

MAT425

Differential Topology

by Brandon Papandrea

The following is based on lecture notes taken during the Winter 2026 offering of MAT425: Differential Topology, at the University of Toronto, with the textbook used being *Differential Topology* by Guillemin & Pollack. The intention is for these notes to be a polished version of my own lecture notes that allows me to revise and look over the material multiple times, and thus should not be considered a primary source for learning about differential topology.

Contents

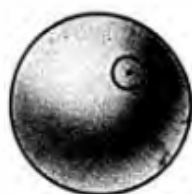
1	Manifolds and Smooth Maps	1
1.1	Definitions	1
1.2	Derivatives and Tangent Spaces	3
1.3	The Inverse and Implicit Function Theorems	7
1.3.1	The Inverse Function Theorem	7
1.3.2	The Implicit Function Theorem	7
1.4	Immersions and Submersions	8
1.4.1	Immersions	8
1.4.2	Submersions	9
1.4.3	The Preimage Theorem	9
1.5	Embeddings	11
1.6	Measure Zero and Sard's Theorem	12
1.7	Transversality	16
1.8	Homotopy and the Stability Theorem	20

Manifolds and Smooth Maps

1.1 DEFINITIONS

Differential Topology is a field of math that specializes in the study of what are called “smooth manifolds.” Before we get anywhere, we first need to define what these are.

The cool thing about calculus is that it is built upon the geometry of Euclidean space. What this means is that if something “looks like” Euclidean space, the same techniques will apply. Such spaces are called **manifolds**, spaces where if you zoom in far enough, you will get something that looks a lot like Euclidean space. If, upon zooming in, we find that the space looks like Euclidean space in k dimensions, then we say that the space is a **k -dimensional manifold**, or just a **k -manifold**. Some notable examples of 2-manifolds include the sphere and torus.



Sphere



Torus

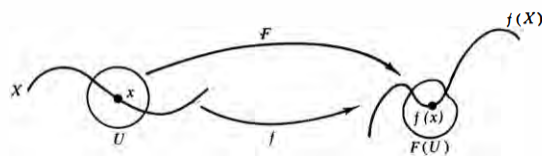
Let’s make this idea more precise. First, we need to define the idea of smoothness.

Definition 1.1. A mapping $f : U \rightarrow \mathbb{R}^m$, where $U \subset \mathbb{R}^n$ is open, is called **smooth** if it has continuous partial derivatives of all orders.

This is great when our domains are open subsets, but this isn’t always true. To rectify this, we need to use a local definition:

Definition 1.2. A mapping $f : X \rightarrow \mathbb{R}^m$, $X \subset \mathbb{R}^n$ is an arbitrary subset, is called **smooth** if, for each $x \in X$, there is an open set $U \subset \mathbb{R}^n$ and a smooth map $F : U \rightarrow \mathbb{R}^m$ such that $F|_{U \cap X} = f$. In other words, f can be locally extended to a smooth map on open sets.

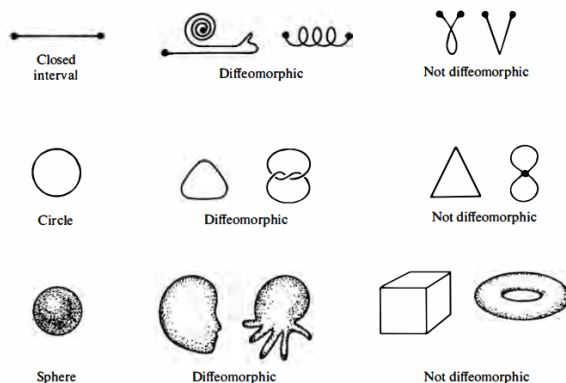
In other words, we define smoothness to be a local property: a map is smooth if it is smooth in a neighbourhood of every point in the domain. This is in contrast to it being a global property, where we consider X as a single, unified object.



With smooth now defined, we move to defining when two spaces looks “the same.”

Definition 1.3. A smooth map $f : X \rightarrow Y$ of subsets of Euclidean space is a **diffeomorphism** if it is a bijection with a smooth inverse $f^{-1} : Y \rightarrow X$. We call X, Y **diffeomorphic** if there is a diffeomorphism between them.

For our purposes, two diffeomorphic spaces sets are basically the same set, just situated differently in the ambient space. Below are some examples of diffeomorphic and non-diffeomorphic spaces.



Finally, we are able to define what a manifold actually is.

Definition 1.4. Let $X \subset \mathbb{R}^n$. We call X a **k -dimensional manifold** if it is locally diffeomorphic to \mathbb{R}^k . That is, for each $x \in X$, there is a neighbourhood V of x that is diffeomorphic to an open set $U \subseteq \mathbb{R}^k$. We call the corresponding diffeomorphism $\phi : U \rightarrow V$ a **parameterization** of V , while the inverse $\phi^{-1} : V \rightarrow U$ is called a **coordinate system** on V .

The map ϕ^{-1} can be written as k smooth functions in a tuple (x_1, \dots, x_k) , called **coordinate functions**. Sometimes these are called **local coordinates** on V , and one writes a point of V as this tuple. There is some subtly here, as these functions identify V with U implicitly. A point $v \in V$ is identified with

$$(x_1(v), \dots, x_k(v))$$

We will also write the dimension of X , the dimension of the space it is locally diffeomorphic to, as just $\dim X$.

Example 1. Consider the circle $S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$. We claim this is a 1-dimensional manifold.

We will parameterize it with 4 maps. First suppose (x, y) is such that $y > 0$, then the map

$$\phi_1(x) = (x, \sqrt{1-x^2})$$

maps $I = [-1, 1]$ bijectively to the upper semicircle. The inverse map, which just projects (x, y) to the x -coordinate, is also smooth. Thus, ϕ_1 is a parameterization. A similar parameterization, ϕ_2 , for the lower semicircle where $y < 0$, is defined as $\phi_2(x) = (x, -\sqrt{1-x^2})$. These maps give local parameterizations for all points of S^1 except $(1, 0)$ and $(-1, 0)$. To get those, we use the maps $\phi_3(y) = (\sqrt{1-y^2}, y)$ and $\phi_4 = (-\sqrt{1-y^2}, y)$ to send I to the left and right semi-circles.

We have thus shown the circle is a 1-dimensional manifold by parameterizing it with 4 maps to between it and a 1-dimensional interval. Note that this can also be done with 2 maps using stereographic projection. In general, S^n is an n -dimensional manifold, which can be proven in a similar way.

Example 2. Given manifolds X, Y in \mathbb{R}^N and \mathbb{R}^M respectively, one can form a new manifold by taking their cartesian product $X \times Y \subset \mathbb{R}^{M+N}$.

Suppose X, Y have dimension k, l respectively. For each $x \in X$ we have an open set $W \subset \mathbb{R}^k$ and a local parameterization $\phi : W \rightarrow X$ at x . Similarly, for each $y \in Y$, we have an open set $U \subset \mathbb{R}^l$ and a local parameterization $\psi : U \rightarrow Y$ at y . Then we can define a map $\phi \times \psi : W \times U \rightarrow X \times Y$ by

$$(\phi \times \psi)(w, u) = (\phi(w), \psi(u))$$

$W \times U$ is open in \mathbb{R}^{k+l} , and more importantly, $\phi \times \psi$ is a local parameterization of $X \times Y$ at (x, y) . Thus, $X \times Y$ is a manifold of dimension $k + l$.

