Chapter 7

Fubini's Theorem

The product measure of two outer measures which is again an outer measure is defined in Section 1. In Section 2 Fubini's theorem which relates the integral with respect to the product measure to the iterated integrals with respect to its factor measures. Section 3, 4 and 5 contain applications of Fubini's theorem to three different topics, namely, Rademarcher's theorem on the differentiability of Lipschitz continuous functions, layer cake representation and the convolution of functions.

7.1 The Product Measure

Let μ and ν be outer measures on the non-empty sets X and Y respectively. We define the **product measure** of μ and ν on the product set $X \times Y$ as, for $E \subset X \times Y$,

$$\begin{aligned} & (\mu \times \nu)(E) \\ = & \inf \left\{ \sum_{j=1}^{\infty} \mu(A_j)\nu(B_j) : E \subset \bigcup_{j=1}^{\infty} A_j \times B_j,\ A_j\ \mu\text{-measurable},\ B_j\ \nu\text{-measurable.} \right\} . \end{aligned}$$

It is understood that $\phi \times \phi = \phi$ so that $(\mu \times \nu)(\phi) = 0$. It is also straightforward to check

$$(\mu \times \nu) \left(\bigcup_{j=1}^{\infty} E_j \right) \le \sum_{j=1}^{\infty} (\mu \times \nu)(E_j), \quad \forall E_j \subset X \times Y, j \ge 1.$$

Hence $\mu \times \nu$ is an outer measure on $X \times Y$.

In the following we study how to evaluate $\mu \times \nu$ in terms of μ and ν. Introduce the following notations:

$$\mathcal{P}_0 = \Big\{A \times B : A \ \mu\text{-measurable and } B \ \nu\text{-measurable}\Big\},$$

$$\mathcal{P}_1 = \Big\{R : R = \bigcup_{j=1}^{n} A_j \times B_j,\ 1 \le n \le \infty,\ A_j \times B_j \in \mathcal{P}_0\Big\}, \text{ and}$$

$$\mathcal{P}_2 = \Big\{R : R = \bigcap_{j=1}^{n} R_j,\ 1 \le n \le \infty,\ R_j \in \mathcal{P}_1\Big\}.$$

Elements in $\mathcal{P}_0$ are called measurable rectangles. Clearly $\mathcal{P}_0 \subset \mathcal{P}_1 \subset \mathcal{P}_2$. We also set

$$\mathcal{F} = \Big\{R : \text{ For } \nu\text{-a.e. } y,\ x \mapsto \chi_R(x,y) \text{ is } \mu\text{-measurable and } y \mapsto \int \chi_R(x,y)\, d\mu(x) \text{ is } \nu\text{-measurable.}\Big\}$$

Note that the map

$$y \mapsto \int \chi_R(x,y)\, d\mu(x)$$

is defined almost everywhere in Y. Since ν is a complete measure and so every null set is measurable, it can be extended to be a measurable function in Y. Moreover, the integral

$$\int_Y \int_X \chi_R(x,y) d\mu(x) \nu(y)$$

is independent of the extension.

For $R \in \mathcal{F}$, we can define

$$\rho(R) = \int_Y \left(\int_X \chi_R(x,y)\, d\mu(x) \right) d\nu(y).$$

We will show in a series of lemmas that $\mathcal{P}_0, \mathcal{P}_1$, and $\mathcal{P}_2 \subset \mathcal{F}$ and they are $\mu \times \nu$-measurable. Moreover,

$$(\mu \times \nu)(R) = \rho(R),$$

for $R \in \mathcal{P}_1$ or $R \in \mathcal{P}_2$ provided in the latter R satisfies $\rho(R) < \infty$.

Lemma 7.1. $\mathcal{P}_0 \subset \mathcal{F}$ *and*

$$\rho(A \times B) = \mu(A)\nu(B), \quad A \times B \in \mathcal{P}_0.$$

Proof. The key observation is

$$\chi_{A \times B}(x,y) = \chi_A(x)\chi_B(y), \quad \forall (x,y) \in X \times Y.$$

For $y \in B$,

$$x \mapsto \chi_{A\times B}(x,y) = \chi_A(x) \text{ is } \mu\text{-measurable.}$$

For $y \notin B$,

$$x \mapsto \chi_{A\times B}(x,y) \equiv 0 \text{ is } \mu\text{-measurable.}$$

Next,

$$y \mapsto \int_X \chi_{A\times B}(x,y)\, d\mu(x) = \mu(A)\chi_B(y) \text{ is } \nu\text{-measurable.}$$

We have shown that $\mathcal{P}_0 \subset \mathcal{F}$. We also have

$$\begin{aligned}
\rho(A\times B) &= \int_Y \left(\int_X \chi_{A\times B}\, d\mu \right) d\nu \\
&= \int_Y \left(\int_X \chi_A(x)\chi_B(y)\, d\mu(x) \right) d\nu(y) \\
&= \mu(A)\nu(B).
\end{aligned}$$

□

Lemma 7.2. $\mathcal{P}_1 \subset \mathcal{F}$ *and*

$$\rho(R) = \sum_1^\infty \mu(A_j)\nu(B_j), \quad \textit{whenever } R = \overset{\circ}{\bigcup} A_j \times B_j, A_j \times B_j \in \mathcal{P}_0.$$

We have put a circle on top of the union sign to indicate that this is a union of pairwise disjoint sets.

The following fact will be used several times in the subsequent development: Each $R \in \mathcal{P}_1$ can be expressed as a countable disjoint union of measurable rectangles. Indeed, it follows from the observation

$$A_2 \times B_2 \setminus A_1 \times B_1 = A_2 \times (B_2 \setminus B_1) \overset{\circ}{\bigcup} (A_2 \setminus A_1) \times (B_2 \cap B_1).$$

Proof. Let $R \in \mathcal{P}_1$. Then

$$R = \overset{\circ}{\bigcup_{j=1}} A_j \times B_j, \quad A_j \times B_j \in \mathcal{P}_0.$$

Then

$$\chi_R = \sum_{j=1}^\infty \chi_{A_j \times B_j}.$$

Let

$$\varphi_n = \sum_{j=1}^n \chi_{A_j \times B_j}.$$

From Lemma 7.1, each $\varphi_n \in \mathcal{F}$. As $\chi_R(x,y) = \lim_{n\to\infty} \varphi_n(x,y)$. For ν a.e. y, $x \mapsto \chi_R(x,y)$ is μ-measurable. By monotone convergence theorem,

$$\int_X \chi_R(x,y)\, d\mu = \lim_{n\to\infty} \int_X \varphi_n(x,y)\, d\mu.$$

As $y \mapsto \displaystyle\int \varphi_n(x,y)\, d\mu(x)$ is ν-measurable, $y \mapsto \displaystyle\int \chi_R(x,y)\, d\mu(x)$ is also ν-measurable. We have shown that $R \in \mathcal{F}$.

