Major theorems in analysis

Shrivathsa Pandelu

October 29, 2020

Contents

1 Inverse Function Theorem

Notation: Vectors in $\mathbb{R}^n$ are column vectors.

Theorem 1. *Let A be open in $\mathbb{R}^n$; let $f\colon A \to \mathbb{R}^n$ be of class $\mathcal{C}^r$. If $Df(x)$ is non-singular at the point a of A, there is a neighbourhood U of the point a such that f carries U in a one-to-one fashion onto an open set V of $\mathbb{R}^n$ and the inverse is of class $\mathcal{C}^r$.*

Lemma 1. *Let A be open in $\mathbb{R}^n$; let $f\colon A \to \mathbb{R}^n$ be of class $\mathcal{C}^r$. If $Df(a)$ is non-singular; then there exists an $\alpha > 0$ such that the inequality*

$$|f(x_0) - f(x_1)| \geq \alpha |x_0 - x_1|$$

holds for all x_0, x_1 in some open cube $C(a;\epsilon)$ centred at a. If follows that f is one-to-one on this open cube.

Proof. Let $E = Df(a)$. Consider the linear transformation determined by E. We have for any x_0, x_1,

$$|x_0 - x_1| = |E^{-1}(Ex_0 - Ex_1)| \leq |E^{-1}||Ex_0 - Ex_1|$$

where $|E^{-1}|$ is the sum of squares norm of the matrix E^{-1}. Take $2\alpha = 1/|E^{-1}|$

Take $H = f - Ex$, then $DH(a) = 0$, and by (assuming) mean-value theorem, for the ith component, $|H_i(x_0) - H_i(x_1)| = |DH_i(c)(x_0 - x_1)|$ for some c on the line between x_0, x_1. Because $DH(a) = 0$, we choose, by continuity, an open cube $C(a;\epsilon)$ around a such that $|DH(x)| < \alpha/\sqrt{n}$ for $x \in C(a;\epsilon)$. This means that each row vector of $DH(c)$ has norm $< \alpha/\sqrt{n}$, therefore $|H(x_0)-H(x_1)| \leq \alpha|x_0-x_1|$. Now, we can apply mean value theorem because the cube is convex and

$$\begin{aligned}\alpha|x_0 - x_1| &\geq |H(x_0) - H(x_1)| \\ &= |f(x_0) - f(x_1) + Ex_1 - Ex_0| \\ &\geq |Ex_0 - Ex_1| - |f(x_0) - f(x_1)| \\ &\geq 2\alpha|x_0 - x_1| - |f(x_0) - f(x_1)|\end{aligned}$$

The lemma follows. □

Theorem 2. *Let A be open in $\mathbb{R}^n$; let $f\colon A \to \mathbb{R}^n$ be of class $\mathcal{C}^r$; let $B = f(A)$. If f is one-to-one on A and if $Df(x)$ is non-singular for $x \in A$, then the set B is open in $\mathbb{R}^n$ and the inverse function $g\colon B \to A$ is of class $\mathcal{C}^r$.*

Proof. Step 1: Derivative at local extremum

We first prove that if $\phi\colon A \to \mathbb{R}$ is differentiable and has a local minimum (or maximum) at $x_0 \in A$, then $D\phi(x_0) = 0$. To say that it has a local minimum is to say that in a neighbourhood around x_0, $\phi(x_0)$ is minimum. So, in some neighbourhood U around x_0, for $x \in U, \phi(x) \geq \phi(x_0)$. Then, it is clear that all directional derivatives are 0. Since ϕ is differentiable, this means that $D\phi(x_0) = 0$.

Step 2: B is open

Let $a \in A, b = f(a)$. Choose a (closed) rectangle Q in A whose interior contains a. The boundary of Q is a compact set and since f is injective, its image is compact, closed and disjoint from b. Choose a $\delta > 0$ such that the ball of radius 2δ around b is disjoint from $f(\partial Q)$. Choose any $c \in B(b;\delta)$ and consider the function $|f(x) - c|^2$ on Q. This is a continuous function and Q is compact, hence it attains a minimum.

Since $|b - c|^2 < \delta^2$, the minimum so attained is less than δ^2, hence the minimum is not attained in ∂Q. Thus, the minimum is attained in the open interior of Q at some x_0, therefore x_0 is a point of local minimum (the minimum attained is a global minimum on Q). Using Step 1, the derivative of $|f - c|^2$ is zero at x_0. This derivative is

$$2Df(x)\begin{bmatrix}f_1(x_0) - c_1 & \dots & f_n(x_0) - c_n\end{bmatrix}^T = 0.$$

Using the fact that Df is non-singular, we conclude that $f(x_0) = c$ and the minimum attained is zero. Therefore, $B = f(A)$ containes the open ball $B(b;\delta)$ aroung $b = f(a)$. Since $b \in B$ was arbitrary, B is open.

Step 3: g is continuous

Because f is injective, we have an inverse function $g\colon B \to A$. To show continuity, we need to show that the inverse of open sets in A under g are open in B, this is the same as saying that the image of open sets in A under f are open in B, which follows from Step 2.

Step 4: g is differentiable

Let $b \in B$ with $g(b) = a$, and let $E = Df(a)$. Consider

$$G(k) = \frac{g(b+k) - g(b) - E^{-1}k}{|k|}$$

defined on a deleted neighbourhood of zero (sufficiently small so that $g(b+k)$ is defined; recall that B is open). If we show that this has zero as its limit as $k \to 0$, then g is differentiable at b with derivative E^{-1}. Define $\Delta(k) = g(b+k) - g(k)$ in the previously chosen deleted neighbourhood of 0.

From the previous lemma, there is an open cube C around a such that for $x, y \in C, |f(x)-f(y)| \geq \alpha|x - y|$. Because f is open, we choose $\epsilon > 0$ to be small enough so that $b + k \in f(C)$ when $|k| < \epsilon$, then

$$|b + k - b| \geq \alpha|g(b+k) - g(k)| = \alpha|\Delta(k)|.$$

Therefore, $|\Delta(k)|/|k| \leq 1/\alpha$ is bounded near a neigbourhood of zero.

Now, for $0 < |k| < \epsilon$, we have

$$\begin{aligned} G(k) &= \frac{\Delta(k) - E^{-1}k}{|k|} \\ &= -E^{-1}\left[\frac{k - E\Delta(k)}{|\Delta(k)|}\right]\frac{|\Delta(k)|}{|k|} \end{aligned}$$

where we use the fact that g is one-to-one, so $\Delta(k) \neq 0$ for $k \neq 0$. It suffices to show that the expression in the bracket goes to zero because the other terms are constants or bounded.

The expression in the bracket is

$$\frac{f(g(b+k)) - f(g(b)) - E(g(b+k) - g(b))}{|g(b+k) - g(b)|}.$$

Since g is continuous, as $k \to 0, \Delta(k) \to 0$ and f is differentiable at $g(b) = a$ with derivative E, therefore the above ratio goes to zero. Therefore $G(k) \to 0$ as $k \to 0$, hence g is differentiable at b with derivative $Df(g(b))^{-1}$.

Step 5: g is of class $\mathcal{C}^r$

From the previous step we know that the derivative $Dg(b) = Df(g(b))^{-1}$. Note that the entries of the inverse of a matrix are rational functions of the entries of the matrix. We proceed by induction on r. Suppose f is $\mathcal{C}^1$, then the entries of Df are continuous, and because g is continuous, the entries of $Dg(b)$ are also continuous. Thus, g is also $\mathcal{C}^1$.

Suppose we have proved that if f is $\mathcal{C}^{r-1}$, then so is g. Assume f is $\mathcal{C}^r$, then it is also $\mathcal{C}^{r-1}$, hence g is $\mathcal{C}^{r-1}$. The entries of Df are $\mathcal{C}^{r-1}$, therefore the entries of Dg are $\mathcal{C}^{r-1}$, hence g is $\mathcal{C}^r$. The proof follows by induction. □

Proof of Theorem 1. By the lemma above, there is a neighbourhood U_0 of a where f is one-to-one. Because $\det Df(x)$ is a continuous function, there is a neighbourhood U_1 of a where it is nonzero, i.e., Df is non-singular. Then taking $U = U_0 \cap U_1$, we can apply the previous theorem for $f\colon U \to \mathbb{R}^n$. The theorem follows. □

1.1 Norms on matrices

Given an $n \times n$ real matrix $A = (a_{ij})$, we consider two norms

- $\|A\|_2 = \sqrt{\sum a_{ij}^2}$
- $\|A\| = \sup\{\|Ax\| : \|x\| = 1\} = \sum_{x\neq 0} \frac{\|Ax\|}{\|x\|}$

Here the norm of a vector $x \in \mathbb{R}^n$ is the usual Pythagorean norm. The first norm is the usual Pythagorean norm of vectors when seeing A as an element of $\mathbb{R}^{n^2}$. The second norm, called the operator norm, gives a sense of how big the standard sphere becomes under the action of A. The operator norm is finite because $x \mapsto Ax$ is a continuous function on $\mathbb{R}^n$ and the sphere is compact.

- $\|A\| \leq \|A\|_2$

  Proof. Suppose the rows of A are $v_1, \ldots, v_n$, then Ax has entries $v_1 \cdot x, \ldots, v_n \cdot x$ where $\cdot$ is the standard dot product. Then, applying Cauchy-Schwartz inequality,

  $$\|Ax\|^2 = |v_1 \cdot x|^2 + \cdots + |v_n \cdot x|^2 \leq (\|v_1\|^2 + \cdots + \|v_n\|^2)\|x\| = \|A\|_2\|x\|.$$

  it follows that $\|A\| \leq \|A\|_2$ □

- For matrices A, B, we have $\|AB\|_2 \leq \|A\|_2\|B\|_2$ and $\|AB\| \leq \|A\|\|B\|$.

Proof. Suppose the rows of A are $r_1, \ldots, r_n$ and the columns of B are $c_1, \ldots, c_n$, then applying Cachy-Schwartz inequality,

$$\|AB\|_2^2 = \sum \|Ac_i\|^2 \leq \|A\|_2^2 \|B\|_2^2$$

and the first part follows. Next for $x \neq 0$, observe that

$$\|ABx\| \leq \|A\| \|Bx\| \leq \|A\| \|B\| x$$

therefore $\|AB\| \leq \|A\| \|B\|$. □

- Both norms satisfy the triangle inequality. They obviously follow the other axioms of a norm, so they deserve the name "norm".

  Proof. Triangle inequality for $\|\cdot\|_2$ follows from the triangle inequality on $\mathbb{R}^{n^2}$. For matrices A, B and $x \in \mathbb{R}^n$ observe that $\|(A+B)x\| \leq \|Ax\| + \|Bx\|$. Therefore, $\|\cdot\|$ also satisfies the triangle inequality. □

- Both norms are continuous functions $M_n(\mathbb{R}) \to \mathbb{R}$ where we treat $M_n(\mathbb{R})$ as $\mathbb{R}^{n^2}$ with the usual topology.

  Proof. $\|\cdot\|_2$ is a continuous because it is the square root of a non negative polynomial function. To show continuity of $\|\cdot\|$, we use the sequential definition. Suppose A_n converge to A in $\|\cdot\|_2$, then

$$\|A_n\| = \|A_n - A + A\| \leq \|A_n - A\| + \|A\| \leq \|A_n - A\|_2 + \|A\|$$
$$\|A\| = \|A - A_n + A_n\| \leq \|A - A_n\| + \|A_n\| \leq \|A - A_n\|_2 + \|A_n\|$$

  It follows that $\|A_n\|$ converges to $\|A\|$. □

There are other possible norms on matrices, but for our purposes we will need only these two.

1.2 Operator theoretic proof

We provide an independent proof of the inverse function theorem.

Theorem 3. *Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be a $\mathcal{C}^r, r \geq 1$ function. For $x^* \in \mathbb{R}^n$, suppose $Df(x^*)$ is invertible. Then there are open sets $U, V \subseteq \mathbb{R}^n$ such that $x^* \in U, f|_U \to V$ is invertible with inverse f^{-1} of calss $\mathcal{C}^r$ and $D(f^{-1})(y) = (Df(x))^{-1}$ where $y = f(x)$.*

Proof. By replacing f with $Df(x^*)^{-1} f$ if necessary, we may assume $Df(x^*) = I$. Observe that this function is $\mathcal{C}^r$ and obtaining an inverse for $Df(x^*)^{-1} f$ gives us an inverse for f as well. For any $y \in \mathbb{R}^n$, define

$$\begin{aligned} T_y \colon \mathbb{R}^n &\to \mathbb{R}^n \\ x &\mapsto x - f(x) + y \end{aligned}$$

Then T_y is of class $\mathcal{C}^r$ with $DT_y = I - Df$, so $DT_y(x^*) = 0$ for every y. Since $\|DT_y\|$ is a continuous function, for a given y, there is a neighbourhood, which we can take to be convex, U of x^* such that

$$\|DT_y(x)\| \leq 1/2, x \in U.$$

For $x \in U$, if $Df(x)$ is singular, say $Df(x)z = 0$ for some nonzero z, then $DT_y(x)z = z \implies \|DT_y(x)\| \geq 1$ which is impossible. Therefore, Df is nonsingular on U.

Now, by the fundamental theorem of calculus and the convexity of U,

$$\begin{aligned}\|T_y(x_1) - T_y(x_0)\| &= \|\int_{t=0}^{1} DT_y(x_0 + (x_1 - x_0)t)(x_1 - x_0)dt\| \\ &\leq \int_0^1 \|DT_y(x_0 + (x_1 - x_0)t)(x_1 - x_0)\| \text{ (to be proved)}\end{aligned}$$

hence,

$$\|T_y(x_1) - T_y(x_0)\| \leq \frac{1}{2}\|x_1 - x_0\|, x_0, x_1 \in U \tag{1}$$

Now, $\|Df(x^*)^{-1}\| = \|I\| = 1$. On U, Df is invertible, so $Df(x)^{-1}$ is continuous and we may shrink U so that $\|Df(x)^{-1}\| \leq 2$ on U. Set $V = f(U)$. Equation 1 means that f is injective on U, so $f^{-1}\colon V \to U$ is well-defined. So far the y in T_y was arbitrary.

