Chapter 12

Manifolds and Sard’s Theorem

This chapter introduces basic concepts—‘manifold,’ ‘tangent vector,’ ‘smooth map,’ ‘derivative’—from differential topology. Those who have not been introduced to manifolds before may find the treatment rather terse, and the books by Milnor [Mil65] and Guillemin and Pollack [GP74] are highly recommended. On the other hand, the definitions presented below are the central concepts of the subject, and are basically nothing more than an intuitively natural extensions of familiar ideas of multivariable calculus, with many obvious applications, and those who have already been exposed to these definitions, and perhaps also the additional concepts developed in Chapter ??, will probably be able to advance more rapidly in their study of those books, which pursue the subject to much greater depth.

12.1 Manifolds

The manifold concept generalizes the notions of “curve” and “surface.” It is of fundamental importance in modern mathematics and physics. Among the many reasons for this is that manifolds are the natural host spaces of a “coordinate free” multivariable calculus that does not presume a space with a “flat” geometry.

**Definition 12.1.1** A set $M \subset \mathbb{R}^k$ is an $m$-dimensional $C^r$ ($1 \leq r \leq \infty$) manifold if, for each $p \in M$, there is an open set $U \subset \mathbb{R}^m$ and a $C^r$ injection $\varphi : U \to M$ whose inverse $\varphi^{-1} : \varphi(U) \to U$ has a $C^r$ extension to an open $W \subset \mathbb{R}^k$ such that $M \cap W = \text{image } \varphi$. In this circumstance
we say that \( \varphi \) is a \( C^r \) **parameterization** of \( \varphi(U) \), and that \( \varphi^{-1} \) is a \( C^r \) **coordinate chart** for \( \varphi(U) \).

Note that a \( C^r \) manifold is a \( C^s \) manifold for any \( 1 \leq s < r \). Also, although this definition would make sense if \( r = 0 \), the standard definition of a \( C^0 \) manifold is weaker. Restricting to \( r \geq 1 \) avoids the technical complications arising from the possibility (e.g., the Alexander horned sphere [Ale24]) that a \( C^0 \) manifold may have what is known as a “wild” embedding in a Euclidean space.

The most basic examples of \( C^r \) manifolds are:

(i) A set \( S \subset \mathbb{R}^k \) is a 0-dimensional \( C^\infty \) manifold if and only if each \( p \in S \) has a neighborhood \( W \) such that \( S \cap W = \{ p \} \).

(ii) Any open set \( U \subset \mathbb{R}^m \) is an \( m \)-dimensional \( C^\infty \) manifold: let \( \varphi := \text{Id}_U \).

(iii) If \( U \subset \mathbb{R}^m \) is open and \( \phi : U \rightarrow \mathbb{R}^{k-m} \) is \( C^r \), then \( \text{Gr}(\phi) = \text{image} \varphi \) for \( \varphi(x) := (x, \phi(x)) \), so \( \text{Gr}(\phi) \) is an \( m \)-dimensional \( C^r \) manifold.

Note that the empty set is an \( m \)-dimensional \( C^\infty \) manifold for any \( m \). For the most part this aspect of the definition works well. There are results for which this case is exceptional, and sometimes authors do not make nonemptiness an explicit hypothesis, expecting the reader to see that the assertion in question would be silly when \( M = \emptyset \).

The implicit function theorem gives conditions under which a level set of a \( C^r \) function is the graph of a \( C^r \) function and thus a \( C^r \) manifold. This principle generates many important examples, including objects from economic analysis such as indifference curves, isoprofit curves, and so forth. The converse—that any \( C^r \) manifold is locally a level set of a \( C^r \) function—is often useful. In detail:

**Lemma 12.1.2** For a set \( M \subset \mathbb{R}^k \) the following are equivalent:

(a) for each \( p \in M \), there is a neighborhood of \( Y \subset \mathbb{R}^k \) of \( p \) and a \( C^r \) function \( \Phi : Y \rightarrow \mathbb{R}^{k-m} \) such that \( D\Phi(p) \) has rank \( k-m \) and

\[
M \cap Y = \Phi^{-1}(0).
\]

(b) \( M \) is an \( m \)-dimensional \( C^r \) manifold.

**Proof:** ((a) \( \Rightarrow \) (b)) Fixing \( p \in M \), suppose that \( Y \) and \( \Phi : Y \rightarrow \mathbb{R}^{k-m} \) are as above. Let \( e_1, \ldots, e_k \) be the standard unit basis of \( \mathbb{R}^k \). There exist
$k-m$ of these basis vectors whose images under $D\Phi(p)$ are linearly independent, so, by reindexing, we can assume without loss of generality that $D\Phi(p)e_{m+1}, \ldots, D\Phi(p)e_k$ are linearly independent. Let $p = (x_0, y_0)$ where $x_0$ and $y_0$ are the projections of $p$ on the coordinate subspaces of the first $m$ and last $p-m$ coordinates respectively. The implicit function implies the existence of an open neighborhood $U \subset \mathbb{R}^m$ of $x_0$, an open neighborhood $V \subset \mathbb{R}^{k-m}$ of $y_0$, and a $C^r$ function $\varphi : U \to V$ such that $\varphi(x_0) = y_0$ and

