These notes are based on Analysis on Manifolds by James R Munkres.

Until we get to the inverse function theorem, we use the operator norm a lot, so until then vertical bars means operator norm and determinant will be denoted with Det.

Note that it is pretty annoying that there are so many definitions of integrals. For this course we use the riemann integral. The Lebesgue integral was needed for our work on probability but we do not need it for this.

Lemma 1: Let f be a function from $\mathbb{R}^m \to \mathbb{R}^n$. We say f is continuously differentiable in an open set U if for each x in U, the matrix Df exists, and for each ε>0 we have that $|u - v| < \delta \Rightarrow |Df(u) - Df(v)| < \varepsilon$ using the operator norm for some δ>0. Then f is continuously differentiable in U if and only if all its partial derivatives are continuous in U.

Proof:

Let e be the standard basis for $\mathbb{R}^n$ and b the standard basis for $\mathbb{R}^m$. Then suppose f is continuously differentiable. Then we have

$$\left|\frac{\delta x_i}{\delta b_j} - \frac{\delta y_i}{\delta b_j}\right| = \left|D_{ji}(x) - D_{ji}(y)\right|$$

And $\left|D_{ji}(x) - D_{ji}(y)\right| = \left|\left(Df(x) - Df(y)\right)e_j . b_i\right| \leq \left|\left(Df(x) - Df(y)\right)e_j\right| \leq \left|\left(Df(x) - Df(y)\right)\right|$

By definition of the operator norm and the fact that projecting something onto a basis vector decreases the length. Hence, if Df(x) and Df(y) are within ε for some δ, then so are each of the partial derivatives.

Conversely, suppose each of the partial derivatives are continuous. Then by differential equations the derivative Df exists so we just need to show that it is continuous. We will take screenshots from other cambridge notes here, recalling vectors and matrices knowledge. The double bars here mean the operator norm/length of a vector.

For any linear map $A \in L(\mathbb{R}^n; \mathbb{R}^m)$ represented by (a_{ij}) so that $A\mathbf{h} = a_{ij}h_j$, then for $\mathbf{x} = (x_1, \cdots, x_n)$, we have

$$\|A\mathbf{x}\|^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} x_j \right)^2$$

By Cauchy-Schwarz, we have

$$\begin{aligned} &\leq \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij}^2 \right) \left(\sum_{j=1}^{n} x_j^2 \right) \\ &= \|\mathbf{x}\|^2 \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2. \end{aligned}$$

Taking square roots, using $|Ax| \leq |A||x|$ and dividing by |x| gives

$$\|A\| \le \sqrt{\sum\sum a_{ij}^2}.$$

Applying this to $A = D\mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{y})$, we get

$$\|D\mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{y})\| \le \sqrt{\sum\sum (D_j f_i(\mathbf{x}) - D_j f_i(\mathbf{y}))^2}.$$

Thus since all partial derivatives are continuous, it is easy to see that if we make the distance between x and y small enough, the operator norm of the derivative difference will get as small as we want so our function is continuously differentiable.

Lemma 2 (Contraction mapping theorem in $\mathbb{R}^n$): We define $f: \mathbb{R}^n \to \mathbb{R}^n$ to be a contraction if there is a $0 \le \lambda < 1$ such that it is always the case that $|f(x) - f(y)| \le \lambda |x - y|$. Then if f is a contraction, there is a unique x such that $f(x) = x$.

Proof:

First note that if there is an x such that $f(x) = x$ it is unique since if we also had $f(y) = y$ then we know that $|x - y| = |f(x) - f(y)| \le \lambda |x - y|$, so $|x - y| = 0$ so $x = y$.

Note also that since Cauchy sequences in $\mathbb{R}$ converge (See Analysis I notes), it is intuitive that a cauchy sequence in $\mathbb{R}^n$ defined by "elements get arbitrarily close to eachother" is the same as being cauchy in each component, and convergence component-wise is the same as convergence distance-wise, so cauchy sequences in $\mathbb{R}^n$ converge.

To prove existence, we must pick x_0 and keep applying f, ie we define $x_{n+1} = f(x_n)$. Then this sequence is cauchy, because $|x_{n+1} - x_n| = |f(x_n) - f(x_{n-1})| \le \lambda |x_n - x_{n-1}|$ so by induction we know that $|x_{n+1} - x_n| \le \lambda^n |x_1 - x_0|$. Since this is true for all n, we have, by the triangle inequality

$$|x_m - x_n| \le \sum_{j=n}^{m-1} |x_{j+1} - x_j| \le \sum_{j=n}^{m-1} \lambda^j |x_1 - x_0| = \frac{\lambda^n}{1 - \lambda} |x_1 - x_0| \to 0$$

By geometric series. Therefore this converges to some x. Since it is easy to see that f is continuous by definition, and $f(x_n) = x_{n+1}$, we may take the limit on both sides to get that f(x)=x. This should remind you of the work we did in level 4 with cobweb diagrams.

The only assumption we used is that cauchy sequences converged, so the function can be on any such domain provided its image is contained in that domain, not only $\mathbb{R}^n$. In particular, it works on closed balls in $\mathbb{R}^n$, since cauchy sequences there converge to some point in $\mathbb{R}^n$, but closed balls have the property that the limit of any sequence of points in the closed ball is in the closed ball (This is because any point outside the closed ball is not on the boundary and thus has a ball around it of points outside the closed ball so if the sequence converged to that point since there is a ball around that point the sequence could never go into this is a contradiction), so cauchy sequences in closed balls converge to a point in the closed ball.

Lemma 3: Suppose $f: [a, b] \to \mathbb{R}^n$ is continuous on [a,b] and differentiable on (a,b), and there is M such that $|Df(t)| \le M$ for all t in (a,b). Then $|f(b) - f(a)| \le M(b - a)$. Furthermore, this lemma holds if the domain is an open ball in $\mathbb{R}^m$.

Proof: Note that the operator norm for 1*n matrices is the same as the length of that matrix as a vector – if we dot product it by a unit vector then the maximum result is the length of the original vector, achieved when they are linearly dependent.

We will prove the first part of the lemma and then reduce the second part to the first part.

Let $v := f(b) - f(a)$. Define $g(t) := v.f(t) = \sum_{i=1}^{m} v_i f_i(t)$. Since each f is differentiable and the v's are constant, $g'(t) = \sum_{i=1}^{m} v_i f_i{}'(t)$. By the cauchy schwartz inequality,

$$|g'(t)| = \left|\sum_{i=1}^{m} v_i f_i{}'(t)\right| \leq |v| \left|\sum_{i=1}^{m} f_i'^2(t)\right|^{\frac{1}{2}} = |v||Df(t)| \leq M|v|$$

Applying the ordinary mean value theorem to g gives

$$g(b) - g(a) = g'(t)(b-a)$$

For some t in (a, b). By definition of g, $v.\big(f(b) - f(a)\big) = g'(t)(b-a)$. Then by definition of v,

$$|f(b) - f(a)|^2 = |g'(t)(b-a)| \leq (b-a)|g'(t)| \leq M(b-a)|f(b) - f(a)|$$

Since the result is trivial if $f(b) - f(a) = 0$, we can divide through by $|f(b) - f(a)|$ to get the result.

Now let $a \in \mathbb{R}^m$ and $f: B_r(a) \to \mathbb{R}^n$ be differentiable with $|Df(x)| \leq m$.

Fix $b_1, b_2 \in B_r(a)$. Then $tb_1 + (1-t)b_2$ is the line that is b_2 when t=0 and b_1 when t=1. Since we are talking about a ball here and balls are clearly convex meaning a line between 2 points in the ball is in the ball, it means that for t in [0,1] $tb_1 + (1-t)b_2 \in B_r(a)$. Now consider $g(t) := f(tb_1 + (1-t)b_2)$, then g is differentiable by the chain rule with derivative $Df(tb_1 + (1-t)b_2)(b_1 - b_2)$ and therefore by the $|Ax| \leq |A||x|$ property of operator norms, $|Dg(t)| \leq |Df(tb_1 + (1-t)b_2)||b_1 - b_2| \leq M|b_1 - b_2|$. By the first part of the lemma, $|f(b_1) - f(b_2)| = |g(1) - g(0)| \leq M|b_1 - b_2|$ so done.