Moreover,

$$\begin{aligned}
\rho(R) &= \int_Y \left(\int_X \chi_R(x,y)\, d\mu(x) \right) d\nu(y) \\
&= \int_Y \left(\int_X \lim_{n\to\infty} \varphi_n\, d\mu(x) \right) d\nu(y) \\
&= \int_Y \left(\lim_{n\to\infty} \int_X \varphi_n\, d\mu(x) \right) d\nu(y) \quad \text{(Monotone convergence theorem)} \\
&= \int_Y \left(\sum_{j=1}^{\infty} \mu(A_j)\chi_{B_j}(y) \right) d\nu(y) \\
&= \sum_{j=1}^{\infty} \int_Y \mu(A_j)\chi_{B_j}(y)\, d\nu(y) \quad \text{(Monotone convergence theorem)} \\
&= \sum_{j=1}^{\infty} \mu(A_j)\nu(B_j).
\end{aligned}$$

□

Lemma 7.3. *For $E \subset X \times Y$,*

$$(\mu \times \nu)(E) = \inf \left\{ \rho(R) : E \subset R,\ R \in \mathcal{P}_1 \right\}.$$

In particular, for $A \times B \in \mathcal{P}_0$,

$$(\mu \times \nu)(A \times B) = \mu(A)\nu(B) = \rho(A \times B).$$

Proof. Let $R \in \mathcal{P}_1$, $E \subset R$ and express $R = \mathring{\bigcup} A_j \times B_j$ where $A_j \times B_j \in \mathcal{P}_0$. Using the definition of $\mu \times \nu$,

$$\begin{aligned}
(\mu \times \nu)(E) &\le \sum_{j=1}^{\infty} \mu(A_j)\nu(B_j) \\
&= \rho(R). \quad \text{(Lemma 7.2)}
\end{aligned}$$

Taking infimum over all these R gives

$$(\mu \times \nu)(E) \le \inf \{\rho(R) : E \subset \mathbb{R},\ R \in \mathcal{P}_1\}.$$

On the other hand, for each n, there is some $R_n \in \mathcal{P}_1$, $E \subset R_n$, such that $R_n = \overset{\circ}{\bigcup} A_j^n \times B_j^n$, and

$$\begin{aligned}(\mu \times \nu)(E) + \frac{1}{n} &\ge \sum_{j=1}^{\infty} \mu(A_j^n)\nu(B_j^n) \\ &= \rho(R_n) \\ &\ge \inf \{\rho(R) : E \subset R,\ R \in \mathcal{P}_1\},\end{aligned}$$

and the inequality

$$(\mu \times \nu)(E) \ge \inf \{\rho(R) : E \subset \mathbb{R},\ R \in \mathcal{P}_1\}$$

follows after letting $n \to \infty$.

Now, for $A \times B \in \mathcal{P}_0$ and any $R \in \mathcal{P}_1$, $A \times B \subset R$,

$$\begin{aligned}(\mu \times \nu)(A \times B) &\le \mu(A)\nu(B) \\ &= \rho(A \times B) (\text{Lemma 7.1}) \\ &\le \rho(R), \quad (\text{since } \chi_{A\times B} \le \chi_R)\end{aligned}$$

so $\rho(A \times B)$ realizes the infimum of ρ over all $R \in \mathcal{P}_1$, $A \times B \subset R$. □

Lemma 7.4. *$\mathcal{P}_1$ and $\mathcal{P}_2$ consist of $\mu \times \nu$-measurable sets. For $R \in \mathcal{P}_1$,*

$$(\mu \times \nu)(R) = \sum_j \mu(A_j)\nu(B_j) = \rho(R).$$

Proof. We claim that $A \times B \in \mathcal{P}_0$ is $\mu \times \nu$-measurable. According to the definition, we need to prove

$$(\mu \times \nu)(T) \ge (\mu \times \nu)(T \cap A \times B) + (\mu \times \nu)(T \setminus A \times B), \quad \forall T \subset X \times Y.$$

To see this, for every $R \in \mathcal{P}_1$ with $T \subset R$, we have

$$\begin{aligned}&(\mu \times \nu)(T \cap A \times B) + (\mu \times \nu)(T \setminus A \times B) \\ \le\ &(\mu \times \nu)(R \cap A \times B) + (\mu \times \nu)(R \setminus A \times B) \\ \le\ &\rho(R \cap A \times B) + \rho(R \setminus A \times B) \quad (\text{By Lemma 7.3 and } R \cap A \times B,\ R \setminus A \times B \in \mathcal{P}_1) \\ =\ &\rho(R) \quad (\text{Since } \chi_{R\cap A\times B} + \chi_{R\setminus A\times B} = \chi_R)\end{aligned}$$

Taking infimum over all these R yields the desired result.

As all $\mu\times\nu$-measurable sets form a σ-algebra, $\mathcal{P}_0$, $\mathcal{P}_1$ and $\mathcal{P}_2$ consist of $\mu\times\nu$-measurable sets.

Write $R = \overset{\circ}{\bigcup} A_j \times B_j$, $A_j \times B_j \in \mathcal{P}_0$. We have

$$\begin{aligned}
&(\mu \times \nu)(R) \\
&= \sum_j (\mu \times \nu)(A_j \times B_j) \quad (R \text{ is } \mu \times \nu\text{-measurable}) \\
&= \sum_j \rho(A_j \times B_j) \quad (\text{Lemma 7.3}) \\
&= \rho(R). \quad (\text{Lemma 7.2})
\end{aligned}$$

□

Lemma 7.5. *Let $R \in \mathcal{P}_2$. Suppose that $R = \bigcap_{j=1}^{\infty} R_j$, $R_j \in \mathcal{P}_1$, and $\rho(R_1) < \infty$. Then $R \in \mathcal{F}$ and*

$$(\mu \times \nu)(R) = \rho(R).$$

Proof. For each $n \geq 1$, let $R^n = \bigcap_{j=1}^{n} R_j \in \mathcal{P}_1$ (check!), so $R^n \in \mathcal{F}$. That means, ν-a.e. y, $x \mapsto \chi_{R^n}(x, y)$ is μ-measurable. Using

$$\chi_R(x, y) = \lim_{n\to\infty} \chi_{R^n}(x, y), \quad \forall (x, y),$$

$x \mapsto \chi_R(x, y)$ is μ-measurable, for ν-a.e. y.

Next, $\chi_{R_1} - \chi_{R^n} \uparrow \chi_{R_1} - \chi_R$ as $n \to \infty$. By Lebsegue's monotone convergence theorem, for each fixed y,

$$\int (\chi_{R_1} - \chi_{R^n})\, d\mu \to \int (\chi_{R_1} - \chi_R)\, d\mu \text{ as } n \to \infty.$$

So,

$$y \mapsto \int_X (\chi_{R_1} - \chi_R)\, d\mu(x) \text{ is } \nu\text{-measurable.}$$

Using

$$\int \chi_R\, d\mu = \int \chi_{R_1} d\mu - \int (\chi_{R_1} - \chi_R)\, d\mu,$$

$$y \mapsto \int \chi_R(x, y)\, d\mu(x) \text{ is } \nu\text{-measurable.}$$

Note that we have used the fact that $\rho(R_1) < \infty$ and hence $\int \chi_{R_1} d\mu$ is finite and measurable for ν-a.e.y. We have shown that $\mathcal{P}_2 \subset \mathcal{F}$.

