarea formula. Given a Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}^m$, the coarea states that the integral of $(n-m)$-dimensional measure of level sets of f can be obtained by integrating the Jacobian. This is a kind of "curvilinear" generalization of Fubini's theorem.

Theorem 5.27 (Coarea formula)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a Lipschitz continuous function, $n \geq m$. Then for each $\mathcal{L}^n$-measurable set $A \subset \mathbb{R}^n$,*

$$\int_A Jf(x)\,dx = \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap f^{-1}\{y\})\,dy.$$

Like in Section 5.6, we fix a Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}^m$, $n \geq m$ and introduce some technical lemmata.

Lemma 5.28. *Let $A \subset \mathbb{R}^n$ be $\mathcal{L}^n$-measurable. Then*

(i) $A \cap f^{-1}\{y\}$ is $\mathcal{H}^{n-m}$-measurable for $\mathcal{L}^m$-a.e. $y \in \mathbb{R}^m$,

(ii) the function $y \mapsto \mathcal{H}^{n-m}(A \cap f^{-1}\{y\})$ is $\mathcal{L}^m$ measurable, and

$$\int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap f^{-1}\{y\})\,dy \leq \frac{\alpha_{n-m}\alpha_m}{\alpha_n}[f]_{C^{0,1}}^m \mathcal{L}^n(A).$$

Proof. **Step I.** For each $k \in \mathbb{N}$, there exists closed balls $(B_j^k)_{j=1}^\infty$ such that

$$A \subset \bigcup_{j=1}^\infty B_j^k, \quad \operatorname{diam} B_j^k \leq \frac{1}{k}, \quad \sum_{j=1}^\infty \mathcal{L}^n(B_j^k) \leq \mathcal{L}^n(A) + \frac{1}{k}.$$

We define

$$g_j^k = \alpha_{n-m}\left(\frac{\operatorname{diam} B_j^k}{2}\right)^{n-m} \mathbb{1}_{f(B_j^k)}, \quad j,k = 1,2,\cdots,$$

which is $\mathcal{L}^m$-measurable. Then for all $y \in \mathbb{R}^m$, we have $A \cap f^{-1}\{y\} \subset \bigcup_{j=1,\, y\in f(B_j^k)}^\infty B_j^k$, and

$$\mathcal{H}_{\frac{1}{k}}^{n-m}(A \cap f^{-1}\{y\}) \leq \sum_{j=1}^\infty g_j^k(y).$$

Thus, using Fatou's Lemma and the isodiametric inequality,

$$\begin{aligned}
\int_{\mathbb{R}^m}^* \mathcal{H}^{n-m}(A \cap f^{-1}\{y\})\,dy &= \int_{\mathbb{R}^m}^* \lim_{k\to\infty} \mathcal{H}_{\frac{1}{k}}^{n-m}(A \cap f^{-1}\{y\})\,dy \leq \int_{\mathbb{R}^m} \liminf_{k\to\infty} \sum_{j=1}^\infty g_j^k(y)\,dy \\
&\leq \liminf_{k\to\infty} \int_{\mathbb{R}^m} \sum_{j=1}^\infty g_j^k(y)\,dy = \liminf_{k\to\infty} \sum_{j=1}^\infty \alpha_{n-m}\left(\frac{\operatorname{diam} B_j^k}{2}\right)^{n-m} \mathcal{L}^m(f(B_j^k)) \\
&\leq \liminf_{k\to\infty} \sum_{j=1}^\infty \alpha_{n-m}\left(\frac{\operatorname{diam} B_j^k}{2}\right)^{n-m} \alpha_m \left(\frac{\operatorname{diam}(f(B_j^k))}{2}\right)^m \\
&\leq \frac{\alpha_{n-m}\alpha_m}{\alpha_n}[f]_{C^{0,1}}^m \liminf_{k\to\infty} \sum_{k=1}^\infty \mathcal{L}^n(B_j^k) \leq \frac{\alpha_{n-m}\alpha_m}{\alpha_n}[f]_{C^{0,1}}^m \mathcal{L}^n(A),
\end{aligned}$$

where $\int^*$ is the upper integral. This automatically establishes (iii) once we prove (ii).