$$\{ (x, \varphi(x)) : x \in U \} = \Phi^{-1}(0) \cap (U \times V).$$

((b) $\Rightarrow$ (a)) Let $\varphi : U \to M$ be a parameterization of $\varphi(U)$, and let $W \subset \mathbb{R}^k$ be open with $W \cap M = \varphi(U)$. Since $\varphi^{-1}$ is also $C^r$, the chain rule implies that the matrix of $D\varphi(\varphi^{-1}(p))$ has full rank. Therefore some $m \times m$ submatrix has full rank, and after relabelling of coordinates we may assume that it is the submatrix associated with the basis vectors $e_1, \ldots, e_m$. This submatrix is the matrix of $D(\pi \circ \varphi)(\varphi^{-1}(p))$ where $\pi$ is the projection of $\mathbb{R}^k$ onto the first $m$ coordinates. The inverse function theorem implies that $\pi \circ \varphi$ is locally invertible, with $C^r$ inverse, near $\pi(p)$. Let $\psi : V \to U$ be such an inverse, where $V$ is a neighborhood of $\pi(p)$ in the coordinate subspace of the first $m$ coordinates. Let

$$\Phi : (V \times \mathbb{R}^{k-m}) \cap W \to \mathbb{R}^{k-m}$$

be the function $\Phi(v, z) = z - \varphi(\psi(v))$. There may be points in $(V \times \mathbb{R}^{k-m}) \cap W \cap M$ that are not in $\Phi^{-1}(0)$ but the implicit function theorem implies that this will not be a problem if we restrict $\Phi$ to a sufficiently small neighborhood $Y$ of $p$.

12.2 $C^r$ Maps, Tangent Vectors, and Derivatives

There is a category whose objects are $C^r$ manifolds, with the following morphisms:

**Definition 12.2.1** If $M \subset \mathbb{R}^k$ and $N \subset \mathbb{R}^l$ are $C^r$ manifolds, then a $C^r$ map from $M$ to $N$ is a function $f : M \to N$ such that there is a neighborhood $W \subset \mathbb{R}^k$ of $M$ on which is defined a $C^r$ function $F : W \to \mathbb{R}^l$ that extends $f$ in the sense $f$ and $F$ agree on $M \cap W$.

(In fact this is the standard definition of what it means for a function $f : S \to \mathbb{R}^l$ to be $C^r$, where $S$ may be any subset of $\mathbb{R}^k$.) Note that a $C^r$ map $f : M \to N$ is a $C^s$ map for any $1 \leq s < r$. 


12.2. $C^r$ Maps, Tangent Vectors, and Derivatives

We quickly verify that, indeed, $C^r$ manifolds and $C^r$ maps constitute a category. Of course the identity function on a $C^r$ manifold is obviously a $C^r$ map. If $f : M \to N$ is $C^r$, $P \subset \mathbb{R}^r$ is another $C^r$ manifold, and $g : N \to P$ is a $C^r$ map, then $g \circ f : M \to P$ is $C^r$, simply because the composition of a $C^r$ extension of $f$ to a neighborhood of $M$ with a $C^r$ extension of $g$ to a neighborhood of $N$ is $C^r$ as a matter of multivariable calculus.

There are many notions of “derivative” in mathematics. Invariably the term refers to a linear approximation of a function that is accurate “up to first order.” The first step in defining the derivative of a $C^r$ map between manifolds is to specify the vector spaces that serve as the linear approximation’s domain and range. Visually, the tangent space of a manifold $M \subset \mathbb{R}^k$ at a point $p$ is the linear subspace of $\mathbb{R}^k$ that is parallel to the tangent plane of $M$ at $p$.

**Definition 12.2.2** If $M \subset \mathbb{R}^k$ is an $m$-dimensional $C^1$ manifold, $p \in M$, $\varphi : U \to M$ is a $C^1$ parameterization where $U \subset \mathbb{R}^m$ is open, and $\varphi^{-1}(p) = p$, then the **tangent space** of $M$ at $p$ is

$$T_pM := \text{image}(D\varphi(\varphi^{-1}(p))).$$

To see that this definition does not depend on the choice of $\varphi$, consider that for any another $\varphi'$ we have

$$\text{image}(D\varphi'(\varphi'^{-1}(p))) = \text{image}(D(\varphi \circ \varphi^{-1} \circ \varphi')(\varphi'^{-1}(p)))$$

$$\subset \text{image}(D\varphi(\varphi^{-1}(p)))$$

by the chain rule, and of course the opposite inclusion holds by symmetry.

**Definition 12.2.3** If $f : M \to \mathbb{R}^l$ is a $C^1$ function and $F$ is a $C^1$ extension of $f$ to a neighborhood of $p$, the **derivative of $f$ at $p$**, denoted by $Df(p)$, is the linear function

$$DF(p)|_{T_pM} : T_pM \to \mathbb{R}^l.$$ 

To see that this definition does not depend on the choice of extension observe that, for any $v \in T_pM$, there is some $w \in \mathbb{R}^m$ such that $v = D\varphi(x_0)w$, so

$$DF(p)v = DF(p)(D\varphi(x_0)w) = D(F \circ \varphi)(x_0)w = D(f \circ \varphi)(x_0)w.$$ 

[Insert Independence of Extension figure here.]
**Lemma 12.2.4** If \( f : M \to N \) is a \( C^1 \) map, where \( M \subset \mathbb{R}^k \) and \( N \subset \mathbb{R}^l \) are \( C^1 \) manifolds, then, for each \( p \in M \), the image of \( Df(p) \) is contained in \( T_{f(p)}N \).