By repeatedly using Lebsegue's monotone convergence theorem,

$$\int \left(\int (\chi_{R_1} - \chi_{R^n})\, d\mu \right) d\nu \to \int \left(\int (\chi_{R_1} - \chi_R)\, d\mu \right) d\nu,$$

that is,

$$\rho(R_1) - \rho(R^n) \to \rho(R_1) - \rho(R),$$

or

$$\rho(R^n) \to \rho(R) \text{ as } n \to \infty.$$

On the other hand, by applying monotone convergence theorem to $\mu \times \nu$,

$$\int_{X\times Y} (\chi_{R_1} - \chi_{R^n})\, d(\mu \times \nu) \to \int_{X\times Y} (\chi_{R_1} - \chi_R)\, d(\mu \times \nu),$$

that is,

$$(\mu \times \nu)(R_1) - (\mu \times \nu)(R^n) \to (\mu \times \nu)(R_1) - (\mu \times \nu)(R),$$

or

$$(\mu \times \nu)(R^n) \to (\mu \times \nu)(R). \quad (\text{Use } (\mu \times \nu)(R_1) \le \rho(R_1) < \infty).$$

From $\rho(R^n) = (\mu \times \nu)(R^n)$, we get $\rho(R) = (\mu \times \nu)(R)$. □

Our last lemma is concerned with a regularity property of the product measure. It shows every set can be approximated from outside by a measurable set in any product measure.

Lemma 7.6. *For $E \subset X \times Y$, $\exists R \in \mathcal{P}_2$, $E \subset R$ such that*

$$(\mu \times \nu)(E) = (\mu \times \nu)(R).$$

Proof. If $(\mu \times \nu)(E) = \infty$, take $R = X \times Y$. If $(\mu \times \nu)(E) < \infty$, for each $n \ge 1$, there exists an $R^n \in \mathcal{P}_1$ such that

$$\begin{aligned}(\mu \times \nu)(E) + \frac{1}{n} &\ge \rho(R^n) \quad (\text{Lemma 7.3})\\ &\ge \rho(R)\end{aligned}$$

if we take $R = \bigcap_{n=1}^{\infty} R^n \in \mathcal{P}_2$. Letting $n \to \infty$, $\mu \times \nu(E) \ge \rho(R) = \mu \times \nu(R)$, done. □

We point out some properties of the product measure.

- The product space $(X \times Y) \times Z$ can be identified with $X \times (Y \times Z)$ and written as $X \times Y \times Z$. For measure μ, ν and λ on X, Y and Z respectively, the product measures $(\mu \times \nu) \times \lambda$ and $\mu \times (\nu \times \lambda)$ are well-defined on $X \times Y \times Z$. It is an exercise to show that these two measures coincides and thus we can write it as $\mu \times \nu \times \lambda$. This is the distributional law for the measure product. The same property extends to the product of finitely many product as well.

- The product measure of Borel measures is again Borel. The product measure of Radon measures is again Radon.

- We have $\mathcal{L}^{n-m} \times \mathcal{L}^m = \mathcal{L}^n$ for $0 < m < n$. In [EG] the n-dimensional Lebsegue measure is defined to be the nth times product $\mathcal{L}^1 \times \cdots \times \mathcal{L}^1$. It is again an exercise to show that this definition coincides with our definition in Chapter 3.

7.2 Fubini's Theorem

Theorem 7.7 (**Fubini's Theorem**). *Let μ and ν be σ-finite outer measures on X and Y respectively.*

(a) For any non-negative $\mu \times \nu$-measurable function f,

$$x \mapsto f(x,y) \text{ is } \mu\text{-measurable for } \nu\text{-a.e.} y, \quad \text{and}$$

$$y \mapsto \int_X f(x,y)\, d\mu(x) \text{ is } \nu\text{-measurable.}$$

Moreover,

$$\int_{X\times Y} f(x,y)\, d(\mu\times\nu)(x,y) = \int_Y \left(\int_X f(x,y)\, d\mu(x) \right) d\nu(y).$$

(b) (a) holds for $f \in L^1(\mu \times \nu)$.

Part (b) was first formulated by Tonelli and is also called *Tonelli's theorem.*

Before the proof of this theorem, it is worth to look at how close we are from this goal. Taking $f = \chi_R$ where R is measurable, the integral formula in Fubini's theorem becomes

$$(\mu \times \nu)(R) = \rho(R).$$

We have shown that this formula is valid for $R \in \mathcal{P}_2$ assuming $(\mu \times \nu)(X \times Y)$ is finite. We have inches to go, namely, to improve it to all measurable R. In this regard we need Lemma 7.6. In the following we take μ and ν to be finite.

Proof of Fubini's Theorem. (a) Let R be $\mu \times \nu$-measurable. We claim that $R \in \mathcal{F}$ first. Indeed, by Lemma 7.6, there exists an $R_1 \in \mathcal{P}_2$, $R \subset R_1$, such that $(\mu \times \nu)(R) = (\mu \times \nu)(R_1)$. As R is measurable,

$$(\mu \times \nu)(R_1 \setminus R) = (\mu \times \nu)(R_1) - (\mu \times \nu)(R) = 0.$$

Fix $R_2 \in \mathcal{P}_2$, $R_1 \setminus R \subset R_2$, such that $(\mu \times \nu)(R_2) = (\mu \times \nu)(R_1 \setminus R) = 0$. By Lemma 7.5,

$$\begin{aligned} 0 &= (\mu \times \nu)(R_2) \\ &= \rho(R_2) \\ &= \int \left(\int \chi_{R_2}\, d\mu \right) d\nu. \end{aligned}$$

Thus, for ν-a.e. y,

$$\int \chi_{R_2}\, d\mu = 0,$$

so

$$\int \chi_{R_1 \setminus R}\, d\mu = 0,$$

too. It means the set $\{x : (x, y) \in R_1 \setminus R\}$ is of μ-measure zero for ν-a.e. y. As every set of measure zero is measurable here, $\chi_{R_1 \setminus R}$ is μ-measurable for ν-a.e. y. So

$$\chi_R = \chi_{R_1} - \chi_{R_1 \setminus R}$$

is μ-measurable for all ν-a.e. y. Next,

$$\begin{aligned} \int \chi_R\, d\mu &= \int \chi_{R_1}\, d\mu - \int \chi_{R_1 \setminus R}\, d\mu \\ &= \int \chi_{R_1}\, d\mu. \end{aligned}$$

$R_1 \in \mathcal{P}_2 \subset \mathcal{F}$ means

$$y \mapsto \int \chi_{R_1}\, d\mu$$

is ν-measurable, so is

$$y \mapsto \int \chi_R\, d\mu.$$

We shown that $R \in \mathcal{F}$ for every $\mu \times \nu$-measurable R.

Next,

$$\begin{aligned} \int \chi_R\, d(\mu \times \nu) &= (\mu \times \nu)(R) \\ &= (\mu \times \nu)(R_1) \quad \text{(Lemma 7.6)} \\ &= \rho(R_1) \quad \text{(Lemma 7.5)} \\ &= \int \left(\int \chi_{R_1}\, d\mu \right) d\nu \\ &= \int \left(\int \chi_R\, d\mu \right) d\nu, \end{aligned}$$

or, if you like,

$$(\mu \times \nu)(R) = \rho(R)\ , \quad R \text{ measurable}\ .$$

Starting from this formula, we can pass f for simple functions and then non-negative measurable to obtain the integral formula in (a).