Definition: We define a set A is open relative to, or in, another set B if each point in A has the property that it has an open ball around it such that the intersection of that open ball with B is contained in A.

Lemma 4: If a function is continuous then pre-images of open sets in the codomain space are open in the domain space.

Let W be an open subset of the codomain, then we want to show $f^{-1}(W)$ is open. For each ε>0 small enough such that W contains $B_\varepsilon\big(f(x)\big)$ (possible since W is open), there is δ>0 such that $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$, so $f(y) \in B_\varepsilon\big(f(x)\big) \subseteq W$ if $y \in B_\delta(x)$. Since this is true for each y in $B_\delta(x)$ we have that $f\big(B_\delta(x)\big) \subseteq W$. Therefore we have that $B_\delta(x) \subseteq f^{-1}\big(B_\varepsilon(f(x))\big)$, and since x was an arbitrary element in the pre-image of W, this means an open ball around x is contained in $f^{-1}\big(B_\varepsilon(f(x))\big)$ which is in $f^{-1}(W)$, so $f^{-1}(W)$ is open due to every arbitrary point having an open ball around it contained in $f^{-1}(W)$.

Note: A function with a continuous inverse only has the reverse implication (images of open sets are open) if it actually has a continuous inverse, ie it is a bijection and we are talking about openness in the codomain space, and we cannot just expand to larger spaces.

Lemma 5 (Inverse function theorem): Let f be a continuously differentiable function $\mathbb{R}^n \to \mathbb{R}^n$. Suppose $U \in \mathbb{R}^n$ is open. Let $f: U \to \mathbb{R}^n$ be continuously differentiable. Suppose that for some a in U Df(a) is an invertible matrix. Then there exist open sets V and W in $\mathbb{R}^n$ such that V is a subset of U a is in V, f(a) is in W, and W is the image of f(V), and f from V to W is a bijection with a continuously differentiable inverse. Furthermore, for each y in W, $Df^{-1}(y) = Df\left(f^{-1}(y)\right)^{-1}$.

Proof:

Lets replace f with its transformation by an invertible linear map, namely $\left(Df(a)\right)^{-1}f$. Since linear maps that are invertible are bijections and linear maps have a constant derivative, then by the chain rule this is a continuously differentiable bijection if f is. Therefore we will replace $Df(a)$ with I, or rotate and stretch our heads a bit. By continuity of Df, there is an open ball around a with radius r>0 such that for each x in this ball, $|Df(x) - I| < \frac{1}{2}$. Since there is an open ball around a contained in U by openness of U, we shall make r small enough such that our open ball around a is contained in U. We will denote the open ball with radius r around a as $B_r(a)$ and similarly for other open balls, and we will let $W = B_{\frac{r}{2}}f(a), V = f^{-1}(W) \cap B_r(a)$, where f^{-1} means pre-image. However, we need to prove V is open.

Note that a finite intersection of open sets is open – for each point in the intersection there are open balls around that point such that they are contained in each of the sets in the intersection, and the intersection of these open balls is an open ball, so an open ball around any point in the intersection is in the intersection and thus the intersection is open. So it remains to show that $f^{-1}(W)$ is open. To do this, use the previous lemma.

We want to show that $f: V \to W$ is a bijection. To do this we will show that for each y in W there is a unique x in V such that f(x)=y. Consider the function $T(x) = x - f(x) + y$, then if T(x)=x this means that f(x)=y so we want to show that T(x) has a unique point x in V such that T(x)=x, and to do this we will use the contraction mapping theorem.

Let $h(x) = x - f(X)$, then $Dh(x) = I - Df(x)$. By how we defined R, we know $|Dh(x)| \leq \frac{1}{2}$. Therefore by lemma 3 we know $|h(x_1) - h(x_2)| \leq \frac{1}{2}|x_1 - x_2|$ inside $B_r(a)$. Since y is independent of x, we know that $|T(x_1) - T(x_2)| \leq \frac{1}{2}|x_1 - x_2|$. T is defined on the closed ball $\overline{B_r(a)}$ – the line on top means the ball is closed and not open, so the contraction mapping theorem gives us the desired result if we can show that the codomain of T is contained in $\overline{B_r(a)}$. But this is immediate, because

$$\begin{aligned}|T(x) - a| = |x - f(x) + y - a| &= \left|x - f(x) - \left(a - f(a)\right) + \left(y - f(a)\right)\right| \\ &\leq |h(x) - h(a)| + |y - f(a)|\end{aligned}$$

Since y is in W which is by definition $B_{\frac{r}{2}}f(a)$, and $|h(x) - h(a)| \leq \frac{1}{2}|x - a| < \frac{1}{2}r$ by above and the fact that x is in $B_r(a)$, we know that $|T(x) - a| \leq |h(x) - h(a)| + |y - f(a)| < \frac{r}{2} + \frac{r}{2} = r$, which means the range of T is contained in $\overline{B_r(a)}$ so f is a bijection by the contraction mapping theorem.

Therefore the only thing that remains to prove is that the inverse of f is continuously differentiable.

Let $g = f^{-1}: W \to V$ which we now know exists. Then by the triangle inequality,

$$|x_1 - x_2| \leq |f(x_1) - f(x_2)| + \left|\left(x_1 - f(x_1)\right) - \left(x_2 - f(x_2)\right)\right|$$

Therefore

$$|x_1 - x_2| - |f(x_1) - f(x_2)| \leq \left|\left(x_1 - f(x_1)\right) - \left(x_2 - f(x_2)\right)\right| = |h(x_1) - h(x_0)| \leq \frac{1}{2}|x_1 - x_2|$$

Therefore $|f(x_1) - f(x_2)| \geq \frac{1}{2}|x_1 - x_2|$

Therefore $|x_1 - x_2| = \left|f\left(g(x_1)\right) - f\left(g(x_2)\right)\right| \geq \frac{1}{2}|g(x_1) - g(x_2)|$

Therefore $|g(x_1) - g(x_2)| \leq 2|x_1 - x_2|$ so g is continuous.

Note that by the chain rule, since $f(g(y)) = y, I = Df\left(g(y)\right)Dg(y)$ so if g were differentiable its derivative would be given by $Df\left(g(y)\right)^{-1}$ exactly as in the statement of the lemma, and also dg would be continuous as $Df\left(g(y)\right)$ is continuous due to being the composition of continuous functions. So it remains to show that $Df\left(g(y)\right)^{-1}$ is actually the derivative of g, ie that g is differentiable. We first need to check that $Df(x)$ is invertible for every x in $B_r(a)$. Since, $|Df(x) - I| \leq \frac{1}{2}, Df(x)v = 0$ implies that $|v| = |Df(x)v - v| \leq |Df(x) - I||v| \leq \frac{1}{2}|v|$ by the standard operator norm property we always use and thus v would be 0, so Df has trivial kernel and thus is invertible so the inverse actually exists.

Let $h \coloneqq g(y + k) - g(y)$, in other words $f(x + h) - f(x) = k$ for k small. Since g is injective on an open set W, k can be made small enough such that $B_k(y) \in W$ and thus small enough that whenever |k| is not 0, h is also not 0. Since g is continuous, h goes to 0 as |k| goes to 0. To check the definition of a derivative, we need to check that

$\frac{h - Df\left(g(y)\right)^{-1}k}{|k|}$ goes to 0 as k goes to 0 then we will be done.

$$\frac{h - Df\left(g(y)\right)^{-1}k}{|k|} = \frac{Df(x)^{-1}(Df(x)h - k)}{|k|} = -Df(x)^{-1}\frac{(k - Df(x)h)}{|k|}$$
$$= -Df(x)^{-1}\frac{f(x + h) - f(x) - Df(x)h}{|h|}\frac{|h|}{|k|}$$
$$= -Df(x)^{-1}\frac{f(x + h) - f(x) - Df(x)h}{|h|}\frac{|g(y + k) - g(y)|}{|k|}$$

As k goes to 0, so does h. The first factor is constant for any fixed x, and the third factor is bounded by 2 by above, so we just need to show that $\frac{f(x+h) - f(x) - Df(x)h}{|h|}$ goes to 0. But f is differentiable and this is literally the definition of the derivative Df(x) that that goes to 0. So done.

