4 The Integral

4.1 Measurable Functions

Before we can define the integral, we need to spend a bit of time talking about the sort of functions we can integrate.

First, we want to get some notational conventions out of the way. We'll often need to talk about the *extended real number line* $\overline{R} = [-\infty, \infty]$. Most of the algebra with ∞ does what you probably think it should by this point; but it's important to note that sometimes $0 \cdot \infty$ is undefined and other times it's 0.

In order to do integrals, We want to take functions where we can approximate the output in some reasonable sense: if we look at all the values where f takes on a value "near" a, the set we get will be sensible. We thus define:

Definition 4.1. Let X be a set and $\mathcal{M}$ a σ-algebra on X. Let $f : X \to \overline{\mathbb{R}}$. We say that f is $\mathcal{M}$-*measurable* if, for all $t\overline{R}$, the set $f^{-1}([-\infty, t])$ is $\mathcal{M}$-measurable.

Another way of expressing this is that for all $t \in \overline{\mathbb{R}}$, we have $\{x : f(x) \leq t\} \in \mathcal{M}$.

Exercise 4.2. *Let $A \subset X$. Prove that the characteristic function χ_A is $\mathcal{M}$-measurable if and only if $A \in \mathcal{M}$.*

Exercise 4.3. *Let $\mathcal{M} = \{\varnothing, X\}$ and $\mathcal{N} = 2^X$. Describe explicitly the sets of $\mathcal{M}$-measurable functions and of $\mathcal{N}$-measurable functions.*

You might ask why we specficially look at $[-\infty, t]$ and not $[-\infty, t)$ or $(t, \infty]$ or something. The answer is that it doesn't matter.

Proposition 4.4. *Let $\mathcal{M}$ be a σ-algebra and $f : X \to \overline{\mathbb{R}}$. Then the following are equivalent:*

1. *f is measurable*
2. *$f^{-1}([-\infty, t)) \in \mathcal{M}$ for any $t \in (-\infty, \infty]$*
3. *$f^{-1}((t, \infty]) \in \mathcal{M}$ for any $t \in \overline{\mathbb{R}}$*
4. *$f^{-1}((t, \infty,]) \in \mathcal{M}$ for any $t \in [-\infty, \infty)$*
5. *$f^{-1}(\{-\infty\}) \in \mathcal{M}$, $f^{-1}(\{\infty\}) \in \mathcal{M}$, and $f^{-1}(E) \in \mathcal{M}$ for every Borel set $E \subset \mathbb{R}$.*

Proof. $\square$

Proposition 4.5. *Assume $f, g : X \to \mathbb{R}$ are $\mathcal{M}$-measurable, and $\phi : \mathbb{R} \to \mathbb{R}$ is Borel measurable. Then*

1. $\phi \circ f$ *is $\mathcal{M}$-measurable.*
2. *If $f \neq 0$ then $\frac{1}{f}$ is $\mathcal{M}$-measurable.*
3. *If $0 < p < \infty$ then $|f|^p$ is $\mathcal{M}$-measurable.*
4. $f + g$ *is $\mathcal{M}$-measurable.*
5. fg *is $\mathcal{M}$-measurable.*

Proof. 1. If E is a Borel set, then $\phi^{-1}(E)$ is Borel, and thus $f^{-1}(\phi^{-1}(E)) \in \mathcal{M}$.

2. Excercise. Prove that $\phi(t) = \frac{1}{t}$ is Borel measurable and then conclude this result.

3. The function $\phi(t) = |t|^p$ is continuous, and thus Borel measurable.

4. This one takes a small amount of work.

   We know that $f(x) + g(x) < t$ if and only if $f(x) < t - g(x)$, if and only if there is a $r \in \mathbb{Q}$ such that $f(x) < r < t - g(x)$. So we can write

$$(f+g)^{-1}([-\infty, t)) = \bigcup_{r \in \mathbb{Q}} f^{-1}((-\infty, r)) \cap g^{-1}((-\infty, t - r)).$$