One final thing to define in this section is the idea that manifolds can sit inside larger manifolds.

Definition 1.5. If X, Z are manifolds in \mathbb{R}^N and $Z \subset X$, we call Z a **submanifold** of X .

As a trivial example, any manifold X is a submanifold of its ambient space \mathbb{R}^N . Similarly, every open set of X is a submanifold of X . As a non-trivial example, S^m is a submanifold of S^n whenever $m < n$.

1.2 DERIVATIVES AND TANGENT SPACES

We move to talking about derivatives. Recall from multivariate calculus that for a smooth map $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and point x in the domain, the **directional derivative** in the direction of $h \in \mathbb{R}^n$ is

$$df_x(h) = \lim_{t \rightarrow 0} \frac{f(x + th) - f(x)}{t}$$

If we fix x , we get a linear map $df_x : \mathbb{R}^n \rightarrow \mathbb{R}^m$ mapping vectors h to their corresponding directional derivative. We call this map the **derivative** of f at x . This map is oftentimes represented by a matrix of partial derivatives called the Jacobian matrix of f at x :

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(x) & \cdots & \frac{\partial f_n}{\partial x_n}(x) \end{pmatrix}$$

For our purposes, we will think of the derivative as a linear map.

This is a very natural definition, and we can see this in the chain rule. Let $U \subset \mathbb{R}^n, V \subset \mathbb{R}^m$ be open, with smooth maps $f : U \rightarrow V, g : V \rightarrow \mathbb{R}^l$. Then by the chain rule, for $x \in U$,

$$d(g \circ f)_x = dg_{f(x)} \circ df_x$$

This gives us a commutative diagram of derivative maps:

$$\begin{array}{ccccc} \mathbb{R}^n & \xrightarrow{df_x} & \mathbb{R}^m & \xrightarrow{dg_{f(x)}} & \mathbb{R}^l \\ & \searrow & & \nearrow & \\ & d(g \circ f)_x & & & \end{array}$$

Note that, as df_x gives a linear approximation to f at x , if $f : U \rightarrow \mathbb{R}^m$ is a linear map, L , then $df_x = L$ for all x . In particular, the inclusion map $\iota : U \rightarrow \mathbb{R}^n$ at any x is just the identity map of \mathbb{R}^n .

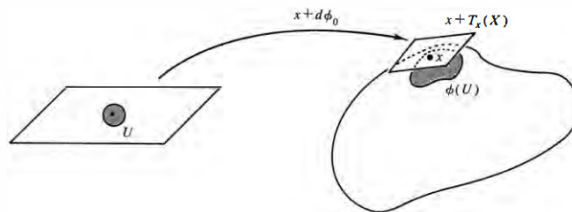
Since the derivative is the best linear approximation of a mapping, we can use them to find the linear space that best approximates a manifold X at a point x . Let X be a manifold in \mathbb{R}^N and $\phi : U \rightarrow X$ a local parameterization at x , with U open in \mathbb{R}^k . Assume that $\phi(0) = x$. Then the best linear approximation to ϕ at 0 is

$$u \mapsto \phi(0) + d\phi_0(u) = x + d\phi_0(u)$$

The image of this map is a space centered at x . If we ignore the translation to x , we can define what this space is.

Definition 1.6. The **tangent space** of X at x is the image of $d\phi_0 : \mathbb{R}^k \rightarrow \mathbb{R}^N$, denoted TX_x . It is a vector subspace of \mathbb{R}^N whose translation to x , $x + TX_x$, is the closest flat approximation to X at x .

A point $v \in TX_x$ is called a **tangent vector to X at x** , though it is more natural to view v as a vector from x to $x + v$.



Because this definition is based on the local parameterization, one should ask if the tangent space depends on the choice of local parameterization. Suppose that $\psi : V \rightarrow X$ is another choice with $\psi(0) = x$. $U \cap V \neq \emptyset$, so without loss of generality we can shrink both sets so that $\psi(U) = \psi(V)$. The map

$$h = \psi^{-1} \circ \phi : U \rightarrow V$$

is a diffeomorphism (composition of diffeomorphisms), and $\phi = \psi \circ h$. Differentiating gives us

$$d\phi_0 = d\psi_0 \circ dh_0$$

Thus, the image of $d\phi_0$ is in $d\psi_0$. Reversing the roles of ϕ, ψ gives the other direction, meaning the images are the same, and TX_x is well-defined.

The dimension of the tangent space is the same as the dimension of the manifold. Suppose $\dim X = k$, and let $W \subset \mathbb{R}^N$ be open. Let $\Phi' : \mathbb{R}^N \rightarrow \mathbb{R}^k$ be smooth and extend ϕ^{-1} to all of \mathbb{R}^N . $\Phi' \circ \phi$ is the identity map on U . Smoothness of ϕ^{-1} , and the chain rule, then imply that

$$\text{id} = d\Phi_x \circ d\phi_0$$

on \mathbb{R}^k . Hence, $d\phi_0 : \mathbb{R}^k \rightarrow TX_x$ is an isomorphism, so the dimensions are the same.

We are now ready to define the derivative of a map on arbitrary manifolds $f : X \rightarrow Y$. Assuming $f(x) = y$, it should be a map of tangent space $df_x : TX_x \rightarrow TY_y$. It must satisfy two things. First, if the map is in Euclidean space, we should just get the usual derivative. Second, the chain rule must be satisfied.

With all the usual terms and definitions, and assuming that $\phi(0) = x, \psi(0) = y$, we have a commutative diagram:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \phi \uparrow & & \uparrow \psi \\ U & \xrightarrow{h=\psi^{-1} \circ f \circ \phi} & V \end{array}$$

We already know what $d\phi_0, d\psi_0, dh_0$ must be, and the chain rule tells us that taking derivatives turns our above diagram into a diagram of linear maps:

$$\begin{array}{ccc} TX_x & \xrightarrow{df_x} & TY_y \\ d\phi_0 \uparrow & & \uparrow d\psi_0 \\ \mathbb{R}^k & \xrightarrow{dh_0} & \mathbb{R}^l \end{array}$$

Recalling that $d\phi_0$ is an isomorphism, we get exactly one acceptable definition of df_x :

Definition 1.7. Let $f : X \rightarrow Y$ be a map of manifolds, with $f(x) = y$. Let ϕ, ψ be local parameterizations of x, y respectively, on open sets $U \subset \mathbb{R}^k, V \subset \mathbb{R}^l$ respectively. Then the **derivative of f at x** is the linear map

$$df_x : TX_x \rightarrow TY_y \quad df_x(v) = (d\psi_0 \circ dh_0 \circ d\phi_0^{-1})(v)$$

Again, we need to check that this definition does not depend on the choice of parameterization. Suppose that ϕ', ψ' parameterize x, y as well, and without loss of generality assume the images of ϕ, ϕ' and ψ, ψ' are equal. We get new diffeomorphisms

$$\Phi = \phi'^{-1} \circ \phi \quad \Psi = \psi'^{-1} \circ \psi$$

with $\Phi(0) = \Psi(0) = 0$. Setting $h' = \psi'^{-1} \circ f \circ \phi'$, we get that

$$h = \Psi^{-1} \circ h' \circ \Phi$$

so at 0, we get that $dh_0 = d\Psi_0^{-1} \circ dh'_0 \circ d\Phi_0$. Thus,

$$\begin{aligned} df_x &= d\psi_0 \circ dh_0 \circ d\phi_0^{-1} \\ &= d\psi_0 \circ d\Psi_0^{-1} \circ dh'_0 \circ d\Phi_0 \circ d\phi_0^{-1} \\ &= d\psi'_0 \circ dh'_0 \circ d\phi_0'^{-1} \end{aligned}$$

as required.

