8. Integration theory

Integration theory is set up for a general measure just as for Lebesgue measure. We first define the integral on simple functions, then on bounded measurable functions of compact support, then on nonnegative measurable functions and finally on all measurable functions.

We have the following definitions:

- A measure μ is complete if every subset of a measure zero set is measurable (and necessarily has measure zero)
- A measure space $(X, \mathcal{M}, \mu)$ is σ-*finite* if X can be written as a countable union of measurable sets, each with finite measure.

For the remainder of this section, we'll assume our measure spaces are complete and σ-finite wherever convenient.

- Let $(X, \mathcal{M}, \mu)$ be a measure space. A function $f : X \to [-\infty, \infty]$ is measurable if the inverse image of every open set is measurable. (It is sufficient to require that the inverse image of $[-\infty, a)$ is measurable for each $a \in \mathbb{R}$.)
- A function $g : X \to \mathbb{C}$ is measurable if its real and imaginary parts are.
- A function is simple if it a finite linear combination of characteristic functions of measurable sets.

Then we have the properties:

- If f_n are measurable functions, then $\sup f_n$, $\inf f_n$, $\limsup f_n$, $\liminf f_n$ are measurable;
- If $f(x) = \lim f_n(x)$ then f is measurable;
- If f and g are measurable, then powers f^k are measurable, $k \geq 1$, and if they are both finite-valued then $f + g$ and fg are measurable.
- if $f = g$ a.e. with respect to μ and f is measurable then g is measurable.

We also have approximation properties:

- If f is nonnegative and measurable, then there exists a sequence of nonnegative simple functions ϕ_n such that $\phi_n \leq \phi_{n+1}$ and $\lim_n \phi_n(x) = f(x)$;

  Take ϕ_n to the largest simple function, with $\phi_n \leq f$, and values in $2^{-n}\{0, 1, 2, \ldots, 2^{2n}\}$. Explicitly, this is

$$\phi_n = \sum_{k=0}^{2^{2n}} k2^{-n} \chi_{f^{-1}[k2^{-n},(k+1)2^{-n})} + 2^n \chi_{f^{-1}[(2^{2n}+1)2^{-n},\infty)}.$$

- If f is measurable, then there exists a sequence of simple functions ϕ_n such that $|\phi_n| \leq |\phi_{n+1}|$ and $\lim_n \phi_n(x) = f(x)$.

  Write $f = f_+ + f_-$, and use the approximation above.

Defining the integral

As with the Lebesgue integral, we first define the integral on simple functions, then on bounded measurable functions defined on a set of finite measure, then on nonnegative measurable functions and finally on measurable functions. In each case, we check that the integral satisfies four conditions:

- Linearity: for $a, b \in \mathbb{R}$,

$$\int (af + bg) = a \int f + b \int g.$$

- Additivity: if E and F are disjoint measurable subsets of X then

$$\int_E f + \int_F f = \int_{E \cup F} f.$$

- Monotonicity: if $f \leq g$, then

$$\int f \leq \int g.$$

- Triangle inequality:

$$\Big| \int f \Big| \leq \int |f|.$$

Step 1. The integral of a simple function $f = \sum_i a_i 1_{E_i}$ is defined to be

$$\int f = \sum_i a_i \mu(E_i).$$

It is crucial that this formula is independent of the representation of f as a simple function (this uses the finite additivity of μ). We also define

$$\int_E f = \int 1_E f.$$

Then we can check that the integral for simple functions satisfies all four conditions above.

Step 2. Bounded functions supported on a set of finite measure E. To define the integral here we need Egorov's theorem. Egorov's theorem essentially says that a sequence of measurable functions converging pointwise actually converges *uniformly*, away from a set of arbitrarily small measure.

Theorem 8.1 (Egorov)**.** *Let f_n be measurable functions converging pointwise on E to f. Then for any $\epsilon > 0$ there is a subset A_ϵ of E of measure at least $\mu(E) - \epsilon$ such that $f_n \to f$ uniformly on A_ϵ.*

Proof: Consider sets

$$C_M^\eta = \{x \mid |f_n(x) - f(x)| < \eta \text{ for all } n \geq M\}.$$

For any fixed $\eta > 0$, every x is in some C_M^η, and hence $\cup_M C_M^\eta = E$. Since C_M^η is an increasing family of sets, $\mu(C_M^\eta) \to \mu(E)$, by countable additivity.

Now we each n, we choose M_n so that with $B_n = (C_{M_n}^{2^{-n}})^c$, the measure of B_n is less that $2^{-n}\epsilon$. Note that we now have $|f_k(x) - f(x)| < 2^{-n}$ for $x \notin B_n$ and $k \geq M_n$.

