LECTURE 8: THE THEOREMS OF SARD AND WHITNEY

1. Sard's theorem

Let $f : M \to N$ be a smooth map. Recall that $q \in N$ is a regular value if for any $p \in f^{-1}(q)$, df_p is surjective. (Note that by this definition, any point $q \notin f(M)$ is a regular value.) Any $q \in N$ which is not a regular value is called a *critical value* of f

Today we will prove that almost all points in N are regular:

Theorem 1.1 (Sard's theorem)**.** *For any smooth map $f : M \to N$, the set of all critical values is of measure zero in N.*

Remark. The theorem does not claim that the set of critical points in M is a measure zero subset. In fact, if we consider a constant map $f(p) \equiv q_0 \in N$, then any point in M is a critical point. However, the set of critical points contains only one point in this case, which is of course of measure zero.

We need to explain the words "of measure zero" in the theorem. Recall

Definition 1.2. A subset $A \subset \mathbb{R}^n$ is of *measure zero* if for any $\varepsilon > 0$, there exists a countable union of open sets $U_i \in \mathbb{R}^n$ so that $A \subset \cup_i U_i$ and $\sum_i \text{volume}(U_i) < \varepsilon$.

Example. Any countable set is a measure zero set. In particular, $\mathbb{Q}$ has measure zero.

For measure zero sets, we have the following properties:

1. A countable union of measure zero sets is a measure zero set.
2. If A is a measure zero set, and f is smooth, then $f(A)$ is a measure zero set.
3. Fubini's theorem: Let A be a countable union of compact sets in $\mathbb{R}^m$ such that $A \cap (\{c\} \times \mathbb{R}^{m-r})$ has measure zero in $\mathbb{R}^{m-r}$ for all $c \in \mathbb{R}^r$. Then A has measure zero in $\mathbb{R}^m$.

Since any manifold M admits an atlas that contains only countable many charts, and each chart identify an open set in M with an open set in $\mathbb{R}^n$, it is reasonable to define

Definition 1.3. We say $A \subset M$ is a *measure zero* set if for any chart (φ, U, V) of M, $\varphi(A \cap U)$ is a measure zero set in V.

As a corollary of theorem 1.1, we see that if $f(M)$ is not a measure zero set in N, then for almost all point $q \in f(M)$, $f^{-1}(q)$ is a smooth submanifold of M. Another very important corollary is

Corollary 1.4. *If $f : M \to N$ is smooth and $\dim M < \dim N$, then $f(M)$ has measure zero in N.*

Proof of Sard's theorem. Since the definition of measure zero is local, it suffices to prove the theorem for the case where $M = U \subset \mathbb{R}^m$ and $N = V \subset \mathbb{R}^n$ are Euclidean open subsets. The theorem is certainly true for $m = 0$. We will proceed to prove that the theorem is true for m assuming that it is true for $m-1$. Let C be the set of all critical points of f. Denote

$$C_j = \{x \in U \mid \partial^\alpha f(x) = 0 \text{ for all } |\alpha| \le j\}.$$

Obviously for any k,

$$f(C) = f(C \setminus C_1) \cup f(C_1 \setminus C_2) \cup \cdots \cup f(C_{k-1} \setminus C_k) \cup f(C_k).$$

Following Milnor, we will divide the proof into three steps:

Step 1: $f(C \setminus C_1)$ has measure zero.
Step 2: $f(C_i \setminus C_{i+1})$ has measure zero.
Step 3: $f(C_k)$ has measure zero for $k > \frac{m}{n} - 1$.

Proof of step 1. For each $x \in C \setminus C_1$, we will find an open set $V \ni x$ such that $f(V \cap C)$ has measure zero. Since $C \setminus C_1$ is covered by countably many of such open sets (by second-countability), this implies $f(C \setminus C_1)$ is of measure zero.

Since $x \notin C_1$, there is some partial derivative, say $\frac{\partial f_1}{\partial x^1}$, is not zero at x. Consider

$$h : U \to \mathbb{R}^m, \quad h(x) = (f_1(x), x^2, \cdots, x^m).$$

Then dh_x is non-singular, so h maps a neighborhood V of x diffeomorphically onto an open set V' in $\mathbb{R}^m$. The composition $g = f \circ h^{-1}$ will then map V' into $\mathbb{R}^n$ so that the critical values of g is $f(V \cap C)$.

Note that the map g we constructed is of the form

$$g(t, x^2, \cdots, x^m) = (t, y^2, \cdots, y^n).$$

Sor for each t, g induces a map $g_t : (\{t\} \times \mathbb{R}^{m-1}) \cap V' \to \{t\} \times \mathbb{R}^{n-1}$. Moreover,

$$\left(\frac{\partial g_i}{\partial x^j}\right) = \begin{pmatrix} 1 & 0 \\ * & \left(\frac{\partial (g_t)_i}{\partial x^j}\right) \end{pmatrix}.$$

It follows that a point of $\{t\} \times \mathbb{R}^{m-1}$ is critical for g_t if and only if it is critical for g. By induction, Sard's theorem is true for $m-1$, so the set of critical values of g_t has measure zero. Finally by applying Fubini's theorem, we see that the set of critical values of g is of measure zero.

