A Short Course on deRham Cohomology

Richard Koch

October 7, 2021

Contents

Chapter 1

Introduction

The aim of this course is to prove the Lefshetz Fixed Point Theorem. This result is used in a second set of notes on Compact Lie Groups to prove that any two maximal tori are conjugate.

The notes are based on the classic book *Differential Forms in Algebraic Topology* by Raoul Bott and Loring W. Tu, but cover only the first portion of that book. Anyone looking at these notes should obtain that book and read the clear and beautiful treatment there and the remaining two thirds of the book not covered here at all.

The notes exist because I wanted to tie down the Lefshetz Formula, which is only treated in an exercise of Bott and Tu. Their book omits a few other details, like the proof that a good cover always exists, and the treatment of Thom Transversality, but has a wealth of additional information I didn't need in a direct march to the Lefshetz Formula.

We first outline the development. Suppose we have a C^∞ manifold M. We define differential forms on M and the $\mathbf{d}$ operator, and prove Poincare's lemma. As a consequence, we get an exact sequence of sheaves

$$0 \to R \to \Lambda^0 \to \Lambda^1 \to \ldots \to \Lambda^n \to 0$$

The corresponding sequence of global sections is not necessarily exact. It's cohomology groups are the *de Rham cohomology groups of M:*

$$H^k(M) = \frac{\{k \text{ forms } \omega \mid d\omega = 0\}}{\{\omega = d\lambda \mid \lambda \text{ is a } k-1 \text{ form}\}}$$

If $f : M \to N$ is a C^∞ map, f induces a pullback map from forms on N to forms on M: $\omega \to f^\star(\omega)$ and this induces a map $f^\star : H^k(M) \leftarrow H^k(N)$. We will prove that when f

and g are C^∞ maps homotopic by a C^∞ homotopy, then $f^\star = g^\star$ on deRham cohomology groups.

The vector space $H^0(M)$ contains all locally constant real valued function. Such functions are constant on connected components of M. So $H^0(M)$ has dimension the number of connected components of M.

If ω is an n-form with compact support on an open set in R^n, we can compute

$$\int \omega = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} w_{12\dots n}(x_1, \dots, x_n)\; dx_1 \dots dx_n$$

If we change to new coordinates, ω changes by multiplying ω by the Jacobian of the coordinate change, and the integral changes by multiplying the integrand by the absolute value of this Jacobian. By definition, M is *oriented* if it has a coordinate cover by coordinates related to each other by positive Jacobians. If we insist on integrating using oriented coordinate systems, then integration of n-forms is well-defined on an oriented manifold.

If M is oriented and compact, we can choose a finite partition of unity φ_i subordinate to an oriented coordinate cover and define

$$\int_M \omega = \sum \int_{\mathcal{U}_i} \varphi_i \omega$$

This is easily seen to be independent of the partitian of unity. Moreover, Stokes formula shows that

$$\int_M d\lambda = 0$$

Consequently, integration over M defines a linear map

$$H^n(M) \to R$$

and we will prove that this map is an isomorphism when M is connected. If M has finitely many components, each with an orientation, we get a similar isomorphism $H^n(M) \to \sum R$ where the sum is taken over the components.

Next we obtain the Mayer-Vietoris sequence. Suppose $\mathcal{U}$ and $\mathcal{V}$ are open subsets of M. These induce inclusion maps $\mathcal{U} \cap \mathcal{V} \xrightarrow{i_1} \mathcal{U} \xrightarrow{j_1} \mathcal{U} \cup \mathcal{V}$ and $\mathcal{U} \cap \mathcal{V} \xrightarrow{i_2} \mathcal{V} \xrightarrow{j_2} \mathcal{U} \cup \mathcal{V}$, which in turn induce maps in cohomology going the reverse directions:

$$H^k(\mathcal{U} \cup \mathcal{V}) \xrightarrow{j_1^\star} H^k(\mathcal{U}) \xrightarrow{i_1^\star} H^k(\mathcal{U} \cap \mathcal{V})$$

and

$$H^k(\mathcal{U} \cup \mathcal{V}) \xrightarrow{j_2^\star} H^k(\mathcal{V}) \xrightarrow{i_2^\star} H^k(\mathcal{U} \cap \mathcal{V})$$

Then it is possible to define maps

$$H^k(\mathcal{U} \cup \mathcal{V}) \xleftarrow{D} H^{k-1}(\mathcal{U} \cap \mathcal{V})$$

making the following sequence exact.

$$\cdots \xleftarrow{i_1^\star - i_2^\star} H^{k+1}(\mathcal{U} \cup \mathcal{V}) \xleftarrow{D} H^k(\mathcal{U} \cap \mathcal{V}) \xleftarrow{j_1^\star + j_2^\star} H^k(\mathcal{U}) \oplus H^k(\mathcal{V}) \xleftarrow{i_1^\star - i_2^\star} H^k(\mathcal{U} \cup \mathcal{V}) \xleftarrow{D} \cdots$$

We will use this result to compute the cohomology of spheres, and to prove that if M is compact, then each $H^k(M)$ is finite dimensional.

There is a natural product on forms, the wedge product. It is easy to see that this product induces a multiplication $H^i(M) \otimes H^j(M) \to H^{i+j}(M)$.

We then use the Mayer-Vietoris sequence to prove two classical results from algebraic topology, the Kunneth Formula and the Poincare Duality Theorem. Suppose M and N are C^∞ manifolds. We can define a map

$$\sum_{i+j=k} H^i(M) \otimes H^j(N) \to H^{i+j}(M \times N)$$

as follows: the projections $p_M : M \times N \to M$ and $p_N : M \times N \to N$ induce maps $H^i(M) \xrightarrow{p_M^\star} H^i(M \times N)$ and $H^j(N) \xrightarrow{p_N^\star} H^j(M \times N)$. Follow these maps with a wedge product $H^i(M \times N) \otimes H^j(M \times N) \to H^{i+j}(M \times N)$. Our first theorem, *the Kunneth Formula*, asserts that the sum of such maps is an isomorphism.

As for the duality theorem, Poincare originally defined the *homology groups* by triangulating the space; in particular for surfaces he cut the surface into a finite number of points, lines, and triangles. Poincare then noticed that for each such triangulation there is a *dual triangulation*: replace each triangle by the point in its center, and replace each line between triangles by the dual line between the centers of these triangles, and finally replace each vertex by the triangle formed by these dual lines. This ultimately led Poincare to prove that for an $n-$dimensional oriented manifold, H_k and H_{n-k} are isomorphic.

We will use the Mayer-Vietoris sequence to prove a cohomological version of this result. Suppose M is a compact, oriented manifold. If ω represents an element of $H^i(M)$ and τ represents an element of $H^{n-i}(M)$ then $\omega \wedge \tau$ induces an element of $H^n(M)$ and we can integrate this element to form the real number

$$\int_M \omega \wedge \tau$$

This gives a bilinear map

$$H^i(M) \otimes H^{n-i}(M) \to R$$

The duality theorem asserts that this map is non-degenerate. If ω represents a non-zero element of $H^i(M)$, then there is a $\tau \in H^{n-i}(M)$ such that this real number is non-zero.

It follows that our map induces a one-to-one map from $H^i(M)$ to the dual space of $H^{n-i}(M)$, and a one-to-one map from $H^{n-i}(M)$ to the dual space of $H^i(M)$. Because these vector spaces are finite dimensional, it immediately follows that both of these maps are isomorphisms, and thus that $H^i(M)$ and $H^{n-i}(n)$ have the same dimension.

Notice carefully that these spaces are not canonically isomorphic. Instead $H^i(M)$ and the *dual space of* $H^{n-i}(M)$ are canonically isomorphic. This is the correct way to use the Poincare duality theorem.

Next suppose that K is a compact oriented submanifold of M of dimension k. If ω is a k-form on M, the inclusion map $i : K \to M$ induces a k-form $i^\star(\omega)$ on K, and this form can be integrated over K to form a number $\int_K i^\star(\omega)$. In this way we obtain a map

$$H^k(M) \to R$$

induced by

$$\omega \to \int_K i^\star(\omega)$$

This map belongs to the dual space of $H^k(M)$ and thus by Poincare duality corresponds to an element

$$\tau_K \in H^{n-k}(M)$$

We say this is the class *dual to* K.

Suppose next that we have two compact oriented submanifolds K and L of M, of dimensions k and l. We can form τ_K and τ_L, the elements dual to these submanifolds, and then form their wedge product $\tau_K \wedge \tau_L$ of degree $(n-k)+(n-l)$. Finally, we can form the Poincare dual of this product, which is a linear map $H^{k+l-n}(M) \to R$. Amazingly, this map is defined by the intersection $K \cap L$. Thus the wedge product in deRham cohomology corresponds to intersections of submanifolds.

However, there are subtle complications. For instance, $K \cap L$ need not be a manifold. If K and L are submanifolds of a manifold M which intersect at a point p, we say that these spaces intersect *transversally* at p if their tangent spaces at p satisfy $T_p(K) + T_p(L) = T_p(M)$. This sum will usually not be direct; instead the intersection of the two tangent spaces should be the tangent space of $K \cap L$. If the intersection is transversal, then the intersection of $T_p(K)$ and $T_p(L)$ will have dimension $k + l - n$. It K and L intersect transversally at each of their intersection points, the implicit function theorem shows that their intersection is a submanifold of dimension $k + l - n$.

Notice that the de-Rham element defined by a submanifold depends on $i : K \to M$ and the induced map $H^k(K) \xrightarrow{i^\star} H^k(M)$, which only depends on i up to homotopy. So we are free to modify K by a homotopy. According to the *Thom transversality theorem*, if K and L are submanifolds of M, we can find an arbitrarily small homotopy of i_K so that after applying this homotopy, the two submanifolds intersect transversally. Incidentally, if $k + l - n < 0$, this homotopy will lead to submanifolds K and L which do not intersect at all.

We will prove this Thom transversality theorem, and then prove that intersection corresponds to wedge product.

These ideas can then be applied to prove the Lefshetz Fixed Point Theorem. Suppose $f : M \to M$ is a C^∞ map from the compact oriented manifold M to itself. Then f induces $f^\star : H^k(M) \leftarrow H^k(M)$ and we can compute the *Lefshetz number*

$$L(f) = \sum (-1)^k trace\left(H^k(M) \xleftarrow{f^\star} H^k(M)\right)$$

The *Lefshetz Fixed Point Theorem* asserts that this number is the number of fixed points of f, if they are properly counted.

To count properly, we must first replace f by a homotopic map which has only finitely many fixed points, each transversal. This last condition means that $f^\star$ at a fixed point never leaves a non-zero vector fixed. Then we compute $\det(f^\star - I)$. If this number is positive, the fixed point counts positively, and otherwise it counts negatively.

Lefshetz's theorem follows by applying the previous intersection theory to $M \times M$. Select the diagonal submanifold of $M \times M$ and the graph of f in $M \times M$. These submanifolds intersect exactly at fixed points of f. By Thom's theorem, we can find a small homotopy so all intersection points are transverse. It then follows that these points are isolated fixed points, and $\det(f^\star - I) \neq 0$. Thus their intersection consists of a finite number of points with signs. We can compute the sum of these signs cohomologically using the wedge product, and this calculation will yield the Lefshetz number of f.

To repeat, the primary reference for these notes is Raoul Bott and Loring W. Tu's wonderful book *Differential Forms in Algebraic Topology*, published by Springer in their Graduate Texts in Mathematics Series. This book sometimes skips over easy but crucial results, like the finite dimensionality of the deRham groups for compact manifolds. I mainly wanted to place emphasis on the Lefshetz result.

Chapter 2

Partitions of Unity

2.1 Construction of C^∞ Functions

A central idea in topology is to cut spaces into simple pieces and then control reassembly of the pieces. In the deRham theory, this cutting is done using partitions of unity.

In complex analysis, a small piece of a holomorphic function completely determines the function: the theory is very rigid. We now show that this is false for C^∞ functions: their theory is very flabby.

Suppose $(x_1, \ldots, x_n)$ is a local coordinate system with coordinates defined at least on the ball of radius 3 about the origin. We will find a C^∞ function in these coordinates which is identically one on the ball of radius 1, rapidly falling to zero between the balls of radii 1 and 2, and identically zero beyond radius 2.

Start with

$$f(x) = \begin{cases} e^{-1/x^2} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

This function is certainly C^∞ except possibly at the origin. We will prove that all derivatives are continuous near the origin and zero at the origin. Notice first that $f^{(k)} = e^{-1/x^2} P(1/x)$ for $x > 0$ where P is a polynomial that depends on k, by easy induction. If we define $f^{(k)}(0) = 0$, then

$$f^{(k+1)}(0) = \lim_{h \to 0} \frac{e^{-1/h^2} P(1/h) - 0}{h}$$

and this vanishes because $\lim_{h \to 0} \frac{1}{h^m e^{1/h^2}} = 0$ for any positive integer m. Details are left to the reader.

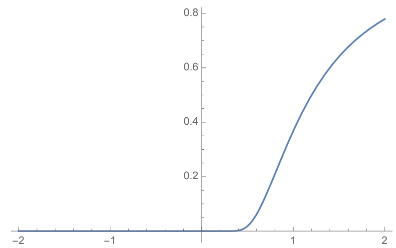

Figure 2.1: $f(x)$

Let $g(x) = f(x)f(1-x)$

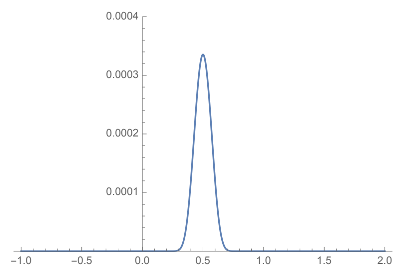

Figure 2.2: $g(x)$

Let $h(x) = \frac{\int_0^x g(t)\ dt}{\int_0^1 g(t)\ dt}$

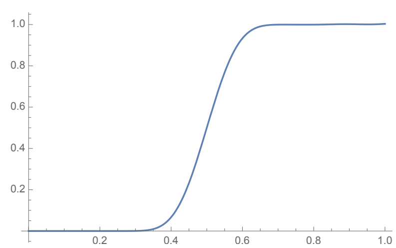

Figure 2.3: $h(x)$

Let $k(x) = h(2 - x)$

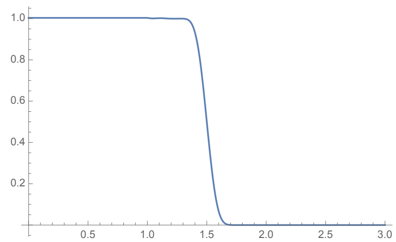

Figure 2.4: $k(x)$

Let $\varphi(x_1, \ldots, x_n) == k(\sqrt{x_1^2 + \ldots + x_n^2}$

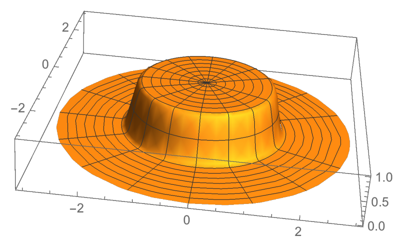

Figure 2.5: $k(x)$

2.2 Partitions of Unity

From now on, all C^∞ manifolds are assumed to have a countable basis. Clearly, we can select this basis such that each open basis set is in one coordinate system in which the coordinates are centered at the origin and extend out to radius 3. We can also assume that when we restrict these open sets so the coordinates only extend out to radius 1, the resulting open sets again form a basis.

With the technicalities out of the way, we come to the central idea. A *partiton of unity* is a countable collection of C^∞ functions φ_i on a C^∞ manifold M, such that

- $0 \leq \varphi_i \leq 1$
- each point $m \in M$ has an open neighborhood $\mathcal{V}$ such that only finitely many φ_i are nonzero somewhere in $\mathcal{V}$
- $\sum \varphi_i = 1$ on M

Notice that the third condition makes sense because of the second condition.

When constructing a partition of unity, we can replace the third condition with the weaker assertion that whenever $m \in M$, there is an i such that $\varphi_i(m) \neq 0$. Indeed, given such a system, define $\psi_i = \frac{\varphi_i}{\sum_j \varphi_j}$.

When we use partitions of unity, we always start with an open cover $\mathcal{U}_\alpha$ of M, whose open sets define "the small pieces of a dissection of M". We then introduce a partition of unity φ_i such that the set where a particular φ_i is non-zero is contained in one $\mathcal{U}_\alpha$. It may intersect others, but it is wholly inside some particular $\mathcal{U}_\alpha$.

In the discussion that follows, it is useful to consider the case where M is the union $\mathcal{U} \cup \mathcal{V}$ of two open circles shown below.

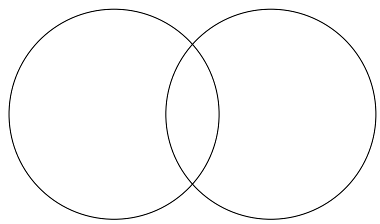

Figure 2.6: $\mathcal{U} \cup \mathcal{V}$

Theorem 1 *If $\mathcal{U}_\alpha$ is an open cover, we can construct a partition of unity φ_i subordinate to this cover and indexed by a different index set, such that each φ_i has compact support.*

Proof: Since manifolds are locally compact, each $m \in M$ has an open neighborhood $\mathcal{U}_m$ with compact closure. Our manifolds have a countable basis. Throw away basic sets which are not inside any $\mathcal{U}_m$. The remaining sets still form a countable basis, because any open $\mathcal{U}$ is the union of $\mathcal{U} \cap \mathcal{U}_m$ over all $m \in \mathcal{U}$ and $\mathcal{U} \cap \mathcal{U}_m$ is a union of basic subsets. Therefore we can assume that each basic subset has compact closure.

We now claim that we can find compact sets K_i and open sets $\mathcal{U}_i$ with $M = \cup K_i$ and

$$K_1 \subset \mathcal{U}_1 \subset K_2 \subset \mathcal{U}_2 \subset \ldots$$

Indeed make a list of basic sets and let K_1 be the closure of the first of these. This is a compact set, so it is covered by a finite union of basic open sets taken in order until enough are chosen. Call the closure of this set K_2. It is compact, so it is covered by a finite union of basic open sets taken in order from the first until enough are selected. Continue.

We now sketch the rest of the proof before filling in gaps. Cover K_1 by a finite number of "coordinate bump functions," all with support in $\mathcal{U}_1$. Cover $K_2 - Int(K_1)$ by a finite number of similar bump functions, all with support in $\mathcal{U}_2$. Cover $K_3 - Int(K_2)$ by bump functions, all with support in $\mathcal{U}_3 - K_1$. Cover $K_4 - Int(K_3)$ with bump functions, all with support in $\mathcal{U}_4 - K_2$. Continue. Only finitely many bump functions have support intersecting K_1. Only finitely many bump function supports intersect K_2. Etc. Since the union of the K_i is all of M, every point of M has an open neighborhood on which only finitely many bump functions are non-zero, and every point of M is in a set in which at least some bump functions are nonzero. This concludes the proof, modulo clarification of the construction of bump functions.

The previous step reduces the argument down to the situation $K \subset \mathcal{U}$ where K is a single compact set inside a single open set. If $m \in K$, we can find a coordinate system near m taking m to the origin and defined in some small open neighborhood about m which is inside $\mathcal{U}$. Recall that we started with an open cover $\mathcal{U}_\alpha$; by shrinking the coordinate system for m, we can assume it is entirely contained in one such $\mathcal{U}_\alpha$. By magnifying just the image of these coordinates in R^n, but not the coordinates in M, we can suppose the coordinate image contains a ball of radius 3, and thus construct a bump function near m which is 1 very near m and vanishes before it reaches the limits of the small coordinate neighborhood. Finally, using compactness, a finite number of these bump functions are non-zero on all of K. QED.

Theorem 2 *If $\mathcal{U}_\alpha$ is a countable open cover, we can construct a partition of unity subordinate to this cover and indexed by the α, with φ_α only non-zero inside $\mathcal{U}_\alpha$.*

Remark: In short, if we want the index set of the partition of unity to be the index set of the covering, then we cannot require that each φ_i have compact support. The previous picture gives an example where this restriction holds.

Proof: The second theorem follows easily from the first. First select a partition of unity φ_i subordinate to the covering such that each φ_i has compact support. The α are countable, so order them and work by induction. For the first α_1, consider all φ_i with support in $\mathcal{U}_{\alpha_1}$ and take their sum. For the next α_2, consider all unused φ_i with support in $\mathcal{U}_{\alpha_2}$ and take their sum. Continue. Since each φ_i has support in some $\mathcal{U}_\alpha$, eventually all φ_i will be used. The result clearly follows.

Chapter 3

deRham Cohomology

3.1 Differential Forms

Suppose $m \in M$ is a point in a C^∞ manifold. A k *tensor* at m is a function

$$T(X_1, \ldots, X_k) \to R$$

from ordered k-tuples of tangent vectors to R, linear in each argument if the other arguments are held fixed.

For example, a 1-tensor at m is just a dual tangent vector at m. A 2-tensor is a map $T(X, Y)$ linear in each variable. Such a map can be decomposed into symmetric and skew-symmetric pieces. This decomposition is independent of coordinate choices; the two pieces are irreducible and cannot be further decomposed:

$$T(X,Y) = S(X,Y) + \Lambda(X,Y) = \frac{T(X,Y) + T(Y,X)}{2} + \frac{T(X,Y) - T(Y,X)}{2}$$

Tensors of higher degree decompose into several pieces, but the most important are the symmetric tensors $S(X_1, \ldots, X_k)$ which are invariant under permutation of the tangent vectors, and the skew-symmetric tensors $\Lambda(X_1, \ldots, X_k)$ which are invariant up to sign $\mathrm{sgn}(\sigma)$ under a permutation σ of the vectors.

We make the set of symmetric tensors of all degrees, and the set of skew-symmetric tensors of all degrees, into algebras by defining products on the two sets. In both cases, the product just multiplies the values of two tensors together, but the result then has to be forced to be symmetric, or skew. We denote the two products by "$\odot$" and "$\wedge$" and define them as follows:

$$(T_1 \odot T_2)(X_1, \ldots, X_k, X_{k+1}, \ldots, X_l) = \frac{1}{(k+l)!} \sum_\sigma T_1(X_{\sigma(1)}, \ldots, X_{\sigma(k)}) T_2(X_{\sigma(k+1)}, \ldots, X_{\sigma(k+l)})$$

$$(T_1 \wedge T_2)(X_1, \ldots, X_k, X_{k+1}, \ldots, X_l) = \frac{1}{(k+l)!} \sum_\sigma \mathrm{sgn}(\sigma) T_1(X_{\sigma(1)}, \ldots, X_{\sigma(k)}) T_2(X_{\sigma(k+1)}, \ldots, X_{\sigma(k+l)})$$

Once we have these definitions, it is common for authors to spend time carefully developing and proving theorems which assert that the symmetric tensors form an associative, commutative algebra, and that the skew-symmetric tensors form an associative, anti-commutative algebra. These results then lead to a natural basis for the algebras in terms of a basis of the tangent space. Interested readers can invent these theorems and their somewhat tricky proofs for themselves, or look them up in books. I find it better to just state the final coordinate forms and proceed immediately to the deeper theory, and that is what we will do below.

Here's a final word about the symmetric case. A symmetric tensor $S(X_1, \ldots, X_k)$ is completely determined by the expressions $S(X, \ldots, X)$, which are homogeneous polynomials of degree k in the coefficients of the vectors. At several spots in differential geometry, a natural polynomial expression appears, and geometers reinterpret it as a sum of symmetric tensors of degrees one through n, and then use these tensors to construct crucial geometric objects. But these ideas must be left to another course, so for now we concentrate on skew-symmetric tensors.

A skew-symmetric tensor of degree k is usually called a *differential form of degree* k. It is common for these differential forms to be defined not just at one point m, but instead at each $m \in \mathcal{U}$ for an open set $\mathcal{U}$, or on all of M.

Suppose we have a coordinate neighborhood, with coordinates $x_1, \ldots, x_n$. Then at each point, a basis of the tangent space is given by $\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}$. We denote the dual basis by $dx_1, \ldots, dx_n$. Thus $dx_i \left(\frac{\partial}{\partial x_j} \right) = \delta_{ij}$.

Then it turns out that a basis of the k-forms is given by $dx_{i_1} \wedge dx_{i_2} \wedge \ldots \wedge dx_{i_k}$ where $i_1 < i_2 < \ldots < i_k$. Thus every k-form can be written uniquely as

$$\omega = \sum_{i_1 < \ldots < i_k} \omega_{i_1 \ldots i_k} dx_{i_1} \wedge \ldots \wedge dx_{i_k}$$

where the $\omega_{i_1 \ldots i_k} dx_{i_1}$ are real numbers, or functions of the coordinates if the form is defined in an open neighborhood. By definition, ω is C^∞ if each $\omega_{i_1 \ldots i_k}(x_1, \ldots, x_n)$ is C^∞.

The rules for dealing with the wedges are simple: we allow such expressions even if two indices are equal or the indices are not increasing, but then we can interchange two indices provided we change the sign of the expression, and in particular if two indices are equal, then the expression is zero.

The notation suggests the correct rule for changing coordinates. Suppose we have new coordinates $y_i(x_1, \ldots, x_n)$. It is easy to check from our definition that $dy_i = \sum \frac{\partial y_i}{\partial x_j} dx_j$. So

$$\sum_{i_1 < \ldots < i_k} \omega_{j_1, \ldots, j_k}(y_1, \ldots, y_n) dy_{i_1} \wedge \ldots \wedge dy_{i_k} =$$

$$\sum_{i_1 < \ldots < i_k} \sum_{i_1, \ldots, i_k} \frac{\partial y_{i_1}}{\partial x_{j_1}} \cdots \frac{\partial y_{i_k}}{\partial x_{j_k}} \; \omega_{j_1, \ldots, j_k}\Big(y_i(x_1, \ldots, x_n)\Big) dx_{j_1} \wedge \ldots \wedge dx_{j_k}$$

Here the $j_1, \ldots, j_k$ need not be increasing, so the wedge rules must be applied.

3.2 Tangent Vectors

Suppose $m \in M$ is a point in a C^∞ manifold. There is an easy coordinate-free definition of the tangent vectors at m, given by identifying a vector X with a resulting directional derivative $X(f)$.

