Sard's Theorem

1 Pre-Requisites

1.1 Critical values

Let $f : M \to N$ be a function between two manifolds and consider Tf the induced map on the tangent bundles. A point $x \in M$ is called a *critical point* if $T_x f : T_x M \to T_{f(x)} N$ is not surjective; the set of such critical points we shall denote by C. A critical value is any point in N that is the image of a critical values, i.e. $y \in f(C)$.

Lemma 1.1. *The set of critical points is closed.*

Proof. Let $x \in M$ be a regular value (i.e. not critical). The question is local so we can assume we have taken co-ordinate patches around both x and $f(x)$. As $T_x f$ has maximal rank we can find a linear map $A : \mathbb{R}^m \to \mathbb{R}^{n-k}$ such that the map $\phi : U \subset \mathbb{R}^n \to \mathbb{R}^n$ given by

$$y \mapsto \begin{bmatrix} f(y) \\ Ay \end{bmatrix}$$

has an invertible Jacobian at x. We then apply the inverse function theorem to ϕ to find a neighbourhood of x on which ϕ is invertible. It is clear that on this neighbourhood Tf must have rank n $\square$

Let C_k denote the set of points where all partial derivatives of f of degree $\leq k$ vanish. In general higher order derivatives do not transform as tensors however if all lower partial derivatives vanish then it is a co-ordinate independent question to ask whether the k^{th} one does. Thus C_k is a well defined subset of M.

Lemma 1.2. $C \supset C_1 \supset \cdots \supset C_k \supset C_{k+1} \supset \cdots$. *Moreover* C_k *is closed.*

Proof. The inclusions are obvious so we just turn to the question of closedness of C_k. As above this is a local question so we work in a co-ordinate system. For $x \notin C_k$ let $r < k$ be the first term in which $f(x)$ has a non-zero partial derivative of degree r, Then using Taylor's theorem we have

$$f(x+h) = f(x) + \sum_{|I|=r} f_I(x) h^I + R(x,h)$$

where $R(x,h) \sim q(x,h) h^{r+1}$. For a given compact neighbourhood of x we can bound all the partial derivatives of R of degree less than r so that we have the r^{th} partial derivative of f at $x+h$ (for suitable small h) is $f_I(x)$ plus a term that goes as $Na(h + h^2 + \cdots + h^r) \leq \frac{Nah}{1-h}$. For a suitably small choice of h we then see that there is a neighbourhood of x where this r^{th} partial derivative doesn't vanish hence the result. $\square$

1.2 Cubical measure

Let $Z \subset \mathbb{R}^n$ be a set. A cubical cover of Z is a countable collection of n-dimensional cubes $\{S_i\}$ such that $Z \subset \cup_i S_i$. Note that as $\mathbb{R}^n$ can be covered by a countable collection every such set has a cubical cover. We define a map μ that takes a cubical cover and returns $\sum_i l_i^n \in [0, \infty]$ where l_i is the length of the cube S_i.

Definition 1.3. *The (cubical) measure of* Z *is the infimum of* $\mu(S) \mid S$ *is a cubical cover of* Z.

Proposition 1.4. *Let* S *be a cube, then its measure is equal to its volume.*

Proof. Using the cover $\{S\}$ we see that its measure is $\leq l^n$. Conversely let $\{S_i\}$ cover S. We know that any finite subset which covers S must have volume greater than or equal the volume of S. Now assume we have an infinite set $\{S_i\}$. For a given $\delta > 0$ consider $\{((1+\delta)S_i)^\circ\}$ - that is for each cube we extends its length slightly and and then take the interior. This is an open cover of S and hence has a finite subcover. The volume of the finite subcover is $\leq (1+\delta)\mu(\{S_i\})$ but greater than the volume of S, i.e. $\text{vol}(S) \leq (1+\delta)^n \mu(\{S_i\})$ - as δ was arbitrary we must have $\text{vol}(S) \leq \mu(\{S_i\})$. $\square$

It follows that any open set has non-zero measure as it contains a cube.