Note that it is only in 1 dimension that derivative invertible on the whole domain implies inverse on that domain, here we only know there is a local inverse.

Lemma 6 (Implicit function theorem): Let A be open in $\mathbb{R}^{k+n}$ and let $f: A \to \mathbb{R}^n$ be continuously differentiable. Write f in the form f(x,y) with $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^n$. Suppose that (a,b) is a point in A such that $f(a, b) = 0$ and Det $\left(\frac{\delta f}{\delta y}\right)(a, b) \neq 0$. Then there is a neighbourhood B of a (open set containing a) in $\mathbb{R}^k$ such that $g: B \to \mathbb{R}^n$, $g(a) = b$ and $f\left(x, g(x)\right) = 0$ and g is continuously differentiable and there is a unique g with this property.

Before we prove this, we will give an example to explain what this is saying.

Suppose $n = 2, k = 1$ so x=(u) and y=(v,w), $f(x,y) = (u^2 + v^2 - 1, u^2 + w^2 - 1)$, and A is $B_2(0,0)$. Then at $(a,b) = (0.6, 0.8, 0.8)$: $a = (0.6), b = (0.8, 0.8)$, $\mathrm{Det}\left(\frac{\delta f}{\delta y}\right)(a,b) =$

$$\mathrm{Det}\begin{pmatrix} \frac{\delta(u^2+v^2-1)}{\delta v} & \frac{\delta(u^2+v^2-1)}{\delta w} \\ \frac{\delta(u^2+w^2-1)}{\delta v} & \frac{\delta(u^2+w^2-1)}{\delta w} \end{pmatrix}\big(0.6,(0.8,0.8)\big) = \mathrm{Det}\begin{pmatrix} 2v & 0 \\ 0 & 2w \end{pmatrix}\big(0.6,(0.8,0.8)\big) = \mathrm{Det}\begin{pmatrix} 1.6 & 0 \\ 0 & 1.6 \end{pmatrix} \neq 0$$

Now the theorem claims that there is a g from $\mathbb{R} \to \mathbb{R}^2$ such that $g(0.6) = (0.8, 0.8), f\big(x, g(x)\big) = 0$. Lets solve for this directly, $(u^2 + g_1^2 - 1, u^2 + g_2^2 - 1) = 0$, hence $g = \left(\sqrt{1-u^2}, \sqrt{1-u^2}\right)$ works in this case, not the negative square root by the $g(0.6) = (0.8, 0.8)$ condition. g here is continuously differentiable.

Proof:

Define $F(x,y) \coloneqq \big(x, f(x,y)\big)$ as a function from $\mathbb{R}^{k+n} \to \mathbb{R}^{k+n}$. Then the matrix DF in block form is given by

$$\begin{pmatrix} \frac{\delta x}{\delta x} & \frac{\delta x}{\delta y} \\ \frac{\delta f}{\delta x} & \frac{\delta f}{\delta y} \end{pmatrix} = \begin{pmatrix} I_k & 0 \\ \frac{\delta f}{\delta x} & \frac{\delta f}{\delta y} \end{pmatrix}$$

By vectirs and matrices the determinant of this is $Det\left(\frac{\delta f}{\delta y}\right)$ which is not 0 by assumption. Therefore we can apply the inverse function theorem to F. Since $f(x,y) = 0$ at the point (a,b) by assumption, it means that F is one to one in an open set U×V containing (a,0) and the inverse function to F is continuously differentiable. We have $(x,y) = G\big(x, f(x,y)\big)$. G preserves the first k coordinates and therefore G applied only to the last n coordinates has an inverse h by the inverse function theorem so we can write $G(x,z) = \big(x, h(x,z)\big)$ for continuously differentiable h. $G(x,0) = \big(x, h(x,0)\big)$ and since G is an inverse to F we have $(x,0) = F\big(x, h(x,0)\big) = \Big(x, f\big(x, h(x,0)\big)\Big)$ by definition of F, and since the first k coordinates are preserved, we have $0 = f\big(x, h(x,0)\big)$. Set $g(x) = h(x,0)$ then we are done (almost). We have that $(a,b) = F^{-1}(a,0) = G(a,0) = \big(a, h(a,0)\big)$ so $g(a) = b$. g is unique because we know that any function satisfying the conditions of g is b at a. Call the open interval in question that g's range is V. Since g is continuous, there is a neighbourhood B around a such that everything in B is mapped into V by g. If there was another function k satisfying the conditions of g we would have that $F\big(x, k(x)\big) = (x,0)$, so $\big(x, k(x)\big) = G(x,0) = \big(x, h(x,0)\big) = \big(x, g(x)\big)$ so k and g agree on B. Therefore there is a neighbourhood around a (in fact it is B) such that we have a unique g with this property, so done.

Definition: A subset of $\mathbb{R}^n$ has measure 0 if it can be covered by open sets whose total n-dimensional volume is less than ε for any ε>0. Intuitively this means "volume 0".

Lemma 7: Recall that a function is riemann integrable if overestimating or underestimating it with rectangles lets us get as close as we want and we proved that in 1D this is satisfied for continuous functions. Let Q be a rectangle in $\mathbb{R}^n$ and let $f: Q \to \mathbb{R}$ be a bounded function. Let D be the set of points of Q such that f is not continuous. Then f is integrable over Q if and only if D has measure 0.

Proof:

Pick M such that |f(x)|<M for x in Q (possible by boundedness).

First, assume D has measure 0. Fix ε>0 and define $\varepsilon' := \frac{\varepsilon}{2M+2vol(Q)}$. Cover D by open rectangles $Q_1, Q_2, Q_3, \ldots$ with total volume less than ε', possible as D has measure 0. For each point a in Q\D pick an open rectangle Q_a containing a such that (by continuity of f at a) $|f(x)-f(a)| < \varepsilon'$ for $x \in Q_a \cap Q$. Since Q is closed and bounded we can pick a finite subcover $Q_1, Q_2, Q_3, \ldots Q_k, Q_{a_1}, Q_{a_2}, \ldots, Q_{a_n}$ that covers Q. The first k may not cover D but it turns out that that will not matter. Set $Q_j' := Q_{a_j}$, then we have a cover of k+n rectangles of Q by $Q_{1\ldots k}, Q_{1\ldots n}'$, where $Q_{1\ldots k}$ has total volume less than ε'. The rectangles Q_j' satisfy $|f(x)-f(y)| \leq |f(x)-f(a_j)| + |f(a_j)-f(y)| \leq 2\varepsilon'$ for x,y in $Q_j' \cap Q$. For convenience replace each rectangle by its intersection with Q. Now using the end points of each rectangle define a partition of Q into a grid such that we cut it at each coordinate, ie each rectangle is made up of grid lines. Divide the rectangles in the partition into R so that each rectangle in R lies in one of the Q rectangles and R' such that each rectangle in R' lies in one of the Q' rectangles, or alternatively R' is everything not in R. Then we have

$$\sum_{r\in R} \max_r(f) - \min_r(f)\, vol(r) \leq 2M \sum_{r\in R} vol(r)$$

$$\sum_{r\in R'} \max_r(f) - \min_r(f)\, vol(r) \leq 2\varepsilon' \sum_{r\in R} vol(r)$$

Both by definition.

Now $\sum_{r\in R} vol(r) < \varepsilon'$ because the total volume of the Q rectangles is less than that by assumption, and $\sum_{r\in R} vol(r) \leq vol(Q)$. Thus

$$\sum_{r\in P} \max_r(f) - \min_r(f)\, vol(r) = 2M \sum_{r\in R} vol(r) + 2\varepsilon' \sum_{r\in R} vol(r) \leq 2M\varepsilon' + 2Q\varepsilon' = \varepsilon$$

So done.

Now we want to prove that if our function is integrable then D has measure 0.

Define the oscillation of f at a to be $\lim_{\delta\to 0}\left(\max_{B_\delta(a)}(f) - \min_{B_\delta(a)}(f)\right)$. Then this is clearly 0 if and only if f is continuous by definition of continuity and definition of the limit. Now we will prove the converse – f riemann integrable implies D has measure 0.