Then let $B = \cup B_n$ and $A_\epsilon = E \setminus B$, and we see that $f_n \to f$ uniformly on A_ϵ since for every n there exists

M_n such that $|f_k(x) - f(x)| < 2^{-n}$, for $x \in A_\epsilon$ and $k \geq M_n$. Moreover, $\mu(B) \leq \sum_n 2^{-n}\epsilon = \epsilon$. $\square$

Using this theorem, we see that if f is a bounded function on E, with $|f|$ bounded by M say, and if ϕ_n are simple functions converging pointwise to f, with each $|\phi_n| \leq M$, then the integrals $\int \phi_n$ converge. It is not hard to check that the limit of the integrals is independent of the choice of sequence ϕ_n (if it did depend on the choice, we could splice together two sequences with integrals converging to different limits, and obtain a sequence whose integrals did not converge). The integral of f is defined to be

$$\int f = \lim_n \int_E \phi_n.$$

One then checks that the integral so defined has the properties of linearity, additivity, monotonicity and the triangle inequality.

In fact, repeating the argument above using Egorov's theorem proves the Bounded Convergence Theorem (a baby version of the Dominated Convergence Theorem):

Theorem 8.2. *Suppose that f_n is a sequence of measurable functions that are all supported on a fixed set E of finite measure, and uniformly bounded by M. If $f_n \to f$ a.e., then*

$$\int f_n \to \int f.$$

Step 3. Nonnegative measurable functions. Notice that we trivially have the property that, if f is bounded and supported on a set of finite measure E, then we have

$$\int f = \sup \int g$$

where the sup is over all bounded measurable g supported on E with $g \leq f$. This follows from monotonicity, showing that $\int g \leq \int f$ for any such g, together with the fact that we may take $g = f$.

We use this to define the integral for nonnegative measurable functions. That is, if f is nonnegative and measurable, we define $\int f$ to be the sup of $\int g$ over all bounded functions g such that $0 \leq g \leq f$ and g is supported on a set of finite measure. Notice that the value of the integral might be $+\infty$ (we use the convention that the sup of a set that is not bounded above is $+\infty$).

Again you should check that the integral so defined has the properties of linearity, additivity, monotonicity and the triangle inequality. These properties are all straightforward except for the linearity property.

It is straightforward to check that

$$\int cf = c \int f, \quad c \geq 0$$

for nonnegative measurable functions, so it suffices to check that

$$\int (f_1 + f_2) = \int f_1 + \int f_2$$

for such functions. Given g_i such that $0 \le g_i \le f_i$ and g_i is bounded and supported on a set of finite measure, the function $g = g_1 + g_2$ has the same property with respect to $f_1 + f_2$. Then

$$\int g_1 + \int g_2 = \int g \le \int (f_1 + f_2),$$

and taking the sup over all g_1, g_2 shows that

$$\int f_1 + \int f_2 \le \int (f_1 + f_2).$$

For the reverse inequality, suppose that $0 \le g \le f_1 + f_2$ and g is bounded by M and supported on a set of finite measure E. Define g_i by

$$g_i(x) = \begin{cases} 0 \text{ if } x \notin E \\ f_i(x) \text{ if } f_i(x) \le M \\ M \text{ if } f_i(x) > M. \end{cases}$$

Then $g_1 + g_2 \ge g$ pointwise, giving

$$\int g \le \int g_1 + \int g_2 \le \int f_1 + \int f_2$$

and taking the sup over all such g gives the opposite inequality.

Now in this setting it is no longer the case that $f_n \to f$ a.e. implies that $\int f_n \to \int f$. You have already seen examples, e.g. $f = 0$ and $f_n = n1_{[0,1/n]}$ on the real line. However, an inequality holds:

Lemma 8.3 (Fatou's lemma)**.** *Suppose that $f \geq 0$, and that the sequence of functions f_n is nonnegative and converges to f a.e. Then*

$$\liminf_n \int f_n \geq \int f.$$

• Mass in the integral can 'bubble off' and escape in the limit. However, since f_n are nonnegative, only positive amounts of mass can be lost.