- V is open (this actually shows that f is an open map)

  Let $x_0 = f^{-1}(y_0)$ for some $y_0 \in V$. Pick a closed ball $K = \overline{B_\delta(x_0)}$ around x_0 contained in U. We claim that $B_{\delta/2}(y_0) \subseteq V$. In fact, let $y \in B_{\delta/2}(x_0)$, then for any x

  $$\begin{aligned}\|T_y(x) - x_0\| &\leq \|T_y(x) - T_y(x_0)\| + \|T_y(x_0) - x_0\| \\ &\leq \frac{1}{2}\|x - x_0\| + \|y - y_0\| \\ &\leq \delta\end{aligned}$$

  Therefore, T_y maps K to itself and is a contraction (by 1). Hence T_y has a fixed point (by Banach fixed point theorem), which means that there is an $x \in K$ such that $T_y(x) = x \implies f(x) = y$.

- f^{-1} is continuous:

  Given $f(x_0), f(x_1)$ we have

  $$\begin{aligned}\|f^{-1}(f(x_1)) - f^{-1}(f(x_0))\| = \|x_1 - x_0\| &= \|T_{y^*}(x_1) + f(x_1) - T_{y^*}(x_0) - f(x_0)\| \\ &\leq \|T_{y^*}(x_1) - T_{y^*}(x_0)\| + \|f(x_1) - f(x_0)\| \\ &\leq \frac{1}{2}\|x_1 - x_0\| + \|f(x_1) - f(x_0)\|\end{aligned}$$

  where y^* is arbitrary. So,

  $$\|x_1 - x_0\| \leq 2\|f(x_1) - f(x_0)\| \tag{2}$$

  making f^{-1} continuous.

- f^{-1} is of class $\mathcal{C}^r$:

  As in the previous proof, the first step is to show that the derivative of f^{-1} is of a specific form. Fix $y_0 = f(x_0) \in V$. Observe that for $y = f(x) \in V$,

  $$f^{-1}(y_0) - f^{-1}(y) - [Df(x_0)]^{-1}(y_0 - y) = -[Df(x_0)]^{-1}(f(x_0) - f(x) - Df(x_0)(x_0 - x))$$

  therefore, by the choice of U,

  $$\|f^{-1}(y_0) - f^{-1}(y) - [Df(x_0)]^{-1}(y_0 - y)\| \leq 2\|f(x_0) - f(x) - Df(x_0)(x_0 - x)\|.$$

  It suffices to show that the right side divided by $\|y_0 - y\|$ tends to 0 as $y \to y_0$.

  $$\begin{aligned}\frac{\|f(x_0) - f(x) - Df(x_0)(x_0 - x))\|}{\|y_0 - y\|} &= \frac{\|f(x_0) - f(x) - Df(x_0)(x_0 - x))\|\|x_0 - x\|}{\|x_0 - x\|\|y_0 - y\|} \\ &\leq 2\frac{\|f(x_0) - f(x) - Df(x_0)(x_0 - x))\|}{\|x_0 - x\|} \text{ (by 2)}\end{aligned}$$

  and the right side goes to 0 as $x \to x_0$ because f is differentiable.

  Now, using the same argument as before, we can inductively show that f is indeed of class $\mathcal{C}^r$.

$\square$

Lemma 2. *Suppose $f\colon [0,1] \to \mathbb{R}^n$ is integrable, then $\|\int_0^1 f\| \leq \int_0^1 \|f\|$.*

Proof. Let $f = (f_1, \ldots, f_n)$ and $v = \int_0^1 f$. Then,

$$\|v\|^2 = \sum_{i=1}^{m} (\int_0^1 f_i)(\int_0^1 f_i) = \sum_{i=0}^{n} \int_0^1 (\int_0^1 f_i) f_i = \int_0^1 \sum_{i=1}^{m} (\int_0^1 f_i) f_i \leq \int_0^1 \|v\| \|f\|$$

giving us the desired inequality. $\square$

This proof generalizes to infinite dimensional Banach spaces when differentiability is replaced by Frechet differentiability. The core of the proof lies in choosing the transformation T_y whose fixed points are precisely points in $f^{-1}(y)$ and showing that T_y is a contraction of a suitable domain. Note that T_y translates x by $f(x) - y$, so we are perturbing x by some error which when 0, gives us a fixed point from $f^{-1}(y)$.

The inverse function theorem above requires f to be at least $\mathcal{C}^1$ (we used the fact that norms of matrices is continuous in both proofs above). However, it turns out that this assumption can be weakened, see [6].

1.3 Constant Rank Theorem

Theorem 4. *Suppose $f\colon U \to V$ is a $\mathcal{C}^r$ map where $U \overset{open}{\subseteq} \mathbb{R}^n$ and $V \overset{open}{\subseteq} \mathbb{R}^m$. Suppose in a neighbourhood U_1 of $p \in U, Df$ has constant rank k. Then, there is a diffeomorphism G of an open neighbourhood of p with an open set in $\mathbb{R}^n$ and a diffeomorphism F of an open neighbourhood of $f(p)$ with an open set in $\mathbb{R}^m$ such that $F \circ f \circ G^{-1}(r_1, \ldots, r_n) = (r_1, \ldots, r_k, 0, \ldots, 0)$.*

Note that the rank of the differential is $\leq \min\{n, m\}$, so the statement is valid. Secondly, the inverse function theorem follows from this theorem because we'll have $f = F^{-1}G$, a diffeomorphism. If at a point p, the rank is k, then some $k \times k$ submatrix of its derivative is invertible, so in a neighbourhood, the rank is at least k. When $k = n = m$, then the rank in a neighbourhood is n.

Proof. In $Df(p)$, we'll assume that the first $k \times k$ submatrix is invertible. If this is not the case, then we may either compose f with some permutations, or repeat the following proof with some notational changes and then permute $\mathbb{R}^n, \mathbb{R}^m$ afterwards. Write $f = (f_1, \ldots, f_m)$. By continuity, we obtain a neighbourhood $U_2 \subseteq U_1$ around p where the first $k \times k$ submatrix of the derivative is invertible and since the rank is exactly k on U_1, the rank is exactly k on U_2 as well. Consider

$$\begin{aligned} G\colon U_2 &\to \mathbb{R}^n \\ (x_1, \ldots, x_n) &\mapsto (f_1, \ldots, f_k, x_{k+1}, \ldots, x_n) \end{aligned}$$

Then on U_2, G is differentiable and of class $\mathcal{C}^r$ with an invertible derivative. By inverse function theorem, we obtain a neighbourhood $p \in U_3 \subseteq U_2$ where the restriction of G is a diffeomorphism.

Consider the map

$$\begin{aligned} f \circ G^{-1}\colon G(U_3) &\to \mathbb{R}^m \\ (y_1, \ldots, y_n) &\mapsto (y_1, \ldots, y_k, z_{k+1}, \ldots, z_m) \end{aligned}$$

This map is also differentiable of class $\mathcal{C}^r$. By the chain rule, since G is a diffeomorphism, its derivative has rank k (because f has rank k on U_3). We have

$$D(f \circ G^{-1}) = \begin{bmatrix} I_k & 0 \\ A & B \end{bmatrix}$$

where A, B are some matrices. Since it should have rank k, we conclude that $B = 0$.

Here we have an open set $G(U_3)$ on which we have defined functions $z_{k+1}, \ldots, z_m$. Since the derivatives of these functions with respect to $y_{k+1}, \ldots, y_n$ is zero, intuitively this means that these functions depend only on the first k coordinates. We can make this exact as follows. Replace $G(U_3)$ with an open ball B around $G(p)$ (the point around which the proof revolves), then G^{-1} restricted to B is a diffeomorphism. Now, suppose $q_1, q_2 \in B$ and have the same first k coordinates. Then, since B is convex, and z_i (which are defined on B) are differentiable, we may apply the mean value theorem to conclude that $z_i(q_1) = z_i(q_2), i \geq k+1$.

Let $W = \{x \in V : \Pi(x) \in \Pi(B)\}$ where Π is the projection onto the first k coordinates (there is a slight abuse of notation in that the Π on the left takes m-coordinates, but the one on the right takes n). This is an open set because the projection map is open and continuous (and because V is open). Moreover, because $G(p) \in B$, we have $f(p) \in W$. On W consider the function

$$\begin{aligned} F\colon W &\to \mathbb{R}^m \\ (u_1, \ldots, u_m) &\mapsto (u_1, \ldots, u_k, u_{k+1} - z_{k+1}, \ldots, u_m - z_m) \end{aligned}$$

where $z_i(u_1, \ldots, u_m)$ is defined as $z(u_1, \ldots, u_k, v_{k+1}, \ldots, v_n)$ where $v_{k+1}, \ldots, v_n$ are chosen so that $(u_1, \ldots, v_n) \in B$. By the arguments above, the value of z_i is independent of the choice of $v_{k+1}, \ldots, v_n$.

More accurately, we take

$$z_i(u_1, \ldots, u_m) = z_i(u_1, \ldots, u_k, f(p)_{k+1}, \ldots, f(p)_n),$$

where $f(p)_{k+1}, \ldots, f(p)_n$ are the last $n-k$ coordinates of $f(p)$, then it is easy to see that this z_i are differentiable. Moreover, we get

$$Df = \begin{bmatrix} I_k & 0 \\ A & I_{m-k} \end{bmatrix}$$

Therefore, F is a local diffeomorphism. Take a neighbourhood $\tilde{V}$ around $f(p)$ such that the restriction of F to $\tilde{V}$ is a diffeomorphism. Now, we pull everything back to $G(U_3)$.

Since $\tilde{V}$ is open, its preimage under the continuous $f \circ G^{-1}$ is open in B, and its preimage under G is open in $G^{-1}B$, call it $\tilde{U}$, so $\tilde{U} = G^{-1}((f \circ G^{-1})^{-1}(\tilde{V}))$. Then

$$\begin{tikzcd}
 & U \arrow[r, "f"] & V & \\
 & U_1 \arrow[u, hook] & W \arrow[u, hook] \arrow[r, "F"] & \mathbb{R}^m \\
\mathbb{R}^n & U_2 \arrow[l, "G"'] \arrow[u, hook] & \tilde{V} \arrow[u, hook] \arrow[r, "F", "\text{diff}"'] & F(\tilde{V}) \\
G(U_3) & U_3 \arrow[l, "G", "\text{diff}"'] \arrow[u, hook] & & \\
B \arrow[u, hook] \arrow[r, "G^{-1}", "\text{diff}"'] & G^{-1}(B) \arrow[u, hook] & & \\
(f \circ G^{-1})^{-1}(\tilde{V}) \arrow[u, hook] \arrow[r, "G^{-1}", "\text{diff}"'] & \tilde{U} \arrow[u, hook] \arrow[uuur, "f"] & &
\end{tikzcd}$$

This large diagram shows that there is a neighbourhood $\tilde{U}$ of p and a neighbourhood $\tilde{V}$ of $f(p)$ and diffeomorphisms G, F of these sets into open sets in $\mathbb{R}^n, \mathbb{R}^m$ respectively such that $F \circ f \circ G^{-1}(r_1, \ldots, r_n) = (r_1, \ldots, r_k, 0, \ldots, 0)$. Note that the diffeomorphisms F, G are of class $\mathcal{C}^r$. $\square$

The constant rank theorem of course applies more generally to maps between manifolds because it is a lcoal theorem. We define the following

Definition 1. *Let M be a smooth m-manifold, N a smooth n-manifold and $F\colon M \to N$ a smooth map. Then*

- *F is said to be an immersion if the derivative is injective everywhere, i.e., rank$(dF_p) =$ dim M.*
- *F is said to be a submersion if the derivative is surjective everywhere, i.e., rank$(dF_p) =$ dim N.*
- *F is said to be an embedding if the it is an immersion and F is a topologocial embedding, i.e., homeomorphic to its image in N.*

Theorem 5. *(Global Rank Theorem) let $F\colon M \to N$ be a smooth map of constant rank.*

- *If F is injective, then it is an immersion.*
- *If F is surjective, then it is a submersion.*
- *If F is bijective, then it is a diffeomorphism.*

Proof. Take a point $p \in M$, and restrict F to charts around $p, F(p)$, then there are open sets U, V around $p, F(p)$ respectively and diffeomorphisms $G\colon U \to \mathbb{R}^m, F\colon V \to \mathbb{R}^n$ such that $F \circ f \circ G^{-1}(r_1, \ldots, r_m) = (r_1, \ldots, r_k, 0, \ldots, 0), (r_1, \ldots, r_m) \in G(U)$ where k is the constant rank.

If F is injective, then we must have $k = m$, therefore F is an immersion.

If F is surjective, then **Author doesn't understand the proof yet, we are facing technical difficulties** □

2 Fubini's Theorem

A note on integration: We first consider Riemann integrals as in [1]. Here we first define the integral of a function over rectangles, i.e., sets of the type $Q = [a_1, b_1] \times \cdots \times [a_n, b_n]$ where $a_i \leq b_i$. Given a continuous function f over Q, its upper integral $\overline{\int_Q} f$ is defined as the infimum over all partitions P of Q (obtained by partitioning each $[a_i, b_i]$ and looking at the subrectangles they determine) of the upper sums. So, if Q is partitioned into rectangles $R_1, \ldots, R_m$, then $M_{R_i}(f) = \sup_{R_i} f$, and the upper sum $U(f; P)$ is $\sum M_{R_i}(f)v(R_i)$ where v is the volume of R_i given as the product of lenghts of its sides.

The lower integral $\underline{\int_Q} f$ is defined as the supremum over all partitions P of Q of the lower sums. With R_i as above, the lower sum $L(f; P)$ is $\sum m_{R_i}(f)v(R_i)$ where $m_{R_i}(f) = \inf_{R_i}(f)$. Using refinements of partitions, it is possible to show that $\underline{\int_Q} f \leq \overline{\int_Q} f$. The integral is said to exist if both these quantities are finite and equal.

