3.2 Sard's theorem

The fundamental idea which allows us to prove that transversality is a generic condition is a the theorem of Sard showing that critical values of a smooth map $f : M \longrightarrow N$ (i.e. points $q \in N$ for which the map f and the inclusion $\iota : q \hookrightarrow N$ fail to be transverse maps) are *rare*. The following proof is taken from Milnor, based on Pontryagin.

The meaning of "rare" will be that the set of critical values is of *measure zero*, which means, in $\mathbb{R}^m$, that for any $\epsilon > 0$ we can find a sequence of balls in $\mathbb{R}^m$, containing $f(C)$ in their union, with total volume less than ϵ. Some easy facts about sets of measure zero: the countable union of measure zero sets is of measure zero, the complement of a set of measure zero is dense.

We begin with an elementary lemma describing the behaviour of measure-zero sets under differentiable maps.

Lemma 3.15. *Let $I^m = [0,1]^m$ be the unit cube, and $f : I^m \longrightarrow \mathbb{R}^n$ a C^1 map. If $m < n$ then $f(I^m)$ has measure zero. If $m = n$ and $A \subset I^m$ has measure zero, then $f(A)$ has measure zero.*

Proof. If $f \in C^1$, its derivative is bounded on I^m, so for all $x, y \in I^m$ we have

$$||f(y) - f(x)|| \leq M||y - x||, \tag{58}$$

for a constant[3] $M > 0$ depending only on f. So, the image of a ball of radius r in I^m is contained in a ball of radius Mr, which has volume proportional to r^n.

If $A \subset I^m$ has measure zero, then for each ϵ we have a countable covering of A by balls of radius r_k with total volume $c_m \sum_k r_k^m < \epsilon$. We deduce that $f(A_i)$ is covered by balls of radius Mr_k with total volume $M^n c_n \sum_k r_k^n$; since $n \geq m$ this goes to zero as $\epsilon \to 0$. We conclude that $f(A)$ is of measure zero.

If $m < n$ then f defines a C^1 map $I^m \times I^{n-m} \longrightarrow \mathbb{R}^n$ by pre-composing with the projection map to I^m. Since $I^m \times \{0\} \subset I^m \times I^{n-m}$ clearly has measure zero, its image must also. $\square$

Remark 3.16. If we considered the case $n < m$, the resulting sum of volumes may be larger in $\mathbb{R}^n$. For example, the projection map $\mathbb{R}^2 \longrightarrow \mathbb{R}$ given by $(x, y) \mapsto x$ clearly takes the set of measure zero $y = 0$ to one of positive measure.

A subset $A \subset M$ of a manifold is said to have measure zero when its image in each chart of an atlas has measure zero. Lemma 3.15, together with the fact that a manifold is second countable, implies that the property is independent of the choice of atlas, and that it is preserved under equidimensional maps:

Corollary 3.17. *Let $f : M \to N$ be a C^1 map of manifolds where $\dim M = \dim N$. Then the image $f(A)$ of a set $A \subset M$ of measure zero also has measure zero.*

---

[3]This is called a Lipschitz constant.

Corollary 3.18 (Baby Sard)**.** *Let $f : M \to N$ be a C^1 of manifolds where* $\dim M < \dim N$*. Then $f(M)$ (i.e. the set of critical values) has measure zero in N.*

Remark 3.19. Note that this implies that space-filling curves are not C^1.

Now we investigate the measure of the critical values of a map $f : M \to N$ where $\dim M = \dim N$. The set of critical points need not have measure zero, but we shall see that

> **The variation of f is constrained along its critical locus since this is where Df drops rank. In fact, the set of critical *values* has measure zero.**

Theorem 3.20 (Equidimensional Sard)**.** *Let $f : M \to N$ be a C^1 map of n-manifolds, and let $C \subset M$ be the set of critical points. Then $f(C)$ has measure zero.*

Proof. It suffices to show result for the unit cube mapping to Euclidean space (using second countability, we can cover M by countable collection of charts $(U_i, \varphi_i)_{i\in I}$ with the property that $(\varphi_i^{-1}(I^n))_{i\in I}$ covers M. Since a countable union of measure zero sets is measure zero, we obtain the result). Let $f : I^n \longrightarrow \mathbb{R}^n$ a C^1 map, and let M be the Lipschitz constant for f on I^n, i.e.