Proof: Take any $0 \leq g \leq f$ which is bounded and supported on a set of finite measure, and define $g_n = \min(g, f_n)$. Then by the bounded convergence theorem, $\int g_n \to \int g$. Since $g_n \leq f_n$, we have

$$\lim \int g_n = \liminf \int g_n \leq \liminf \int f_n.$$

Hence $\int g$ is less than $\liminf \int f_n$, and taking the supremum over g gives the result. □

A corollary is

Theorem 8.4 (Monotone convergence theorem)**.** *Let* $f \geq 0$*, and let* f_n *be an increasing sequence of measurable functions converging to* f*. Then*

$$\lim \int f_n = \int f.$$

Step 4. General measurable functions. In this case, we cannot integrate all such functions; we restrict to the class of **integrable** functions f, for which

$$\int |f| < \infty.$$

To define the integral for real functions, we write $f = g_1 - g_2$, where $g_i \geq 0$. This can be done, for example, by taking $g_1 = \max(f, 0)$ and $g_2 = -\min(f, 0)$. Then we define

$$\int f = \int g_1 - \int g_2.$$

The integrals on the RHS are defined in Step 3. One needs to check that this is independent of the representation $f = g_1 - g_2$. But if also $f = h_1 - h_2$, where $h_i \geq 0$, then $g_1 + h_2 = g_2 + h_1$. By linearity of the integral in Step 3, we find that

$$\int g_1 + \int h_2 = \int g_2 + \int h_1,$$

which shows the value of $\int g_1 - \int g_2$ is independent of the choice of g_i.

The integral of an integrable complex-valued function $f = g + ih$, where g, h are real, necessarily integrable functions, is defined to be $\int g + i \int h$.

We can then check that the integral on integrable functions is linear, additive, monotonic and satisfies the triangle inequality.

The most important convergence theorem in integration theory is the dominated convergence theorem. To prove it we start with a lemma:

Lemma 8.5. *Let g be an integrable function.*

(i) Given $\epsilon > 0$, there exists a set E of finite measure such that

$$\int_{E^c} |g| \leq \epsilon.$$

(ii) Given $\epsilon > 0$, there exists $M > 0$ such that, with $A = \{x \mid |g(x)| > M\}$, we have

$$\int_A |g| \leq \epsilon.$$

To prove (i), define

$$E_n = \{x \mid |g(x)| \geq 1/n\}.$$

Then E_n has finite measure, since $\int |g| \geq \mu(E_n)/n$. Let $g_n = |g| 1_{E_n}$. Then $g_n \to |g|$ monotonically, so by the monotone convergence theorem,

$$\int g_n \to \int |g|.$$

Thus, by taking n large enough, we have

$$\int_{E_n} |g| = \int g_n \geq \int |g| - \epsilon.$$

To prove (ii), we define A_n to be the set where $|g| \geq n$, and let $g_n = |g| 1_{A_n}$. Then, since the measure of the set where

$|g| = \infty$ is zero, we have $|g| - g_n \to |g|$ a.e. Therefore, by the monotone convergence theorem,

$$\int g_n \to 0.$$

Thus taking n sufficiently large, we have $\int g_n \leq \epsilon$. and thus $\int_{A_n} |g| \leq \epsilon$.

Using this it is quite straightforward to prove the dominated convergence theorem:

Theorem 8.6. *Suppose that the sequence f_n of measurable functions converges to f a.e. , and $|f_n| \leq g$ for some nonnegative integrable function g. Then*

$$\int |f_n - f| \to 0, \text{ and hence } \int f_n \to \int f.$$

Proof: Using the previous lemma, choose a set E of finite measure such that $g \leq M$ on E, and such that

$$\int_{E_c} g < \epsilon.$$

Then,

$$\int |f_n - f| = \int_E |f_n - f| + \int_{E^c} |f_n - f|.$$

By the bounded convergence theorem,

$$\int_E |f_n - f| \to 0.$$

On E^c, we estimate $|f_n - f| \leq 2g$, and see that

$$\int_{E^c} |f_n - f| \leq \int_{E^c} 2g \leq 2\epsilon.$$

Thus, $\limsup \int |f_n - f| \leq 2\epsilon$, and since this is true for all $\epsilon > 0$ we obtain the result. □

8.1. **Product measures and Fubini's theorem.**

Let $X = X_1 \times X_2$ and suppose that $(X_1, \mathcal{M}_1, \mu_1)$ and $(X_2, \mathcal{M}_2, \mu_2)$ are two measure spaces. Can we define a measure μ on X with the property that $\mu(A \times B) = \mu_1(A)\mu_2(B)$ for all $A \in \mathcal{M}_1$ and $B \in \mathcal{M}_2$?

We can do this by defining a premeasure on an algebra of subsets of X, namely the algebra $\mathcal{A}$ consisting of finite unions of disjoint rectangles, which by definition are sets of the form $A \times B$, where $A \in \mathcal{M}_1$ and $B \in \mathcal{M}_2$. This is an algebra: the complement of $A \times B$ is $(A \times B^c) \cup (A^c \times B) \cup (A^c \times B^c)$, while the union of two rectangles is the disjoint union of at most 6 rectangles.