Remark. The very same procedure also establish that

$$\int_{\mathbb{R}^m}^* \mathcal{H}^s(A \cap f^{-1}\{y\})\,dy \leq \frac{\alpha_s\alpha_m}{\alpha_{s+m}}[f]_{C^{0,1}}^m \mathcal{H}^{s+m}(A), \quad 0 \leq s \leq n-m. \tag{5.16}$$

Step II. We first assume that A is compact. Fix $t \geq 0$, and for each $k \in \mathbb{N}$, define

$$U_k = \Bigg\{ y \in \mathbb{R}^m : \text{there exists finitely many open sets } S_1, \cdots, S_l \text{ such that } A \cap f^{-1}\{y\} \subset \bigcup_{j=1}^{l} S_l, \\ \operatorname{diam} S_j \leq \frac{1}{k} \text{ for each } j = 1, \cdots, l, \text{ and } \sum_{j=1}^{l} \alpha_{n-m} \left(\frac{\operatorname{diam} S_j}{2} \right)^{n-m} \leq t + \frac{1}{k} \Bigg\}.$$

For each $y \in U_k$, assume $A \cap f^{-1}\{y\} \subset \bigcup_{j=1}^{l} S_l$ as above. Then by continuity of f and compactness of A, the image $f(A \backslash \bigcup_{j=1}^{l} S_l)$ is a compact set not containing $\{y\}$, and for $z \in \mathbb{R}^m$ sufficiently close to y, we have $z \notin f(A \backslash \bigcup_{j=1}^{l} S_l)$, and $A \cap f^{-1}(z) \subset \bigcup_{j=1}^{l} S_l$. Hence U_k is an open set.

Step III. We claim that

$$\{y \in \mathbb{R}^m : \mathcal{H}^{n-m}(A \cap f^{-1}\{y\}) \leq t\} = \bigcap_{k=1}^{\infty} U_k. \tag{5.17}$$

To establish this, we note that if $\mathcal{H}^{n-m}(A \cap f^{-1}\{y\}) \leq t$, then $\mathcal{H}_{\delta}^{n-m}(A \cap f^{-1}\{y\}) \leq t$ for each $\delta > 0$. For any $k \in \mathbb{N}$, choose $0 < \delta < \frac{1}{k}$. Then there exists sets $(S_j)_{j=1}^{\infty}$ such that $\bigcup_{j=1}^{\infty} S_j \supset A \cap f^{-1}\{y\}$, $\operatorname{diam} S_j \leq \delta < \frac{1}{k}$ for all $j \in \mathbb{N}$ and

$$\sum_{j=1}^{\infty} \alpha_{n-m} \left(\frac{\operatorname{diam} S_j}{2} \right)^{n-m} < t + \frac{1}{k}.$$

We may assume S_j's are open by replacing S_j with $\bigcup_{x \in S_j} B(x, \epsilon_j)$ with sufficiently small ϵ_j's. Since $A \cap f^{-1}\{y\}$ is compact, a finite subcollection $\{S_1, \cdots, S_l\}$ covers $A \cap f^{-1}\{y\}$, and hence $y \in S_k$ for all $k \in \mathbb{N}$.

On the other hand, if $y \in \bigcap_{k=1}^{\infty} U_k$, one have $\mathcal{H}_{\frac{1}{k}}^{n-m}(A \cap f^{-1}\{y\}) \leq t + \frac{1}{k}$ for each $k \in \mathbb{N}$. Letting $k \to \infty$ gives $\mathcal{H}^{n-m}(A \cap f^{-1}\{y\}) \leq t$. Thus we finish the proof of (5.17). Consequently, the set on the left side is Borel, and $y \mapsto \mathcal{H}^{n-m}(A \cap f^{-1}\{y\})$ is a Borel measurable function.

Step IV. If $A \subset \mathbb{R}^n$ is open, there exist increasing compact sets $K_1 \subset K_2 \subset \cdots \subset A$ such that $A = \bigcup_{j=1}^{\infty} K_j$. Then for each $y \in \mathbb{R}^m$,

$$\mathcal{H}^{n-m}(A \cap f^{-1}\{y\}) = \lim_{j \to \infty} \mathcal{H}^{n-m}(K_j \cap f^{-1}\{y\}).$$

Hence $y \mapsto \mathcal{H}^{n-m}(A \cap f^{-1}\{y\})$ is a Borel measurable function for every open subset $A \subset \mathbb{R}^n$.