Here we use a dumb trick called polarization. We know that $f \cdot f$ is measurable for any measurable f, by (3). So we write

$$fg = \frac{1}{4}(f+g)^2 - \frac{1}{4}(f-g)^2.$$

Since $f + g$ and $f - g$ are measurable, this whole function is measurable. □

Proposition 4.6. *Suppose $f_k : X \to \overline{\mathbb{R}}$ is $\mathcal{M}$-measurable for all $k \in \mathbb{N}$. Then the following functions are all $\mathcal{M}$-measruable:*

- $\sup_k f_k$
- $\inf_k f_k$
- $\limsup_{k \to \infty} f_k$
- $\liminf_{k \to \infty} f_k$

- $\lim_{k\to\infty} f_k$, *if the pointwise limit exists.*

Proof. We can write

$$\{x : \sup f_k(x) \leq t\} = \bigcap_k \{x : f_k(x) \leq t\}.$$

The right-hand side is an intersection of measurable sets since each f_k is measurable, so the left-hand side is measurable. Similarly

$$\{x : \inf f_k(x) \geq t\} = \bigcap_k \{x : f_k(x) \geq t\}.$$

Then we know that $\limsup f_k = \inf \sup f_k$, and $\liminf f_k = \sup \inf f_k$. Since both sup and inf are measurable, so are these.

Finally, if $\lim f_k$ exists, then $\lim f_k = \limsup f_k = \liminf f_k$ is measurable.

□

4.2 Simple Functions

Definition 4.7. A *simple* function from X to $\overline{\mathbb{R}}$ is any function which assumes finitely many values. Thus we can write

$$s = \sum_{k=1}^{m} \alpha_k \chi_{A_k}$$

where the sets A_k are disjoint and the numbers $\alpha_k \in \overline{\mathbb{R}}$ are distinct.

Exercise 4.8. *A simple function s is measurable if and only if each set A_k is measurable.*

Definition 4.9. Let $a \in \overline{\mathbb{R}}$. We define

$$a_+ = \begin{cases} a & a \geq 0 \\ 0 & a < 0 \end{cases}$$

$$a_- = \begin{cases} 0 & a \geq 0 \\ -a & a < 0 \end{cases}$$

We call these the *positive part* and *negative part* of a.

We observe that $a = a_+ - a_-$ and $|a| = a_+ + a_-$. A silly but useful observation is that $a_+ a_- = 0$.

We can extend this definition for functions: if $f : X \to \overline{\mathbb{R}}$ then $f_+(x) = (f(x))_+$ and $f_-(x) = (f(x))_-$.

Exercise 4.10. *If f is $\mathcal{M}$-measurable, then so are f_+ and f_-.*

It's easy to see that the limit of a sequence of simple measurable functions is measurable; this follows directly from proposition 4.6. Much less obvious is that the converse of this statement is also true: every measurable function is the limit of a sequence of simple measurable functions.

That means that the measurable functions are precisely the closure of simple measureable functions under pointwise limits. A function is measurable if and only if it is the limit of $s_k = \sum_{i=1}^{m_k} \alpha_{i,k}\chi_{A_{i,k}}(x)$.

Theorem 4.11. *Suppose $f : X \to \overline{\mathbb{R}}$ is $\mathcal{M}$-measurable. Then there is a sequence of $\mathcal{M}$-measurable simple functions $s_1, s_2, \dots$ that converge pointwise to f on X. That is, $\lim_{k\to\infty} s_k(x) = f(x)$ for every $x \in X$.*

If $f \geq 0$, we may choose the sequence such that $0 \leq s_1 \leq s_2 \leq \dots$. We may always choose the sequence such that $|s_1| \leq |s_2| \leq \dots$.

Proof. First we prove the case where $f \geq 0$. We define s_k through the following complicated-looking formula:

$$s_k(x) = \begin{cases} \frac{i}{2^k} & \frac{i}{2^k} \leq f(x) < \frac{i+1}{2^k} \leq k \\ k & k \leq f(x) \end{cases}$$

This formula does two things. First, the maximum possible value we give s_k is k, and the only values we allow are those that are integer multiples of $\frac{1}{2^k}$. Thus there are $k2^k+1$ possible values of s_k, so it is simple.