We define our premeasure on the disjoint union $\cup_j A_j \times B_j$ by setting

$$\mu_0(\cup_j A_j \times B_j) = \sum_j \mu_1(A_j)\mu_2(B_j).$$

We have to check that this is independent of the representation as a disjoint union of rectangles and that it satisfies countable additivity: whenever $C \in \mathcal{A}$ is the countable disjoint union of rectangles $A_j \times B_j$, then

$$\mu_0(C) = \sum_j \mu_0(A_j \times B_j).$$

It is enough to do this for rectangles $A \times B$. We have for every $x_1 \in X_1$

$$1_A(x_1)\mu_2(B) = \sum_j 1_{A_j}(x_1)\mu_2(B_j)$$

using countable additivity of μ_2. Then integrating in x_1 and using the monotone convergence theorem, we obtain

$$\mu_0(A \times B) = \sum_j \mu_0(A_j \times B_j).$$

The premeasure μ_0 generates a measure μ on the σ-algebra $\mathcal{M}$ generated by $\mathcal{A}$. This defines the product measure $(X, \mathcal{M}, \mu)$.

We shall now prove a Fubini theorem for this product measure. First, we prove a special case. For any set $E \subset X$ we define E^{x_2} to be

$$\{x_1 \in X_1 \mid (x_1, x_2) \in E\},$$

i.e. the slice through E with fixed second coordinate x_2.

Proposition 8.7. *Assume that μ_1 and μ_2 are both complete and σ-finite. Suppose that $E \subset X$ is measurable. Then for almost every $x_2 \in X_2$, E^{x_2} is measurable w.r.t. μ_1, and*

$$\int_{X_2} \mu_1(E^{x_2}) d\mu_2 = \mu(E).$$

Moreover, if $E \in \mathcal{A}_{\sigma\delta}$, then the same is true with 'almost every' replaced by 'every'.

Proof: We first prove the second statement. Thus, assume that $E \in \mathcal{A}_{\sigma\delta}$. In fact, we first suppose that $E \in \mathcal{A}_\sigma$, i.e. is a countable union of rectangles. Without loss of generality, these rectangles $E_j = A_j \times B_j$ are disjoint. The conclusion is obvious for a single rectangle. Noting that $E_j^{x_2}$ are disjoint measurable sets in X_1, countable additivity of μ_1 and the monotone convergence theorem show that the LHS is equal to

$$\sum_j \int \mu_1(E_j^{x_2}) d\mu_2 = \sum_j \mu_1(A_j)\mu_2(B_j),$$

which is equal to the RHS by countable additivity of μ.

Now for any set $E \in \mathcal{A}_{\sigma\delta}$ with $\mu(E)$ finite and also each slice has $\mu_1(E^{x_2})$ finite, we can write it as a countable intersection of $\mathcal{A}_\sigma$ sets E_j, which we may are assume are decreasing, and then $\mu(E_j) \to \mu(E)$. Then the sets $E_j^{x_2}$ are a decreasing family with intersection E^{x_2}, which is therefore measurable; moreover, if we define $f_j(x_2) = \mu_1(E_j^{x_2})$, and $f(x_2) = \mu_1(E^{x_2})$, then f_j is a nonincreasing family of finite measurable functions with $f_j \to f$. (We needed to assume that each x_2 slice had finite measure here; Stein & Shakarchi seems to make a mistake here.) Hence f is measurable, and by the monotone convergence theorem,

$$\lim_j \int_{X_2} f_j(x_2) d\mu_2(x_2) = \int_{X_2} f(x_2) d\mu_2(x_2).$$

However, the LHS is $\lim_j \mu(E_j)$ by our first result, which converges to $\mu(E)$ as we saw above. This establishes the result for $E \in \mathcal{A}_{\sigma\delta}$ with $\mu(E)$ finite and each x_2 slice finite. To treat the general case, we take increasing sequences F_j, G_j in X_1, X_2 respectively, of sets of finite measure whose union is X_i, and define $E_j = E \cap (F_j \times G_j)$. We apply the result to each E_j, and use the monotone convergence theorem on the LHS and countable additivity on the RHS to deduce the result.

To prove the result for general E, we first show for sets E of measure zero. Then there exists $F \in \mathcal{A}_{\sigma\delta}$ with $E \subset F$ and $\mu(F) = 0$. The result already proved then shows that $\mu_1(F^{x_2})$ is zero for a.e. x_2, hence by completeness of μ_1, E^{x_2} is measurable for a.e. x_2 (with measure zero). This establishes the result for E with measure zero. In general, $E \subset G$ where $G \in \mathcal{A}_{\sigma\delta}$ and $Z = G \setminus E$ has measure zero. Combining the results proved for Z and for G, we obtain the result for E. $\square$

As a corollary, we see that a set E of measure zero in X has slices E^{x_2} which have μ_1 measure zero except on a set of μ_2 measure zero.