Step V. If $A \subset \mathbb{R}^n$ is $\mathcal{L}^n$-measurable and $\mathcal{L}^n(A) < \infty$, by Borel regularity and outer regularity, there exists open sets $V_1 \supset V_2 \supset \cdots \supset A$ such that $\mathcal{L}(V_j \backslash A) \to 0$ and $\mathcal{L}^n(V_1) < \infty$. Then

$$\mathcal{H}^{n-m}(V_j \cap f^{-1}\{y\}) \leq \mathcal{H}^{n-m}(A \cap f^{-1}\{y\}) + \mathcal{H}^{n-m}((V_j \backslash A) \cap f^{-1}\{y\}).$$

Consequently,

$$\begin{aligned} &\limsup_{j \to \infty} \int_{\mathbb{R}^m} \left| \mathcal{H}^{n-m}(V_j \cap f^{-1}\{y\}) - \mathcal{H}^{n-m}(A \cap f^{-1}\{y\}) \right| dy \\ &\leq \limsup_{j \to \infty} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}((V_j \backslash A) \cap f^{-1}\{y\})\, dy \leq \limsup_{j \to \infty} \frac{\alpha_{n-m}\alpha_m}{\alpha_n} [f]_{C^{0,1}}^m \mathcal{L}^n(V_j \backslash A) = 0. \end{aligned}$$

Hence for $\mathcal{L}^n$-a.e. $y \in \mathbb{R}^n$, we have $\mathcal{H}^{n-m}((V_j \backslash A) \cap f^{-1}\{y\}) \downarrow 0$, and so $A \backslash f^{-1}\{y\}$ is $\mathcal{H}^{n-m}$-measurable. Also,

$$\mathcal{H}^{n-m}(V_j \cap f^{-1}\{y\}) \to \mathcal{H}^{n-m}(A \cap f^{-1}\{y\}) \quad \text{for } \mathcal{L}^n\text{-a.e. } y \in \mathbb{R}^n.$$

By Step IV, it follows that $y \mapsto \mathcal{H}^{n-m}(A \cap f^{-1}\{y\})$ is $\mathcal{L}^n$-measurable.

Step VI. Finally, if $A \subset \mathbb{R}^n$ is $\mathcal{L}^n$-measurable and $\mathcal{L}^n(A) = \infty$, we just write A as an increasing union of bounded $\mathcal{L}^n$-measurable sets and apply Steps I and V to estabilish (i) and (ii). ☐

Lemma 5.29. *Let $t > 1$, and assume $h : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous. Set*

$$P = \{x \in \mathbb{R}^n : Dh(x) \text{ exists, } Jh(x) > 0\}.$$

Then there is a countable collection $(F_k)_{k=1}^{\infty}$ of Borel subsets of $\mathbb{R}^n$ such that

(i) $\mathcal{L}^n(P \backslash \bigcup_{k=1}^{\infty} F_k) = 0$,

(ii) $h|_{F_k}$ *is one-to-one for each* $k = 1, 2, \cdots$*; and*

(iii) For each $k = 1, 2, \cdots$*, there exists a nonsingular symmetric linear map* $S_k : \mathbb{R}^n \to \mathbb{R}^n$ *such that*

$$[(h|_{F_k})^{-1} \circ S_k]_{C^{0,1}} \le t, \quad [S_k^{-1} \circ (h|_{F_k})]_{C^{0,1}} \le t, \quad \text{and} \quad t^{-n}|\det S_k| \le Jh|_{F_k} \le t^n |\det S_k|.$$

Proof. **Step I.** We first select Borel sets $(E_K)_{k=1}^{\infty}$ and symmetric automorphisms $(T_k)_{k=1}^{\infty}$ as in Lemma 5.25:

(a) $P \subset \bigcup_{k=1}^{\infty} E_k$,

(b) $h|_{E_k}$ is one-to-one for each $k = 1, 2, \cdots$; and

(c) For each $k = 1, 2, \cdots$,

$$[(h|_{E_k}) \circ T_k^{-1}]_{C^{0,1}} \le t, \quad [T_k \circ (h|_{E_k})^{-1}]_{C^{0,1}} \le t, \quad \text{and} \quad t^{-n}|\det T_k| \le Jh|_{E_k} \le t^n |\det T_k|.$$

According to the properties (b) and (c), the inverse map $(h|_{E_k})^{-1}$ is also Lipschitz continuous. By Theorem 5.19, there exists a Lipschitz continuous mapping $h_k : \mathbb{R}^n \to \mathbb{R}^n$ such that $h_k = (h|_{E_k})^{-1}$ on $h(E_k)$.