**Proof:** Suppose that \( V \subset \mathbb{R}^n \) is open, \( \varphi : U \to M \) is a \( C^r \) parameterization of a neighborhood of \( p \), and \( \psi : V \to N \) is a \( C^r \) parameterization of a neighborhood of \( f(p) \). Applying the chain rule twice,

\[
Df(p)(T_pM) = Df(p)(\text{image}(D\varphi(\varphi^{-1}(p)))) \\
= \text{image}(D(f \circ \varphi)(\varphi^{-1}(p))) \\
= \text{image}(D(\psi \circ \psi^{-1} \circ f \circ \varphi)(\varphi^{-1}(p))) \\
\subset \text{image}(D\psi(\psi^{-1}(f(p)))) = T_{f(p)}N.
\]

\[\square\]

In virtually every version of differential calculus the chain rule is the most important basic result. We expect that many readers have seen the following result, and at worst it is a suitable exercise, following from the chain rule of multivariable calculus without trickery, so we give no proof.

**Proposition 12.2.5** If \( M \subset \mathbb{R}^k \), \( N \subset \mathbb{R}^l \), and \( P \subset \mathbb{R}^m \) are \( C^1 \) manifolds, and \( f : M \to N \) and \( g : N \to P \) are \( C^1 \) maps, then, at each \( p \in M \),

\[
D(g \circ f)(p) = Dg(f(p)) \circ Df(p).
\]

By analogy with functions from \( \mathbb{R} \) to itself, it seems natural to try to regard the derivative of a \( C^1 \) map \( f : M \to N \) as a function in its own right. Briefly, the **tangent space** of \( M \) is

\[
TM := \{ (p,v) \in M \times \mathbb{R}^k : p \in M \text{ and } v \in T_pM \}
\]

and the **tangent map** \( Tf : TM \to TN \) is the function

\[
Tf : (p,v) \mapsto (f(p),Df(p)v).
\]

As a subset of \( \mathbb{R}^k \subset \mathbb{R}^k \), \( TM \) is a metric space, and (Problem ????) the continuous differentiability of \( f \) implies that \( Tf \) is a continuous function. If \( f \) is \( C^r \) for \( r \geq 2 \), then (Problem ????) \( TM \) is a \( C^{r-1} \) manifold and \( Tf \) is a \( C^{r-1} \) map. If \( g : N \to P \) is another \( C^1 \) map, where \( P \subset \mathbb{R}^m \) is a \( C^1 \) manifold, then (Problem ????) \( T(g \circ f) = Tg \circ Tf \). This is the least trivial aspect of the verification that, if \( r \geq 2 \), \( T \) is a **covariant functor from the category**
of $C^r$ manifolds and $C^r$ maps to the category of $C^{r-1}$ manifolds and $C^{r-1}$ maps.

Injections, surjections, and bijections occur frequently throughout mathematics, of course. For smooth maps there is a more sophisticated terminology, reflecting the fact that the properties of interest may hold locally.

**Definition 12.2.6** If $f : M \to N$ is a $C^1$ map, where $M \subseteq \mathbb{R}^k$ is an $m$-dimensional $C^1$ manifold and $N \subseteq \mathbb{R}^l$ is an $n$-dimensional $C^1$ manifold, then $f$ is:

(a) an **immersion** if rank $Df(p) = m$ for all $p \in M$;

(b) an **embedding** if it is an injective immersion;

(c) a **submersion** if rank $Df(p) = n$ for all $p \in M$;

(d) a **local diffeomorphism** if it is an immersion and a submersion;

(e) a **diffeomorphism** if it is a bijective local diffeomorphism.

The next result, which asserts that local diffeomorphisms are locally invertible, is basically nothing more than a restatement of the inverse function theorem, and its proof is a matter of passing to coordinate charts, then applying the familiar version of that result.

**Lemma 12.2.7** If $f : M \to N$ is a $C^r$ local diffeomorphism, then for each $p \in M$ there are neighborhoods $W \subseteq M$ of $p$ and $Z \subseteq N$ of $f(p)$ such that $f|_W$ is a bijection between $W$ and $Z$, and $(f|_W)^{-1}$ is $C^r$.

**Proof:** Assume that $M$ and $N$ are $m$-dimensional. Let $\varphi : U \to M$ and $\psi : V \to N$ be $C^r$ parameterizations with $p \in \varphi(U)$ and $f(p) \in \psi(V)$. Since

$$D\psi(\psi^{-1}(f(p))) \circ D(\psi^{-1} \circ f \circ \varphi)(\varphi^{-1}(p)) = D(f \circ \varphi)(\varphi^{-1}(p)),$$

the image of $D(\psi^{-1} \circ f \circ \varphi)(\varphi^{-1}(p))$ must be all of $\mathbb{R}^m$. The inverse function theorem implies the existence of neighborhoods $W' \subseteq U$ of $\varphi^{-1}(p)$ and $Z' \subseteq V$ of $\psi^{-1}(f(p))$ such that $(\psi^{-1} \circ f \circ \varphi)^{-1}|_{Z'}$ is a $C^r$ function with image $W'$. Letting $W = \varphi(W')$ and $Z = \psi(Z')$, $f^{-1}_Z : Z \to W$ is $C^r$ because it is the composition $\varphi \circ (\psi^{-1} \circ f \circ \varphi)^{-1} \circ \psi^{-1}|_Z$. 

\[\square\]
Typically a category has a notion of isomorphism, and the properties that one studies are those that are preserved by isomorphism. Differential topology may be thought of as the study of properties of $C^r$ manifolds and $C^r$ maps that are unaffected if a space is replaced by another manifold that is diffeomorphic to it, or a map is replaced by its compositions with a diffeomorphism. Of course the sensibility of this depends on the following fact, which follows immediately from the last result.

