REVIEW OF LEBESGUE MEASURE AND INTEGRATION

CHRISTOPHER HEIL

These notes will briefly review some basic concepts related to the theory of Lebesgue measure and the Lebesgue integral. We are not trying to give a complete development, but rather review the basic definitions and theorems with at most a sketch of the proof of some theorems. These notes follow the text *Measure and Integral* by R. L. Wheeden and A. Zygmund, Dekker, 1977, and full details and proofs can be found there.

1. OPEN, CLOSED, AND COMPACT SUBSETS OF EUCLIDEAN SPACE

Notation 1.1. $\mathbb{N} = \{1, 2, 3, \dots\}$ is the set of natural numbers, $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ is the set of integers, $\mathbb{Q}$ is the set of rational numbers, $\mathbb{R}$ is the set of real numbers, and $\mathbb{C}$ is the set of complex numbers. $\mathbb{R}^d$ is real d-dimensional Euclidean space, the space of all vectors $x = (x_1, \dots, x_d)$ with $x_1, \dots, x_d \in \mathbb{R}$.

On occasion, we formally use the *extended real number line* $\mathbb{R} \cup \{-\infty, \infty\} = [-\infty, \infty]$, but it is important to note that ∞ is a formal object, not a number. To write $a \in [-\infty, \infty]$ means that either a is a finite real number or a is one of $\pm\infty$. We write $|a| < \infty$ to mean that a is a finite real number. Note that there is no analogue of the extended reals when we consider complex numbers; there's no obvious "∞" or "$-\infty$."

We declare some arithmetic conventions for the extended real numbers: $\infty + \infty = \infty$, $1/0 = \infty$, $1/\infty = 0$, and $0 \cdot \infty = 0$. The symbols $\infty - \infty$ are undefined, i.e., they have no meaning.

The empty set is denoted by $\emptyset$. Two sets A, B are *disjoint* if $A \cap B = \emptyset$. A collection $\{A_k\}$ of sets are disjoint if $A_j \cap A_k = \emptyset$ whenever $j \neq k$.

The *real part* of a complex number $z = a + ib$ is $\mathrm{Re}\,(z) = a$, and the *imaginary part* is $\mathrm{Im}\,(z) = b$. The *complex conjugate* of $z = a + ib$ is $\bar{z} = a - ib$. The *modulus*, or *absolute value*, of $z = a + ib$ is $|z| = \sqrt{z\bar{z}} = \sqrt{a^2 + b^2}$. $\square$

For concreteness, we will use the Euclidean distance on $\mathbb{R}^d$ in these notes. However, all the results of this section are valid with respect to any norm on $\mathbb{R}^d$.

Definition 1.2.

(a) The *Euclidean norm* on $\mathbb{R}^d$ is

$$|x| \;=\; (x_1^2 + \cdots + x_d^2)^{1/2}.$$

The *distance* between x, $y \in \mathbb{R}^d$ is $|x - y|$.

Date: Revised July 30, 2007.

(b) Suppose that $\{x_n\}_{n\in\mathbb{N}}$ is a sequence of points in $\mathbb{R}^d$ and that $x \in \mathbb{R}^d$. We say that x_n *converges* to x, and write $x_n \to x$ or $x = \lim_{n\to\infty} x_n$, if

$$\lim_{n\to\infty} |x_n - x| = 0. \qquad \square$$

Using this definition of distance, we can now define open and closed sets in $\mathbb{R}^d$ and state some of their basic properties.

Definition 1.3. Let $E \subseteq \mathbb{R}^d$ be given.

(a) E is *open* if for each point $x \in E$ there is some open ball

$$B_r(x) \;=\; \{y \in \mathbb{R}^d : |x-y| < r\}$$

centered at x that is completely contained in E, i.e., $B_r(x) \subseteq E$ for some $r > 0$.

(b) A point $x \in \mathbb{R}^d$ is a *limit point* of E if there exist points $x_n \in E$ that converge to x, i.e., such that $x_n \to x$.

(c) The *complement* of E is

$$E^{\mathrm{C}} \;=\; \mathbb{R}^d \backslash E \;=\; \{x \in \mathbb{R}^d : x \notin E\}.$$

(d) E is *closed* if its complement E^{C} is open. It can be shown that E is closed if and only if E contains all its limit points.

(e) If E is any subset of $\mathbb{R}$, then its *closure* $\bar{E}$ is the smallest closed set that contains E. It can be shown that

$$\bar{E} \;=\; E \cup \{x \in \mathbb{R}^d : x \text{ is a limit point of } E\}.$$

(f) E is *dense* in $\mathbb{R}^d$ if $\bar{E} = \mathbb{R}^d$. For example, the set $\mathbb{Q}$ of all rational numbers is a dense subset of $\mathbb{R}$.

(g) E is *bounded* if it is contained in some ball with finite radius, i.e., if there is some $r > 0$ such that $E \subseteq B_r(0)$. $\square$

The following notion of compact sets is very important.

Definition 1.4. Let $E \subseteq \mathbb{R}^d$ be given.

(a) An *open cover* of E is any collection $\{U_\alpha\}_{\alpha\in J}$ of open sets such that $E \subseteq \bigcup_\alpha U_\alpha$. The index set J may be finite, countable, or uncountable, i.e., there may be finitely many, countably many, or uncountably many open sets U_α in the collection.

(b) E is *compact* if *every* open cover $\{U_\alpha\}_{\alpha\in J}$ of E contains a finite subcover. That is, E is compact if whenever we choose open sets U_α such that $E \subseteq \bigcup_\alpha U_\alpha$, then there exist *finitely many* indices $\alpha_1, \dots, \alpha_k \in J$ such that $E \subseteq U_{\alpha_1} \cup \cdots \cup U_{\alpha_k}$. $\square$

Theorem 1.5. Let $E \subseteq \mathbb{R}^d$ be given.

(a) (Heine–Borel Theorem) E is compact if and only if it is both closed and bounded.

(b) (Bolzano–Weierstrass Theorem) If E is compact, then every countable sequence of points $\{x_n\}_{n\in\mathbb{N}}$ with $x_n \in E$ has a convergent subsequence (even if the original sequence does not converge). That is, there exist indices $n_1 < n_2 < \dots$ and a point $x \in \mathbb{R}^d$ so that $x_{n_k} \to x$. (Note that x is then a limit point of E, and therefore $x \in E$ since E is closed.) □

Theorem 1.6. Let $E, F \subseteq \mathbb{R}^d$ be given.

(a) If $E \subseteq F$, then $\bar{E} \subseteq \bar{F}$.

(b) If E and F are compact sets then $E + F = \{x + y : x \in E \text{ and } y \in F\}$ is compact.

(c) If E and F are bounded sets (not necessarily compact), then $\overline{E+F} \subseteq \bar{E} + \bar{F}$.

Proof. (a) We just have to show that every limit point of E is a limit point of F. So, suppose that x is a limit point of E. Then there exist points $x_n \in E$ such that $x_n \to x$. However, $E \subseteq F$, so each x_n is also an element of F. Therefore x is a limit point of F by definition.

(b) Suppose E and F are both compact. Then E and F are both bounded, so they are contained in some finite balls centered at the origin, say $E \subseteq B_r(0)$ and $F \subseteq B_s(0)$. Hence $E + F \subseteq B_{r+s}(0)$, so $E + F$ is bounded.