(b) When $f \in L^1(\mu \times \nu)$, apply (a) to f_+ and f_- separately to get the desired result. $\square$

The theorem is usually used in this way. Given a $\mu \times \nu$-measurable function f, we apply (a) to $|f|$ and see if

$$\int_{X\times Y} |f| \, d(\mu\times\nu) = \int_Y \Big(\int_X |f| \, d\mu \Big) \, d\nu.$$

If the right hand side is finite, then $|f| \in L^1(\mu\times\nu)$ and we can now use (b) to conclude that

$$\int_{X\times Y} f \, d(\mu\times\nu) = \int_Y \left(\int_X f \, d\mu \right) \, d\nu,$$

so the double integral can be evaluated by using an iterated integral.

We discuss an example to illustrate the role of σ-finiteness in Fubini's theorem.

Example 7.1. Let $\mathcal{L}^1$ and $\mathbf{c}$, the counting measure, be defined on $\mathbb{R}$ and consider the product measure $\mathcal{L}^1 \times \mathbf{c}$ on $\mathbb{R}^2 = \mathbb{R}^1 \times \mathbb{R}^1$. Consider $f = \chi_D$ where $D = \{(x,y) : x = y\}$. We claim that f is $\mathcal{L}^1 \times \mathbf{c}$-measurable, or, D is $\mathcal{L}^1 \times \mathbf{c}$-measurable. For $k \geq 1$, the sets $A_k = \bigcup_{j\in\mathbb{Z}} \big[j/k, (j+1)/k\big] \times \big[j/k, (j+1)/k\big]$, are countable unions of measurable rectangles so are $\mathcal{L}^1 \times \mathbf{c}$-measurable. As $D = \bigcap_{k=1}^{\infty} A_k$, D is also $\mathcal{L}^1 \times \mathbf{c}$-measurable.

Suppose Fubini's theorem holds for f. We have

$$\begin{aligned}\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y) \, d\mathcal{L}^1(x) \right) \, d\mathbf{c}(y) &= \int_{\mathbb{R}^2} f \, d(\mathcal{L}^1 \times \mathbf{c}) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y) \, d\mathbf{c}(y) \right) \, d\mathcal{L}^1(x).\end{aligned}$$

For a fixed y, $x \mapsto f(x,y) = \chi_{\{y\}}(x)$ is $\mathcal{L}^1$-measurable and

$$\int_{\mathbb{R}} f(x,y) \, d\mathcal{L}^1(x) = \int_{\mathbb{R}} \chi_{\{y\}}(x) \, d\mathcal{L}^1(x) = 0.$$

Thus

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y) \, d\mathcal{L}^1(x) \right) \, d\mathbf{c}(y) = 0.$$

On the other hand, for any fixed x,

$$y \mapsto f(x,y) = \chi_{\{x\}}(y)$$

is $\mathbf{c}$-measurable and

$$\begin{aligned}\int_{\mathbb{R}} f(x,y) \, d\mathbf{c}(y) &= \int_{\mathbb{R}} \chi_{\{x\}}(y) \, d\mathbf{c}(y) \\ &= \mathbf{c}\,\{x\} \quad \text{(definition of integral)} \\ &= 1,\end{aligned}$$

which implies

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)\, d\mathbf{c}(y) \right) d\mathcal{L}^1(x) = \infty,$$

contradiction holds!

In this example, all assumptions Theorem 7.7(a) are satisfied except $\mathbf{c}$ is not σ-finite.

7.3 Rademacher's Theorem

Recall that a function f defined on a set $E \subset \mathbb{R}^n$ is Lipschitz continuous in E if there exists some $M > 0$ such that

$$|f(x) - f(y)| \le M\,|x-y|\,, \quad \forall x, y \in E.$$

A function f defined in an open set $G \subset \mathbb{R}^n$ is called differentiable at $x \in G$ if there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}$ such that

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

It is well-known that when f is differentiable at x, the partial derivatives $\dfrac{\partial f}{\partial x_j}(x)$, $j = 1, \ldots, n$, must exist and $Lh = \displaystyle\sum_{j=1}^n h_j \frac{\partial f}{\partial x_j}(x)$.

Theorem 7.8 (**Rademacher's Theorem**). *Every locally Lipschitz continuous function in $\mathbb{R}^n$ must be differentiable almost everywhere.*

A function is locally Lipschitz continuous in some E if it is Lipschitz continuous in every compact subset of E. Since differentiability is a local property, in the following proof we may assume that f is Lipschitz continuous, that is,

$$|f(x) - f(y)| \le M\,|x-y|\,, \quad \forall x, y \in \mathbb{R}^n.$$

Proof. We claim: For each direction v, there exists a set $S_v \subset \mathbb{R}^n$, $\mathcal{L}^n(S_v) = 0$, such that

$$D_v f(x) = \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t}, \quad \forall x \in \mathbb{R}^n \setminus S_v.$$

Indeed, let

$$\overline{D}_v f(x) = \overline{\lim_{t\to 0}} \frac{f(x+tv) - f(x)}{t}, \qquad \text{and}$$

$$\underline{D}_v f(x) = \underline{\lim_{t\to 0}} \frac{f(x+tv) - f(x)}{t}.$$

Then $\underline{D}_v f$ and $\overline{D}_v f$ are measurable and bounded by M. Let $S_v = \{x \in \mathbb{R}^n : \underline{D}_v f < \overline{D}_v f\}$. For any line L parallel to v, we claim that $S_v \cap L$ has $\mathcal{L}^1$-measure zero for each such L. WLOG, let $v = e_1 = (1, 0, \dots, 0)$ and $\varphi(t) = f(x + te_1) = f(x_1 + t, x')$. For fixed (x_1, x'), $t \mapsto (x_1 + t, x')$ is the line L, φ is Lipschitz continuous and hence absolutely continuous on $\mathbb{R}$, hence is differentiable a.e. t, that is,

$$\overline{D}_v f(x + te_1) = \underline{D}_v f(x + te_1), \quad \text{a.e. } t.$$

We conclude that for a.e. x on L, $D_v f(x) = \overline{D}_v f(x) = \underline{D}_v f(x)$. Using Fubini's theorem

$$\mathcal{L}^n(S_v) = \int_{\mathbb{R}^{n-1}} \left(\int_{\mathbb{R}} \chi_{S_v}(x, x') \, d\mathcal{L}^1(x) \right) d\mathcal{L}^{n-1}(x') = 0.$$

Next, we introduce notations $D_j f = D_{e_j} f, j = 1, \dots, n$. Let

$$T_v = \left\{ x \in \mathbb{R}^n : D_v f(x),\ D_j f(x),\ j = 1, \dots, n, \text{ exist and } D_v f(x) = \sum_{j=1}^n v_j D_j f(x). \right\}$$