Proof of step 2. For each $x \in C_i \setminus C_{i+1}$, one can find some i^{th} partial derivative of f, denoted by w, that vanishes on C_i but has a first derivative, say $\frac{\partial w}{\partial x^1}$, that does not vanish. Again the map

$$h : U \to \mathbb{R}^m, \quad h(x) = \big(w(x), x^2, \cdots, x^m\big)$$

maps a neighborhood V of x diffeomorphically onto an open set V'. By construction, h carries $C_i \cap V$ into the hyperplane $\{0\} \times \mathbb{R}^{m-1}$. Again we consider the map $g = f \circ h^{-1}$. Then the critical points of g of type C_i are all in the hyperplane $\{0\} \times \mathbb{R}^{m-1}$. Let

$$\bar{g} : (\{0\} \times \mathbb{R}^{m-1}) \cap V' \to \mathbb{R}^n$$

be the restriction of g. By induction, the set of critical values of $\bar{g}$ is of measure zero in $\mathbb{R}^n$. Moreover the critical points of g of type C_i are obviously critical points of $\bar{g}$. It follows that the image of these critical points of g is of measure zero. Therefore, $f(C_i \cap V)$ is of measure zero. Since $C_i \setminus C_{i+1}$ can be covered by countable many such sets V, $f(C_i \setminus V_{i+1})$ is of measure zero.

Proof of step 3. Let $Q \subset U$ be a cube whose sides are of length δ. We will prove that for $k > \frac{m}{n} - 1$, $f(C_k \cap Q)$ has measure zero. Since C_k can be covered by countably many such cubes, this implies $f(C_k)$ has measure zero.

From Taylor's theorem, the compactness of Q and the definition of C_k, we see that

$$f(x+h) = f(x) + R(x,h),$$

where $|R(x,h)| < a|h|^{k+1}$ for $x \in C_k \cap Q, x+h \in Q$, and the constant a depends only on f and Q. Now we subdivide Q into r^m cubes whose sides are of length $\frac{\delta}{r}$. Let Q_1 be a cube of subdivision that contains a point $x \in C_k$. Then any point of Q_1 can be written as $x+h$ with $|h| < \sqrt{m}\frac{\delta}{r}$. It follows that $f(Q_1)$ lies in a cube with sides of length $\frac{b}{r^{k+1}}$ centered about $f(x)$, where $b = 2a(\sqrt{m}\delta)^{k+1}$ is a constant. So $f(C_k \cap Q)$ is contained in the union of at most r^m cubes having total volume

$$\mathrm{Vol} \le r^m (\frac{b}{r^{k+1}})^n = b^n r^{m-(k+1)n}.$$

Since $k > \frac{m}{n} - 1$, we see $\mathrm{Vol} \to 0$ as $r \to \infty$. It follows that $f(C_k \cap Q)$ is of measure zero. □

2. The Whitney embedding theorems

Let M be a smooth manifold of dimension m. A natural question is: which manifolds can be embedded into $\mathbb{R}^N$ as smooth submanifolds?

Theorem 2.1 (The Whitney embedding theorem: easiest version)**.** *Any compact manifold M can be embedded into $\mathbb{R}^N$ for sufficiently large N.*

Proof. Let $\{\varphi_i, U_i, V_i\}_{1 \le i \le k}$ be a finite set of coordinate charts on M so that $\mathcal{U} = \{U_i \mid 1 \le i \le k\}$ is an open cover of M. Let $\{\rho_i \mid 1 \le i \le k\}$ be a partition of unity subordinate to $\mathcal{U}$. Let $\tilde{\varphi}_i = \rho_i \varphi_i : M \to \mathbb{R}^m$. Define

$$\Phi : M \to \mathbb{R}^{k(m+1)}, \quad p \mapsto (\tilde{\varphi}_1(p), \cdots, \tilde{\varphi}_k(p), \rho_1(p), \cdots, \rho_k(p)).$$

We claim that Φ is an injective map. In fact, suppose $\Phi(p_1) = \Phi(p_2)$. Take an index i so that $\rho_i(p_1) = \rho_i(p_2) \ne 0$. Then $p_1, p_2 \in \mathrm{supp}(\rho_i) \subset U_i$. It follows that $\varphi_i(p_1) = \varphi_i(p_2)$. So we must have $p_1 = p_2$ since φ_i is bijective.

Next let's prove that Φ is an immersion. In fact, for any $X_p \in T_pM$,

$$\begin{aligned} d\Phi_p(X_p) = (&X_p(\rho_1)\varphi_1(p) + \rho_1(p)(d\varphi_1)_p(X_p), \cdots, \\ &X_p(\rho_k)\varphi_k(p) + \rho_k(p)(d\varphi_k)_p(X_p), X_p(\rho_1), \cdots, X_p(\rho_k)). \end{aligned}$$

It follows that if $d\Phi_p(X_p) = 0$, then $X_p(\rho_i) = 0$ for all i, and thus $\rho_i(p)(d\varphi_i)_p(X_p) = 0$ for all i. Pick an index i so that $\rho_i(p) \ne 0$. We see $(d\varphi_i)_p(X_p) = 0$. Since φ_i is a diffeomorphism, we conclude that $X_p = 0$. So $d\Phi$ is injective.