Let D_m be the set of points where f has oscillation $\geq \frac{1}{m}$, then D is the union of all such D_m. Therefore if each D_m has measure 0 then we can cover it by open sets of total volume $\varepsilon 2^{-m}$ so that combining these covers gives a cover of D by open sets of total volume ε so we just need to show each D_m has measure 0. Fix any m and ε>0. Pick any partion P of Q for which the difference between its upper estimate and lower estimate for the integral is within $\frac{\varepsilon}{2m}$, possible as f is integrable. Let D_m' be the intersection of D_m with the boundary of P and $D_m'' = D_m - D_m'$. D_m' has measure 0 as it is literally the boundary of rectangles so we can certainly cover it with total volume less than $\frac{\varepsilon}{2}$ by open sets. For D_m'' let $R_{1,\ldots,k}$ be the rectangles containing points of D_m''. Given i, R_i has a point in D_m'', and since this point (a) is not on the boundary of a rectangle there is a δ such that the cube with side length 2δ centered around a is contained in R_i. Now by definition of oscillation $\frac{1}{m} \leq \max_c(f) - \min_c(f) \leq \max_{R_i}(f) - \min_{R_i}(f)$

so $\sum_{i=1}^{k}\left(\max_{R_i}(f)-\min_{R_i}(f)\right)vol(R_i)\le\sum_{i=1}^{k}\frac{vol(R_i)}{m}<\frac{\varepsilon}{2m}$ so those rectangles have total volume less than $\frac{\varepsilon}{2}$ so done.

Remark: If f is non-negative and the integral of f over Q is 0 then f is continuous everywhere but a set of measure 0, and if f were >0 somewhere where it is continuous it would be >0 in a neighbourhood so the integral would be >0, so if the integral of a non-negative function over a rectangle Q is 0 then f is 0 everywhere on Q except a set of points of measure 0. Also if f vanishes except on a point of measure 0 then its integral exists, and on any rectangle in a partition containg this measure 0 set, it contains a point not in that measure 0 set where the function is 0, and thus 0 is between its overestimate and underestimate, so the overestimate and underestimate of all partitions in the riemann integral has 0 between them so if the integral exists it is 0.

Definition: Let S be a bounded set in a rectangle $Q\in\mathbb{R}^n$. Suppose S has a boundary with measure 0, then the integral over S of a function f(x) is the integral over Q of $f(x)1_S(x)$, which is integrable if f is integrable and $1_S(x)$ is integrable, hence why we want the boundary to have measure 0, so no weird fractal curves with boundary that has area. Write $f_S(x)\coloneqq f(x)1_S(x)$

Definition: A set is rectifiable if it is bounded and its boundary has measure 0, and for such a set its volume is equal to the integral of 1 over that set.

Definition: A set is compact if it is closed and bounded. Sometimes compact is defined as "Every cover has a finite subcover" or "Every sequence has a convergent subsequence" but until future analysis courses we stick with "Closed and bounded" and note that we know from earlier work that closed and bounded implies the other 2 definitions.

In the past we have defined improper integrals as the limit. We are now going to define the extended integral. You will see from this definition that it agrees with the idea of an improper integral being a limit and that the lebesgue integral is actually more general than even this.

Definition: If f is non-negative on a set A then we define the extended integral of f over A to be the supremum of $\int_D f$ as D ranges over all compact rectifiable subsets of A when it exists. If f is not non-negative define it as the sum of the integral of the positive parts minus the integral of minus the negative parts. f is therefore integrable in the extended sense if and only if we have absolute convergence.

Lemma 8: Let S be a subset of $\mathbb{R}^n$ and let f,g be functions from S to $\mathbb{R}$. Then define the two functions $F(x)\coloneqq\max(f(x),g(x))$ and $G(x)\coloneqq\min(f(x),g(x))$. Then if f and g are both continuous at x_0 so are F and G, and this implies that if they are both integrable over S so are f and g as their discontinuities are just the union of 2 measure 0 sets by integrability of f and g. In particular if g=0 this allows us (for bounded functions on bounded domains) to split the integral into positive and negative parts.

Proof: Consider first the case where $f(x_0)=g(x_0)=r$. Fix ε>0 and pick by continuity a δ>0 such that $|f(x)-r|<\varepsilon$ and $|g(x)-r|<\varepsilon$ for $|x-x_0|<\delta$ and x∈S. It then follows automatically by definition of F and G that $|F(x)-F(x_0)|<\varepsilon$ and similarly for G so they are continuous. On the otherhand, suppose without loss of generality that $f(x_0)>g(x_0)$, then since f and g are both continuous we can find a neighbourhood such that $f(x_0)>g(x_0)$ still holds and F and G take on f and g respectively, thus F and G are continuous at x_0 by continuity of f and g.

Lemma 9: For any rectangle Q in $\mathbb{R}^n$ there exists an infinitely differentiable non-negative function $\mathbb{R}^n \to \mathbb{R}$ that is non-zero if and only if the input is in Q.

Proof: If n=1, then $f(x) = \begin{cases} 0: x \leq 0 \\ e^{-\frac{1}{x}}\ otherwise \end{cases}$ is infinitely differentiable at everywhere but 0, and at 0 one checks by a l'hoptial's rule argument that it is also differentiable. Then $f(x)f(1-x)$ has the desired property on (0,1) and can be scaled appropriately. We can take products of scaled versions of this function on the different coordinates to get the result we want.

We will now prove a lemma that seems quite useless but it will be important later.

Lemma 10: Let A be the union of a collection of open sets O in $\mathbb{R}^n$. Then there exists a countable collection of rectangles Q contained in A such that their interiors cover A, each one is contained in one of the O's, and each point of A has a neighborhood (Neighborhoods are open in maths by the way) that intersects only finitely many of the rectangles Q.

Proof: Note that an infninite union of open sets is open. It is possible to find a sequence of compact sets $D_1, D_2, \ldots$ such that each is contained inside the interior of the next one and their union is A. This is because if $D_n = \{|x| \leq n\} \cap \{B_{\frac{1}{n}}(x) \in A\}$ then these are each bounded by being in $\{|x| \leq n\}$ and closed as they are the intersection of a closed set with $\{B_{\frac{1}{n}}(x) \in A\}$ which is closed as it is the complement of a set that is open: The complement is open due to being the union of infinitely many open balls around the boundary of A. These can be made to be strictly included in eachother if we replace it with $\{|x| \leq C\} \cap \{B_{\frac{D}{n}}(x) \in A\}$ where C is large enough that D_1 is non-empty and D is small enough that an open ball of radius D is in A. The union of all D_n is exactly equal to A as it is equal to $\{|x| \leq \infty\} \cap \{B_{>0}(x) \in A\}$ and A is open.

So pick such a sequence of sets and define $B_i \coloneqq B_i - Interior(B_{i-1})$, then this is the intersection of a compact set with the complement of an open set which is thus compact. Also, B_i is disjoint from D_{i-2} as D_{i-2} is in the interior of D_{i-1} which is not in B_i. Therefore for each $x \in B_i$ we can pick a cube centered at x that is contained in one of the elements of A and disjoint from D_{i-2}. Since B_i is compact, finitely many of these cubes cover it. For each i, pick such a finite cover, and now we have a countable collection of cubes contained in one of the O's, and each point of A is in $Int(D_i)$ and since $Int(D_i)$ is open it has a neighbourhood contained in $Int(D_i)$ for some i so this neighbourhood covers cubes that lie in one of the first i+1 finite collections so all conditions are satisfied.

We will now give a seemingly useless definition that will seem useful later.

Definition: Let A be the union of a collection of open sets in $\mathbb{R}^n$. A partition of unity is a sequence of infinitely differentiable non-negative functions $\phi_i(x)$ with the following properties:

1. They are 0 except on a set of points with the property that this set and its boundary (which we call the support of ϕ_i) is compact and contained in an element of A.
2. Each point of A has a neighborhood that intersects only finitely many supports of ϕ_i
3. $\sum_{i=1}^{\infty} \phi_i(x) = 1$ everywhere in A

The actual definition of a partition of unity varies but it is the same general idea. This is the definition we will use. Under some definitions the definition above is actually "An infinitely differentiable partition of unity with compact supports subordinate to A" where the last part means the supports are contained in a single element of A and not just A itself.