2.1 Analytic; special case

Theorem 6. *Let $Q = A \times B$ where A is a rectangle in $\mathbb{R}^k$ and B a rectangle in $\mathbb{R}^n$. Let $f\colon Q \to \mathbb{R}$ be a bounded function; write f in the form $f(x, y)$ where $x \in A, y \in b$. For each $x \in A$, consider the lower and upper (Riemann) integrals*

$$\underline{\int_{y \in B}} f(x, y) \text{ and } \overline{\int_{y \in B}} f(x, y).$$

If f is integrable over Q, then these two functions of x are integrable over A, and

$$\int_Q f = \int_{x \in A} \underline{\int_{y \in B}} f(x, y) = \int_{x \in A} \overline{\int_{y \in B}} f(x, y).$$

Proof. For $x \in A$, set

$$\underline{I}(x) = \underline{\int_{y \in B}} f(x, y) \text{ and } \overline{I}(x) = \overline{\int_{y \in B}} f(x, y).$$

Assuming f is integrable over Q, we show that $\underline{I}, \overline{I}$ are integrable over A. Let P be any partition of Q, then P consists of partitions P_A of A and P_B of B, and we write $P = (P_A, P_B)$, such that if R_A

is a rectangle determined by P_A and R_B by P_B, then $R_A \times R_B$ is a rectangle determined by P and all rectangles determined by P arise this way.

Step 1: $L(f; P) \leq L(\underline{I}; P_A)$

Consider a general subrectangle $R_A \times R_B$ determined by P. For any $x_0 \in A$, we have $m_{R_A \times R_B}(f) \leq m_{R_B}(f(x_0, y))$. Holding x_0, R_A fixed, we multiply by $v(R_B)$ and sum over R_B to get

$$\sum_{R_B} m_{R_A \times R_B}(f) \leq L(f(x_0, y); P_B) \leq \underline{\int_B} f(x_0, y) = \underline{I}(x_0).$$

Since the inequality holds for any $x_0 \in R_A$, we have $\sum_{R_B} m_{R_A \times R_B}(f) \leq m_{R_A}(\underline{I})$. Multiply by $v(R_A)$ and sum to get

$$L(f; P) \leq L(\underline{I}; P_A).$$

An exactly similar process shows $U(f; P) \geq U(\overline{I}; P_A)$.

Step 2: Finishing the proof

Along with the inequalities above, we have $U(\underline{I}; P_A) \leq U(\overline{I}; P_A)$ and $L(\underline{I}; P_A) \leq L(\overline{I}; P_A)$. If f is integrable, then $L(f; P), U(f; P)$ can be made arbitrarily close. Because $U(\overline{I}; P_A), L(\overline{I}; P_A)$ are sandwiched between the lower and upper sums of f, they too can be made arbitrarily close and therefore $\overline{I}$ is integrable and moreover, for the same reason its integral is equal to $\int_Q f$. Similarly, $\underline{I}$ is integrable over A and its integral is $\int_Q f$. □

2.2 Construction of product measures

In this subsection, we prove the general Tonelli-Fubini theorems. Most of the material is taken from [2]. The first thing we need to do is to construct the product measure. Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are two measured spaces. The σ-algebra on the Cartesian product $X \times Y$ is the one generated by $A \times B$ where $A \in \mathcal{M}, B \in \mathcal{N}$ and is denoted $\mathcal{M} \otimes \mathcal{N}$. We seek to obtain a measure $\mu \otimes \nu$ on $X \times Y$ that is such that $\mu \otimes \nu(A \times B) = \mu(A)\nu(B)$ when $A \in \mathcal{M}, B \in \mathcal{N}$.

Fubini's theorem is ultimately trying to integrate a function on the Cartesian product as a step-wise integral. In order to generalise it, we need to know how to integrate indicator functions on the Cartesian product and this is the same as obtaining a product measure. However, we cannot always have such a product measure.

For example, consider $X = [0, 1]$ with the standard Borel measure, and $Y = [0, 1]$ with the counting measure and discrete σ-algebra. The diagonal D is a closed subset when considered in the Borel σ-algebra on $[0, 1]^2$, hence it is measurable. Since $X \times Y$ has a finer σ-algebra, D is measurable in $X \times Y$. Now, we look at its measure in the following ways, writing χ_D for the indicator on D,

$$\int_Y (\int_X \chi_D d\mu) d\nu = \int_Y 0 d\nu = 0$$
$$\int_X (\int_Y \chi_D d\nu) d\nu = \int_X 1 d\nu = 1$$

Suppose generally, $C \in \mathcal{M} \otimes \mathcal{N}$. We have

$$\int_Y (\int_X \chi_C d\mu) d\nu = \int_Y \mu(t^y(C)) d\nu \text{ and } \int_X (\int_Y \chi_C d\nu) d\mu = \int_X \nu(t_x(C)) d\mu$$

where $t_x(C) = \{y \in Y : (x, y) \in C\}$ and $t^y(C) = \{x \in X : (x, y) \in C\}$. Define $i_x \colon Y \to X \times Y$ sending $y \mapsto (x, y)$ and similarly $j_y \colon X \to X \times Y$, for $x \in X, y \in Y$. Then $t_x(C) = i_x^{-1}(C), t^y(C) = j_y^{-1}(C)$. We need the maps

$$x \mapsto \nu(t_x(C)), y \mapsto \mu(t^y(C))$$

to be measurable.

Before we proceed, we prove a few results.

Lemma 3. *Let $\mathcal{E}$ be the collection of all finite union of measurable rectangles in $X \times Y$, i.e., finite unions of sets of the type $A \times B, A \in \mathcal{M}, B \in \mathcal{N}$. Then $\mathcal{E}$ is an algebra of sets, i.e., it is closed under complements, finite intersection, finite union and contains $\emptyset, X \times Y$.*

Proof. Clearly $\mathcal{E}$ contains $\emptyset$ (as an empty union, or one can explicitly force this) and $X \times Y$. Moreover, by definition, it is closed under finite unions. Suppose we have $A \times B, C \times D$, then their intersection is $A \cap C \times B \cap D$, therefore the finite intersection of rectangles is another rectangle. It now follows that $\mathcal{E}$ is closed under finite intersections. To show closure under complement, it suffices to show that the complement of a rectangle is in $\mathcal{E}$, by the closure under finite intersections.

Clearly, the complement of $A \times B$ is given by $(X \setminus A) \times B \cup A \times (X \setminus B) \cup (X \setminus A) \times (X \setminus B)$, and is therefore an element of $\mathcal{E}$ (keep in mind that σ-algebras are closed under complements, countable intersections and unions). $\square$

Theorem 7. *(Monotone class lemma) Let X be a set, and $\mathcal{O}$ a collection of subsets which is closed under countable ascending unions and countable descending intersections. Such a collection is called a monotone class. If $\mathcal{A}$ is an algebra of sets contained in $\mathcal{O}$, then $\mathcal{O}$ contains the σ-algebra generated by $\mathcal{A}$.*

Proof. First suppose $\mathcal{B}$ is a monotone class and a set algebra. Then we show that $\mathcal{B}$ is a σ-algebra. All one needs to show is that it is stable under countable unions, because $\mathcal{B}$ is closed under complements. If $A_1, A_2, \ldots,$ is a countable subcollection of sets in $\mathcal{B}$, then their union is the countable ascending union $A_1 \subseteq A_1 \cup A_2 \subseteq \ldots$, therefore is an element of $\mathcal{B}$ because it is a monotone class (the finite unions are in $\mathcal{B}$ because it is a set algebra).

Our method is to show that the smallest monotone class containing $\mathcal{A}$ is a set algebra as well. Let $\mathcal{O}'$ be the smallest monotone class containing $\mathcal{A}$ in $\mathcal{O}$ (it can be obtained as the intersection of all monotone classes; check that arbitrary intersection of monotone classes is itself a monotone class). We need to show that $\mathcal{O}'$ is a set algebra. Our proof will heavily rely on the minimality of $\mathcal{O}'$, thus to show that a certain property P holds, we consider the set of all elements in $\mathcal{O}'$ where P holds and show that this subcollection is a monotone class containing $\mathcal{A}$.

We start with closure under complements. Let $\mathcal{G} = \{A \in \mathcal{O}' : X \setminus A \in \mathcal{O}'\}$. Clearly $\mathcal{A} \subseteq \mathcal{G}$ and $\mathcal{G}$ is closed under complements. If $A_1 \subseteq A_2 \subseteq \ldots$ is an ascending collection in $\mathcal{G}$, then $(\cup A_n)^c = \cap A_n^c$ is a descending intersection of sets in $\mathcal{O}'$ (because each $A_n \in \mathcal{G}$), therefore is an element of $\mathcal{O}'$, hence $\cup A_n \in \mathcal{G}$. Similarly, $\mathcal{G}$ is closed under countable descending intersections. Therefore, $\mathcal{G}$ is a monotone class containing $\mathcal{A}$, and by minimality, $\mathcal{G} = \mathcal{O}'$ and is closed under complements.

It suffices to show that $\mathcal{O}'$ is closed under finite unions and the problem is reduced to showing that whenever $A, B \in \mathcal{O}'$, we also have $A \cup B \in \mathcal{O}'$. Fix $A \in \mathcal{O}'$, and look at $\mathcal{G}_1 = \{B \in \mathcal{O}' : A \cup B \in \mathcal{O}'\}$. It is clear that $\mathcal{G}_1$ is a monotone class, but need not contain $\mathcal{A}$. Therefore, we first consider $A \in \mathcal{A}$, in this case $\mathcal{A} \subseteq \mathcal{G}_1$, therefore $\mathcal{G}_1 = \mathcal{O}'$. Since $A \in \mathcal{A}$ was arbitrary, this means that whenever $A \in \mathcal{A}, B \in \mathcal{O}'$, we have $A \cup B \in \mathcal{O}'$.

After this brief stop, we can take $A \in \mathcal{O}'$ arbitrary and consider $\mathcal{G}_1$ as above. Then, it is a monotone class containing $\mathcal{A}$, therefore $\mathcal{G}_1 = \mathcal{O}'$ and since $A \in \mathcal{O}'$ was arbitrary, we conclude that $\mathcal{O}'$ is closed under finite unions. Therefore, $\mathcal{O}'$ is a monotone class and a set algebra, hence a σ-algebra. It therefore contains $\sigma(\mathcal{A})$. However, note that $\sigma(\mathcal{A})$ is also a monotone class containing $\mathcal{A}$, therefore $\sigma(\mathcal{A}) = \mathcal{O}' \subseteq \mathcal{O}$ as required. $\square$

Now we proceed to the construction of product measure. We state the following easy lemma.

Lemma 4. *For a fixed $x \in X$, we have*

$$\begin{aligned} t_x((X \times Y) \setminus C) &= Y \setminus t_x(C) \\ t_x(\cup_{i \in I} C_i) &= \cup_{i \in I} t_x(C_i) \\ t_x(\cap_{i \in I} C_i) &= \cap_{i \in I} t_x(C_i) \end{aligned}$$

where I is any indexing set.

Theorem 8. *Let $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ be measured spaces.*

1. *For any $C \in \mathcal{M} \otimes \mathcal{N}, x \in X, t_x(C) \in \mathcal{N}$, in other words, i_x is measurable.*
2. *If we assume Y is σ-finite (i.e., countable union of finite measure spaces), the non negative function $x \mapsto \nu(t_x(C))$ is $\mathcal{M}$-measurable for every $C \in \mathcal{M} \otimes \mathcal{N}$. Note that by part 1, $t_x(C)$ is a measurable set.*
3. *If Y is σ-finite, then the map*

$$\begin{aligned} \Pi \colon \mathcal{M} \otimes \mathcal{N} &\to [0, \infty] \\ C &\mapsto \int_X \nu(t_x(C)) d\mu \end{aligned}$$

is a measure on $\mathcal{M} \otimes \mathcal{N}$.

Our method of proof will involve that the statements hold when C comes from some monotone class containing the set algebra of rectangles. The monotone class lemma then tells that these statements hold for sets coming from the σ-algebra generated by the rectangles which is precisely $\mathcal{M} \otimes \mathcal{N}$. Note that because rectangles are particularly simple measurable sets, the statements are not going to be hard to show for rectangles. Moreover, because measures behave well with ascending and descending sets (because of convergence theorems like monotone convergence theorem and dominated convergence theorem), ascending unions and descending intersections should similarly not provide any difficulties. Thus, we have all the necessary ingredients to prove the statements above.

Proof. 1. IF $C = A \times B$ is a rectangle, then $i_x^{-1}(A \times B) = B$ if $x \in A$, else it is empty and in both cases, it is measurable. Therefore, the σ-algebra $\{C \subseteq X \times Y | i_x^{-1}(C) \in \mathcal{N}\}$ contains all rectangles, hence it contains $\mathcal{M} \otimes \mathcal{N}$.

2. Let $\mathcal{O} = \{C \subseteq X \times Y : x \mapsto \nu(t_x(C)) \text{ is measurable }\}$. If $C = A \times B$ is a rectangle, then the corresponding map is $x \mapsto \nu(B)$ if $x \in A$ and 0 if not. Therefore, $\mathcal{O}$ contains all rectangles.

   Step 1: Assume $\nu(Y) < \infty$

   If $C \in \mathcal{O}$, take $D = X \times Y \setminus C$. Then $\nu(t_x(D)) = \nu(Y \setminus t_x(C)) = \nu(Y) - \nu(t_x(C))$ is well defined and measurable, therefore $\mathcal{O}$ is closed under complements. Now, suppose $C_n, n \geq 1$ are disjoint elements of $\mathcal{O}$, then

$$\nu(t_x(\sqcup_{n\geq 1} C_n)) = \sum_{n\geq 1} \nu(t_x(C_n)).$$

   The right side is the monotone limit of measurable functions, hence measurable. Similarly, $\mathcal{O}$ is closed under ascending countable unions. By closure under complements, it is closed under descending countable intersections.

   Now, any finite union of rectangles can be written as a finite disjoint union of rectangles (simply look at all the regions in the Venn diagrams), therefore $\mathcal{O}$ is a monotone class containing the set algebra generated by rectangles. By the monotone class lemma, we conclude that $\mathcal{O} \supseteq \mathcal{M} \otimes \mathcal{N}$.