To show that $E + F$ is closed, suppose that z is any limit point of $E + F$. This means that there are points $z_n \in E + F$ which converge to z. By definition, $z_n = x_n + y_n$ for some $x_n \in E$ and $y_n \in F$. WE DO NOT KNOW whether x_n and y_n will converge to some points x and y! However, we do know that $\{x_n\}_{n\in\mathbb{N}}$ is a sequence of points in E and that E is compact. Therefore, there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ which does converge to some $x \in E$. For simplicity of notation, write $x'_k = x_{n_k}$ and $y'_k = y_{n_k}$. Now, $\{y'_k\}_{k\in\mathbb{N}}$ is a sequence of points in F and F is compact, so there must be a subsequence $\{y'_{k_j}\}_{j\in\mathbb{N}}$ which converges to some $y \in F$. Note that since $x'_k \to x$, it is still true that $x'_{k_j} \to x$. Again for simplicity write $x''_j = x'_{k_j}$ and $y''_j = y'_{k_j}$. Then we have $x''_j \to x$ and $y''_j \to y$. Therefore $x''_j + y''_j \to x + y$. However, remember where these points came from: $x''_j = x_{n_{k_j}}$ and $y''_j = y_{n_{k_j}}$. Therefore $x''_j + y''_j \to z$ since $x_n + y_n \to z$. So it must be the case that $z = x + y$. Thus $z \in E + F$, so $E + F$ contains all its limit points, and therefore is closed. Since $E + F$ is both closed and bounded, it is compact.

(c) Suppose that E and F are bounded sets. Then $\bar{E}$ and $\bar{F}$ are closed and bounded sets, hence compact. Certainly $E + F \subseteq \bar{E} + \bar{F}$, so we just have to show that every limit point of $E + F$ is in $\bar{E} + \bar{F}$.

So, suppose that z is a limit point of $E + F$. Then there exist points $z_n \in E + F$ such that $z_n \to z$. By definition, $z_n = x_n + y_n$ for some $x_n \in E$ and $y_n \in F$. Since $\bar{E}$ and $\bar{F}$ are compact, we can imitate the argument of part b and find convergent subsequences x''_j and y''_j, i.e., $x''_j \to x \in \bar{E}$ and $y''_j \to y \in \bar{F}$ for some $x \in \bar{E}$ and $y \in \bar{F}$ (not necessarily $x \in E$ or $y \in F$). Therefore $x + y = \lim x''_j + \lim y''_j = \lim\,(x''_j + y''_j) = z$, so $z \in \bar{E} + \bar{F}$, as desired. □

2. MEASURE THEORY

Our goal in this section is to assign to each subset of $\mathbb{R}^d$ a "size" or "measure" that generalizes the concept of area or volume from simple sets to arbitrary sets. However, we will see that this cannot be done for all sets without introducing some strange pathologies, and therefore we must restrict the definition of measure to a subclass of well-behaved "measurable sets."

Definition 2.1.

(a) A *cube* or *rectangular box* in $\mathbb{R}^d$ is a set of the form

$$Q \;=\; [a_1,b_1]\times\cdots\times[a_d,b_d].$$

The *volume* of this cube is

$$\mathrm{vol}(Q) \;=\; (b_1-a_1)\cdots(b_d-a_d).$$

(b) The *exterior Lebesgue measure* of an arbitrary set $E \subseteq \mathbb{R}^d$ is

$$|E|_e \;=\; \inf\biggl\{\sum_k \mathrm{vol}(Q_k) : \text{all countable sequences of cubes } Q_k \text{ with } E \subseteq \bigcup_k Q_k\biggr\}.$$

The exterior measure of a set lies in the range $0 \le |E|_e \le \infty$. Allowing the possibility of infinite exterior measure, every subset of $\mathbb{R}^d$ has a uniquely defined nonnegative exterior measure. ☐

Example 2.2.

(a) A seemingly "obvious" fact is that if Q is a cube in $\mathbb{R}^d$ then $|Q|_e = \mathrm{vol}(Q)$. Since Q covers itself with one cube, it does follow immediately from the definition that $|Q|_e \le \mathrm{vol}(Q)$. However, the other inequality is not so trivial to prove. More generally, if $Q_1,\dots,Q_n$ are disjoint cubes, then it can be shown that

$$|Q_1\cup\cdots\cup Q_n|_e \;=\; \mathrm{vol}(Q_1)+\cdots+\mathrm{vol}(Q_n).$$

(b) $|\mathbb{R}^d|_e = \infty$.

(c) If $S \subseteq \mathbb{R}^d$ contains only countably many points then $|S|_e = 0$. For example, the set of rational numbers in $\mathbb{R}$ has zero exterior measure, i.e., $|\mathbb{Q}|_e = 0$.

(d) The Cantor set C is an example of a subset of $\mathbb{R}$ which contains uncountably many points yet has exterior measure $|C|_e = 0$. The Cantor set is also closed and equals its own boundary. ☐

Definition 2.3. A property which holds except on a set of exterior measure zero is said to hold *almost everywhere* (abbreviated a.e.). For example, if

$$g(x) \;=\; \begin{cases} 1, & x \text{ rational},\\ 0, & x \text{ irrational},\end{cases} \tag{2.1}$$

then we say that $g = 0$ a.e. since the set of points $\{x \in \mathbb{R} : g(x) \neq 0\}$ is countable and therefore has measure zero.

The prefix *essential* is often applied to a property that holds a.e. For example, the *essential supremum* of a function $f\colon E \to [-\infty, \infty]$ is

$$\operatorname*{ess\,sup}_{x\in E} f(x) \;=\; \inf\{M : f(x) \le M \text{ a.e.}\}.$$

Thus, for the function g given in equation (2.1) we have

$$\sup_{x\in\mathbb{R}} g(x) \;=\; 1 \qquad \text{while} \qquad \operatorname*{ess\,sup}_{x\in\mathbb{R}} g(x) \;=\; 0. \quad \square$$

Here are some basic properties of exterior measure.

Lemma 2.4. Let $E,\, F \subseteq \mathbb{R}^d$ be given.

(a) If $E \subseteq F$, then $|E|_e \le |F|_e$.

(b) If $E_1, E_2, \ldots \subseteq \mathbb{R}^d$, then $\left|\bigcup E_k\right|_e \le \sum_k |E_k|_e$.

(c) If $E \subseteq \mathbb{R}^d$ and $\varepsilon > 0$, then there exists an open set $U \supseteq E$ such that $|U|_e \le |E|_e + \varepsilon$. (Note that we also have $|E|_e \le |U|_e$ by part a.) $\square$

Remark 2.5. We might expect in Lemma 2.4(b) that if the sets $E_1, E_2, \ldots$ are disjoint, then we would actually have $|\cup E_k|_e = \sum_k |E_k|_e$. Yet it can be shown that this is FALSE in general: there exist *disjoint* sets $E_1, E_2, \ldots \subseteq \mathbb{R}^d$ such that $|\cup E_k|_e < \sum_k |E_k|_e$.