Now we can prove

Theorem 8.8 (Fubini-Tonelli). *Let X be as above. Suppose that $f(x_1, x_2)$ is a nonnegative measurable function on X. Then*

(i) the slice function f^{x_2} defined by $f^{x_2}(x_1) = f(x_1, x_2)$ is measurable for a.e. x_2;

(ii) $x_2 \mapsto \int_{X_1} f^{x_2}(x_1) d\mu_1(x_1)$ is measurable on X_2;

(iii)

$$\int_{X_2} \Big(\int_{X_1} f^{x_2}(x_1) d\mu_1 \Big) d\mu_2 = \int_{X_1 \times X_2} f d\mu. \tag{8.1}$$

Moreover, if f is integrable, rather than nonnegative, on X, then the conclusions are

(i) the slice function f^{x_2} defined by $f^{x_2}(x_1) = f(x_1, x_2)$ is integrable (in particular, measurable) for a.e. x_2;

(ii) $x_2 \mapsto \int_{X_1} f^{x_2}(x_1) d\mu_1(x_1)$ is integrable on X_2;

(iii) (**??**) *holds.*

Proof: First, suppose that $f = 1_E$ for some μ-measurable set $E \subset X$. Then the result is precisely given by the previous Proposition. Therefore the result holds for simple functions f by linearity of the integral. Now suppose that f is nonnegative. Take an increasing sequence of simple functions f_n converging to f. Then $f_n^{x_2}$ converges monotonically to f^{x_2}, so by the MCT, we have

$$\int_{X_1} f_n^{x_2}(x_1) d\mu_1 \to \int_{X_1} f^{x_2} d\mu_1(x_1)$$

for every value of x_2. Applying MCT again, we find that

$$\int_{X_2} \Big(\int_{X_1} f_n^{x_2}(x_1) d\mu_1 \Big) d\mu_2$$

converges to

$$\int_{X_2} \Big(\int_{X_1} f^{x_2}(x_1) d\mu_1 \Big) d\mu_2.$$

On the RHS, applying MCT once again shows that

$$\int_{X_1 \times X_2} f_n d\mu \to \int_{X_1 \times X_2} f d\mu.$$

Since the theorem holds for each f_n, this shows that it also holds for f.

The second conclusion follows by applying the first to the positive and negative parts of f separately (or the positive real, negative real, positive imaginary and negative imaginary parts separately if f is complex-valued). $\square$

Example. Integration in polar coordinates. Let $(S^{n-1}, \mathcal{M}_{S^{n-1}}, d\theta_n)$ denote the standard $(n-1)$-sphere with its usual σ-algebra (the Lebesgue measurable sets in any smooth coordinate chart) and measure, and let $(\mathbb{R}_{\geq}, \mathcal{M}_{\mathbb{R}_{\geq}}, r^{n-1}dr)$ denote the half-line with the Lebesgue measurable sets and the measure r^{n-1} times Lebesgue measure dr. We can show that the product measure space is naturally identified with Lebesgue measure on $\mathbb{R}^n$. Then Fubini-Tonelli justifies integration in polar coordinates.

8.2. **Pushforward of measures.**

Suppose that $(X, \mathcal{M}, \mu)$ is a measure space and that $(Y, \mathcal{C})$ is a measurable space (that is, a set with a σ-algebra of sets). We say that $F : X \to Y$ is measurable if $F^{-1}(E) \in \mathcal{M}$ whenever $E \in \mathcal{C}$.

In this situation, we can define an induced measure on Y, the pushforward measure $F\mu$, as follows:

$$(F\mu)(E) = \mu(F^{-1}(E)).$$

It is straightforward to check that this is countably additive on $\mathcal{C}$.

Proposition 8.9. *Let $f : Y \to \mathbb{R}$ be measurable. Then,*

$$\int_X (f \circ F) d\mu = \int_Y f d(F\mu)$$

in the sense that when the RHS exists, so does the LHS and then they are equal.

Proof: This is true for $f = 1_E$, the characteristic function of $E \in \mathcal{C}$, by definition of $F\mu$. By linearity it is true for all simple functions. We then show that it is true for nonnegative functions using the MCT, as in the Fubini proof above, and finally, for all integrable functions. $\square$

A very important case is the following: let $R \subset \mathbb{R}^n$ be a closed rectangle, and let $F : R \to \mathbb{R}^n$ be a C^1 diffeomorphism onto its image, i.e. a C^1 function with a C^1 inverse $G : F(R) \to R$.

Theorem 8.10. *The pushforward of Lebesgue measure $d\lambda$ under F is equal to $d\lambda |\det DF|^{-1}$, or equivalently, the pushforward of $|\det DF| \cdot d\lambda$ is equal to $d\lambda$. Consequently, we have the change of variable formula*

$$\int_R (f \circ F)(x) |\det DF(x)| d\lambda(x) = \int_{F(R)} f d\lambda. \tag{8.2}$$

The proof is given in the notes for interest, but will not be covered in class.