   Step 2: Assume Y is σ-finite

   Write $Y = \cup Y_n$ where Y_n are sets of finite measure. We can adjust the Y_i so as to be disjoint. Given any $C \in \mathcal{M} \otimes \mathcal{N}$, we can write $C = \sqcup C \cap Y_i$, so $\nu(t_x(C)) = \sum_{n\geq 1} \nu(t_x(C \cap Y_i))$. By Step 1, each $\nu(t_x(C \cap Y_i))$ is measurable, and $\nu(t_x(C))$ is their monotone limit, therefore it is measurable.

3. By part 2 Π is defined because the integrand is measurable. Suppose $\{C_n\}_{n\geq 1}$ is a collection of disjoint measurable sets, then

$$\nu(t_x(\sqcup_{n\geq 1} C_n)) = \sum_{n\geq 1} \nu(t_x(C_n)).$$

   By Monotone convergence theorem, we have $\Pi(\sqcup_{n\geq 1} C_n) = \sum_{n\geq 1} \Pi(C_n)$. It follows that Π is a measure on $(X \times Y, \mathcal{M} \otimes \mathcal{N})$. $\square$

Lemma 5. *(Uniqueness of product measure) Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces. Suppose μ_1, μ_2 are measures on $(X \times Y, \mathcal{M} \otimes \mathcal{N})$ such that $\mu_1(C) = \mu_2(C)$ whenever $C = A \times B$ is a rectangles. Further suppose there is a sequence of disjoint measurable rectangles $C_n \times D_n$ such that $X \times Y = \sqcup(C_n \times D_n)$ with $\mu_i(C_n \times D_n) < \infty, i = 1, 2$. Then $\mu_1 = \mu_2$.*

Proof. Let $\mathcal{O} = \{C \in \mathcal{M} \otimes \mathcal{N} : \mu_1(C) = \mu_2(C)\}$. Then $\mathcal{O}$ contains the set algebra of rectangles and by monotone convergence is closed under ascending countable unions. First suppose $\mu_i(X \times Y) < \infty, i = 1, 2$, then $\mathcal{O}$ is closed under complements as well, hence closed under descending countable intersections. Therefore, $\mathcal{O}$ contains all of $\mathcal{M} \otimes \mathcal{N}$ by the monotone class lemma.

In the general case, we have the equality of restrictions $\mu_1|_{C_n \times D_n} = \mu_2|_{C_n \times D_n}$ by the previous case. Given any measurable C, we have

$$\mu_1(C) = \sum \mu_1(C \cap (C_n \times D_n)) = \sum \mu_2(C \cap (C_n \times D_n)) = \mu_2(C).$$

Therefore, $\mu_1 = \mu_2$. □

Now, if $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ are both σ-finite, then using the theorem above we know that for $C \in \mathcal{M} \otimes \mathcal{N}, t^y(C) \in \mathcal{M}$ and $y \mapsto \mu(t^y(C))$ is measurable. Moreover, $C \mapsto \int_Y \mu(t^y(C))d\nu$ is a measure on $\mathcal{M} \otimes \mathcal{N}$. Observe that for rectangles $C = A \times B$,

$$\int_X \nu(t_x(C))d\mu = \mu(A)\nu(B) = \int_Y \mu(t^y(C))d\nu.$$

Theorem 9. *(Existence of product measure) Let $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ be two σ-finite measure spaces. Then there is a unique measure $\mu \otimes \nu$ on $\mathcal{M} \otimes \mathcal{N}$ such that on rectangles $A \times B, \mu \otimes \nu(A \times B) = \mu(A)\nu(B)$.*

Proof. For $C \in \mathcal{M} \otimes \mathcal{N}$, set

$$\mu_1(C) = \int_X \nu(t_x(C))d\mu$$
$$\mu_2(C) = \int_Y \mu(t^y(C))d\nu$$

Then $\mu_1 = \mu_2$ on rectangles. If we take $X = \sqcup_n X_n, Y = \sqcup_m Y_m$ where X_n, Y_m are of finite measure, then $X \times Y = \sqcup_{m,n} X_n \times Y_m$ is σ-finite under μ_1, μ_2. Since μ_1, μ_2 agree on rectangles, they are equal. Moreover, by the uniqueness, μ_1 is the unique measure θ satisfying $\theta(A \times B) = \mu(A)\nu(B)$ because any such θ will make $X \times Y$ into a σ-finite space and agrees with μ_1 on rectangles. □

Remark. The product of two σ-finite measures is again σ-finite, so we may take n-fold prodcts. Moreover, by the uniqueness, we have $(\mu_1 \otimes \mu_2) \otimes \mu_3 = \mu_1 \otimes (\mu_2 \otimes \mu_3)$, i.e., the product is associative, because these two agree on rectangles $A_1 \times A_2 \times A_3$ which generates the σ-algebra on three-fold Cartesian products (as shown below).

[Suppose $(X_1, \mathcal{M}_1), (X_2, \mathcal{M}_2)$ and $(X_3, \mathcal{M}_3)$ are three measurable spaces. Then we claim that $\mathcal{M} := \mathcal{M}_1 \otimes (\mathcal{M}_2 \otimes \mathcal{M}_3) = (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3 =: \mathcal{N}$ and that they are both the σ-algebra $\mathcal{L}$ generated by cubes $A_1 \times A_2 \times A_3, A_i \in \mathcal{M}_i, i = 1, 2, 3$. First, $\mathcal{M}$ is the σ-algebra generated by $A \times B$ where $A \in \mathcal{M}_1, B \in \mathcal{M}_2 \otimes \mathcal{M}_3$. Therefore $\mathcal{L} \subseteq \mathcal{M}$.

Next, for any $A \in \mathcal{M}_1$ look at $\{C \subseteq X_2 \times X_3 : A \times C \in \mathcal{L}\}$. This is a σ-algebra and contains all measurable rectangles in $X_2 \times X_3$, therefore $A \times C \in \mathcal{L}$ for every $C \in \mathcal{M}_2 \times \mathcal{M}_3$. It follows that $\mathcal{L} = \mathcal{M}$. Similarly $\mathcal{N} = \mathcal{L} = \mathcal{M}$. Therefore, the product of σ-algebras is associative.]

2.3 Tonelli-Fubini theorems

Theorem 10. *(Tonelli-Fubini theorems) Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measures spaces and $\mu \otimes \nu$ the product measure on $X \times Y$.*

1. *(Tonelli) If $f\colon X \times Y \to [0,\infty]$ is measurable, then the non negative functions*

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

are measurable and

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

2. *(Fubini) If $f\colon X\times Y \to \mathbb{C}$ is in $L^1(\mu\otimes\nu)$, then for ν-almost all $y\in Y$, the function $t^y(f)\colon x \mapsto f(x,y)$ is measurable and μ-integrable on X, and for μ-almost all $x\in X$, the function $t_x(f)\colon y\mapsto f(x,y)$ is measurable and ν-integrable and the functions*

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

are respectively μ,ν-integrable and

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

3. *If $f\colon X\times Y \to \mathbb{C}$ is such that $x\mapsto \int_Y |f(x,y)|d\nu(y)$ is μ-integrable, then $f\in L^1(\mu\otimes\nu)$.*

Proof. First observe that $f\mapsto t_x(f)$ (and similarly $f\mapsto t^y(f)$) satisfies the following

- $t_x(\alpha f+\beta g) = \alpha t_x(f)+\beta t_x(g)$
- $t_x(f)^{\pm} = t_x(f^{\pm})$ where $f^+ = \max\{f,0\}, f^- = \max\{-f,0\}$ for real valued f
- $Re(t_x(f)) = t_x(Re(f))$ and $Im(t_x(f)) = t_x(Im(f))$
- $f\le g \implies t_x(f)\le t_x(g)$
- $f_n\to f$ pointwise a.e., then $t_x(f_n)\to t_x(f)$ pointwise a.e.

1. The statement is true for $f=\chi_C$ for $C\in\mathcal{M}\otimes\mathcal{N}$, and by linearity it holds for non negative step functions. Now, let f be a non negative measurable function and $\{s_n\}_{n\ge 0}$ a sequence increasing to f. Then for any fixed $x\in X, t_x(s_n)$ increases to $t_x(f)$, therefore by monotone convergence, $t_x(f)$ is measurable and

$$\int_Y t_x(f)d\nu = \lim\int_Y t_x(s_n)d\nu.$$

Now $\int_Y t_x(s_n)d\nu$ is a measurable function on X, and increases to $\int_Y t_x(f)d\nu$, therefore applying the monotone convergence theorem again,

$$\int_X\int_Y t_x(f)d\nu d\mu = \lim\int_X\int_Y t_x(s_n)d\nu d\mu$$
$$= \lim\int_{X\times Y} s_n d(\mu\otimes\nu) \text{ by the case of step functions}$$
$$= \int_{X\times Y} fd(\mu\otimes\nu) \text{ monotone convergence theorem}$$

Interchanging the roles of X, Y we get the other equation.

2. Let $f \in L^1(\mu \otimes \nu)$. The function $t_x(f)$ is a composition of measurable functions, therefore it is a measurable function on Y. By Tonelli's theorem we have

$$\int_X \int_Y |f| d\nu d\mu = \int_{X \times Y} |f| d(\mu \otimes \nu) < \infty.$$

Therefore the function $x \mapsto \int_Y |f(x,y)| d\nu$ is a nonnegative measurable function with finite integral over X, thus the integrand $\int_Y |t_x(f)| d\nu$ is finite for μ-almost every x, therefore $t_x(f) \in L^1(\nu)$ for μ-almost all $x \in X$.

Assume f is real valued, so $\int_Y t_x(f) d\nu = \int_Y t_x(f^+) d\nu - \int_Y t_x(f^-) d\nu$, this makes sense only when $t_x(f)$ is absolutely integrable. When $t_x(f) \in L^1(\nu)$, we have $t_x(f^+), t_x(f^-) \in L^1(\nu)$ because they are dominated by $t_x(f)$. By the first part, we know that $x \mapsto \int_Y t_x(f^\pm) d\nu$ is measurable on X, therefore, $x \mapsto \int_Y t_x(f) d\nu$ is also measurable (note that it is defined almost everywhere because $t_x(f) \in L^1(\nu)$ for almost all x; where it is not defined (because of $\infty - \infty$ type problems) make it zero).

Next, we also have

$$\int_X |\int_Y f(x,y) d\nu| d\mu \leq \int_X \int_Y |f| d\nu d\mu < \infty.$$

Therefore, $x \mapsto \int_Y f d\nu$ is μ-integrable. Finally, applying Tonelli's theorem

$$\begin{aligned}
\int_X \int_Y f d\nu d\mu &= \int_X \int_Y t_x(f^+) d\nu d\mu - \int_X \int_Y t_x(f^-) d\nu d\mu \\
&= \int_{X \times Y} f^+ d(\mu \otimes \nu) - \int_{X \times Y} f^- d(\mu \otimes \nu) \\
&= \int_{X \times Y} f d(\mu \otimes \nu)
\end{aligned}$$

Interchanging the roles of X, Y gives the other equality. By taking the real part and imaginary parts separately, the statement holds when f is complex valued.

3. By Tonelli's theorem, we have

$$\int_{X \times Y} |f| d(\mu \otimes \nu) = \int_X \int_Y |f| d\nu d\mu < \infty.$$

Therefore, $f \in L^1(\mu \otimes \nu)$. □

If we don't assume absolute integrability for Fubini's theorem, we have counterexamples. Consider $X = Y = (0,1)$ with Borel measure and

$$f(x,y) = \frac{x^2 - y^2}{(x^2 + y^2)^2}$$

which is continuous, hence measurable. Applying Tonelli's theorem

$$\begin{aligned}
\int_{X \times Y} |f| &= \int_X \int_Y |f| dy dx \\
&\geq \int_X \int_0^x \frac{x^2 - y^2}{(x^2 + y^2)^2} dy dx \\
&= \int_X \frac{1}{2x} = \infty
\end{aligned}$$

We also have

$$\int_X \int_Y f = \int_X \frac{1}{x^2 + 1} dx = \frac{\pi}{4}$$
$$\int_Y \int_X f = \int_Y \frac{-1}{y^2 + 1} dy = -\frac{\pi}{4}$$

3 Riesz-Markov-Kakutani Representation Theorem

Let X be a topological space. Given a continuous function $f\colon X \to \mathbb{C}$, define its support to be $\mathrm{supp}(f) = \overline{\{x : f(x) \neq 0\}}$. The collection of compactly supported functions, $C_C(X)$ form a complex vector space. Since continuous functions on compact sets are bounded, the sup norm $\|\cdot\|_\infty$ is seen to be a vector space norm on $C_C(X)$.

Lemma 6. *Let X be a topological space, μ a Borel measure finite on compact sets, then*

$$\begin{aligned}\Lambda\colon C_C(X) &\to \mathbb{C}\\ f &\mapsto \int_X f d\mu\end{aligned}$$

is well defined, linear and nonnegative (i.e., if $f \in C_C(X)$ is non negative, then $\Lambda(f) \geq 0$).

Proof. Because μ is finite on compact sets and $f \in C_C(X)$ is bounded and supported on compact sets, it is easy to see that Λ is well defined. Clearly Λ is linear and nonnegative. □

The Riesz-Markov-Kakutani representation theorem allows one to construct Borel measures from linear functionals like Λ above. Under suitable conditions on X, one can obtain functions that behave like characteristic functions of compact sets. Once we have such functions, then we can define, using the linear functional, a measure on compact sets. We can then extend this measure to open sets if we can suitably approximate open sets with compact sets inside them. Of course, such a notion would require X to be Hausdorff and locally compact.

Definition 2. *Let X be a Hausdorff space, and μ a Borel measure. Then μ is said to be*

- *Outer regular if for every Boreal measurable $E, \mu(E) = \inf\{\mu(U) : E \subseteq U, U\,open\}$.*
- *Inner regular or tight if for every open $U, \mu(U) = \sup\{\mu(K) : K \subseteq U, K \text{ compact }\}$.*
- *Inner regular on finite measure sets and open sets if the condition of inner regularity hold on open sets as well as measurable E of finite measure.*
- *Locally finite if every $x \in X$ has a neighbourhood U of finite measure.*

μ is said to be a Radon measure if it is outer regular, inner regular on open sets and locally finite.