Example 1.5. *Consider the line $y = 0$ in $\mathbb{R}^2$, call it H. If you consider the sequence of intervals $I_n = [-n, n] \times \{0\}$, then $\cup I_n = H$. We cover $I_n - I_{n-1}$ by cubes of length $\frac{1}{m_n}$ of which there are $2m_n$ so this cubical cover has volume*

$$\sum_n \frac{2}{m_n}$$

and by a suitable choice of m_n for any $\epsilon > 0$ we can find a cubical cover with volume $< \epsilon$. It follows that $\mu(H) = 0$. This can be extended to any hyperplane in any dimension.

Lemma 1.6. *The countable union of a sets of measure zero has measure zero.*

Proof. Let C_i be a countable collection of sets of measure zero indexed by the non-zero natural numbers. Then for a given $\epsilon > 0$ for each i choose a covering $S_{i,j}$ such that $\sum_j \mu(S_{i,j}) < 2^{-i}\epsilon$. It follows that $\{S_{i,j}\}$ cover $\cup_i C_i$ and have measure $< \epsilon$. □

Lemma 1.7. *Let $U \subset \mathbb{R}^m$ be open and $C \subset U$ have measure zero. Then for any diffeomorphism $\phi : U \to V \subset \mathbb{R}^m$ $\phi(C)$ has measure zero.*

Proof. Take the notation from the question, firstly for any $x \in U$ we can write $\phi(x+h) = \phi(x)+\sum_i R_i(x,h)h_i$. Now if we take some compact set $K \subset U$, then for $x, y \in S \cap K$ for S a cube of length l we have

$$|\phi(y) - \phi(x)| \leq A_K\, l\, \sqrt{n}$$

where A_K is a constant that doesn't depend on C or S. It follows that $\phi(S)$ lies in a cube of length $A_S\, l$. So if we take some cubical cover of $C \cap K$, say S_i such that $\sum \mu(S_i) < \epsilon$ it follows $\sum \mu(\phi(S_i)) < A_k\epsilon$ and hence $\phi(C \cap K)$ has measure zero. The general result follows by applying 1.6 and taking a countable compact cover of U. □

It follows that a set having measure zero is an intrinsic property of a differentiable manifold that doesn't depend on any choice of local co-ordinates.

2 Main Results

Theorem 2.1 (Sard's Thereom)**.** *Let $f : M \to N$ be a smooth function; then the critical values have measure zero.*

We will prove the theorem by the route of a few lemmas - one will see that the smoothness condition can be weakened but the differentiability required is a function of the dimension of the two manifolds. In terms of notation C denotes all critical points, C_k denotes the set of points where all partial derivatives of order $\leq k$ vanish; clearly $C \supset C_1 \supset \cdots \supset C_k \supset \cdots$, and all functions will be assumed to have the required level of differentiability.

Lemma 2.2. *Let $f : U \subset \mathbb{R}^m \to \mathbb{R}^n$. If $k > \frac{m}{n} - 1$ then $f(C_k)$ has measure zero.*

Proof. We can cover U by a countable number of cubes of length l so by lemma 1.6 it is sufficient to provide it for $C_k \cap S$ where S is such a cube. Taking the Taylor expansion of f around a point $x \in C_k$ we have that

$$f(x + h) = f(x) + \sum_{|I|=k+1} R_I(x, h)h^I$$

so it follows that on S we have $|f(x + h) - f(x)| = A_{S,k}|h|^{k+1}$ for some constant $A_{S,k}$. We now partition the cube into r^m cubes each of length l/r. Let x be an element of C_k that lies in one of these smaller cubes, call it S'. If $y \in S'$ is another point then we have that $|x - y| \leq \sqrt{m}\frac{l}{r}$, and so

$$|f(y) - f(x)| \leq A_{S,k} \left(\sqrt{m}\frac{l}{r}\right)^{k+1}$$

and $f(S')$ lies in a cube of length $A_{S,k}\sqrt{m^{k+1}/n}(l/r)^{k+1}$. There are at most r^m such cubes and hence the $f(S \cap C_k)$ can be covered with cubes of total volume less than or equal to

$$r^m \cdot \left(A_{S,k}\sqrt{\frac{m^{k+1}}{n}} \left(\frac{l}{r}\right)^{k+1}\right)^n = a\, r^{m-n(k+1)}$$

where the constant a depends on S, k, n and m which are all fixed. It follows that if $m - n(k+1) < 0$, i.e. the conditions of the lemma, we can choose r large enough such that $f(S \cap C_k)$ is covered by cubes with total volume arbitrarily small, i.e. it has measure zero. $\square$

Corollary 2.3. *Sard's theorem holds where the domain is an open subset of* $\mathbb{R}$.