$$||f(x) - f(y)|| \leq M|x - y|, \quad \forall x, y \in I^n. \tag{59}$$

Let c be a critical point, so that the image of $Df(c)$ is a proper subspace of $\mathbb{R}^n$. Choose a hyperplane containing this subspace, translate it to $f(c)$, and call it H. Then

$$d(f(x), H) \leq ||f(x) - f_c^{\mathrm{lin}}(x)||, \tag{60}$$

where $f_c^{\mathrm{lin}}(x) = f(c) + D_c f(x - c)$ is the linear approximation to f at c. By the definition of the derivative, for each $c \in C$, we have that $\forall \epsilon > 0, \exists \delta > 0$ such that

$$||f(x) - f_c^{\mathrm{lin}}(x)|| < \epsilon ||x - c|| \text{ for all } x \text{ s.t. } ||x - c|| < \delta.$$

Because f is C^1 and C is compact, we conclude that $\forall \epsilon > 0, \exists \delta > 0$ such that the inequality above holds for all $c \in C$.

Now we apply this: if $c \in C$ and $||x - c|| \leq \delta$, then $f(x)$ is within a distance $\epsilon\delta$ from H and within a distance $M\epsilon$ of $f(c)$, so lies within a paralellepiped of volume

$$(2\epsilon\delta)(2M\delta)^{n-1}. \tag{61}$$

Now subdivide I^n into h^n cubes of edge length h^{-1} with h sufficiently large that $h^{-1}\sqrt{n} < \delta$. Apply the argument for each small cube, in which $||x - c|| \leq h^{-1}\sqrt{n} < \delta$. The number of cubes containing critical points is at most h^n, so this gives a total volume for $f(C)$ less than

$$(2\epsilon h^{-1}\sqrt{n})(2Mh^{-1}\sqrt{n})^{n-1}(h^n). \tag{62}$$

Since ϵ can be chosen arbitrarily small, $f(C)$ has measure zero. ☐

The argument above will not work for $\dim N < \dim M$; we need more control on the function f. In particular, one can find a C^1 function $I^2 \longrightarrow \mathbb{R}$ which fails to have critical values of measure zero. (Hint: find a C^1 function $f : \mathbb{R} \to \mathbb{R}$ with critical values containing the Cantor set $C \subset [0,1]$. Compose $f \times f$ with the sum $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and note that $C + C = [0,2]$.) As a result, Sard's theorem in general requires more differentiability of f.

Theorem 3.21 (Big Sard's theorem)**.** *Let $f : M \longrightarrow N$ be a C^k map of manifolds of dimension m, n, respectively. Let C be the set of critical points. Then $f(C)$ has measure zero if $k > \frac{m}{n} - 1$.*

Proof. As before, it suffices to show for $f : I^m \longrightarrow \mathbb{R}^n$. We do an induction on m – note that the theorem holds for $m = 0$.

Define $C_1 \subset C$ to be the set of points x for which $Df(x) = 0$. Define $C_i \subset C_{i-1}$ to be the set of points x for which $D^j f(x) = 0$ for all $j \leq i$. So we have a descending sequence of closed sets:

$$C \supset C_1 \supset C_2 \supset \cdots \supset C_k. \tag{63}$$

We will show that $f(C)$ has measure zero by showing

1. $f(C_k)$ has measure zero,
2. each successive difference $f(C_i \backslash C_{i+1})$ has measure zero for $i \geq 1$,
3. $f(C \backslash C_1)$ has measure zero.

Step 1: For $x \in C_k$, Taylor's theorem gives the estimate

$$||f(x+t) - f(x)|| \leq c||t||^{k+1}, \tag{64}$$

where c depends only on I^m and f.