To be precise, let C_m^∞ be the set of *germs* of C^∞ functions at m. A *tangent vector at* m is a linear map $X : C_m^\infty \to R$ which satisfies

$$X(fg) = X(f)g(m) + f(m)X(g)$$

In particular, $X(f)$ is defined if f is C^∞ in some open neighborhood of m, and $X(f) = X(g)$ if $f = g$ on some smaller open neighborhood of m.

Suppose $x_1, \ldots, x_n$ is a local coordinate system near m and $(X_1, \ldots, X_n)$ are n real numbers. Define a tangent vector $X = \sum X_i \frac{\partial}{\partial x_i}$ via the following formula. This clearly satisfies the definition for a tangent vector.

$$X(f) = \sum X_i \frac{\partial f}{\partial x_i}(m_1, \ldots m_n)$$

Conversely, every tangent vector has this form. Indeed, suppose f is C^∞ near m. Then

$$f(x) - f(m) = \int_0^1 \frac{d}{dt} f(tx + (1-t)m) \; dt = \sum \int_0^1 \frac{\partial f}{\partial x_i}(tx + (1-t)m) \; dt \; (x_i - m_i)$$

Apply X to both sides of this formula. Note that $X(1) = X(1 \cdot 1) = X(1) \cdot 1 + 1 \cdot X(1) = 2X(1)$, so $X(1) = 0$ and for any constant c, $X(c) = 0$. Thus X applied to the left side of the displayed formula is $X(f)$. The right side of this formula is a sum of products, where the first term in each product is an integral and the second term is $x_i - m_i$. The value of the integral when $x = m$ is $\frac{\partial f}{\partial x_i}(m)$ and the value of the second term when $x = m$ is zero. So the product rule applied to the right side gives

$$\sum \frac{\partial f}{\partial x_i}(m) X(x_i - m_i)$$

But $X(x_i - m_i) = X(x_i) - X(m_i) = X(x_i)$, which is a real number we shall call X_i. We have thus proved that for any f, $X(f)$ has the following desired form:

$$X(f) = \sum X_i \frac{\partial f}{\partial x_i}(m)$$

3.3 The d Operator

Suppose f is C^∞ on an open set $\mathcal{U}$. We define a differential 1-form df on $\mathcal{U}$ by

$$df(X) = X(f)$$

Recall that a 1-form is a cotangent vector, and thus is determined by its value on any tangent vector. Our formula determines this value.

Next we determine the formula for df in local coordinates. We know that $df = \sum \omega_i \, dx_i$ for certain real coefficients ω_i. Recall that the dx_i are a dual basis to the basis $\frac{\partial}{\partial x_i}$. Thus

$$df(X) = \left(\sum \omega_i dx_i \right) \left(\sum X_j \frac{\partial}{\partial x_i} \right) = \sum \omega_i X_i$$

and by definition this is

$$df(X) = X(f) = \sum X_i \frac{\partial f}{\partial x_i}$$

We conclude that $\omega_i = \frac{\partial f}{\partial x_i}$ and thus

$$df = \sum \frac{\partial f}{\partial x_i} \, dx_i$$

So our d operator is the *gradient operator* of advanced calculus. Notice that its value is a cotangent vector rather than a standard vector; the two objects transform differently under a coordinate change.

Theorem 3 *Let Λ^k be the vector space of k-forms on a manifold. There is a unique way to define differential operators*

$$d : \Lambda^k \to \Lambda^{k+1}$$

such that the following properties hold:

1. *on 0-forms, d is the previously defined map $f \to df$*
2. $d(\lambda \wedge \tau) = d\lambda \wedge \tau + (-1)^{\deg \lambda} \lambda \wedge d\tau$
3. $d^2 = 0$

Proof: The second and third properties imply that $d(dx_{i_1} \wedge \ldots \wedge dx_{i_k}) = 0$. So

$$d\left(\sum_{i_1 < \ldots < i_k} \omega_{i_1 \ldots i_k} dx_{i_1} \wedge \ldots \wedge dx_{i_k}\right) = \sum_j \sum_{i_1 < \ldots < i_k} \frac{\partial \omega_{i_1 \ldots i_k}(x_1, \ldots, x_n)}{\partial x_j} dx_j \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_k}$$

Note that k is out of order with the $x_{i_1}, \ldots, x_{i_k}$, so our *wedge rules* need to be applied to write the final expression in proper form.

I leave the verification that this formula has the desired properties to the reader. In particular, the proof that $d^2 = 0$ is worth the effort.

Given that the formula satisfies the conditions of the theorem, it follows that it is independent of coordinate changes. Indeed, if we have two coordinate systems, then we have two formulas which both satisfy these conditions, but we proved that the conditions uniquely determine **d**.

3.4 Examples

Special cases of these **d** operators occur in advanced calculus, and in the theory of electricity and magnetism. In these applications, a standard coordinate system has already been chosen, so the distinction between vectors and dual vectors is usually disregarded.

Dimension 1: Here 0-forms $\omega(x)$ and 1-forms $\omega(x)dx$ are essentially functions and the sequence $\Lambda^0 \to \Lambda^1 \to \ldots \to \Lambda^n$ becomes

$$\{\text{functions}\} \xrightarrow{\frac{d}{dx}} \{\text{functions}\}$$

Dimension 2: Here 0-forms $\omega(x, y)$ and 2-forms $\omega_{12}(x, y) dx \wedge dy$ are essentially functions, and 1-forms $\omega_1(x, y)dx + \omega_2(x, y)dy$ are essentially vector fields. The first **d** operator is the gradient, and the second sends $\omega_1(x, y)dx + \omega_2(x, y)dy$ to

$$\frac{\partial \omega_1}{\partial x} dx \wedge dx + \frac{\partial \omega_1}{\partial y} dy \wedge dx + \frac{\partial \omega_2}{\partial x} dx \wedge dy + \frac{\partial \omega_2}{\partial y} dy \wedge dy = \left(\frac{\partial \omega_2}{\partial x} - \frac{\partial \omega_1}{\partial y}\right) dx \wedge dy$$

So the standard sequence becomes

$$\{\text{functions } f\} \xrightarrow{\frac{\partial f}{dx}, \frac{\partial f}{dx}} \{\text{vector fields } (X, Y)\} \xrightarrow{\frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}} \{\text{functions } g\}$$

Dimension 3: Here 0-forms and 3-forms are essentially functions, and 1-forms and 2-forms are essentially vector fields. Indeed in this second case a 1-form is $X dx + Y dy + Z dz$ and a 2-form is $X dy \wedge dx + Y dz \wedge dx + Z dx \wedge dy$. A short calculation then shows that the **d** operators become curl, gradient, and divergence:

$$\{\text{functions } f\} \xrightarrow{\text{grad}} \{\text{vector fields } (X, Y, Z)\} \xrightarrow{\text{curl}} \{\text{vector fields } (X, Y, Z)\} \xrightarrow{\text{div}} \{\text{functions } g\}$$

Dimension 4: Here 0-forms and 4-forms are essentially functions, and 1-forms and 3-forms are essentially vector fields. For the first time we get a new kind of object, 2-forms, with six components. If we agree to call the coordinates x, y, z, and t and label the six components E_x, E_y, E_z and B_x, B_y, B_z, then the 2-form can be identified with the electromagnetic field and each of the **d** maps becomes a standard operator in Maxwell's theory of electricity and magnetism. Indeed half of Maxwell's equations four equations appear naturally, and the other half appear with just a little more work.

Intermission: It is useful to summarize Maxwell's theory of electricity and magnetism before continuing. According to this theory, space is permeated with two vector fields, the electric field E and the magnetic field B. If a particle with charge e and velocity vector v moves in this field, the force on the particle is

$$F = e\ E + \frac{e}{c}\ v \times B$$

The cross product term may be unexpected. When I was a student, I worked at the Harvard cyclotron. The cyclotron has a gigantic magnet with poles about eight feet across separated by a gap of a foot; this magnet created a field B pointing downward from the top pole to the bottom one. Protons were injected at very high speed horizontally into the space between the magnets. This caused a force $v \times B$ pointing inward toward the center, which forced the particles to rotate in a circle. Twice in each rotation, a mechanism increased the energy of the protons, until they were ejected with enormous energy. That cyclotron existed because of the term $v \times B$.

If you took a course in electricity and magnetism, you may be familiar with D and H rather than E and B. When solid material is placed in an electrical field, the electrons in the material are slightly displaced by the field, and this creates a force from the material which modifies the electric and magnetic fields. When the theory of electricity and magnetism is developed piece by piece from experiment, these modified fields are taken into account, leading to formulas about D and H. In empty space there are no such materials, and the more fundamental fields E and B take their place.

Getting back to the theory, we have explained how particles react to the electromagnetic field. We next explain how the field is created from these particles. Putting the two ideas together, we find that moving charges create a field, which expands outward with the speed of light and eventually sets other particles into motion. The resulting theory automatically satisfies the theory of relativity.

To describe the creation of the fields, we no longer think of particles as individual point charges, but instead imagine they are spread out in space. So suppose the charge density of the particles is given by a function ρ, and the charge and velocity density is given by a vector field J. Then the Maxwell equations which govern the theory are the following:

$$\text{div } B = 0$$

$$\text{curl } E = -\frac{1}{c}\frac{\partial B}{\partial t}$$

$$\text{div } E = 4\pi\rho$$

$$\text{curl } B = \frac{1}{c}\frac{\partial E}{\partial t} + \frac{4\pi}{c} J$$

Dimension 4, continued:

In the theory of relativity, the form $x^2+y^2+z^2-c^2t^2$ plays a significant role. Consequently, it is convenient to think of the basis vectors for Λ^1 as $dx, dy, dz, c\ dt$. This explains factors of c scattered about the following formulas.

Let us start with the map **d** from 3-forms to 4-forms. It is convenient to write a 3-form as a pair (ρ, J) where ρ is a function and J is a vector field. This expression is an abbreviation for the following 3-form; notice the extra c terms promised in the previous paragraph.

$$\rho\ dx \wedge dy \wedge dz + c\ J_x dy \wedge dz \wedge dt + c\ J_y dz \wedge dx \wedge dt + c\ J_z dx \wedge dy \wedge dt$$

A brief calculation shows that **d** applied to this form gives the following 4-form:

$$-\frac{\partial \rho}{\partial t} dx \wedge dy \wedge dz \wedge dt + c \text{ div } (J)\ dx \wedge dy \wedge dy \wedge dt = \left(-\frac{1}{c}\frac{\partial \rho}{\partial t} + \text{div } J\right) c\ dx \wedge dy \wedge dz \wedge dt$$

This corresponds to the function $-\frac{1}{c}\frac{\partial \rho}{\partial t} + \text{div } J$. So **d** from 3-forms to 4-forms maps

$$(\rho, J) \rightarrow -\frac{1}{c}\frac{\partial \rho}{\partial t} + \text{div } J$$

We explain the significance of this result later.

Next consider the map **d** from 2-forms to 3-forms. We think of a 2-form as a pair of vector fields (E, B) called the electic and magnetic fields. Specifically

$$cE_x dx \wedge dt + cE_y dy \wedge dt + cE_z dz \wedge dt + B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy$$

A brief calculation then shows that **d**: $(E, B) \to (\rho, J)$ maps this form to the 3-form

$$(E, B) \to \Big(\mathrm{div} B, \frac{1}{c}\frac{\partial B}{\partial t} + \ \mathrm{curl}\ E\Big)$$

In particular, if ω is the 2-form representing the electromagnetic field, the equation $d\omega = 0$ is equivalent to the two equations $\mathrm{div} B = 0$ and $\mathrm{curl} E = -\frac{1}{c}\frac{\partial B}{\partial t}$, which are the first two Maxwell equations.

Now consider the map **d** from 1-forms to 2-forms. It is convenient to write a 1-form as a pair (ϕ, A) where ϕ is a function and A is a vector field. So $\omega = c\ \phi dt + A_x dx + A_y dy + A_z dz$ and then a brief calculation shows that **d:** $(\phi, A) \to (E, B)$ is the map

$$(\phi, A) \to \Big(-\frac{1}{c}\frac{\partial A}{\partial t} + \mathrm{grad}\ \phi, \mathrm{curl}\ A\Big)$$

We always have $d^2 = 0$. A consequence is that if we are given ω, then we can find λ with $d\lambda = \omega$ only if $d\omega = 0$. In the next section, we will discover that the converse is often true as well. Let us apply that converse to electromagnetic theory. The 2-form $\omega = (E, B)$ satisfies $d\omega = 0$ by half of Maxwell's equations. So we expect to be able to find a 1-form $\lambda = (\phi, A)$ with $d\lambda = \omega$. Expanding and using the above displayed formula, this would give

$$E = \mathrm{grad}\ \phi - \frac{1}{c}\frac{\partial A}{\partial t} \qquad \text{and} \qquad B = \mathrm{curl}\ A$$

In fact, the physicists do exactly this, calling (ϕ, A) the *vector potential* for the electromagnetic field (for historical reasons, physicists often replace ϕ by $-\phi$). An advantage is that only four components need be determined, rather than the six components of the electromagnetic field.

Finally, consider the map from functions to 1-forms. Call our function f and our 1-form (ϕ, A). Then the map is

$$f \to \left(\frac{1}{c}\frac{\partial f}{\partial t}, \mathrm{grad}\ f\right)$$

If we modify the vector potential by the element on the right, we will not change the electromagnetic field because $\mathbf{d}^2 = 0$. Physicists call the addition of this expression to the vector potential a *gauge transformation*, and ingeniously choose f to simplify some calculations. Notice that if the sign of ϕ is changed, then the gauge transformation becomes

$$\phi \to \phi - \frac{1}{c}\frac{\partial f}{\partial t} \qquad \text{and} \qquad A \to A + \mathrm{grad}\ f$$

So far, the entire theory of differential forms works for an arbitrary C^∞ manifolds, with arbitrary local coordinates. This means, incidentally, that the electricity and magnetism discussed so far is invariant under arbitrary coordinate changes: Galileon relativity, Einstein relativity with Lorentz transformations, and even general curvilinear coordinates.

But the deeper theory of Riemannian geometry introduces a metric tensor which is positive-definite, and explores consequences of this metric. Similarly, Einstein introduced a metric tensor which is nondegenerate but not positive-definite, $dx^2+dy^2+dz^2-c^2dt^2$, and required that physics be invariant under transformations preserving this metric.

It turns out that in either case, the metric allows us to define a map $\star : \Lambda^k \to \Lambda^{n-k}$. If dx, dy, dz are orthogonal coordinates in 3-space, then $\star(dx) = dy \wedge dz, \star(dy) = dz \wedge dx$, and $\star(dz) = dx \wedge dy$. We essentially used this idea earlier in discussing the three dimensional theory.

Because of the extra c and the sign change, $\star$ is slightly more complicated in the 4-dimensional relativistic case. We only describe $\star : \Lambda^2 \to \Lambda^2$. It turns out that

$$cE_x dy \wedge dz + cE_y dz \wedge dx + cE_z dx \wedge dy + B_x dx \wedge dt + B_y dy \wedge dt + B_z dz \wedge dt \xrightarrow{\star}$$

$$-cB_x dy \wedge dz - cB_y dz \wedge dx - cB_z dx \wedge dy + E_x dx \wedge dt + E_y dy \wedge dt + E_z dz \wedge dt$$

In our earlier terminology, the star operator sends (E, B) to $(-B, E)$. Let us apply $\mathbf{d}$ to this new 2-form. We obtain the answer from our earlier calculation by interchanging E and B and changing one sign:

$$(-B, E) \to \left(\text{div } E, \frac{1}{c}\frac{\partial E}{\partial t} - \text{curl } B\right)$$

According to the second pair of Maxwell equations, these terms equal

$$\left(4\pi\rho, -\frac{4\pi}{c}J\right)$$

Recall again that $\mathbf{d}^2 = 0$. It follows that $\mathbf{d}$ applied to $\left(\rho, -\frac{1}{c}J\right)$ must be zero. We computed the $\mathbf{d}$-map on 3-forms at the beginning of this discussion, so we can just read off the answer:

$$-\frac{1}{c}\frac{\partial \rho}{\partial t} + \text{div }\left(-\frac{1}{c}J\right) = 0 \qquad \text{and so} \qquad \frac{\partial \rho}{\partial t} + J = 0$$

This last result is a famous equation in electricity and magnetism called *the continuity equation*. It implies that charge is never created, but simply moves around. Indeed $\frac{d\rho}{dt}$ measures the increase of the charge in a region, and J measures the flow of curve out of this region, so $\frac{d\rho}{dt} = -J$ asserts that the increase of charge is entirely due to this flow.

We have shown that every 4-dimensional **d** map arises in a natural way in the theory of electricity and magnetism. Of course Maxwell invented the theory in a completely different way starting with the experiments of Faraday and others. It came as a great surprise that the theory explained light as waves in the electromagnetic field; the original equations contained more mundane constants in place of c.

Faraday's experiments did not suggest the term $\frac{1}{c}\frac{\partial E}{\partial t}$ in the final Maxwell equation, so initially it was not part of Maxwell's theory. When Maxwell performed our last calculation, he found that $\frac{\partial \rho}{\partial t} + J$ was not zero, and thus that charge could be created out of thin air. Finding this improbable, he discovered that he could fix the problem by adding the extra term.

In his lectures on Physics, Richard Feynman has an interesting remark about this step. He writes:

> It was not yet customary in Maxwell's time to think in terms of abstract fields. Maxwell discussed his ideas in terms of a model in which the vacuum was like an elastic solid. He also tried to explain the meaning of his new equation in terms of the mechanical model. There was much reluctance to accept his theory, first because of the model, and second because there was at first no experimental justification. Today, we understand better that what counts are the equations themselves and not the model used to get them. We may only question whether the equations are true or false. This is answered by doing experiments, and untold numbers of experiments have confirmed Maxwell's equations. If we take away the scaffolding he used to build it, we find that Maxwell's beautiful edifice stands on its own. He brought together all of the laws of electricity and magnetism and made one complete and beautiful theory.

3.5 Maps Applied To Vectors and k-Forms

Let $\varphi : M \to N$ be a C^∞ map from one manifold to another, and suppose $p \in M$. If X is a tangent vector at p, we define $\varphi_\star(X)$ to be the tangent vector at $\varphi(p)$ defined by

$$\varphi_\star(X)(f) = X(f \circ \varphi)$$

The notation has been chosen to make the obvious coordinate form of this map obvious. Indeed let x_i be coordinates on M and y_j be coordinates on N, and let the coordinate form of φ be given by functions $y_j(x_1, \ldots, x_n)$. Suppose $X = \sum X_i \frac{\partial}{\partial x_i}$. Then

$$\varphi_\star(X)(f \circ \varphi) = \sum X_i \frac{\partial}{\partial x_i} \left(f(y_j(x_1, \ldots, x_n)) \right) = \sum X_i \frac{\partial f}{\partial y_j} \frac{\partial y_j}{\partial x_i}$$

from which we conclude that

$$\varphi_\star(X) = \sum_j \left(\sum_i X_i \frac{\partial y_j}{\partial x_i} \right) \frac{\partial}{\partial y_j}$$

If we let p vary and try to define $\varphi_\star$ of a vector field on M, we run into trouble because some points in N may not be images of anything in M, and other points in N may come from more than one point of M. So there is no such concept in the theory.

Now suppose $X_1, \ldots, X_k$ are tangent vectors at $p \in M$, and ω is a k-form at $\varphi(p)$. (We have slightly changed the notation; X_i is a vector, not a component of a vector.) We can then define a pullback form $\varphi^\star(\omega)$ at $p \in M$ by

$$\varphi^\star(\omega)(X_1, \ldots, X_k) = \omega(\varphi^\star(X_1), \ldots, \varphi^\star X_k)$$

This is clearly a $k-$form at p.

But this time, letting p vary makes perfect sense, so if ω is a k-form on all of N, then $\varphi^\star(\omega)$ is a k-form on all of M. This is one of the reasons that dealing with dual vectors and forms is nicer than dealing with vectors and their tensor products.

The notation immediately gives the coordinate form of this pullback map. Indeed it is easy to see that $dy_j = \sum \frac{\partial y_j}{\partial x_i} dx_i$, and so

$$\varphi^\star \left(\sum_{j_1 < \ldots < j_k} \omega_{j_1, \ldots, j_k}(y_j) dy_{j_1} \wedge \ldots \wedge dy_{j_k} \right) =$$

$$\sum_{j_1 < \ldots < j_k} \sum_{i_1, \ldots, i_k} \frac{\partial y_{j_1}}{\partial x_{i_1}} \ldots \frac{\partial y_{j_k}}{\partial x_{i_k}} \omega_{j_1, \ldots j_k}(y_j(x_1, \ldots, x_n) dx_{i_1} \wedge \ldots \wedge dx_{i_k}$$

In this sum, we must sum over all possible x_i and apply the "wedge rules" to simplify to increasing x_i form.

Theorem 4

$$\varphi^\star(\omega \wedge \tau) = \varphi^\star(\omega) \wedge \varphi^\star(\tau)$$

$$d\varphi^\star = \varphi^\star d$$

Proof: The first case is left to the reader and follows easily from the earlier sum defining the wedge product. To prove the second result, then, it suffices to prove that $d\varphi^\star(f) = \varphi^\star(df)$ and $d\varphi^\star(dy_j) = \varphi^\star(ddy_j) = 0$. But

$$d\varphi^\star(f) = d(f \circ \varphi) = \sum_{i,j} \frac{\partial f}{\partial y_i} \frac{\partial y_i}{\partial x_j} dx_j \qquad \varphi^\star(df) = \varphi^\star \left(\sum \frac{\partial f}{\partial y_i} dy_i \right) = \sum_i \sum_j \frac{\partial y_j}{\partial x_i} \frac{\partial f}{\partial y_i} dx_i$$

and

$$d\varphi^\star(dy_j) = d\left(\sum_i \frac{\partial y_j}{\partial x_i} dx_i\right) = \sum_{i,k} \frac{\partial^2 y_j}{\partial x_i \partial x_k} dx_k \wedge dx_i$$

This expression is zero because the second partial is symmetric in i, k and the wedge is skew-symmetric in i, k.

3.6 Poincare's Lemma

If ω is a k-form, we cannot find a $k-1$-form λ with $d\lambda = \omega$ unless $d\omega = 0$ because $d^2 = 0$. The equation $d\lambda = \omega$ is a differential equation, and $d\omega = 0$ is an integrability condition for this equation. According to Poincare's lemma, it is the only integrability condition; there may be other global topological requirements before we can find λ, but no other local differential requirements.

Theorem 5 *Suppose ω is a k-form defined on an open rectangular box U of R^n. If $d\omega = 0$, then there is a $k-1$-form λ on $\mathcal{U}$ with $d\lambda = \omega$.*

Proof: Call the coordinates $x_1, \ldots, x_n$ and write $\omega = dx_1 \wedge \tau_1 + \tau_2$ where no term in τ_2 has a dx_1. Each coefficient of τ_1 can be integrated in the x_1 direction; replace these coefficients by their indefinite integrals and call the result λ. If we compute $d\lambda$, we obtain all the terms in $dx_1 \wedge \tau_1$ and other terms, none of which involve dx_1. Consequently we may write $\omega = d\lambda + \omega_1$ where no term in ω_1 has dx_1 and $d\omega_1 = 0$. But if a differential form ω_1 has no terms containing dx_1 and $d\omega_1 = 0$, then the coefficients of ω_1 must be independent of x_1. So now we have written $\omega = d\lambda + \omega_1$ where ω_1 is a form in R^{n-1}.

Continue this argument by induction. This time $\omega_1 = d\lambda_1 + \omega_2$ where $d\omega_2 = 0$ and the coefficients of ω_2 are functions only of x_3 and higher. Continue. Eventually ω is a sum of $d\lambda_i$. QED.

3.7 The deRham Cohomology Groups

If ω is a k-form on a manifold M and $d\omega = 0$, it does not necessarily follow that there is a $k-1$-form λ on M with $d\lambda = \omega$. For example, suppore $M = R^2 - (0, 0)$ and let

$$\omega = \frac{-y}{x^2 + y^2} dx + \frac{x}{x^2 + y^2} dy$$

The easiest way to see that $d\omega = 0$ is to notice that it is the gradient of $\arctan \frac{y}{x}$ when $x \neq 0$ and the gradient of $-\arctan \frac{x}{y}$ when $y \neq 0$. In both cases, then, our form is the gradient of $\arctan \theta$ where θ is the standard angle of polar coordinates. But this is not a single valued function. Indeed, if ω were globally the gradient of λ, then λ would have to be a function which continuously increases as we circle the singularity at the origin, which is impossible.

So ω only equals $d\lambda$ if it satisfies $d\omega = 0$ and additional global topological conditions. The deRham cohomology groups capture these additional global conditions.

Definition 1 *Let M be a C^∞ manifold. The deRham cohomology groups are the groups (actually real vector spaces)*

$$H^k(M) = \frac{\{k \text{ forms } \omega \mid d\omega = 0\}}{\{\omega = d\lambda \mid \lambda \text{ is a } k-1 \text{ form}\}}$$

Definition 2 *Let $\varphi : M \to N$ be a C^∞map. Since $d\varphi^\star = \varphi^\star d$, $\varphi^\star$ induces a map*

$$\varphi^\star : H^k(M) \leftarrow H^k(N)$$

Example 1: Suppose $k = 0$. Then $H^0(M) = \{0 \text{ functions } f \mid df = 0\}$. Such a function is locally constant, and thus constant on each connected component of M. So if there are only finitely many components,

$$H^0(M) = \bigoplus_{\text{components of } M} R$$

Example 2: Suppose $k = n$. Consider a typical n-form

$$\omega = \omega_{1,\ldots,n}(x_1, \ldots, x_n) dx_1 \wedge dx_2 \wedge \ldots \wedge dx_n$$

If this $n-$ form has compact support, and exists in a single coordinate system which we can imagine extends in all directions to infinity, we can form

$$\int \ldots \int_{-\infty}^{\infty} \omega_{1,\ldots,n} dx_1 \ldots dx_n$$

If we change to new coordinates $y_1, \ldots, y_n$, then the $n - form$ changes through multiplication by

$$\det \left(\frac{\partial y_j}{\partial x_i} \right)$$

and the integral changes through multiplication by

$$\left| \det \left(\frac{\partial y_j}{\partial x_i} \right) \right|$$

Thus the integral is almost independent of coordinate changes, and we can fix that with a definition.

Definition 3 *An* orientation *on a C^∞ manifold M is an open cover by coordinate systems such that whenever $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ are systems in the orientation,*

$$\det\left(\frac{\partial y_j}{\partial x_i}\right) > 0$$

Two orientations are equivalent *if their union is again an orientation. If M has an orientation, it is said to be* orientable.

