LECTURE 7: SARD'S THEOREM

1. Proof of Sard's Theorem

Now we are ready to prove

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.*

Note that if $n = \dim N = 0$, then there is no critical points (and thus no critical values) at all. So in what follows we may assume $n > 0$.

Proof. Since we may take at most countable many coordinate charts U_i covering M so that each $f(U_i) \subset X_i$ for some coordinate charts X_i of N, and since a countable union of measure zero sets is still of measure zero, it suffices to prove the theorem for the case that $M = U \subset \mathbb{R}^m$ and $N = X \subset \mathbb{R}^n$ are Euclidean open subsets.

First observe that if $m < n$, then the theorem holds trivially. In fact, in this case we can prove that the whole image $f(U)$ is of measure zero in $\mathbb{R}^n$. To see this, we identify U with the subset $U \times \{0\}$ inside $U \times \mathbb{R}^{n-m} \subset \mathbb{R}^n$. Obviously $U \times \{0\}$ has measure zero in $\mathbb{R}^n$. Now define a map $\tilde{f} : U \times \mathbb{R}^{n-m} \to \mathbb{R}^n$ via $\tilde{f}(x, y) = f(x)$. Then $f(U) = \tilde{f}(U \times \{0\})$. ($\tilde{f}$ is smooth since it is the composition of two smooth maps: the projection map $\pi : \mathbb{R}^n \to \mathbb{R}^m$ and $f : U \to \mathbb{R}^n$.) Since $\tilde{f}$ is a smooth between spaces of the same dimension, and $U \times \{0\}$ is of measure zero in $\mathbb{R}^n$, we conclude that the image $\tilde{f}(U \times \{0\})$ is of measure zero in $\mathbb{R}^n$.

We will proceed by induction. The theorem is certainly true for $m = 0$, since any countable set has measure zero. 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, then we need to show that $f(C)$ is of measure zero in N. Denote

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

Obviously for any positive integer k,

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

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

Step 1: $f(C \backslash C_1)$ has measure zero.
Step 2: $f(C_i \backslash C_{i+1})$ has measure zero for each i.
Step 3: $f(C_k)$ has measure zero for large k, say for $k \ge \frac{m}{n}$.

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

Note that if $n = 1$, then a point x is a critical point of f if and only if $\frac{\partial f}{\partial x^i}(x) = 0$ for all i. It follows $C\setminus C_1 = \emptyset$, so $f(C\setminus C_1)$ must be of measure zero. So we may assume $m \geq n > 1$. 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. According to the inverse mapping theorem, h maps a neighborhood U_x of x diffeomorphically onto an open set V in $\mathbb{R}^m$. The composition $g = f \circ h^{-1}$ will then map V into $\mathbb{R}^n$. Moreover, since h^{-1} is a diffeomorphism on V, dh^{-1} is a linear isomorphism everywhere in V. So the set of critical values of g is exactly $f(U_x \cap C)$.

Note that the map g we constructed is of the form

$$g(t, x^2, \cdots, x^m) = (t, g_2, \cdots, g_n).$$

So 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,

$$dg = \begin{pmatrix} 1 & 0 \\ * & \left(\frac{\partial (g_t)_i}{\partial x^j}\right)_{i,j\geq 2} \end{pmatrix}.$$

It follows that a point in $(\{t\} \times \mathbb{R}^{m-1}) \cap V$ is critical for g_t if and only if it is critical for g. But by the induction hypothesis, Sard's theorem is true for $m-1$, i.e. is true for each g_t. So the set of critical values of g_t has measure zero in $\{t\} \times \mathbb{R}^{n-1}$. 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 multi-index α with $|\alpha| = i$, so that

- the partial derivative $w := \partial^\alpha f$ vanishes on C_i,
- at least one first order partial derivative of w, say $\frac{\partial w}{\partial x^1}$, does not vanish at x.