Lemma 11: A partition of unity with the properties listed above always exists.

Proof:

Let $Q_1, Q_2, Q_3, \ldots$ be a sequence of rectangles in A satisfying the conditions of lemma 10. For each i, define $\psi_i: \mathbb{R}^n \to \mathbb{R}$ to be an infinitely differentiable function that is zero everywhere except the interior of Q_i where it is positive, possible by lemma 9, and the supports of these satisfy the conditions in condition 1 of the definition. In fact the only condition in the definition not satisfied by these functions is condition 3. Note that since finitely many of these functions are non-zero, the sum of all of them converges and we will call this $\lambda(x)$. Because each x in A has a neighbourhood around it where $\lambda(x)$ is the sum of a particular set of finitely many infinitely differentiable functions, $\lambda(x)$ is infinitely differentiable. It is also never 0 since the interiors of the rectangles Q cover A so that contributes a positive amount to the sum. Then $\phi_i = \frac{\psi_i}{\lambda}$ satisfies all the conditions of the lemma.

Lemma 12: Suppose I have K' compact inside an open set V in $\mathbb{R}^n$ and K' has measure zero boundary. Then there exists a compact set K with measure zero boundary in V such that K' is in the interior of K and a continuous function g on V that is 1 on K and 0 outside K'.

Proof:

For each x∈K' pick an axis aligned cube Q centered around it that is contained in V, possible as V is open. Then these cover K'. By compactness, pick a finite subcover $Q_1, Q_2, \ldots, Q_m$. Then each of these cubes and their boundary are contained in V and they cover K'. We can define K as the finite union of these cubes and their boundaries, then K is compact (finite union of compact sets), and K' is in the interior of K and K is in V. K has a measure 0 boundary because each cube has a measure 0 boundary and there are finitely many cubes so their total boundary is measure 0. So K exists.

Define the closed set $F := \mathbb{R}^n \backslash Interior(K)$. Then since K' is in the interior of K, $K' \cap F = \emptyset$ (the empty set).

Note that the function "distance to the nearest point in A" is continuous, this seems pretty legit but the justification is that $|d(x,A) - d(y,A)| \leq |x - y|$ because for any a in A, we have that by the triangle inequality $d(x,A) \leq |x - a| \leq |x - y| + |y - a|$, so $d(x,A) \leq |x - y| + d|y - a|$, so we can swap x and y to get the reverse inequality and we get $|d(x,A) - d(y,A)| \leq |x - y|$, then this is continuous because for any ε just pick δ=ε.

Note that since K is compact and F is closed and disjoint from K, the continuous function $d(x,F)$ attains a minimum on K by the extreme value theorem. If this were 0 anywhere, F and K would touch, so there is a δ that equals its minimum.

Now define $f(x) := \frac{d(x,F)}{d(x,F)+d(x,K')}$. Then this never has a 0 denominator as that would require a point to be in F and K at the same time, and it is continuous. Lets check the required values.

If x is in K' then $f(x) = \frac{d(x,F)}{d(x,F)+0} = 1$. If x is not in K then it is in f so $f(x) = \frac{0}{stuff} = 0$, so f is the function we want.

Lemma 13 (Change of variables theorem): This is essentially a generalization of integration by substitution to higher dimensions. The theorem says:

Let $U, V \in \mathbb{R}^n$ be open and let $\Phi: U \to V$ be a continuously differentiable bijection with continuously differentiable inverse. Write $J_\Phi(x) := |Det(D\Phi(x))|$. Suppose f is continuous on V, and that K' is a compact set contained in V, then we have that

$\int_{K'} f(y)dy = \int_{\Phi^{-1}(K')} f(\Phi(x))J_\Phi(x)dx.$

In fact this happens for any function f integrated on a compact set K' with f and Φ continuous on an open set around a compact set K'.

We will define K to be a compact set in V such that K' is in its interiorand use lemma 12 and multiply by that corresponding function to replace f with a function supported on a compact set K and prove the theorem for this case.

Proof:

We prove this for $f \geq 0$ because then we can do $f^+ - f^-$ after as both of these are continuous on a compact set and thus integrable. Note that $L := \Phi^{-1}(K)$ is compact because it is the continuous image of a closed set so it is closed and it is bounded by the extreme value theorem as it is the image of a continuous function on a compact set. Therefore everything happens inside L and K since f is 0 outside of that.

Since Φ is continuously differentiable, for any ε>0 there is a δ>0 such that whenever |x-a|<δ we have that $|D\Phi(x) - D\Phi(a)| < \varepsilon$. Now make all your cubes have diameter <δ. Then for any x and a in a cube Q, $\Phi(x) = \Phi(a) + A(x - a) + r(x)$ where $A = D\Phi(a)$. If $g(t) := \Phi(a + t(x - a))$ on [0,1] then the normal fundamental theorem of calculus implies

$\Phi(x) - \Phi(a) = \int_0^1 g'(t)dt = \int_0^1 D\Phi(a + t(x - a))(x - a)dt$ by the fact that the directional derivative is the derivative times the direction. Therefore

$$r(x) = \int_0^1 D\Phi(a + t(x - a) - A)(x - a)dt$$

$$|r(x)| = \int_0^1 |D\Phi(a + t(x - a) - A)||x - a|dt \leq \varepsilon|x - a| \leq \varepsilon * \delta$$

By the triangle inequality. So Φ can be uniformly approximated by linear functions on each of the cubes (This gives vibes of a theorem I proved in the differential equations course, but this is slightly different).

Let $T(x) := \Phi(a) + A(x - a)$, then $T(Q)$ is a parallelepiped and by standard determinant facts that we proved in level 6, $Vol(T(Q)) = |\det(A)|vol(Q)$. If Q is small enough, and $r := \varepsilon * \delta$, then if we cpmsoder P_r: the set of points within a distance r of a parallelepiped P, $Vol(P_r) = Vol(P)(1 + O(\varepsilon))$. Therefore $\Phi(Q) \subseteq T(Q)_r$ by our bound on $\Phi(x)$ so $Vol(\Phi(Q)) \leq Vol(T(Q))(1 + O(\varepsilon))$. By the same argument applied to Φ^{-1}, $Vol(\Phi(Q)) = Vol(T(Q))(1 + O(\varepsilon))$.

Now let Q be a cube with diameter at most δ and let suppose that $K \subset \Phi(Q)$. Then f is uniformly continuous as well so we can pick δ small enough so that $\int_{\Phi(q)} f = \left(f\left(\Phi(x_q)\right) + O(\varepsilon)\right)\left(vol(\Phi(q))\right)$ $= \left(f\left(\Phi(x_q)\right) + O(\varepsilon)\right)\left(|\det(A_q)|vol(q)(1 + O(\varepsilon))\right)$. Since $vol(q) = O(\delta^n)$, the integral is exactly $\left(f\left(\Phi(x_q)\right)\right)\left(|\det(A_q)|vol(q)\right) + O(\varepsilon\delta^n)$.

Now let f be supported on more than just a single rectangle, ie on K.

Since L is compact we can cover it by finitely many rectangles Q_i such that the local result applies on each. Let η_i be a partition of unity on U subordinate to Q_i and define $f_i(y) \coloneqq f(y)\eta_i(\Phi^{-1}(y))$. Then since this has support on a single Q, and since we have a partition of unity, we have $f(y) = \sum_{i=1}^{\infty} f_i(y)$ for all y in U and each f_i is continuos because it is the product of continuous functions. By the theorem for one rectangle that we proved above,

$$\int_V f_i(y)\,dy = \int_{Q_i} f_i(\Phi(x))\,J_\Phi(x)\,dx.$$

But $f_i(\Phi(x)) = f(\Phi(x))\,\eta_i(x)$. So

$$\int_V f_i(y)\,dy = \int_{Q_i} f(\Phi(x))\,\eta_i(x)\,J_\Phi(x)\,dx.$$

Where these integrals have error $O(\varepsilon\delta^n)$

Since f is 0 outside L and $\sum \eta_i = 1$ on L, we can sum over all finitely many i to deduce the desired result – Each error has $Vol(Q_i)O(\varepsilon)$ and since the volume of L is constant the total error is $O(\varepsilon)$ which goes to 0, so done.