We need to check two things. First, does the sequence s_k converge to f? For large k, we have $|f(x) - s_k(x)| < \frac{1}{2^k}$, so the sequence converges pointwise. (In exercise 4.12 you will prove that this convergence is uniform if the function f is bounded).

Now is each s_k measurable? We have that $s_k(x) = \frac{i}{2^k}$ when $\frac{i}{2^k} \leq f(x) < \frac{i+1}{2^k}$, so

$$s_k^{-1}\left\{\frac{i}{2^k}\right\} = f^{-1}\left(\left[\frac{i}{2^k}, \frac{i+1}{2^k}\right)\right)$$

and the latter set is measurable because f is measurable. The only other possible value of s_k is k, which happens when $k \leq f(x)$; then we have

$$s_k^{-1}\{k\} - f^{-1}\left([k, \infty]\right)$$

and again this set is measurable since f is measurable. Thus f is the pointwise limit of a sequence of simple measurable functions.

For a general function f, we can just leverage the previous result, in a way that we'll use a lot. We have a sequence of functions $0 \leq s_1 \leq s_2 \leq \dots$ converging to f_+, and a sequence $0 \leq t_1 \leq t_2 \leq \dots$ converging to f_-. Then the sequence $s_k - t_k$ converges to f.

□

Exercise 4.12. *If $f : X \to \mathbb{R}$ is measurable and bounded, prove that it is the* uniform *limit of a sequence of measurable functions.*

This result has one simple consequence that isn't strictly speaking about measurable functions, but which will be extremely useful to us. Remember we said that we can approximate any Lebesgue measurable set with a Borel set: a Lebesgue measurable set is a Borel set union a set of measure zero. This means that we can approximate a Lebesgue measurable function with a Borel measurable function.

Theorem 4.13. *Suppose $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is Lebesgue measurable. Then there is a Borel measurable function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ such that $\{x : f(x) \neq g(x)\}$ has measure zero.*

Proof. As usual, start by assuming $f \geq 0$. There is an increasing sequence $0 \leq s_1 \leq s_2 \leq \dots$ of Lebesgue measurable simple functions s_k that converge to f. Then for each k, we can write

$$s_k = \sum_{i=1}^{m_k} \alpha_{i,k} \chi_{A_{i,k}}$$

whre each $A_{i,k}$ is a Lebesgue measurable set. Then there is a Borel set $E_{i,k}$ such that $\lambda(A_{i,k} \setminus E_{i,k}) = 0$. Define

$$t_k = \sum_{i=1}^{m_k} \alpha_{i,k} \chi_{E_{i,k}}.$$

This is a simple, Borel measurable function such that $0 \leq t_k \leq s_k$ and $t_k = s_k$ except on a set N_k of measure zero.

Define $g = \sup_k t_k$; this is Borel measurable since it's the supremum of Borel measurable functions. Then $g(x) = f(x)$ unless $x \in (A_{i,k} \setminus E_{i,k})$ for some i. But

$$\lambda\left(\bigcup_{i=1}^{m_k} A_{i,k} \setminus E_{i,k}\right) = \sum_{i=1}^{m_k} \lambda(A_{i,k} \setminus E_{i,k}) = 0.$$

Now suppose f is any function. We have shown that we can approximate f_+ with some Borel measurable g_+, and can approximate f_- with some Borel measurable g_-. Then $g = g_+ - g_-$ is a Borel measurable function, and $g(x) = f(x)$ except on a set of measure zero. □

And now, with those preliminaries completed, we are ready to start defining the integral.

For the moment, we'll let S be the set of Lebesgue-measurable simple functions $s : \mathbb{R}^n \to [0, \infty)$.