The model situation is when F is an invertible linear map:

Proposition 8.11. *Suppose that F is an invertible linear map. Then $F(d\lambda) = |\det F|^{-1} d\lambda$; that is, the image of any measurable set E under F has measure $|\det F| d\lambda(E)$.*

Proof: Any square matrix can be written as the product of a finite number of the following 'elementary' matrices: diagonal matrices; permutation matrices; and matrices of the form

$$M_c = \begin{pmatrix} 1 & 0 & 0 & \dots \\ c & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

To see this, note that multiplying a matrix by the one above on the left has the effect of adding c times row 1 to row 2. By combining with permutation matrices, we can add c times any row to any other. If we multiply on the right instead, we can do column operations instead. Hence we can apply row and column operations to any matrix, and eventually reduce it to a diagonal matrix.

So to prove the theorem, it suffices to show for the elementary matrices, since both the determinant and the volume-magnification factor are multiplicative. The result is obvious for diagonal and permutation matrices since they map rectangles to rectangles with the correct measure ratio. So it is enough to prove for M_c above. To do this it suffices to treat dimension 2.

So consider the effect of M_c on a rectangle $R = [0, A] \times [0, B]$, where $A, B > 0$. This gets mapped by M_c to a parallelogram P with sides from $(0, 0)$ to (A, cA) and from $(0, 0)$ to $(0, B)$. But P can be covered by n rectangles

$$\left[A\frac{j}{n}, A\frac{j+1}{n}\right] \times \left[cA\frac{j}{n}, cA\frac{j+1}{n} + B\right],$$

and contains the n rectangles

$$\left[A\frac{j}{n}, A\frac{j+1}{n}\right] \times \left[cA\frac{j+1}{n}, cA\frac{j}{n} + B\right].$$

Thus the measure of P is estimated by

$$n \cdot \frac{A}{n}\left(B - \frac{cA}{n}\right) \leq \mu(P) \leq n \cdot \frac{A}{n}\left(B + \frac{cA}{n}\right),$$

and hence $\mu(P) = AB = \mu(R)$. □

For the next lemma, let Q be a closed rectangle, centred at the origin, such that the ratio between the longest and shortest side is ≤ 2.

Lemma 8.12. *Let F be a C^1 map from Q to $\mathbb{R}^n$, satisfying $F(0) = 0$ and*

$$\|DF(x) - A\| \leq \epsilon,$$

for sufficiently small ϵ and fixed invertible linear map $A : \mathbb{R}^n \to \mathbb{R}^n$. (Here we use the operator norm on matrices:

$$\|A\| = \sup_{|x|=1} |Ax|. \;)$$

Let $\epsilon' = 2\sqrt{n}\|A^{-1}\|\,\epsilon$. Then, the image $F(Q)$ contains $(1-\epsilon')AQ$, and is contained in $(1+\epsilon')AQ$.

Proof: We first prove this with A equal to the identity. In that case, we compute for $x \in Q$

$$\begin{aligned} F(x) - x &= \int_0^1 \frac{d}{dt}(F(tx) - tx)\,dt \\ &= \int_0^1 (DF_{tx}(x) - x)\,dt \text{ , so} \\ |F(x) - x| &\leq \int_0^1 \epsilon|x|\,dt = \epsilon|x|. \end{aligned}$$

Now let c_1 be half the length of the shortest side of Q and let $c_2 = \max_{x\in Q}|x|$. By the condition on Q we have $2\sqrt{n}c_1 > c_2$. Therefore, $F(x)$ is in the ϵc_2 enlargement of Q. This is contained in $(1+2\sqrt{n}\epsilon)Q$ since the sides of this rectangle are at least $2\sqrt{n}c_1\epsilon \geq c_2\epsilon$ from the corresponding sides of Q. Hence, $F(x) \in (1+\epsilon')Q$.

To show that $F(Q)$ covers $(1-\epsilon')Q$, we observe that $F(\partial Q)$ is disjoint from $(1-\epsilon')Q$ by the argument above. For sufficiently small ϵ, the condition on DF ensures that this is invertible. Then, by the inverse function theorem, F is locally a diffeomorphism, and therefore sends small open balls to open sets. It follows that F maps the interior of Q to interior points of $F(Q)$, and therefore the boundary of $F(Q)$ is contained in $F(\partial Q)$.

Assume for a contradiction that there exists $x_0 \in (1-\epsilon')Q$ not in the image of F. Consider the line segment $tx_0, t \in [0,1]$. Then this goes from $0 = F(0) \in F(Q)$ to $x_0 \notin F(Q)$ without intersecting $\partial F(Q)$, which is a contradiction.