Remark: If M is connected, it is easy to prove that it either has no orientations, or else exactly two orientations up to equivalence.

Example 2, continued: Let M be an oriented, compact manifold. Choose a coordinate cover by oriented coodinate systems, and choose a partition of unity φ_i subordinate to this covering. Since M is compact, we can assume that there are only finitely many φ_i, each with compact support inside a coordinate system defining the orientation. If ω is an n-form, define

$$\int_M \omega = \sum_i \int \ldots \int \varphi_i(x_1, \ldots, x_n)\ \omega_{1,\ldots,n}(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

It is easy to check that this expression is independent of the particular coordinates defining the orientation and of the choice of partition of unity. Indeed, given two such choices, take the union of the coordinate systems in both, and replace the two partitions of unity by the set of products of their elements. This is a new choice, and it is easy to prove that it gives the same integral as either of the original choices.

Theorem 6 *If M is compact and oriented, and λ is an $n-1$-form, $\int_M d\lambda = 0$. Consequently, integration induces a well-defined map*

$$H^n(M) \xrightarrow{\int_M \omega} R$$

Remark: If M is connected, we will later prove that this map is an isomorphism.

Proof: Since $\sum \varphi_i = 1$ and this is a finite sum,

$$\sum \varphi_i d\lambda = d\lambda = d(\sum \varphi_i \lambda) = \sum d(\varphi_i \lambda)$$

Consequently

$$\int_M \sum \varphi_i d\lambda = \int_M \sum d(\varphi_i \lambda) = \sum \int_M d(\varphi_i \lambda)$$

so it suffices to prove the theorem in one coordinate system in which λ has compact support.

But $\int d\lambda$ in one coordinate system equals

$$\int \dots \int_{-\infty}^{\infty} \sum_i \frac{\partial}{\partial x_i} \lambda_{1\dots\hat{i}\dots n}(x_1, \dots, x_n)\; dx_i \wedge dx_1 \wedge \dots \wedge \hat{dx_i} \wedge \dots \wedge dx_n$$

where the hat indicates an omitted symbol. We are free to perform the n required integrations in any order. For the ith term, integrate first with respect to x_i. Since λ has compact support, this will involve integrating the derivative of a function from a spot where the function equals zero to another spot where the function equals zero. Hence that particular integral is zero. Similarly each term separately integrates to zero.

3.8 Homotopy Invariance of Induced Maps

Definition 4 *Suppose M and N are C^∞ manifolds and let φ_0 and φ_1 be C^∞ maps from M to N. We say these maps are C^∞homotopic if there is a C^∞ map*

$$h : (-\epsilon, 1 + \epsilon) \times M \to N$$

such that $h(0, m) = \varphi_0(m)$ and $h(1, m) = \varphi_1(m)$.

Theorem 7 *Suppose φ_0 and $\varphi_1 : M \to N$ are C^∞homotopic. Then*

$$\varphi_0^\star = \varphi_1^\star : H^k(M) \leftarrow H^k(N)$$

Proof: Let i_0 and $i_1 : M \to *(-\epsilon, 1 + \epsilon) \times M$ be the maps sending m respectively to $0 \times m$ and $1 \times m$. Notice that i_0 and i_1 are homotopic in a trivial manner. The following sequences

$$\varphi_0 : M \xrightarrow{i_1} (-\epsilon, 1 + \epsilon) \times M \xrightarrow{h} N$$

$$\varphi_1 : M \xrightarrow{i_0} (-\epsilon, 1 + \epsilon) \times M \xrightarrow{h} N$$

induce

$$\varphi_0^\star : H^k(M) \xleftarrow{i_0^\star} H^k((-\epsilon, 1 + \epsilon) \times M) \xleftarrow{h^\star} H^k(N)$$

$$\varphi_1^\star : H^k(M) \xleftarrow{i_1^\star} H^k((-\epsilon, 1 + \epsilon) \times M) \xleftarrow{h^\star} H^k(N)$$

These diagrams show that it suffices to prove that $i_0^\star = i_1^\star$. The central idea of the proof of this fact is to construct a series of maps $L^k : \Lambda^{k-1}(M) \leftarrow \Lambda^k((-\epsilon, 1 + \epsilon) \times M)$ with the property that

$$d \circ L + L \circ d = i_1^\star - i_0^\star$$

Suppose we know these maps. If $\omega \in \Lambda^k((-\epsilon, 1 + \epsilon) \times M)$ defines an element of deRham cohomology, then $d\omega = 0$ and consequently

$$(d \circ L + L \circ d)\omega = d(L\omega) = i_1^\star(\omega) - i_0^\star(\omega)$$

so that these last two elements are equivalent and define the same element of $H^k(M)$.

Next we define L. If ω is a k-form on $(-\epsilon, 1+\epsilon) \times M$, then this form has terms involving dt and other terms with no dt. Ignore the second kind of term, and map the first kind $\omega_{i_1,\ldots,i_{k-1}}(t, x_1, \ldots, x_n)\ dt \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}}$ to

$$\left(\int_0^1 \omega_{i_1,\ldots,i_{k-1}}(t, x_1, \ldots, x_n) dt \right) dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}}$$

To finish the proof, it suffices to compute $dL + Ld$ on a fixed k-form. Such forms are sums of terms, and it suffices to consider each individual term. Consider first the case when a term has no dt and thus equals $\omega(t, x_1, \ldots, x_n) dx_{i_1} \wedge \ldots \wedge dx_{i_k}$. Then L of this term is zero, so $dL\omega = 0$. We must compute $Ld\omega$. The only term of $d\omega$ not killed by L is $\frac{\partial \omega(t,x_1,\ldots,x_n)}{\partial t} dt \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_k}$ and L of this term is

$$\left(\int_0^1 \frac{\partial \omega(t, x_1, \ldots, x_n)}{\partial t} dt \right) dx_{i_1} \wedge \ldots \wedge dx_{i_k} =$$

$$\omega(1, x_1, \ldots, x_n) dx_{i_1} \wedge \ldots \wedge dx_{i_k} - \omega(0, x_1, \ldots, x_n) dx_{i_1} \wedge \ldots \wedge dx_{i_k} = i_1^\star \omega - i_0^\star \omega$$

Finally suppose a term has a dt and thus equals $\omega(t, x_1, \ldots, x_n) dt \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}}$. Then

$$Ld\omega = L\left(-\sum_k \frac{\partial \omega}{\partial x_k} dt \wedge dx_k \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}} \right) = -\sum_k \left(\int_0^1 \frac{\partial \omega}{\partial x_k} dt \right) dx_k \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}}$$

and

$$dL\omega = d\left(\int_0^1 \omega\ dt \right) dx_1 \wedge \ldots \wedge dx_{i_{k-1}} =$$

$$\left(\int_0^1 \frac{\partial \omega}{\partial t} dt \right) dt \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}} + \sum_k \left(\int_0^1 \frac{\partial \omega}{\partial x_k} dt \right) dx_k \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}}$$

So

$$dL\omega + Ld\omega = \left(\int_0^1 \frac{\partial \omega}{\partial t} dt \right) dt \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}} =$$

$$\omega(1, x_1, \ldots, x_n) dt \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}} - \omega(0, x_1, \ldots, x_n) dt \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}} = i_1^\star(\omega) - i_0^\star(\omega)$$

Chapter 4

The Mayer-Vietoris Sequence

4.1 The Sequence

The following important theorem allows us to compute cohomology groups of M by stitching small pieces of M together and examining the changing behavior of the groups.

Suppose $\mathcal{U}$ and $\mathcal{V}$ are open subsets of M. These induce inclusion maps $\mathcal{U} \cap \mathcal{V} \xrightarrow{i_1} \mathcal{U} \xrightarrow{j_1} \mathcal{U} \cup \mathcal{V}$ and $\mathcal{U} \cap \mathcal{V} \xrightarrow{i_2} \mathcal{V} \xrightarrow{j_2} \mathcal{U} \cup \mathcal{V}$, which in turn induce maps in cohomology going the reverse directions:

$$H^k(\mathcal{U} \cap \mathcal{V}) \xleftarrow{i_1^\star} H^k(\mathcal{U}) \xleftarrow{j_1^\star} H^k(\mathcal{U} \cup \mathcal{V})$$

and

$$H^k(\mathcal{U} \cap \mathcal{V}) \xleftarrow{i_2^\star} H^k(\mathcal{V}) \xleftarrow{j_2^\star} H^k(\mathcal{U} \cup \mathcal{V})$$

Theorem 8 (Mayer-Vietoris) *It is possible to define maps D as below*

$$H^k(\mathcal{U} \cup \mathcal{V}) \xleftarrow{D} H^{k-1}(\mathcal{U} \cap \mathcal{V})$$

making the following sequence exact.

$$\cdots \xleftarrow{i_1^\star - i_2^\star} H^{k+1}(\mathcal{U} \cup \mathcal{V}) \xleftarrow{D} H^k(\mathcal{U} \cap \mathcal{V}) \xleftarrow{j_1^\star - j_2^\star} H^k(\mathcal{U}) \oplus H^k(\mathcal{V}) \xleftarrow{i_1^\star + i_2^\star} H^k(\mathcal{U} \cup \mathcal{V}) \xleftarrow{d} \cdots$$

Part 1 of Proof: Select a partition of unity $\varphi_{\mathcal{U}}$ and $\varphi_{\mathcal{V}}$ for $\mathcal{U} \cup \mathcal{V}$ so that $\varphi_{\mathcal{U}}$ is non-zero only in $\mathcal{U}$ and $\varphi_{\mathcal{V}}$ is nonzero only in $\mathcal{V}$. (See the last result in the section on partitions of unity for details.)

Using this partition of unity, we prove that each of the following sequences is exact:

$$0 \leftarrow \Lambda^k(\mathcal{U} \cap \mathcal{V}) \xleftarrow{j_1^\star - j_2^\star} \Lambda^k(\mathcal{U}) \oplus \Lambda^k(\mathcal{V}) \xleftarrow{i_1^\star + i_2^\star} \Lambda^k(\mathcal{U} \cup \mathcal{V}) \leftarrow 0$$

Much of this is trivial. Exactness on the right is the assertion that a form ω on $\mathcal{U} \cup \mathcal{V}$ vanishes just in case its restrictions to $\mathcal{U}$ and $\mathcal{V}$ both vanish. Exactness in the center is the assertion that a form on $\mathcal{U}$ and a separate form on $\mathcal{V}$ can be glued together to make a form on $\mathcal{U} \cup \mathcal{V}$ just in case both of their restrictions to $\mathcal{U} \cap \mathcal{V}$ are equal.

Finally, exactness at the left is proved using our partition of unity. Suppose ω is a form on $\mathcal{U} \cap \mathcal{V}$. Then $\varphi_{\mathcal{V}}\, \omega$ modifies ω on the intersection so that it extends to zero on the rest of $\mathcal{U}$ and remains C^∞. Similarly $\varphi_{\mathcal{U}}\, \omega$ modifies ω in the intersection so it can be extended to zero on the rest of $\mathcal{V}$ while remaining C^∞. Take the element in $\Lambda^k(\mathcal{U}) \oplus \Lambda^k(\mathcal{V})$ which equals $\varphi_{\mathcal{V}}\, \omega$ in the left term and $-\varphi_{\mathcal{U}}\, \omega$ in the right term. It maps by $j_1^\star - j_2^\star$ to $(\varphi_{\mathcal{V}} + \varphi_{\mathcal{U}})\, \omega = \omega$ on $\mathcal{U} \cap \mathcal{V}$.

Intermission: Some readers may still be dubious that the extension of $\varphi_{\mathcal{V}}\, \omega$ on $\mathcal{U} \cap \mathcal{V}$ to be zero on the rest of $\mathcal{U}$ is still C^∞. We will fill in details of that argument. Recall that we started with a countable partition of unity φ_i such that each element has compact support inside either $\mathcal{U}$ or $\mathcal{V}$, and such that every point has an open neighborhood on which only finitely many φ_i are nonzero. We normalized so $\sum \varphi_i = 1$. Then we let $\varphi_{\mathcal{U}}$ be the sum of all φ_i with support inside $\mathcal{U}$ and $\varphi_{\mathcal{V}}$ be the sum of the remaining elements.

Consider then $\varphi_{\mathcal{V}}\, \omega = \sum_{\text{some } i} \varphi_i \omega$. Each φ_i in this sum has compact support in $\mathcal{V}$ and thus vanishes before it reaches the boundary of $\mathcal{V}$ and thus before it reaches the portion of $\mathcal{U}$ which is not in $\mathcal{U} \cap \mathcal{V}$. So $\varphi_i \omega$ remains C^∞ if it is extended from $\mathcal{U} \cap \mathcal{V}$ to be zero in the rest of $\mathcal{U}$. The sum of these extensions is C^∞ because each point has an open neighborhood where only finitely many of these terms is nonzero.

Part 2 of Proof:

Surprisingly, the rest of the proof is abstract homological algebra and diagram chasing. By definition, a cohomological complex is a chain of vector spaces and d maps such that d^2 is always zero.

$$\ldots A^3 \xleftarrow{d} A^2 \xleftarrow{d} A^1 \xleftarrow{d} A^0 \leftarrow 0$$

For any such complex, we can form the cohomology groups (actually vector spaces) just as we did in the deRham theory.

Suppose we have a diagram as on the next page consisting of vectical complexes A, B, C and horizontal short exact sequences, such that the diagram commutes. Then it is possible to define $D : H^k(C) \xrightarrow{D} H^{k+1}(A)$ making the analogue of the Mayer-Vietoris sequence exact.

$$\begin{array}{ccccccccc}
 & & \uparrow d & & \uparrow d & & \uparrow d & & \\
0 & \longleftarrow & C^{k+1} & \xleftarrow{j} & B^{k+1} & \xleftarrow{i} & A^{k+1} & \longleftarrow & 0 \\
 & & \uparrow d & & \uparrow d & & \uparrow d & & \\
0 & \longleftarrow & C^{k} & \xleftarrow{j} & B^{k} & \xleftarrow{i} & A^{k} & \longleftarrow & 0 \\
 & & \uparrow d & & \uparrow d & & \uparrow d & & \\
0 & \longleftarrow & C^{k-1} & \xleftarrow{j} & B^{k-1} & \xleftarrow{i} & A^{k-1} & \longleftarrow & 0 \\
 & & \uparrow d & & \uparrow d & & \uparrow d & &
\end{array}$$

$$\leftarrow H^{k+1}(A) \xleftarrow{D} H^k(C) \xleftarrow{j} H^k(B) \xleftarrow{i} H^k(A) \leftarrow$$

This is one of those arguments where you cannot go wrong if you keep your wits about you. We will try a few cases and leave the rest to the reader. First we define D. Suppose $c \in C^k$ induces an element in $H^k(C)$. Then $dc = 0$ and $c = j(b)$. By commutativity of the diagram, $j(db) = 0$, so $db = i(a)$ where $a \in A^{k+1}$. This a defines $Dc \in H^{k+1}(A)$. But we need to show that $da = 0$ to prove that a defines an element of cohomology. Clearly $ddb = 0$, so by commutativity $ida = 0$. Since i is one-to-one, $da = 0$.

Let us prove exactness at $H^k(B)$. First $ji = 0$ because $ji = 0$ in the complex diagram. Conversely, suppose $b \in B^k$ represents an element of $H^k(B)$ which maps to zero in $H^k(C)$. Then $j(b) = dc$ for some $c \in C^{k-1}$. This c equals $j\hat{b}$ for some $\hat{b} \in B^{k-1}$. Then $jd\hat{b} = djb = dc = jb$. So $b \in B^k$ and $d\hat{b} \in B^k$ map to the same element, and therefore their difference comes from A^k. So $b - d\hat{b} = ia$. This element a represents an element of $H^k(A)$ because $ida = dia = d(b - d\hat{b}) = 0$ and i is one-to-one. But b and $b - d\hat{b}$ represent the same element of $H^(B)$, and this element comes from $H^k(A)$, proving exactness. Whew.

Remark: For later use, it is convenient to have an explicit definition of D in the Mayer-Vietoris case. We are to start with a form ω defining an element of $H^k(\mathcal{U} \cap \mathcal{V})$. The general definition calls for writing this element as $j_1(\omega_1) - j_2(\omega_2)$ where ω_1 is a form on $\mathcal{U}$ and ω_2 is a form on $\mathcal{V}$. Indeed $\omega = \varphi_{\mathcal{V}}\omega - (-\varphi_{\mathcal{U}})\omega$. We are then told to take d of this element and discover that it comes from a form on $\mathcal{U} \cup \mathcal{V}$. Indeed $d\omega = 0 = d\left(\varphi_{\mathcal{V}}\omega\right) + d\left(\varphi_{\mathcal{U}}\omega\right)$ on $\mathcal{U} \cap \mathcal{V}$, so the two forms agree up to sign and we can define

$$D\omega = \begin{cases} \;\; d\ \varphi_{\mathcal{V}}\omega & \text{on } \mathcal{U} \\ -d\ \varphi_{\mathcal{U}}\omega & \text{on } \mathcal{V} \end{cases}$$

4.2 Cohomology Groups of Spheres

It M is a single point, clearly $H^0(M) = R$ and $H^k(M) = 0$ for $k \neq 0$.

Suppose $\mathcal{U}$ is contractible to a point. We then get maps

$$\{\text{point}\} \xrightarrow{i} \mathcal{U} \xrightarrow{j} \{\text{point}\} \xrightarrow{i} \mathcal{U}$$

where the map ji is the identity, and the map ij is homotopic to the identity. It follows that both compositions in the following sequence as isomorphisms, and thus $H^k(\mathcal{U})$ has the same cohomology as a point.

$$H^k(\{point\}) \xleftarrow{i^\star} H^k(\mathcal{U}) \xleftarrow{j^\star} H^k(\{point\}) \xleftarrow{i^\star} H^k(\mathcal{U})$$

Now consider the various spheres. Start with S^1, the ordinary circle. Let $\mathcal{U}$ be the upper half of the circle plus a little extra at the endpoints, and let $\mathcal{V}$ be the lower half of the circle plus a little extra. Then $\mathcal{U}$ and $\mathcal{V}$ are contractible and have the cohomology of a point. Also $\mathcal{U} \cap \mathcal{V}$ consists of small open intervals about $(-1, 0)$ and $(1, 0)$, and thus is contractible to two points. We easily write down the Mayer-Vietorus sequence as follows:

$$0 \oplus 0 \leftarrow H^(S^1) \xleftarrow{D} R \oplus R \xleftarrow{j_1^\star - j_2^\star} R \oplus R \xleftarrow{i_1^\star + i_2^\star} H^0(S^1) \leftarrow 0$$

We know that $H^0(S^1) = R$. The image in the right copy of $R \oplus R$ must have dimension one, so the kernel of the map from this space to the left copy of $R \oplus R$ must have dimension one; it follows that $H^1(S^1) = R$. Clearly all higher $H^k(S^1)$ are zero because there are no higher forms in one dimension.

Theorem 9 *The only non-zero cohomology groups of the n-dimensional sphere are*

$$H^0(S^n) = R \qquad\qquad H^n(S^n) = R$$

Proof: Suppose n is at least two. Let $\mathcal{U}$ be the upper hemisphere of the sphere, plus a little more, and let $\mathcal{V}$ be the lower hemisphere, plus a little more. Then $\mathcal{U} \cap \mathcal{V}$ is an open neighborhood of the equator, and is contractible to this equator, which equals S^{n-1}. So the Meyer-Vietoris sequence is

$$0 \leftarrow H^n(S^n) \leftarrow H^{n-1}(S^{n-1}) \leftarrow 0 \oplus 0 \leftarrow H^{n-1}(S^n) \leftarrow H^{n-2}(S^{n-1}) \leftarrow 0 \oplus 0$$
$$\leftarrow \ldots \leftarrow H^1(S^n) \leftarrow H^0(S^{n-1}) \leftarrow R \oplus R \leftarrow R \leftarrow 0$$

Work by induction on n starting with $n = 2$. At the right of the sequence, S^{n-1} is connected so $H^0(S^{n-1}) = R$. We conclude that $H^1(S^n) = 0$. The sequence shows that for larger k, $H^k(S^n)$ is isomorphic to $H^{k-1}(S^{n-1})$. The result immediately follows.

Remark: Here is a famous application:

Theorem 10 (Brouwer) *Let B^n be the unit ball in R^n, $n \geq 1$. Suppose $f : B^n \to B^n$ is continuous. Then f leaves at least one point fixed.*

Proof:

Our theory deals with C^∞ maps, so the first step is to replace f with a similar C^∞ map.

Assume the theorem false for f. Then $||f(x) - x||$ has a positive minimum value c on the closed ball. Use the Stone-Weierstrass theorem to approximate f on the ball by a polynomial $P(x)$, so $||f(x) - P(x)|| < c/3$. Replace $P(x)$ by $P(x)/(1 + c/3)$, noting that the resulting map sends the closed ball back inside itself. We claim that $P(x)/(1 + c/3)$ also has no fixed points in the ball. And, of course, this function is C^∞.

Indeed

$$||f(x) - P(x)/(1 + c/3)|| \leq ||f(x) - P(x)|| + ||P(x) - P(x)/(1 + c/3)|| \leq$$

$$c/3 + (1 + c/3)(1 - 1/(1 + c/3)) \leq 2c/3.$$

So

$$||x - P(x)/(1 + c/3)|| \geq ||x - f(x)|| - ||f(x) - P(x)/(1 + c/3)|| \geq c - 2c/3 = c/3 > 0$$

From now on, replace $P(x)/(1+c/3)$ by $f(x)$, but assume that f is C^∞ in a neighborhood of the ball.

Since f has no fixed points, we can define a map $g : B^n \to S^{n-1}$ by drawing a line from $f(x)$ to x and continuing this line until it hits the boundary of the ball at $g(x)$. Notice that $g : S^{n-1} \to S^{n-1}$ is the identity. We thus obtain $S^{n-1} \xrightarrow{i} B^n \xrightarrow{g} S^{n-1}$ and this composition is the identity. So

$$H^{n-1}(S^{n-1}) \xleftarrow{i^\star} H^{n-1}(B^n) \xleftarrow{g^\star} H^{n-1}(S^{n-1})$$

is the identity map. This is impossible because the left and right groups are R and the group in the center is 0.

Remark: To make this completely rigorous, we must replace B^n with a slightly larger contractible open set, since our theory deals with manifolds and B^n isn't a manifold. Our map f comes from a polynomial and is certainly defined on a slightly larger open set. The remaining details are left to the reader.

Remark: We proved that $H^1(S^1) = R$. If (r, θ) are polar coordinates in the plane, then θ is a coordinate system on the circle. The generator of $H^1(S^1)$ is $d\theta$. Notice that θ is not globally defined on the circle; said another way, it is a multiple-valued function. Therefore $d\theta$ is not zero in cohomology.

Remark: Clearly $R^2 - \{0\}$ has the homotopy type of S^1 and thus the same cohomology groups. This time the generator of H^1 is

$$\omega = \frac{-y}{x^2 + y^2}\, dx + \frac{x}{x^2 + y^2}\, dy$$

4.3 Good Coverings, Part 1

We now seek to prove several crucial results about deRham cohomology using the Mayer-Vietoris sequence. In all cases, we will start with a finite covering of M by contractible open $\mathcal{U}_i$ and work by induction to prove the results for $\mathcal{U}_1 \cup \ldots \cup \mathcal{U}_k$ until we reach the complete union. To make this method work, it is not enough that the $\mathcal{U}_i$ be contractible; all possible intersections $\mathcal{U}_{i_1} \cap \ldots \cap \mathcal{U}_{i_j}$ also need to be empty or contractible.

Definition 5 *A* good covering *of a C^∞ manifold M is a covering by open sets such that*

- *each open set is diffeomorphic to R^n*
- *each finite intersection $\mathcal{U}_{i_1} \cap \ldots \cap \mathcal{U}_{i_k}$ is either empty or else diffeomorphic to R^n*

The question then arises whether such a cover exists. To get an idea of how to proceed, consider the torus T^2 and think of it as R^2 modulo the integer lattice, so two points in R^2 are equivalent if their components differ by integers. In this case, we can let the $\mathcal{U}_i$ be open disks smaller than the lattice. These disks are convex: any two points can be joined by a straight line in the disk. It then follows trivially that intersections are also convex. In particular, these intersections are star-shaped and thus diffeomorphic to R^n.

Unfortunately, it doesn't make sense to talk about convex open sets in a manifold. However, if we give our manifold a Riemannian structure, then we can replace straight lines by geodesics. Using these geodesics, we will prove that any second countable C^∞ manifold has a good cover. This result is false for arbitrary topological manifolds, so the proofs we give are somewhat subtle.

If M is a C^∞ manifold with a countable base, it is possible to define a Riemannian metric on M. Such a metric is a positive definite inner product $\langle X, Y \rangle$ on each tangent space such that the metric varies from point to point in a C^∞ manner.

Once such a metric is present, we can define the length of a parameterized curve $\gamma(t)$, where $a \le t \le b$, by

$$\int_a^b \sqrt{\left\langle \frac{d\gamma}{dt}, \frac{d\gamma}{dt} \right\rangle}\, dt$$

We then obtain *geodesics* by applying the calculus of variations to this length integral. This yields a differential equation whose solutions minimize length between points, at least locally. In practice we minimize *energy* rather than *length*, and consequently solutions minimize length locally *and* are traced with fixed speed.

$$\frac{d^2\gamma_i}{dt^2} + \sum \Gamma^i_{j_1 j_2} \frac{d\gamma_{j_1}}{dt}\frac{d\gamma_{j_2}}{dt} = 0$$

To be clear, a *geodesic* is a solution of this equation, whether or not it minimizes curve length. On the sphere S^2, for example, geodesics are great circles, and the minimal path between two points is the short portion of the great circle through them. But the opposite longer portion is still a geodesic even though it does not minimize length. If the two points are poles, then infinitely many great circles pass through them, all minimizing length.

If we multiply the parameter t by a constant, we get another geodesic with the same length which just moves faster or slower. It is thus convenient to talk about equivalence classes of geodesics modulo such reparameterization.