**Lemma 12.2.8** If $f : M \to N$ is a $C^r$ diffeomorphism, then $f^{-1} : N \to M$ is also a $C^r$ diffeomorphism.

In view of the close relationship between the inverse function theorem and the implicit function theorem, and the importance of the latter result, one might guess that there is also a version of the implicit function theorem for manifolds, and that it is applied frequently. A point $p$ is a **regular point** of a $C^r$ map $f : M \to N$ if the image of $Df(p)$ is all of $T_{f(p)}N$. (Note that regularity depends on the specification of the range in that, if $N \subset \mathbb{R}^t$, a regular point of $f$ will not typically be regular when we reinterpret $f$ as a function from $M$ to $\mathbb{R}^t$.) A point $q \in N$ is a **regular value** if every point of $f^{-1}(q)$ is a regular point of $f$. This terminology is a bit paradoxical insofar as a point is necessarily a regular value when it is *not* a “value” in the sense that $f^{-1}(q) = \emptyset$.

**Definition 12.2.9** If $M$ and $P$ are $C^r$ manifolds, and $P \subset M$, then $P$ is said to be a $C^r$ **submanifold** of $M$. If the dimensions of $M$ and $P$ are $m$ and $p$ respectively, then $m - p$ is the **codimension** of $P$.

**Theorem 12.2.10** (Regular Value Theorem) If $q$ is a regular value of a $C^r$ map $f : M \to N$, where $M \subset \mathbb{R}^k$ and $N \subset \mathbb{R}^t$ are $m$- and $n$-dimensional manifolds, then $f^{-1}(q)$ is a $C^r$ submanifold of $M$, and its codimension is $n$.

**Proof:** Fix a point $p \in f^{-1}(q)$. (If $f^{-1}(q) = \emptyset$, then the claim holds trivially.)

Let $\varphi : U \to M$ and $\psi : V \to N$ be $C^r$ parameterizations with $p \in \varphi(U)$ and $f(\varphi(U)) \subset \psi(V)$. Without loss of generality assume that $\varphi^{-1}(p) = 0$ and $\psi^{-1}(f(p)) = 0$. Let $\xi$ be a $C^r$ extension of $\varphi^{-1}$ to a neighborhood $W \subset \mathbb{R}^k$ of $p$.

By Lemma 12.1.2 there is a neighborhood $Y \subset \mathbb{R}^k$ and a $C^r$ function $\Phi : Y \to \mathbb{R}^{k-m}$, for which $p$ is a regular point, such that $M \cap Y = \Phi^{-1}(0)$.  


12.3. SETS OF MEASURE ZERO

Replacing $W$ and $Y$ with their intersection if need be, assume that $W = Y$. Define $\Psi : W \to \mathbb{R}^{k-m} \times \mathbb{R}^n$ by

$$\Psi(z) = (\Phi(z), \psi^{-1}(f(\varphi(\xi(z))))).$$

Clearly $\psi^{-1}(0) = f^{-1}(q)$.

The definition of a parameterization implies that $p$, $\varphi^{-1}(p)$ and $f(p)$ are regular points of $\xi$, $\varphi$, and $\psi^{-1}$ respectively, and $p$ is a regular point of $f$ by hypothesis, so the chain rule implies that $\varphi^{-1}(p)$ is a regular point of $\psi^{-1} \circ f \circ \varphi \circ \xi$. In addition, the rank of the restriction of $D(\psi^{-1} \circ f \circ \varphi \circ \xi)$ to the kernel of $D\Phi(p)$ is $n$, so the rank of $D\Psi(p)$ is $k - m + n$. The claim now follows from Lemma 12.1.2.

\[ \blacksquare \]

12.3 Sets of Measure Zero

Our main goal in this chapter, Sard’s theorem, asserts that certain sets “have measure zero.” As most readers know, there is something called ‘Lebesgue measure’ which assigns nonnegative numbers to certain subsets of Euclidean spaces, but we wish to avoid making that theory a prerequisite, so we will develop a self-contained theory of sets of measure zero. Informally, a set $S \subset \mathbb{R}^k$ has measure zero if it is possible to cover $S$ with a countable collection of open balls of arbitrarily small total volume. But we do not wish to go to the trouble of stating and justifying a definition of the volume of a ball, so we adopt definitions that slide around this issue.