Definition 4.14. Let $s \in S$, with $s = \sum_{k=1}^m \alpha_k \chi_{A_k}$ where the A_k are disjoint measurable sets. Then the *integral* of s is

$$\int s \, d\lambda = \sum_{k=1}^m \alpha_k \lambda(A_k).$$

Here we use the convention that $0 \cdot \infty = 0$. If $\alpha_k = 0$, it doesn't matter if $\lambda(A_k)$ is infinite. And when we allow ∞-valued functions, we'll ignore that as long as it happens on a set of measure 0.

We can always assume that $\bigcup A_k = \mathbb{R}^n$ if that's convenient; if it isn't true, we can always define $A_{m+1} = \left(\bigcup_{k=1}^m A_k\right)^C$ and $\alpha_{m+1} = 0$, and nothing substantive will change.

It's not *immediately* clear that this definition is well-defined; there is more than one way to describe a simple function like this. But we will prove that it is well-defined in the next proposition.

Before we do that, though, it's worth emphasizing the ways this is similar to the Riemann integral. We can look at the Riemann integral as approximating functions below by a series of step functions. So any finite Riemann sum will add up a finite collection of heights-times-widths.

Here the α_k plays the role of the height, and the $\lambda(A_k)$ plays the role of the width. But we get some extra flexibility by not requiring our A_k to all be intervals; this flexibility is given by all the work we did to define the Lebesgue measure in section 2.

Proposition 4.15.

1. *$\int s \, d\lambda$ is well-defined, and doesn't depend on the measurable sets we choose to divide $\mathbb{R}^n$ into.*
2. *$0 \leq \int s \, d\lambda \leq \infty$.*
3. *If $0 \leq c < \infty$ is a constant, then $\int cs \, d\lambda = c \int s \, d\lambda$.*
4. *If $s, t \in S$, then $\int (s+t) \, d\lambda = \int s \, d\lambda + \int t \, d\lambda$.*
5. *If $s, t \in S$ and $s \leq t$, then $\int s \, d\lambda \leq \int t \, d\lambda$.*

Proof. We're going to prove (5) first, and that's going to give us most of the rest for free.

Suppose we have $s, t \in S$ with $s \leq t$. Then we have representations

$$s = \sum_{k=1}^m \alpha_k \chi_{A_k} \qquad t = \sum_{j=1}^n \beta_j \chi_{B_j}.$$

We assume that $\bigcup A_k = \bigcup B_j = \mathbb{R}^n$. So we've partitioned $\mathbb{R}^n$ two ways: into the A_k and into the B_j. We can mutually refine these partitions: the sets $A_k \cap B_j$ are all disjoint, and their union is $\mathbb{R}^n$. Then we can write

$$\int s\, d\lambda = \sum_{k=1}^{m} \alpha_k \lambda(A_k) = \sum_{j,k}^{n,m} \alpha_k \lambda(A_k \cap B_j)$$
$$\int t\, d\lambda = \sum_{k=1}^{n} \beta_j \lambda(B_k) = \sum_{j,k}^{n,m} \beta_j \lambda(A_k \cap B_j).$$

We claim that for each j, k, then $\alpha_k \lambda(A_k \cap B_j) \leq \beta_j \lambda(A_k \cap B_j)$. If $\lambda(A_k \cap B_j) = 0$, then this is trivially true. If $\lambda(A_j \cap B_j) > 0$, then there is some $x \in A_j \cap B_j$. Then $s(x) = \alpha_k$ and $t(x) = \beta_j$, but $s \leq t$, so $\alpha_k \leq \beta_j$, which proves our claim.

But then $\alpha_k \lambda(A_k \cap B_j) \leq \beta_j \lambda(A_k \cap B_j)$ for every j, k, and thus $\int s\, \lambda \leq \int t\, d\lambda$ by definition.

Now this by itself proves that our definition is well-posed. For suppose we have $s = t$ as just two different ways of representing the same underlying function. Then $s \leq t$ and also $t \leq s$, so $\int s\, d\lambda \leq \int t\, d\lambda$ and also $\int t\, d\lambda \leq \int s\, d\lambda$.