Proof. We split this into two cases:

(i) The codomain has dimension 1: In this case the critical points are exactly C_1 and so the result is true by 2.2.

(ii) The codomain has dimension > 1: Here $C = U$ ($U \subset \mathbb{R}$ is the domain). We split U as $(U - C_1) \sqcup C_1$. We know from 2.2 that $f(C_1)$ has measure 0 and from 1.2 we $V = U - C_1$ is open. Let $x \in V$ then one of the coordinate functions has non-zero differential at x, suppose it is f_1. As $f_1'(x) \neq 0$ we can find a neighbourhood V_x of x and a local inverse to f_1, call it ϕ, on V_x by the inverse function theorem. The map $g = f \circ \phi^{-1}$ clearly has the same critical values as f so it suffices to consider g. The function g has the form $t \mapsto (t, g_2(t), \ldots, g_m(t))$, in particular it is an embedding of V_x. Using the submanifold coordinate proposition we can find, by potentially refining V_x to W_x, a diffeomorphism of $\mathbb{R}^m$, ψ, such that $\psi \circ g(t) = (t, 0, \ldots, 0)$ which shows $\psi g(W_x)$ has measure 0, and by 1.7 so does $g(W_x)$. Hence for each $x \in V$ we can find a neighbourhood W_x such that $g(W_x)$ has measure zero. We can cover V by a countable number of such open subsets, by second countability, hence by 1.6 it follows $g(V)$ has measure zero and so the result.

$\square$

Lemma 2.4. *Let* $f : U \subset \mathbb{R}^m \to \mathbb{R}^n$*, then* $f(C - C_1)$ *has measure 0.*

Proof. We shall prove this by induction on m with the base case holding by 2.3. Now for the inductive step; let $x \in C - C_1$, we can then assume, by relabelling the co-ordinates, that $\frac{\partial f_1}{\partial y_1}(x) \neq 0$. Then the function $\phi : U \to \mathbb{R}^m$ by $y \mapsto (f_1(y), y_2, \ldots, y_m)$ it has invertible Jacobian at x and thus we can invoke the inverse function theorem to get an inverse to ϕ locally defined around $\phi(x)$. The function $g = f \circ \phi^{-1}$ has the same critical values so we can use this instead. Now $g(y_1, \ldots, y_m) = (y_1, \ldots)$ and so a point (t, z) is critical for g iff z is a critical point of $g_t(z)$ where $g_t(z) = g(t, z)$. We can now apply the inductive hypothesis[1] to get that g_t has measure zero for each t. We can now apply Fubini's theorem, on some countable compact subcover if necessary, to conclude that $f(U)$ has measure zero. $\square$

Lemma 2.5. *If* $k \geq 1$ *then* $f(C_k - C_{k+1})$ *has measure 0.*

Proof. The proof follows a similar method to above and will be by induction with base case $m = 1$ covered by 2.3. Let $x \in C_k - C_{k-1}$ then there is some k^{th} partial derivative of one of the f_i whose derivative with respect to one of the y_i is non-zero at x. Let us denote the partial derivative by u and by relabelling the y_i we may assume that $\frac{\partial u}{\partial y_1}(x) \neq 0$. We can then apply the inverse function theorem to $y \mapsto \phi(y) = (u(y), y_2, \ldots, y_m)$ and so we have the function $g = f \circ \phi^{-1}$ that has the same values as f on this neighbourhood of x, call it V. We have that $\phi(C_k \cap V) \subset \{0\} \times \mathbb{R}^{m-1}$ and so we can apply the inductive hypothesis to get the result. $\square$

Proof of Sard's Theorem. Let $f : M \to N$ be as in the statement of the theorem. We can find a countable number of co-ordinate patches $U_i \subset M$, $V_i \subset N$ with $f|_{U_i} : U_i \to V_i$. It follows from 1.6 and 1.7 that it is sufficient to prove the result for open subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$. The result thus follows by using the above results. $\square$

[1] Most proofs go down to $m = 0$ here but not sure that is valid for applying Fubini's theorem