Now we treat the case of general A. Let $F' = A^{-1} \circ F$. Then $\|DF' - \mathrm{Id}\,\| \leq \epsilon\|A^{-1}\|$, since $\|B_1B_2\| \leq \|B_1\|\|B_2\|$. From what we just proved, we get

$$(1-\epsilon'')Q \subset F'(Q) \subset (1+\epsilon'')Q$$

with $\epsilon'' = 2\sqrt{n}\epsilon\|A^{-1}\|$. Applying A on the left we find

$$(1-\epsilon'')AQ \subset F(Q) \subset (1+\epsilon'')AQ,$$

as required. □

Lemma 8.13. *Suppose that $\|DF(x) - A\| < \epsilon$ for some $0 < \epsilon < \|A\|/2$. Then there exists a C depending only on dimension n such that*

$$|\det DF(x) - \det A| < C\epsilon\|A\|^{n-1}.$$

Proof: Suppose that DF and A differed only in the first row. Then we could expand the determinant along the first row and find that

$$\det DF - \det A = \sum_{j=1}^{n} ((DF)_{1j} - A_{1j})p_j(A)$$

where $p_j(A)$ is a polynomial of degree $n-1$ in the entries of A from rows $2 \ldots n$. This immediately gives the estimate, since $|p_j(A)| \leq C\|A\|^{n-1}$. In general, we can let A_j be the matrix with the first j rows from DF and the remaining rows from A. Apply the estimate above for $\det A_j - \det A_{j-1}$, and use the fact that $\|A_j\| \leq \|A\| + \epsilon \leq 2\|A\|$. □

Finally we prove the theorem. Since DF is continuous on the compact set R, it is uniformly continuous. Therefore, there exists δ such that $\|DF_x - DF_y\| < \epsilon$ whenever $|x - y| < \delta$. Moreover, since both DF and $D(F^{-1})$ are continuous, there are bounds

$$\|DF_x\| \leq M_1, \quad \|D(F^{-1})_{F(x)}\| \leq M_{-1}, \quad x \in R.$$

Choose a decomposition of R into a finite number of disjoint rectangles Q_i such that the longest side is at most twice the shortest side, and such that the diameter of Q_i is less than δ. Let c_i be the centre of Q_i, and let $A_i = DF_{c_i}$. Then on each Q_i we have, by Proposition **??** and Lemma **??**, with $\epsilon' = 2\sqrt{n}M_{-1}\epsilon$,

$$(1 - \epsilon')^n |\det A_i| \lambda(Q_i) \leq \lambda(F(Q_i)) \leq (1 + \epsilon')^n |\det A_i| \lambda(Q_i).$$

Now define $d_i = \lambda(F(Q_i))/\lambda(Q_i)$. Then

$$(1 - \epsilon')^n |\det A_i| \leq d_i \leq (1 + \epsilon')^n |\det A_i|.$$

Now let x_i be any point in Q_i. Using Lemma **??**, we have

$$(8.3) \qquad \begin{aligned}(1-\epsilon')^n\big(|\det DF_{x_i}| - C\epsilon M_1^{n-1}\big) &\leq d_i \\ &\leq (1+\epsilon')^n\big(|\det DF_{x_i}| + C\epsilon M_1^{n-1}\big).\end{aligned}$$

Now for $\epsilon = 2^{-n}$ we choose a decomposition Q_i^n as above. Define the function g_n to be equal to $d_i = d_i^n$ on each Q_i^n as above. By construction, we have

$$\int_R g_n d\lambda = \lambda(F(R)),$$

since $F(R)$ is equal to the disjoint union of the $F(Q_i^n)$. On the other hand, (**??**) shows that g_n is uniformly bounded and converges pointwise to $|\det DF|$. The DCT shows that

$$\lim_n \int_R g_n d\lambda = \int_R |\det DF_x| d\lambda.$$

Therefore,

$$d\lambda(F(R)) = \int_R |\det DF_x| d\lambda.$$

Thus, Lebesgue measure $d\lambda$ and the pushforward $F(|\det DF|d\lambda)$ agree on $F(R)$, and therefore on $F(R')$ for any subrectangle $R' \subset R$. Since any open set O in R is a countable union of rectangles, the same is true for $F(O)$ for all open $O \subset R$. Since F is assumed to be a diffeomorphism, $F(O)$ runs over all open sets in $F(R)$. Now observe that the family of sets for which the two measures agree forms a σ-algebra, so they must agree on the σ-algebra generated by open sets, i.e. the Borel sets. Moreover, it is not hard to see that the pushforward of a set of measure zero has measure zero (this is true for all Lipschitz F), so the two measures agree on all sets which differ from a Borel set by a measure zero set, i.e. all Lebesgue measurable sets.