Likewise, in Lemma 2.4(c) we might expect that since $E \subseteq U$ and $|E|_e \le |U|_e \le |E|_3 + \varepsilon$, the set $U\backslash E$ should have small exterior measure. Specifically, we expect that $|U\backslash E|_e \le \varepsilon$. Yet this is also FALSE in general! Consequently, for such sets we have $|(U\backslash E) \cup E|_e = |U|_e \le |E|_e + \varepsilon < |E|_e + |U\backslash E|_e$ even though U is the union of the two disjoint sets $U\backslash E$ and E. $\square$

The problem is that, in some sense, the definition of exterior measure is too inclusive. All sets have an exterior measure, even though there exist some very strange sets that behave in unexpected ways (the existence of such strange sets is a consequence of the Axiom of Choice). One way to handle this problem is to restrict our attention to sets which are "well-behaved" with respect to exterior measure. This leads us to make the following definition.

Definition 2.6. A set $E \subseteq \mathbb{R}^d$ is *Lebesgue measurable*, or simply *measurable*, if given any $\varepsilon > 0$ there exists an open set $U \supseteq E$ such that $|U\backslash E|_e \le \varepsilon$.

If E is measurable, then its *Lebesgue measure* is its exterior measure, and is denoted $|E| = |E|_e$. $\square$

There exists sets that are not measurable. However, as the proof of this fact relies on the Axiom of Choice, it is nonconstructive, i.e., it simply says that such sets exist but does not explicitly display one. Typically, the sets we encounter are all measurable, and almost all operations that we perform on measurable sets leave their measurability intact.

Lemma 2.7. (a) All open subsets of $\mathbb{R}^d$ are measurable.

(b) All closed subsets of $\mathbb{R}^d$ are measurable.

(c) Countable unions of measurable sets are measurable. That is, if $E_1, E_2, \dots$ are measurable, then so is $\bigcup_k E_k$.

(d) Countable intersections of measurable sets are measurable. That is, if $E_1, E_2, \dots$ are measurable, then so is $\bigcap_k E_k$.

(e) The complement of a measurable set is measurable. That is, if E is measurable, then so is E^{C}.

(f) All sets with exterior measure zero are measurable. That is, if $|E|_e = 0$, then E is measurable.

Proof. (f) Suppose that $|E|_e = 0$, and let $\varepsilon > 0$ be given. Then by Lemma 2.4, there exists an open set $U \supseteq E$ such that $|U|_e \le |E|_e + \varepsilon = 0 + \varepsilon = \varepsilon$. Therefore, since $U \backslash E \subseteq U$, we have by Lemma 2.4(a) that $|U\backslash E|_e \le |U|_e \le \varepsilon$. Hence E is measurable by definition. □

Theorem 2.8. Let E and $E_1, E_2, \dots$ be measurable subsets of $\mathbb{R}^d$.

(a) $\left|\bigcup E_k\right| \le \sum_k |E_k|$.

(b) If $E_1, E_2, \dots$ are disjoint, then $\left|\bigcup E_k\right| = \sum_k |E_k|$.

(c) If $E_1 \subseteq E_2$ and $|E_2| < \infty$, then $|E_1 \backslash E_2| = |E_1| - |E_2|$.

(d) If $E_1 \subseteq E_2 \subseteq \cdots$, then $\left|\bigcup E_k\right| = \lim\limits_{k\to\infty} |E_k|$.

(e) If $E_1 \supseteq E_2 \supseteq \cdots$ and $|E_1| < \infty$, then $\left|\bigcap E_k\right| = \lim\limits_{k\to\infty} |E_k|$.

(f) If $h \in \mathbb{R}^d$ and we define $E + h = \{x + h : x \in E\}$, then $|E + h| = |E|$.

(g) If $T \colon \mathbb{R}^d \to \mathbb{R}^d$ is linear, then $|T(E)| = |\det(T)|\,|E|$. □

Here are some final attempts to illustrate the way in which measurable sets are "well-behaved."

Definition 2.9. Let $E \subseteq \mathbb{R}^d$ be arbitrary. The *inner measure* of $E \subseteq \mathbb{R}^d$ is

$$|E|_i \;=\; \sup\{|F| : \text{all closed sets } F \text{ such that } F \subseteq E\}.$$

Compare this definition of inner measure to Lemma 2.4(c), which implies that the exterior measure of E is given by

$$|E|_e \;=\; \inf\{|U| : \text{all open sets } U \text{ such that } U \supseteq E\}.$$

Theorem 2.10. If $|E|_e < \infty$, then E is measurable if and only if $|E|_e = |E|_i$. $\square$

Theorem 2.11 (Carathéodory's Criterion)**.** Let $E \subseteq \mathbb{R}^d$ be given. Then E is measurable if and only if for every set $A \subseteq \mathbb{R}^d$ we have

$$|A|_e \;=\; |A \cap E|_e + |A \backslash E|_e. \quad \square$$

From now on, when we are given a set $E \subseteq \mathbb{R}^d$ *we implicitly assume that it is measurable* unless specifically stated otherwise.

3. THE LEBESGUE INTEGRAL

Our goal in this section is to define the integral of most real- or complex-valued functions, including functions for which the Riemann integral is not defined.

Definition 3.1. Let $E \subseteq \mathbb{R}^d$, and consider a function mapping E to the extended nonnegative reals, i.e., $f\colon E \to [0,\infty]$.

(a) The *graph* of f is

$$\Gamma(f,E) \;=\; \bigl\{(x,f(x)) \in \mathbb{R}^{d+1} : x \in E,\ f(x) < \infty\bigr\}.$$

(b) The *region under the graph of* f is the set $R(f,E)$ of all points $(x,y) \in \mathbb{R}^{d+1}$ with $x \in E$ and $y \in \mathbb{R}$ and such that $0 \le y \le f(x)$ if $f(x) < \infty$, or $0 \le y < \infty$ if $f(x) = \infty$. □

We begin by defining the integral of a nonnegative, real-valued function. Later we will extend this definition to general real- or complex-valued functions.

Definition 3.2. Let E be a measurable subset of $\mathbb{R}^d$, and suppose that $f\colon E \to [0,\infty]$.

(a) We say that f is a *measurable function* if $R(f,E)$ is a measurable subset of $\mathbb{R}^{d+1}$. It can be shown that f is a measurable function if and only if $\{x \in \mathbb{R}^d : f(x) \ge \alpha\}$ is a measurable subset of $\mathbb{R}^d$ for each $\alpha \in \mathbb{R}$. Sums, products, and limits of measurable functions are measurable.

(b) If f is a measurable function, then the *Lebesgue integral of f over E* is the measure of the region under the graph of f as a subset of $\mathbb{R}^{d+1}$, i.e.,

$$\int_E f \;=\; \int_E f(x)\,dx \;=\; |R(f,E)|.$$

If the set E is understood, then we may write simply $\int f = \int_E f$. Note that the integral of a nonnegative f lies in the range

$$0 \;\le\; \int_E f \;\le\; \infty. \qquad \square$$

From now on, when we are given a function f *we implicitly assume that it is measurable* unless specifically stated otherwise.

There are many equivalent ways to define the Lebesgue integral. I prefer the one given in Definition 3.2 because it captures the intuition of what an integral should mean, i.e., the integral should represent the "area under the graph" of f. Many texts begin by defining the integral of *step functions*, i.e., functions which take only finitely many distinct values. It is clear how to define the integral of a step function. Then, an arbitrary function f is written as a limit of step functions and the Lebesgue integral of f is defined to be the limit of the integrals of the step functions.

Here are some basic properties of integrals of nonnegative functions.

Theorem 3.3. Let $E \subseteq \mathbb{R}^d$ be given, and suppose that $f,\, g\colon E \to [0,\infty]$.