Definition 12.3.1 A set $S \subset \mathbb{R}^m$ is said to have measure zero if, for any $\varepsilon > 0$, there is a sequence $\{(x_j, \delta_j)\}^\infty_{j=1}$ in $\mathbb{R}^k \times (0, 1)$ such that

$$S \subset \bigcup_j B_{\delta_j}(x_j) \quad \text{and} \quad \sum_j \delta_j^m < \varepsilon.$$

The unit ball contains any cube centered at the origin with side length $2m^{-1/2}$, and it is contained in any cube centered at the origin with side length 2. A closed rectangle in $\mathbb{R}^m$ is a set of the form

$$\{ s_1v_1 + \cdots + s_mv_m : a_1 \leq s_1 \leq b_1, \ldots, a_m \leq s_m \leq b_m \}$$

where $v_1, \ldots, v_m$ is an orthonormal basis of $\mathbb{R}^m$. We define its volume to be $\prod_j (b_j - a_j)$. It is easy to see (and tedious to prove formally, so we will not do it) that any rectangle can be covered with a collection of cubes with only slightly greater total volume. Thus:
Lemma 12.3.2  For $S \subset \mathbb{R}^m$ the following conditions are equivalent:

(a) $S$ has measure zero;

(b) for any $\varepsilon > 0$, $S$ can be covered by a countable collection of cubes of total volume less than $\varepsilon$;

(c) for any $\varepsilon > 0$, $S$ can be covered by a countable collection of rectangles of total volume less than $\varepsilon$.

Condition (a) has the desirable property of being obviously unaffected by orthogonal changes of coordinates. (Lemma 12.4.2 below shows that it is unaffected by any linear change of coordinates.) Condition (c) is typically easiest to verify. The rectangles in a cover of $S$ can be expanded slightly without changing their volume by very much (e.g., by a factor of two) so if $S$ has measure zero there is a countable collection of rectangles, with arbitrarily small total volume, whose interiors cover $S$. If $S$ is compact, then there is a finite covering of this sort.

The following property of sets of measure zero occurs frequently in proofs.

Lemma 12.3.3  If $S_1, S_2, \ldots \subset \mathbb{R}^m$ are sets of measure zero, then $S_1 \cup S_2 \cup \ldots$ has measure zero.

Proof:  For given $\varepsilon$ take the union of a countable cover of $S_1$ by rectangles of total volume $< \varepsilon/2$, a countable cover of $S_2$ by rectangles of total volume $< \varepsilon/4$, etc.

The next result is the key to most applications of Sard’s theorem. The length of the proof is a measure of the price we pay for working with an ad hoc theory of sets of measure zero. In a “proper” treatment of Lesbesgue measure it would emerge naturally and effortlessly in the course of laying out the foundational results.

Proposition 12.3.4  If $S \subset \mathbb{R}^m$ has measure zero, then the complement of $S$ is dense.

Proof:  If the complement of $S$ is not dense, then $S$ contains a compact rectangle, so it suffices to show that a rectangle (with all side lengths positive) cannot be covered by countably many rectangles of total volume less than the volume of the rectangle. If such a covering existed its rectangles
could be slightly expanded to obtain another such covering, this time by open rectangles, which would necessarily have a finite subcover. Thus we may assume a finite covering by rectangles. These rectangles are cartesian products of intervals. For each dimension there are finitely many numbers appearing as endpoints of the respective intervals, and these numbers subdivide that dimension into a finite collection of bounded intervals and two intervals that are each unbounded in one direction. Each rectangle in the covering is a union of “small” rectangles formed by taking cartesian products of the bounded intervals, and the distributive law of arithmetic implies that its volume is the sum of the volumes of these “small” rectangles. Thus the sum of the volume of the rectangles in the covering is greater than the sum of the volumes of the “small” rectangles in their union, which is greater than the sum of the volumes of the “small” rectangles in the minimal covering of the given rectangle by “small” rectangles, which is not less than the volume of the given rectangle by virtue of the distributive law.

The remainder of this section is devoted to a technical result that will be required in the proof of Sard’s theorem. For $S \subseteq \mathbb{R}^m$ and $t \in \mathbb{R}$ let

$$S_1(t) := \{(x_2, \ldots, x_m) \in \mathbb{R}^{m-1} : (t, x_2, \ldots, x_m) \in S\}$$

be the $t$-slice of $S$. The next result is a special case of Fubini’s theorem from measure theory.

**Lemma 12.3.5** Suppose that $S \subseteq \mathbb{R}^m$ is a compact set such that for each $t \in \mathbb{R}$, $S_1(t)$ has measure zero. Then $S$ has measure zero.

**Proof:** Choose a compact interval $[a_1, b_1]$ such that $S \subseteq [a_1, b_1] \times \mathbb{R}^{m-1}$. Fixing $\varepsilon > 0$, for each $t \in [a_1, b_1]$ choose a finite collection of rectangles $R_1(t), \ldots, R_{s(t)}(t) \subseteq \mathbb{R}^{m-1}$ of total volume less than $\varepsilon$ whose interiors cover $S_1(t)$. Since $S$ is compact, there is an interval $I(t) = [a(t), b(t)]$ containing $t$ in its interior such that for all $t' \in I(t)$, $S_1(t')$ is contained in the union of the interiors of $R_1(t), \ldots, R_{s(t)}(t)$. Since $[a_1, b_1]$ is compact, we can choose $t_1, \ldots, t_r$ such that the interiors of $I(t_1), \ldots, I(t_r)$ cover $[a_1, b_1]$, and by discarding redundant intervals, we may assume that this cover is minimal. If $t$ was an element of three of these intervals, then, by taking the interval with the smallest left endpoint and the interval with the largest right endpoint, we would find that at least one of the three intervals was redundant. Thus each $t$ is contained in either one or two of the intervals, and it follows that

$$\{(I(t_j) \cap [a_1, b_1]) \times R_h(t_j) : 1 \leq j \leq r, 1 \leq h \leq s(t_j)\}$$

is a cover of $S$ by rectangles of total volume less than $2(b_1 - a_1)\varepsilon$. 