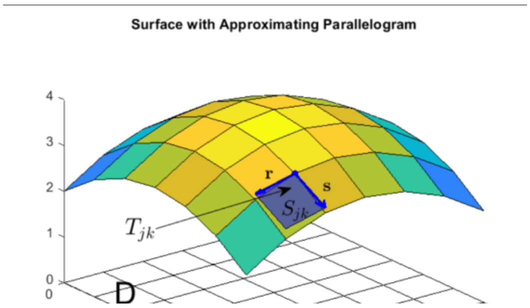
We will now talk about manifolds, but really these are embedded manifolds in $\mathbb{R}^n$ and not general manifolds.

Definition: Let $(x_1, x_2, \dots, x_k)$ be a k-tuple of elements in $\mathbb{R}^n$, then the k-dimensional volume of the parallelepiped spanned by these k vectors is given by a function $V(x_1, x_2, \dots, x_k)$

Definition: Let A be open in $\mathbb{R}^k$ and let $k \le n$ and let $f: A \to \mathbb{R}^n$ be continuously differentiable. We call the image of A under f a parametrized-manifold of dimension k which we call Y_α. We define the volume of this (which is surface area in the case k=2, n=3) to be $\int_A V(D\alpha)$, ie the integral over A of the function V of $D\alpha$.

Why the above definition agrees with how we normally think of area/volume:

Idea: Something like adding up the areas of the tangent rectangles like in this image (ignore the text in the image).



Suppose A has been partitioned into rectangles P and one such rectangle R is

$[a_1, a_1 + h_1] \times [a_2, a_2 + h_2] \times \dots \times [a_k, a_k + h_k]$, then if the h's are small, $\alpha(R)$ is a slightly curved rectangle contained in Y. Consider the edge having endpoints $a, a + h_i e_i$. And its image under α. By the derivative, this is approximately $v_i \coloneqq D\alpha(a).\, h_i e_i = h_i \left(\frac{\delta\alpha}{\delta x_i}\right)$. Therefore we want $V(v_1, v_2 \dots v_k)$, which equals $V\left(\frac{\delta\alpha}{\delta x_1}, \frac{\delta\alpha}{\delta x_2}, \ldots, \frac{\delta\alpha}{\delta x_k}\right) h_1 h_2 \ldots h_k = V\big(D\alpha(a)\big)vol(R)$ so when we sum this over all R and let the partition get finer we do get the integral we claimed at the beginning. This is well defined because the derivative and thus the volume is continuous and thus integrable.

There is a problem which is that this depends on α. This is something which we will address later to ensure this is well defined.

We also define the integral with respect to volume over the surface. The idea is we want to weight by a factor of f(x) or integrate f(x) over the surface. We write

$$\int_{Y_a} f = \int_A (f \circ \alpha) V(D\alpha)$$

Lets get a feel for what this is really saying. If we want to integrate the function |x| over a surface, then we have

$$\int_{Y_a} |x|$$

Now note that

$$\int_A |\alpha| V(D\alpha)$$

Has all of the small rectangle areas weighted by $|\alpha|$, but $|\alpha|$ is just |x| in the world where the surface lives.

This definition is important for later but we now define a k-manifold in $\mathbb{R}^n$ which is a stronger condition than being a parametrized-manifold of dimension k.

Definition: Suppose that $0 < k < n$ and M is a subset of $\mathbb{R}^n$ such that for each p in M there is a set V open in M containing p that is open in M and a set U that is open in $\mathbb{R}^k$ and a continuously differentiable map $\alpha: U \to V$ that has a rank k derivative everywhere and a continuous inverse, then M is a k-manifold without boundary in $\mathbb{R}^n$ and α is called a coordinate patch on M about p. Lets get a feel for this definition.

Suppose k=1 and n=2, then $\alpha = (t^3, t^2)$ gives a surface with a cusp at (0,0) so it is not smooth which is why we need the rank 1 condition. However, $\alpha = (t, t^2)$ works as it has rank 1 despite one of the derivatives being 0 – it is just a parabola.

Also, the condition that we have a continuous inverse is important because $(\sin(2t)\,|\cos(t)|, \sin(2t)\sin(t))$ fails only that condition (points near 0 in M can map under this inverse to points near either $0, \pi, \frac{\pi}{2}$, and then it maps the open interval $(0, \pi)$ to a figure eight in the plane).

Definition: If S is a subset of $\mathbb{R}^k$ we say f is continuously differentiable on S if there is an open set containing S such that f can be extended to a function g from U that is continuously differentiable.

Lemma 14: Let S be a subset of $\mathbb{R}^k$ and $f: S \to \mathbb{R}^n$. If for each x in S there is a neighborhood U_x of x and a continuously differentiable function $g_x: U_x \to \mathbb{R}^n$ that agrees with f on $U_x \cap S$ then f is continuously differentiable on S.

Proof:

Cover S by countably many neighborhoods U_x and let A be their union and ϕ_i be a partition of unity on A subordinate to U_x. For each i, choose a neighborhood U_x containing the support of ϕ_i – possible as we have a partition of unity and let g_i denote the function $g_x: U_x \to \mathbb{R}^n$ which exists by assumption. Then $\phi_i g_i$ is 0 outside a closed subset of U_x. This function combined with the function 0 everywhere else is thus a continuously differentiable function $h_i(x)$. If we sum over all $h_i(x)$ to get a function g(x) then by the partition of unity property that the finitely many ϕ's add to 1 and the assumption of the lemma that $g_i = f$ on U_x, g agrees with f everywhere and is continuously differentiable everywhere for the same reason as in the proof of lemma 11. Therefore f is continuously differentiable on S due to being defined and continuously differentiable on an infinite union of open neighborhoods U_x which is an open set around S.

The above is actually a k-manifold without boundary – if k=1 this is a line and if k=2 this is a surface. We can allow for boundaries if instead of $\mathbb{R}^k$ we work in the upper half space of $\mathbb{R}^k$ including the boundary of that space (H^k) and force U to be open in either $\mathbb{R}^k$ or H^k. We still require V to be open in M in the definition. We also require it to be the case that α can be extended to a continuously differentiable function on an open set containing U.

We define H_+^k as H^k without its boundary and B^k as the k-1-dimensional boundary of H^k.

Definition: The boundary of a k-manifold in $\mathbb{R}^n$ is not the boundary in the usual sense, it is the points p such that we do not have a continuously differentiable function α from U to V with U open in $\mathbb{R}^k$ and V open in M. If a k-manifold is M we denote the boundary ∂M.

Intuition: By the next lemma, it will follow that by the inverse function theorem, we can essentially break the manifold down into an open cover where we parametrize each open piece by an invertible continuously differentiable function and then glue them together.

Lemma 15: Let M be a k-manifold in $\mathbb{R}^n$ with coordinate patches $a_0: U_0 \to V_0, a_1: U_1 \to V_1$ such that $W \coloneqq V_0 \cap V_1$ is non-empty. Let $a_0^{-1}(W) = W_0, a_1^{-1}(W) = W_1$. Then $a_1^{-1} \circ a_0: W_0 \to W_1$ is continuously differentiable with non-singular derivative.

Proof: We just need to show that if α is a coordinate patch on M then its inverse is continuously differentiable, and thus so is $a_1^{-1} \circ a_0$ by the chain rule. Since we will have that $a_0^{-1} \circ a_1$ is continuously differentiable as well, we know then that the derivative is invertible everywhere.