Claim: $\mathcal{L}^n(\mathbb{R}^n \setminus T_v) = 0$. Indeed, let $\varphi \in C_c^1(\mathbb{R}^n)$. We have

$$\int \frac{f(x + tv) - f(x)}{t} \varphi(x) \, d\mathcal{L}^n(x) = -\int f(x) \frac{\varphi(x - tv) - \varphi(x)}{t} \, d\mathcal{L}^n(x).$$

By Lebsegue's dominated convergence theorem,

$$\int D_v f(x) \varphi(x) \, d\mathcal{L}^n(x) = -\int f(x) D_v \varphi(x) \, d\mathcal{L}^n(x), \quad \forall v,\ |v| = 1.$$

Taking $v = e_j$,

$$\int D_j f(x) \varphi(x) \, d\mathcal{L}^n(x) = -\int f(x) \frac{\partial \varphi}{\partial x_j}(x) \, d\mathcal{L}^n(x), \quad j = 1, \dots, n.$$

We have, for given $v = (v_1, \dots, v_n)$,

$$\begin{aligned}
\int D_v f(x) \varphi(x) \, d\mathcal{L}^n(x) &= -\int f(x) D_v \varphi(x) \, d\mathcal{L}^n(x) \\
&= -\int f(x) \sum_j v_j \frac{\partial \varphi}{\partial x_j} \, d\mathcal{L}^n(x) \\
&= -\sum_j v_j \int f(x) \frac{\partial \varphi}{\partial x_j} \, d\mathcal{L}^n(x) \\
&= \sum_j v_j \int D_j f(x) \varphi(x) \, d\mathcal{L}^n(x) \\
&= \int \left(\sum_j v_j D_j f(x) \right) \varphi(x) \, d\mathcal{L}^n(x), \quad \forall \varphi \in C_c^1(\mathbb{R}^n),
\end{aligned}$$

which implies that $D_v f(x) = \sum_{j=1}^{n} v_j D_j f(x)$ almost everywhere in the set $\mathbb{R}^n \setminus (S_v \cup S_{e_1} \cup \cdots \cup S_{e_n})$. Hence $\mathcal{L}^n(\mathbb{R}^n \setminus T_v) = 0$ for every direction v.

For v, $|v| = 1$, $x \in \mathbb{R}^n$, $t \neq 0 \in \mathbb{R}$, set

$$Q(x, v, t) = \frac{f(x+tv) - f(x)}{t} - \sum_{j=1}^{n} v_j D_j f(x).$$

We are going to show that for every $\varepsilon > 0$, there corresponds some $\delta > 0$, such that

$$|Q(x, v, t)| < \varepsilon, \quad \text{a.e. } x,\ \forall v,\ |v| = 1,\ |t| < \delta. \tag{7.1}$$

We first fix a countable, dense set $\{v_k\}_1^\infty$ in S^{n-1}. Let

$$T = \bigcap_{k=1}^{\infty} T_{v_k}.$$

Then $\mathcal{L}^n(\mathbb{R}^n \setminus T) \leq \sum_{k=1}^{\infty} \mathcal{L}^n(\mathbb{R}^n \setminus T_{v_k}) = 0$. We claim (7.1) holds for all $x \in T$.

Given $\varepsilon > 0$, we can find finitely many points $v^1, \ldots, v^m$ on S^{n-1} such that for each v, there exists one of these points, say, v^k, satisfying $\left|v - v^k\right| < \varepsilon/2M(1+\sqrt{n})$. We have

$$\begin{aligned} &\left|Q(x,v,t) - Q(x,v^k,t)\right| \\ &= \left|\frac{f(x+tv) - f(x+tv^k)}{t} - \sum_{j=1}^{n}(v_j - v_j^k) D_j f(x)\right| \\ &\leq M\left|v - v^k\right| + \sqrt{n} M \left|v - v^k\right| \\ &= M(1+\sqrt{n})\left|v - v^k\right| \\ &< \frac{\varepsilon}{2}\,. \end{aligned} \tag{7.2}$$

For $x \in T$, since $Q(x, v^k, t) \to 0$ as $t \to 0$ for $k = 1, \ldots, m$, we can find a δ such that

$$\left|Q(x, v^k, t)\right| < \frac{\varepsilon}{2}, \quad \forall t,\ |t| < \delta. \tag{7.3}$$

Putting (7.2) and (7.3) together, for every $x \in T$ and $\varepsilon > 0$, there is a δ such that

$$\begin{aligned} |Q(x,v,t)| &\leq \left|Q(x,v,t) - Q(x,v^k,t)\right| + \left|Q(x,v^k,t)\right| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \quad \forall t,\ |t| < \delta,\ \text{and } \forall v, |v| = 1. \end{aligned}$$

Finally, for $x \in T$, we set $v = (y-x)/|y-x|$ for $y \neq x$, and $t = |y-x|$. By (8.1), for every ε, there is a δ such that

$$\begin{aligned} \varepsilon &> \Big|\frac{f(y)-f(x)}{|y-x|} - \sum_{j=1}^{n} \frac{(y_j-x_j)}{|y-x|} D_j f(x)\Big| \\ &= \frac{1}{|y-x|}\Big|f(u) - f(x) - \sum_{j=1}^{n}(y_j-x_j)D_j f(x)\Big|, \end{aligned}$$

for all $y, |y-x| < \delta$. The proof of Rademarcher's theorem is completed. □

7.4 The Layer Cake Representation

We begin with a lemma.

Lemma 7.9. *Let f be a non-negative, Lebsegue measurable function in $\mathbb{R}^n$. The set*

$$A = \{(x,t) \in \mathbb{R}^n \times [0,\infty) : f(x) \geq t \geq 0\},$$

is $\mathcal{L}^{n+1}$-measurable.

Proof. When $f = \chi_E$ for some measurable E,

$$A = E \times [0,1] \cup (R^n \setminus E) \times \{0\}$$

is measurable. Next, when $f = \sum_j \alpha_j \chi_{E_j}$ where E_j's are disjoint and measurable, and $\alpha_j > 0$,

$$A = \bigcup_j E_j \times [0,\alpha_j] \cup (\mathbb{R}^n \setminus \bigcup_j E_j) \times \{0\}$$

is measurable. In general, the lemma follows from using simple functions to approximate f. □

It is easy to deduce that the set

$$B = \{(x,t) \in \mathbb{R}^n \times [0,\infty) : f(x) > t > 0\}$$

is also measurable.

Let f be a non-negative measurable function in some measure space $(X, \mathcal{M}, \mu)$. We use $\{f > t\}$ to denote the set $\{x \in X : f(x) > t\}$. We are going to establish a general formula, which includes

$$\int_X f d\mu = \int_0^\infty \mu\{f > t\} d\mathcal{L}^1(t),$$

as a special case. Imagining the set $\{(x,t) : 0 \le t \le f(x)\}$ as a cake, the set $\{f > t\}$ is its layer at height t. This formula asserts that the integral of f can be computed by an integration over its layers. (Incidentally, since $\mu\{f > t\}$ is an increasing real-valued function provided it is finite and this is true, for instance, when f is integrable. As every increasing function is Riemann integrable, the integral on the right hand side of this formula is in fact an improper Riemann integral. Interestingly it shows that the abstract integral of f with respect to some μ can be defined by an Riemann integral in terms of its cake layers.)