Since $h_k \circ h(x) = x$ for $x \in E_k$, by Theorem 3.7, we have $Dh_k(h(x))Dh(x) = \mathrm{Id}$ for $\mathcal{L}^n$-a.e. $x \in E_k$. Hence $Jh_k(h(x))Jh(x) = 1$ for $\mathcal{L}^n$-a.e. $x \in E_k$. By property (c), we know that $Jh_k(h(x)) > 0$ for $\mathcal{L}^n$-a.e. $x \in E_k$, and by Lipschitz continuity of h, it holds $Jh_k(x) > 0$ for $\mathcal{L}^n$-a.e. $x \in h(E_k)$.

Step II. For each $k \in \mathbb{N}$, we apply Lemma 5.25 on h_k and $h(E_k)$ to select Borel sets $(G_j^k)_{j=1}^{\infty}$ and symmetric non-singular linear maps $(R_j^k)_{j=1}^{\infty}$ such that

(d) $\mathcal{L}^n\left(h(E_k) \subset \bigcup_{j=1}^{\infty} G_j^k\right) = 0$,

(e) $h_k|_{G_j^k}$ is one-to-one for each $k = 1, 2, \cdots$; and

(f) For each $k = 1, 2, \cdots$,

$$[(h_k|_{G_j^k}) \circ (R_j^k)^{-1}]_{C^{0,1}} \le t, \quad [R_j^k \circ (h_k|_{G_j^k})^{-1}]_{C^{0,1}} \le t, \quad \text{and} \quad t^{-n}|\det R_j^k| \le Jh_k|_{G_j^k} \le t^n |\det R_j^k|.$$

We take

$$F_j^k = E_k \cap h^{-1}(G_j^k), \quad S_j^k = (R_j^k)^{-1}, \quad j, k = 1, 2, \cdots.$$

Note that

$$h_k\left(h(E_k) \backslash \bigcup_{j=1}^{\infty} G_j^k\right) = h^{-1}\left(h(E_k) \backslash \bigcup_{j=1}^{\infty} G_j^k\right) = E_k \backslash \bigcup_{j=1}^{\infty} F_j^k,$$

by Lipschitz continuity of h_k, we have $\mathcal{L}^n(E_k \backslash \bigcup_{j=1}^{\infty} F_j^k) = 0$. Recalling property (a), we establish (i). Also, property (b) implies that $h|_{F_j^k}$ is one-to-one, which establishes (ii). Finally, note that

$$[(h|_{F_j^k})^{-1} \circ S_j^k]_{C^{0,1}} = [(h|_{F_j^k})^{-1} \circ (R_j^k)^{-1}]_{C^{0,1}} \le [(h_k|_{G_j^k}) \circ (R_j^k)^{-1}]_{C^{0,1}} \le t,$$

and

$$[(S_j^k)^{-1} \circ (h|_{F_j^k})]_{C^{0,1}} = [R_j^k \circ (h|_{F_j^k})]_{C^{0,1}} \le [R_j^k \circ (h_k|_{G_j^k})^{-1}]_{C^{0,1}} \le t.$$

Since $Jh_k(h(x))Jh(x) = 1$ for $\mathcal{L}^n$-a.e. $x \in F_j^k$, the property (f) implies

$$t^{-n}|\det S_j^k| = t^{-n}|\det R_j^k|^{-1} \le Jh|_{F_j^k} \le t^n|\det R_j^k|^{-1} = t^n|\det S_j^k|.$$

Thus we establish (iii) and complete the proof. $\square$

Now we are prepared to proof the coarea formula. We define

$$\Lambda(n, n-m) = \{\lambda : \{1, 2, \cdots, n-m\} \to \{1, 2, \cdots, n\} \,|\, \lambda \text{ is strictly increasing}\}.$$

It is clear that there are $\binom{n}{m}$ elements in the set. For each $\lambda \in \Lambda(n, n-m)$, we define $P_\lambda : \mathbb{R}^n \to \mathbb{R}^{n-m}$ by

$$P_\lambda(x) = (x_{\lambda(1)}, x_{\lambda(2)}, \cdots, x_{\lambda(n-m)}).$$

Proof of Theorem 5.27. In the view of Lemma 5.29, we may assume that Df and Jf exist for all $x \in A$ and that $\mathcal{L}^n(A) < \infty$.