We also need to check that our definition satisfies the chain rule.

Theorem 1.8 (Chain Rule). *If $f : X \rightarrow Y, g : Y \rightarrow Z$ are smooth maps of manifolds, then*

$$d(g \circ f)_x = dg_{f(x)} \circ df_x$$

Proof. Let $\eta : W \rightarrow Z$ at $z = g(y)$, with W open in \mathbb{R}^m and $\eta(0) = z$, be a parameterization of Z at z . We get a commutative diagram

$$\begin{array}{ccccc} X & \xrightarrow{f} & Y & \xrightarrow{g} & Z \\ \uparrow \phi & & \uparrow \psi & & \uparrow \eta \\ U & \xrightarrow{h=\psi^{-1} \circ f \circ \phi} & V & \xrightarrow{j=\eta^{-1} \circ g \circ \psi} & W \end{array}$$

and which reduces to

$$\begin{array}{ccc} X & \xrightarrow{g \circ f} & Z \\ \uparrow \phi & & \uparrow \eta \\ U & \xrightarrow{j \circ h} & W \end{array}$$

Thus, by definition,

$$d(g \circ f)_x = d\eta_0 \circ d(j \circ h)_0 \circ d\phi_0^{-1}$$

The standard chain rule applies for $j \circ h$, so $d(j \circ h)_0 = (dj)_0 \circ (dh)_0$. Hence,

$$d(g \circ f)_x = (d\eta_0 \circ dj_0 \circ d\psi_0^{-1}) \circ (d\psi_0 \circ dh_0 \circ d\phi_0^{-1}) = dg_y \circ df_x$$

as required. □

1.3 THE INVERSE AND IMPLICIT FUNCTION THEOREMS

Before we start talking about the topology of manifolds, we need to study the local behavior of smooth maps.

1.3.1 THE INVERSE FUNCTION THEOREM

Suppose X, Y are manifolds of the same dimension, and $f : X \rightarrow Y$ is smooth. The simplest behavior this map can have at a point x , is that it sends neighbourhoods of x to neighbourhoods of $f(x)$ via a diffeomorphism. We call such maps **local diffeomorphisms**. How can we check if a map is a local diffeomorphism? A necessary condition of such maps is that the derivative df_x is an isomorphism of tangent spaces. Suprisingly, this is also a sufficient condition:

Theorem 1.9 (The Inverse Function Theorem). *Let $f : X \rightarrow Y$ be smooth such that df_x at x is an isomorphism. Then f is a local diffeomorphism at x .*

This is a powerful theorem, the map df_x can be thought of as a matrix of numbers, and this matrix is invertible (hence yielding an isomorphism), precisely when it has non-zero determinant. Thus, checking for a local property of smooth maps amounts to just showing that a matrix has a non-zero determinant!

Proofs of the Inverse Function Theorem when X, Y are open subsets of Euclidean space are abundant. Generalizing to manifolds amounts to using local parameterization.

Proof of Inverse Function Theorem. Take parameterizations $\phi : U \rightarrow X, \psi : V \rightarrow Y$ around x and $f(x)$ and set $h = \psi \circ f \circ \phi^{-1}$. By assumption, $df_x = d\psi_0 \circ dh_0 \circ d\phi_0^{-1}$ is an isomorphism, so dh_0 must be an isomorphism. The Inverse Function Theorem on Euclidean space applies, meaning there is an open neighbourhood $W \subset \phi(U)$ of $\phi^{-1}(x)$ such that $h(W)$ is an open neighbourhood of $\psi^{-1}(f(x))$, and h restricts to a diffeomorphism $W \rightarrow h(W)$. It then follows that $f = \psi^{-1} \circ h \circ \phi$ is a diffeomorphism from $\phi^{-1}(W)$, an open neighbourhood of x , to $\psi^{-1}(h(W))$, an open neighbourhood of $f(x)$. \square

Another way of viewing the inverse function theorem is to say that, when df_x is an isomorphism, one can choose local coordinates x_1, \dots, x_k around x and y such that f looks like the identity near x . In general, we say two maps $f : X \rightarrow Y$ and $f' : X' \rightarrow Y'$ are **equivalent** if there are diffeomorphisms $\phi : X' \rightarrow X, \psi : Y' \rightarrow Y$ such that the resulting diagram commutes. From this angle, we can say that if df_x is an isomorphism, when f is locally equivalent, at x , to the identity.

1.3.2 THE IMPLICIT FUNCTION THEOREM

Unfortunately, our maps will sometimes go between manifolds with different dimensions. For example, suppose $f : X \rightarrow Y$ goes between manifolds of dimension k, l respectively, such that $l < k$. Can we say anything about the local behavior of f ? Yes! In fact, we get another fairly important theorem.

Theorem 1.10 (The Implicit Function Theorem). *Let $f : X \rightarrow Y$ be a map between manifolds of dimension $n + m$ and n respectively, and let $x \in X$. Suppose that the matrix of partial derivatives corresponding to the first n variables at x is invertible. Then there exist open sets*

$U_1 \subset \mathbb{R}^n, U_2 \subset \mathbb{R}^m$ with $(x_1, \dots, x_n) \in U_1$ and a map $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ such that for every $x' \in U_1$, there is a unique point $\phi(x') \in U_2$ such that

$$f(x', \phi(x')) = f(x)$$

In other words, near x , f acts like projection onto the first n coordinates. Like before, the proof essentially boils down to the proof in Euclidean space, which can easily be found. The proof for manifolds then amounts to using local parameterizations.

Something else that should be noted is that the implicit and inverse function theorems are equivalent, meaning you can use one to prove the other.

1.4 IMMERSIONS AND SUBMERSIONS

1.4.1 IMMERSIONS

What is the best local behavior $f : X \rightarrow Y$ can have when $\dim X < \dim Y$? The best we can expect is for df_x to be an injective map.

Definition 1.11. If $f : X \rightarrow Y$ is such that $\dim X < \dim Y$ and df_x is injective, then we say f is an **immersion** at x .

The **canonical immersion** is the inclusion map $(a_1, \dots, a_k) \rightarrow (a_1, \dots, a_k, 0, \dots, 0)$ from $\mathbb{R}^k \rightarrow \mathbb{R}^l$ for $k \leq l$.

The canonical immersion is special. Up to diffeomorphism, all immersions are locally the canonical one.