Definition 6 *An open set $\mathcal{U}$ is said to be* geodesically convex *if any pair p, q of points in $\mathcal{U}$ can be joined by exactly one shortest geodesic, up to equivalence, and this geodesic lies entirely in $\mathcal{U}$.*

Theorem 11 *Every point p in a Riemannian manifold M has a geodesically convex open neighborhood.*

Remark: This result was first proved in 1932 by G. H. C. Whitehead. The proof below is taken from differential geometry notes by Ben Andrews at Australian National University, available on the web.

Remark: Note that finite intersections of geodesically convex open sets are either empty or else again geodesically convex. Consequently they satisfy the main condition for a good covering. In part 2, we will prove that each is diffeomorphic to R^n.

Proof: The geodesic equation is second order, so its solutions are determined by two boundary conditions, $\gamma(0) = q$ and $\frac{d\gamma}{dt}(0) = X$. The general form of the local existence theorem applies to these boundary conditions, and guarantees that there are solutions $\gamma_{q,X}(t)$ defined for $q \in \mathcal{U}$, $||X|| < \delta$, $|t| < \eta$ and C^∞ in all of these variables. Here $\mathcal{U}$ is an open neighborhood of p.

Suppose λ is a constant, and notice that $\gamma(\lambda t)$ also satisfies the geodesic equation, but with boundary condition $\gamma'(0) = \lambda X$. In other words, $\gamma_{p,X}(\lambda t) = \gamma_{p,\lambda X}(t)$. Hence we may shrink the size of X by shrinking δ and simultaneously increasing η. So assume that $\eta = 2$ and geodesics given by the existence theorem are defined for $|t| < 2$.

Fix q to be the initial p of the theorem, and define $exp(X) : \mathcal{D} \subset T_p(M) \to M$ by $exp(X) = \gamma_{p,X}(1)$. This map is defined on a domain $\mathcal{D}$ consisting of tangent vectors with norm less than δ.

The exponential map is easy to understand geometrically. It maps straight lines through the origin in the tangent space to geodesics through p in M because $\exp(tX) = \gamma_{p,tX}(1) = \gamma_{p,X}(t)$.

We claim that the exponential map is a local diffeomorphism from an open neighborhood of $0 \in T_p(M)$ to an open set in M. This will follow from the inverse function theorem if we can show that $exp^\star(0)$ is nonsingular. To show this, take a path $\tau(t)$ in $T_p(M)$ with derivative X at $t = 0$. Form the path $exp(\tau(t))$ and take the derivative of this path at $t = 0$. This derivative is $\exp^\star(X)$. But one easy τ is $\tau(t) = tX$ and then $exp(tX) = \gamma_X(t)$ and the derivative of this path at $t = 0$ is X. So $exp^\star$ is the identity map and the inverse function theorem applies.

From now on, assume δ is small enough that the exponential map defined on the ball of radius δ about the origin in $T_p(M)$ is a diffeomorphism.

Lemma 1 *The geodesics $exp(tX), 0 \le t \le t_1 \le 1$ minimize the length of any differentiable curve from one endpoint to the other. They are the unique geodesics of smallest length between their endpoints, up to equivalence.*

Idea of the proof: These geodesics are the images of radial lines in a polar coordinate system on $T_p(M)$ and an induced polar coordinate system on the image of the ball of radius δ. The length of any radial line is the length of the geodesic, and other lines between the endpoints are longer because they also move in the angular direction.

Details of the proof: Think of S^{n-1} as the vectors of unit length in $T_p(M)$. Define a map $S^{n-1} \times (0, \delta) \to T_p(M)$ by $s \times t \to t\ s$. This defines a polar coordinate system on a neighborhood of 0 in $T_p(M)$ and exp maps this to a polar coordinate system on an open neighborhood of p in M. The key fact we need about these coordinates is that curves on M formed by fixing s and moving t are perpendicular to curves on M formed by fixing t and moving s.

To show this, consider first the geodesic $\gamma(t) = exp(tX)$. The tangent vectors along this curve give a vector field along the curve. Let ∇ be the covariant derivative associated with the Riemannian metric and recall that the tangent field of a geodesic is parallel, i.e., $\nabla \gamma'(t) = 0$. Also $\frac{d}{dt} < \gamma'(t), \gamma'(t) >= 2 < \nabla \gamma'(t), \gamma'(t) >= 0$, so the length of these tangent vectors is constant.

Now let $s(u)$ be a path in $S^{n-1} \subset T_p(M)$ and consider the surface $\lambda(t, s) = exp(ts(u))$. If we fix t, we have a curve in s and we can take $\frac{\partial}{\partial s}$ to get a tangent field on the surface. Similarly by fixing s we have a curve in t and can form $\frac{\partial}{\partial t}$. Then

$$\frac{d}{dt} < \frac{\partial \lambda}{\partial t}, \frac{\partial \lambda}{\partial s} >=< \nabla_t \frac{\partial \lambda}{\partial t}, \frac{\partial \lambda}{\partial s} > + < \frac{\partial \lambda}{\partial t}, \nabla_t \frac{\partial \lambda}{\partial s} >$$

The first of these terms vanishes because each curve is a geodesic, so

$$\frac{d}{dt} < \frac{\partial \lambda}{\partial t}, \frac{\partial \lambda}{\partial s} >=< \frac{\partial \lambda}{\partial t}, \nabla_t \frac{\partial \lambda}{\partial s} >$$

Recall that

$$\nabla_X Y - \nabla_Y X = [X, Y]$$

In our case the partials commute and the bracket is zero, so

$$\frac{d}{dt} < \frac{\partial \lambda}{\partial t}, \frac{\partial \lambda}{\partial s} >=< \frac{\partial \lambda}{\partial t}, \nabla_s \frac{\partial \lambda}{\partial t} >= \frac{1}{2} \frac{d}{ds} < \frac{\partial \lambda}{\partial t}, \frac{\partial \lambda}{\partial t} >$$

But the length of $\frac{\partial \lambda}{\partial t}$ is constant in t and equals the length of $s(u)$, i.e., 1, when $t = 0$. So the expression on the right is zero and $< \frac{\partial \lambda}{\partial t}, \frac{\partial \lambda}{\partial s} >$ is constant. When $t = 0$, $\lambda(t, s) = exp(ts(u)) = p$ and the partial with respect to u is zero. Thus $< \frac{\partial \lambda}{\partial t}, \frac{\partial \lambda}{\partial s} >$ is always zero.

The proof of the lemma follows immediately. We sketch the idea with a picture and then provide an equivalent formula.

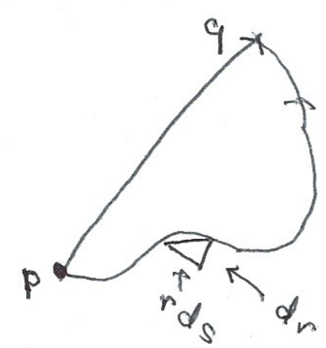

Figure 4.1: Radial and Spherical Components

The picture above shows a radial curve from p to q and another more general curve between these points. Each small segment of the general curve consists of two perpendicular motions, one in the radial direction and one in the perpendicular spherical direction. If we just add up the radial changes, we get the length of the radial geodesic, so the general curve is longer. If the general curve backtracks in the radial direction before reaching q, it is longer still because length is the sum of $|dr|$ rather than the sum of dr. If the radial curve leaves the image of the exponential map before returning to end at q, it is longer still because exp is a diffeomorphism on a ball of radius δ in $T_p(M)$ and the general curve would therefore cross the image of this the boundary sphere of this ball and have length greater than δ.

Symbolically, our general curve can be defined by introducing a parameter u and letting both t and s be functions of u, as in $\lambda(t(u)s(u))$. Recall that when t is fixed, the derivative of this expression was $\frac{\partial\lambda}{\partial s}$. When t also depends on u, the derivative is $\frac{\partial\lambda}{\partial t}\frac{dt}{du} + \frac{\partial\lambda}{\partial s}$. Since these vectors are orthogonal, Symbolically, this argument amounts to the formula

$$\int_0^{t_1} \Big|\Big|\frac{d}{du}\lambda(t(u)s(u))\Big|\Big|\ dt = \int_0^{t_1} \Big|\Big|\frac{\partial\lambda}{\partial t}\frac{dt}{du} + \frac{\partial\lambda}{\partial s}\Big|\Big| du \geq \int_0^{t_1} \Big|\Big|\frac{\partial\lambda}{\partial t}\frac{dt}{du}\Big|\Big|\ dt$$

This last expression is the length of just the radial pieces of the curve, which equals the length of the geodesic if there is no backtracking. This completes the proof of the lemma.

Lemma 2 *If $p \in M$, there is an open neighborhood of p in which any two points can be jointed by a unique geodesic of smallest length.*

Proof: Apply the inverse function theorem again, but this time to the full map

$$\{\ (q, Y)\ |q \in \mathcal{U}, Y \in T_q(M), ||Y|| < \delta\} \to M \times M$$

given by $(q, Y) \to q \times \gamma_{q,Y}(1)$. Here $\mathcal{U}$ is an open neighborhood of p, and the norm of Y is computed using the Riemannian metric. The derivative of this map at $p \times 0$ is again an isomorphism, so the inverse function theorem gives an open neighbborhood of $p \times 0$ on which the map is a diffeomorphism to its open image $\mathcal{W}$ in $M \times M$.

In particular, if we fix q, then the map is a diffeomorphism to the slice of this image in $M \times M$ with first element q, which is an open set in M. It follows from our previous analysis that for each fixed q, the geodesic $exp_{p,Y}(t)$ minimizes distance from q to its endpoint and is the unique such geodesic, whenever the ending t is less than or equal to one.

Let $\mathcal{U}_\epsilon$ be the ball of points in M whose distance from p is less than ϵ. This $\mathcal{U}$ is the set of endpoints of geodesics $exp(tX)$ for $t < \epsilon$. Select ϵ small enough that $\mathcal{U}_\epsilon \times \mathcal{U}_\epsilon$ is an open neighborhood of the image of $(p, 0)$ inside $\mathcal{W}$. Our map is onto this set. So whenever q and r are in $\mathcal{U}_\epsilon$, there is a geodesic from q to r which minimizes distance between these points, and that geodesic is unique.

This proves the lemma.

To complete the proof of the main theorem of this section, we must still show that if ϵ is small enough, and if care was used in the previous constructions, then the geodesic from q to r whose existence is guaranteed by lemma 2 is entirely inside $\mathcal{U}_\epsilon$.

Before giving the details of this final argument, we sketch the idea. Recall that coordinates can be chosen near a point p making the Riemannian metric $g_{ij}dx_idx_j$ equal δ_{ij} at

p. Said another way, any Riemannian metric is infinitesimally Euclidean in appropriate coordinates.

We can produce a neighborhood $\mathcal{U}_\epsilon$ which is close to Euclidean. So the geodesics from q to r which interest us essentially look like those in the following picture.

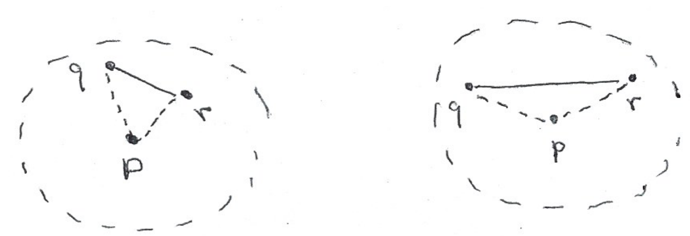

Figure 4.2: Geodesics from q to r

Notice that these geodesics, which are straight lines in Euclidean geometry, are furthest from p at the endpoints, and closer to p between these points. Since our $\mathcal{U}_\epsilon$ is a ball of radius ϵ as measured in the Riemannian metric, these geodesics are entirely within the ball. We will show that when we choose coordinates making $g_{ij}(p) = \delta_{ij}$, then in a sufficiently small ϵ ball the distance from a point on the geodesic to p in convex and thus furthest from p at one or the other endpoint. This will complete the proof.

Select an orthonormal basis for $T_p(X)$ and consider geodesically normal coordinates near p on M generated by this basis. Thus if $X = \sum x_i e_i$, we assign $(x_1, \ldots, x_n)$ as coordinates of $exp(X)$. Notice that the length of the geodesic $\gamma_{p,X}(t), 0 \leq t \leq 1$ is the integral of the length of the tangent vector to the curve between 0 and 1. These tangent vectors are parallel along the curve and all have the same length, namely $||X||$, so this distance is $||X|| = \sqrt{\sum x_i^2}$.

Next consider the Christoffel symbols Γ_{ij}^k. We claim they all vanish at the origin. Indeed $\frac{\partial}{\partial x_i}$ is the tangent to the geodesic $exp(te_i)$ and this vector field is parallel along the geodesic, so $\nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_i} = 0$ along this entire geodesic, and thus certainly at the origin.

Very temporarily, we write ∂_i for $\frac{\partial}{\partial x_i}$. Then $\nabla_{\partial_i} \partial_i = \nabla_{\partial_j} \partial_j = \nabla_{\partial_i + \partial_j}(\partial_i + \partial_j) = 0$ because all are derivatives along geodesics of their tangent fields. At the origin, we have $\nabla_{\partial_i} \partial_j + \nabla_{\partial_j} \partial_i = 0$. But $\nabla_X Y - \nabla_Y X = [X, Y]$ and $[\partial_i, \partial_j] = 0$. We conclude that at the origin, $\nabla_{\partial_i} \partial_j = 0$, but this expression equals $\sum_k \Gamma_{ij}^k \partial_k$. So all $\Gamma_{ij}^k = 0$ at the origin.

Recall that we have an ϵ ball about p, and any q and r in this ball can be joined by a unique geodesic of smallest length, $\gamma(t)$. But we do not know that this geodesic stays in the ball as it moves between its endpoints. That is what we want to prove. We will actually prove that the second derivative of the distance from p to $\gamma(t)$ squared is positive. So this distance squared is convex up, and the largest value that this function takes must be at an endpoint (why?), where the distance is at most ϵ. So the distance is always less than ϵ and the geodesic remains in our ball, as desired.

Here, then, is the calculation. Write $\gamma(t)$ in coordinates as $(x_1(t), \ldots, x_n(t))$.

$$\frac{d^2}{dt^2} d(p, \gamma(t))^2 = \frac{d^2}{dt^2} \sum x_i(t)^2 = 2\frac{d}{dt} \sum x_i(t) \frac{dx_i}{dt} =$$

$$2 \sum \left(\frac{dx_i}{dt} \right)^2 + 2 \sum x_i(t) \frac{d^2 x_i}{dt^2}$$

However, γ is a geodesic, so it satisfies

$$\frac{d^2 x_i}{dt^2} + \sum \Gamma^i jk \frac{dx_j}{dt} \frac{dx_k}{dt} = 0$$

and therefore this expression equals

$$2 \sum \left(\frac{dx_i}{dt} \right)^2 - 2 \sum x_i(t) \left(\sum \Gamma^i jk \frac{dx_j}{dt} \frac{dx_k}{dt} \right)$$

Recall that $\Gamma^i_{jk} = 0$ at the origin. We will show in the "cleanup" phase of the proof that we can guarantee

$$\left| \sum x_i \Gamma^i_{jk} \xi^j \xi^k \right| < \frac{1}{2} \sum (\xi^k)^2$$

by choosing $\sum (x_k)^2 < 2\eta$ for an appropriate $\eta > 0$. Putting this into the above calculations, we obtain

$$\frac{d^2}{dt^2} d(p, \gamma(t))^2 = 2 \left(\sum \left(\frac{dx_i}{dt} \right)^2 - \sum x_i \left(\sum \Gamma^i_{jk} \frac{dx_j}{dt} \frac{dx_k}{dt} \right) \right) \geq 2 \left(\sum \left(\frac{dx_i}{dt} \right)^2 - \frac{1}{2} \sum \left(\frac{dx_i}{dt} \right)^2 \right) \geq 0$$

and the proof is complete.

Here's the cleanup phase! Once we bound $||x||$, the terms $x_i \Gamma^i_{jk}$ can be made as small as we like by sticking close to p. Say each such term is at most K. Also $|\xi^j \xi^k| \leq max(\xi^i)^2 \leq \sum (\xi^i)^2$, so $\left| \sum x_i \Gamma^i_{jk} \xi^j \xi^k \right|$ is at most

$$(\text{number of terms}) \times K \times \sum (\xi^i)^2$$

and we can make the coefficient at most $\frac{1}{2}$ by sticking close enough to p.

But there is one more complication. We are working in a ball $\mathcal{U}_\epsilon$ centered at p. We want to prove that the geodesic γ, proved to exist earlier, stays in this ball. To do that, we differentiated using a coordinate system for the ball. But how can we do that if our γ might leave the ball?

Notice that γ is the shortest possible path from q to r, and q and r are in $\mathcal{U}_\epsilon$. One possible path from q to r is the geodesic from q back to p, followed by the geodesic from p to r. The maximal length of this path is 2ϵ. So we will be in fine shape if our calculations work for paths of length at most 2ϵ.

Look, then, at the entire argument of this section. The argument can be rearranged as follows. First, consider the following map.

$$\{\ (q, Y)\ | q \in \mathcal{U}, Y \in T_q(M), ||Y|| < \delta\} \to M \times M$$

By the existence theorem for differential equations, this map will be defined on some open neighborhood $\mathcal{U}$ of p and for some positive δ. Then by the inverse function theorem, it will be a diffeomorphism on a smaller $\mathcal{U}$ and smaller δ.

Next choose orthonormal coordinates for $T_p(M)$ and choose a number, which we will call 2ϵ, such that geodesic normal coordinates coming from the exponential map exist for geodesics of length 2ϵ. In particular, each radial geodesic is the unique shortest geodesic from p to its endpoint for lengths up to 2ϵ. Choose this 2ϵ small enough that the ball of radius 2ϵ is inside the smaller $\mathcal{U}$ of the previous paragraph.

Then look at the required estimate for the Γ^i_{jk} and shrink ϵ if necessary so this estimate holds in the ball of radius 2ϵ.

Then shrink ϵ even further so $\mathcal{U}_\epsilon \times \mathcal{U}_\epsilon \subset M \times M$ is in the image of the diffeomorphism of the first paragraph of this summary.

Notice now that any two points q, r in $\mathcal{U}_\epsilon$ can be joined by a geodesic starting at q and minimal and with length at most 2ϵ. Hence our calculation applies and shows that such a geodesic is entirely inside $\mathcal{U}_\epsilon$. Whew. QED.

4.4 Good Coverings, Part 2

We now want to prove that any geodesically convex set is diffeomorphic to R^n. The geodesically convex sets produced by the previous section are balls of radius ϵ about p and the result is obvious for them. But we want to apply the result to intersections of such balls, and these no longer need be balls. However, if $p \in \mathcal{U}$, then every point in $\mathcal{U}$ can be connected to p by a unique geodesic, so the sets are "star-shaped." But the picture of a star-shaped open set on the next page shows that it is not obviously diffeomorphic to R^n.

Theorem 12 *Every geodesically star-shaped open set is diffeomorphic to R^n.*

Figure 4.3: Star Shaped

Remark: An open set is *geodesically star-shaped* with respect to a point p if every point in the open set can be connected to p by a geodesic in $\mathcal{U}$ which is the unique shortest geodesic joining the two points.

In this case, we can consider the exponential map from $T_p(M)$ to $\mathcal{U}$ and discover that $\mathcal{U}$ is diffeomorphic to its preimage in $T_p(M)$. So without loss of generality, we may suppose that M is R^n and $\mathcal{U}$ is star-shaped with respect to Euclidean geometry, i.e., every point can be connected to p by a straight line in $\mathcal{U}$. So it suffices to prove

Theorem 13 *Let $\mathcal{U}$ be an open set in R^n, $0 \in \mathcal{U}$, such that $\mathcal{U}$ is star-shaped with respect to 0. Then $\mathcal{U}$ is diffeomorphic to R^n.*

Remark: Before giving details, we sketch the idea. Typically, star shaped regions are proved contractible by pulling the set back to the center along radial lines at constant speeds. The previous picture shows that this cannot work in all cases, since adjacent radial lines have different lengths and thus will be assigned different speeds. So instead, we vary the speed at which points on a given radial line move during the homotopy. We do this by finding a function φ on $\mathcal{U}$ so that $\varphi(q)$ gives the speed of a point on a radial line as it passes near q during the homotopy.

Remark: The following proof is from the book "Calcul Differentiel" by Gonnord and Tonel, 1998. The proof is a translation by Erwann Aubry. See https://mathoverflow.net/questions/4468/what-are-the-open-subsets-of-mathbbrn-that-are-diffeomorphic-to-mathbb/212595#212595.

Lemma 3 *There is a C^∞ function φ on R^n which is positive on $\mathcal{U}$ and zero off $\mathcal{U}$.*

Proof: Our arguments on partitions of unity easily give this result, but we sketch the argument again.

Start by finding a sequence $K_1 \subset \mathcal{U}_1 \subset K_2 \subset \mathcal{U}_2 \subset \ldots$ with union $\mathcal{U}$, where the K_i are compact and the $\mathcal{U}_i$ are open.

Consider the set of open balls $B_i \subset C_i$ with the same rational center and rational radii $b_i < c_i$ and such that the closure of C_i is in $\mathcal{U}$. These B_i form a countable basis for the topology of $\mathcal{U}$. Moreover, each B_i supports a C^∞ bump function on all of R^n which is identically one on B_i and vanishes off C_i.

Cover K_1 by a finite number of B_i whose corresponding C_i are in $\mathcal{U}_1$. Cover K_2 by a finite number of B_i whose corresponding C_i are in $\mathcal{U}_2$. Cover $K_3 - \mathcal{U}_1$ by a finite number of B_i whose corresponding C_i are in $\mathcal{U}_3 - K_1$. Continue.

In the end we have a countable number of B_i and corresponding bump functions such that each K_i is in the support of only finitely many functions and yet the sum of all these functions is nonzero on all of $\mathcal{U}$ and certainly zero off this set.

QED.

Next we introduce the central idea of the proof. Define $f : \mathcal{U} \to R^n$ by

$$f(x) = \left[1 + \left(\int_0^{||x||} \frac{dt}{\varphi(t\frac{x}{||x||})}\right)^2\right] \cdot x$$

This map leaves the origin fixed and sends a ray through x to another ray through x. The key point is that it maps each maximal ray in $\mathcal{U}$ to a full ray from 0 to ∞. Why? When φ is large, the integral is small, but the extra "1" guarantees that the output ray moves anyway. When φ is small, the integral is large, but if the input ray exists at both endpoints, the output ray still only moves a finite amount. The key point is that if the input ray ends at a boundary point of $\mathcal{U}$, then φ and all of its derivatives vanish at this boundary point, so just before the boundary, φ is very small and almost flat. This forces the integral to diverge, and thus the output ray goes to infinity.

Here are the details. We might as well study just one ray, say the ray through the real number x. Then

$$f(x) = \left[1 + \left(\int_0^x \frac{dt}{\varphi(t)}\right)^2\right] \cdot x$$

Even the final x is irrelevant and only the parameter inside the square brackets matters. The only question is whether this expression goes to infinity if $x = A$ is a boundary point of $\mathcal{U}$. But we can apply the mean value theorem to φ at A to obtain $\varphi(t) - \varphi(A) = \varphi'(\xi)(t - A)$. Since $\varphi(A) = 0$, we get $\varphi(t) = \varphi'(\xi)(t - A)$. We know that $\varphi(t) > 0$ and $t < A$, so $\varphi'(\xi)$ is negative.

We also know that $\varphi'(A) = 0$, so $\varphi'(\xi)$ must be bounded in absolute value near A, say by a positive constant c. So $\varphi(t) < c|t - A|$ and near A the integral is bounded below by

$$\int_{A-\delta}^{A} \frac{dt}{c|t - A|)}$$

and this blows up.

To prove that the map is a diffeomorphism, it is enough to prove that it has a local C^∞ inverse, and this holds if the Jacobian of the map is non-zero. Since our map preserves the spherical coordinate of rays we need only look at the radial component. But the derivative of $f(x) = \left[1 + \left(\int_0^x \frac{dt}{\varphi(t)}\right)^2\right] \cdot x$ with respect to x is always positive. QED.

4.5 Cohomology Groups of Compact M Are Finite Dimensional

Theorem 14 *Let M be a compact C^∞ manifold. Then each $H^k(M)$ is finite dimensional.*

Proof: This theorem is true more generally for any M with a finite good cover. The idea of the proof is to inductively show that $\mathcal{U}_1 \cup \ldots \cup \mathcal{U}_k$ has finite dimensional cohomology. Taking more and more unions, we ultimately reach M itself.

Thus the theorem to be proved by induction says that a manifold M, compact or not, has finite dimensional cohomology if it has a good cover by k open sets. We prove this by induction on the number k of such sets. If there is only one, then the set is contractible and we are done.

Consider now the induction step when $\mathcal{U} = \mathcal{U}_1 \cup \ldots \cup \mathcal{U}_k$ has finite dimensional cohomology and we form $\mathcal{U} \cup \mathcal{V}$. Note that $\mathcal{U} \cap \mathcal{V}$ satisfies the induction step because it is covered by the geodesically convex open sets $\mathcal{U}_i \cap \mathcal{V}$. So in the Mayer-Vietoris sequence, all terms have finite dimensional cohomology except $\mathcal{U} \cup \mathcal{V}$. Since these are isolated in the sequence, their cohomology groups must also be finite dimensional.

For example, suppose $A \stackrel{j}{\leftarrow} B \stackrel{i}{\leftarrow} C$ is exact and A and C are finite dimensional. Then the image of i is finite dimensional in B and equal to the kernel of j. So B modulo this kernel injects into A and must be finite dimensional. Thus B itself must be finite dimensional. QED.

4.6 Products in deRham Cohomology

Recall that we can compute the wedge product $\omega \wedge \lambda$ of an i-form and a j-form. It is easy to see that $\lambda \wedge \omega = (-1)^{i\ j} \omega \wedge \lambda$ and $d(\omega \wedge \lambda) = (d\omega) \wedge \lambda + (-1)^i \omega \wedge (d\lambda)$. It immediately follows that the wedge product induces a product in cohomology:

$$H^i(M) \times H^j(M) \xrightarrow{\triangle} H^{i+j}(M)$$

4.7 The Kunneth Formula

Suppose M and N are C^∞ manifolds. The projection maps $\pi_M : M \times N \to M$ and $\pi_N : M \times N \to N$ induce cohomology mapw $\pi_M^\star : H^i(M) \to H^i(M \times N)$ and $\pi_N^\star : H^j(N) \to H^j(M \times N)$. Combining these maps with the wedge product then produces a map $\sum_{i+j=k} H^i(M) \otimes H^j(N) \to H^k(M \times N)$.