First, suppose U is open in H^k but not $\mathbb{R}^k$. Then by assumption, we can extend α to a continuously differentiable function β on a set U' open in $\mathbb{R}^k$ as this was part of the definition above of a k-manifold. Let x_0 be in U and $p_0 \coloneqq a(x_0)$ be in W. Then $D\alpha(x_0)$ has rank k so some k rows of this matrix are independent. So let $\pi: \mathbb{R}^n \to \mathbb{R}^k$ project $\mathbb{R}^n$ onto the k coordinates corresponding to the rows which are independent. Then $g \coloneqq \pi \circ \beta$ maps U' into $\mathbb{R}^k$ with $Dg(x_0)$ non-singular (as it consists of these k independent rows). By the inverse function theorem, there is an open set S of $\mathbb{R}^k$ about x_0 such that g has a continuously differentiable inverse. Now consier the map $h = g^{-1} \circ \pi$, then this is a continuously differentiable function on some open set. Note that $W \cap U = U_0$ is open in U since it is the pre-image of an open set in M under a continuous function - V_0 is open in M and thus in V. This means there is an open set A of $\mathbb{R}^n$ such that $A \cap V = V_0$ by definition of "open in". By intersecting A with $\pi^{-1}(g(S))$, we gurantee that it still contains p_0 and we can redefine V_0 and the domain of a_0 accordingly wile ensuring that A is in the domain of h so that what we are about to do makes sense as we need g^{-1} to be defined. If $p \in V_0$ then we let $x = a^{-1}(p)$. We then can prove that h is the inverse we want: $h(p) = h\big(a(x)\big) = g^{-1}\left(\pi\big(a(x)\big)\right) = g^{-1}\big(g(x)\big) = x = a^{-1}(p)$, so h is the continuously differentiable function with non singular derivative that we wanted. If U is open in $\mathbb{R}^k$ it's the same argument but without the need to extend to β.

Lemma 16: If a set K in $\mathbb{R}^n$ has the property that every open cover has a finite subcover then it is closed and bounded, and therefore these definitions of compactness are equivalent in $\mathbb{R}^n$.

Proof:

Boundedness is easy – if there exists a finite subcover the set is covered by a bounded thing and thus bounded, so we just need to prove it is closed. Let z be in $\mathbb{R}^n \backslash K$ (the complement of K). Each x in K is not z so we can finx disjoint open balls around x and around z. These open balls around these x's form an open cover with a finite subcover by assumption. But then the open balls around z intersected is still an open ball since there are finitely many, and they are disjoint from our cover of K and thus not in K. Since z was an arbitrary point not in K, it follows that $\mathbb{R}^n \backslash K$ is open and thus K is closed.

Remark: In $\mathbb{R}^n$ we now have shown in this website (In Levels 6,7,8) that $1 \Rightarrow 3 \Rightarrow 1$ and $1 \Rightarrow 2$ and we know that $2 \Rightarrow 1$ since if the set is not closed or not bounded we can definitely find a sequence with no convergent subsequence, where here 1, 2 and 3 mean

1. A set is closed and bounded
2. Every sequence of points in the set has a subsequence converging to a point in the set
3. Every open cover of the set has a finite subcover

Therefore all 3 of these definitions are equivalent in $\mathbb{R}^n$ and are what we call being compact.

Corollary: Images of compact sets under continuous functions are compact. Let B be the image and find an open cover of B. Then the pre-image of this open cover is an open cover by the property that pre-images of open sets under continuous functions are open. This covers the domain A which is compact so find a finite subcover and re-apply the function to get a finite subcover of B. We will use this fact in the proof of the next lemma.

Lemma 17: If f is a one-to-one function from A to B with $A, B \in \mathbb{R}^n$ with Df non-singular for all x in a, and A is open in $\mathbb{R}^n$, then B is open in $\mathbb{R}^n$.

Proof:

We note that if a function $\mathbb{R}^n \to \mathbb{R}$ has a local minimum then all partial derivatives are 0 and thus the derivative there is 0.

Given b in B pick a rectangle Q in A whose interior contains $a = f^{-1}(b)$ (possible because A is open). The boundary of Q is compact so f(Bd(Q)) is also compact by the last corollary. Because f is assumed to be one-to-one, every point in f(Bd(Q)) is disjoint from b. Because f(Bd(Q)) is closed its complement is open so there is a δ>0 such that $B_{2\delta}(b)$ is disjoint from f(Bd(Q)). Given c in $B_\delta(b)$. Then we want to show that c is f(x) for some x in Q, ie the image of the rectangle and its inside is filled in and cannot tear or puncture. Now consider $\phi(x) := |f(x) - c|^2$, then Q is compact so by the extreme value theorem it takes on a minimum in Q at some point x in Q. $\phi(a) := |f(a) - c|^2 = |b - c|^2 < \delta^2$ so the minimum value is at less than δ^2. It also does not occur on the boundary of Q because the boundary is a distance more than δ away from c. Therefore, since the minimum occurs on the interior, it is a local minimum, so the derivative of ϕ there canishes. By pythagoras, $\phi(x) = \sum_{k=1}^{n} (f_k(x) - c_k)^2$, so by the chain rule $D_j\phi(x) = \sum_{k=1}^{n} 2(f_k(x) - c_k)D_j f_k(x)$ where I mean the j'th partial derivative. We know this partial derivative is 0 so we can write this as $2[(f_1(x) - c_1) \dots (f_n(x) - c_n)].Df(x) = 0$. But Df(x) is invertible by hypothesis. Multiply both sides on the right by its inverse to get that f=c follows, so done.

Lemma 18: A point p in a k-manifold M with coordinate patch $\alpha: U \to V$ about p is a boundary point if and only if we have that both U is not open in $\mathbb{R}^k$ and $\alpha^{-1}(p) \in B^k$.

Proof:

From the definition, if U is open in $\mathbb{R}^k$ we are not at a boundary point. Also, if U is open in H^k and $\alpha^{-1}(p) \in H_+^k$ then set $U_0 \coloneqq U \cap H_+^k$ and $V_0 \coloneqq \alpha(U_0)$. Then α with domain restricted to U_0 is a coordinate patch with domain open in $\mathbb{R}^k$ so we are not a boundary point. Therefore we have proven the "only if" part of the lemma.

So lets prove the if part. Let $\alpha: U \to V$ be a coordinate patch about p and suppose that U is open in H^k and $p = \alpha(x)$ with $x \in B^k$. Then assume there is a coordinate patch $\alpha_1: U_1 \to V_1$ about p with U_1 open in $\mathbb{R}^k$. Then V_1 and V contain p and are open in M due to the definition of a coordinate patch. Therefore $W \coloneqq V_1 \cap V$ is open in M. But then $\alpha^{-1}(W)$ is open in H^k and contains $\alpha^{-1}(p)$, and $\alpha_1^{-1}(W)$ is open in $\mathbb{R}^k$. Therefore $\alpha^{-1} \circ \alpha_1: \alpha_1^{-1}(W) \to \alpha^{-1}(W)$ is a continuously differentiable bijective function with continuously differentiable inverse and full rank derivative (by lemma 15) carrying $\alpha_1^{-1}(W)$ to $\alpha^{-1}(W)$. We know that $\alpha^{-1}(W)$ is not open in $\mathbb{R}^k$ since U is in H^k and contains a boundary point, but also by lemma 17 $\alpha^{-1}(W)$ is open in $\mathbb{R}^k$ which is a contradiction.

Lemma 19: If the boundary of a k-manifold that actually has a boundary is non-empty it is a k-1 manifold without boundary.

Proof:

Let p be in ∂M. Let $\alpha: U \to V$ be a coordinate patch on M about p. Then U is open in H^k and can be extended to a set U' open in $\mathbb{R}^k$ and $p = a(x_0)$ for some $x_0 \in B^k$. By the previous lemma, each point of $U \cap H_+^k$ is mapped to an interior point of M, and each point of $U' \cap B_k$ is mapped to a point of ∂M. Thus the restriction of α to $U' \cap B_k$ is a bijection onto the open set $V \cap \partial M$ of ∂M. It is continously differentiable because α is and its derivative has rank k-1 because it is just the derivative of the original α with the last column removed. Its inverse is continuous because it equals the restriction to V_0 of the continuous function α^{-1}. It is without boundary because U' is open.

We note that the condition that the manifold has a boundary is because we need the derivative to be of full rank on the boundary which is guranteed if the boundary is part of the manifold.

Problem: These notes were based on a book that used a terrible definition of k-manifold by not allowing to go to quarter or eighth space or anything which means the version of the "master theorems" proved in that book that we are working up to in these notes is useless for computation due to not allowing for piecewise smooth things but only totally smooth things. This means I will have to do an approximation lemma at the end of these notes, I really hope it is easy.