To formulate a more general result, let φ be an increasing function in $[0,\infty]$ satisfying

(i) $\varphi(t) \to \varphi(\infty) \le \infty$ as $t \to \infty$,

(ii) $\varphi(0) = 0$,

(iii) φ is absolutely continuous on $[0,a]$, $\forall a \in (0,\infty)$.

Proposition 7.10. *Let μ be an outer measure on X and φ be given as above. For any non-negative μ-measurable function f in X,*

$$\int_X \varphi \circ f \, d\mu = \int_0^\infty \mu\{f > t\}\, \varphi'(t)\, d\mathcal{L}^1(t).$$

Note that $\varphi \circ f$ is μ-measurable and $\mu\{f > t\}\varphi'(t)$ is increasing and hence Lebsegue measurable. As a result, both integrals are well-defined.

Proof. Letting

$$A = \{(x,t) \in X \times [0,\infty) : f(x) > t\}$$

be measurable by Lemma 7.9, observe that

$$\mu\{f > t\} = \int_X \chi_A(x,t)\, d\mu(x) \ .$$

We have

$$\begin{aligned}
\int_X \varphi(f(x))\ d\mu &= \int_X \int_0^{f(x)} \varphi'(t)\, d\mathcal{L}^1(t) d\mu(x). \\
&= \int_X \int_0^\infty \chi_A(x,t) \varphi'(t)\, d\mathcal{L}^1(t) d\mu(x) \\
&= \int_0^\infty \int_X \chi_A(x,t)\, d\mu(x) \varphi'(t) d\mathcal{L}^1(t) \quad \text{(Fubini's theorem)} \\
&= \int_X \mu\{f > t\} \varphi'(t)\ d\mathcal{L}^1(t).
\end{aligned}$$

Note that in the first step, when $f(x) < \infty$,

$$\int_0^{f(x)} \varphi'(t)\, d\mathcal{L}^1(t) = \varphi(f(x)) - \varphi(0) = \varphi(f(x)) \ ,$$

by the fundamental theorem of calculus. If $f(x) = \infty$,

$$\begin{aligned}\int_0^{\infty} \varphi'(t)\, d\mathcal{L}^1(t) &= \lim_{a\to\infty} \int_0^{a} \varphi'(t)\, d\mathcal{L}^1(t) \quad \text{(monotone convergence theorem)} \\ &= \varphi(\infty) \quad (\text{ by (i) }) \\ &= \varphi(f(x)).\end{aligned}$$

Clearly, the proposition follows. □

Taking $\varphi(z) = z$, then $\varphi'(z) = 1$ and we cover the layer cake representation before.

We give an application of this formula to maximal functions.

Recall that for an $\mathcal{L}^n$-measurable f, its maximal function is defined to be

$$(Mf)(x) = \sup_{B\in\mathcal{B}_x} \frac{1}{\mathcal{L}^n(B)} \int_B |f| \ d\mathcal{L}^n,$$

where $\mathcal{B}_x$ is the collection of all closed balls containing x. (In fact, since the Lebesgue measure of ∂B is 0, you may take B to be an open ball.) We know that

- $Mf = M|f|$.
- Mf is $\mathcal{L}^n$-measurable.
- The weak L^1-estimate

  $$\mu\{Mf > t\} \le \frac{C}{t} \|f\|_{L^1} \ , \quad \forall t > 0$$

  where C is a dimensional constant, holds for $f \in L^1(\mathbb{R}^n)$.

Proposition 7.11. *Let $f \in L^p(\mathbb{R}^n)$, $p > 1$. We have*

$$\|Mf\|_{L^p} \le C \|f\|_{L^p} \ , \quad \forall f \in L^p(\mathbb{R}^n),$$

where C depends only on n and p.

Proof. We apply the formula in Proposition 7.10 by taking $\varphi(z) = z^p, p \in (1, \infty)$,

$$\int_{\mathbb{R}^n} f^p\, d\mathcal{L}^n = \int_0^{\infty} \mu\{f > t\}\, pt^{p-1}\, d\mathcal{L}^1(t)$$

where $f \geq 0$. We replace f by an L^1-function g as follows. First, for fixed $\alpha \in (0,1)$ and $t > 0$, set

$$g(x) = \begin{cases} f(x), & f(x) > \alpha t \\ 0, & f(x) \leq \alpha t, \end{cases}$$

and $h = f - g$. Then $g \in L^1(\mathbb{R}^n)$ and

$$\begin{aligned}
\int g \, d\mathcal{L}^n &\leq \int_{\{f > \alpha t\}} f \, d\mathcal{L}^n \\
&= \int_{\{f > \alpha t\}} \frac{f^p}{f^{p-1}} \, d\mathcal{L}^n \\
&\leq \frac{1}{(\alpha t)^{p-1}} \int f^p \, d\mathcal{L}^n \\
&< \infty.
\end{aligned}$$

Moreover, from $f = g + h$,

$$\begin{aligned}
Mf &\leq Mg + Mh \\
&\leq Mg + \alpha t
\end{aligned}$$

implies that if $Mf > t$ then $Mg > (1-\alpha)t$, so

$$\{Mf > t\} \subset \{Mg > (1-\alpha)t\}.$$

Using the weak L^1-estimate for $g \in L^1(\mathbb{R}^n)$

$$\begin{aligned}
\mathcal{L}^n \{Mf > t\} &\leq \mathcal{L}^n \{Mg > (1-\alpha)t\} \\
&\leq \frac{C}{(1-\alpha)t} \|g\|_{L^1} \\
&= \frac{C}{(1-\alpha)t} \int_{A_{\alpha t}} f \mathcal{L}^n,
\end{aligned}$$

where $A_t = \{f > \alpha t\}$. By Proposition 7.10,

$$\begin{aligned}
\int (Mf)^p \, d\mathcal{L}^n &= \int_0^\infty \mathcal{L}^n \{Mf > t\} \, pt^{p-1} \, d\mathcal{L}^1(t) \\
&\leq \int_0^\infty \{Mg > (1-\alpha)t\} \, pt^{p-1} \, d\mathcal{L}^1(t) \\
&\leq \int_0^\infty \frac{Cp}{(1-\alpha)t} \int_{A_{\alpha t}} f \, d\mathcal{L}^n \, t^{p-1} \, d\mathcal{L}^1(t).
\end{aligned}$$

Letting $A = \{(x,t) : f(x) > \alpha t\}$,

$$
\begin{aligned}
\int (Mf)^p \, d\mathcal{L}^n &\le \frac{Cp}{1-\alpha} \int_0^\infty \int_{A_{\alpha t}} f(x) \, d\mathcal{L}^n(x) \, t^{p-2} \, d\mathcal{L}^1(t) \\
&= \frac{Cp}{1-\alpha} \int_0^\infty \int \chi_A(x,t) f(x) \, d\mathcal{L}^n(x) \, t^{p-2} \, d\mathcal{L}^1(t) \\
&= \frac{Cp}{1-\alpha} \int \int_0^\infty \chi_A(x,t) t^{p-2} \, d\mathcal{L}^1(t) \, f(x) \, d\mathcal{L}^n(x) \\
&= \frac{Cp}{1-\alpha} \int \int_0^{\frac{f(x)}{\alpha}} t^{p-2} \, d\mathcal{L}^1(t) \, f(x) \, d\mathcal{L}^n \\
&= \frac{Cp}{(1-\alpha)(p-1)\alpha^{p-1}} \int f^p(x) \, d\mathcal{L}^n.
\end{aligned}
$$