Theorem 15 (Kunneth) *If one of M and N is compact, then the map below is an isomorphism. More generally, it is an isomorphism if one of M and N has a finite good cover.*

$$\sum_{i+j=k} H^i(M) \otimes H^j(N) \to H^k(M \times N)$$

Proof: As in the previous proof, we assume that M is covered by a finite number of contractible open sets with the property that the intersection of some of these sets is always contractible or empty. We work by induction on the number of open sets in the covering. If we have just one set, then $\mathcal{U} \times N$ can be deformed to N, so both sides have the same cohomology, which is implied by the Kunneth formula when $H^0(\mathcal{U}) = R$ and $H^k(\mathcal{U}) = 0$ for $k \geq 1$.

Tensoring an exact sequence of vector spaces over R with a fixed real vector space preserves exactness. This essentially gives the elaborate diagram on the next page, where $\mathcal{U}$ is the union of k open sets of the good covering and $\mathcal{V}$ is one additional open set of the covering.

We must prove that this diagram commutes. The only question is in the square at extreme left because it involves the D operator used to define the long exact sequence for the Mayer-Vietoris sequence.

After the diagram commutes, all vertical arrows are isomorphisms by induction except those for $H^i(\mathcal{U} \cup \mathcal{V}) \otimes H^j(N)$, so these maps are also isomorphisms by the 5-lemma. This completes the proof of the induction step. But we must fill in these two details.

We do this on the page following the elaborate diagram.

$$
\begin{tikzcd}
\sum H^{i+1}(\mathcal{U}\cup\mathcal{V})\otimes H^j(N) \arrow[d, "\wedge"] & \sum H^i(\mathcal{U}\cap\mathcal{V})\otimes H^j(N) \arrow[l] \arrow[d, "\wedge"] & \begin{array}{c}\sum H^i(\mathcal{U})\otimes H^j(N)\\ \oplus \\ \sum H^i(\mathcal{V})\otimes H^j(N)\end{array} \arrow[l] \arrow[d, "\wedge"] & \sum H^{i+1}(\mathcal{U}\cup\mathcal{V})\otimes H^j(N) \arrow[l] \arrow[d, "\wedge"] \\
H^{k+1}(\mathcal{U}\cup\mathcal{V}\times N) & H^k((\mathcal{U}\cap\mathcal{V}\times N) \arrow[l] & \begin{array}{c}H^k(\mathcal{U}\times N)\\ \oplus \\ H^k(\mathcal{V}\times N)\end{array} \arrow[l] & H^k(\mathcal{U}\cup\mathcal{V}\times N) \arrow[l]
\end{tikzcd}
$$

Lemma 4 (The 5 Lemma) *Suppose we have a commutative diagram with exact horizontal sequences as below, and suppose f, g, i, j are isomorphisms. Then h is also an isomorphism.*

$$
\begin{tikzcd}
A \arrow[d, "f"] & B \arrow[l, "\varphi_B"'] \arrow[d, "g"] & C \arrow[l, "\varphi_C"'] \arrow[d, "h"] & D \arrow[l, "\varphi_D"'] \arrow[d, "i"] & E \arrow[l, "\varphi_E"'] \arrow[d, "j"] \\
A' & B' \arrow[l, "\varphi_{B'}"'] & C' \arrow[l, "\varphi_{C'}"'] & D' \arrow[l, "\varphi_{D'}"'] & E' \arrow[l, "\varphi_{E'}"']
\end{tikzcd}
$$

Proof: Suppose $h(c) = 0$. Then $\varphi_{C'}h(c) = 0$, so $g\varphi_C(c) = 0$. Since g is an isomorphism, $\varphi_C(c)$ is zero, so $c = \varphi_D d$ for $d \in D$. Then $\varphi_{D'}i(d) = 0$, so $i(d)$ comes from $e' \in E'$. By isomorphism of the rightmost down arrow, there exists $e \in E$ with $j(E) = e'$. So $\varphi_{E'}j(e) = i(d) = i\varphi_E e$. Since i is an isomorphism, $\varphi_E(e) = d$. So $c = \varphi_D(d) = \varphi_D\varphi_E(e) = 0$.

Suppose $c' \in C'$. Then $\varphi_{C'}c' = g(b)$ because g is an isomorphism. But $\varphi_{B'}g(b) = \varphi_{B'}\varphi_{C'}c' = 0$, so $f\varphi_B(b) = 0$. But f is an isomorphism, so $\varphi_B(b) = 0$, and hence $b = \varphi_C(c)$. Then $\varphi_{C'}h(c) = \varphi_{C'}(c')$, so $c' - h(c) = \varphi_{D'}(d')$ for some $d' \in D'$. Since i is an isomorphism, $d' = i(d)$ for $d \in D$. But then $h\varphi_D(d) = c' - h(c)$ and so $h(\varphi_D(d) + c) = c'$. Hence h is onto. QED.

Remark: Finally we prove the left square on the previous large diagram is commutative. We start with an element of $\sum H^i(\mathcal{U} \cap \mathcal{V}) \otimes H^j(N)$ in the upper right corner. We can deal with each element of this sum separately, so ignore the sum sign. A representative of the element looks like $\omega \otimes \tau$. This element maps left to an element of $H^{i+1}(\mathcal{U} \cup \mathcal{V}) \otimes H^j(N)$ via the D map, and a representative of the image is $d(\varphi_{\mathcal{V}}\omega) \otimes \tau$ on $\mathcal{U}$ and $d(-\varphi_{\mathcal{U}}\omega) \otimes \tau$ on $\mathcal{V}$. We then map downward, getting $d(\varphi_{\mathcal{V}}\omega) \wedge \tau$ on $\mathcal{U}$ and $d(-\varphi_{\mathcal{U}}\omega) \wedge \tau$ on $\mathcal{V}$.

If we go around the square the other way, then we take the wedge first and then compute D. This gives $d\left(\varphi_{\mathcal{V}}(\omega \wedge \tau)\right)$ on $\mathcal{U}$ and $d\left(-\varphi_{\mathcal{U}}(\omega \wedge \tau)\right)$ on $\mathcal{V}$.

The two expressions are equal because $d(\omega \wedge \tau) = (d\omega) \wedge \tau + (-1)^{\deg \omega}\omega \wedge d\tau$ and in our case $d\tau = 0$.

However, an alert reader will notice that $\varphi_{\mathcal{V}}$ and $\varphi_{\mathcal{U}}$ came from a partition of unity on $\mathcal{U} \cup \mathcal{V}$ in the first case, and a partition of unity on $(\mathcal{U} \times N) \cup (\mathcal{V} \times N)$ in the second case.

We could form a partition of unity for $(\mathcal{U} \cup \mathcal{V}) \times N$ by starting with a partition of unity τ_i for $(\mathcal{U} \cup \mathcal{V})$ with compact supports, each support in either $\mathcal{U}$ or $\mathcal{V}$, and finding a similar partition of unity σ_j for N. Then the set of all products $\tau_i\sigma_j$ is a partition of unity for the product of the two spaces. Then we'd find a partition of unity with just two elements, one for $\mathcal{U} \times N$ and one for $\mathcal{V} \times N$, in the standard way. Namely, let the first function be the sum of all $\tau_i\sigma_j$ with the support of τ_i inside $\mathcal{U}$, and let the second function be the sum of all the remaining products. But then the first product is $\varphi_{\mathcal{U}}$ and the second is $\varphi_{\mathcal{V}}$, neither dependent on N, and so identifying the partitions of unity for $\mathcal{U} \cup \mathcal{V}$ and for $(\mathcal{U} \times \mathcal{V}) \times N$ is justified. QED.

Example 1: The standard torus is $S^1 \times S^1$, so the Kunneth Formula tells us that

$$H^0(\text{torus}) = R$$

$$H^1(\text{torus}) = R \oplus R$$

$$H^2(\text{torus}) = R$$

Example 2: The cohomology groups for $S^3 \times S^5 \times S^8$ are the following: $H^k = 0$ except for $k = 0, 3, 5, 8, 11, 13, 16$. All of these particular groups are R except H^8, which is $R \oplus R$.

4.8 Poincare Duality

Suppose M is an oriented compact C^∞ manifold of dimension n. Recall that integration over M induces a map

$$H^n(M) \xrightarrow{\int_M \omega} R$$

Suppose $\omega \in H^k(M)$ and $\tau \in H^{n-k}(M)$. Then we can form $\omega \wedge \tau$ and apply the previous map to get a number

$$\int_M \omega \wedge \tau$$

Theorem 16 (Poincare Duality) *If M is an oriented, compact, n-dimensional C^∞ manifold, the map*

$$H^k(M) \times H^{n-k}(M) \xrightarrow{\int_M \omega \wedge \tau} R$$

is nondegenerate. Hence if the cohomology class represented by ω is not zero, we can find a class τ such that the above integral is not zero.

Corollary 1 *If M is an oriented, compact, n-dimensional, C^∞ manifold,*

$$\dim H^k(M) = \dim H^{n-k}(M)$$

Proof: The map defined above induces a map $H^k(M) \to \left(H^{n-k}(M)\right)^\star$ and the duality theorem asserts that this map is one-to-one. Since these cohomology groups are finite dimensional, each has the same dimension as its dual. So

$$\dim H^k(M) \leq \dim \left(H^{n-k}(M)\right)^\star = \dim H^{n-k}(M)$$

For the same reason, $\dim H^{n-k}(M) \leq \dim H^k(M)$ and the result follows.

Corollary 2 *The induced maps below are isomorphisms:*

$$H^k(M) \to \left(H^{n-k}(M)\right)^\star$$

$$H^{n-k}(M) \to \left(H^k(M)\right)^\star$$

Proof: This follows from the previous argument.

Remark Although $H^k(M)$ and $H^{n-k}(M)$ have the same dimension and thus as isomorphic, there is no canonical isomorphism between them. So it is better to think of each as canonically isomorphic to the dual of the other.

Remark: We will prove this profound result using our standard argument involving good coverings and the Mayer-Vietoris theorem, but there is a catch. The theorem is false for non-compact manifolds, and the intermediate manifolds in the Mayer-Vietoris argument are non-compact.

So we have to generalize the theorem to cover certain non-compact manifolds. This requires introducing a new version of the deRham cohomology groups.

Definition 7 *A differential form ω has* compact support *if the closure of the set where it is non-zero is compact.*

Remark: Notice that when ω has compact support, so does $d\omega$. This allows us to define the deRham cohomology groups with compact support, $H_c^k(M)$ in the standard way, but using only forms with compact support. Notice that $H_c^k(M) = H^k(M)$ if M is compact.

Notice that $H^i(M) \otimes H_c^j(M) \to H_c^{i+j}(M)$ is well-defined. Indeed if $\omega \in H^i(M)$ is an arbitrary i form, and $\tau \in H_c^j(M)$ is a form with compact support, then $\omega \wedge \tau$ has compact support.

Notice also that if M is oriented but possibly non-compact,

$$H_c^n(M) \xrightarrow{\int_m \omega} R$$

is well-defined, because we can integrate a form with compact support over any set.

Theorem 17 (Generalized Poincare Duality) *If M is an oriented, n-dimensional C^∞ manifold with a finite good covering, then the map*

$$H^k(M) \times H_c^{n-k}(M) \xrightarrow{\int_M \omega\wedge\tau} R$$

is nondegenerate. Hence if the cohomology class represented by ω is not zero, we can find a class τ such that the above integral is not zero, and if τ represents a non-zero element, we can find ω such that the above integral is not zero.

Remark We will prove this by generalizing the Mayer-Vietoris sequence to compact cohomology, and using the standard induction on the number of open sets in the good covering. To get started, we need to prove the result when M is a single good open set. This will require a lemma:

Lemma 5 $H^n_c(R^n) \xrightarrow{\int_m \omega} R$ *is an isomorphism, and all other* $H^k_c(R^n)$ *are zero.*

Remark: We postpone the proof to the following section, but notice that Poincare duality is than a corollary for one good open set. Indeed, the only nonzero $H^k(R^n)$ is $H^0(R^n)$, the set of constant functions on R^n, and $H^0(R^n) \times H^n_c(R^n) \to R$ is the map $f \times \omega \to \int_M f \times \omega$, which is clearly nondegenerate by the above results.

Next we need Mayer-Vietoris for compactly supported cohomology. Suppose $i : \mathcal{U} \to \mathcal{V}$ is an inclusion map. This map induces $\Lambda^k_c(\mathcal{U}) \to \Lambda^k_c(\mathcal{V})$ because if ω is a $k - form$ with compact support in $\mathcal{U}$, then it can be extended to the rest of $\mathcal{V}$ by zero.

If, then $\mathcal{U}$ and $\mathcal{V}$ are open sets, we get a sequence, unexpectedly going backward:

$$0 \to \Lambda^k_c(\mathcal{U} \cap \mathcal{V}) \xrightarrow{j_1 - j_2} \Lambda^k_c(\mathcal{U}) \oplus \Lambda^k_c(\mathcal{V}) \xrightarrow{i_1 + i_2} \Lambda^k_c(\mathcal{U} \cup \mathcal{V}) \to 0$$

This sequence is exact. Exactness at the left is trivial. The composition of the two maps at the center is zero. Conversely suppose ω is a form with compact support on $\mathcal{U}$ and τ is a form with compact support on $\mathcal{V}$ and when these forms are extended by zero then $\omega + \tau = 0$. Then ω is non-zero if and only if τ is non-zero, and both can only happen in $\mathcal{U} \cap \mathcal{V}$ and on this set one form is the negative of the other. This proves exactness in the middle.

Finally, we must prove that the map on the right is onto. Find a partition of unity $\varphi_{\mathcal{U}}$ and $\varphi_{\mathcal{V}}$. If ω has compact support on $\mathcal{U} \cap \mathcal{V}$, consider $\varphi_{\mathcal{U}}\omega$ and $\varphi_{\mathcal{V}}\omega$. We claim these have compact support in $\mathcal{U}$ and $\mathcal{V}$ respectively. Their sum is $(\varphi_{\mathcal{U}} + \varphi_{\mathcal{V}})\omega = \omega$, so the map on the right is onto.

Why does $\varphi_{\mathcal{U}}\omega$ have compact support in $\mathcal{U}$? Since the two φ are a partition of unity, they are defined on $\mathcal{U} \cup \mathcal{V}$, as is ω. The support of the product is contained in the union of the support of $\varphi_{\mathcal{U}}$ and the support of ω, and thus is in $\mathcal{U}$. Finally, this support is a closed subset of a compact set, being contained in the support of ω. But a closed subset of a compact set is compact.

It follows as earlier that there is an induced long exact sequence in cohomology with compact supports. On the next page, we place this sequence below the regular Mayer-Vietoris sequence. An equivalent way to write the diagram is to replace the vector spaces in the bottom sequence by their duals, and this can be an easier way to see what needs to be done. We also show that diagram.

$$
\begin{tikzcd}
H^{i+1}(\mathcal{U}\cup\mathcal{V}) & H^i(\mathcal{U}\cap\mathcal{V}) \arrow[l] & \begin{matrix} H^i(\mathcal{U}) \\ \oplus \\ H^i(\mathcal{V}) \end{matrix} \arrow[l] & H^i(\mathcal{U}\cup\mathcal{V}) \arrow[l] \\
\otimes & \otimes & \otimes & \otimes \\
H_c^{n-i-1}(\mathcal{U}\cup\mathcal{V}) \arrow[r] \arrow[d, "\int_{\mathcal{U}\cup\mathcal{V}}"] & H_c^{n-i}((\mathcal{U}\cap\mathcal{V}) \arrow[r] \arrow[d, "\int_{\mathcal{U}\cap\mathcal{V}}"] & \begin{matrix} H_c^{n-i}(\mathcal{U}) \\ \oplus \\ H_c^{n-i}(\mathcal{V}) \end{matrix} \arrow[r] \arrow[d, "\int_{\mathcal{U}} + \int_{\mathcal{V}}"] & H_c^{n-i}(\mathcal{U}\cup\mathcal{V}) \arrow[d, "\int_{\mathcal{U}\cup\mathcal{V}}"] \\
R & R & R & R
\end{tikzcd}
$$

$$
\begin{tikzcd}
H^{i+1}(\mathcal{U}\cup\mathcal{V}) \arrow[d] & H^i(\mathcal{U}\cap\mathcal{V}) \arrow[l] \arrow[d] & \begin{matrix} H^i(\mathcal{U}) \\ \oplus \\ H^i(\mathcal{V}) \end{matrix} \arrow[l, "j_1^\star - j_2^\star"'] \arrow[d] & H^i(\mathcal{U}\cup\mathcal{V}) \arrow[l, "i_1^\star + i_2^\star"'] \arrow[d] \\
\left(H_c^{n-i-1}(\mathcal{U}\cup\mathcal{V})\right)^\star & \left(H_c^{n-i}((\mathcal{U}\cap\mathcal{V})\right)^\star \arrow[l, "D^\star"'] & \begin{matrix} \left(H_c^{n-i}(\mathcal{U})\right)^\star \\ \oplus \\ \left(H_c^{n-i}(\mathcal{V})\right)^\star \end{matrix} \arrow[l, "j_1^\star - j_2^\star"'] & \left(H_c^{n-i}(\mathcal{U}\cup\mathcal{V})\right)^\star \arrow[l, "I_1^\star + i_2^\star"']
\end{tikzcd}
$$

When we earlier proved that $H^k(M)$ is finite dimensional for compact M. we actually proved more. Namely we proved by induction that if we have a finite cover by good sets, then all $H^k(\mathcal{U}_{i_1} \cup \ldots \cup \mathcal{U}_{i_k})$ are finite dimensional. Now that we have a Mayer-Vietoris sequence for compactly supported cohomology, the same proof shows that compactly supported cohomology groups are finite dimensional on manifolds with a finite good cover. This proof definitely uses our result about $H_c^k(R^n)$, which has been stated above but not yet proved.

Note that if $U \to V \to W$ is an exact sequence of finite dimensional vector spaces, then $U^\star \leftarrow V^\star \leftarrow W^\star$ is also exact. The proof is easy. So the bottom sequence in the second diagram is exact.

The proof of Poincare duality is now easy. We will prove that the vertical arrows in the second diagram are all isomorphisms, and that will complete the proof. We do this by induction in the number of open sets in a good cover of $\mathcal{U}$. If we have a single $\mathcal{U}$, it is diffeomorphic to R^n and the result follows from explicit calculations of cohomology of R^n. Otherwise we apply the 5-lemma to the diagram. Four vertical arrows are isomorphisms by induction, so the left vertical arrow is also an isomorphism.

But this still does not complete the argument because we need to show that the diagram is commutative. This is easy for the two squares on the right. For example, take the square on the right and start with ω on $\mathcal{U} \cup \mathcal{V}$. Mapping down gives a map from $H^{n-i}(\mathcal{U} \cup \mathcal{V})$ to R given by $\int_{\mathcal{U} \cup \mathcal{V}} \omega \wedge \tau$. On the other hand, mapping ω to the left gives $\omega|_{\mathcal{U}}$ and $\omega|_{\mathcal{V}}$. Mapping these down gives two maps which could be applied to the images of τ in $H^{n-i}(\mathcal{U})$ and in $H^{n-i}(\mathcal{V})$. By an earlier lemma, these images are $\varphi_{\mathcal{U}} \tau$ and $\varphi_{\mathcal{V}} \tau$. The sum of the two maps is then

$$\int_{\mathcal{U}} \omega|_{\mathcal{U}} \wedge \varphi_{\mathcal{U}} \tau + \int_{\mathcal{V}} \omega|\mathcal{V} \wedge \varphi_{\mathcal{V}} \tau$$

The first integrand vanishes off $\mathcal{U}$ and the second vanishes off $\mathcal{V}$, so we can perform both integrals over $\mathcal{U} \cup \mathcal{V}$ and this clearly gives

$$\int_{\mathcal{U} \cup \mathcal{V}} \omega \wedge (\varphi_{\mathcal{U}} + \varphi_{\mathcal{V}}) \tau = \int_{\mathcal{U} \cup \mathcal{V}} \omega \wedge \tau$$

We leave the even easier proof in the middle square to the reader.

Finally we tackle the left square of the bottom diagram. Start with ω on the top-right corner of the square. It maps down to a map from $H^{n-i}(\mathcal{U} \cap \mathcal{V})$ to R. If τ is an element of this group, the map assigns $\int_{\mathcal{U} \cap \mathcal{V}} \omega \wedge \tau$ to this element.

The same ω maps left by the D map, which writes $\omega = \varphi_{\mathcal{V}} \omega + \varphi_{\mathcal{U}} \omega$; the first extends to all of $\mathcal{U}$ and the second extends to all of $\mathcal{V}$; on the intersection, $d\omega = 0$, so $d\varphi_{\mathcal{V}} \omega = -d\varphi_{\mathcal{U}} \omega$. Define $D\omega$ as $d\varphi_{\mathcal{V}} \omega$ on $\mathcal{U}$ and as $-d\varphi_{\mathcal{U}} \omega$ on $\mathcal{V}$.

This drops down to a map defined on $H^{n-i-1}(\mathcal{U} \cup \mathcal{V})$. Let σ define an element of this group. Then applying the map gives

$$\int_{\mathcal{U} \cup \mathcal{V}} D\omega \wedge \sigma$$

On the bottom of this square, a map from dual spaces goes left because the map from cohomology groups goes right. So we need to suppose $D\sigma = \tau$ and prove that $\int_{\mathcal{U} \cap \mathcal{V}} \omega \wedge \tau = \int_{\mathcal{U} \cup \mathcal{V}} D\omega \wedge \sigma$.

Said another way, our task is to prove that

$$\int_{\mathcal{U} \cap \mathcal{V}} \omega \wedge D\sigma = \int_{\mathcal{U} \cup \mathcal{V}} D\omega \wedge \sigma$$

Thus we need to compute $D\sigma$ where σ has compact support in $\mathcal{U} \cup \mathcal{V}$. Recall that we write $\sigma = \varphi_{\mathcal{U}}\sigma + \varphi_{\mathcal{V}}\sigma$ where the left extends to all of $\mathcal{U}$ and the right extends to all of $\mathcal{V}$. This sum is in $\Lambda^k(\mathcal{U}) \oplus \Lambda^k(\mathcal{V})$. We map these to $d\varphi_{\mathcal{U}}\sigma$ and $d\varphi_{\mathcal{V}}\sigma$ and notice that the sum of these terms is $d\sigma$ and thus vanishes on the intersection. So the first term and the negative of the second term agree on $\mathcal{U} \cap \mathcal{V}$ and come from applying $j_1 - j_2$ to an element of cohomology for $\mathcal{U} \cup \mathcal{V}$. This element equals $d\varphi_{\mathcal{U}}\sigma$ on $\mathcal{U}$ and $d\varphi_{\mathcal{V}}\sigma$ on $\mathcal{V}$.

Since σ has compact support inside $\mathcal{U} \cap \mathcal{V}$, integrating either of these expressions over $\mathcal{U} \cup \mathcal{V}$ is the same as integrating over $\mathcal{U} \cap \mathcal{V}$. Since the two expressions agree on this intersection, we can pick one of them and just integrate that. So

$$\int_{\mathcal{U}\cup\mathcal{V}} \omega \wedge D\sigma = \int_{\mathcal{U}\cap\mathcal{V}} \omega \wedge d(\varphi_{\mathcal{V}}\sigma)$$

However, $d\sigma = 0$, so this last integral is just

$$\int_{\mathcal{U}\cap\mathcal{V}} \omega \wedge (d\varphi_{\mathcal{V}}) \wedge \sigma = (-1)^{\deg \omega} \int_{\mathcal{U}\cap\mathcal{V}} (d\varphi_{\mathcal{V}}) \wedge \omega \wedge \sigma$$

Now $d\omega = 0$, so $d(\varphi_{\mathcal{V}}\omega) = d(\varphi_{\mathcal{V}}) \wedge \omega + 0$, and we can rewrite the result as follows. Note that ω has compact support in $\mathcal{U} \cap \mathcal{V}$, so it makes no difference whether we integrate over $\mathcal{U} \cap \mathcal{V}$ or $\mathcal{U} \cup \mathcal{V}$

$$(-1)^{\deg \omega} \int_{\mathcal{U}\cup\mathcal{V}} (d\varphi_{\mathcal{V}}\omega) \wedge \sigma$$

On the other hand, $D\omega = d\varphi_{\mathcal{V}}\omega = -d\varphi_{\mathcal{V}}\omega$, so this integral is

$$(-1)^{\deg \omega} \int_{\mathcal{U}\cup\mathcal{V}} (D\omega) \wedge \sigma$$

as required; notice that we need only worry about $\mathcal{U} \cap \mathcal{V}$ because ω and $d\omega$ vanish off this set.

Remark: The only trouble with this last calculation is that we only obtained commutativity up to a sign. But this do not matter in applying the 5-lemma, as the reader can readily check.

Corollary 3 *Suppose M is a connected, compact, oriented, n-dimensional C^∞ manifold. Then the map*

$$\int_m \omega : H^n(M) \to R$$

is an isomorphism.

Remark: This corollary follows immediately from Poincare Duality. But our proof of duality isn't yet complete; the final details will be proved in the next section. Thus we are not allowed to use the corollary in that section.

4.9 $H_c^k(R^n)$

Notice that R^0 is a single point, so $H_c^0(R^0) = 0$ and all other $H_c^k(R^0) = 0$. We will now prove by induction on n that

Theorem 18 *For $n \geq 1$*

- $H_c^n(R^n) = R$
- $H_c^k(R^n) = 0$ *for all other* k

Proof: We will prove that there are natural isomorphisms $H_c^k(R^n \times R) \to H_c^{k-1}(R^n)$. The above result follows immediately.

Define a map $f : \Lambda^k(R^n \times R) \to \Lambda^{k-1}(R^n)$ as follows. A form on $R^n \times R$ is a sum of terms of the form $\omega(x_1, \ldots, x_n, t) dx_{i_1} \wedge \ldots \wedge dx_{i_k}$. If none of the dx_i are dt, let

$$\omega(x_1, \ldots, x_n, t) dx_{i_1} \wedge \ldots \wedge dx_{i_k} \to 0$$

If one of the dx_i, say the last, is dt, let

$$\omega(x_1, \ldots, x_n, t) dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}} \wedge dt \to \left(\int_{-\infty}^{\infty} \omega(x_1, \ldots, x_n, t)\ dt \right) dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}}$$

The integral exists because our forms have compact support. Clearly the forms in the image of f still have compact support.