Case I: $A \subset \{Jf > 0\}$. For each $\lambda \in \Lambda(n, n-m)$, we write $f = p \circ h_\lambda$, where $h_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ is defined by $h(x) = (f(x), P_\lambda(x))$, and $p : \mathbb{R}^n \to \mathbb{R}^m$ is the standard projection $p(x_1, \cdots, x_m, x_{m+1}, x_n) = (x_1, \cdots, x_m)$. We set

$$A_\lambda = \{x \in A : \det Dh_\lambda(x) \neq 0\} = \{x \in A : P_\lambda|_{Df(x)^{-1}\{0\}} \text{ is injective}\}.$$

Then $A = \bigcup_{\lambda \in \Lambda(n,n-m)} A_\lambda$. Hence for simplicity we may assme $A = A_\lambda$ for some $\lambda \in \Lambda(n, n-m)$.

Step I.1. We fix $t > 1$ and apply Lemma 5.29 to $h = h_\lambda$ to obtain disjoint Borel sets $\{F_k\}_{k=1}^\infty$ and nonsingular symmetric linear maps $\{S_k\}_{k=1}^\infty$ satisfying properties (i)-(iii) in Lemma 5.29. We set $G_k = A \cap F_k$, and claim that

$$t^{-n}[\![p \circ S_k]\!] \leq Jf|_{G_k} \leq t^n [\![p \circ S_k]\!]. \tag{5.18}$$

Since $f = p \circ h$,

$$Df = p \circ Dh = p \circ S_k \circ S_k^{-1} \circ Dh = (p \circ S_k) \circ D(S_k^{-1} \circ h).$$

By Lemma 5.29 (iii),

$$[(h|_{G_k})^{-1} \circ S_k]_{C^{0,1}} \leq t, \quad [S_k^{-1} \circ (h|_{G_k})]_{C^{0,1}} \leq t. \tag{5.19}$$

Then for all $x, y \in G_k$,

$$t^{-1} \leq \frac{\left|S_k^{-1} \circ h(x) - S_k^{-1} \circ h(y)\right|}{|(h|_{G_k})^{-1} \circ S_k(S_k^{-1} \circ h(x)) - (h|_{G_k})^{-1} \circ S_k(S_k^{-1} \circ h(y))|} = \frac{\left|S_k^{-1} \circ h(x) - S_k^{-1} \circ h(y)\right|}{|x-y|} \leq t$$

Hence

$$t^{-1} \leq \lambda_{\min}(D(S_k^{-1} \circ h)) \leq \lambda_{\max}(D(S_k^{-1} \circ h)) \leq t \quad \text{on } G_k. \tag{5.20}$$

Now we fix $x \in G_k$, and take the polar decomposition $Df = S \circ U^*$ and $p \circ S_k = T \circ V^*$, where $S, T \in \mathbb{R}^{m \times m}$ are symmetric and $U, V \in \mathbb{R}^{n \times m}$ are orthogonal. Then $S = T \circ V^* \circ D(S_k^{-1} \circ h) \circ U$. Since $G_k \subset A \subset \{Jf > 0\}$, we have $\det S \neq 0$, and so $\det T \neq 0$. Then for all $z \in \mathbb{R}^m$, by (5.20),

$$|T^{-1} \circ Sz| = |V^* \circ D(S_k^{-1} \circ h) \circ Uz| \leq |D(S_k^{-1} \circ h) \circ Uz| \leq t|Uz| = t|z|.$$

Therefore $\|T^{-1}S\| \leq t$, and

$$Jf = |\det S| = |\det T| \left|\det(T^{-1}S)\right| \leq t^n \left|\det T\right|.$$

On the other hand, by (5.20),

$$|S^{-1} \circ Tz| = |U^* \circ D(S_k^{-1} \circ h)^{-1} \circ Vz| \leq |D(S_k^{-1} \circ h)^{-1} \circ Vz| \leq t|Vz| = t|z|.$$

Therefore $\|S^{-1}T\| \leq t$, and

$$Jf = \frac{1}{|\det S^{-1}|} = \frac{|\det T|}{|\det(S^{-1}T)|} \geq t^{-n} \left|\det T\right|.$$

Then we finish the proof of (5.3).