We conclude that

$$
\|Mf\|_{L^p} \le \left[\frac{Cp}{(1-\alpha)(p-1)\alpha^{p-1}} \right]^{\frac{1}{p}} \|f\|_{L^p}, \quad \forall \alpha \in (0,1).
$$

□

In fact, by minimizing α, we get an explicit constant

$$
C = \min_{\alpha \in (0,1)} \left[\frac{Cp}{(1-\alpha)(p-1)\alpha^{p-1}} \right]^{\frac{1}{p}} = (Cepq)^{\frac{1}{p}},
$$

where q is conjugate to p, see [R1] for details.

7.5 Convolution

Convolution is a product between two functions. It appears in two contexts. First, it is well-known that the Fourier transform of the pointwise product of two functions equals to the pointwise product of the Fourier transform of these functions. It plays a fundamental role in harmonic analysis. Second, as it will be explained shortly, convolution can be used to construct various approximation of identity, so it is quite useful in approximation.

The definition of convolution involves an integration of the product of two integrable functions. It is not clear at all why this product makes sense. Fubini's theorem is in an essential use to justify the well-definiteness of the convolution.

Proposition 7.12. *Let $f, g \in L^1(\mathbb{R}^n)$. Then for a.e. x,*

$$
\int_{\mathbb{R}^n} |f(x-y)g(y)| \, d\mathcal{L}^n(y) < \infty.
$$

For those x, define the convolution of f and g by

$$(f * g)(x) = \int_{\mathbb{R}^n} f(x-y)g(y)\, d\mathcal{L}^n(y).$$

Then $f * g \in L^1(\mathbb{R}^n)$ and

$$\|f * g\|_{L^1} \le \|f\|_{L^1} \|g\|_{L^1}.$$

It will be understood that the integral is over $\mathbb{R}^n$ when the domain of integration is not specified in this section.

Proof. First, we claim that $(x, y) \mapsto f(x, y)g(y)$ is measurable in $\mathbb{R}^{2n}$. Recall that every measurable function is equal a.e. to a Borel function. We replace f and g by such Borel functions $\widetilde{f}$ and $\widetilde{g}$. Consider

$$(x, y) \mapsto \widetilde{f}(x-y)\widetilde{g}(y).$$

Let $\varphi(x, y) = x - y$. The map $(x, y) \mapsto \widetilde{f}(x-y) = \widetilde{f} \circ \varphi(x, y)$. As $\widetilde{f}$ is Borel and φ is continuous (hence measurable), $\widetilde{f}(x-y)$ is measurable. So is $\widetilde{f}(x-y)\widetilde{g}(y)$. As $f(x-y)g(y)$ differs from $\widetilde{f}(x-y)\widetilde{g}(y)$ on a set a of measure zero, it is also measurable.

Next, applying the first part of Fubini's theorem to the measurable function $|f(x-y)g(y)|$, we have

$$\begin{aligned}\int_{\mathbb{R}^{2n}} |f(x-y)g(y)|\, d\mathcal{L}^n(x) d\mathcal{L}^n(y) &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |f(x-y)|\, |g(y)|\, d\mathcal{L}^n(x) \right) d\mathcal{L}^n(y) \\ &= \|f\|_{L^1} \|g\|_{L^1} \\ &< \infty.\end{aligned}$$

We conclude that $f(x-y)g(y) \in L^1(\mathbb{R}^{2n})$. Moreover, exchanging the order of integration, we have

$$\int \left(\int |f(x-y)|\, |g(y)|\, d\mathcal{L}^n(y) \right) d\mathcal{L}^n(x) = \|f\|_{L^1} \|g\|_{L^1} < \infty,$$

$$\int |f(x-y)|\, |g(y)|\, d\mathcal{L}^n(y) < \infty$$

for a.e. x. Therefore, for a.e. x, the convolution $f * g$ is well-defined and finite. Moreover,

$$\begin{aligned}\int |f * g(x)|\, d\mathcal{L}^n(x) &= \int \left| \int f(x-y)g(y)\, d\mathcal{L}^n(y) \right| d\mathcal{L}^n(x) \\ &\le \int \left(\int |f(x-y)|\, |g(y)|\, d\mathcal{L}^n(y) \right) d\mathcal{L}^n(x) \\ &= \int \left(\int |f(x-y)|\, |g(y)|\, d\mathcal{L}^n(x) \right) d\mathcal{L}^n(y) \\ &= \|f\|_{L^1} \|g\|_{L^1}.\end{aligned}$$

□

By the same reason, $g * f$ exists and

$$\begin{aligned}(g * f)(x) &= \int g(x-y) f(y)\, d\mathcal{L}^n(y) \\ &= \int g(y) f(x-y)\, d\mathcal{L}^n(y) \quad \text{(change of variables)} \\ &= (f * g)(x),\end{aligned}$$

whenever $f * g$ or $g * f$ is well-defined at x.

For f, g and h in $\mathcal{L}^1(\mathbb{R}^n)$, one can verify that

$$(f * g) * h(x) = f * (g * h)(x), \quad \text{for a.e. } x.$$

There are several Young's inequality. The following one is concerned with convolution of functions.

Proposition 7.13 (**Young's Inequality**). *Let $f \in L^1(\mathbb{R}^n)$ and $g \in L^p(\mathbb{R}^n)$, $p \in (1, \infty)$. Then a.e. x,*

$$\int |f(x-y) g(y)|\, d\mathcal{L}^n(y) < \infty,$$

and $f * g \in L^p(\mathbb{R}^n)$ with

$$\|f * g\|_{L^p} \le \|f\|_{L^1} \|g\|_{L^p}.$$

Proof. Assume $g \in L^1(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$ first. For $\varphi \in C_c(\mathbb{R}^n)$,

$$\begin{aligned}\left| \int \varphi(x) \left(\int |f(x-y) g(y)|\, d\mathcal{L}^n(y) \right) d\mathcal{L}^n(x) \right| &= \left| \int \varphi(x) \left(\int |f(y) g(x-y)|\, d\mathcal{L}^n(y) \right) d\mathcal{L}^n(x) \right| \\ &= \left| \int \left(\int \varphi(x) |f(y)| |g(x-y)|\, d\mathcal{L}^n(x) \right) d\mathcal{L}^n(y) \right| \\ &\le \int |f(y)| \|\varphi\|_{L^q} \|g\|_{L^p}\, d\mathcal{L}^n(y) \quad \left(\frac{1}{p} + \frac{1}{q} = 1\right) \\ &= \|f\|_{L^1} \|g\|_{L^p} \|\varphi\|_{L^q}.\end{aligned}$$