Notice that $f(d\omega) = d(f(\omega))$. Indeed, if ω has no dt, then the only term that matters when computing $f(d\omega)$ is the term when we differentiate with respect to t and

$$f(d\omega) = f\left(\frac{\partial \omega}{\partial t} (-1)^k dx_{i_1} \wedge \ldots \wedge dx_{i_k} \wedge dt \right) = (-1)^k \left(\int_{-\infty}^{\infty} \frac{\partial \omega}{\partial t} dt \right) dx_{i_1} \wedge \ldots \wedge dx_{i_k}$$

Since ω has compact support, the integral vanishes and so $f(d\omega) = 0$. Trivially $d(f(\omega)) = 0$.

If ω contains a dt, then the only terms in $d\omega$ which matter when computing $f(d\omega)$ are those involving partials with respect to d_i because $dt \wedge dt = 0$. Then f replaces each of these partials with the integral of the partial with respect to t. As for $d(f(\omega))$, it first integrates ω with respect to t and then takes partials of these integrals. The integral is a definite integral with respect to t, so the coefficient no longer depends on t. All other terms involve and integral and a partial, which can be done in either order. So $f(d\omega) = d(f(\omega))$.

Define a second map $g : \Lambda^{k-1}(R^n) \to \Lambda^k(R^n \times R)$ as follows. Let $\varphi(t)$ be a C^∞ function with compact support and integral one, and form $\varphi(t)\ dt$. Define $g(\omega) = \omega \wedge \varphi(t)\ dt$. Clearly $\omega \wedge \varphi(t) dt$ has compact support. Note that $g(d\omega) = d(g(\omega)$. Indeed both $g(d(\omega))$ and $d(g(\omega))$ have terms generated by differentiation by x_j where $x_j \neq t$. The terms on both sides are the same because it doesn't matter if we wedge the answer with $\varphi(t)\ dt$ before or after differentiating. On the other hand, neither side has terms obtained by differentiating a coefficient with respect to t. That is because the original ω doesn't depend on t, and when we differentiate $\varphi(t)\ dt$ with respect to t, we get $dt \wedge dt = 0$.

It follows that f induces $f : H^k_c(R^n \times R) \to H^{k-1}_c(R^n)$ and g induces $g : H^{k-1}_c(R^n) \to H^k_c(R^n \times R)$. We have $f(g(\omega)) = f(\omega \wedge \varphi(t)\ dt) = \omega \left(\int \varphi(t)\ dt \right) = \omega$.

Unfortunately, $g \circ f$ is not the identity, but we will show that it is cohomologically equivalent to the identity. This is the only difficult step.

Define $K : \Lambda^k_c(R^n \times R) \to \Lambda^{k-1}_c(R^n \times R)$ as follows. As earlier, let t be the coordinate of the last R. If a term ω in a form has no dt, let $K(\omega) = 0$. If a term has a dt, then recall the function $\varphi(t)$ introduced in the definition of g above, and let

$$K(\omega) = \left(\int_{-\infty}^{t} \omega(x_1, \ldots, x_n, u)\ du - \int_{-\infty}^{\infty} \omega(x_1, \ldots, x_n, u)\ du \int_{-\infty}^{t} \varphi(u)\ du \right) dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}}$$

We claim that $dK - Kd = \mathrm{id} - g \circ f$ up to sign. It immediately follows that $g \circ f$ is the identity in cohomology, and thus that f and g are isomorphisms.

First, suppose that a term ω has no dt. Then

$$(dK - Kd)\omega = -(-1)^k K \frac{\partial \omega}{\partial t} dx_{i_1} \wedge \ldots \wedge dx_{i_k} \wedge dt =$$

$$(-1)^{k-1} \left(\int_{-\infty}^{t} \frac{\partial \omega}{\partial u} du - \int_{-\infty}^{\infty} \frac{\partial \omega}{\partial u}\ du \int_{-\infty}^{t} \varphi(u) du \right) dx_{i_1} \wedge \ldots \wedge dx_{i_k} =$$

$$(-1)^{k-1} \omega\ dx_{i_1} \wedge \ldots \wedge dx_{i_k} = (-1)^{k-1} (\mathrm{id} - g \circ f)(\omega)$$

Finally, suppose a term ω has a dt term. Then

$$(dK - Kd)\omega = d \left[\left(\int_{-\infty}^{t} \omega(x_1, \ldots, x_n, u)\ du - \int_{-\infty}^{\infty} \omega(x_1, \ldots, x_n, u)\ du \int_{-\infty}^{t} \varphi(u)\ du \right) dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}} \right]$$

$$-K \left(\sum_j \frac{\partial \omega}{\partial x_j} dx_j \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}} \wedge dt \right) =$$

$$\sum_j \left[\left(\int_{-\infty}^{t} \frac{\partial \omega}{\partial x_j}(x_1, \ldots, x_n, u)\ du - \int_{-\infty}^{\infty} \frac{\partial \omega}{\partial x_j}(x_1, \ldots, x_n, u)\ du \int_{-\infty}^{t} \varphi(u)\ du \right) dx_j \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}} \right] +$$

$$\left[\omega - \left(\int_{-\infty}^{\infty} \omega(x_1, \ldots, x_n, u)\ du \right) \varphi(t) \right] dt \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}} -$$

$$\sum_j \left[\left(\int_{-\infty}^{t} \frac{\partial \omega}{\partial x_j}(x_1, \ldots, x_n, u)\ du - \int_{-\infty}^{\infty} \frac{\partial \omega}{\partial x_j}(x_1, \ldots, x_n, u)\ du \int_{-\infty}^{t} \varphi(u)\ du \right) dx_j \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}} \right] =$$

$$(-1)^{k-1} \left[\omega - \left(\int_{-\infty}^{\infty} \omega(x_1, \ldots, x_n, u)\ du \right) \varphi(t) \right] dx_{i_1} \wedge \ldots \wedge dx_{i_{k-1}} \wedge dt = (-1)^{k-1} (\mathrm{id} - g \circ f)(\omega)$$

QED.

Chapter 5

The Lefshetz Fixed Point Theorem (1)

5.1 The Easy Lefshetz Fixed Point Theorem

At this point we are about half way to our ultimate goal. We understand the cohomological language used to capture the central ideas, but other hard work is required.

Solomon Lefshetz was born in Moscow, but his parents moved to Paris shortly afterward. He later received an engineering degree in Paris, and then emigrated in 1905 to the United States. In 1907 he lost both hands in an industrial accident, so he switched to mathematics. His first position was in Nebraska, and then he spent 11 years at Kansas. There he produced fundamental work in algebraic geometry and algebraic topology, and was recruited to Princeton in 2024, where he worked until retirement.

In the early 1900's, algebraic topology emerged as a separate field, with notable results like the Brouwer Fixed Point theorem and the theory of degrees of maps $f : S^n \to S^n$. During this period, Lefshetz found a beautiful generalization of both of these results. Suppose X is a nice space, say a finite simplicial complex, and $f : X \to X$ is a continuous map. Such an f induces vector space homomorphisms $f_\star : H_k(X, R) \to H_k(X, R)$, and the trace of each of these homomorphisms is a real number. Form the alternating sum

$$L(f) = \sum (-1)^k \text{trace}(f_\star : H_k \to H_k)$$

This number has become known as the *Lefshetz number of f*. Lefshetz proved:

Theorem 19 (Easy Lefshetz Fixed Point Theorem) *If X is a finite simplicial complex and $f : X \to X$ is a continuous map with $L(f) \neq 0$, then f has a fixed point.*

For example, suppose X is the closed $n-$dimensional ball. Then $H_0(X,R) = R$ and $H_k(X,R) = 0$ for $k > 0$: moreover, $f_\star : H_0(X,R) \to H_0(X,R)$ is the identity map. So for any f we have $L(f) = 1$, and f has a fixed point. Therefore Lefshetz' result generalizes the Brouwer Fixed Point theorem.

Suppose $X = S^n$. Then $f_\star : H_0 \to H_0$ is the identity and by definition $f_\star : H_n \to H_n$, which is a map $R \to R$, is multiplication by a number called the *degree* of f, or $deg(f)$. So $L(f) = 1+(-1)^n deg(f)$. It follows that f has a fixed point unless $\deg(f) = (-1)^{n-1}$. Since an antipodal map A has no fixed points, $deg(A) = (-1)^{n-1}$. In particular, the antipodal map is not homotopic to the identity if n is even. If n is odd, the antipodal map *is* homotopic to the identity via

$$h(t,x) = ((\cos t)x_0 + (\sin t)x_1, -(\sin t)x_0 + (\cos t)x_1,$$

$$(\cos t)x_2 + (\sin t)x_3, -(\sin t)x_2 + (\cos t)x_3, \ldots)$$

Suppose X is a connected Lie group. Note that the Lefshetz number is a homotopy invariant. The Lefshetz number of the identity map is just the Euler characteristic $\chi(X)$. If τ is a path starting at the identity and ending at another point, then left translation by τ is a homotopy from the identity to a map with no fixed points. Consequently the Lefshetz number of either map is zero. It follows that the Euler characteristic of a compact connected Lie group must be zero. For instance, the only compact 2-manifold which can be given a Lie group structure is the torus.

Outline of the Proof of the Easy Lefshetz Theorem: The easy version of the theorem is not difficult to prove. Put a metric on X, and assume $f : X \to X$ has no fixed points. Then there is a $\delta > 0$ such that every $x \in X$ moves by at least δ, that is, $d(x, f(x)) \geq \delta$.

By the simplicial approximation theorem, the map f is homotopic, via a homotopy which moves each point by less than $\delta/2$, to a simplicial map, provided we replace the simplicial complex by a sufficiently subdivided subcomplex. Hence we may assume that f is a simplicial map which moves every simplex. Note that f then induces maps $f_\star : C_k \to C_k$ which move every simplex and thus have trace zero.

Because each C_k is finite dimensional, we can form a Lefshetz number of the chain complex $C_0 \leftarrow C_1 \leftarrow C_2 \leftarrow \ldots$, using the obvious formula

$$\sum (-1)^k \text{trace}(C_k(X) \to C_k(X))$$

This number is zero since all traces are zero.

The proof of the easy Lefshetz theorem then reduces to a simple lemma: the Lefshetz number of the complex $f_\star : C_k \to C_k$ equals the Lefshetz number of the induced homology groups $f_\star : H_k \to H_k$.

5.2 The Hard Lefshetz Fixed Point Theorem

Suppose $f : [0,1] \to [0,1]$ is a continuous map. We can draw the graph of f in the plane, and superimpose the graph of $g(x) = x$. This gives the pictures below, for various f. In a movie he once made about the fixed point theorem, Lefshetz said that a substantial portion of his career was a consequence of studying this picture.

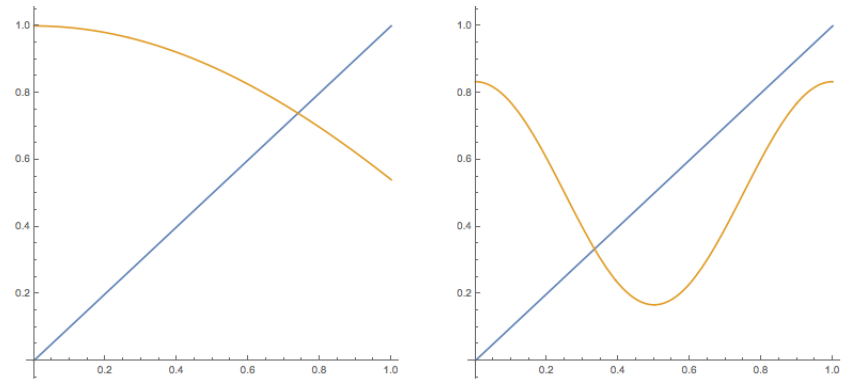

Figure 5.1: $f : [0,1] \to [0,1]$

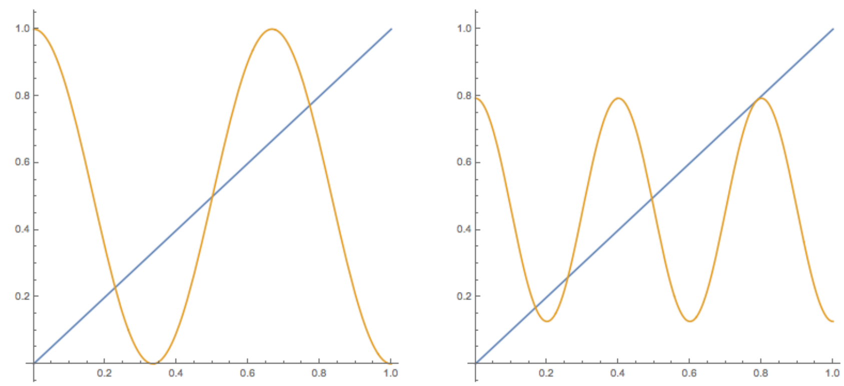

Figure 5.2: $f : [0,1] \to [0,1]$

Note that at a fixed point, $f(x) = x = g(x)$, so the fixed points are the points where the two graphs intersect. Each map above has at least one fixed point, and indeed according to the Brouwer Fixed Point theorem, each continuous map has a fixed point.

The picture suggests a simple proof of the Brouwer Fixed Point theorem in this case. If $f(0) = 0$ or $f(1) = 1$, then we certainly have a fixed point, so suppose neither result is true. Then the diagonal line divides the box $[0,1] \times [0,1]$ into two pieces and f starts in one of the pieces and ends in the other. Thus it must cross the diagonal at least once.

The first three pictures, and the previous rough proof, suggest that a more powerful result is true. Since the graph starts above the diagonal line and ends below the line, it must cross the line an odd number of times. Recall, incidentally, that the Lefshetz number of f is 1. Perhaps the Lefshetz number counts the number of fixed points and therefore crossings, except that a crossing from top to bottom counts positively and a crossing from bottom to top counts negatively. This turns out to be true and its generalization to all compact manifolds is the *Hard Lefshetz Theorem.*

But the final picture shows that we must be careful, because the graph of f can touch the line without crossing it. If we have two differentiable curves in the plane which meet at a point p, we say that they cross *transversally* at p if the tangent lines to the two curves at p have different slopes. If we have two differentiable curves in the plane, then after an arbitrarily small homotopy of one of the curves, all intersection points will be transversal. Since the Lefshetz number is a homotopy invariant, homotopies will not change its value. The theorem we are after says that the Lefshetz number counts fixed points, i.e., intersections of the graph of f with the diagonal line, provided we count with a sign which is determined by the way the tangents to the graphs cross when touching and provided all fixed points are transversal.

All of this generalizes. If $f : M \to M$ is a smooth map from a compact, oriented manifold to itself and the Lefshetz number of f is not zero, then f has a fixed point. Indeed, $L(f)$ counts the number of fixed points, provided we assign ± 1 to each fixed point in a manner to be described shortly, and provided all fixed points are transversal in a sense to be described. We will outline the proof here, and then fill in theoretical details in future chapters. The essential idea of the proof is to replace $[0, 1] \times [0, 1]$ by $M \times M$ and replace the graphs of $f(x)$ and $g(x)$ by $\{p \times f(p) \in M \times M\}$ and the diagonal $\{p \times p \in M \times M\}$. Each of these is a submanifold of $M \times M$ and these submanifolds induce cohomology classes in $H^\star(M \times M)$.

This works more generally. Suppose K and L are any two submanifolds of M which intersect transversally. Then $K \cap L$ is again a submanifold, so it induces a third cohomology class. The crucial theorem says that the class attached to this intersection is the cup product of the classes attached to K and L.

In the special case of the fixed point theorem, the intersection is just the finitely many fixed points of f, and the cohomology class of this intersection counts these fixed points with signs. And that is the content of the Lefshetz theorem.

We now describe this story in more detail.

Let M be a fixed compact, oriented, C^∞ manifold. Suppose K is also compact, oriented, and C^∞ and suppose $f : K \to M$ is a C^∞ map. If M has dimension n and K has dimension k, then f induces a map $f^\star : H^k(K) \leftarrow H^k(M)$. We earlier used the orientation of K to define an integral of k-forms on K and resulting map $R \xleftarrow{\int_K \omega} H^k(K)$. Putting these maps together, we have

$$R \xleftarrow{\int_K \omega} H^k(K) \xleftarrow{f^\star} H^k(M)$$

Thus we get an element of the dual space $\left(H^k(M)\right)^\star$ associated with the manifold K. By the Poincare duality theorem, this element is duel to an element of $H^{n-k}(M)$, called the cohomology class dual to K.

In particular, these considerations hold when K is a submanifold of M. Note that the initial map from $R \leftarrow H^k(M)$ is completely defined by K and the inclusion map f, but the dual element is only defined up to cohomology, and thus has many representatives. The crucial fact we will prove in the following sections is that one representative is an $n-k$ form which has support in an arbitrarily small open neighborhood of K.

Suppose now that we have two compact, oriented, C^∞ manifolds of dimensions k and l, where L is a submanifold of M and $f : K \to M$ is a C^∞ map. The image of f and L will intersect if $k+l$ is sufficiently large; and we would like to understand the generic situation. If $k + l < n$, then locally K defines a k dimensional surface in M and L defines an l dimensional surface and there is at least one extra dimension not in the direction of either surface. By pulling K in this direction, we can entirely remove the intersection. Thus if $k + l < n$, we do not expect that K and L will intersect. For instance, two lines in R^3 should not intersect generically.

But suppose that $k + l \geq n$. At an intersection point, the two surfaces may share some tangent directions. For instance, if two 2-dimensional surfaces intersect in R^3, we expect them to intersect along a curve, so their tangent spaces will share the line tangent to that curve. The two surfaces might be completely tangent at the intersection point, so that the sum of their tangent planes is two-dimensional; generically, however, we expect the surfaces to be as disjoint as possible, intersecting only in a curved line, and with tangent planes intersecting in a line.

If K and L meet at p, we say they meet *transversally* if $T_p(K)+T_p(L) = T_p(M)$, so that they fill out as many dimensions as possible. For this to happen, we must have $k+l \geq n$.

We will prove the following:

Theorem 20 (Thom Transversality Theorem) *If $f : K \to M$ is C^∞, and if $L \subset M$ is a submanifold, both compact, and if $k < n$ and $l < n$, then after an arbitrarily small homotopy of f, K is also a submanifold of M and the two submanifolds meet transversally.*

In this case, $K \cap L$ is also a compact submanifold of M, of dimension $n - k - l$.

Remark: Note that if $k + l < n$, then this theorem asserts that after an arbitrarily small homotopy, K and L do not intersect.

Finally, we will prove:

Theorem 21 (Main Theorem of Intersection Theory) *Suppose K, f, and L are as above, then after an arbitrarily small homotopy of f, K is also a submanifold and it intersects L transversally. Then if d_K, d_L, and $d_{K\cap L}$ are dual to the submanifolds of dimensions $n-k$, $n-l$, and $k+l-n$, we have*

$$d_K \wedge d_L = d_{K\cap L}$$

Remark: Note that $d_K \wedge d_L$ has degree $(n-k)+(n-l) = 2n-(k+l) = n-[(k+l)-n]$ which is the correct dimension for a form dual to a manifold of dimension $(k+l)-n$.

Note that $K \cap L$ is commutative in K and L, while $d_K \wedge d_L$ is commutative or anticommutative depending on the degrees of d_K and d_L. Hence this theorem is really true "up to sign."

5.3 Calculation of the Lefshetz Number

The previous section described difficult assertions about intersection theory in cohomology. But the calculations needed to obtain the Lefshetz theorem from them are fairly straightforward and we'll do that calculation now.

Assume the fundamental theorem of intersection theory. Then the number of such points, properly counted, is the wedge product of the class dual to the graph of f and the class dual to the diagonal. In this section, we will compute this wedge product and show that it equals the Lefshetz number of f.

Let $\{\omega_i\}$ and $\{\tau_j\}$ be dual bases for $H^\star(M)$, and let $\pi_1, \pi_2 : M \times M \to M$ be the obvious projections. Then $\{\pi_1^\star(\omega_i) \wedge \pi_2^\star(\tau_j)\}$ is a basis of $H^\star(M \times M)$ by the Kunnuth Fomula. The Poincare dual of the diagonal $[\Delta]$ can therefore be written

$$\sum \alpha_{ij} \pi_1^\star(\omega_i) \wedge \pi_2^\star(\tau_j)$$

We now compute the integral of $\pi_1^\star(\tau_k) \wedge \pi_2^\star(\omega_l)$ over Δ (the switch of order is deliberate). The map $i : M \to \Delta$ by $m \to m \times m$ is a diffeomorphism, and $i \circ \pi_1$ and $i \circ \pi_2$ are both the identity, so the integral of our element over Δ is the integral of $i^\star$ of the element over M, and thus $\int_M \tau_k \wedge \omega_l$. Thus it equals $(-1)^{|\tau_k||\omega_l|}\delta_{kl}$.

On the other hand, by definition of the Poincare dual of $[\Delta]$, we have

$$\int_\Delta \pi_1^\star(\tau_k)\times\pi_2^\star(\omega_l) = \int_{M\times M} \pi_1^\star(\tau_k)\wedge\pi_2^\star(\omega_l)\wedge\eta_\Delta = \sum_{ij} \int_{M\times M} \pi_1^\star(\tau_k)\wedge\pi_2^\star(\omega_l)\wedge\alpha_{ij}\pi_1^\star(\omega_i)\wedge\pi_2^\star(\tau_j)$$

$$= \sum_{ij} \int_{M\times M} \alpha_{ij}.(-1)^{|\omega_i||\omega_l|}\pi_1^\star(\tau_k \wedge \omega_i) \wedge \pi_2^\star(\omega_l \wedge \tau_j)$$

$$= \sum_{ij} \int_{M\times M} \alpha_{ij}.(-1)^{|\omega_i|(|\omega_l|+|\tau_k|)}\pi_1^\star(\omega_i \wedge \tau_k) \wedge \pi_2^\star(\omega_l \wedge \tau_j)$$

In this integral, the term $\pi_1^\star(\omega_i \wedge \tau_k)$ involves the coefficients of the first M in $M \times M$ and the remaining term involves coefficients of the second M in the product. Consequently we can integrate the first term over the first M and integrate the second term over the second M and multiply. The first integral is δ_{ik} and the second is δ_{lj} and when we sum over i and j, we obtain

$$\alpha_{kl}(-1)^{|\omega_k|(|\omega_l|+|\tau_k|)}$$

Comparing the two calculations, we conclude that $(-1)^{|\tau_k||\omega_l|}\delta_{kl} = \alpha_{kl}(-1)^{|\omega_k|(|\omega_l|+|\tau_k|)}$, and therefore that $\alpha_{kl} = (-1)^{|\omega_k|}\delta_{kl}$. So $d_\Delta = \sum_i (-1)^{|\omega_i|}\pi_1^\star(\omega_i) \wedge \pi_2^\star(\tau_i)$.

Proof, continued: We now repeat essentially the same calculation to obtain d_G.

As before, we compute the integral of $\pi_1^\star(\tau_k) \wedge \pi_2^\star(\omega_l)$ over G. The map $i : M \to \Delta$ by $m \to m \times f(m)$ is a diffeomorphism, and $i \circ \pi_1 = id$ and $i \circ \pi_2 = f$. Let $f^\star(\omega_i) = \sum \beta_{ij}\omega_j$ where this is a sum over j so $|\omega_i| = |\omega_j|$. Then

$$\int_G \pi_1^\star(\tau_k) \wedge \pi_2^\star(\omega_l) = \int_M i^\star\pi_1^\star(\tau_k) \wedge i^\star\pi_2^\star(\omega_l) = \int_M \tau_k \wedge f^\star(\omega_l) = \int_M \tau_k \wedge \sum_i \beta_{li}\omega_i$$

This equals

$$\int_M \sum_i (-1)^{|\tau_k||\omega_i|}\beta_{li}\omega_i \wedge \tau_k = (-1)^{|\tau_k||\omega_k|}\beta_{lk}$$

Write the Poincare dual of $[G]$ as $d_G = \sum \gamma_{ij}\pi_1^\star(\omega_i) \wedge \pi_2^\star(\tau_j)$. By definition of the Poincare dual of $[G]$, we have

$$\int_G \pi_1^\star(\tau_k)\times\pi_2^\star(\omega_l) = \int_{M\times M} \pi_1^\star(\tau_k)\wedge\pi_2^\star(\omega_l)\wedge d_G = \sum_{ij}\int_{M\wedge M} \pi_1^\star(\tau_k)\wedge\pi_2^\star(\omega_l)\wedge\gamma_{ij}\pi_1^\star(\omega_i)\wedge\pi_2^\star(\tau_j)$$

$$= \sum_{ij} \int_{M\times M} \gamma_{ij}.(-1)^{|\omega_i||\omega_l|}\pi_1^\star(\tau_k \wedge \omega_i) \wedge \pi_2^\star(\omega_l \wedge \tau_j)$$

$$= \sum_{ij} \int_{M\times M} \gamma_{ij}.(-1)^{|\omega_i(||\omega_l|+|\tau_k|)}\pi_1^\star(\omega_i \wedge \tau_k) \wedge \pi_2^\star(\omega_l \wedge \tau_j)$$

In this integral, the term $\pi_1^\star(\omega_i \wedge \tau_k)$ involves the coefficients of the first M in $M \times M$ and the remaining term involves coefficients of the second M in the product. Consequently we can integrate the first term over the first M and integrate the second term over the second M and multiply. The first integral is δ_{ik} and the second is δ_{lj} and when we sum over i and j, we obtain

$$\gamma_{kl}(-1)^{|\omega_k|(|\omega_l|+|\tau_k|)}$$

Comparing results, we obtain $(-1)^{|\tau_k||\omega_k|}\beta_{lk} = \gamma_{kl}(-1)^{|\omega_k|(|\omega_l|+|\tau_k|)}$. So $\gamma_{kl} = (-1)^{|\omega_k||\omega_l|}\beta_{lk}$ and

$$d_G = \sum (-1)^{|\omega_i||\omega_j|}\beta_{ji}\pi_1^\star(\omega_i) \wedge \pi_2^\star(\tau_j)$$

Proof, concluded: Now form the following integral, where again $i : M \to G$ is the map $i(m) = m \times f(m)$, a diffeomorphism:

$$\int_\Delta d_G = \int_M \sum_{ij} (-1)^{|\omega_i||\omega_j|}\beta_{ji} i^\star \pi_1^\star(\omega_i) \wedge i^\star \pi_2^\star(\tau_j)$$

This equals

$$\sum_{ij} (-1)^{|\omega_i||\omega_j|}\beta_{ji} \int_M \omega_i \wedge \tau_j = \sum_i (-1)^{|\omega_i|}\beta_{ii} = L(f)$$

However, if Δ and G meet transversally, then the definition of the dual class to Δ gives

$$\int_\Delta \omega_G = \int_{M \times M} \omega_G \wedge d_\Delta$$

for all ω_G and consequently

$$L(f) = \int_\Delta d_G = \int_{M \times M} d_G \wedge d_\Delta$$

so the intersection number is the Lefshetz number of f up to sign.