Given that the definition is well posed, items (2) and (3) are fairly clear. So we just have to prove (4). But by the logic from above, we have

$$\begin{aligned}
s + t &= \sum_{j,k}^{n,m} (\alpha_k + \beta_j) \chi_{A_k \cap B_j} \\
\int (s+t)\, d\lambda &= \sum_{j,k}^{n,m} (\alpha_k + \beta_j) \lambda(A_k \cap B_j) \\
&= \sum_{j,k}^{n,m} \alpha_k \lambda(A_k \cap B_j) + \sum_{j,k}^{n,m} \beta_j \lambda(A_k \cap B_j) \\
&= \int s\, d\lambda + \int t\, d\lambda.
\end{aligned}$$

□

4.3 The Integral of Non-Negative Functions

We can now integrate simple functions, which are the measure theory analogues of our finite Riemann sums from the Riemann integral. Now we want to extend this as far as possible.

The essential idea is this: we can compute the integrals of simple functions. Since every measurable function is the limit of simple functions, we can define the integral of a measurable function to be the limit of the integrals of the simple functions.

This definition is quite simple, and it's genuinely shocking how well it works.

Definition 4.16. Let $f : \mathbb{R}^n \to [0, \infty]$ be measurable. We define the *(Lebesgue) integral* of f to be
$$\int f\,d\lambda = \sup\left\{\int s\,d\lambda : s \leq f, s \in S\right\}.$$

Exercise 4.17. *Prove that our two definitions of the integral coincide if f is a measurable simple function. In particular, prove that if $f : \mathbb{R}^n \to [0, \infty]$ is a measurable simple function with $0 \leq \alpha_k \leq \infty$, then*
$$\int f\,d\lambda = \sum_{k=1}^{m} \alpha_k \lambda(A_k).$$

We now want to prove an analogue of proposition 4.15 for this more general integral. Most of the statements just follow immediately from the definition:

1. $\int f\,d\lambda$ is well defined (since every set has a supremum);
2. $0 \leq \int f\,d\lambda \leq \infty$
3. $\int cf\,d\lambda = c\int f\,d\lambda$
5. If $f \leq g$ then $\int f\,d\lambda \leq \int g\,d\lambda$.

However, it's highly non-trivial to prove that $\int (f+g)\,d\lambda = \inf f\,d\lambda + \int g\,d\lambda$.

One half of this is easy. We have that
$$\begin{aligned}
\int (f+g)\,d\lambda &= \sup\left\{\int s\,d\lambda : s \leq f+g\right\} \\
&= \sup\left\{\int (s+t)\,d\lambda : s+t \leq f+g\right\} \\
&= \sup\left\{\int s\,d\lambda + \int t\,d\lambda : s+t \leq f+g\right\}.
\end{aligned}$$

But while $s \leq f, t \leq g$ implies that $s+t \leq f+g$, the converse isn't true. So
$$\left\{\int s\,d\lambda + \int t\,d\lambda : s+t \leq f+g\right\} \supsetneq \left\{\int s\,d\lambda + \int t\,d\lambda : s \leq f, t \leq g\right\}$$
and thus
$$\int (f+g)\,d\lambda \geq \int f\,d\lambda + \int g\,d\lambda.$$
This is basically because our definition doesn't apply to *any* sequence of simple functions approaching f, but just sequences approaching from below. (This is similar to a definition of Riemann sum that only uses lower sums.)

There are various ways to prove the converse to this statement, many of which we could work out right now. One example is to show that the supremum over $s \leq f$ is the same as the infimum over $t \geq f$. But there are some major results that we want to prove anyway that will give us this result as a simple corollary.

In particular, one of the primary advantages of the Lebesgue integral formulation is that it allows us to interchange limits and integrals relatively freely.