We can now give an inductive characterization of sets of measure zero. For $S \subset \mathbb{R}^m$ let $\Pi_1(S) \subset \mathbb{R}$ be the set of $t$ such that $S_1(t)$ is not a set of measure zero. We begin with the compact case.

**Lemma 12.3.6** A compact $S \subset \mathbb{R}^m$ has measure zero if and only if $\Pi_1(S)$ has measure zero.

**Proof:** Let $S$ be contained in the rectangle $[a_1, b_1] \times \cdots \times [a_m, b_m]$.

If $\Pi_1(S)$ has measure zero, then, for any $\varepsilon > 0$, it can be covered by the interiors of finitely many intervals $I_1, \ldots, I_r$ of total length less than $\varepsilon$. The last result implies that

$$S \setminus \left( \bigcup_{j=1}^r (\text{int} I_j) \times [a_2, b_2] \times \cdots \times [a_m, b_m] \right)$$

has measure zero, hence can be covered by finitely many rectangles of total volume less than $\varepsilon$, and the cartesian products of $I_1, \ldots, I_r$ with $[a_2, b_2] \times \cdots \times [a_m, b_m]$ have total volume less than $\varepsilon(b_2 - a_2) \cdots (b_m - a_m)$. Thus $S$ has measure zero.

Now suppose that $S$ has measure zero. For $\delta > 0$ let $\Pi_1^\delta(S)$ is the set of $t$ such that $S_1(t)$ cannot be covered by countable many rectangles with total volume less than $\delta$. Then

$$\Pi_1(S) = \bigcup_{\delta > 0} \Pi_1^\delta(S).$$

Our goal is to show that $\Pi_1(S)$ has measure zero, so, by Lemma 12.3.3, we may assume that there is some $\delta$ such that $\Pi_1^\delta(S)$ is not a set of measure zero, meaning that there is some $\gamma > 0$ such that $\Pi_1^\gamma(S)$ cannot be covered by countably many intervals of total volume less than $\gamma$. Fix $\varepsilon < \min\{\delta^2, \gamma^2\}$. There is a finite cover $R^1, \ldots, R^s$ of $S$ by rectangles of total volume less than $\varepsilon$. Since $\Pi_1^\gamma(R_1 \cup \cdots \cup R_s)$ is the set of $t \in [a_1, b_1]$ such that the sum of the $(m - 1)$-dimensional volumes of the rectangles $R_i^\gamma(t)$ is greater than $\sqrt{\varepsilon}$, clearly $\Pi_1^\gamma(R_1 \cup \cdots \cup R_s)$ must be a finite union of disjoint intervals of total length less than $\sqrt{\varepsilon}$. But $\delta > \sqrt{\varepsilon}$ and $S \subset R_1 \cup \cdots \cup R_s$, so

$$\Pi_1^\delta(S) \subset \Pi_1^\gamma(R_1 \cup \cdots \cup R_s),$$

contrary to the assumption that $\Pi_1^\gamma(S)$ cannot be covered by countably many intervals of total volume less than $\gamma$. 


12.4. MEASURE ZERO SUBSETS OF MANIFOLDS

The inductive characterization of sets of measure zero holds in much greater generality than the statement of the following result suggests. Among other things, it is true for countable unions of intersections of closed and open sets, by Lemma 12.3.3. But our later work does not require such generality.

**Proposition 12.3.7** If \( S \subset \mathbb{R}^m \) is the intersection of a closed set \( C \) and an open set \( U \), then \( S \) has measure zero if and only if \( \Pi_1(S) \) has measure zero.

**Proof:** Let \( A^1, A^2, \ldots \) be a countable collection of compact rectangles that cover \( U \). Then (a) \( \iff \) (b) \( \iff \) (c) \( \iff \) (d) where (a)-(d) are:

(a) \( S \) has measure zero;

(b) each \( C \cap A^j \) has measure zero;

(c) each \( \Pi_1(C \cap A^j) \) has measure zero;

(d) \( \Pi_1(S) \) has measure zero.

The equivalence of (a) and (b) follows from Lemma 12.3.3, and the equivalence of (b) and (c) is the special case established above. Clearly Lemma 12.3.3 also implies that

\[
\Pi_1(S) = \Pi_1(C \cap A^1) \cup \Pi_1(C \cap A^2) \cup \ldots ,
\]

so that (again by Lemma 12.3.3) (c) and (d) are equivalent.

12.4 Measure Zero Subsets of Manifolds

In transporting the notion of a subset of measure zero from Euclidean space to manifolds we follow the familiar pattern of using an arbitrary parameterization to perform the translation.

**Definition 12.4.1** If \( M \subset \mathbb{R}^k \) is an \( m \)-dimensional \( C^1 \) manifold, then \( S \subset M \) has \( m \)-dimensional measure zero if \( \varphi^{-1}(S) \) has measure zero whenever \( U \subset \mathbb{R}^m \) is open and \( \varphi : U \rightarrow M \) is a \( C^1 \) parameterization.
In order for this to be sensible, it should be the case that \( \varphi(S) \) has measure zero whenever \( \varphi : U \to M \) is a \( C^1 \) parameterization and \( S \subset U \) has measure zero. That is, it must be the case that if \( \varphi' : U' \to M \) is another \( C^1 \) parameterization, then \( \varphi'^{-1}(\varphi(S)) \) has measure zero. This follows from the application of the following result to \( \varphi'^{-1} \circ \varphi \).