Theorem 1.12 (Immersion Theorem). Let $f : X \rightarrow Y$ be an immersion at x and $y = f(x)$. Then there are local coordinates around x and y such that

$$f(x_1, \dots, x_k) = (x_1, \dots, x_k, 0, \dots, 0)$$

Proof. Choose local parameterizations $\phi : U \rightarrow X, \psi : V \rightarrow Y$ for x, y , with $\phi(0) = x, \psi(0) = y$. We need to augment the map $h = \psi^{-1} \circ f \circ \phi$ to apply the Inverse Function Theorem.

dg_0 is injective, so we may apply a change of basis to \mathbb{R}^l to make it a matrix of the form

$$\begin{pmatrix} I_k \\ 0 \end{pmatrix}$$

We can now define a map $G : U \times \mathbb{R}^{l-k} \rightarrow \mathbb{R}^l$ by

$$G(x, z) = h(x) + (0, z)$$

so apply h to the first k coords, and leave the rest unchanged. G maps an open set of \mathbb{R}^l to \mathbb{R}^l , and dG_0 is just the identity matrix, so the Inverse Function Theorem applies. Thus, G is a local diffeomorphism of \mathbb{R}^l at 0. We also get that $\psi \circ G$ is a local diffeomorphism at 0, and so we may use it as a local parameterization of Y at y . By construction, $h = G \circ (\text{canonical immersion})$, so if we make U, V sufficiently small, we get the following commutative diagram:

$$\begin{array}{ccc}
 X & \xrightarrow{f} & Y \\
 \uparrow \phi & & \uparrow \psi \circ G \\
 U & \xrightarrow{\text{canonical immersion}} & V
 \end{array}$$

and the claim is proven. \square

Corollary 1. *If f is an immersion at x , it is an immersion in a neighbourhood of x .*

1.4.2 SUBMERSIONS

Now suppose that $\dim X \geq \dim Y$. Now the best we can say about df_x is that it is surjective.

Definition 1.13. *If $f : X \rightarrow Y$ is such that df_x is surjective, we say f is a **submersion** at x . The **canonical submersion** is just projection $(x_1, \dots, x_k) \rightarrow (x_1, \dots, x_l)$ where $k \geq l$.*

Like immersions, every submersion is locally equivalent to the canonical submersion near x .

Theorem 1.14 (Submersion Theorem). *Suppose $f : X \rightarrow Y$ is a submersion at x and $y = f(x)$. Then there are local coordinates around x and y such that $f(x_1, \dots, x_k) = (x_1, \dots, x_l)$.*

Proof. The proof is similar to that of the Immersion Theorem. We now define $G : U \rightarrow \mathbb{R}^k$ by

$$G(a) = (h(a), a_{l+1}, \dots, a_k)$$

Inverse Function Theorem then applies and we get a local diffeomorphism G^{-1} of a neighbourhood U' of 0 to U . $\phi \circ G^{-1}$ is also a local diffeomorphism, and $h = (\text{canonical isomorphism}) \circ G$, so $h \circ G^{-1}$ is the canonical submersion. We get a similar commutative diagram and the claim follows. \square

Corollary 2. *If f is a submersion at x , it is a submersion in a neighbourhood of x .*

1.4.3 THE PREIMAGE THEOREM

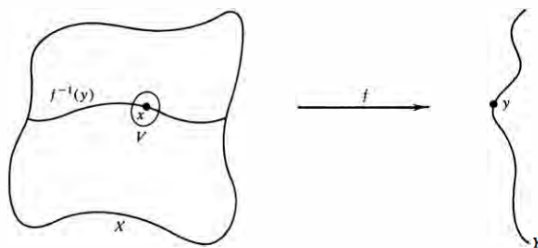
The most valuable insights of these classification theorems is the geometric nature of solutions. Let $f : X \rightarrow Y$ and $y \in Y$. Then we call the set $f^{-1}(y)$ the **preimage** of y . This isn't always a nice object, but suppose f is a submersion at $x \in f^{-1}(y)$. By the Submersion Theorem we can choose local coordinates around x and y such that

$$f(x_1, \dots, x_k) = (x_1, \dots, x_l)$$

and y corresponds to $(0, \dots, 0)$. Thus, near x , $f^{-1}(y)$ is just the set of points

$$\{(0, \dots, 0, x_{l+1}, \dots, x_k)\}$$

In more formal terms, let V be the neighbourhood of x on which these coordinates are defined. Then $f^{-1}(y) \cap V$ is the set of points for which the first l coordinates of x are 0. The functions x_{l+1}, \dots, x_k form a coordinate system on $f^{-1}(y) \cap V$, a relatively open subset of $f^{-1}(y)$. This proves an important theorem:



Theorem 1.15 (Preimage Theorem). *If $f : X \rightarrow Y$ is a smooth map and $y \in Y$ is such that df_x is surjective at all $x \in f^{-1}(y)$, then $f^{-1}(y)$ is a submanifold of X with dimension $\dim X - \dim Y$.*

We are thus led to a new definition:

Definition 1.16. *For a map $f : X \rightarrow Y$, a point $y \in Y$ is a **regular value** for f if df_x is surjective at all $x \in f^{-1}(y)$. It is called a **critical value** if it is not regular. The point x for which df_x is surjective is called a **regular point**, and called a **critical point** if it is not regular.*

The Preimage Theorem is very powerful, in that it allows us to create many submanifolds without needing to create local parameterizations.

Example 3 (The Orthogonal Group). *We let $M(n) \cong \mathbb{R}^{n^2}$ be the space of $n \times n$ matrices with real-valued entries; this is a manifold of dimension n^2 . The subspace $O(n)$, called the **orthogonal group**, is the group of matrices satisfying the equation*

$$AA^t = I$$

In other words, it is the group of linear transformation on \mathbb{R}^n that preserve distance. We claim that $O(n)$ is a manifold.

For any matrix A , the matrix AA^t is symmetric, since

$$(AA^t)^t = (A^t)^t A^t = AA^t$$

The vector space of such $n \times n$ matrices, $S(n)$, is a submanifold of $M(n)$ of dimension $n(n+1)/2$ (we only need to define elements on or above the diagonal). Moreover, the map

$$f : M(n) \rightarrow S(n) \quad f(A) = AA^t$$

is smooth. Now, since $AA^t = I$ for all $A \in O(n)$, we know that

$$O(n) = f^{-1}(I)$$

hence by the Preimage Theorem, we only need to show that I is a regular value of f . The derivative of f at A is

$$\begin{aligned}
 df_A(B) &= \lim_{s \rightarrow 0} \frac{f(A + sB) - f(A)}{s} \\
 &= \lim_{s \rightarrow 0} \frac{(A + sB)(A + sB)^t - AA^t}{s} \\
 &= \lim_{s \rightarrow 0} \frac{AA^t + sBA^t + sAB^t + s^2BB^t - AA^t}{s} \\
 &= \lim_{s \rightarrow 0} BA^t + AB^t + sBB^t \\
 &= BA^t + AB^t
 \end{aligned}$$