(a) $\int_E 1 = |E|$.

(b) If $f \le g$, then $\int_E f \le \int_E g$.

(c) If $E_1 \subseteq E_2$, then $\int_{E_1} f \le \int_{E_2} f$.

(d) If $E_1, E_2, \dots$ are disjoint sets in $\mathbb{R}^d$ and $E = \cup E_k$, then $\int_E f = \sum_k \int_{E_k} f$.

(e) $\int_E (f+g) = \int_E f + \int_E g$.

(f) (Tchebyshev's Inequality) If $\alpha > 0$, then $|\{x \in E : f(x) > \alpha\}| \le \frac{1}{\alpha} \int_E f$.

(g) $f = 0$ a.e. on E if and only if $\int_E f = 0$.

(h) If $f = g$ a.e., then $\int_E f = \int_E g$.

Proof. (a) If $f(x) = 1$ for all $x \in E$ then $R(f,E) = \{(x,y) : x \in E,\, 0 \le y \le 1\} = E \times [0,1]$. The seemingly obvious but nontrivial fact that $|E \times F| = |E|\,|F|$ then implies that $|R(f,E)| = |E|$.

(b) Since $R(f,E) \subseteq R(g,E)$, we have $\int_E f = |R(f,E)| \le |R(g,E)| = \int_E g$.

(c) This follows similarly from the fact that $R(f,E_1) \subseteq R(f,E_2)$.

(d) Note that the sets $R(f,E_k)$ are disjoint and that $R(f,E) = \cup R(f,E_k)$. Therefore, this part follows from Theorem 2.8(b).

(e) This is another "obvious" property that is not trivial to prove using the definition of Lebesgue integral that we have chosen. The proof is not difficult, but it is rather long and technical. The idea is that it is easy to prove if f and g are step functions, and that arbitrary functions can be approximated by step functions.

(f) Let $F = \{x \in E : f(x) > \alpha\}$. Then

$$\int_E f \;\ge\; \int_F f \;\ge\; \int_F \alpha \;=\; \alpha\,|F|.$$

(g) If $f = 0$ a.e. on E then $\int_E f = |R(f,E)| = 0$.

Conversely, suppose that $\int_E f = 0$. Let $F = \{x \in E : f(x) > 0\}$ be the set of points where f is strictly positive. We have to show that $|F| = 0$. Now, for each $\alpha > 0$, we have by Tchebyshev's Inequality that

$$0 \;\le\; |\{x \in E : f(x) > \alpha\}| \;\le\; \frac{1}{\alpha} \int_E f \;=\; 0.$$

In particular, the set $F_n = \{x \in E : f(x) > 1/n\}$ has measure zero for each $n > 0$. However, F is the union of the countably many sets $F_1, F_2, \dots$, so $0 \le |F| \le \sum_n |F_n| = 0$.

(h) If $f = g$ a.e., then $f - g = 0$ a.e., so

$$\int_E f - \int_E g \;=\; \int_E (f-g) \;=\; 0. \qquad \square$$

Note that Theorem 3.3(h) says that if two functions f and g are equal except on a set of measure zero, then their integrals are equal. Hence, given a function f we can change the values of f on any set of measure zero without changing the integral of the f. Hence, whenever we are concerned only with integrals, we typically do not distinguish between two functions that are equal except on a set of measure zero.

Now we can extend the definition of Lebesgue integral to more general functions.

Definition 3.4 (Lebesgue Integral for Real-Valued Functions)**.** Let f be a real-valued function $f\colon E \to [-\infty, \infty]$. We write f as a difference of two nonnegative functions by defining

$$f^+(x) \;=\; \begin{cases} f(x), & f(x) \geq 0, \\ 0, & f(x) < 0, \end{cases} \qquad \text{and} \qquad f^-(x) \;=\; \begin{cases} 0, & f(x) \geq 0, \\ |f(x)|, & f(x) < 0, \end{cases}$$

so that

$$f \;=\; f^+ - f^- \qquad \text{and} \qquad |f| \;=\; f^+ + f^-.$$

We say that f is measurable if both f^+ and f^- are measurable, and in this case we define the Lebesgue integral of f to be

$$\int_E f(x)\,dx \;=\; \int_E f^+(x)\,dx \;-\; \int_E f^-(x)\,dx,$$

as long as this does not have the form $\infty - \infty$ (in that case, the integral is undefined). $\square$

Since $0 \leq f^+, f^- \leq |f| = f^+ + f^-$, it follows from Definition 3.4 that

$$\begin{aligned} \int_E f \text{ exists as a finite real value} \quad &\iff \quad \int_E f^+ < \infty \quad \text{and} \quad \int_E f^- < \infty \\ &\iff \quad \int_E f^+ \;+\; \int_E f^- < \infty \\ &\iff \quad \int_E |f| < \infty. \end{aligned}$$

Note that if $|f(x)| = \infty$ on a set with positive Lebesgue measure then $\int |f| = \infty$. Equivalently, if $\int_E |f| < \infty$ then $|f(x)| < \infty$ a.e. However, it is possible to have $|f(x)| < \infty$ a.e. yet still have $\int_E |f| = \infty$. For example, if $f(x) = 1$ for all $x \in \mathbb{R}$, then $\int f = \infty$.

Definition 3.5 (Lebesgue Integral for Complex-Valued Functions)**.** Suppose that $f\colon E \to \mathbb{C}$. Split f into real and imaginary parts by writing $f = \mathrm{Re}\,(f) + i\,\mathrm{Im}\,(f)$. Then we define the Lebesgue integral of f to be

$$\int_E f \;=\; \int_E \mathrm{Re}\,(f) \;+\; i \int_E \mathrm{Im}\,(f),$$

as long as both integrals on the right are defined and finite. $\square$

Note that $|\mathrm{Re}\,(f)|,\, |\mathrm{Im}\,(f)| \leq |f| \leq |\mathrm{Re}\,(f)| + |\mathrm{Im}\,(f)|$. Therefore

$$\int_E |f| < \infty \quad \iff \quad \int_E |\mathrm{Re}\,(f)|,\; \int_E |\mathrm{Im}\,(f)| < \infty.$$

Consequently,

$$\int_E f \text{ exists as a complex number} \qquad \Longleftrightarrow \qquad \int_E |f| < \infty.$$

Definition 3.6. We say that a function $f\colon E \to [-\infty, \infty]$ or $f\colon E \to \mathbb{C}$ is *integrable* if $\int_E |f| < \infty$. The collection of all integrable functions on E is called $L^1(E)$. That is, if we are dealing with real-valued functions then

$$L^1(E) \;=\; \Big\{f\colon E \to [-\infty, \infty] \;:\; \int_E |f| < \infty\Big\},$$

or if we are dealing with complex-valued functions then

$$L^1(E) \;=\; \Big\{f\colon E \to \mathbb{C} \;:\; \int_E |f| < \infty\Big\}.$$

The choice of real-valued versus complex-valued functions is usually clear from context. In either case, $L^1(E)$ is a vector space under the usual operations of function addition and multiplication by scalars, and

$$\|f\|_1 = \int_E |f|$$

defines a norm on this space, meaning that:

(a) $0 \le \|f\|_1 < \infty$ for all $f \in L^1(\mathbb{R})$,

(b) $\|f\|_1 = 0$ if and only if $f = 0$ a.e.,

(c) $\|cf\|_1 = |c|\,\|f\|$ for all $f \in L^1(\mathbb{R})$ and all scalars c, and

(d) $\|f+g\|_1 \le \|f\|_1 + \|g\|_1$ for all $f,\, g \in L^1(\mathbb{R})$. □

Note that in property (a), we only have that $\|f\|_1 = 0$ implies $f = 0$ a.e., not that $f = 0$. In this sense $\|f\|_1$ does not quite satisfy the requirements of a norm (instead, it is only a seminorm). On the other hand, we have declared that we will not distinguish between two functions that are equal a.e., and with this identification we do have that $\|f\|_1$ satisfies all the requirements of a norm. In other words, we regard any function that is 0 a.e. as being "the" zero element of $L^1(\mathbb{R})$, and if $f = g$ a.e. then we regard f and g as being the "same" element of $L^1(\mathbb{R})$. To be more precise, we are really taking the elements of $L^1(\mathbb{R})$ to be equivalence classes of functions that are equal a.e. This distinction between equivalence classes of functions and the functions themselves is not usually an issue, and we will ignore it.