**Lemma 12.4.2** If \( U \subset \mathbb{R}^m \) is open, \( f : U \to \mathbb{R}^m \) is \( C^1 \), and \( S \subset U \) has measure zero, then \( f(S) \) has measure zero.

**Proof:** Let \( R \subset U \) be a compact rectangle. Since \( U \) can be covered by countably many such rectangles, it suffices to show that \( f(S \cap R) \) has measure zero. Let \( B := \max_{x \in R} \| Df(x) \| \), where

\[
\| Df(x) \| := \max_{\| v \| = 1} \| Df(x)v \|
\]

is the operator norm. For any \( x, y \in R \) we have

\[
\| f(x) - f(y) \| = \left\| \int_0^1 Df((1 - t)x + ty)(y - x) \, dt \right\|
\leq \int_0^1 \| Df((1 - t)x + ty) \| \cdot \| y - x \| \, dt \leq B \| y - x \|.
\]

If \( \{(x_j, \delta_j)\}_{j=1}^\infty \) is a sequence such that

\[
S \cap R \subset \bigcup_j B_{\delta_j}(x_j) \quad \text{and} \quad \sum_j \delta_j^m < \varepsilon,
\]

then

\[
f(S \cap R) \subset \bigcup_j B_{B\delta_j}(x_j) \quad \text{and} \quad \sum_j (B\delta_j)^m < B^m \varepsilon.
\]

Clearly the basic properties of sets of measure zero in Euclidean spaces—countable unions of sets of measure zero have measure zero, and the complement of a set of measure zero is dense—extend, by straightforward verifications, to subsets of manifolds of measure zero. In these and similar cases it is common to cite such results in arguments concerning manifolds, even when the statement of the result refers only to Euclidean spaces.
12.5 Sard’s Theorem

Let \( f : N \to M \) be a \( C^1 \) function. A point \( p \in N \) is a critical point of \( f \) if it is not a regular point, i.e., the rank of \( Df(x) \) is less than the dimension of \( M \). A point \( q \in M \) is a critical value if it is the image of a critical point. The result of this section is:

**Theorem 12.5.1** (Morse-Sard Theorem) If \( M \) and \( N \) are \( m \)- and \( n \)-dimensional \( C^r \) manifolds, \( f : N \to M \) is a \( C^r \) map, and \( r > n - m \), then the set of critical values of \( f \) has measure zero.

*Proof:* Clearly [CITE?] it suffices to prove that if \( U \subset \mathbb{R}^n \) is open, \( f : U \to \mathbb{R}^m \) is \( C^r \), and \( r > n - m \), then the set of critical values of \( f \) has measure 0. If \( m = 0 \), then \( f \) has no critical points, and if \( n = 0 \), then the image of \( f \) is a single point. In either case the claim holds, so, by induction, we may assume that the claim has been established with \((n, m)\) replaced by either \((n - 1, m - 1)\) or \((n - 1, m)\).

Let \( C \) by the set of critical points of \( f \). For each \( i \geq 1 \) let \( C_i \) be the set of points in \( U \) at which all partial derivatives of \( f \) up to order \( i \) vanish. We will show that:

(i) \( f(C \setminus C_i) \) has measure 0;

(ii) \( f(C_i \setminus C_{i+1}) \) has measure zero for all \( i \geq 1 \);

(iii) if \( ms \geq n \), then \( f(C_s) \) has measure zero.

*Proof of (i):* Fix a point \( x_0 \in C \setminus C_1 \). An open ball \( B \subset \mathbb{R}^m \) will be said to be rational (this terminology is specific to this proof, and not standard) if the coordinates of its center and its radius are rational numbers. We will show that \( x_0 \) is contained in a rational ball \( B \) such that \( B \cap C_1 = \emptyset \) and \( f(C \cap B) \) has measure zero. Since there are countably many rational balls, Lemma 12.3.3 will then imply that \( f(C \setminus C_1) \) has measure zero.

Some partial derivative of \( f \) does not vanish at \( x_0 \); by reindexing we may assume that it is \( \partial f_1 / \partial x_1 \). Let \( h : U \to \mathbb{R}^n \) be the function

\[
h(x) := (f_1(x), x_2, \ldots, x_n).
\]

The matrix of partial derivatives of \( h \) at \( x_0 \) is triangular, with nonzero entries on the diagonal, so \( x_0 \) is a regular point of \( h \). The inverse function theorem implies the existence of a neighborhood of \( x_0 \) on which \( f \) is a \( C^r \)
diffeomorphism. Let $B$ be a rational ball that contains $x_0$ and is contained in this neighborhood, and define $g := f \circ h^{-1}|_{f(B)}$. Observe that

$$g(y) = (y_1, g_2(y), \ldots, g_m(y))$$

for all $y \in f(B)$, so that the first row of the matrix of partial derivatives of $g$ at $y$ is $(1, 0, \ldots, 0)$. Thus $x \in B$ is a critical point of $f$ if and only if $y := h(x)$ is a critical point of $g$, and in turn this is true if and only if $y$ is a critical point of the restriction $g_{y_1}$ of $g$ to $f(B) \cap \{y_1\} \times \mathbb{R}^{m-1}$. By the induction hypothesis, for each $y_1 \in \mathbb{R}$, the set of critical values of $g_{y_1}$ is a set of measure zero in $\{y_1\} \times \mathbb{R}^{m-1}$, so the claim follows from Lemma 12.3.5.