On a locally compact Hausdorff space, μ is said to be regular if it is outer regular, inner regular on open and finite measure sets and finite on compact sets. Note that if μ is locally finite, then it is finite on compact sets and under the assumption of local compactness, the converse holds as well.

Theorem 11. *(Riesz-Markov-Kakutani representation theorem) Let X be a locally compact Hausdorff space and $\Lambda\colon C_C(X) \to \mathbb{C}$ a nonnegative linear map.*

1. *There exists a σ-algebra $\mathcal{M}$ finer than the Borel σ-algebra $\mathcal{B}$ on X and a complete measure μ on $\mathcal{M}$ which is finite on compact sets such that $\Lambda(f) = \int_X f d\mu \,\forall\, f \in C_C(X)$.*
2. *There is a unique regular measure μ such that $\Lambda(f) = \int_X f d\mu$.*
3. *If in addition X has the property that every open set is a countable union of compact sets, then the measure μ is unique even without the regularity assumption.*

3.1 Urysohn lemma and applications

Before we proceed, we need some theorems from topology. To keep things short, I shall only state the results. The proofs can be found in [4]. Although the results here have been stated for locally compact Hausdorff spaces, they hold more generally for normal spaces. Now, locally compact T_2 spaces are not normal (eg. the Tychonoff plane with an edge deleted), however by the one point compactification, they are subspaces of compact T_2 spaces.

Lemma 7. *Locally compact Hausdorff spaces are regular.*

This regularity allows us to separate compact spaces by open sets which allows us to prove

Theorem 12. *(Urysohn's lemma) Let X be a locally compact T_2 space, C a compact subspace contained in an open set U, then there is a continuous function $f: X \to [0,1]$ such that $f|_C = 1$ and is 0 outside U.*

The proof given in [4] is under the assumption of normality, however since we are dealing with a compact subspace C, the same proof applies here. Alternately, as mentioned above, we may move to the one point compactification of X, obtain such a function, and then restrict it to X. However, under local compactness, we can make sure that f has compact support by simply first obtaining a neighbourhood V of C with compact closure contained in U, and then shrinking (see below for more on this) V further, and obtaining the Urysohn function.

Theorem 13. *(Tietze extension theorem): Let X be a locally compact T_2 space, $A \subseteq X$ a compact subspace. Suppose $f: A \to [a,b]$ is continuous, then there is an extension $g: X \to [1,b]$ restricting to f on A.*

Once again, the proof in [4] is under the assumption of normality and it applies here because we are dealing with a compact subset A.

Theorem 14. *(Finite partition of unity): Let $\{U_1, \dots, U_n\}$ be an open cover of a normal space X, then there is a partition of unity dominated by $\{U_n\}$, i.e., a collection $\phi_1, \dots, \phi_n : X \to [0,1]$ such that $\sum \phi_i = 1$ and supp(ϕ_1) $\subseteq U_i$.*

Proof. In a normal space, Urysohn's lemma applies when C above is replaced by a closed set. Now, we have for example $X \setminus (U_2 \cup \cdots \cup U_n) \subseteq U_1$ and we can obtain a function that behaves like a characteristic function. This way we get a bunch of functions, however it is not guaranteed that they add to nonzero values at any $x \in \cap U_n$. So, we need to enlarge our closed sets, or rather shrink our open sets so that such intersection points are taken care of. Once we have n functions that add to a nonzero positive value everywhere, then we can normalize it by dividing by the sum.

Step 1. Shrinking the cover

$A = X \setminus (U_2 \cup \cdots \cup U_n)$ is a closed subset in U_1, so we obtain open V_1 such that $A \subseteq V_1 \subseteq \overline{V_1} \subseteq U_1$, then $\{V_1, U_2, \dots, U_n\}$ is an open cover of X. Having obtained the cover $\{V_1, \dots, V_k, U_{k+1}, \dots, U_n\}$, we can obtain an open V_{k+1} such that $X \setminus (V_1 \cup \cdots \cup V_k \cup U_{k+2} \cup \cdots \cup U_n \subseteq V_{k+1} \subseteq \overline{V_{k+1}} \subseteq U_{k+1}$.

Continuing this process, we obtain a cover $\{V_1, \dots, V_n\}$ of X such that $\overline{V_i} \subseteq U_i$.

Step 2. Obtaining the partition

Shrink the cover once again to get $\{W_1, \dots, W_k\}$ such that $\overline{W}_i \subseteq V_i$. For each i, using Urysohn's lemma, obtain a continuous $f_i : X \to [0,1]$ such that f_i is 1 on W_i and 0 outside V_i, so that supp$(f_i) \subseteq \overline{V_i}$. Thus, these V_i act as buffers to prevent the support overflowing U_i. Since W_i cover X, we have $\sum f_i \neq 0$ everywhere. Taking $\phi_i = f_i / \sum f_i$ gives us our partition of unity subordinate to $\{U_i\}$. □

Lemma 8. *Let X be a locally compact Hausdorff space.*

1. *For any compact $K \subseteq X$ and open neighbourhood V of K, there is an $f \in C_C(X)$ such that $\chi_K \leq f \preceq \chi_V$ where $f \preceq \chi_V$ means $f \leq \chi_V$ and supp(f) $\subseteq V$.*
2. *For disjoint compact subsets K_1, K_2, there is an $f \in C_C(X)$ such that $0 \leq f \leq 1$ and f is 0 on $K_1, 1$ on K_2.*
3. *Let K be compact, $V_1, \dots, V_n$ open such that $K \subseteq V_1 \cup \cdots \cup V_n$. For any $g \in C_C(X)$, there exist $g_i \in C_C(X)$ such that supp(g_i) $\subseteq V_i$ and $\sum g_i(x) = g(x) \forall x \in K$. In addition, if $g \geq 0$, then g_i can also be chosen to be nonnegative.*

Proof. 1. Since K is compact and X is locally compact, we obtain a U such that $K \subseteq U \subseteq \overline{U} \subseteq V$ and $\overline{U}$ being compact. Applying Urysohn's lemma, obtain a continuous function which is 1 on K and 0 outside U, then supp$(f) \subseteq \overline{U}$, hence is compact.

2. This follows by 1. by considering $K_2 \subseteq X \setminus K_1$ (which is open because X is T_2).

3. K is a compact Hausdorff space, hence normal. The sets $V_i \cap K$ are open in K and cover K, so we can obtain a finite partition of unity subordinate to $\{V_1 \cap K, \ldots, V_n \cap K\}$, let it be $\phi_1, \ldots, \phi_n$ (let us say that we either consider only those V_i such that $V_i \cap K \neq \emptyset$, or we consider the corresponding ϕ_i to be 0). We apply the Tietze extension theorem to obtain $\psi_i \colon X \to [0,1]$ extending ϕ_i. However, we have no control over how ψ_i looks outside K, so we multiply it by appropriate "indicator" functions.

   Now, $\mathrm{supp}(\phi_i) \subseteq V_i \cap K$ is a closed subset of K, hence compact. By applying Urysohn's lemma, we obtain $\theta_i \colon X \to [0,1]$ such that θ_i is 1 on $\mathrm{supp}(\phi_i)$ with $\mathrm{supp}(\theta_i) \subseteq V_i$. Now look at $\theta_i\psi_i \colon X \to [0,1]$. Clearly, $\mathrm{supp}(\theta_i\psi_i) \subseteq \mathrm{supp}(\theta_i) \cap \mathrm{supp}(\psi_i) \subseteq V_i$. Next, for any $x \in K$,

$$\sum \theta_i(x)\psi_i(x) = \sum \theta_i(x)\phi_i(x) = \sum \phi_i(x) = 1$$

   because when $\phi_i(x) \neq 0, x \in \mathrm{supp}(\phi_i)$, hence $\theta_i(x) = 1$.

   Set $f_i = \theta_i\psi_i \colon X \to [0,1]$, then f_i is continuous, $\mathrm{supp}(f_i) \subseteq V_i, \sum f_i = 1$ on K. Given $g \in C_C(X)$, we can take $g_i = gf_i$. □

3.2 Proof of RMKRT

Let X be a locally compact T_2 space and $\Lambda \colon C_C(X) \to \mathbb{C}$ a nonnegative linear functional. Suppose μ, ν are two regular Borel measures such that $\int_X f d\mu = \Lambda(f) = \int_X f d\nu \,\forall\, f \in C_C(X)$. In order to show that $\mu = \nu$, it suffices to show that they agree on open sets by outer regularity. Set

$$\mu^+(U) = \sup\{\Lambda(f) : 0 \leq f \preceq \chi_U\} \in [0,\infty].$$

Clearly, $\mu^+(U) \leq \mu(U), \nu(U)$. Next we use inner regularity on open sets. Given any compact $K \subseteq U$, we obtain an $f \in C_C(X)$ such that $\chi_K \leq f \preceq \chi_U$, then $\mu(K) \leq \Lambda(f) \leq \mu^+(U) \leq \nu(U)$ and hence $\mu(U) \leq \nu(U)$. Similarly we obtain the other inequlity proving that $\mu = \nu$ everywhere. This proves uniqueness.

In order to construct the measure, we force regularity conditions. In fact, given Λ as above, we first define an outer measure μ^+ on all subsets of X as follows. When U is open, define

$$\mu^+(U) = \sup\{\Lambda(f) : f \in C_C(X), 0 \leq f \preceq \chi_U\}.$$

For any other subset E, define

$$\mu^+(E) = \inf\{\mu^+(U) : E \subseteq U, U \text{ open}\}.$$

Clearly, μ^+ is monotonic on open sets, hence the two definitions of μ^+ agree on their common domain, i.e., the open sets. Next we define an inner measure μ^- by

$$\mu^-(E) = \sup\{\mu^+(K) : K \subseteq E, K \text{ compact}\}$$

where E is any subset. We then consider $\mathcal{M} = \{E \subseteq X : \mu^+(E) = \mu^-(E)\}$ and $\mu \colon \mathcal{M} \to [0,\infty]$ by $\mu(E) = \mu^+(E) = \mu^-(E), E \in \mathcal{M}$ (since Λ is nonnegative, μ is nonnegative). We will show that $\mathcal{M}$ is a σ-algebra finer than the Borel σ-algebra, that μ is a measure and that $\Lambda(f) = \int_X f d\mu, \forall f \in C_C(X)$.

Some initial observations:

- Monotonicity : We have shown that μ^+ is monotonic on open sets, but it is also clear that it is monotonic in general. Similarly, μ^- is also monotonic.
- Finiteness on compact sets: Suppose K is a compact set. By local compactness, we obtain a neighbourhood V of K with compact closure and applying regularity, obtain an open W such that $K \subseteq W \subseteq \overline{W} \subseteq V$. Since V has compact closure, so does W. Since $\overline{W}$ is compact, we may obtain $f \colon X \to [0,1]$ such that $f \in C_C(X), \chi_{\overline{W}} \leq f \preceq \chi_V$. Then given any $0 \leq g \preceq \chi_W, g \in C_C(X)$, we have $g \leq f$, hence $\Lambda(g) \leq \Lambda(f) < \infty$ by nonnegativity of Λ, hence, $\mu^+(W) < \infty$ and therefore $\mu^+(K) < \infty$.

- $\mathcal{M}$ contains compact sets. This follows from monotonicity. Similarly, by monotoncity, $\mu^-(E) \leq \mu^+(E)$ for every $E \subseteq X$.

- $E \in \mathcal{M}$ if and only if for every $\epsilon > 0$, there are compact K, open U such that $K \subseteq E \subseteq U$ and $\mu^+(U) - \mu^+(K) < \epsilon$.

Step 1. Outer measure is countably subadditive:

First, suppose U_1, U_2 are open and $f \in C_C(X)$ with $f \preceq \chi_{U_1 \cup U_2}$. Then $\text{supp}(f) \subseteq U_1 \cup U_2$ and we can get hold of $f_1, f_2 \in C_C(X)$ with $0 \leq f_i \preceq \chi_{U_i}$ such that on $\text{supp}(f), f = f_1 + f_2$. Since f_i are nonnegative, we get $f \leq f_1 + f_2$, hence $\Lambda(f) \leq \Lambda(f_1) + \Lambda(f_2) \leq \mu^+(U_1) + \mu^+(U_2)$, hence $\mu^+(U_1 \cup U_2) \leq \mu^+(U_1) + \mu^+(U_2)$. By induction, the result holds for finite unions of open sets.

Suppose $U = \cup_{n\geq 1} U_n$ is a countable union of open sets. Given any $0 \leq f \preceq U$, we obtain a finite subcover of $\text{supp}(f)$, say $\text{supp}(f) \subseteq U_1 \cup \cdots \cup U_n$. Then,

$$\Lambda(f) \leq \mu^+(U_1 \cup \cdots \cup U_n) \leq \sum_{n\geq 1} \mu^+(U).$$

Therefore μ^+ satisfies countable subadditivity for open sets.

Next, suppose $E = \cup E_n$ is a countable union of subsets of X. If any $\mu^+(E_i) = \infty$ or $\sum \mu^+(E_n) = \infty$, then countable subadditivity of μ^+ is obvious, so assume $\sum \mu^+(E_n) < \infty$. In this case, given $\epsilon > 0$ obtain open sets $U_i \supseteq E_i$ such that

$$\mu^+(U_i) \leq \mu^+(E_i) + \epsilon/2^i.$$

Then $E \subseteq \cup U_n$ and

$$\mu^+(\cup U_n) \leq \sum \mu^+(U_n) \leq \sum \mu^+(E_n) + \epsilon$$

where we were able to separate the sums because they are all convergent. It follows that $\mu^+(E) \leq \sum \mu^+(E_n) + \epsilon$ and since $\epsilon > 0$ was arbitrary, countable subadditivity follows.

Step 2. For compact $K, \mu(K) = \inf\{\Lambda(f) : f \in C_C(X), \chi_K \leq f\}$:

Let $m(K) = \inf\{\Lambda(f) : f \in C_C(X), \chi_K \leq f\}$. Given an open $U \supseteq K$, by Urysohn's lemma, there is an $f \in C_C(X)$ such that $\chi_K \leq f \preceq \chi_U$, hence $\Lambda(f) \leq \mu^+(U)$ which means that $m(K) \leq \mu^+(U)$. Since U was arbitrary, we get $m(K) \leq \mu(K)$.