Definition 3.7. In analogy to $L^1(E)$, given $1 \le p < \infty$ we define

$$L^p(E) \;=\; \Big\{f\colon E \to [-\infty, \infty] \;:\; \int_E |f(x)|^p\,dx < \infty\Big\},$$

or make the obvious adjustment if we are dealing with complex-valued functions. It can be shown that $L^p(E)$ is a vector space, and that

$$\|f\|_p \;=\; \left(\int_E |f(x)|^p\,dx\right)^{1/p}$$

is a norm on this space (again in the sense of identification functions that are equal almost everywhere). In fact, $L^p(E)$ is complete in this norm (all Cauchy sequences converge), and is therefore Banach space. If $|E| < \infty$ and $1 \le p < q < \infty$ then $L^q(E) \subseteq L^p(E)$. This inclusion fails if E has infinite measure.

For the case $p = \infty$, we define

$$L^\infty(E) \;=\; \bigl\{f\colon E \to [-\infty,\infty] \;:\; \operatorname*{ess\,sup}_{x\in E} |f(x)| < \infty\bigr\}.$$

Then $L^\infty(E)$ is a Banach space with respect to the norm

$$\|f\|_\infty \;=\; \operatorname*{ess\,sup}_{x\in E} |f(x)|.$$

If $|E| < \infty$, then $L^\infty(E) \subseteq L^p(E)$ for each $1 \le p < \infty$, and in fact $L^\infty(E) = \bigcap_{p\ge 1} L^p(E)$.

For the case $p = 2$,

$$\langle f, g\rangle \;=\; \int_E f(x)\,\overline{g(x)}\,dx$$

defines an inner product on $L^2(E)$. Thus $L^2(E)$ is a Hilbert space as well as a Banach space. Out of all the exponents $1 \le p \le \infty$, only $L^2(E)$ is a Hilbert space. □

Here is a fundamental inequality for the L^p norms.

Theorem 3.8 (Hölder's Inequality)**.** Let $E \subseteq \mathbb{R}^d$ be measurable, and fix $1 \le p \le \infty$. If $f \in L^p(E)$ and $g \in L^{p'}(E)$ then $fg \in L^1(E)$, and

$$\|fg\|_1 \le \|f\|_p\,\|g\|_{p'}.$$

For $1 < p < \infty$, this inequality is

$$\int_E |fg| \;\le\; \left(\int_E |f|^p\right)^{1/p} \left(\int_E |g|^{p'}\right)^{1/p'}.$$

For $p = 2$, Hölder's inequality is known as the *Schwarz*, *Cauchy–Schwarz*, or *Cauchy–Bunyakowski–Schwarz* inequality. It has the form

$$\int_E |fg| \;\le\; \left(\int_E |f|^2\right)^{1/2} \left(\int_E |g|^2\right)^{1/2}.$$

4. SWITCHING INTEGRALS

Suppose that $f(x,y)$ is a function of two variables, with x varying through a domain $E \subseteq \mathbb{R}^m$ and y varying through a domain $F \subseteq \mathbb{R}^n$. Suppose also that we want to integrate f over the entire domain

$$E \times F \;=\; \{(x,y) \in \mathbb{R}^{m+n} : x \in E,\, y \in F\}.$$

In this case, it is very important which set we integrate over first! In general, it is NOT true that if we integrate over x first and y second, we will get the same result as if we integrate over y first and x second.

Fubini's and Tonelli's Theorems give two conditions under which we can safely exchange the order of integration. Fubini's Theorem says that we can the switch integrals if f is an *integrable* function, and Tonelli's Theorem says that we can switch the integrals if f is a *nonnegative* function. If neither theorem applies, then it is possible that $\int_E \int_F f(x,y)\,dx\,dy \neq \int_F \int_E f(x,y)\,dy\,dx$!

Theorem 4.1 (Fubini's Theorem)**.** If any one of the possible double integrals of $|f(x,y)|$ is finite, then interchanging the order of integration of $f(x,y)$ is allowed. That is, if

(a) $\displaystyle\int_E \left(\int_F |f(x,y)|\,dy\right) dx < \infty$, or

(b) $\displaystyle\int_F \left(\int_E |f(x,y)|\,dx\right) dy < \infty$, or

(c) $\displaystyle\iint_{E\times F} |f(x,y)|\,(dx\,dy) < \infty$,

then

$$\int_E \int_F f(x,y)\,dy\,dx \;=\; \int_F \int_E f(x,y)\,dx\,dy \;=\; \iint_{E\times F} f(x,y)\,(dx\,dy). \tag{4.1}$$

Moreover, in this case $g(x) = \displaystyle\int_F f(x,y)\,dy$ is well-defined for almost every x, and g is an integrable function of x, i.e., $g \in L^1(E)$. Similarly, $h(y) = \int_E f(x,y)\,dx$ is well-defined for almost every y, and is an integrable function of y, i.e., $h \in L^1(F)$. □

Theorem 4.2 (Tonelli's Theorem)**.** If $f(x,y) \geq 0$ a.e., then interchanging the order of integration of $f(x,y)$ is allowed, i.e.,

$$\int_E \int_F f(x,y)\,dy\,dx \;=\; \int_F \int_E f(x,y)\,dx\,dy. \quad \Box \tag{4.2}$$

Note that the hypotheses of Fubini's Theorem imply that the integrals in equation (4.1) are finite real or complex numbers. However, the integrals in equation (4.2) may be finite real numbers or ∞. The equality in (4.2) means that if one side is finite then the other side is finite as well, and if one is infinite then the other is infinite as well.

Because a series can be viewed as a "discrete integral," there are analogues of Fubini's and Tonelli's theorems that apply to the problem of interchanging an integral and a sum or interchanging two summations (see Corollary 5.3).

5. SWITCHING INTEGRALS AND LIMITS

Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of functions that converge pointwise almost everywhere, i.e., there is a function f such that $f_n(x) \to f(x)$ for almost every x. The following example shows that this does NOT imply that the integrals of f_n must converge to the integral of f, i.e., it need NOT be true that $\displaystyle\lim_{n\to\infty}\int_E f_n = \int_E \lim_{n\to\infty} f_n$!