Proposition 4.18 (Lebesgue Monotone Convergence Theorem). *Let $f_1, f_2, \dots : \mathbb{R}^n \to \overline{\mathbb{R}}$ be measurable such that*

$$0 \leq f_1 \leq f_2 \leq \dots.$$

Then

$$\lim_{k\to\infty} \int f_k \, d\lambda = \int \left(\lim_{k\to\infty} f_k \right) d\lambda.$$

4.4 Integrating non-non-negative functions

4.5 Integrating over sets other than $\mathbb{R}^n$

If X is any set, and $f : X \to \overline{\mathbb{R}}$, we define

$$\sum_{x\in X} f(x) = \sup \left\{ \sum_{x\in F} f(x) : F \subseteq X \text{ is finite} \right\}.$$

Think about why we need this definition; why this is case complicated if $X \neq \mathbb{N}$?

If $X = \mathbb{N}$, prove that $\sum_{n\in\mathbb{N}} f(x) = \sum_{k=1}^{\infty} f(k)$.

Let $(X, \mathcal{M}, \mu)$ be a measure space, and let $\overline{\mu}$ be the completion of μ. If $f : X \to \overline{\mathbb{R}}$ is μ-measurable, we know it must also be $\overline{\mu}$-measurable. Prove that $\int f \, d\overline{\mu} = \int f \, d\mu$.

(Conversely, if g is $\overline{\mathcal{M}}$-measurable, it need not be $\mathcal{M}$-measurable. But there is a $\mathcal{M}$-measurable function f such that $f(x) = g(x)$ almost everywhere, and then $\int g \, d\overline{\mu} = \int f \, d\mu$.)

Exercise 4.19. *Let $E \in \mathcal{M}$ and assume $\lambda(E) = 0$. Prove that every function defined on E is measurable, and that $\int_E f \, d\mu = 0$ for any f defined on E.*

4.6 Two-place functions

Differentiating under the integral sign blah blah

—

Let $l, m \in \mathbb{N}$, and set $n = l + m$. We can decompose $\mathbb{R}^n$ by writing $\mathbb{R}^n = \mathbb{R}^l \times \mathbb{R}^m$. If we have a point $z \in \mathbb{R}^n$, we will write $z = (x, y)$, where $x_i = z_i$ and $y_{i-l} = z_i$.

Then we can view functions on $\mathbb{R}^n$ as two-place functions on $\mathbb{R}^l$ and $\mathbb{R}^m$. We have $f(z) = f(x, y)$, and if we fix some specific $y_0 \in \mathbb{R}^m$ then we have a function $f_{y_0} : \mathbb{R}^l \to \mathbb{R}$ defined by $f_{y_0}(x) = f(x, y_0)$. We can similarly define $f_{x_0} : \mathbb{R}^m \to \mathbb{R}$ by $f_{x_0}(y) = f(x_0, y)$.

We call these functions f_y and f_x the *sections* of f determined by y or x. They're essentially the cross-sections we use to graph functions in multivariable calculus.

We'll find those sections especially interesting if f is the characteristic function of some $A \subseteq \mathbb{R}^n$. Then we have

$$f_y(x) = \begin{cases} 1 & (x, y) \in A \\ 0(x, y) \in A^C. & \end{cases}$$

Then f_y is the characteristic function of *some* subset of $\mathbb{R}^l$, and we write

$$A_y = \{x \in \mathbb{R}^l : (x, y) \in A\} = (\chi_A)_y^{-1}(\{1\}).$$

Thus by definition, we have $\chi_{A_y} = (\chi_A)_y$. We call the set A_y the *section* of A determined by y.

If we have a function $f : \mathbb{R}^n$, for any fixed $y \in \mathbb{R}^m$ the function $f_y : \mathbb{R}^l \to \mathbb{R}$ may or may not be integrable. If it is, we will write $F(y) = \int_{\mathbb{R}^l} f_y(x)\, d\lambda(x)$.

If f_y is integrable for almost every $y \in \mathbb{R}^m$, then this gives us a function $F : \mathbb{R}^m \to \overline{\mathbb{R}}$. That "almost" is important, since it's fairly hard to guarantee that f_y is integrable for *every* y.