Proof of (ii): As above, it is enough to show that an arbitrary $x_0 \in C_i \setminus C_{i+1}$ is contained in a rational ball $B \subset U$ such that $C_{i+1} \cap B = \emptyset$ and $f(C_i \cap B)$ has measure zero. Choose a partial derivative $\frac{\partial^{k+1} f}{\partial x_{s_1} \cdots \partial x_{s_i}}(x)$ that does not vanish at $x_0$. Without loss of generality assume that $s_{i+1} = 1$, and define $h : U \to \mathbb{R}^n$ by

$$h(x) := \left(\frac{\partial f}{\partial x_{s_1}}(x), \ldots, \frac{\partial f}{\partial x_{s_i}}(x), x_2, \ldots, x_n\right).$$

The matrix of partial derivatives of $h$ at $x_0$ is triangular with nonzero diagonal entries, so, as above, there is a rational ball $B \subset U$ of rational center and rational radius that contains $x_0$ on which $h$ is a $C^r$ diffeomorphism. Let $g := f \circ h^{-1}|_{f(B)}$. Repeated applications of the chain rule (or Taylor’s theorem, if you like) show that all partial derivatives of $f$ up to order $i$ vanish at $x \in B$ if and only if all partial derivatives of $g$ up to order $i$ vanish at $y := h(x)$. Moreover, all such points $y$ must lie in the plane $\{0\} \times \mathbb{R}^{n-1}$, so the claim follows from induction on $n$.

Proof of (iii): Let $I \subset U$ be a compact cube. We will show that $f(C_i \cap I)$ has measure zero. Since $U$ can be covered by countably many compact cubes, this suffices to establish the claim.

Since $I$ is compact and the partials of $f$ of order $s$ are continuous, Taylor’s theorem implies that for every $\varepsilon > 0$ there is $\delta > 0$ such that $\|f(x + h) - f(x)\| \leq \varepsilon\|h\|^s$ whenever $x, x + h \in I$ with $x \in C_k$ and $\|h\| < \delta$. Let $L$ be the side length of $I$. For each integer $d > 0$ divide $I$ into $d^n$ subcubes of side length $L/d$. If a subcube contains a point $x \in C_k$, then its image is contained in the ball of radius $\varepsilon(\sqrt{n}L/d)^s$ centered at $f(x)$, and the cube of sidelength $2\varepsilon(\sqrt{n}L)^s$ centered at this point. There are $d^n$ subcubes of $I$, each one of which may or may not contain a point in $C_i$, so, when $\sqrt{n}L/d < \delta$, $f(C_i \cap I)$ is contained in a finite union of cubes of total volume at most $(2(\sqrt{n}L)^s)^m \varepsilon^n d^{n-ms}$. 

This result, and its proof, raises several questions and issues. First of all, the inequality \( r > n - m \) operates in the argument by implying that \( r > (n - 1) - (m - 1) \) and \( r > (n - 1) - m \), as required by the induction, and also by implying that \( rm \geq n \). But it remains to be shown that the inequality \( r > n - m \) is best possible. Problem ??? asks you to settle the matter.

In another direction, it seems intuitive that the image of the set of points \( x \in N \) at which the rank of \( Df(x) \) is less than \( n - 1 \) should be “smaller” than the image of the set of critical. For \( 0 \leq p < n \) let \( R_p \) be the set of points \( x \in N \) such that the rank of \( Df(x) \) is less than or equal to \( p \). One might guess that the dimension of \( f(R_p) \) is less than \( p + 1 \), but of course we need to define what we mean by this.

**Definition 12.5.2** A set \( S \subset \mathbb{R}^k \) has \( p \)-dimensional Hausdorff measure zero if, for any \( \varepsilon > 0 \), there is a sequence \( \{(x_j, \delta_j)\}_{j=1}^\infty \) such that

\[
S \subset \bigcup_j B_{\delta_j}(x_j) \quad \text{and} \quad \sum_j \delta_j^p < \varepsilon.
\]

Note that this definition makes perfect sense even if \( p \) is not an integer! The most general and sophisticated version of Sard’s theorem, due to Federer, states that if \( 1 \leq p \leq m \), then \( f(R_p) \) has \( \alpha \)-dimensional measure zero for all \( \alpha > p + \frac{m-p}{r} \). A beautiful informal introduction to the circle of ideas surrounding these concepts, which is the branch of analysis called geometric measure theory, is given by [Mor88]. The proof itself is in ??? of [Fed69]. This reference also gives a complete set of counterexamples showing this result to be best possible.

**Problems**

**Problem 12.1:** Suppose that for each pair \((n, m)\) of nonnegative integers, \( \rho(n, m) \) is a positive integer, and the following conditions are satisfied: (a); (b) if \( m \) and \( n \) are both positive, then \( \rho(n, m) \leq \rho(n - 1, m - 1) \); (c) if \( n \) is positive, then \( \rho(n, m) \leq \rho(n - 1, m) \); (d) if ???, then \( m\rho(n, m) \geq n \). Prove that \( \rho(n, m) \leq \max\{1, n - m\} \).