Next, given an $f \in C_C(X)$ such that $\chi_K \leq f$, for $0 < \alpha < 1$ let $U_\alpha = \{x : f(x) > \alpha\}$. Then U_α is an open set by continuity of f and $K \subseteq U_\alpha$ for every $0 < \alpha < 1$. Fix one such α, then for any $0 \leq g \leq \chi_{U_\alpha}, g \in C_C(X)$, we have

$$\alpha g \leq f \implies \alpha\Lambda(g) \leq \Lambda(f) \implies \alpha\mu^+(U_\alpha) \leq \Lambda(f).$$

Then, $\alpha\mu^+(K) \leq \Lambda(f)$. Since $0 < \alpha < 1$ is arbitrary, we get $\mu(K) \leq \Lambda(f)$, hence $\mu(K) \leq m(K)$.

Step 3. Open sets are in $\mathcal{M}$:

Given $\epsilon > 0$, choose $f \in C_C(X)$ such that $\mu^+(U) - \epsilon \leq \Lambda(f)$. Consider the compact set $K = \text{supp}(f)$. Now if $K \subseteq W$ and W is open, then $\Lambda(f) \leq \mu^+(W)$, hence by taking infimum, $\Lambda(f) \leq \mu(K)$. Therefore, $\mu^-(U) = \mu^+(U)$.

Step 4. $\mathcal{M}$ is stable under countable disjoint union and μ is countably additive:

This step is very similar to the countably subadditivity property. Let K_1, K_2 be disjoint compact subsets, then we can find disjoint open sets V_1, V_2 such that $K_i \subseteq V_i$. Obtain $f_1, f_2 \in C_C(X)$ such that $\chi_{K_i} \leq f_i \preceq \chi_{V_i}$. Now, given any $f \in C_C(X)$ such that $\chi_K \leq f$ where $K = K_1 \cup K_2$ is compact, we have

$$ff_1 + ff_2 \leq f \implies \mu(K_1) + \mu(K_2) \leq \Lambda(ff_1) + \Lambda(ff_2) \leq \Lambda(f)$$

where we have used the fact that $ff_i \in C_C(X)$ and $ff_i \geq \chi_{K_i}$ because both f, f_i satisfy that inequality and the fact that $f_1 + f_2 \leq 1$ on X because they have disjoint supports. We have also used the result from Step 2. Applying Step 2 again gives $\mu(K_1) + \mu(K_2) \leq \mu(K)$ and we already know that μ is subadditive. By induction, we can extend this additivity for finite union of compact sets.

Next, suppose we are given disjoint subsets E_n of X. If any $\mu^-(E_n) = \infty$, then there are compact subsets K_n such that $K_n \subseteq E_n \subseteq \sqcup E_n = E$ and $n \leq \mu^+(E_n)$. It follows that $\mu^-(E) = \sum \mu^-(E_n) = \infty$. So, assume that each E_n has finite inner measure. Given $\epsilon > 0$, obtain compact K_n such that

$$\mu^+(K_n) \geq \mu^-(E_n) - \epsilon/2^n.$$

Then, for any $N \geq 1$,

$$\mu^-(E) \geq \mu^+(\cup_{i=1}^N K_i) = \sum_{i=1}^N \mu^+(K_n) \geq \sum_{i=1}^N \mu^-(E_n) - \epsilon.$$

Taking $N \to \infty$, we get $\mu^-(E) \geq \sum_{n\geq 1} \mu^-(E_n) - \epsilon$. Since $\epsilon > 0$ was arbitrary, we have

$$\sum_{n\geq 1} \mu^-(E_n) \leq \mu^-(E) \leq \mu^+(E) \leq \sum_{n\geq 1} \mu^+(E_n).$$

When $E_n \in \mathcal{M}$, the ends are equal, therefore $E \in \mathcal{M}$.

Step 5. $\mathcal{M}$ is a σ-algebra:

Suppose $E_1, E_2 \in \mathcal{M}$, we show that $E_1 \setminus E_2 \in \mathcal{M}$, this will then show that $\mathcal{M}$ is stable under complements and by the previous step, stable under countable intersections. It is clear that $\mathcal{M}$ contains $\emptyset, X$.

Given $E_1, E_2 \in \mathcal{M}, \epsilon > 0$ obtain open sets V_1, V_2 and compact sets K_1, K_2 such that $K_i \subseteq E_i \subseteq V_i$ and $\mu(V_i) - \mu(K_i) < \epsilon$. Since $V_i \setminus K_i$ is open, it is in $\mathcal{M}$, and by additivity, we get

$$\mu(V_i) - \mu(K_i) = \mu(V_i \setminus K_i) < \epsilon.$$

Now,

$$E_1 \setminus E_2 \subseteq V_2 \setminus K_2 \cup K_1 \setminus V_2 \cup V_1 \setminus K_1.$$

At the same time, $K_1 \setminus V_2$ is a compact set contained in $E_1 \setminus E_2$, hence

$$\mu^+(E_1 \setminus E_2) \leq \mu^+(K_1 \setminus V_2) + 4\epsilon \leq \mu^-(E_1 \setminus E_2) + 4\epsilon.$$

Since $\epsilon > 0$ was arbitrary, we conclude that $E - 1 \setminus E_2 \in \mathcal{M}$, therefore $\mathcal{M}$ is a σ-algebra.

At this stage, we have a σ-algebra $\mathcal{M}$ that is finer than $\mathcal{B}$. The function μ on $\mathcal{M}$ is clearly a measure which is by construction, finite on compact sets, outer regular, inner regular on open sets (and any other $E \in \mathcal{M}$). Therefore, μ is a regular measure. By local compactness, it is also a Radon measure. Furthermore, it is a complete measure, for if $E \in \mathcal{M}$ with $\mu(E) = 0$, then for any subset $F \subseteq E, \mu^+(F) \leq \mu^+(E) = 0$, hence $\mu^-(F) = \mu^+(F) = 0 \implies F \in \mathcal{M}$.

Step 6. $\Lambda(f) = int_X f d\mu, \forall f \in C_C(X)$:

Take any $f \in C_C(X)$ with compact support K, then f must be bounded by some $M < \infty$, hence $\int_X |f| d\mu = \int_K |f| d\mu \leq M\mu(K) < \infty$, therefore f is absolutely integrable. Since both $\Lambda, \int_X$ are $\mathbb{C}$-linear, it suffices to look at nonnegative $f \in C_C(X)$. Moreover, by dividing by M, we may suppose $M = 1$, i.e. $f\colon X \to [0,1]$.

First suppose $s = \sum_{i=1}^n \alpha_i \chi_{E_i}$ is a nonnegative step function where $E_1, \ldots, E_n$ are disjoint measurable sets and $s \leq f$. Fix $\epsilon > 0$ and choose compact $K_i \subseteq E_i$ such that $\mu(E_i) - \epsilon \leq \mu(K_i)$. Next, since $K_1, \ldots, K_n$ are disjoint compact sets, we can find functions $f_1, \ldots, f_n\colon X \to [0,1]$ with disjoint compact supports such that $f_i|_{K_i} = 1$. Then it follows that $f_1 + \cdots + f_n \leq 1$ everywhere.

Now, on K_i, we have $\alpha_i \leq f f_i$, which gives $\alpha_i \mu(K_i) \leq \Lambda(f f_i)$ because when $\alpha_i \neq 0, f f_i/\alpha_i \geq \chi_{K_i}$. Therefore (because f is nonnegative),

$$\sum_{i=1}^n \alpha_i \mu(K_i) \leq \Lambda(\sum_{i=1}^n f f_i) \leq \Lambda(f).$$

By the choice of K_i, we get $\int_X s d\mu - (\sum \alpha_i)\epsilon \leq \Lambda(f)$ and $\epsilon > 0$ was arbitrary. It then follows that $\int_X f d\mu \leq \Lambda(f)$.

Take $E_i = f^{-1}((\frac{i}{n}, \frac{i+1}{n}]), -1 \le i \le n-1$. These E_i are disjoint and since f is continuous, they are measurable. Given $\epsilon > 0$, choose open sets V_i such that $E_i \subseteq V_i, \mu(V_i) \le \mu(E_i) + \epsilon$. By replacing V_i with $V_i \cap f^{-1}((\frac{i}{n}, \frac{i+1}{n} + \epsilon))$ we may assume that $V_i \subseteq f^{-1}((\frac{i}{n}, \frac{i+1}{n} + \epsilon))$.

Then $K \subseteq \cup E_i \subseteq \cup V_i$, hence there are nonnegative functions $g_1, \dots, g_n \in C_C(X)$ with $\operatorname{supp}(g_i) \subseteq V_i$ and $\sum g_i = 1$ on K. Since $\operatorname{supp}(f) = K, f = \sum f g_i$ on X. Then

$$\Lambda(f) = \sum \Lambda(f g_i) \le \sum (\frac{i+1}{n} + \epsilon)\Lambda(g_i)$$

because $fg_i = 0$ outside V_i and on $V_i, fg_i \le ((i+1)/n + \epsilon)g_i$, therefore $fg_i \le ((i+1)/n + \epsilon)g_i$. Now, we also have $\Lambda(g_i) \le \mu(V_i)$, therefore

$$\Lambda(f) \le \sum (\frac{i+1}{n} + \epsilon)(\mu(E_i) + \epsilon).$$

Since $\epsilon > 0$ was arbitrary, we get

$$\begin{aligned}\Lambda(f) &\le \sum \frac{i+1}{n}\mu(E_i) \\ &= \int_X s_n d\mu + \frac{1}{n}\sum \mu(E_i) \\ &\le \int_X s_n d\mu + \frac{\mu(K)}{n}\end{aligned}$$

where s_n is the step function attaining i/n on $E_i, i \ge 1$ and 0 on $E_{-1}, X \setminus K$. Now, $s_n \le f$, therefore $\int_X s_n d\mu \le \int_X f d\mu$. Letting $n \to \infty$, we get $\Lambda(f) \le \int_X f d\mu$ as required.

We have thus constructed a regular measure on a σ-algebra $\mathcal{M}$ finer than the Borel σ-algebra such that $\Lambda(f) = \int_X f d\mu, \forall f \in C_C(X)$. Note that the regularity was forced by construction. Finally, we arrive at the third part of the representation theorem.

Theorem 15. *Let X be a locally compact T_2 space such that every open set is a countable union of compact sets. Let $\Lambda \colon C_C(X) \to \mathbb{C}$ be a linear nonnegative map and suppose μ is any Borel measure such that $\Lambda(f) = \int_X f d\mu, \forall f \in C_C(X)$, then μ is regular. By the uniqueness proved earlier, μ is unique.*

Proof. From the preceding parts, we have a Borel regular (and Radon) measure ν such that $\Lambda(f) = \int_X f d\nu, \forall f \in C_C(X)$. It suffices to show that $\mu = \nu$ on the Borel σ-algebra $\mathcal{B}$. Let V be an open set, write $V = \cup_{n \ge 1} K_n$ as a union of comapct sets, we may assume $K_{n-1} \subseteq K_n$. These K_n, together with Urysohn lemma will help us approximate χ_V using functions in $C_C(X)$.

Choose $f_n \in C_C(X)$ such that $\chi_{K_n} \le f_n \le \chi_V$. Because $K_n \subseteq K_{n+1}, f_n \le f_{n+1}$. Moreover, because $V = \cup K_n, \lim f_n = \chi_V$. Therefore, applying monotone convergence theorem,

$$\mu(V) = \lim \int_X f_n d\mu = \lim \Lambda(f_n) = \lim \int_X f_n d\nu = \nu(V).$$

Next, let $E \in \mathcal{B}$ with $\nu(E) < \infty$. Given $\epsilon > 0$, using regularity of ν, obtain compact K, open V such that $K \subseteq E \subseteq V$ and $\nu(V) - \nu(K) < \epsilon$. Since $V \setminus K$ is open, $\mu(V \setminus K) = \nu(V \setminus K) = \nu(V) - \nu(K) < \epsilon$.

$$\begin{aligned}\mu(E) \le \mu(V) = \nu(V) \le \nu(E) + \epsilon \\ \nu(E) \le \nu(V) = \mu(V) = \mu(V \setminus K) + \mu(K) \le \mu(E) + \epsilon\end{aligned}$$

It follows that $\nu(E) = \mu(E)$.

Lastly, let $E \in \mathcal{B}$ with $\nu(E) = \infty$. In this case, write $X = \cup_{n \ge 1} K_n$ as a union of compact sets with $K_n \subseteq K_{n+1}$. Then each $E \cap K_n$ has finite measure with respect to ν, so $\nu(E \cap K_n) = \mu(E \cap K_n) < \infty$. By monotonicity, $\nu(E \cap K_n) \to \nu(E)$ and $\mu(E \cap K_n) \to \mu(E)$, and it follows that $\mu(E) = \nu(E) = \infty$. □

Corollary 1. *Let X is a locally compact T_2 space such that all open sets are countable union of compact sets. Suppose μ_1, μ_2 are two Borel measures finite on compact sets such that $\int_X f d\mu_1 = \int_X f d\mu_2, \forall f \in C_C(X)$. Then $\mu_1 = \mu_2$*

Proof. Since μ_1 is finite on compact sets, it defines a

$$\begin{aligned} \Lambda_1 \colon C_C(X) &\to \mathbb{C} \\ f &\mapsto \int_X f d\mu_1 \end{aligned}$$

This is clearly a linear nonnegative map, hence by the theorem μ_1 is regular and equal to the regular measure constructed from Λ_1. Similarly, μ_2 is regular and equal to the regular measure constructed from a similarly defined Λ_2 where the integration is done with respect to μ_2. However, by the assumption, $\Lambda_1 = \Lambda_2$, therefore $\mu_1 = \mu_2$. □

3.3 A note on lebesgue measure

The Riesz-Markov-Kakutani representation theorem allows us to construct the Lebesgue measure. One starts with the usual Riemann integration defined in terms of the limit of tagged sums or Darboux upper and lower sums. This defines a well defined linear nonnegative functional on the space of compactly supported continuous functions. The measure on obtains from this is a completion of the Borel measure, i.e. the Lebesgue measure. There are alternate ways to arrive at the Lebesgue measure. For example, in [3], one starts with elementary sets which are finite unions of boxes/rectangles in $\mathbb{R}^n$ and defines a elementary measure.