Step I.2. We repeatedly apply (5.18), (5.19) and Lemma 5.21 (ii) to obtain

$$\begin{aligned}
t^{-3n+m} & \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(G_k \cap f^{-1}\{y\})\, dy = t^{-3n+m} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}((h|_{G_k})^{-1}(h(G_k) \cap p^{-1}\{y\}))\, dy \\
&\le t^{-2n} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(S_k^{-1}(h(G_k) \cap p^{-1}\{y\}))\, dy = t^{-2n} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(S_k^{-1} \circ h(G_k) \cap (p \circ S_k)^{-1}\{y\})\, dy \\
&= t^{-2n} [\![p \circ S_k]\!]\, \mathcal{L}^n(S_k^{-1} \circ h(G_k)) \le t^{-n} [\![p \circ S_k]\!]\, \mathcal{L}^n(G_k) \\
&\le \int_{G_k} Jf\, d\mathcal{L}^n \\
&\le t^n [\![p \circ S_k]\!]\, \mathcal{L}^n(G_k) = t^n [\![p \circ S_k]\!]\, \mathcal{L}^n((h|_{G_k})^{-1} \circ S_k(S_k^{-1} \circ h(G_k))) \le t^{2n} [\![p \circ S_k]\!]\, \mathcal{L}^n(S_k^{-1} \circ h(G_k)) \\
&= t^{2n} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(S_k^{-1} \circ h(G_k) \cap (p \circ S_k)^{-1}\{y\})\, dy = t^{2n} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(S_k^{-1}(h(G_k) \cap p^{-1}\{y\}))\, dy \\
&\le t^{3n-m} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}((h|_{G_k})^{-1}(h(G_k) \cap p^{-1}\{y\}))\, dy = t^{3n-m} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(G_k \cap f^{-1}\{y\})\, dy
\end{aligned}$$

Since $\mathcal{L}^n(A \setminus \bigcup_{k=1}^\infty G) = 0$, we sum on k, use Lemma 5.28, and let $t \downarrow 1$ to conclude

$$\int_A Jf\, d\mathcal{L}^n = \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap f^{-1}\{y\})\, dy.$$

Case II: $A \subset \{Jf = 0\}$. We fix $0 < \epsilon < 1$ and define

$$g(x,y) = f(x) + \epsilon y, \quad p(x,y) = y, \quad x \in \mathbb{R}^n, y \in \mathbb{R}^m.$$

Then $Dg = (Df, \epsilon\, \mathrm{Id}_{m \times m})$, and similar to our proof of Theorem 5.23, $\epsilon^m \le Jg \le C_{f,m}\epsilon$ for some constant $C_{f,m}$ depending on f and m only. We observe that

$$\begin{aligned}
\int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap f^{-1}\{y\})\, dy &= \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap f^{-1}\{y - \epsilon w\})\, dy \quad \text{for all } w \in \mathbb{R}^m \\
&= \frac{1}{\alpha_m} \int_{B(0,1)} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap f^{-1}\{y - \epsilon w\})\, dy\, dw. \qquad (5.21)
\end{aligned}$$

We set $D = A \times B(0,1) \subset \mathbb{R}^{n+m}$. For any $y \in \mathbb{R}^m$ and $w \in B(0,1)$, we have $(x,z) \in D \cap g^{-1}\{y\} \cap p^{-1}\{w\}$ if and only if $x \in A, z \in B(0,1), f(x) + \epsilon = y$ and $z = w$, if and only if $x \in A$, $z = w \in B(0,1)$ and $f(x) = y - \epsilon w$. Hence

$$D \cap g^{-1}\{y\} \cap p^{-1}\{w\} = \begin{cases} (A \cap f^{-1}\{y - \epsilon w\}) \times \{w\}, & w \in B(0,1) \\ \emptyset, & w \notin B(0,1). \end{cases}$$