Is this map $df_A : T_A M(n) \rightarrow T_{f(A)} S(n)$ surjective when $A \in f^{-1}(I)$? We know that

$$T_A M(n) = M(n), \quad T_{f(A)} S(n) = S(n)$$

so it suffices to show that $df_A : M(n) \rightarrow S(n)$ is surjective when $A \in O(n)$: for all $C \in S(n)$, there's a $B \in M(n)$ such that

$$BA^t + AB^t = C$$

C is symmetric, so we may write

$$C = \frac{1}{2}C + \frac{1}{2}C^t$$

and solve $BA^t = \frac{1}{2}C$, since the transpose will correspond to the other term. We get

$$BA^t = \frac{1}{2}C \implies BA^t A = \frac{1}{2}CA \implies B = \frac{1}{2}CA$$

Indeed,

$$df_A(B) = (1/2)CAA^t + A(1/2CA)^t = 1/2C + 1/2C^t = C$$

as desired. $O(n)$ is a submanifold of $M(n)$, of dimension

$$\dim M(n) - \dim S(n) = n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}$$

1.5 EMBEDDINGS

Given an arbitrary immersion $f : X \rightarrow Y$, there is no guarantee that the image of f will be a submanifold of Y . We know from the Immersion Theorem that f maps neighbourhoods of any x diffeomorphically to $f(W)$ in Y . The problem is that these subsets, $f(W)$, need not be open, so while points in the image have parameterizable subsets, they need not be neighbourhoods.

Take, for example, this map twisting a circle into a figure eight. This is an immersion map $f : S^1 \rightarrow \mathbb{R}^2$, but the figure eight is not a manifold. The problem is the crossing point, which cannot be parameterized.

Does this mean that the immersion must be injective for the image to be a manifold? Nope! We can turn this figure eight map into an injective immersion $f : \mathbb{R}^1 \rightarrow \mathbb{R}^2$, and we will still get an issue at the crossing point.

A more interesting example is a loop around the torus. We define a map

$$g : \mathbb{R}^1 \rightarrow S^1 \quad g(t) = (\cos 2\pi t, \sin 2\pi t)$$

that loops the real line around the circle. Then define the map $\mathbb{R}^2 \rightarrow S^1 \times S^1$ by

$$G(x, y) = (g(x), g(y))$$

which is a local diffeomorphism of the plane to the torus. We may also think of the torus as the product of gluing opposite sides of a square together. We define a new map h by restricting G to a line through the origin with irrational slope, which results in an injective immersion of \mathbb{R}^1 onto the torus. The problem now is that the image of h is dense in the torus, so we can't create any neighbourhoods!

These counterexamples do not mean all hope is lost. We just need some additional assumptions. Notice that the above examples map many points near infinity to a small parts of the image. We express points near infinity as the exterior of a compact set. What we want, then, is for points near infinity near X should be mapped near infinity in Y .

Definition 1.17. A map $f : X \rightarrow Y$ is **proper** if the preimage of every compact set in Y is compact in X .

An injective, proper immersion is called an **embedding**.

This is enough to draw global conclusions from our local immersion theorem:

Theorem 1.18. An embedding $f : X \rightarrow Y$ maps X diffeomorphically onto a submanifold of Y .

Proof. We prove that the image of an open set $W \subseteq X$ is open in $f(X)$. Suppose $f(W)$ is not open. Then it contains a limit point y , meaning there is a sequence of points $y_i \notin f(W)$ that converge to y . The set $\{y, y_i\}$ is compact, so as f is proper, its preimage is compact in X . Each point in this set has one preimage point, x_i for y_i and x for y , so $\{x, x_i\}$ is compact. This set has a convergent subsequence, so we can say that x_i converges to some $z \in X$. Thus,

$$f(x_i) \rightarrow f(z)$$

and as $f(x_i) \rightarrow f(x)$, we conclude by injectivity of f that $z = x$. As W is open, and $x_i \rightarrow x$, we know that for a large enough i , $x_i \in W$, so $f(x_i) = y_i \in f(W)$, a contradiction. Thus, $f(X)$ is a manifold, and we know $f : X \rightarrow f(X)$ is a local diffeomorphism. It is a bijection, so the inverse f^{-1} is well-defined, and since we know it to be locally smooth, we're done. \square

1.6 MEASURE ZERO AND SARD'S THEOREM

If $a = (a_1, \dots, a_n), b = (b_1, \dots, b_n)$ with $a_i < b_i$ for all i , we call $S(a, b)$ the **rectangular solid** consisting of points (x_1, \dots, x_n) with $a_i < x_i < b_i$ for all i . The **volume** of $S = S(a, b)$, denoted $\text{vol}(S)$, is

$$\prod_{i=1}^n (b_i - a_i)$$

Definition 1.19. $A \subset \mathbb{R}^n$ has **measure zero** if for all $\varepsilon > 0$, there is a countable covering of A by solids S_1, S_2, \dots , such that $\sum \text{vol}(S_i) < \varepsilon$.

Subsets of sets of measure zero, and any countable union of sets of measure zero is also of measure zero. From this, many sets of measure zero may be constructed. For instance, consider $\mathbb{R}^k \subset \mathbb{R}^n$ with $k < n$. It suffices to show that for any m , the set

$$[-m, m]^k \subset \mathbb{R}^n$$

has measure zero. We form cubes of the form

$$(-m - \delta, m + \delta) \times (-m - \delta, m + \delta) \times (-\delta, \delta)$$

where $\delta < m$. Then the volume of this cube is given by

$$(2(m + \delta))^2 2\delta = 8(m + \delta)^2 \delta < 8(2m)^2 \delta = 32m^2 \delta$$

setting $\delta = \frac{\varepsilon}{32m^2}$ completes the claim.

The concept of measure zero is extended to manifolds by local parameterization. An arbitrary set $S \subset Y$ is **measure zero** if for all parameterizations ψ of Y , $\psi^{-1}(S)$ has measure zero.

Why do we care about this? The preimage of a regular value of a smooth map $f : X \rightarrow Y$ is a submanifold. The notion of regular is a strong one, and it may be the case that such values occur so irregularly that the Preimage Theorem is of little use. In fact, the opposite is true, thanks to a wonderful theorem of calculus:

Theorem 1.20 (Sard's Theorem). *If $f : X \rightarrow Y$ is a map of smooth manifolds, then the set of critical values of a smooth map has measure zero. In other words, almost every point in Y is a regular value of f .*

Proving Sard will take us some time. We first need to establish that smooth maps preserve measure zero sets:

Proposition 1.21. *Let $U \subset \mathbb{R}^n$ be open and $f : U \rightarrow \mathbb{R}^n$ be smooth. Then if $A \subset U$ has measure zero, so too does $f(A)$*

Proof. Without loss of generality, we assume \bar{A} is compact and contained in U ; this is because A can always be written as some countable union of such subsets. Let W be any open neighbourhood of A where \bar{W} is compact and contained in U .

As \bar{W} is compact, there is a M such that for all $x, y \in \bar{W}$,

$$|f(x) - f(y)| < M|x - y|$$

It then follows that there is a constant M' such that for any cube S in W , $f(S)$ is contained in a cube S' whose volume is at most $M' \text{vol}(S)$.