A special case of the theorem proved at the end of this document states that

$$d_G \wedge d_\Delta = d_{G \cap \Delta}$$

When G and Δ are transverse, this is just a finite set of points $\mathcal{P}$ on $M \times M$, namely all $p \times p$ where p is a fixed point of f. Let g be a function on $M \times M$ which is identically 1. By definition of the dual class,

$$\int_{\mathcal{P}} f = \int_{M \times M} d_{G \wedge \Delta}$$

The first expression is $\sum_{p \in \mathcal{P}} \pm f(p)$ where $\pm$ depends on the orientation assigned to each p. We will determine this sign at the end of these notes. Consequently we have

$$L(f) = \sum_{p \in \mathcal{P}} \pm 1$$

Chapter 6

The Thom Transversality Theorem

6.1 Measure Theory

Recall the beginning of a typical analysis course in measure theory. By definition, an open box is a subset of the form $(a_1, b_1) \times (a_2, b_2) \times \ldots \times (a_n, b_n)$, and the volume of this box is $\prod |b_i - a_i|$. If A is an arbitrary subset of R^n, we can cover it by a countable number of open boxes B_i and form $\sum vol(B_i)$.

Definition 8 *The greatest lower bound of such sums over all countable coverings is called the* measure *of A and denoted $m(A)$.*

This measure generalizes *length*, when $n = 1$, *area*, when $n = 2$, and *volume*, when $n = 3$ and allows us to compute these numbers for much more general sets than those considered in ordinary calculus.

Measure is usually computed using the following results, rather than directly from the definition.

Theorem 22

- $0 \leq m(A) \leq \infty$
- *if $A_1 \subset A_2$, then $m(A_1) \leq m(A_2)$*
- *if B is an open box, then $m(B) = vol(B)$*
- *if T is a translation, then $m(T(A)) = m(A)$*
- *if A is a disjoint union of countably many sets A_i, then $m(A) = \sum m(A_i)$*

Most of these are trivial to prove; one is more difficult and will be proved below. The last property is by far the most important, but it is in a different league: it is not true in general. More about it in a moment.

Proof: We sketch the proof of the third item, which is the only tricky result in the first four items. If suffices to prove the result for closed boxes, because if B is open we can find closed B_1 and B_2 with $B_1 \subset B \subset B_2$; the result for closed boxes would then give $vol(B_1) \leq m(B) \leq vol(B_2)$ and the result would follow by taking limits as B_1 and B_2 approach B.

So assume B closed. Note that $m(B) \leq vol(B)$ since we can cover B by a single rectangle with volume as close to $vol(B)$ as we like.

By compactness, a cover of B by open boxes has a finite subcover. We give the remaining argument in two dimensions, but it clearly generalizes. If we extend the sides of the finite subcover, we obtain a mesh dividing B into rectangles. By simple algebra, the sum of the areas of the mesh rectangles equals the area of B. Each open mesh rectangle belongs to at least one rectangle in the subcover, so the sum of the mesh areas is at most the sum of the areas of the rectangles in the subcover and thus $vol(B) \leq \sum vol(R_i)$. So $vol(B) \leq m(B)$.

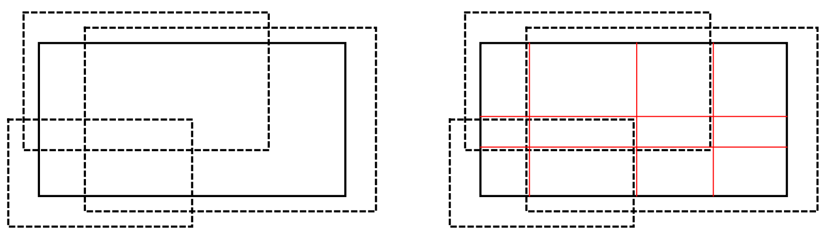

Figure 6.1: Measure equals Volume

Remark: We now turn to the requirement that if A is a disjoint union of countably many sets A_i, then $m(A) = \sum m(A_i)$. This is the essential property of a measure. We must restrict to countable unions because otherwise the result is trivially false, since every set is a disjoint union of isolated points, each of which has measure zero.

Lemma 6 *If $A = \cup A_i$ is a countable union, not necessarily disjoint,*

$$m(A) \leq \sum m(A_i)$$

Proof: Suppose $\epsilon > 0$. Find a countable cover of A_i by rectangles R_{ij} such that

$$\sum_j vol(R_{ij}) \leq m(A_i) + \frac{\epsilon}{2^{i+1}}$$

Then the R_{ij} form a countable cover of $A = \cup A_i$ of total volume $\sum_{ij} vol(R_{ij}) \leq \sum m(A_i)+\epsilon$ so

$$m(A) \leq \sum_{ij} vol(R_{ij}) \leq \sum m(A_i) + \epsilon$$

This holds for all $\epsilon > 0$ and thus also for $\epsilon = 0$, QED.

Remark: If the union in the lemma is disjoint, we might hope that the inequality would become an equality. Examples to be given in the next section show that this is false. So we must restrict attention to certain "nice" sets, called *measurable sets*. It will turn out, however, that virtually all sets are measurable; counterexamples require the axiom of choice.

The definition below is due to Cartheodory. The motivation for this definition will be given in the following section.

Definition 9 (Cartheodory) *A set A is measurable if for any set S we have*

$$m(S) = m(S \cap A) + m(S - A)$$

Theorem 23 *If A_i are a countable family of disjoint measurable sets, then $m(\cup A_i) = \sum m(A_i)$.*

Proof: Let $S = A_1 \cup A_2$ and apply the definition of measurability of A_2. Thus $m(S) = m(S \cap A_2) + m(S - A_2)$ and so $m(A_1 \cup A_2) = m(A_2) + m(A_1)$.

Now apply the definition of measurability of A_3 to $S = A_1 \cup A_2 \cup A_3$:

$$m(S) = m(S \cap A_3) + m(A - A3) = m(A_3) + m(A_1 \cup A_2)$$

So by the previous result, $m(A_1 \cup A_2 \cup A_3) = m(A_1) + m(A_2) + m(A_3)$, and in general $m(\cup A_i) = \sum m(A_i)$ for finite disjoint unions.

Now consider the sum of a countable disjoint union $\sum_{i=1}^{\infty} m(A_i)$. If this sum is infinite, then either one of the terms $m(A_j)$ is infinite or else finite sums $\sum_{i=1}^{n} m(A_i) = m(\cup_{i=1}^{n} A_i)$ grow arbitrarily large. In the first case, $A_j \subset \cup_{i=1}^{\infty} A_i$ and so $m(\cup_{i=1}^{\infty} A_i) = \infty$. In the second case, $A_1 \cup \ldots \cup A_n \subset \cup A_i$ and so $\sum_{i=1}^{n} m(A_i) \leq m(\cup_{i=1}^{\infty} A_i)$; since the left side is unbounded, $m(\cup_{i=1}^{\infty} A_i)$ is infinite, and the result is true.

Finally, suppose $\sum_{i=1}^{\infty} m(A_i)$ converges to a finite sum. Then for any $\epsilon > 0$ we can find N such that $\sum_{i=N+1}^{\infty} m(A_i) < \epsilon$. Then $A_1 \cup \ldots \cup A_N \subset \cup_{i=1}^{\infty} A_i$ and so

$$\sum_{i=1}^{N} m(A_i) \leq m(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} m(A_i) \leq \sum_{i=1}^{N} m(A_i) + \epsilon$$

Since ϵ is arbitrary, $\sum_{i=1}^{\infty}(A_i) = m(\cup_{i=1}^{\infty} A_i)$. QED.

Remark: It remains to show that there are lots of measurable sets.

Theorem 24 *Measurable sets satisfy the following properties:*

- *any set with measure zero is measurable*
- R^n *is measurable; if* A *is measurable, so is* $A^c = R^n - A$
- *if* A_i *is a countable collection of measurable sets, then* $\cup A_i$ *is measurable*
- *Open boxes are measurable*
- *Arbitrary open and closed sets are measurable*

Remark: Using these results, many other measurable sets can be constructed. For example, the boundary of a rectangle or disk has measure zero, so any subset of this boundary has measure zero. Thus if we start with an open rectangle or disk and add any subset of its boundary, the resulting set is still measurable.

Proof: Suppose A has measure zero. Note that $S \cap A \subset A$ has measure zero. Since $S = (S \cap A) \cup (S - A)$, $m(S) \leq m(S \cap A) + m(S - A) = m(S - A) \leq m(S)$ and we are done.

The second item is trivial, for if S is a set, $m(S) = m(S \cap R^n) + m(S - R^n) = m(S) + m(\emptyset)$. Moreover, if S is a set and $m(S) = m(S \cap A) + m(S - A)$ then $S \cap A^c = S - A$ and $S - A^c = S - A$ and so $m(S) = m(S \cap A^c) + m(S - A^c)$.

For the third item, we first prove that finite unions of measurable sets are measurable. It suffices to study the union of two measurable sets A_1 and A_2.

Then

$$S = [S \cap (A_1 \cup A_2)] \cup [S - (A_1 \cup A_2)] = (S \cap A_1) \cup ([S - A_1] \cap A_2) \cup [[S - A_1] - A_2)$$

is a disjoint union, and so we get an inequality which we want to be an equality:

$$m(S) \leq m(S \cap A_1) + m([S - A_1] \cap A_2) + m([S - A_1] - A_2)$$

Applying the definition of measurability of A_1 and A_2 we have

$$m(S) = m(S \cap A_1) + m(S - A_1) = m(S \cap A_1) + m([S - A_1] \cap A_2) + m([S - A_1] - A_2)$$

Hence the right side of the previous inequality is equal to $m(S)$ and thus the inequality is indeed an equality.

Now suppose that A_i is a countable collection of measurable sets. We can replace A_2 by $A_2 - A_1 = A_2 \cap A_1^c$, and replace A_3 by $A_3 - (A_1 \cup A_2) = A_3 \cap (A_1 \cup A_2)^c$, etc. So we can suppose the A_j are disjoint.

We need a lemma to finish the argument:

Lemma 7 *If $A_1, A_2, \ldots, A_n$ are disjoint, measurable sets, then for any S we have*

$$m(S \cap (\cup_{i=1}^n A_i)) = \sum_{i=1}^n m(S \cap A_i)$$

Proof of lemma: We prove this by induction on n. It is trivial for $n = 1$. In the induction step, let $\tilde{S} = S \cap (\cup_{i=1}^{n+1} A_k)$ and apply the definition of measurability to $\tilde{S}$ and A_{n+1}. We have

$$m(S \cap (\cup_{i=1}^{n+1} A_i) = m(S \cap (\cup_{i=1}^{n+1} A_i) \cap A_{n+1}) + m(S \cap (\cup_{i=1}^{n+1} A_i) - A_{n+1})$$

Since the A_k are disjoint this says

$$m(\cup_{i=1}^{n+1} S \cap A_i) = m(S \cap A_{n+1}) + m(\cup_{i=1}^n S \cap A_i)$$

and by induction this is $\sum_{i=1}^{n+1} m(S \cap A_i)$. QED.

Continuation of proof of main theorem: Since finite unions of measurable sets are measurable, and since the A_i are disjoint, we have

$$m(S) = m(S \cap (\cup_{i=1}^n A_i) + m(S - \cup_{i=1}^n A_i) \geq \sum_{i=1}^n m(S \cap A_i) + m(S - \cup_{i=1}^\infty A_i)$$

Looking just at the left and right sides, only one term depends on n, and yet the inequality is true for all finite n. It follows that it remains true as n approaches ∞ and so

$$m(S) \geq \sum_{i=1}^\infty m(S \cap A_i) + m(S - \cup_{i=1}^\infty A_i)$$

But then

$$m(S) \geq \sum_{i=1}^\infty m(S \cap A_i) + m(S - \cup_{i=1}^\infty A_i) \geq m(S \cap (\cup_{i=1}^\infty A_i) + m(S - \cup_{i=1}^\infty A_i) \geq m(S)$$

so every inequality in the final equation is an equality and $\cup_{i=1}^\infty A_i$ is measurable.

We still have items in our main theorem. The fourth item clearly implies the final item, so to complete the proof we must show that open boxes are measurable. The proof depends on another lemma:

Lemma 8 *Suppose A and B are disjoint subsets of R^n, not necessarily measurable, and suppose there is a δ such that whenever $a \in A$ and $b \in B$, the distance $d(a, b) \geq \delta$. Then $m(A \cup B) = m(A) + m(B)$.*

Proof: The definition of Lebesgue measure requires covering a set by open boxes. But we could just as well use closed boxes; the minor details proving that both definitions give the same measure are left to the reader.

We know that $m(A \cup B) \leq m(A) + m(B)$. It suffices to prove that $m(A \cup B) \geq m(A) + m(B)$. Let $\epsilon > 0$ and find a countable covering of $A \cup B$ by closed boxes R_i such that $\sum vol(R_i) < m(A \cup B) + \epsilon$. Subdivide each box into smaller closed boxes such that any two points of a subdivided box are less than δ apart. These smaller boxes still cover $A \cup B$ and the sum of their volumes is unchanged over the original sum. But a smaller box intersects A or B or neither. Divide the boxes into those that cover A and those that cover B or neither. Then $\sum vol(R)$ over those that cover A is larger than $m(A)$ and $\sum vol(R)$ over those that cover B is larger than $m(A)$. It follows that

$$m(A) + m(B) \leq \sum_{A \subset R} vol(R) + \sum_{\text{rest}} vol(R) = \sum vol(R) < m(A \cup B) + \epsilon$$

This holds for all $\epsilon > 0$ and thus $m(A) + m(B) \leq m(A \cup B)$. QED.

Now we are ready to prove that each open box R is measurable, and thus that for any S, $m(S) = m(S \cap R) + m(S - R)$. As usual, it suffices to show that $m(S) \geq m(S \cap R) + m(S - R)$.

Enlarge R slightly to a bigger box R_1. Let T be the "moat" between R and R_1..

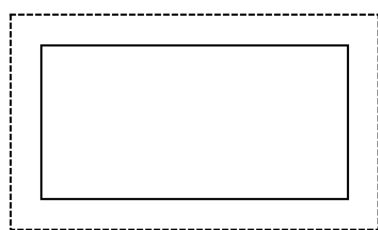

Figure 6.2: R, R_1, and Moat

If $\epsilon > 0$ is given to us, we can do the enlarging so $m(T) < \epsilon$ and thus $m(S \cap T) < \epsilon$.

Then $(S \cap R) \cup (S - R_1) \subset S$, so $m((S \cap R) \cup (S - R_1)) \leq m(S)$. Therefore by the lemma,

$$m(S \cap R) + m(S - R_1) \leq m(S)$$

Also $S - R = (S - R_1) \cup (S \cap T)$, so $m(S - R) \leq m(S - R_1) + m(S \cap T) \leq m(S - R_1) + \epsilon$, so $m(S - R_1) \geq m(S - R) - \epsilon$. Inserting this in the previously displayed inequality gives

$$m(S \cap R) + m(S - R) - \epsilon \leq m(S)$$

This equation holds for all $\epsilon > 0$ and thus

$$m(S \cap R) + m(S - R) \leq m(S)$$

QED.

6.2 Random Remarks on Previous Section

In the previous section, we presented the central results on Lebesgue mesure without interruption. Now some random comments are in order.

Countability and Measure Zero: It is critical in the Lebesgue theory to consider *countable* covers by boxes rather than finite covers. For instance, a finite cover of the set A of rational points in $[0,1]$ by intervals (a,b) must contain everything in $[0,1]$ except possibly any irrational endpoints of such (a,b). Since this exceptional set is finite, the cover will have total length at least 1 and the measure of A would be 1. A similar argument shows that the set of irrational points in $[0,1]$ would have measure 1, and thus additivity of measure would fail or else one of the sets would not be measurable.

But if countable covers are allowed and $\epsilon > 0$, we can enumerate the rational points $q_1, q_2, \ldots$ in $[0,1]$ and cover q_1 by an open interval of length at most $\frac{\epsilon}{2}$, cover q_2 by an open interval of length at most $\frac{\epsilon}{4}$, etc. We obtain a countable cover of the set by open intervals of total length at most ϵ, so the measure of the set of rational points in $[0,1]$ is less than ϵ for all $\epsilon > 0$ and thus 0. In particular, this set is measurable, so its complement in $[0,1]$ is also measurable, and thus the measure of the irrational points in $[0,1]$ is 1.

The argument of the previous paragraph shows that any countable set in R^n has measure zero. The converse of this result is false. For example, consider the Cantor set in R^1, formed by removing the middle third from $[0,1]$, and then removing the middle thirds of the two remaining pieces, etc. The total length of the set removed is

$$\frac{1}{3} + 2\frac{1}{3^2} + 4\frac{1}{3^3} + \ldots = \frac{1}{3}\left[1 + \frac{2}{3} + \left(\frac{2}{3}\right)^2 + \ldots\right] = \frac{1}{3}\frac{1}{1-\frac{2}{3}} = 1$$

and thus the complement in $[0,1]$, the Cantor set, has measure zero. (Note that all of these sets are measurable and thus the argument is rigorous by the previous section.)

The points in $[0,1]$ are associated with infinite "decimals" in base three, like .0221003.... We removed exactly all points with at least one 2 in their expansion, so the Cantor set is associated with base 3 expansions containing only 0's and 2's, and this set is uncountable by the standard argument.

Euclid: Everyone should read at least Book 1 of Euclid. This remarkable document starts with just the axioms — essentially nothing — and proceeds in a logical line to the Pythagorian theorem and its converse. It tells a great story as succinctly as possible.

The book is divided into essentially two parts. The first part is about *congruent figures*, or as Euclid writes, figures which are equal. In modern language, two figures are congruent if one can be mapped to the other by translations, rotations, and reflections, that is, by a Euclidean map. Euclid doesn't develop the theory that way, instead building up the theory from congruence theorems for triangles. But his proof of side-angle-side is telling, because Euclid says to pick up the first triangle by the angle and place it over the angle of the second triangle; the two sides then match, so the third side and the angles at this side also match. This proof has been criticized as "not in the spirit of Euclid" and modern authors often add SAS to the axioms. But it is directly in the spirit of modern geometry, which makes explicit the group of Euclidean motions.

In the middle of the chapter, Euclid begins calling figures equal when they obviously are not congruent, and it soon becomes apparent that *equal* now means *having the same area.* Euclid never calculates an area, no doubt because the Greeks had only rational numbers but knew that not all distances and areas are rational. The only objects in chapter 1 are regions with straight sides, and Euclid says that two such objects are equal if the first can be cut into pieces and these pieces can be reassembled using Euclidean motions to form the second.

Indeed, it can be proved that if two regions with straight sides have the same area, the first can be decomposed into pieces which can be reassembled to form the second. The proof is relatively easy by using Euclid's methods.

These ideas reappeared in 1900 when David Hilbert announced his famous list of problems for the twentieth century. One of the problems asked if this Euclidean approach works in three dimensions. That is, if two three-dimensional regions have plane sides and the same volume, can one of them be decomposed into a finite number of pieces, which are then reassembled to form the second. Hilbert conjectured that the answer is "no". This was the first problem to be solved, and Hilbert's conjecture turned out to be correct.

The previous section on measure contains almost everything necessary to reprove the results in book 1 of Euclid using area directly. The missing ingredient is invariance of Lebesgue measure under rotations.

There are at least two ways to obtain this missing result. The first is to allow boxes with any orientation in the definition of measure. The sides of such a box define orthogonal vectors in R^n and the volume of the box is then the absolute value of the determinant of the row matrix formed by these vectors. Since the "fundamental objects" – the boxes – are invariant under both rotation and translation, it immediately follows that measure is also invariant. It is then only necessary to reprove the results of the previous section with this new definition of measure, but only two results directly mention the boxes, and their proofs are easily modified.

Another approach is more abstract and general. Suppose G is a topological group. A measure on open sets of G is called a Haar measure if it is invariant under left translation. Haar proved that all locally compact topological groups possess such a measure, and von Neumann proved that this measure is unique up to a multiplicative constant. If we know this uniqueness result, the missing proof is easy. Let $m_1(A)$ be Lebesgue measure. Let $m_2(A)$ be the measure obtained by applying a fixed Euclidean orthogonal transformation T to A and then computing its Lebesgue measure. Both are invariant under translations, so $m_2 = \lambda m_1$ for a fixed λ.

On the other hand, $m_1(A) = m_2(A)$ when applied to the standard closed ball of radius 1, so $\lambda = 1$ and $m_1 = m_2$.

Inner and Outer Measure and Cartheodory:

The classical theory of area goes essentially back to the Greeks. Suppose A is a bounded set in R^2. We approximate its area from above by covering it with a finite mesh of closed rectangles. The greatest lower bound of the resulting areas is called the *outer measure* of A and denoted $m^\star(A)$. We also approximate its area from below by finding a mesh of closed rectangles inside A. The least upper bound of the resulting areas is called the *inner measure* of A and denoted $m_\star(A)$. Easily, $m_\star(A) \leq m^\star(A)$. If these expressions are equal, we call their common value the Jordan measure $m(A)$ and say that A is Jordan measurable.

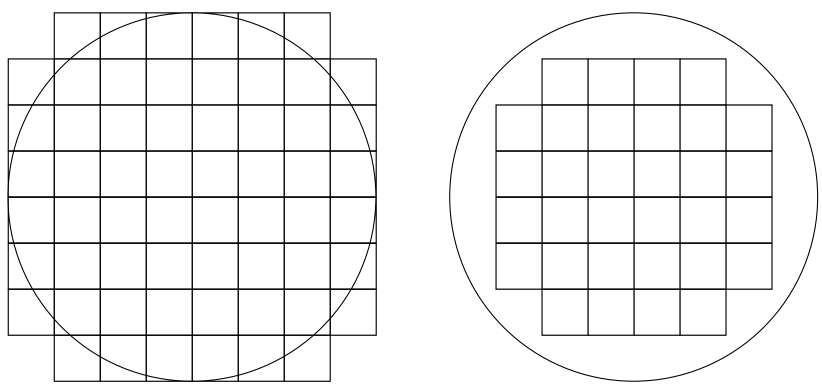

Figure 6.3: Outer Jordan Measure, Inner Jordan Measure

Earlier we gave Lebesgue's definition of Lebesgue outer measure, although we did not introduce the term. Let us temporarily write $m^\star(A)$ for this Lebesgue outer measure. In his original treatment of measure, Lebesgue introduced a term which, in a sense, corresponds to inner measure, and called a set *measurable* if the outer and inner measure were equal. Lebesgue originally worked with bounded subsets and length, so in his context $A \subset I$ where I is a fixed interval. Notice that $m^\star(I - A)$ measures the complement of A by computing the sum of the lengths of intervals which overlap part of A. So the expression

$vol(I) - m^\star(I - A)$ should correspond to a sort of inner measure of A. Lebesgue called A measurable if

$$vol(I) - m^\star(I - A) = m^\star(A)$$

This is similar to Jordan's requirement that outer and inner measure have the same limit. But if we rewrite the condition in the form

$$vol(I) = m^\star(A) + m^\star(I - A) = m^\star(I \cap A) + m^\star(I - A)$$

we see that I is playing the role of S in Cartheodory's definition of measurability. This is the motivation of that definition. Of course it is quite a jump from I to an arbitrary subset S.

Existence of Non-Measurable Sets:

We now construct a non-measurable set in R. Define a relation on R by writing $a \sim b$ if $a - b$ is rational. It is easy to check that this is an equivalence relation. If a represents an equivalence class, then $a + q$ also represents this class whenever $q \in Q$. For each equivalence class, select a representative in $[0, 1]$ and let E be the collection of all of these representatives.

Notice that there are uncountably many equivalence classes, and thus we must make uncountably many choices to construct E. So we are definitely using the *axiom of choice* here!

If $a \in [0, 1]$, then $a \sim e$ for some $e \in E$, so $a - e = q$ for q rational. Clearly $-1 \leq q \leq 1$. Hence

$$[0, 1] \subseteq \cup_{q \in [-1,1]}(E + q) \subseteq [-1, 2]$$

The expression in the middle is a countable union of *disjoint* sets, for if $e, f \in E$ and $e + q_1 = f + q_2$, then $e \sim f$ and so $e = f$.

Recall that Lebesgue measure is translation invariant. If E were measurable, then

$$1 \leq \sum_{q \in [-1,1]} m(E) \leq 3$$

The left inequality can only hold if $m(E) > 0$ and the right inequality can only hold if $m(E) = 0$. So E is not measurable.

Remark: Solovay constructed a model for Zermelo-Frankel set theory without the axiom of choice in which every subset of R^n is measurable. Consequently the construction of a non-measurable set require the axiom of choice.

Remark: We have been using sums over countably many disjoint sets. It is reasonable to hope that measurability problems would vanish if we restricted attention to sums over finitely many disjoint sets. But this hope is dashed by the Banach-Tarski paradox, in which for example a sphere of radius 1 is written as a finite disjoint union of sets A_i and then the A_i are rearranged using Euclidean motions to form a finite disjoint decomposition of the sphere of radius 10^{100}.

6.3 Measure Zero in Manifolds

Take a piece of paper. Fold it, crumple it, and then press the paper down on the desk. You are looking at a map from the paper to the desk, or mathematically a map $f : \mathcal{U} \subset R^2 \to R^2$. Assume this map is C^∞; after all, folds like $(x, y) \to (x, y^2)$ are infinitely differentiable. We are going to prove that most points on the table are covered only by points where f is a local diffeomorphism, so no fold or other singularity occurred over those points. Incidentally, these "good" points on the table include points not covered at all by the map.

What is the meaning of *most* in this statement. Mathematicians often claim that statements are usually true, and assign different meanings to the assertion. Sometimes they mean there are only finitely many exceptions. Sometimes they mean that the exceptions are countable. Sometimes they mean the exceptions form a nowhere dense subset of a complete metric space. In the present situation, we mean that the exceptions form a set of measure zero.