It follows that the pushforward of the measure $|\det DF_x| d\lambda$ is $d\lambda$, as claimed.

8.3. **The Lebesgue-Stieltjes integral.**

The Lebesgue-Stieltjes integral gives a meaning to the expression

$$\int_a^b g(x)dF(x)$$

where F is an non-decreasing function. Roughly, this is supposed to be the limit of expressions of the form

$$\sum_{i=1}^{n} g(t_i)(F(t_i)-F(t_{i-1})), \quad a = t_0 < t_1 < \cdots < t_n = b.$$

Let us say that an non-decreasing function $F(x)$ is normalized if it is right-continuous: that is, that $F(x) = \lim_{y\downarrow x} F(y)$ for all x. Any non-decreasing function can be normalized by changing its values on a countable set of points.

Theorem 8.14. *Let F be a non-decreasing, normalized function on $\mathbb{R}$. Then there is a unique Borel measure μ, often denoted dF, such that $\mu((a, b]) = F(b) - F(a)$ for all $a < b$.*

- If F is C^1 then the measure μ is given by $F'(x)dx$ by the Fundamental Theorem of Calculus.

Before embarking on the proof, consider a simple example: suppose $F(x) = 2x$. Then $dF = 2d\lambda$. On the other hand, the pushforward $F(d\lambda)$ is equal to $\lambda/2$. It seems that what we are looking at here is a sort of 'inverse' to the pushforward operation. And indeed, we can construct μ as a pushforward to a sort of inverse function to F. Of course, in general F will not have an inverse! (since it need be neither one-to-one nor onto). However, we can use the order on $\mathbb{R}$ to define a sort of generalized inverse.

Proof: Let

$$G(y) = \inf\{x \mid F(x) > y\}.$$

Since F is nondecreasing, G is well defined at least on the interval between $\lim_{x\to-\infty} F(x)$ and $\lim_{x\to+\infty} F(x)$, and it is nondecreasing there. It is not hard to see that if F is continuous then G is indeed the inverse to F.

Now consider the measure $\nu = G\lambda$. Let $a < b$ be real numbers. What is $\nu((a,b])$? By definition, it is the Lebesgue measure of the set

$$S = \{y \mid G(y) \in (a,b]\}.$$

This set is

$$\Big\{y \mid \inf\{x \mid F(x) > y\} \in (a,b]\Big\}.$$

The set is an interval by the nondecreasing property of F, so we just have to find its endpoints.

First suppose that $F(b) = F(a)$. Then

- if $y > F(b)$ then the inf is $> b$ (using right cty);
- if $y = F(b)$ then the inf is $\geq b$;
- if $y < F(b)$ then the inf is $\leq a$.

It follows that S is either $\emptyset$ or the singleton set $\{F(b)\}$. In either case, $d\lambda(S) = 0 = F(b) - F(a)$.

Now suppose that $F(a) < F(b)$. Then

- if $y > F(b)$ then the inf is $> b$ (using right cty);
- if $y = F(b)$ then the inf is $\geq b$;
- if $F(a) < y < F(b)$ then the inf is $\leq b$ and $> a$ using right continuity to get "$>$";
- if $y < F(a)$ then the inf is $\leq a$.

Therefore, $(F(a), F(b)) \subset S \subset [F(a), F(b)]$. In any case, we have verified $\lambda(S) = F(b) - F(a)$. Thus ν has the required property. The proof of uniqueness follows standard lines. $\square$

These sorts of measures turn up in the spectral theorem for bounded (non-compact) self-adjoint operators on Hilbert space. This is phrased in terms of a **spectral resolution** on H, i.e. a family of orthogonal projection operators $E(x)$, for $x \in \mathbb{R}$, such that

(i) $E(x)$ is nondecreasing in the sense that the range of $E(x)$ is contained in the range of $E(y)$ if $x \leq y$;

(ii) $E(x)$ is right continuous in the sense that $\lim_{y \downarrow x} E(y)f = E(x)f$ for all $f \in H$;

(iii) There exists an interval $[a, b]$ such that $E(x) = 0$ if $x < a$ and $E(x) = \mathrm{Id}$ if $x > b$.

Then, for all $f, g \in H$, the function $(E(x)f, g)$ is nondecreasing and right continuous.

The spectral theorem for a bounded self-adjoint operator T says there is a spectral resolution $E(x)$ s.t.

$$T = \int_{\mathbb{R}} x \, dE(x)$$

in the sense that

$$(Tf, g) = \int_{\mathbb{R}} x \, d(E(x)f, g)$$

as a Lebesgue-Stieltjes integral.