Now let $\varepsilon > 0$ be given. We cover A with a sequence of cubes S_1, S_2, \dots , each in W with

$$\sum_i \text{vol}(S_i) < \varepsilon$$

Then by mapping each cube by f , we get a covering S'_1, S'_2, \dots , of $f(A)$ by cubes such that

$$\sum_i \text{vol}(S'_i) < M'\varepsilon$$

and taking $\varepsilon \rightarrow 0$ completes the proof. \square

Recalling the first claim about \mathbb{R}^k being measure zero in \mathbb{R}^n , any translation of the cube $[-m, m]^k$ is also measure 0 for all m . Thus, if $U \subset \mathbb{R}^k \subset \mathbb{R}^n$, then as U is a subset of a sufficiently translated and scaled cube, we get that U is measure 0 in \mathbb{R}^n . This, combined with the above proposition, we get a weak form of Sard:

Theorem 1.22 (Mini-Sard's Theorem). *Let U be an open subset of \mathbb{R}^n , and let $f : U \rightarrow \mathbb{R}^m$ be a smooth map. Then if $m > n$, $f(U)$ has measure zero in \mathbb{R}^m .*

We are almost ready to prove Sard, but we need one additional result. First two lemmas:

Lemma 1.23. *Let S_1, \dots, S_n be a covering of $[a, b] \in \mathbb{R}$. Then there is a cover S'_1, \dots, S'_M such that each S'_j is in a S_i , and*

$$\sum_{j=1}^M \text{length}(S'_j) < 2(b-a)$$

Proof. Remove as many overlaps that were occurring in the S_i covering as possible while still maintaining the conditions on s'_j . \square

Lemma 1.24. *Let $I \subset \mathbb{R}$, and define $V_I = I \times \mathbb{R}^{n-1} \subset \mathbb{R}^n$. Let $A \subset \mathbb{R}^n$ be compact, and suppose that $A \cap V_c$ is contained in an open subset U of V_c . Then for any small interval I around $c \in \mathbb{R}$, $A \cap V_I \subset I \times U$.*

Proof. Suppose not. Then there would exist a sequence of points $(x_j, c_j) \in A$ such that $c_j \rightarrow c$, and $x_j \notin U$. But A is a compact, so we may replace x_j with a convergent subsequence, giving us a convergent sequence. We thus get a contradiction. \square

Now we can state and prove our last necessary result:

Theorem 1.25 (Fubini's Theorem for sets of measure zero). *Let A be closed in \mathbb{R}^n such that $A \cap V_c$ has measure zero in V_c for all $c \in \mathbb{R}^k$. Then A has measure zero in \mathbb{R}^n .*

Proof. Without loss of generality, let A be compact. By induction on k , it suffices to show the claim holds for $k = 1$ and $l = n - 1$. Let $I = [a, b]$ such that $A \subset V_I$. For each $c \in I$, choose a covering of $A \cap V_c$ by $n - 1$ dimensional rectangular solids $S_1(c), \dots, S_{N_c}(c)$ with total volume less than some ε . Now choose an interval $J(c) \in \mathbb{R}$ so that $J(c) \times S_i(c)$ covers $A \cap J$ (this is possible by lemma 2). The $J(c)$'s cover $[a, b]$ so by lemma 1 we can replace them with subintervals J'_j with total length less than $2(b-a)$. Each J'_j is in some interval $J(c_j)$, so the solids $J'_j \times S_i(c_j)$ cover A and has total volume less than $2\varepsilon(b-a)$. Take $\varepsilon \rightarrow 0$. \square

Proving Sard's Theorem amounts to another result. We can find a countable collection of open sets (U_i, V_i) , with U_i open in X and V_i in Y , such that the U_i 's cover X , $f(U_i) \subset V_i$, and the U_i 's and V_i 's are diffeomorphic to open sets in \mathbb{R}^n . It thus suffices to prove this:

Theorem 1.26. *Suppose $U \subset \mathbb{R}^n$ is open and $f : U \rightarrow \mathbb{R}^p$ is smooth. Let C be the set of critical points. Then $f(C)$ has measure zero in \mathbb{R}^p .*

This claim clearly holds true for $n = 0$, so assume it holds for $n - 1$. We partition C into nested subsets

$$C \supset C_1 \supset C_2 \supset \cdots$$

where C_i is the set of all $x \in U$ such all partial derivatives of f of order $\leq i$ vanish at x . We need to prove 3 lemmas which will combine to prove the claim:

Lemma 1.27. *$f(C - C_1)$ has measure zero.*

Proof. Around $x \in C - C_1$, we will find an open set V such that $f(V \cap C)$ has measure zero.

Since $x \notin C_1$, there is a partial derivative, say $\frac{\partial f}{\partial x_1}$, that does not vanish at x . Consider the map $h : U \rightarrow \mathbb{R}^n$ given by

$$h(x) = (f_1(x), x_2, \dots, x_n)$$

dh_x is nonsingular, so h maps a neighbourhood V of x diffeomorphically onto some open set V' . The composition $g = f \circ h^{-1}$ will then map V' into \mathbb{R}^p with the same critical values as f , restricted to V .

g is such that it preserves the first coordinate of points in V' . Thus, for each t , we get a map g^t from $(t \times \mathbb{R}^{n-1}) \cap V'$ to $r \times \mathbb{R}^{p-1}$. The derivative of g has a one in the top left, zeroes along the rest of the first row, and the derivative $\frac{\partial g'_i}{\partial x_j}$ in all but the first column, so a point of $t \times \mathbb{R}^{n-1}$ is critical for g^t iff it is critical for g , since the determinant is just $\det \left(\frac{\partial g'_i}{\partial x_j} \right)$. By the induction hypothesis, we apply Sard's theorem and get that the set of critical values of g^t has measure zero. By Fubini's Theorem, we thus conclude that the set of critical values of g has measure zero. \square

Lemma 1.28. *$f(C_k - C_{k-1})$ has measure zero for all $k \geq 1$.*

Proof. For each $x \in C_k - C_{k+1}$, there is a $k + 1$ derivative of f that is not zero. We have a k th derivative of f , say ρ , that vanishes on C_k but has a first derivative $\frac{\partial \rho}{\partial x_1}$ that doesn't vanish. The map $h : U \rightarrow \mathbb{R}^n$

$$h(x) = (\rho(x), x_2, \dots, x_n)$$

maps a neighbourhood V of x diffeomorphically onto an open set V' . h sends $C_k \cap V$ to the hyperplane $0 \times \mathbb{R}^{n-1}$, so the map $g = f \circ h^{-1}$ has all type C^k critical points in this hyperplane. Let $\bar{g} : (0 \times \mathbb{R}^{n-1}) \cap V' \rightarrow \mathbb{R}^p$ be the restriction of g . By induction, the set of critical values of \bar{g} is measure zero. The critical points of g of type C_k are critical points of \bar{g} , so the image of

these critical points is measure zero, thus, $f(C_k \cap V)$ has measure zero. $C_k - C_{k-1}$ is covered by countably many such V , so we're done. \square

Lemma 1.29. *For $k > \frac{n}{p} - 1$, $f(C_k)$ is of measure zero.*

Let $S \subset U$ be a cube whose sides are of length δ . If k is sufficiently large (greater than $n/p - 1$), we will show that $f(C_k \cap S)$ has measure zero.