Example 5.1. Let f_n be defined as follows. For $0 \le x \le 1/n$ the graph of f_n looks like an isosceles triangle with base $[0,1/n]$ and height n. For all other x we set $f_n(x) = 0$. Then $f_n(x) \to 0$ for every $x \in \mathbb{R}$! However, $\int f_n = 1/2$ for every n, so $\int f_n$ does not converge to $\int 0\,dx = 0$. □

Our goal in this section is to give some conditions under which a limit and an integral can be interchanged. The first result, known as the *Monotone Convergence Theorem* or the *Beppo–Levi Theorem*, applies to the case of nonnegative functions that are *monotone increasing a.e.*, i.e., for which

$$0 \;\le\; f_1(x) \;\le\; f_2(x) \;\le\; f_3(x) \;\le\; \cdots \text{ for a.e. } x \in E.$$

Theorem 5.2 (Monotone Convergence Theorem)**.** Let $E \subseteq \mathbb{R}^d$ be given, and suppose that $f_n\colon E \to [0,\infty]$ for $n \in \mathbb{N}$. If the sequence of functions $\{f_n\}_{n\in\mathbb{N}}$ is monotone increasing a.e. on E and if $\displaystyle\lim_{n\to\infty} f_n(x) = f(x)$ for a.e. $x \in E$, then

$$\lim_{n\to\infty}\int_E f_n \;=\; \int_E \lim_{n\to\infty} f_n \;=\; \int_E f.$$

Proof. Note that $R(f_1,E) \subseteq R(f_2,E) \subseteq \cdots$ and that $R(f,E) = \cup R(f_n,E)$. It therefore follows from Theorem 2.8(d) that

$$\int_E f \;=\; |R(f,E)| \;=\; \lim_{n\to\infty}|R(f_n,E)| \;=\; \lim_{n\to\infty}\int_E f_n. \qquad \square$$

Recall that an infinite series is a limit of the partial sums of the series. Hence, we must be careful when switching an integral and an infinite series. As a corollary of the Beppo–Levi Theorem we can prove the following version of Tonelli's Theorem that gives a condition on when an integral and a summation can be interchanged.

Corollary 5.3 (Tonelli's Theorem)**.** Let $E \subseteq \mathbb{R}^d begiven$, and suppose that $f_n \colon E \to [0, \infty]$ for $n \in \mathbb{N}$. Then

$$\int_E \sum_{n=1}^{\infty} f_n \;=\; \sum_{n=1}^{\infty} \int_E f_n.$$

Proof. Set

$$F_N(x) \;=\; \sum_{n=1}^{N} f_n(x) \qquad \text{and} \qquad F(x) \;=\; \sum_{n=1}^{\infty} f_n(x).$$

Then $F_1(x) \le F_x(x) \le \cdots$ for a.e. x, and $\lim_{N\to\infty} F_N(x) = F(x)$ a.e. Therefore, by the Monotone Convergence Theorem,

$$\int_E F \;=\; \lim_{N\to\infty} \int_E F_N \;=\; \lim_{N\to\infty} \int_E \sum_{n=1}^{N} f_n \;=\; \lim_{N\to\infty} \sum_{n=1}^{N} \int_E f_n \;=\; \sum_{n=1}^{\infty} \int_E f_n. \tag{5.1}$$

Note that we were allowed to switch the sum and the integral in equation (5.1) because it was a *finite* sum. □

If the functions f_n are nonnegative but not monotone increasing, then we may not be able to interchange a limit and an integral. However, the following result states that if the functions f_n are all nonnegative, then we do at least have a particular *inequality.*

Theorem 5.4 (Fatou's Lemma)**.** Let $E \subseteq \mathbb{R}^d$ be given, and suppose that $f_n \colon E \to [0, \infty]$ for $n \in \mathbb{N}$. Then

$$\int_E \liminf_{n\to\infty} f_n \;\le\; \liminf_{n\to\infty} \int_E f_n.$$

Consequently, if the f_n converge pointwise almost everywhere, i.e., if $\lim_{n\to\infty} f_n(x) = f(x)$ a.e., and if the integrals converge as well, then

$$\int_E f\,dx \;\le\; \lim_{n\to\infty} \int_E f_n\,dx.$$

Proof. Set $g_n(x) = \inf_{k\ge n} f_k(x)$. Then $g_1(x) \le g_2(x) \le \cdots$, so $\{g_n\}_{n\in\mathbb{N}}$ is monotone increasing for each x. Define

$$f(x) \;=\; \liminf_{n\to\infty} f_n(x) \;=\; \lim_{n\to\infty} \inf_{k\ge n} f_k(x) \;=\; \lim_{n\to\infty} g_n(x).$$

Then by the Monotone Convergence Theorem and the fact that $g_n(x) \le f_n(x)$, we have

$$\int_E f \;=\; \int_E \lim_{n\to\infty} g_n \;=\; \lim_{n\to\infty} \int_E g_n \;=\; \liminf_{n\to\infty} \int_E g_n \;\le\; \liminf_{n\to\infty} \int_E f_n.$$ □

The Lebesgue Dominated Convergence Theorem, or LCDT, is perhaps the most important and useful result of this section. It applies to functions that aren't necessarily nonnegative or monotone increasing, and it is the theorem to use in most cases.

Theorem 5.5 (Lebesgue Dominated Convergence Theorem)**.** Let $E \subseteq \mathbb{R}^d$ be given, and suppose that $f_n \colon E \to [-\infty, \infty]$ for $n \in \mathbb{N}$. Suppose also that the functions $f_n(x)$ converge pointwise almost everywhere, i.e., $\lim_{n\to\infty} f_n(x) = f(x)$ for a.e. $x \in E$. If there is a *single* function g such that:

(a) $|f_n(x)| \le g(x)$ a.e. for every n, and

(b) g is *integrable*, i.e., $\int_E |g| < \infty$,

then

$$\lim_{n\to\infty} \int_E f_n \;=\; \int_E \lim_{n\to\infty} f_n \;=\; \int_E f.$$

In fact, even more is true in this case: f_n converges to f in L^1-norm, i.e.,

$$\lim_{n\to\infty} \|f - f_n\|_1 \;=\; \lim_{n\to\infty} \int_E |f - f_n| \;=\; 0. \quad \square$$

Proof. We will give the proof for nonnegative f_n only, but it can be extended to general f. Suppose that $f_n \ge 0$ a.e. Then by Fatou's Lemma,

$$\int_E f \;=\; \int_E \liminf_{n\to\infty} f_n \;\le\; \liminf_{n\to\infty} \int_E f_n.$$

Further, since $g - f_n \ge 0$ a.e., we can apply Fatou's Lemma to the functions $g - f_n$, to obtain

$$\begin{aligned}
\int_E g - \int_E f \;&=\; \int_E (g - f) \\
&=\; \int_E \liminf_{n\to\infty} (g - f_n) \\
&\le\; \liminf_{n\to\infty} \int_E (g - f_n) \qquad \text{(by Fatou's Lemma)} \\
&=\; \liminf_{n\to\infty} \left(\int_E g - \int_E f_n \right) \\
&=\; \int_E g - \limsup_{n\to\infty} \int_E f_n.
\end{aligned}$$

Therefore,

$$\int_E f \;\le\; \liminf_{n\to\infty} \int_E f_n \;\le\; \limsup_{n\to\infty} \int_E f_n \;\le\; \int_E f.$$

Hence $\lim_{n\to\infty} \int_E f_n$ exists and equals $\int_E f$. $\square$

6. CONVOLUTION

Definition 6.1. Let f and g be real- or complex-value functions with domain $\mathbb{R}^d$. Then the *convolution* of f and g is the function $f * g$ defined by

$$(f * g)(x) \;=\; \int f(y)\, g(x-y)\, dy,$$

whenever this is well-defined. □

Theorem 6.2. If $f,\, g \in L^1(\mathbb{R}^d)$ then $f * g \in L^1(\mathbb{R}^d)$, and $\|f * g\|_1 \le \|f\|_1\, \|g\|_1$.