Then one can define a Jordan outer (inner) measure of a set E by taking the infimum (supremum) of the elementary measures of all elementary sets B containing (contained in) E. This gives us a notion of Jordan measurable sets, namely those whose outer and inner measures coincide. The Riemann integral is closely related to the Jordan measure.

While the Jordan measure takes care of many things, it misses quite a few important sets. To rectify this, one then considers countable union of boxes instead of finite union and defines the Lebesgue outer measure similarly. One defines a set E to be Lebesgue measurable if one can always choose open sets U so that the Lebesgue outer measure of $U \setminus E$ is arbitrarily small. One has to then show that we do indeed obtain σ-algebra and a measure and that this measure is regular and Radon etc.

One can also construct the Lebsgue measure using the Carathéodory's extension theorem.

Theorem 16. *Every open set U in $\mathbb{R}^n$ can be written as a countable union of disjoint comapct sets.*

Proof. First consider the cubes in $\mathbb{Z}^n$. Accept a cube if it is contained in U, tentatively accept it if it intersects U and U^c and reject it otherwise. Divide the tentatively accepted cubes (if any) into cubes of side length $1/2$ of the original and repeat the procedure. This way, we get a countable family of disjoint accepted cubes. The union of these cubes is contained in U and given any $x \in U$, we may obtain a suitable cube of side length $1/2^n$ for some n considered in one of the lattices above, therefore U will be the union of these disjoint accepted cubes. □

3.4 Change of variables

In general it is not possible to integrate directly over manifolds because we do not have measures. Even when the manifold is embedded in a Euclidean space, we cannot use the Lebesgue measure because that will evaluate to zero. In such a situation, what we would like to do is to divide the manifold into pieces where we can integrate. Charts provide a natural domain where integration is possible, but we cannot arbitrarily choose charts because there is a possibility of charts overlapping and counting points more than once.

On parametrized manifolds, we can integrate a function by pulling it back to some open set of Euclidean space and adjusting for the diffeomorphism by scaling the function by the determinant of the derivative. Of course, this needs a justification on why the result so obtained is independent of the parametrization.

Suppose we have a $\mathcal{C}^1$ diffeomorphism $\phi\colon U \to V$ between open sets $U, V \subseteq \mathbb{R}^n$ and we have a function f on V. We would like to know the integral of $f \circ \phi$ over U with respect to the usual lebesgue measure. We will prove that

$$\int_U (f \circ \phi) d\lambda = \int_V f |\det d\phi| d\lambda.$$

In other words, we want to know the push forward of the Lebesgue measure under ϕ from U to V. This pushforward gives us the change of variables theorem. Intuitively, this is clear because the Lebesgue measure depends on the volumes of rectangles in $\mathbb{R}^n$ and at each point, these rectangles are acted upon by the linear transformation $d\phi$ which scales the volume by $|\det d\phi|$. Of course, we first need to prove this. We follow [2]. There is a proof in [1], but it is a little hard to follow and doesn't give this measure theoretic background.

Theorem 17. *(Translational invariance) The Lebesgue meaure λ on $\mathbb{R}^n$ is translationally invariant. Moreover, if μ is any Borel measure on $\mathbb{R}^n$ which is finite on compact sets and invariant under translation, then $\mu = c\lambda$ for some $c \geq 0$, specifically $c = \mu([0,1]^n)$.*

Here, by invariance under translation, we mean that given any $t \in \mathbb{R}^n$ and E Borel/Lebesgue measurable, $\mu(E + t) = \mu(E)$.

Proof. First we prove translational invariance of the Lebesgue measure. Fix $t \in \mathbb{R}^n$ and let T_t denote translation by t i.e., $T_t(x) = x + t$. Set $\mu = T_{t*}\lambda$. Now, because translation is a homeomorphism, μ is finite on compact sets. So, we conclude that μ is a regular/Radon measure. By outer regularity of both μ, λ, it suffices to check $\mu = \lambda$ on open sets.

When $n = 1$, given any open set, we can write it as a countable union of disjoint open intervals. Clearly, the length of intervals doesn't change under translation, so $\mu = \lambda$ on open sets as required. When $n > 1$, we use Tonelli-Fubini theorems. Since translation can be done in steps, so can the pushforward of measures. Therefore, it suffices to consider a translation $T\colon (x_1, \ldots, x_n) \mapsto (x_1 + t, x_2, \ldots, x_n)$. Let U be an open set.

$$\begin{aligned}
\mu(U) = \lambda(T^{-1}U) &= \int_{\mathbb{R}^n} \chi_{T^{-1}U} d\lambda \\
&= \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \chi_U(x_1 + t, x_2, \ldots, x_n) d\lambda_1 d\lambda_{n-1} \\
&= \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \chi_U(x_1, x_2, \ldots, x_n) d\lambda_1 d\lambda_{n-1} \\
&= \int_{\mathbb{R}^n} \chi_U d\lambda \\
&= \lambda(U)
\end{aligned}$$

where λ_1, λ_{n-1} are the Lebesgue measures on $\mathbb{R}, \mathbb{R}^{n-1}$ respectively.

Next, suppose μ is a Borel measure on $\mathbb{R}^n$ finite on compact sets and translationally invariant. We know that μ is then a regular Borel measure. Take $Q = [0,1]^n, c = \mu(Q) < \infty$. Let $\mathcal{K} = \{E \in \mathcal{B} : \mu(E) = c\lambda(E)\}$. Now, suppose $E \in \mathcal{K}$ and F is Borel measurable such that $E = \sqcup_{i=1}^m (t_i + F)$, then

$$\mu(E) = m\mu(F) \text{ and } \lambda(E) = m\lambda(F)$$

by translational invariance. It follows that $F \in \mathcal{K}$. Next, suppose H be a hyperplane obtained by setting some $x_i = a$. We wish to show that $\mu(H) = 0$. By translational invariance, we may set $a = 0$. Take $Q' = Q \cap H$ so that $H = \cup_{m \in \mathbb{Z}^n}(m + Q')$.

We also have $\sqcup_{n \geq 1}(1/n + Q') \subseteq Q$, hence using translational invariance, $\sum_{n \geq 1} \mu(Q') \leq \mu(Q) < \infty$. This is possible only if $\mu(Q') = 0$, hence $\mu(H) = 0$. We similarly have $\lambda(Q') = \lambda(H) = 0$.

Now, $(0,1]^n \in \mathcal{K}$ and this can be tiled with finite translations of $(0,1/2^k]^n$, therefore each $[0,1/2^k]^n \in \mathcal{K}$. Similarly, all rational sided cubes are also in $\mathcal{K}$. Now, every open set in $\mathbb{R}^n$ can be written as a countable union of disjoint cubes of side length $1/2^k$ where $k \geq 1$, hence $\mathcal{K}$ contains all open sets. By regularity, $\mathcal{B} \subseteq \mathcal{K}$, hence $\mu = c\lambda$. $\square$

Lemma 9. *(Linear change of variables) Let $T \in GL_n(\mathbb{R}^n)$, then $T_*(\lambda) = |\det T|^{-1}\lambda$.*

Proof. Let E be Borel measurable, and let $t \in \mathbb{R}^n$, then

$$T_*(E+t) = \lambda(T^{-1}(E+t)) = \lambda(T^{-1}E + T^{-1}t) = T_*(E)$$

because T is linear. Therefore, $T_*\lambda$ is translationally invariant and finite on compact sets (because T is a homeomorphism), hence $T_*\lambda = c\lambda$ where $c = T_*\lambda([0,1]^n)$. Since T is invertible, we can write $T^{-1} = E_1 \dots E_k$ where E_i are elementary measures (involving scaling vectors, swapping two entries and adding one entry to another), it suffices to show that each of them multiply by the corresponding absolute values of determinants.

If E scales a certain coordinate by d, then the measure of the unit cube scales by $|d|$. If E swaps two coordinates, then the unit cube, hence its measure, is invariant, i.e., scaled by 1. So, assume E to be of the third type. We will consider E mapping $x_1 \mapsto x_1 + x_2$ and leaving the other coordinates unchanged.

In this case, the image of $[0,1]^n$ is

$$\begin{aligned} E([0,1]^n) &= \{x_1(e_1+e_2) + x_2 e_2 + \cdots + x_n e_n : 0 \leq x_i \leq 1\} \\ &= \{x_1 e_1 + (x_1+x_2)e_2 + \cdots + x_n e_n : 0 \leq x_i \leq 1, x_1 + x_2 \leq 1\} \\ &\sqcup \{x_1 e_1 + (x_1+x_2)e_2 + \cdots + x_n e_n : 0 \leq x_i \leq 1, x_1 + x_2 > 1\} \end{aligned}$$

Translating the second set by $(0,-1,0,\dots,0)$, the two sets make $[0,1]^n$, therefore by translational invariance of $E_*\lambda$, this also leaves the measure invariant.

It follows that $c = |\det T|^{-1} = |\det T^{-1}|$. $\square$

Theorem 18. *Let U, V be open in $\mathbb{R}^n$ and $\phi\colon U \to V$ a $\mathcal{C}^1$ diffeomorphism. Then $\phi_*(d\lambda(x)) = |J_{\phi^{-1}}(y)|d\lambda(y)$ as measures on V. Here $J_{\phi^{-1}}(y)$ is the determinant of the derivative of ϕ^{-1} at y.*

Set $J(y) = |J_{\phi^{-1}}(y)|, \mu_1 = \phi_*(d\lambda), \mu_2 = J(y)d\lambda$. Both μ_1, μ_2 are finite on compact subsets of V because ϕ is a homeomorphism and J, being continuous, is bounded on compact sets.

Lemma 10. *With the notation as above, $\mu_1(E) \leq \mu_2(E) \forall E \subseteq U$, Borel.*

Proof. Because both μ_1, μ_2 are regular, it suffices to show that $\mu_1(W) \leq \mu_2(W)$ for open sets W. Furthermore, because each open W can be written as an almost disjoint, i.e., intersecting only at the faces, union of cubes, say $W = \sqcup_{n \geq 1} Q_i$, it suffices to show the inequality for cubes. For then,

$$\mu_1(W) \leq \sum \mu_1(Q_i) \leq \sum \mu_2(Q_i) = \mu_2(W).$$

Here the last equality comes from the fact that the intersection lies in some hyperplane and is compact. Since $J(y)$ is bounded on compact sets and λ is zero on the intersection, it will follow that μ_2 is zero on the faces of cubes. Therefore, μ_2 is finitely additive on almost disjoint cubes, and by monotone convergence, it is additive on countable collections of almost disjoint cubes. We cannot say the same for μ_1 because we do not know if the pullback of the intersection will still have Lebesgue measure zero, that requires some argument like Sard's theorem. So, let Q be a cube in V.

Step 1. For any $T \in \mathrm{GL}_r(\mathbb{R}), \mu_1(Q) \leq |\det T|^{-1} M^n \lambda(Q)$ where $M = \sup\{\|T \circ D_y \phi^{-1}\| : y \in Q\}$ where the norm is the operator norm:

First consider $T = I$. Because Q is convex, applying the fundamental theorem of calculus gives us (see a similar result in the section on inverse function theorem),

$$\|\phi^{-1}(x) - \phi^{-1}(y)\| \leq \sup\{D\phi^{-1}(z) : z \in Q\}\|x - y\|$$

hence $\phi^{-1}(Q)$ is contained in a cube whose side length is at most M times the side length of Q. Therefore, $\mu_1(Q) \leq M^n \lambda(Q)$. For an arbitrary T, set $\psi = \phi \circ T^{-1}$. Then

$$M = \sup\{\|D\psi^{-1}\|(z) : z \in Q\} = \sup\{\|T \circ D\phi^{-1}(y)\| : y \in Q\}$$

and

$$\begin{aligned}\mu_1(Q) &= \lambda(\phi^{-1}(Q)) \\ &= \lambda(T^{-1}(T \circ \phi^{-1}(Q))) \\ &= (T_*\lambda)(\psi^{-1}Q) \\ &= |\det T|^{-1}\lambda(\psi^{-1}Q) \leq |\det T|^{-1} M^n \lambda(Q)\end{aligned}$$

Step 2. Controlling M:

Fix $\epsilon > 0$. Now, $y \mapsto D_y\phi^{-1}$ is a continuous function, and since the operator norm is bounded by the Pythagorean norm, we may find, by uniform continuity (for Q is compact) almost disjoint cubes $Q_1, \ldots, Q_m$ such that for each i,

$$\|D_y\phi^{-1} - D_z\phi^{-1}\| < \epsilon \,\forall\, y, z \in Q_i.$$

Because Q_i is compact, the jacobian attains a minimum and we may choose $y_i \in Q_i$, such that $J(y_i) = \min_{y \in Q_i} J(y_i)$. Set $T_i = (D_{y_i}\phi^{-1})^{-1}$, then for any $y \in Q_i$,

$$\begin{aligned}\|T_i D_y \phi^{-1}\| &= \|T_i D_y\phi^{-1} - I + I\| \\ &\leq \|T_i(D_y\phi^{-1} - T_i^{-1})\| + \|I\| \\ &\leq \|T_i\|\|D_y\phi^{-1} - T_i^{-1}\| + 1 \\ &\leq c\epsilon + 1\end{aligned}$$

where $c = \sup_{y \in Q}\|D_y\phi^{-1}\| < \infty$ by continuity and compactness of Q. Note that $c \neq 0$ because ϕ is a diffeomorphism.