By Taylor's Theorem, the compactness of S , and definition of C_k ,

$$f(x+h) = f(x) + R(x, h), \quad |R(x, h)| < a|h|^{k+1}$$

for $x \in C_k \cap S, x+h \in S$, and a a constant depending only on f and S . Subdivide S into r^n cubes whose sides are of length δ/r . Let S_1 be a cube of that subdivision containing $x \in C_k$. Then any point of S_1 may be written as $x+h$, with

$$|h| < \sqrt[n]{n} \left(\frac{\delta}{r} \right)$$

We get that $f(S_1)$ is in a cube of side length b/r^{k+1} centered about $f(x)$, where $b = 2a(\sqrt[n]{n}\delta)^{k+1}$. Hence $f(C_k \cap S)$ is in the union of at most r^n cubes with total volume

$$v \leq r^n \left(\frac{b}{r^{k+1}} \right)^p + b^p r^{n-(k+1)p}$$

As $k+1 > n/p$, v tends to 0 as $r \rightarrow \infty$, so $f(C_k \cap S)$ has measure zero. The claim follows because we can cover C_k by countably many such cubes.

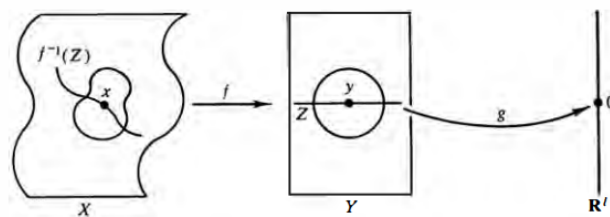
1.7 TRANSVERSALITY

Assuming y is a regular point, we know that the solutions to $f(x) = y$ for a smooth $f : X \rightarrow Y$ is a submanifold of X . What if instead we considered the set of solutions not to a constant equation, but an equation satisfying some smooth condition. In this case, let $Z \supset Y$ be a submanifold and consider the set

$$\{x \in X : f(x) \in Z\} = f^{-1}(Z)$$

when is it that $f^{-1}(Z)$ is a submanifold. Answering this question leads us to a generalization of the concept of regular point, and will be an important theme in the rest of the text.

$f^{-1}(Z)$ is a manifold iff for all $x \in f^{-1}(Z)$, there is a neighbourhood U in X such that $U \cap f^{-1}(Z)$ is a manifold. Because of this, if $y = f(x)$, we can write Z in a neighbourhood of y as the zero set of linearly independent functions g_1, \dots, g_l , where $l = \text{codim}(Z)$. Then near x , $f^{-1}(Z)$ is the zero set of the maps $g_i \circ f$. This means we can reduce our study to the case where Z is a single point. Now, let $g = (g_1, \dots, g_l)$ be the submersion defined at y . Now if W is this neighbourhood of x , we have a map $g \circ f : W \rightarrow \mathbb{R}^l$. Applying what we already know, we get that $(g \circ f)^{-1}(0)$ is a manifold precisely when 0 is a regular value of $g \circ f$.



When is 0 a regular value in this setting? Recall that

$$d(g \circ f)_x = dg_y \circ df_x$$

thus the map $d(g \circ f)_x : T_x(X) \rightarrow \mathbb{R}^l$ is surjective iff dg_y maps the image of df_x onto \mathbb{R}^l . However, $dg_y : T_y(Y) \rightarrow \mathbb{R}^l$ is a surjective linear transformation with kernel $T_y(Z)$. So for dg_y to carry a subspace of $T_y(Y)$ onto \mathbb{R}^l , it must be the case that whatever subspace that is, combined with $T_y(Z)$, span all of $T_y(Y)$. In other words, $g \circ f$ is a submersion at $x \in f^{-1}(Z)$ iff

$$\text{Im}(df_x) + T_y(Z) = T_y(Y)$$

Definition 1.30. The map f is **transversal** to Z if the above equation holds for every $x \in f^{-1}(Z)$. We denote this as $f \pitchfork Z$.

Our above argument also proves our condition for $f^{-1}(Z)$ to be a manifold:

Theorem 1.31. If $f : X \rightarrow Y$ is transversal to a submanifold $Z \subset Y$, then $f^{-1}(Z)$ is a submanifold of X . Moreover, the codimension of $f^{-1}(Z)$ in X is the codimension of Z in Y .

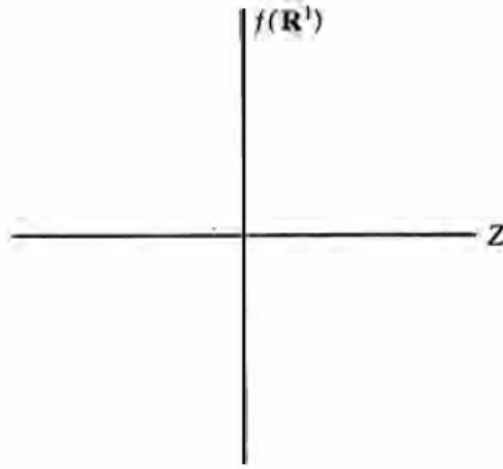
Proof. To prove the codimension claim, suppose that the codimension of Z is l . Then in the above proof we have written the preimage of Z as the zero set of l independent functions. This means the codimension of $f^{-1}(Z)$ in X is l . \square

When $Z = \{y\}$, the tangent space is the zero subspace of $T_y(Y)$, meaning f is transversal to $\{y\}$ when $df_x[T_x(X)] = T_y(Y)$; in other words, when y is a regular value of f .

Example 4. Consider the map

$$f : \mathbb{R} \rightarrow \mathbb{R}^2 \quad f(t) = (0, t)$$

and let Z be the x -axis in \mathbb{R}^2 . f is transversal to Z .



The map

$$g : \mathbb{R} \rightarrow \mathbb{R}^2 \quad g(t) = (t, t^2)$$

fails to be transverse, as its tangent vectors are parallel to the x -axis, meaning the sum of them and the tangent vectors of the x -axis can never span all of \mathbb{R}^2 .



Example 5. Let ι be the inclusion map from $X \subset Y$ to $Z \subset Y$. If $x \in \iota^{-1}(Z)$, it means that $x \in X \cap Z$. Moreover,

$$d\iota_x : T_x(X) \rightarrow T_x(Y)$$

is just inclusion on the tangent spaces. Thus, $i \pitchfork Z$ iff for all $x \in X \cap Z$,

$$T_x(X) + T_x(Z) = T_x(Y)$$

Note that this equation is symmetric in X and Z , leading to a new definition:

Definition 1.32. Two submanifolds $X, Z \subset Y$ are **transversal** submanifolds of Y when the above equation holds, and we write $X \pitchfork Z$.