Then (5.21) becomes

$$\begin{aligned}
\int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap f^{-1}\{y\})\, dy &= \frac{1}{\alpha_m} \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(D \cap g^{-1}\{y\} \cap p^{-1}\{w\})\, dw\, dy \\
&\le \frac{\alpha_{n-m}}{\alpha_n} \int_{\mathbb{R}^m} \mathcal{H}^n(D \cap g^{-1}\{y\})\, dy = \frac{\alpha_{n-m}}{\alpha_n} \int_D Jg\, d\mathcal{L}^{n+m} \\
&\le \frac{\alpha_{n-m}}{\alpha_n} \mathcal{L}^{n+m}(D) C_{f,m}\epsilon = \frac{\alpha_{n-m}\alpha_m}{\alpha_n} \mathcal{L}^n(A) C_{f,m}\epsilon,
\end{aligned}$$

where we use (5.16) and our conclusion in Case I on g in the second line. Letting $\epsilon \downarrow 0$, we conclude that

$$\int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap f^{-1}\{y\})\, dy = 0 = \int_A Jf\, d\mathcal{L}^n.$$

Finally, in the general case, we just split A into $A \cap \{Jf > 0\}$ and $A \cap \{Jf = 0\}$, and apply Cases I and II. $\square$

Theorem 5.30 (Integration over level sets)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a Lipschitz continuous function, $n \geq m$. Then for each $\mathcal{L}^n$-integrable function $g : \mathbb{R}^n \to \mathbb{R}$,*

(i) $g|_{f^{-1}\{y\}}$ is $\mathcal{H}^{n-m}$-integrable for $\mathcal{L}^m$-a.e. $y \in \mathbb{R}^m$, and

(ii)

$$\int_{\mathbb{R}^n} g(x) Jf(x)\, dx = \int_{\mathbb{R}^m} \left[\int_{f^{-1}\{y\}} g \, d\mathcal{H}^{n-m} \right] dy \tag{5.22}$$

Proof. We first assume $g \geq 0$. We may also assume g is Borel measurable by modifying its value on a $\mathcal{L}^n$-null set. Similar to our proof of Theorem 5.26, we can write $g = \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{1}_{A_k}$ for appropriate Borel sets $(A_k)_{k=1}^{\infty}$. By continuity of f, the set $f^{-1}\{y\}$ is closed. Then $g|_{f^{-1}\{y\}}$ is $\mathcal{H}^{n-m}$-measurable. By monotone convergence,

$$\begin{aligned}
\int_{\mathbb{R}^n} g(x) Jf(x)\, dx &= \sum_{k=1}^{\infty} \frac{1}{k} \int_{A_k} Jf(x)\, dx = \sum_{k=1}^{\infty} \frac{1}{k} \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A_k \cap f^{-1}\{y\})\, dy \\
&= \sum_{k=1}^{\infty} \frac{1}{k} \int_{\mathbb{R}^m} \int_{f^{-1}\{y\}} \mathbb{1}_{A_k} d\mathcal{H}^{n-m}\, dy = \int_{\mathbb{R}^m} \int_{f^{-1}\{y\}} \left(\sum_{k=1}^{\infty} \frac{1}{k} \mathbb{1}_{A_k} \right) d\mathcal{H}^{n-m}\, dy \\
&= \int_{\mathbb{R}^m} \left[\int_{f^{-1}\{y\}} g\, d\mathcal{H}^{n-m} \right] dy.
\end{aligned}$$

For the general case that g is a $\mathcal{L}^n$-integrable function, write $g = g^+ - g^-$ and apply the above conclusion. By Lipschitz continuity of f on $\mathbb{R}^n$, Jf is bounded and the left side of (5.22) is finite. Then for $\mathcal{L}^m$-a.e. $y \in \mathbb{R}^m$, $\int_{f^{-1}\{y\}} g^{\pm}\, d\mathcal{H}^{n-m} < \infty$, and $g|_{f^{-1}\{y\}}$ is $\mathcal{H}^{n-m}$ integrable. $\square$

Corollary 5.31 (Polar coordinates)**.** *For each $\mathcal{L}^n$-integrable function $g : \mathbb{R}^n \to \mathbb{R}$,*

$$\int_{\mathbb{R}^n} g\, dx = \int_0^{\infty} \left(\int_{\partial B(0,r)} g\, d\mathcal{H}^{n-1} \right) dr.$$

In particular, for $\mathcal{L}^1$-a.e. $r > 0$,

$$\frac{d}{dr} \int_{B(0,r)} g\, dx = \int_{\partial B(0,r)} g\, d\mathcal{H}^{n-1}.$$

Proof. We let $f(x) = |x|$ in (5.22). Then $Df(x) = \frac{x}{|x|}$ and $Jf(x) = 0$ for every $x \neq 0$. $\square$