Proof. We start by computing:

$$\int |(f*g)(x)|\, dx \;=\; \int \left| \int f(y)\, g(x-y)\, dy \right| dx \;\le\; \iint |f(y)\, g(x-y)|\, dy\, dx \;=\; (*).$$

Because $|f(y)\, g(x-y)| \ge 0$ for all x and y, Tonelli's Theorem allows us to interchange the order of integration in $(*)$. So, we can continue as follows:

$$(*) \;=\; \iint |f(y)\, g(x-y)|\, dx\, dy \;=\; \int |f(y)| \left(\int |g(x-y)|\, dx \right) dy \;=\; (**).$$

Now, since we are integrating over all of $\mathbb{R}^d$, we know that

$$\int |g(x-y)|\, dx \;=\; \int |g(x)|\, dx$$

(this wouldn't necessarily be true if we were integrating on a finite domain). Therefore,

$$(**) \;=\; \int |f(y)| \left(\int |g(x)|\, dx \right) dy \;=\; \int |f(y)|\, \|g\|_1\, dy \;=\; \|g\|_1 \int |f(y)|\, dy \;=\; \|g\|_1\, \|f\|_1.$$

Put it all together and we have shown $\|f * g\|_1 \le \|g\|_1\, \|f\|_1$. □

In fact, Theorem 6.2 is just a special case of the following more general result.

Theorem 6.3 (Young's Convolution Inequality)**.** If $1 \le p \le \infty$ and $f \in L^p(\mathbb{R}^d)$ and $g \in L^1(\mathbb{R}^d)$ then $f * g \in L^p(\mathbb{R}^d)$, and

$$\|f * g\|_p \;\le\; \|f\|_p\, \|g\|_1.$$

Proof. We've already done the case $p = 1$. The case $p = \infty$ is easy, so I will leave it as an exercise. For other p's it is a little tricky.

First let p' be the *dual exponent* to p, i.e., the number which satisfies $\frac{1}{p} + \frac{1}{p'} = 1$. Then write

$$|(f*g)(x)| \;\le\; \int |f(y)\, g(x-y)|\, dy \;=\; \int |f(y)\, g(x-y)^{1/p}|\, |g(x-y)^{1/p'}|\, dy \;=\; (*).$$

Now apply Hölder's inequality to the two parts of the integrand:

$$\begin{aligned}
(*) \;\le\; & \left(\int |f(y)\, g(x-y)^{1/p}|^p \, dy\right)^{1/p} \left(\int |g(x-y)^{1/p'}|^{p'} \, dy\right)^{1/p'} \\
=\; & \left(\int |f(y)|^p \, |g(x-y)| \, dy\right)^{1/p} \left(\int |g(x-y)| \, dy\right)^{1/p'} \\
=\; & \left(\int |f(y)|^p \, |g(x-y)| \, dy\right)^{1/p} \left(\int |g(y)| \, dy\right)^{1/p'} \\
=\; & \|g\|_1^{1/p'} \left(\int |f(y)|^p \, |g(x-y)| \, dy\right)^{1/p}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|f * g\|_p^p = \int |(f * g)(x)|^p \, dx \;&=\; \|g\|_1^{p/p'} \iint |f(y)|^p \, |g(x-y)| \, dy \, dx \\
&=\; \|g\|_1^{p/p'} \iint |f(y)|^p \, |g(x-y)| \, dx \, dy \\
&=\; \|g\|_1^{p/p'} \int |f(y)|^p \left(\int |g(x-y)| \, dx\right) dy \\
&=\; \|g\|_1^{p/p'} \int |f(y)|^p \left(\int |g(x)| \, dx\right) dy \\
&=\; \|g\|_1^{p/p'} \int |f(y)|^p \, \|g\|_1 \, dy \\
&=\; \|g\|_1^{1+p/p'} \, \|f\|_p^p \\
&=\; \|g\|_1^p \, \|f\|_p^p.
\end{aligned}$$

Take pth roots and you're done. □

7. CROSS PRODUCT BASES

Suppose that we have an orthonormal basis for the Hilbert space

$$L^2[a,b] \;=\; \{f\colon [a,b] \to \mathbb{C} \;:\; \|f\|_2 = \left(\int_a^b |f(x)|^2\,dx\right)^{1/2} < \infty\}$$

of functions that are square-integrable defined on the interval $[a,b]$, with inner product

$$\langle f, g\rangle \;=\; \int_a^b f(x)\,\overline{g(x)}\,dx.$$

We will show how to use this orthonormal basis to construct an orthonormal basis for the Hilbert space

$$L^2([a,b]\times[a,b]) \;=\; \{F\colon [a,b]\times[a,b] \to \mathbb{C} \;:\; \|F\|_2 = \left(\int_a^b\int_a^b |F(x,y)|^2\,dx\,dy\right)^{1/2} < \infty\},$$

of functions that are square-integrable on the square $[a,b]\times[a,b]$, under the inner product

$$\langle F, G\rangle \;=\; \int_a^b\int_a^b F(x,y)\,\overline{G(x,y)}\,dx\,dy.$$

Theorem 7.1. Suppose that $\{f_n(x)\}_{n\in\mathbb{N}}$ is an orthonormal basis for $L^2[a,b]$, and define

$$F_{mn}(x,y) \;=\; f_m(x)\,f_n(y).$$

Then $\{F_{mn}(x,y)\}_{m,n\in\mathbb{N}}$ is an orthonormal basis for $L^2([a,b]\times[a,b])$.

Proof. First we check that the functions F_{mn} are indeed orthonormal:

$$\begin{aligned}
\langle F_{mn}, F_{jk}\rangle \;&=\; \int_a^b\int_a^b F_{mn}(x,y)\,\overline{F_{jk}(x,y)}\,dx\,dy \\
&=\; \int_a^b\int_a^b f_m(x)\,f_n(y)\,\overline{f_j(x)}\,\overline{f_k(y)}\,dx\,dy \\
&=\; \int_a^b f_n(y)\,\overline{f_k(y)}\left(\int_a^b f_m(x)\,\overline{f_j(x)}\,dx\right)dy \\
&=\; \int_a^b f_n(y)\,\overline{f_k(y)}\,\langle f_m, f_j\rangle\,dy \\
&=\; \langle f_m, f_j\rangle \int_a^b f_n(y)\,\overline{f_k(y)}\,dy \\
&=\; \langle f_m, f_j\rangle\,\langle f_n, f_k\rangle \\
&=\; \begin{cases} 1, & \text{if } m = j \text{ and } n = k, \\ 0, & \text{if } m \neq j \text{ or } n \neq k. \end{cases}
\end{aligned}$$

This establishes that $\{F_{mn}\}$ is an orthonormal system in $L^2([a,b]\times[a,b])$.