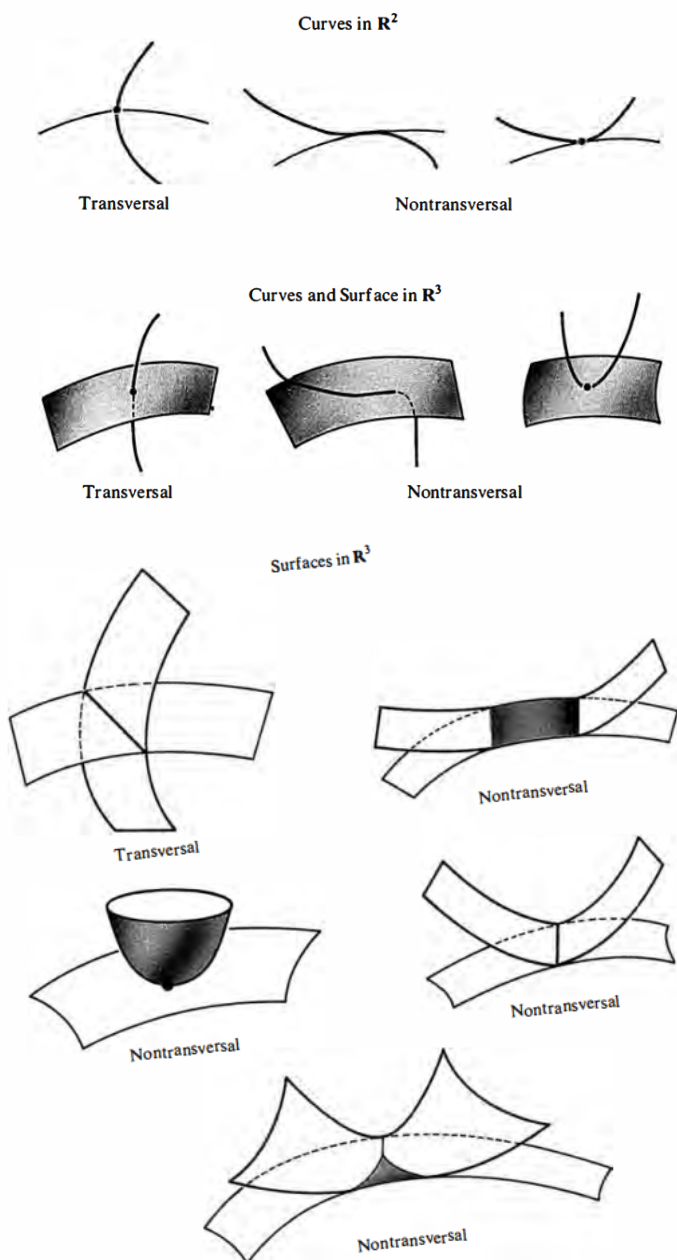
From the above example, we get the following:

Theorem 1.33. The intersection of transversal submanifolds of Y is a submanifold, and

$$\text{codim}(X \cap Z) = \text{codim}(X) + \text{codim}(Z)$$

Proof. Assume X is cut out of Y by k independent functions, while Z is cut out of l . Then their intersection is just the vanishing set of the combined collection of $k + l$ functions, which are all independent from each other. \square

Note that transversality also depends on the ambient space Y . Two coordinate axes may be transversal in \mathbb{R}^2 , but not in \mathbb{R}^3 . A general tip is that if the dimensions of X and Z don't add to that of Y , then they can only be transversal to each other if they don't intersect at all.



1.8 HOMOTOPY AND THE STABILITY THEOREM

We now investigate how certain properties of a smooth map are preserved if that map is deformed in a smooth manner.

Definition 1.34. Two smooth maps $f_0, f_1 : X \rightarrow Y$ are **homotopic**, denoted $f_0 \sim f_1$, if there is a smooth map $F : X \times I \rightarrow Y$ such that $F(x, 0) = f_0(x)$, $F(x, 1) = f_1(x)$. We call F the **homotopy** from f_0 to f_1 .

More generally, we have homotopy maps $f_t = F(x, t)$ for all $t \in [0, 1]$.

Definition 1.35. A property of a smooth map is **stable** if when $f_0 : X \rightarrow Y$ has this property, and $f_t : X \rightarrow Y$ is a homotopy of f_0 , then for some $\varepsilon > 0$, f_t has this property for all $t < \varepsilon$.

Example 6. The property that a curve passes through the origin is not stable, as a simple perturbation of the curve can make it avoid the origin.

On the other hand, transversality is a stable property. Say we have two transverse curves in \mathbb{R}^2 . Then any perturbation of either curve cannot change the tangent vectors at the intersection point in a major way, thus preserving transversality.

Every differential property of maps discussed so far is stable, provided that X is compact:

Theorem 1.36 (The Stability Theorem). *The following classes of smooth maps from a compact manifold X to a manifold Y are stable classes:*

- (i) local diffeomorphism
- (ii) immersion
- (iii) submersion
- (iv) transversal to a submanifold $Z \subset Y$
- (v) embedding
- (vi) diffeomorphism

Proof. (i) is implied by (ii), since a local diffeomorphism is an immersion with the dimensions of the domain and codomain equal. Let f_0 be an immersion and f_t its homotopy. As X is compact, any open neighbourhood of $X \times \{0\}$ in $X \times I$ contains $X \times [0, \varepsilon]$ for sufficiently small ε . It thus suffices to prove that $(x_0, 0)$ has a neighbourhood U in $X \times I$ such that $d(f_t)_x$ is injective for all $(x, t) \in U$. Without loss of generality let X, Y be open subsets of \mathbb{R}^k and \mathbb{R}^l respectively. Injectivity of $d(f_0)_{x_0}$ implies that

$$\left(\frac{\partial(f_0)_i}{\partial x_j}(x_0) \right)$$

has a $k \times k$ submatrix with nonzero determinant. Each partial derivative for f_t is a continuous function with respect to $(x, t) \in X \times I$. The determinant is also continuous, so our submatrix must also have nonzero determinant for all (x, t) in a neighbourhood of $(x_0, 0)$. This proves (ii) and (i) thus follows. The proof of (iii) is similar. (iii) implies (iv) as transversality can be turned into a submersion condition.

To prove (v) it suffices to show that if f_0 is one-to-one, so too is f_t for sufficiently small t . Define a map $G : X \times I \rightarrow Y \times I$ by

$$G(x, t) = (f_t(x), t)$$

and assume towards a contradiction that (v) is false. Then there is a sequence $t_i \rightarrow 0$ and distinct points such that $G(x_i, t_i) = G(y_i, t_i)$. As X is compact we can pass to a convergent subsequence and get $x_i \rightarrow x_0, y_i \rightarrow y_0$. Then

$$G(x_0, 0) = \lim_{i \rightarrow \infty} G(x_i, t_i) = \lim_{i \rightarrow \infty} G(y_i, t_i) = G(y_0, 0)$$

But $G(x_0, 0) = f_0(x_0), G(y_0, 0) = f_0(y_0)$. Thus, $x_0 = y_0$.

The matrix of $dG_{(x_0, 0)}$ is just a matrix with $d(f_0)_{x_0}$ in the top left, a one on the bottom right entry, and 0's along the rest of the bottom row. As $d(f_0)_{x_0}$ is injective, the matrix has k independent rows, and thus $dG_{(x_0, 0)}$ has $k + 1$ independent rows, meaning $dG_{(x_0, 0)}$ is an injective linear map. Thus, G is an immersion around $(x_0, 0)$ and must be injective on some neighbourhood of $(x_0, 0)$. But, for large i , both (x_i, t_i) and (y_i, t_i) are in this neighbourhood, a contradiction.

The proof of (vi) is an exercise. □