Subdivide I^m into h^m small cubes with edge h^{-1}; then any point in in the small cube I_0 containing x may be written as $x + t$ with $||t|| \leq h^{-1}\sqrt{m}$. As a result, $f(I_0)$ is contained by a cube of edge $ah^{-(k+1)}$, with $a = 2cm^{(k+1)/2}$ independent of the small cube size. At most h^m cubes are necessary to cover C_k, and their images have total volume less than

$$h^m(ah^{-(k+1)})^n = a^n h^{m-(k+1)n}. \tag{65}$$

Assuming that $k > \frac{m}{n} - 1$, this tends to 0 as we increase the number of cubes.
Step 2: For each $x \in C_i \backslash C_{i+1}$, $i \geq 1$, there is a $i + 1^{th}$ partial, say wlog $\partial^{i+1} f_1 / \partial x_1 \cdots \partial x_{i+1}$, which is nonzero at x. Therefore the function

$$w(x) = \partial^i f_1 / \partial x_2 \cdots \partial x_{i+1} \tag{66}$$

vanishes on C_i but its partial derivative $\partial w / \partial x_1$ is nonvanishing near x. Then

$$(w(x), x_2, \ldots, x_m) \tag{67}$$

forms an alternate coordinate system in a neighbourhood V around x by the inverse function theorem (the change of coordinates is of class C^k), and we have trapped C_i inside a hyperplane. The restriction of f to $w = 0$ in V is clearly critical on $C_i \cap V$ and so by induction on m we have that $f(C_i \cap V)$ has measure zero. Cover $C_i \setminus C_{i+1}$ by countably many such neighbourhoods V.

Step 3: Let $x \in C \backslash C_1$. Note that we won't necessarily be able to trap C in a hypersurface. But, since there is some partial derivative, wlog $\partial f_1/\partial x_1$, which is nonzero at x, so defining $w = f_1$, we have that

$$(w(x), x_2, \ldots, x_m) \tag{68}$$

is an alternative coordinate system in some neighbourhood V of x (the coordinate change is a diffeomorphism of class C^k). In these coordinates, the hyperplanes $w = t$ in the domain are sent into hyperplanes $y_1 = t$ in the codomain, and so f can be described as a family of maps f_t whose domain and codomain has dimension reduced by 1. Since $w = f_1$, the derivative of f in these coordinates can be written

$$Df = \begin{pmatrix} 1 & 0 \\ * & Df_t \end{pmatrix}, \tag{69}$$

and so a point $x' = (t, p)$ in V is critical for f if and only if p is critical for f_t. Therefore, the critical values of f consist of the union of the critical values of f_t on each hyperplane $y_1 = t$ in the codomain. Since the domain of f_t has dimension reduced by one, by induction it has critical values of measure zero. So the critical values of f intersect each hyperplane in a set of measure zero, and by Fubini's theorem this means they have measure zero. Cover $C \setminus C_1$ by countably many such neighbourhoods. $\square$

Remark 3.22. Note that $f(C)$ is measurable, since it is the countable union of compact subsets (the set of critical values is not necessarily closed, but the set of critical points is closed and hence a countable union of compact subsets, which implies the same of the critical values.)

To show the consequence of Fubini's theorem directly, we can use the following argument. First note that for any covering of $[a, b]$ by intervals, we may extract a finite subcovering of intervals whose total length is $\leq 2|b - a|$. To see this, first choose a minimal subcovering $\{I_1, \ldots, I_p\}$, numbered according to their left endpoints. Then the total overlap is at most the length of $[a, b]$. Therefore the total length is at most $2|b - a|$.

Now let $B \subset \mathbb{R}^n$ be compact, so that we may assume $B \subset \mathbb{R}^{n-1} \times [a, b]$. We prove that if $B \cap P_c$ has measure zero in the hyperplane $P_c = \{x^n = c\}$, for any constant $c \in [a, b]$, then it has measure zero in $\mathbb{R}^n$.

If $B \cap P_c$ has measure zero, we can find a covering by open sets $R^i_c \subset P_c$ with total volume $< \epsilon$. For sufficiently small α_c, the sets $R^i_c \times [c - \alpha_c, c + \alpha_c]$ cover $B \cap \bigcup_{z \in [c-\alpha_c, c+\alpha_c]} P_z$ (since B is compact). As we vary c, the sets $[c - \alpha_c, c + \alpha_c]$ form a covering of $[a, b]$, and we extract a finite subcover $\{I_j\}$ of total length $\leq 2|b - a|$.

Let R^i_j be the set R^i_c for $I_j = [c - \alpha_c, c + \alpha_c]$. Then the sets $R^i_j \times I_j$ form a cover of B with total volume $\leq 2\epsilon|b - a|$. We can make this arbitrarily small, so that B has measure zero.