Again by applying the inverse mapping theorem, we conclude that

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

maps a neighborhood U_x of x diffeomorphically onto an open set V in $\mathbb{R}^m$. (To get a better understand and/or to avoid possible mistakes, you may want to think about the meaning of this map for $m = 1$.) By construction, h carries $C_i \cap U_x$ 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. Then the set of critical points of g of type C_i coincides with the set of critical points of $\bar{g}$. By induction, the set of critical values of $\bar{g}$ is of measure zero in $\mathbb{R}^n$. It follows that the image of the critical points of g of type C_i is of measure zero. Therefore, $f(C_i \cap U_x)$ is of measure zero. Since $C_i \setminus C_{i+1}$ can be covered by countable many such sets U_x, $f(C_i \setminus C_{i+1})$ 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 that $f(C_k)$ 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. □

Remark. On the other hand, the set of critical values could be a dense subset. For example, we can list all rational numbers as $\mathbb{Q} = \{r_1, r_2, \cdots\}$. Then we take a smooth function f_0 defined on $\mathbb{R}$ such that

$$\mathrm{supp}(f_0) \subset (-1/3, 1/3) \quad \text{and} \quad f_0 \equiv 1 \text{ on } (-1/4, 1/4).$$

Let

$$f(x) = \sum_{k=1}^{\infty} r_k f_0(x-k).$$

Then each $k \in \mathbb{N}$ is a critical point of f, and thus the set of critical values of f contains $f(\mathbb{N}) = \mathbb{Q}$.

Remark. Sard's theorem holds for C^r maps between C^r manifolds, provided $r \ge 1 + \max(m-n, 0)$. Moreover, this is sharp. However, the proof above does not work in this more general setting. (Can you tell why?)

2. An alternative form of Sard's Theorem

Next we will give an alternative form of Sard's Theorem. (Reason: In many applications, one has to handle maps between infinitely dimensional manifolds, called Banach manifolds. The original form of Sard's Theorem cannot be extended to this setting because we don't have a Lebesgue measure in the infinite dimensional setting.)

Before we state the alternative form, we first give a definition that can be viewed as the analogue in general topological spaces of the word "of full measure" in measure spaces:

Definition 2.1. A set A in a topological space is called *residual* (or "*second Baire category*") if it is a countable intersection of dense open sets.

For example, the set of all irrational numbers, $\mathbb{R}\backslash\mathbb{Q}$, is a residual set in $\mathbb{R}$. However, the set of rational numbers, $\mathbb{Q}$, is not residual in $\mathbb{R}$. Usually when we say some property is "generic", we mean that property holds for a residual set.

We now prove

Theorem 2.2 (Sard's Theorem, an alternative form). *Let $f : M \to N$ be smooth. Then the set of regular values of f is residual in N.*

[Idea]: Since the set of regular values has full measure, it has to be dense. It remains to show that the set of regular values is "nearly open". This is reasonable, since the set of regular points is open:

$$x \text{ is a regular point} \Leftrightarrow f \text{ is a submersion at } x \Leftrightarrow f \text{ is a submersion near } x.$$

However, this is not enough, since

$$f(\{\text{regular points}\}) \neq \{\text{regular values}\}.$$

However, we do have

$$f(\{\text{critical points}\}) = \{\text{critical values}\}$$

and we know $\{$critical points$\}$ is closed. Unfortunately this is still not enough to deduce that $\{$critical values$\}$ is closed, because in general

$$f(\text{closed set}) \neq \text{closed set}.$$

(See the example at the end of last section.) The solution to this last issue is to apply a standard trick:

$$f(\text{compact set}) = \text{compact set}.$$

Proof. Again we denote the set of critical points of f by C. We will show that every $x \in C$ admits a neighborhood U_x such that when restricted to $\overline{U}_x$, the set of regular values of f, $N\backslash f(C\cap\overline{U_x})$, is dense and open in N. Since C can be covered by at most countable many such open sets $\{U_1, U_2, \cdots\}$, we conclude that the set of regular values

$$N\backslash f(C) = N\backslash\bigcup_i f(C\cap\overline{U_i}) = \bigcap_i N\backslash f(C\cap\overline{U_i}),$$

which is a residual set.

To show the existence of such U_x, we choose a coordinate neighborhood U_x of x so that $\overline{U_x}$ is compact. Since C is closed in M (see the explanation before the proof), we see that $C\cap\overline{U_x}$ is compact. So $f(C\cap\overline{U_x})$ is compact, and thus closed in N. It follows that $N\backslash f(C\cap\overline{U_x})$ is open in N.

On the other hand, according to Sard's theorem we proved, $f(C\cap\overline{U_x})$ has measure zero. So for any $y \in N$ and any small neighborhood V of y in Y, one must have

$$V \cap (N\backslash f(C\cap\overline{U_x})) \neq \emptyset,$$

otherwise $V \subset f(C\cap\overline{U_x})$, which contradicts with the fact that $f(C\cap\overline{U_x})$ has measure zero. It follows that $N\backslash f(C\cap\overline{U_x})$ is dense in N. $\square$