Theorem 4.20 (Tonelli). *Suppose $f : \mathbb{R}^n \to [0, \infty]$ is measurable. Then for almost eveyr $y \in \mathbb{R}^m$, the section $f_y : \mathbb{R}^l \to [0, \infty]$ is measurable, and thus the function $F(y) = \int_{\mathbb{R}^l} f_y(x)\, dx$ is defined for almost every y.*

Further, this function $F : \mathbb{R}^m \to [0, \infty]$ is measurable, and

$$\int_{\mathbb{R}^m} F(y)\, dy = \int_{\mathbb{R}^n} f(z)\, dz.$$

We're not going to prove this, but I will give a quick outline. But first I want to take a minute to convince you that this is exactly the result we used in multivariable calculus. We have

$$\int_{\mathbb{R}^n} f(z)\, dz = \int_{\mathbb{R}^m} F(y)\, dy = \int_{\mathbb{R}^m} \int_{\mathbb{R}^l} f_y(x)\, dx\, dy = \int_{\mathbb{R}^m} \int_{\mathbb{R}^l} f(x, y)\, dx\, dy.$$

Thus the multivariable integral is the same as the iterated integral. And this is why in Math 212 we don't really spend much time thinking about how to do double integrals and two-variable Riemann sums directly; we can always just replace them with iterated one-variable integrals.

This *also* explains why we can interchange the order of integration whenever we want. There's not really any difference between x and y here except the order we write them in. So we could just as easily have

$$\int_{\mathbb{R}^n} f(z)\,dz = \int_{\mathbb{R}^l} F(x)\,dx = \int_{\mathbb{R}^l}\int_{\mathbb{R}^m} f_x(y)\,dy\,dx = \int_{\mathbb{R}^l}\int_{\mathbb{R}^m} f(x,y)\,dy\,dx.$$

Thus we get the same thing no matter which order we integrate in.

Sketch of proof. First, through some fairly tedious work, we show that the result holds for a characteristic function of a bounded set. We prove that

$$\int_{\mathbb{R}^m} \lambda(A_y)\,dy = \lambda(A).$$

That is, if we integrate the measures of each section of A, we get the total measure of A. (Recall this is how we computed volumes in calculus 2!)

After this we show that we can use the increasing function theorem as a lever. If Tonelli's theorem holds for each function in an increasing sequence of functions, it applies to their limit. That is, if $f_{j,y} \to f_y$, then we get a family of functions $F_j(y)$ that converge to $F(y)$ by the increasing convergence theorem. And then we can conclude that

$$\int_{\mathbb{R}^m} F(y)\,dy = \lim_{j\to\infty}\int_{\mathbb{R}^m} F_j(y)\,dy = \lim_{j\to\infty}\int_{\mathbb{R}^m} f_j(z)\,dz = \int_{\mathbb{R}^n} f(z)\,dz.$$

But since the result holds for characteristic functions of bounded sets, we can lever that up to give us any characteristic function, and then any simple function, and then any non-negative function.

$\square$

Tonelli's theorem isn't quite as strong as we'd like, though. It only applies to integrals of non-negative functions. Fortunately, as usual, we can move from non-negative functions to L^1 functions pretty easily.

Theorem 4.21 (Fubini)**.** *Suppose that $f \in L^1(\mathbb{R}^n)$. Then for almost every $y \in \mathbb{R}^m$, the function f_y is in $L^1(\mathbb{R}^l)$, and so the function*

$$F(y) = \int_{\mathbb{R}^l} f_y(x)\,dx$$

is well-defined. Further, this function is (finitely) integrable, and

$$\int_{\mathbb{R}^m} F(y)\,dy = \int_{\mathbb{R}^n} f(z)\,dz.$$

Proof. This proof works in more or less the obvious way. We define $f = f_+ - f_-$. Then for almost all $y \in \mathbb{R}^m$ the sections $f_{+,y}$ and $f_{-,y}$ are measurable, and by Tonelli's theorem we we can define the measurable functions

$$G(y) = \int_{\mathbb{R}^l} f_{-,y}\, dx \qquad H(y) = \int_{\mathbb{R}^l} f_{+,y}\, dx$$

and we get that

$$\int_{\mathbb{R}^m} G\, dy = \int_{\mathbb{R}^n} f_-\, dz \qquad \int_{\mathbb{R}^m} H\, dy = \int_{\mathbb{R}^n} f_+\, dz.$$