Using the density of C_c-functions in $L^q(\mathbb{R}^n)$ and L^p-L^q duality,

$$\begin{aligned}\|f * g\|_{L^p} &= \sup \left\{ \left| \int \varphi(x) (f * g)(x)\, d\mathcal{L}^n(x) \right| : \|\varphi\|_{L^q} \le 1 \right\} \\ &\le \|f\|_{L^1} \|g\|_{L^p}.\end{aligned}$$

We have proved the proposition for $g \in L^1(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$. For $g \in L^p(\mathbb{R}^n)$, letting $g_k = \chi_{B_k(0)} |g|$, then $g_k \in L^1(\mathbb{R}^n)$, $g_k \uparrow |g|$. By Lebsegue's monotone convergence theorem,

$$\begin{aligned}\int \left(\int |f(x-y)| |g(y)|\, d\mathcal{L}^n(y) \right)^p d\mathcal{L}^n(x) &= \int \left(\lim_{k \to \infty} \int |f(x-y)| |g_k(y)|\, d\mathcal{L}^n(y) \right)^p d\mathcal{L}^n(x) \\ &\le \|f\|_{L^1}^p \|g\|_{L^p}^p\end{aligned}$$

and the conclusion follows. ∎

Convolution can be used to construct approximation kernel. As an illustration we prove

Theorem 7.14 (Weierstrass Approximation Theorem). *Let $f \in C_c(\mathbb{R}^n)$. For every $\varepsilon > 0$, there exists a polynomial p such that*

$$|f(x) - p(x)| < \varepsilon,$$

for all x in the support of f.

Proof. WLOG assume $f \in C_c(B_1(0))$. Let

$$Q_k(x) = c_k(1 - |x|^2)^k, \quad k \geq 1.$$

where the constant c_k is chosen so that

$$\int_{B_1(0)} Q_k(x)\, d\mathcal{L}^n(x) = 1,$$

and set $Q_k = 0$ outside $B_1(0)$. Define

$$p_k(x) = \int f(y) Q_k(x - y)\, d\mathcal{L}^n(y).$$

Note that

$$\begin{aligned} p_k(x) &= \int_{B_1(0)} f(y) Q_k(x - y)\, d\mathcal{L}^n(y) \\ &= c_k \int_{B_1(0)} f(y)(1 - |x - y)|^2)^k\, d\mathcal{L}^n(y) \end{aligned}$$

is a polynomial of degree $2k$. To show that p_k well-approximates f we need the following estimate on c_k,

$$c_k \leq C k^{\frac{n}{2}}, \quad \forall k \geq 1.$$

Indeed,

$$\begin{aligned} \frac{1}{c_k} &= \int_{B_1(0)} (1 - |x|^2)^k\, d\mathcal{L}^n(x) \\ &= \int_{S^{n-1}} \int_0^1 (1 - r^2)^k r^{n-1}\, dr d\theta \\ &= \mathcal{H}^{n-1}(S^{n-1}) \int_0^1 (1 - r^2)^k r^{n-1}\, dr. \end{aligned}$$

By the elementary inequality

$$(1 - r^2)^k \geq 1 - kr^2, \quad r \in (0, 1),$$

we have

$$\begin{aligned}
\frac{1}{c_k} &\geq \mathcal{H}^{n-1}(S^{n-1}) \int_0^{\frac{1}{\sqrt{k}}} (1-kr^2)r^{n-1}\,dr \\
&= \mathcal{H}^{n-1}(S^{n-1}) \frac{2}{n(n+2)} \left(\frac{1}{\sqrt{k}}\right)^n,
\end{aligned}$$

which implies the desired estimate. Using this, for $x \in B_1(0) \setminus B_\delta(0)$, $\delta \in (0,1)$,

$$\begin{aligned}
Q_k(x) &\leq c_k(1-\delta^2)^k \\
&\leq Ck^{\frac{n}{2}}(1-\delta^2)^k \to 0 \text{ as } k \to \infty.
\end{aligned}$$

Now,

$$\begin{aligned}
p_k(x) &= \int f(y)Q_k(x-y)\,d\mathcal{L}^n(y) \\
&= \int f(x-y)Q_k(y)\,d\mathcal{L}^n(y) \\
&= \int_{B_1(0)} f(x-y)Q_k(y)\,d\mathcal{L}^n(y).
\end{aligned}$$

We have

$$\begin{aligned}
|p_k(x) - f(x)| &= \left| \int_{B_1(0)} (f(x-y) - f(x))Q_k(y)\,d\mathcal{L}^n(y) \right| \\
&\leq \int_{B_\delta(0)} |f(x-y) - f(x)|\, Q_k(y)\,d\mathcal{L}^n(y) \\
&\quad + \int_{B_1(0)\setminus B_\delta(0)} |f(x-y) - f(x)|\, Q_k(y)\,d\mathcal{L}^n(y).
\end{aligned}$$

As f is uniformly continuous in $\mathbb{R}^n$, for $\varepsilon > 0$, we can find some δ such that

$$|f(x+h) - f(x)| < \varepsilon, \quad \forall x \in \mathbb{R}^n, \text{and } h,\ |h| < \delta.$$

Therefore,

$$\int_{B_\delta(0)} |f(x-y) - f(x)|\, Q_k(y)\,d\mathcal{L}^n(y) \leq \varepsilon.$$

On the other hand,

$$\begin{aligned}
\int_{B_1(0)\setminus B_\delta(0)} |f(x-y) - f(x)|\, Q_k(y)\,d\mathcal{L}^n(y) &\leq 2\sup|f|\,|B_1(0)\setminus B_\delta(0)|\, Ck^{\frac{n}{2}}(1-\delta^2)^k \\
&< \varepsilon,
\end{aligned}$$

for all sufficiently large k, $k \geq k_0$, say. Putting things together,

$$|p_k(x) - f(x)| \leq \varepsilon + \varepsilon = 2\varepsilon,$$

done. □

Corollary 7.15. *Let f be a continuous function on a compact set K in $\mathbb{R}^n$. For $\varepsilon > 0$, there exists a polynomial p such that $f(x) - p(x) < \varepsilon$, $\forall x \in K$.*

Proof. Extend f to be a function in $C_c(B_R(0))$ where R is large such that $K \subset\subset B_R(0)$. Surely there are many ways to do this. By Theorem 7.14, we can find a polynomial p such that $|f(x) - p(x)| < \varepsilon$, $\forall x \in \overline{B_R(0)}$, and this p does the job. □

Comments on Chapter 7. Our discussion on the product measure and Fubini's theorem is taken from [EG]. A treatment on the construction of the product measure from two measure spaces rather than two outer measures can be found in [R1]. Rademarcher's theorem (1919) whose proof depends crucially on Fubini's theorem is the foundation for geometric measure theory. The layer cake representation and convolution of functions are taken from [R1], except the treatment on Weierstrass approximation theorem is modified from "Principles of Mathematical Analysis" of the same author. By the way, Weierstrass approximation theorem should have been covered in some undergraduate analysis course. Indeed, it is in my lecture notes for MATH3060. However, in the past it was never taught due to some mysterious reason. In view of this, I include it here as an application of convolution. Another standard application of Fubini's theorem is the inversion formula for Fourier transform.