The sets $\phi^{-1}(Q_i)$ cover $\phi^{-1}(Q)$, hence by subadditivity,

$$\mu_1(Q) \leq \sum_{i=1}^m \mu_1(Q_i) \leq \sum_{i=1}^m |\det T_i|^{-1}(1 + c\epsilon)^n \lambda(Q_i).$$

Set $s(y) = \sum_{i=1}^m J(y_i)\chi_{Q_i}$. This is a step function and $s \leq J(y)$ by the choice of y_is, hence

$$\mu_1(Q) \leq (1 + c\epsilon)^n \int_Q s d\lambda \leq (1 + c\epsilon)^n \int_Q J d\lambda = (1 + c\epsilon)^n \mu_2(Q).$$

Since $\epsilon > 0$ was arbitrary, we obtain $\mu_1 \leq \mu_2$. □

Proof of change of variables theorem. We finally prove the Change of Variables theorem. Let $f\colon V \to [0, \infty)$ be a measurable function. Then, using the lemma,

$$\int_U f \circ \phi d\lambda(x) = \int_V f d\mu_1(y) \leq \int_V f d\mu_2(y) = \int_V f J d\lambda = \int_V f(y)|J_{\phi^{-1}}(y)| d\lambda.$$

Applying the same to ϕ^{-1} and any $g\colon U \to [0, \infty)$,

$$\int_U g(x)|J_\phi(x)| d\lambda \geq \int_V g \circ \phi^{-1} d\lambda.$$

Put $g = (f \circ \phi)|J_{\phi^{-1}} \circ \phi|$, then we get

$$\int_U f \circ \phi d\lambda \geq \int_V f|J_{\phi^{-1}}(y)| d\lambda.$$

Hence μ_1, μ_2 are both Radon measures on V with $\int_V f d\mu_1 = \int_V f d\mu_2$ for every $f\colon V \to [0, \infty)$, therefore $\mu_1 = \mu_2$ (since we are interested in compactly supported continuous functions, all our functions are going to be bounded). □

4 Stokes' Theorem

Before proving the change of variables theorem, we were left with the question on how to divide a manifold into pieces where we can integrate. Partitions of unity allow us to do exactly that, but unlike compact spaces, we will now need to consider arbitrary sums. However, we would like to have a finite sum at every point and this leads to the notion of local finiteness.

Definition 3. *A $\mathcal{C}^\infty$ partition of unity on a manifold M is a collection of nonnegative $\mathcal{C}^\infty$ functions $\{\rho_\alpha : M \to \mathbb{R}\}_{\alpha \in A}$ such that*

- *the collection of supports $\{supp(\rho_\alpha)\}_{\alpha \in A}$ is locally finite (every point has a neighbourhood intersecting finitely many supports),*
- $\sum \rho_\alpha = 1$

Given an open cover $\{U_\alpha\}_{\alpha \in A}$ of M, we say that $\{\rho_\alpha\}_{\alpha \in A}$ is subordinate to the open cover $\{U_\alpha\}$ if $supp\rho_\alpha \subset U_\alpha$ for every $\alpha \in A$.

Now, we consider integrating forms over manifolds. On the intersection of two charts, the Jacobian may change the sign of the form thus cancelling the integration, for this reason we will also require our manifold to be oriented.

Definition 4. *An atlas on M is said to be oriented if for any two overlapping charts, the transition map has positive Jacobian everywhere on the intersection. If a manifold has an oriented atlas, then it is called oriented.*

4.1 Paracompactness and Partition of unity

This section is based on [4]. When we have an atlas on a manifold, any point may be in infinitely many charts. If we are to get hold of indicator like functions that are supposed to be partitions of unity, then we would like these indicators to be a finite sum at every point. This leads to the notion of a paracompact space.

Definition 5. *Let X be a topological space. A collection $\mathcal{A}$ of subsets of X is said to be locally finite if every point of X has a neighbourhood that intersects only finitely many elements of $\mathcal{A}$. It is said to be countably locally finite if $\mathcal{A}$ can be written as a countable union of collections $\mathcal{A}_n$, each of which is locally finite.*

Definition 6. *A space X is said to be paracompact if every open covering $\mathcal{A}$ of X has a locally finite open refinement $\mathcal{B}$ that covers X.*

Here an open refinement means that every element of $\mathcal{B}$ is an open set and contained in some element of $\mathcal{A}$. The notion of paracompactness is a generalization of compactness and leads to many interesting properties whose proofs can be found in [4].

Theorem 19. *1. Paracompact Hausdorff spaces are normal (so we may apply Urysohn's lemma).*

2. Closed subspace of a paramcompact space is paracompact.

Theorem 20. *Let X be a regular space. The following are equivalent: Every open covering of X has a refinement that is:*

1. *An open covering of X and countaby locally finite.*
2. *A covering of X and locally finite.*
3. *A closed covering of X and locally finite.*
4. *An open covering of X and locally finite.*

Lemma 11. *Let X be a paracompact Hausdorff space, $\{U_\alpha\}_{\alpha\in J}$ be an indexed family of open sets covering X. Then there exists a locally finite indexed family $\{V_\alpha\}_{\alpha\in J}$ of open sets covering X such that $\overline{V_\alpha} \subset U_\alpha$ for each α.*

Proof. Let $\mathcal{A}$ be the collection of open sets A such that $\overline{A}$ is contained in some element of $\{U_\alpha\}$. By regularity, $\mathcal{A}$ is an open cover of X and by paracompactness, we obtain a locally finite open refinement $\mathcal{B}$ of $\mathcal{A}$ covering X. Index elements of $\mathcal{B}$ by some set K, so that elements of $\mathcal{B}$ are of the form $B_\beta, \beta \in K$, and since $\mathcal{B}$ refines $\mathcal{A}$, the axiom of choice gives us a map $f\colon K \to J$ such that $\forall\, \beta \in K, \overline{B_\beta} \subseteq U_{f(\beta)}$.

For each $\alpha \in J$, let V_α denote the union of elements of

$$\mathcal{B}_a l = \{B_\beta : f(\beta) = \alpha\}.$$

By local finiteness, $\overline{V_\alpha} = \cup_{\mathcal{B}_\alpha} \overline{B_\beta} \subseteq U_\alpha$ (see [4]).

Lastly, given any $x \in X$, choose a neighbourhood intersecting finitely many B_β, say for $\beta = \beta_1, \dots, \beta_k$, then that neighbourhood intersects only thos V_α for which $\alpha = f(\beta_i), 1 \le i \le k$. Thus, $\{V_\alpha\}_{\alpha\in J}$ is a locally finite refinement of $\{U_\alpha\}$ such that $\forall \alpha \in J, \overline{V_\alpha} \subseteq U_\alpha$. □

Theorem 21. *Let X be a paracompact Hausdorff space; let $\{U_\alpha\}_{\alpha\in J}$ be an indexed open covering of X. Then there exists a partition of unity on X dominated by $\{U_\alpha\}$.*

Proof. As in the case of finite partition, we shrink the cover twice so that the support doesn't "spill over". Using the previous lemma, obtain locally finite indexed families of open sets $\{W_\alpha\}, \{V_\alpha\}$ covering X such that

$$\overline{W_\alpha} \subset V_\alpha \subset \overline{V_\alpha} \subset U_\alpha.$$

Since X is normal, we can obtain for each α, a continuous function $\psi_\alpha\colon X \to [0,1]$ such that $|psi_\alpha(W_\alpha) = \{1\}$ and $\psi_\alpha(X \setminus V_\alpha) = \{0\}$. It follows that $\mathrm{supp}(\psi_\alpha) \subseteq \overline{V_\alpha} \subset U_\alpha$ and therefore is also locally finite. Furthermore, since $\{W_\alpha\}$ cover X, at each x there is at least one ψ_α which is non zero. By local finiteness of the supports, it is possible to make sense of the formally infinite sum

$$\Psi(x) = \sum_\alpha \psi_\alpha(x).$$

Again by local finiteness, Ψ is continuous. Taking $\phi_\alpha = \psi_\alpha/\Psi$ gives us our partition of unity. □

Now while we have obtained a continuous partition of unity, when dealing with manifolds we will need a smooth partition of unity. It is possible to find a smooth partition of unity and we refer the reader to [5], but the basic idea is to first obtain a smooth bump function over comapct subsets of $\mathbb{R}$. Using this, we obtain bump functions over compact subsets of $\mathbb{R}^n$. Since manifolds are second countable, and locally Euclidean, we can "transfer" these bump functions to the manifold. Again, one obtains a locally finite collection of smooth functions adding to a positive number everywhere and as above one can get a partition of unity.

Bump functions have another use, namely to extend functions defined on some subspace to the entire space, like the Tietze Extension Theorem.

4.2 Integration and Stokes' Theorem

To integrate a form on a manifold, we will first obtain a partition of unity whose supports are contained inside charts. Then, we can pull the forms onto the familiar Euclidean space, integrate the form and add up the results. By the change of variables theorem, the integral doesn't depend on the chart used, however it is not obvious that the integration so obtained is independent of the chosen partition of unity. Indeed, it is independent and the reader is refered to [5] for a proof.

Now, from basic Euclidean analysis, one is familiar with results like the Stokes' theorem which essentially say

$$\int_M d\omega = \int_{\partial M} \omega$$

where M is some n-dimensional manifold and ω an $n-1$-form. Historically, $\partial M, M$ would have been a closed curve bounding its interior in the two plane (Green's Theorem), closed curve bounding a surface in three space (Stoke's Theorem), surface bounding a volume in three space (Gauss Divergence Theorem). We prove this result for arbitrary dimensions. As one expects, using the partition of unity, the task is simplified to proving the statement in ordinary Euclidean space.

We follow [5]. Let M be an oriented manifold with oriented atlas $\{U_\alpha, \phi_\alpha\}$. Denote by $\Omega_C^k(M)$ the vector space of all smooth k-forms with compact support. If (U, ϕ) and (U, ψ) are charts on the same open set and having positive Jacobian, then the integral of the pullback of any $\omega \in \Omega_C^k(M)$ over $\phi(U)$ is the same as that over $\psi(U)$.

Choose a smooth partition of unity $\{\rho_\alpha\}$ subordinate to $\{U_\alpha\}$, all but finitely many $\rho_\alpha \omega$ are zero because ω has compact support. Define

$$\int_M \omega = \sum_\alpha \int_{U_\alpha} \rho_\alpha \omega.$$

One can show that this integral is independent of the oriented atlas and the partition of unity.

Theorem 22. *(Stokes' Theorem on manifolds) For any smooth $(n-1)$-form ω with compact support on an oriented n manifold M,*

$$\int_M d\omega = \int_{\partial M} \omega.$$

Proof. Choose an atlas $\{(U_\alpha, \phi_\alpha\}$ for M in which each U_α is diffeomorphic to either $\mathbb{R}^n$ or $\mathbb{H}^n$ (upper half plane) via an orientation preserving diffeomorphism. Let $\{\rho_\alpha\}$ be a smooth partition of unity subordinate to $\{U_\alpha\}$. Suppose Stokes' Theorem holds for $\mathbb{R}^n, \mathbb{H}^n$, then it also holds for U_α and observe that $(\partial M) \cap U_\alpha = \partial U_\alpha$, including orientations. Therefore,

$$\begin{aligned}
\int_{\partial M} \omega &= \int_{\partial M} \sum_\alpha \rho_\alpha \omega \\
&= \sum_\alpha \int_{\partial M} \rho_\alpha \omega \\
&= \sum_\alpha \int_{\partial U_\alpha} \rho_\alpha \omega \\
&= \sum_\alpha \int_{U_\alpha} d(\rho_\alpha \omega) \\
&= \sum_\alpha \int_M d(\rho_\alpha \omega) \\
&= \int_M d(\sum_\alpha \rho_\alpha \omega) \\
&= \int_M d\omega
\end{aligned}$$

where we have frequently used the fact that $\sum \rho_\alpha = 1$ and the supports are locally finite (that allows us to freely interchange $d, \sum$ for example). So, it suffices to prove Stokes' Theorem for $\mathbb{R}^n, \mathbb{H}^n$. $\square$

We prove the theorem on open sets following [1].

Lemma 12. *Let $k > 1, \eta$ be a $k-1$ form defined on an open set containing the unit cube I^k. Assume that η vanishes at all points of ∂I^k except possibly at points in $IntI^{k-1} \times 0$. Then*

$$\int_{IntI^k} d\eta = (-1)^k \int_{IntI^{k-1}} i^* \eta$$

where $i^*\eta$ is the pullback of η to the lower face under the usual inclusion.

Proof. By linearity of integrals and pullbacks, it suffices to prove the theorem when $\eta = f dx_I$ where I is some ascending $k-1$-tuple, say $dx_I = dx_1 \wedge \cdots \wedge \widehat{dx_j} \wedge \cdots \wedge dx_k$. Then $d\eta = (-1)^{j-1} D_j f dx_1 \wedge \cdots \wedge dx_k$ where $D_j f = \partial f / \partial x_j$.

Then we compute

$$\begin{aligned}\int_{\mathrm{Int} I^k} d\eta &= (-1)^{j-1} \int_{\mathrm{Int} I^k} D_j f \\ &= (-1)^{j-1} \int_{I^k} D_j f \\ &= (-1)^{j-1} \int_{I^{k-1}} \int_{x_j \in I} D_j f (\text{ Fubini's theorem }) \\ &= (-1)^{j-1} \int_{I^{k-1}} (f(x_1, \ldots, 1, \ldots, x_k) - f(x_1, \ldots, 0, \ldots, x_k)\end{aligned}$$

where the $1, 0$ appear in the jth coordinate. By assumption, since f vanishes on the boundary faces except when $x_k = 0$, we get

$$\int_{I^{k-1}} d\eta = \begin{cases} 0 & \text{if } j < k \\ (-1)^k \int_{I^{k-1}} f \circ i & \text{if } j = k \end{cases}$$

This is exactly what we get under the pullback to different faces, i.e., different $x_j = 0$. The theorem follows. □

Remark. Although we have proved it for a unit cube, the proof of the general Stokes' theorem remains true for we may choose our charts to be diffeomorphic to bounded open sets.

References

[1] Analysis on Manifolds, Munkres

[2] Measure and integral(PDF), E. Kowalski

[3] An Introduction to Measure Theory, Terence Tao

[4] Topology, Munkres

[5] An Introduction to Manifolds, Loring W. Tu

[6] For a more general version of the inverse function theorem that doesn't require continuity of the derivative, see Terence Tao's blog and this Math Overflow post.