Since $f \in L^1$, we know both these integrals are finite, which means that G, H are finite almost everywhere. But if G is finite for almost every y, then $\int_{\mathbb{R}^l} f_{-,y}\, dx < \infty$ for almost every y. Similarly, $\int_{\mathbb{R}^l} f_{+,y}\, dx < \infty$ for almost every y. Thus both are finite almost always, and so $f_y \in L^1(\mathbb{R}^l)$ for almost every y.

Further, for almost every y, we can take $F(y) = H(y) - G(y)$ and thus F is integrable. Then we have

$$\int_{\mathbb{R}^m} F\, dy = \int_{\mathbb{R}^m} H\, dy - \int_{\mathbb{R}^m} G\, dy = \int_{\mathbb{R}^n} f_+\, dz - \int_{\mathbb{R}^n} f_-\, dz = \int_{\mathbb{R}^n} f\, dz.$$

□

Proposition 4.22. *If X is a measurable subset of $\mathbb{R}^l$ and Y is a measurable subset of $\mathbb{R}^m$, then $X \times Y$ is a measurable subset of $\mathbb{R}^n$, and $\lambda(X \times Y) = \lambda(X)\lambda(Y)$.*

Proof. We really only need to prove that $X \times Y$ is measurable, since the equality follows from Fubini.

We can write both X and Y as countable unions of sets of finite measure. We can take e.g. $X = \bigcup_{j=1}^\infty (X \cap B_j(x))$. But if $X = \bigcup_{j=1}^\infty X_j$ and $Y = \bigcup_{k=1}^\infty Y_k$ then we can write

$$X \times Y = \bigcup_{j,k=1}^\infty X_j \times Y_k.$$

So we just have to prove that $X_j \times Y_k$ is measurable when X_j, Y_k have finite measure.

Without loss of generality, suppose X, Y have finite measure. We can find $F_1 \subseteq X \subseteq G_1$ and $F_2 \subseteq X \subseteq G_2$ closed and open respectively, with $\lambda(G_1 \setminus F_1), \lambda(G_2 \setminus F_2) < \varepsilon$. Then $F_1 \times F_2 \subseteq X \times Y \subseteq G_1 \times G_2$ closed and open.

Now let's consider the set $G_1 \times G_2 \setminus F_1 \times F_2$. We want to show we can make this as small in measure as we want, because then we can squeeze our set $X \times Y$ between a closed set and an open set. But we can see that

$$G_1 \times G_2 \setminus F_1 \times F_2 \subseteq ((G_1 \setminus F_1) \times G_2) \cup (G_1 \times (G_2 \setminus F_2))$$

This containing set is open. We can estimate[1] its measure:

$$\begin{aligned}\lambda\left(((G_1 \setminus F_1) \times G_2) \cup (G_1 \times (G_2 \setminus F_2))\right) &\leq \lambda(G_1 \setminus F_1)\lambda(G_2) + \lambda(G_1)\lambda(G_2 \setminus F_2) \\ &\leq \varepsilon\lambda(G_2) + \varepsilon\lambda(G_1) \\ &< \varepsilon(\lambda(F_2) + \varepsilon) + \varepsilon(\lambda(F_1) + \varepsilon) \\ &\leq \varepsilon(\lambda(Y) + \lambda(X) + 2\varepsilon).\end{aligned}$$

But this is all we needed: we can make $\lambda(G_1 \times G_2 \setminus F_1 \times F_2)$ as small as we want. And this means that $X \times Y$ is squeezed between an open set and a closed set, and thus is Lebesgue measurable.

$\square$

[1] Analysts use the word "estimate" to mean "we're about to write down like twelve inequalities in a row