Lemma 20: Let $\{v_1, v_2, \dots, v_k\}$ be a list of vectors arranged as columns of a matrix M. Then $V(v_1, v_2, \dots, v_k) = \sqrt{\det(M^T M)}$

Proof:

Recall that in level 6 when we proved the determinant formula coincides with volume, we essentially showed that a function that determines the volume has to have so many properties that it is uniquely determined. We will do a similar approach here.

Define $F(M) \coloneqq \det(M^T M)$, where F is a function on matrices of any dimension (in this case with columns corresponding to the vectors we care about). Let $H: \mathbb{R}^n \to \mathbb{R}^n$ be an orthogonal linear transformation represented by A, then $F(AM) = \det\big((AM)^T(AM)\big) = \det(M^T A^T A M) = \det(M^T M) = F(M)$. Also, if Z is a k by l matrix and y is the n by k matrix $Y = \begin{bmatrix} Z \\ 0 \end{bmatrix}$ then $F(Y) = \det([Z^T\ 0]Y) = \det(Z^T Z) = \det(Z)^2$.

Now note that F is invariant under pre-multiplying by orthogonal transformations. Let A carry $v_1, v_2, \dots, v_k$ onto the subspace which is the natural copy of $\mathbb{R}^k$. Then there is a Z such that A.X is equal to $\begin{bmatrix} Z \\ 0 \end{bmatrix}$ (since k is not more than n here otherwise this is not very interesting as the volume of a >n dimension is always 0 anyway and so is a matrix with more columns (n) than rows (k) multiplied on the right by its transpose which can only have at most k linearly independent rows).

Therefore $F(X) = F(A.X) = \det(\mathrm{Z})^2 \geq 0$ with F 0 if and only if the columns of Z are dependent which occurs if and only if our original k vectors are dependent.

But note that the k-volume should be $|Det(Z)|$, and the thing we are claiming is the square of the k-volume is $\det(\mathrm{Z})^2$ and positive, so we are done.

Lemma 21: We now go back to parametrized manifolds. Let $g: A \to B$ be a continuously differentiable function with A and B open in $\mathbb{R}^k$ such that g has a continuously differentiable inverse. Let $\beta: B \to \mathbb{R}^n$ be continuously differentiable and let y be β(B) in the image sense. Let α=β∘g so that $\alpha: A \to \mathbb{R}^n$ and also $Y = \alpha(A)$. Let $f: Y \to \mathbb{R}$ be continuous, then $\int_B (f \circ \beta) V(D\beta) = \int_A (f \circ \alpha) V(D\alpha)$. Furthermore suppose that these conditions hold on open sets that A and B's closure contain. In particular, for bijective (important so we do not go over the surface more than once) continuously differentiable parametrizations with continuously differentiable inverses, the integral over the parametrized manifold stays the same. Furthermore, one integral exists if the other does.

Proof: By the change of variables theorem which is true for compact sets where the conditions hold on an open set around them, $\int_B (f \circ \beta) V(D\beta) = \int_B \big((f \circ \beta) \circ g\big)(V(D\beta) \circ g)|\det(Dg)| = \int_A (f \circ \alpha)(V(D\beta) \circ g)|\det(Dg)|$. By the chain rule, $D\alpha(x) = D\beta(g(x))Dg(x)$. By the previous lemma, we have that $\left[V\big(D\alpha(x)\big)\right]^2 = \det\Big(Dg(x)^T D\beta\big(g(x)\big)^T D\beta\big(g(x)\big)Dg(x)\Big) = Det\big(Dg(x)\big)^2\left[V\Big(D\beta\big(g(x)\big)\Big)\right]^2$. This means that

$$\int_B (f \circ \beta) V(D\beta) = \int_A (f \circ \alpha)(V(D\beta) \circ g)|\det(Dg)| = \int_A (f \circ \alpha)\big(V(D\alpha)\big)$$

It follows that if α is a single coordinate patch that can cover the support of f, then the integral of f defined in this sense over a manifold is well defined (with the caveat that an integral over a k-manifold is defined over the interior of U in $\mathbb{R}^k$ instead of U – in the other defintion we assumed the domain was open). We now extend this to the global case using partitions of unity, supposing M is a compact manifold.

The good news is that this agrees with the pythagorean intuition of arc length, as that is also working with the tangent area elements and taking an integral of them which gets better with finer partitions.

Lemma 22:

Let M be a compact k-manifold in $\mathbb{R}^n$. Given a covering of M by coordinate patches, there is a finite collection of continuously differentiable functions ϕ_i from $\mathbb{R}^n$ to $\mathbb{R}$ that are always non-negative, sum to 1 in M, and their supports are compact and such that there is a coordinate patch $\alpha: U \to V$ such that $\left((Support(\phi_i)) \cap M\right) \subset V_i$. If then we can multiply f by each of these functions which are individually well defined and independent of patch choice by the previous lemma and add them up since there are finitely many to then get that the integral of f is well defined – in fact we define it by the sum of these. So lets prove this.

Proof:

For each coordinate patch $\alpha: U \to V$ belonging to the given collection, choose an open set $A_V \in \mathbb{R}^N$ such that $A_V \cap M = V$. Let A be the union of these sets and choose a partition of unity on A subordinate to this open covering of A and thus M which thus has a finite subcover. Then on each point in M, finitely many functions in these partitions of unity do not vanish by definition of a partition of unity, and finitely many of these neighborhoods cover M by compactness, so finitely many functions in these partitions of unity do not vanish on M.

Lemma 23:

This one is obvious but we give a proof. A k-1-dimensional manifold in k-1dimensions has measure 0 in the k dimensional volume sense

Proof:

If we can prove it for k dimensions it is in fact trivial for all dimensions **less than** k because, eg we can imagine extending a curve in the x-y-plane by 1 z-unit to get a surface in the z-plane which if it has zero 3-volume means it can be covered by cuboids of total size less than ε for any ε no matter how small, and the mean area of slices of of this covering is ε so some slice has at most that much area so the curve has area 0.

Now lets prove it for surfaces in 3D (the proof easily generalizes to higher dmiensions).

We note that the parametrization can be covered by countably many coordinate patches, since we can consider coordinate patches only involving our manifold M intersected with open balls in $\mathbb{R}^k$ with rational radius and center which there are countably many of. So if we can prove measure 0 for each part of the coordinate patch then by adding $\frac{\varepsilon}{2}$ for the first part, $\frac{\varepsilon}{4}$ for the second, $\frac{\varepsilon}{8}$ for the third, etc, we can get less than ε in total. So we have reduced it to proving it for each part that has a single coordinate patch.

On these parts, by definition we can parametrize them by a continuously differentiable function whose derivative is bounded (we can force this if we make our balls small enough), so if F(u,v) is a parametrization on that part, then concretely, there is an M with $\left|\frac{\partial F(u,v)}{x_i}\right| \leq M$ for all i from 1 to k-1. Now tile $\mathbb{R}^{k-1}$ (ie the domain of the parametrization) by little squares or cubes or hypercubes or whatever of side length δ, and on one such square Q (by derivative bounds), we have that $|F(u,v) - F(u_0 - v_0)| \leq M(|u - u_0| + |v - v_0|) \leq 2M\delta$ for 2-surfaces and in general $(k-1)M\delta$ so the image of F(Q) lives inside a ball of radius $(k-1)M\delta$, in fact it lives inside a cube of volume $\left(2(k-1)\right)^k M^k \delta^k$ (yes I know this is an overestimate but its not wrong), and there are $\frac{C}{\delta^{k-1}}$ for some C such regions that cover U (the domain of that part of the patch). Therefore the total volume of these covering cubes is at most $\left(2(k-1)\right)^k M^k \delta$ and since δ was arbitrary and the rest is constant this gets arbitrarily small so done.

Lemma 23: With the inverse function theorem, if our original function is 2x continuously differentiable so is its local inverse

Proof:

We just need to show that $D(f^{-1})$ is continuously differentiable. By the inverse function theorem this is equal to (At some x) $Df\left(f^{-1}(x)\right)^{-1}$, and if a matrix is invertible and continuously differentiable then since inverse matrices are just polynomial-like functions of the coefficients they are continuously differentiable too as f^{-1} is also continuously differentiable and we assume Df is differentiable, so the whole thing $D(f^{-1})$ is continuously differentiable with respect to x, so f^{-1} is twice continuously differentiable