Now we have to show that this orthonormal system is an orthonormal basis. We have several choices for doing this. One way is to show that $\{F_{mn}\}$ is *complete*, i.e., the only function $F \in L^2([a,b] \times [a,b])$ that is orthogonal to every F_{mn} is the zero function. So, suppose that $F \in L^2([a,b] \times [a,b])$ is such that $\langle F, F_{mn} \rangle = 0$ for every m and n. It would be easy to proceed if it was the case that $F(x,y) = f(x)\, g(y)$ for some functions $f,\, g \in L^2[a,b]$, but it is important to note that only SOME of the functions in $L^2([a,b] \times [a,b])$ can be "factored" in this way. So, we have to be more careful. For a *general* function $F(x,y)$, we begin by computing that

$$\begin{aligned} 0 \;=\; \langle F, F_{mn} \rangle \;&=\; \int_a^b \int_a^b F(x,y)\, \overline{F_{mn}(x,y)}\, dx\, dy \\ &=\; \int_a^b \int_a^b F(x,y)\, \overline{f_m(x)}\, \overline{f_n(y)}\, dx\, dy \\ &=\; \int_a^b \biggl(\int_a^b F(x,y)\, \overline{f_m(x)}\, dx \biggr)\, \overline{f_n(y)}\, dy \\ &=\; \int_a^b h_m(y)\, \overline{f_n(y)}\, dy \\ &=\; \langle h_m, f_n \rangle, \end{aligned} \tag{7.1}$$

where

$$h_m(y) \;=\; \int_a^b F(x,y)\, \overline{f_m(x)}\, dx.$$

Note that $h_m \in L^2[a,b]$ because, by the Schwarz inequality,

$$\begin{aligned} \|h_m\|_2^2 \;=\; \int_a^b |h_m(y)|^2\, dy \;&=\; \int_a^b \biggl| \int_a^b F(x,y)\, \overline{f_m(x)}\, dx \biggr|^2 dy \\ &\le\; \int_a^b \biggl(\int_a^b |F(x,y)|^2\, dx \biggr) \biggl(\int_a^b |f_m(x)|^2\, dx \biggr)\, dy \\ &=\; \int_a^b \biggl(\int_a^b |F(x,y)|^2\, dx \biggr)\, \|f_m\|^2\, dy \\ &=\; \int_a^b \int_a^b |F(x,y)|^2\, dx\, dy \\ &=\; \|F\|_2^2 \;<\; \infty. \end{aligned}$$

Moreover, considering now a fixed m, equation (7.1) says that h_m is orthogonal to every function f_n in the orthonormal basis $\{f_n\}_{n \in \mathbb{N}}$ for $L^2[a,b]$. Therefore $h_m = 0$ a.e.

We still have to show that $F = 0$ a.e. So, for each y let $G_y(x)$ be the function defined by

$$G_y(x) \;=\; F(x,y).$$

As was the case for h_m, you can easily check that for each fixed y, the function G_y is in $L^2[a,b]$ (as a function of x alone). Moreover, since $h(y) = 0$ a.e., we have

$$\langle G_y(x), f_m(x)\rangle \;=\; \int_a^b F(x,y)\,\overline{f_m(y)}\,dx \;=\; h_m(y) \;=\; 0.$$

That is, G_y is orthogonal to every f_m, and since $\{f_m\}$ is an orthonormal basis for $L^2[a,b]$ we therefore conclude that $G_y(x) = 0$ for a.e. x. Since this is true for a.e. y, we conclude that $F(x,y) = 0$ for a.e. (x,y) (OK, there's a little argument to fill about sets in the plane with measure zero, but it's easy). We started with a function $F(x,y)$ that was orthogonal to every $f_m(x)\,f_n(y)$ and showed that this F must be zero a.e., so this shows that $\{f_m(x)\,f_n(y)\}$ is complete, and hence is an orthonormal basis since we have already shown that it is an orthonormal system.

Here is another way of showing that $\{f_m(x)\,f_n(y)\}$ is complete. Instead of showing that only the zero function is orthogonal to every element of this system, we can show that its finite linear span is dense. So, choose any function $F(x,y) \in L^2([a,b]\times[a,b])$. By any one of several arguments, we know that the set of continuous functions is dense in $L^2([a,b]\times[a,b])$. Therefore, there is a continuous function $G(x,y) \in L^2([a,b]\times[a,b])$ such that $\|F-G\|_2 < \varepsilon$. Now subdivide the square $[a,b]\times[a,b]$ into finitely many smaller squares Q_k for $k = 1,\dots N$. Then we can approximate G by a step function

$$H(x,y) \;=\; \sum_{k=1}^{N} c_k\,\chi_{Q_k}(x,y).$$

For example, by taking the squares small enough and letting c_k be the average value of G on the square Q_k, we can make $\|G-H\|_2 < \varepsilon$. Now, each square is a cross product of two intervals: $Q_k = I_k \times J_k$. Therefore,

$$\chi_{Q_k}(x,y) \;=\; \chi_{I_k}(x)\,\chi_{J_k}(y).$$

Since χ_{I_k}, $\chi_{J_k} \in L^2[a,b]$, we can write

$$\chi_{I_k}(x) \;=\; \sum_m a_{mk}\,f_m(x) \qquad \text{and} \qquad \chi_{J_k}(y) \;=\; \sum_n b_{nk}\,f_n(y)$$

for some scalars a_{mk} and b_{nk} (in fact, these scalars are given by inner products, but that doesn't really matter for this argument). Hence,

$$\begin{aligned} H(x,y) \;=\; \sum_{k=1}^{N} c_k\,\chi_{I_k}(x)\,\chi_{J_k}(y) \;&=\; \sum_{k=1}^{N} c_k \left(\sum_m a_{mk}\,f_m(x)\right)\left(\sum_n b_{nk}\,f_n(y)\right) \\ &=\; \sum_m \sum_n \left(\sum_{k=1}^{N} c_k\,a_{mk}\,b_{nk}\right) f_m(x)\,f_n(y) \\ &=\; \sum_m \sum_n d_{mn}\,f_m(x)\,f_n(y), \end{aligned} \tag{7.2}$$

where $d_{mn} = \sum_{k=1}^{N} c_k\,a_{mk}\,b_{nk}$ are some new scalars (note that this is a finite sum, so it is well-defined). But equation (7.2) says that H is an infinite linear combination of the functions $f_m(x)\,f_n(y)$, hence can be approximated to within ε in L^2-norm by a function in $\text{span}\{f_m(x)\,f_n(y)\}$. Therefore F is approximated to within 3ε in L^2-norm by a function in this span. Hence the span is dense, so $\{f_m(x)\,f_n(y)\}$ is complete, and therefore forms an orthonormal basis since we already know that it is an orthonormal system. $\square$

Example 7.2. The set $\{e^{2\pi inx}\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $L^2[0,1]$. Therefore, the collection

$$\{e^{2\pi imx}\,e^{2\pi iny}\}_{m,n\in\mathbb{Z}}$$

is an orthonormal basis for $L^2([0,1]\times[0,1])$. $\square$

Theorem 7.1 can easily be adapted to cover the case of finding an orthonormal basis for $L^2([a,b]\times[c,d])$ or $L^2(\mathbb{R}^2)$, etc. The general principle is that if $\{f_n\}$ is an orthonormal basis for $L^2(\Omega_1)$ and $\{g_n\}$ is an orthonormal basis for $L^2(\Omega_2)$, then $\{f_m(x)\,g_n(y)\}$ is an orthonormal basis for $L^2(\Omega_1\times\Omega_2)$.