Since Φ is an injective immersion, and M is compact, Φ must be an embedding. $\square$

As an application of Sard's theorem, we can prove

Theorem 2.2 (The Whitney embedding theorem: median version)**.** *Any compact manifold M of dimension m can be embedded into $\mathbb{R}^{2m+1}$ and immersed into $\mathbb{R}^{2m}$.*

Proof. Suppose we already have an embedding $\Phi : M \to \mathbb{R}^N$ with $N > 2m+1$. We will show that we can produce an embedding of M in $\mathbb{R}^{N-1}$.

To do so, for any $[v] \in \mathbb{RP}^{N-1}$, we let

$$P_{[v]} = \{u \in \mathbb{R}^N \mid u \cdot v = 0\} \simeq \mathbb{R}^{N-1}$$

be the orthogonal complement of $[v]$ in $\mathbb{R}^N$. Let $\Psi_{[v]} : \mathbb{R}^N \to P_{[v]}$ be the orthogonal projection to this hyperplane. We claim that the set of $[v]$'s for which $\Phi_{[v]} = \Psi_{[v]} \circ \Phi$ is not an embedding has measure zero in $\mathbb{RP}^{N-1}$, hence it is possible to choose $[v]$ so that $\Phi_{[v]}$ is an embedding. Note that if $\Phi_{[v]}$ fails to be an embedding, we must have either $\Phi_{[v]}$ is not injective, or $\Phi_{[v]}$ is not an immersion.

First let's consider $[v]$'s so that $\Phi_{[v]}$ is not injective. Then one can find $p_1 \neq p_2$ so that $\Phi_{[v]}(p_1) = \Phi_{[v]}(p_2)$, i.e. $0 \neq \Phi(p_1) - \Phi(p_2)$ lies in the line $[v]$. In other words, $[v] = [\Phi(p_1) - \Phi(p_2)]$. So $[v]$ must lie in the image of the map

$$\alpha : (M \times M) \setminus \Delta_M \to \mathbb{RP}^{N-1}, \quad (p_1, p_2) \mapsto [\Phi(p_1) - \Phi(p_2)],$$

where $\Delta_M = \{(p,p) \mid p \in M\}$ is the "diagonal" in $M \times M$. Since $(M \times M) \setminus \Delta_M$ is of dimension $2m < N-1$, Sard's theorem implies that the image of α is of measure zero in $\mathbb{RP}^{N-1}$.

Next let's consider $[v]$'s so that $\Phi_{[v]}$ is not an immersion. Then there exists some $p \in M$ and some $0 \neq X_p \in T_pM$ so that $(d\Phi_{[v]})_p(X_p) = 0$, i.e. $(d\Psi_{[v]})_{\Phi(p)}(d\Phi)_p(X_p) = 0$. Since $\Psi_{[v]}$ is linear, $d\Psi_{[v]} = \Psi_{[v]}$. It follows that $0 \neq (d\Phi)_p(X_p)$ is in $[v]$, i.e. $[v] = [(d\Phi)_p(X_p)]$. In other words, $[v]$ lies in the image of

$$\beta : TM \setminus \{0\} \to \mathbb{RP}^{N-1}, \quad (p, X_p) \mapsto [(d\Phi)_p(X_p)],$$

where $TM \setminus \{0\} = \{(p, X_p) \mid X_p \neq 0\}$ is an open submanifold of TM. Again since TM has dimension $2m < N-1$, by Sard's theorem, the image of β is of measure zero.

To see that M can be immersed into $\mathbb{R}^{2m}$, we first embed M into $\mathbb{R}^{2m+1}$, then repeat the last step, with the modification that we choose $X_p \in T_pM$ so that $|X_p| = 1$. $\square$

Remark. More generally one has

- (**The Whitney embedding theorem: regular form**) Any smooth manifold of dimension m can be immersed into $\mathbb{R}^{2m}$ and embedded into $\mathbb{R}^{2m+1}$.
- (**The Strong Whitney embedding theorem**) Any smooth manifold of dimension m can be immersed into $\mathbb{R}^{2m-1}$ and embedded into $\mathbb{R}^{2m}$.
- Any compact orientable m-manifold can be embedded into $\mathbb{R}^{2m-1}$.
- For $m \neq 2^k$, any smooth m-manifold can be embedded into $\mathbb{R}^{2m-1}$. (But if $m = 2^k$, $\mathbb{RP}^m$ cannot be embedded into $\mathbb{R}^{2m-1}$).
- Any smooth m manifold can be immersed into $\mathbb{R}^{2m-a(m)}$, where $a(m)$ is the number of 1's that appear in the binary expansion of m.