From now on, all manifolds are second countable.

Definition 10 *A subset E of a second countable C^∞ manifold M is said to have measure zero if for all coordinate systems $\varphi : \mathcal{U} \to \mathcal{V}$ on M the set $\varphi(\mathcal{U} \cap E) \subset \mathcal{V} \subset R^n$ has measure zero.*

Remark: Luckily, we need not verify this for all coordinate systems. It suffices to verify it for some particular coordinate cover of E; in particular if E is inside a single coordinate system it suffices to verify it in that system. Proving this assertion requires the following lemma, the fact that M has a countable basis, and the fact that a countable union of sets of measure zero again has measure zero.

Lemma 9 *If $E \subset \mathcal{U} \subset R^n$ has measure zero and $\psi : \mathcal{U} \to \mathcal{V} \subset R^n$ is a diffeomorphism, then $\psi(E)$ also has measure zero.*

Proof: Suppose $\mathcal{W}$ is a ball containing two points $(x_1, \ldots, x_n)$ and $(a_1, \ldots, a_n)$. Then

$$\psi^i(x_1, \ldots, x_n) - \psi^i(a_1, \ldots, a_n) = \sum_j \left(\psi^i(x_1, \ldots, x_j, a_{j+1}, \ldots, a_n) - \psi^i(x_1, \ldots, a_j, a_{j+1}, \ldots, a_n)\right)$$

$$= \sum_j \frac{\partial \psi^i}{\partial x_j}(\xi_j)(x_j - a_j)$$

Here we have applied the mean value theorem to the straight line formed by changing just one variable, using the fact that this entire line is in $\mathcal{W}$. Assume $\left|\frac{\partial\psi^i}{\partial x_j}\right| \leq B$ is uniformly bounded in our ball. Then

$$\left|\psi^i(x_1,\ldots,x_n) - \psi^i(a_1,\ldots,a_n)\right| \leq Bn \max_j |x_j - a_j|$$

This result says that in some sense the volume of an image box covering part of $\psi(E)$ is a constant multiple of the volume of a box covering part of E. We just need to clean up the details of this observation.

Observe first that when computing the measure of a set, we can cover it with cubes (generalized squares) rather than boxes (generalized rectangles). Indeed, if we have a box belonging to a cover, we can cut one side into N equal pieces of length $\frac{b_1-a_1}{N}$, and then cut all remaining sides into pieces of the same length, where the number of pieces is chosen to completely cover that side with possible slight overlap of the final piece. This cuts the box into a large number of cubes, with a slightly greater volume due to the overlap along the remaining sides. But since we can select N at will, this extra volume can be made as small as we wish for each individual box. Using the "$\frac{\epsilon}{2^i}$" trick, the total volume of the cubes can be made as close to the total volume of the original boxes as we like.

Now consider a cube in the domain of ψ; let us try to cover the image of this cube by the smallest possible cube in the range of ψ. To do this, let x range over the domain cube until $\psi^i(x)$ is as large as possible, and let a range over the domain cube until $\psi^i(a)$ is as small as possible. The difference gives a possible length for the ith side of the covering cube in the image. By the above inequality, this is no greater than Bn times the side of the domain cube. This result holds for all i, so the cube covering the image will have volume at most $(Bn)^n$ times the volume of the domain cube.

Our final problem is that the bound B need not hold throughout all of $\mathcal{U}$. Find open sets $\mathcal{U}_i$ and compact sets K_i with

$$\mathcal{U}_1 \subset K_1 \subset \mathcal{U}_2 \subset K_2 \ldots$$

such that the union of all of these sets is $\mathcal{U}$. Let $E_k = E \cap \mathcal{U}_k$. Since $E_k \subset E$, it has measure zero and can be covered by a countable number of cubes in $\mathcal{U}_k$ is arbitrarily small total volume. Moreover, $E_k \subset K_k$ on which $\left|\frac{\partial\psi^i}{\partial x_j}\right| \leq B_k$ is bounded. By the above calculations, $\psi(E_k)$ can be covered by a countable number of boxes of arbitrarily small volume, so $\psi(E_k)$ has measure zero. Note that $\psi(E) \subset \cup\psi(E)$, so $m(\psi(E)) \leq \sum m(\psi(E_k))$. Hence $\psi(E)$ has measure zero.

6.4 Sard's Theorem

Definition 11 *Let $f : M \to N$ be a C^∞ map between C^∞ manifolds. A point $p \in M$ is a* singular point *of f if $f_\star : T_p(M) \to T_{f(p)}(N)$ is not onto.*

Theorem 25 (Sard) *Let $f : M \to N$ be a C^∞ map between second countable C^∞ manifolds. Then the image under f of the critical points of f has measure zero in N.*

Remark: Note that the singular points belong to M. Easy examples show that every point of M can be a singular point (for instance, when dim(M) < dim(N)). It is the image of these singular points in N that has measure zero.

In the case dim(M) < dim(N), Sard's theorem says that the entire image of M in N has measure zero. So "space filling curves" which are C^∞ cannot exist.

If dim(M) = dim(N), a singular point is a point where $f_\star$ is not an isomorphism. The inverse function theorem asserts that when it is an isomorphism, f is a local diffeomorphism. So Sard's theorem asserts that for most points $n \in N$, either n is not in the image of f at all or else there are isolated points $m_1, m_2, \ldots$ mapping to n and f is a local diffeomorphism near each of them.

Proof: It suffices to prove the theorem when M is an open subset $\mathcal{U} \subset R^m$ and $f : \mathcal{U} \to R^n$ is C^∞. Indeed in the proof, we will often pick a singular point $p \in \mathcal{U}$ and find a smaller open neighborhood $\mathcal{W}$ of p such that the image of the singular points in $\mathcal{W}$ has measure zero. This is enough because M has a countable basis, and thus we can find one of these countably many open sets $\mathcal{B}$ with $p \in \mathcal{B} \subset \mathcal{W}$. It then follows that the image of the singular points lie in a countable union of sets of measure zero, and thus has measure zero.

We prove the result by induction on m. If $m = 0$, then R^m is a single point and the result is trivial. Assume the result true for $m - 1$.

Let C be the set of critical points in $\mathcal{U}$ and let C_k be the set of critical points where $\frac{\partial^j f^i}{\partial x_{i_1} \ldots \partial x_{i_j}} = 0$ for $j = 1, \ldots, k$. We will prove that

- $f(C - C_1)$ has measure zero
- $f(C_i - C_{i+1})$ has measure zero
- $f(C_k)$ has measure zero for some k

In the first case, suppose $p \in C - C_1$. Then $\frac{\partial f_i}{\partial x_j} \neq 0$ at p for some i, j. Renumber the coordinates in both domain and range so $\frac{\partial f_1}{\partial x_1} \neq 0$ at p. Define a map $\varphi : \mathcal{U} \to R^m$ by $\varphi(x_1, \ldots, x_m) = (f_1(x_1, \ldots, x_m), x_2, \ldots, x_m) = (y_1, \ldots, y_m)$ and notice that the Jacobian at p is non-singular. So this map is a local diffeomorphism which introduces new coordinates on M near p. In this new coordinate system, our original map is $f \circ \varphi^{-1} : \mathcal{W} \subset R^m \to R^n$. The first coordinate of the image of $(y_1, \ldots, y_m)$ under this map is $f_1(x_1, \ldots, x_m) = y_1$. Thus in the new coordinate system, $(y_1, \ldots, y_m) \to (y_1, g(y_1, \ldots, y_m))$ where $g : \mathcal{W} \subset R^m \to R^{n-1}$. For each fixed y_1, g defines a map from an open slice of $\mathcal{W} \subset R^{m-1} \to R^{n-1}$ and we can apply induction to conclude that for each fixed y_1, the image of the singular points in $\mathcal{W}$ with first coordinate y_1 has measure zero in R^{n-1}.

To conclude this step, we need a lemma from measure theory:

Lemma 10 *Let $E \subset R^n$ where coordinates in R^n are denoted $(t_1, \ldots, t_n)$. Suppose that for each fixed t_1, the set of all $(t_1, e_2, \ldots, e_n) \in E$ has measure zero in R^{n-1}. Then E has measure zero in R^n.*

Remark: This result is a special case of Fubini's theorem (or Torelli's theorem) about iterated Lebesgue integrals. Let $\chi(x)$ on R^n be the characteristic function for E. Thus $\chi(x) = 1$ if $x \in E$ and 0 otherwise. Note that $\chi \geq 0$. Torelli's theorem is just Fubini's theorem in the non-negative function case. According to this theorem

$$\int_{R^n} \chi(x, y) = \int_R \left(\int_{R^{n-1}} \chi(x, y) dy \right) dx$$

The hypothesis shows that the interior integral on the right is zero, so the full integral is zero, and thus E has measure zero.

Remark: Shlomo Sternberg, in his book on Differential Geometry, explained how to replace Fubini's theorem in this argument with very elementary measure theory. A footnote attributes the proof to Furstenberg. Here is that argument.

Observe that in the end we study $f \circ \varphi^{-1} : \mathcal{W} \subset R^m \to R^n$. We can replace $\mathcal{W}$ by a smaller neighborhood of p with compact closure K in $\mathcal{W}$. Let $S \subset K$ be the set of all singular points of f (rather than just the image of $C - C_1$) and notice that S is closed, hence compact. Indeed if q_i are singular and q_i converges to q, then the matrices $f^\star(q_i)$ converge to the matrix $f^\star(q)$. If $f^\star(q)$ is onto R^n, then we can find vectors $v_1, \ldots, v_n$ which map to the elementary basis vectors of R^n by $f^\star(q)$, so in the limit these vectors must be linearly independent for large $f^\star(q_i)$.

Let $E = f(S)$. Our induction hypothesis then shows that each slice of E with fixed first component has measure 0. Notice that S is closed and so compact, so E is compact. Here then is the elementary fact from measure theory:

Lemma 11 *Let E be a compact subset of R^n and suppose each slice in R^{n-1} formed by intersecting E with points with fixed first component has measure zero. Then E has measure zero.*

Proof: Fix $\epsilon > 0$. Consider the slice of R^n obtained by setting $x_1 = \xi$. The intersection of E with this slice has measure zero, so we can cover it by countably many open boxes B_i in R^{n-1} of total volume less than ϵ. We claim that we can find $a < \xi < b$ such that the boxes $(a, b) \times B_i$ cover those points in E with first coordinate between a and b. If this were false, then we could find a sequence of points $e_i = t_i \times f_i$ in E where t_i converge to ξ and f_i is not in the open union of the B_j. This sequence would then have a convergent subsequence in E, and the limit would be $\xi \times f$ for f outside the union of the boxes, and so not in our slice.

Notice that the extended box $(a, b) \times B_i$ has volume $(b - a)vol(B_i)$ and the total volume of the extended boxes is at most $(b - a)\epsilon$.

Since E is compact, its first components form a closed set E_1 in a finite interval $[A, B]$ The set E_1 is covered by the open intervals (a, b) obtained in the previous paragraph. By compactness, a finite number of these intervals cover E_1. We can choose A and B and the intervals so each interval is inside $[A, B]$. Note that the intervals cover E_1 but this set need not be connected, so they don't necessarily "join up" to cover $[A, B]$. This doesn't matter.

Begin throwing away intervals until no interval in the cover can be omitted and still cover E_1. Order these intervals by the order of their first elements. Then they are also ordered by their last elements, for if $b_{i+1} < b_i$ then $(a_{i+1}, b_{i+1}) \subset (a_i, b_i)$ and we can omit the second element.

The key observation is then that no three intervals contain a point in common. Otherwise, name the three intervals from left to right, $(c_1, d_1), (c_2, d_2), (c_3, d_3)$ and call their common point p. Notice that the middle interval is covered by the union of the first and last and thus redundant. Indeed, $c_3 < p < d_1$.

So the total length of the intervals is at most $2(B - A)$ and the total volume of the resulting boxes in R^n is at most $2(B - A)\epsilon$. This can be made arbitrarily small, and thus E has measure zero. QED.

Induction proof continued: Next let $p \in C_i - C_{i+1}$. Hence all partial derivatives of f of order i or smaller vanish, but some derivative of order $i + 1$ does not vanish. By renumbering the coordinates on R^m, we can suppose

$$\frac{\partial}{\partial x_1}\left(\frac{\partial^i f_k}{\partial x_{j_1} \ldots \partial x_{j_i}}\right) \neq 0$$

We argue as before. Let $\varphi : \mathcal{U} \to R^m$ by

$$\varphi(x_1, \ldots, x_m) = \left(\frac{\partial^i f_k}{\partial x_{j_1} \ldots \partial x_{j_i}}, x_2, \ldots, x_m\right) = (y_1, \ldots, y_m)$$

Since the Jacobian of this map is not zero at p, this gives new local coordinates near p. Note that φ maps C_i to $0 \times R^{m-1} \subset R^m$ since at any point $q \in C_i$ the first coordinate of φ vanishes. In the new coordinate system our map is $f \circ \varphi^{-1}$ and is defined on an open subset $\mathcal{W} \subset R^m$; it is difficult this time to determine the formula for f but we do not need it.

If we intersect C_i with our new coordinate system, φ maps it to $0 \times R^{m-1}$ and $f \circ \varphi^{-1}$ on the image of C_i is induced by a map from $\mathcal{W} \cap 0 \times R^{m-1}$ to R^n. Thus by induction the image of the critical points of this map has measure zero. But the image of C_i under

φ maps to critical points of $f \circ \varphi^{-1}$ since C_i is contained in the set of critical points of f.

Induction proof concluded: The proof of our third claim follows a different path. Suppose I^m is a cube with side length δ inside U. We claim there is a constant c depending only on f and I^m such that whenever $a \in C_k \cap I^m$ and $a+h \in I^m$, we have $||f(a+h)-f(a)|| \leq c||h||^{k+1}$. This will be proved using Taylor's theorem; assume it for a moment.

We will use this inequality to prove that when k is large enough, $f(C_k \cap I^m)$ has measure zero. Let p be a positive integer. Subdivide I^m into p^m sub-cubes of side $\frac{\delta}{p}$. Suppose one of these sub-cubes contains a point $a \in C_k \cap I^m$ and suppose $a+h$ is in this same sub-cube. Then $||f(a+h)-f(a)|| \leq c||h||^{k+1}$. The largest possible value for $||h||$ is the diameter of the sub-cube, which is $\sqrt{m}\frac{\delta}{p}$ because

$$||(a_1, \ldots, a_m) - (b_1, \ldots, b_m)|| = \sqrt{\sum (a_i - b_i)^2} \leq \sqrt{\sum \left(\frac{\delta}{p}\right)^2} = \sqrt{m}\frac{\delta}{p}$$

The triangle inequality then shows that the maximum distance between any two points in our sub-cube is

$$2c\left(\sqrt{m}\frac{\delta}{p}\right)^{k+1}$$

Therefore, the image of the sub-cube is contained in a cube in R^n with this as side length; the volume of this cube is

$$\left[2c\left(\sqrt{m}\frac{\delta}{p}\right)^{k+1}\right]^n$$

Notice carefully that this estimate only applies to subcubes containing points $a \in C_k$. The total number of possible sub-cubes is p^m, so $f(C_k \cap I^m)$ can be covered by cubes of total volume

$$p^m\left[2c\left(\sqrt{m}\frac{\delta}{p}\right)^{k+1}\right]^n = (2c)^n(\sqrt{m}\delta)^{n(k+1)}p^{m-n(k+1)}$$

Most of the terms in this expression depend only on the original I^m. The subdivision into sub-cubes is completely determined by the integer p, which can be as large as we wish. But this expression goes to zero as soon as $m - n(k+1) < 0$, i.e., $(k+1) > \frac{m}{n}$. It follows that for all such k, $f(C_k \cap I^m)$ has measure zero.

Since $\mathcal{U}$ can be covered by a countable number of cubes I^m (with varying side length δ), $f(C_k)$ has measure zero.

To complete the proof, we need only give the details on the Taylor theorem inequality. Note that f is vector valued, but it suffices to prove the inequality when f is an ordinary function because

$$||f(a+h)-f(a)|| = \sqrt{\sum (f_i(a+h)-f_i(a))^2} \leq \sqrt{n}\max|f_i(a+h)-f_i(a)|$$

So assume that f is a real-valued function and recall the development of Taylor series with remainder:

$$f(a+h)-f(a) = \int_0^1 \frac{d}{dt}\left[f(a_1+th_1,\ldots,a_m+th_m)\right] dt = \int_0^1 \sum \frac{\partial f}{\partial x_i}(a_1+th_1,\ldots,a_m+th_m)h_i\, dt$$

$$= -\int_0^1 \sum \frac{\partial f}{\partial x_i}(a_1+th_1,\ldots,a_m+th_m)\frac{d}{dt}h_i(1-t)\ dt$$

Integrate by parts to get

$$-\sum \frac{\partial f}{\partial x_i}(a_1+th_1,\ldots,a_m+th_m)h_i(1-t)\Big|_0^1 - \int_0^1 \sum \frac{\partial^2 f}{\partial x_{i_1}\partial x_{i_2}}(a_1+th_1,\ldots,a_m+th_m)h_{i_1}h_{i_2}\frac{d}{dt}\frac{(1-t)^2}{2}\ dt$$

$$= \sum \frac{\partial f}{\partial x_i}(a_1,\ldots,a_m)\ h_i - \int_0^1 \sum \frac{\partial^2 f}{\partial x_{i_1}\partial x_{i_2}}(a_1+th_1,\ldots,a_m+th_m)h_{i_1}h_{i_2}\frac{d}{dt}\frac{(1-t)^2}{2}\ dt$$

Continuing, we reach

$$f(a+h)-f(a) = \sum \frac{\partial f}{\partial x_i}(a)\ h_i + \ldots + \frac{1}{k!}\sum \frac{\partial^k f}{\partial x_{i_1}\ldots\partial x_{i_k}}(a)\ h_{i_1}\ldots h_{i_k}$$

$$-\int_0^1 \sum \frac{\partial^{k+1} f}{\partial x_{i_1}\ldots\partial x_{i_{k+1}}}(a_1+th_1,\ldots,a_m+th_m)h_{i_1}\ldots h_{i_{k+1}}\frac{d}{dt}\frac{(1-t)^{k+1}}{(k+1)!}\ dt$$

In our case, $a \in C_k$ and all those partial derivatives in the series vanish, so we are left with just the remainder term. All those partials in the integral are continuous on I^m and thus bounded by a constant c_1. Each h_i is bounded in absolute value by $||h||$, so each product of $k+1$ of these terms is bounded by $||h||^{k+1}$. The remaining integral is

$$\int_0^1 \frac{d}{dt}\frac{(1-t)^{k+1}}{(k+1)!}\ dt = -\frac{(1-t)^{k+1}}{(k+1)!}\Big|_0^1 = \frac{1}{(k+1)!}$$

There are m^{k+1} terms in the sum, and we can thus let $c = c_1\frac{m^{k+1}}{(k+1)!}$ The fraction here is one of the terms in the convergent series for e^m, so these terms are bounded with a bound that depends on m but does not depend on k. Thus c depends only on f and I^m. QED.

6.5 Transversality

Definition 12 *Let K and L be submanifolds of a manifold M which intersect at a point p. We say they* intersect transversally *if $T_p(K)+T_p(L) = T_p(M)$. Note that this sum need not be direct.*

If K and L intersect transversally at each common point, we say the submanifolds are transverse *and write $K \pitchfork L$.*

Remark: The intuition behind this definition is that transverse intersections are generic. If an intersection is transverse, it will still be transverse after small deformations. But if an intersection is not transverse, then an arbitrarily small deformation can remove it entirely or else make it transverse.

Notice that no intersection is transverse if $dim(K) + dim(L) < dim(M)$. That is because there is room to pull K and L apart and remove the intersection. For instance, if two lines in R^3 meet, one can be raised slightly to avoid the intersection.

One of my sources for this material is an honors undergraduate thesis written by Jonathan Michael Bloom at Harvard in 2004. He included the following pictures to illustrate transversal and non-transversal intersections. A footnote adds that the illustrations come from Guillemin and Pollack's *Differential Topology*.

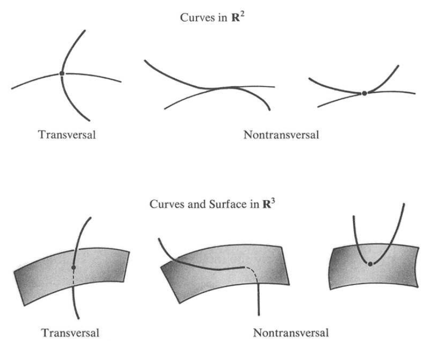

Figure 6.4: Examples One

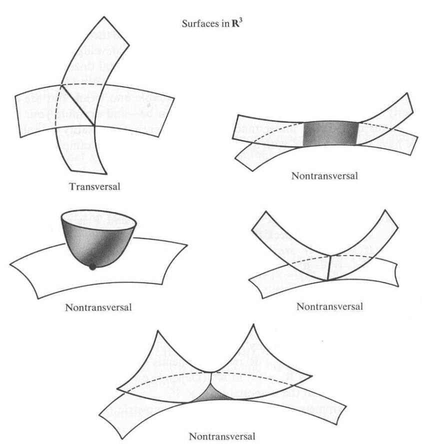

Figure 6.5: Examples Two

6.6 Immersions, Submersions, and Intersections

Definition 13 *Let $f : M \to N$ be a smooth map. We say f is an* immersion *if $f^\star : T_p(M) \to T_{f(p)}(N)$ is always one-to-one. We say f is a* submersion *if $f^\star : T_p(M) \to T_{f(p)}(N)$ is always onto.*

Theorem 26 *If $f : M \to N$ is an immersion, then $dim(M) \leq dim(N)$ and we can find local coordinates on M near any p and local coordinates on N near $f(p)$ such that*

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

If f $f : M \to N$ is a submersion, then $dim(M) \geq dim(N)$ and we can find local coordinates on M near any p and local coordinates on N near $f(p)$ such that

$$f(x_1, \dots, x_m) = (x_1, \dots, x_n) \qquad \textit{for } n \leq m$$

Proof in the case of an immersion: Pick local coordinates on $p \in \mathcal{U} \subset M$ and $f(p) \in \mathcal{V} \subset N$. We can make a linear transformation of the coordinates on $\mathcal{V}$ so $f^\star(T_p(M))$ is the subspace of the tangent space $T_{f(p)}(N)$ spanned by the first m elementary basis vectors. Consider the change of coordinates map on $\mathcal{U}$ defined by

$$(x_1, \dots, x_m) \to (f_1(x_1, \dots, x_m), \dots, f_m(x_1, \dots, x_m))$$

Note that f has n coordinate functions, but we are only using the first m functions. The Jacobian of this map is nonzero at p by assumption, so the map defines new coordinates $y_1, \dots, y_m$ on $\mathcal{U}$. If we use these new coordinates, the map f has the form

$$(y_1, \dots, y_m) \to (y_1, \dots, y_m, f_{m+1}(y_1, \dots, y_m), \dots, f_n(y_1, \dots, y_m))$$

To avoid an abundance of letters, assume this is the original system, so our map f is locally

$$(x_1, \dots, x_m) \to (x_1, \dots, x_m, f_{m+1}(x_1, \dots, x_m), \dots, f_n(x_1, \dots, x_m))$$

This is the expression of f using coordinates on M and our original coordinate system on N. Now make a change of coordinates on N by mapping

$$(x_1, \dots, x_m, \dots x_n) \to (x_1, \dots, x_m, x_{m+1} - f_{m+1}(x_1, \dots, x_m), \dots, x_n - f_n(x_1, \dots, x_m))$$

The Jacobian is non-zero at $f(p)$ by assumption, so these are new coordinates, and in these new coordinates

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

Proof in the case of a submersion: Pick local coordinates near p and call the coordinates x_i on M and y_j on N. By making linear coordinate changes in these variables we can assume that $f^\star$ maps $\frac{\partial}{\partial x_i} \in T_p(M)$ to $\frac{\partial}{\partial y_i} \in T_{f(p)}(N)$ for $1 \leq i \leq n$. Notice that $n \leq m$ so the remaining basis vectors for $T_p(M)$ can map where they like.

Introduce new coordinates on M by

$$(y_1, \ldots, y_m) = (f_1(x_1, \ldots, x_m), \ldots, f_n(x_1, \ldots, x_m), x_{n+1}, \ldots, x_m)$$

The Jacobian of this map at p is non-zero, so these are indeed new coordinates. The map f is given in these new coordinates by $(y_1, \ldots, y_n, y_{n+1}, \ldots, y_n) \to (x_1, \ldots x_m) \to (f_1(x), \ldots, f_n(x)) = (y_1, \ldots, y_n)$.

Remark: We immediately use this theorem to prove a central point about transversal intersections:

Theorem 27 *Let K and L be transverse submanifolds of M. Suppose the dimensions of these manifolds are k, l, and m. Then $K \cap L$ is also a submanifold of M, of dimension $(k + l - m)$. In particular, if $k + l < m$ the intersection is empty, and if $k + l = m$ the intersection consists of isolated points.*

Proof: It suffices to work locally. Let $\mathcal{V}$ be a coordinate neighborhood on L and $\mathcal{W}$ be a coordinate neighborhood on M and let $\varphi : \mathcal{V} \to \mathcal{W}$ be the injection map which makes L a submanifold of M. By the above immersion theorem, we can pick local coordinates on L and M so the inclusion map is

$$(x_1, \ldots, x_l) \to (x_1, \ldots, x_l, 0, \ldots, 0)$$

Now let $p \in K \cap L$. Since $K \subset M$ is a submanifold, we can find an open neighborhood $\mathcal{U}$ of p in K which maps by the inclusion $K \subset M$ into the the coordinate neighborhood on M obtained in the previous paragraph. Let $g : \mathcal{U} \to R^{m-l}$ be this map followed by projection onto the last $m - l$ coordinates. Notice that a point in $\mathcal{U}$ belongs to $K \cap L$ exactly when this projection maps to 0.

By assumption, K and L meet transversally at p, so $T_p(K) + T_p(L) = T_p(M)$. In coordinates, $T_p(L)$ maps to the subspace generated by $\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_l}$ in $T_p(M)$. So $\frac{\partial}{\partial x_{l+1}}, \ldots, \frac{\partial}{\partial x_n}$ must be expressible as elements of $T_p(K)+T_p(L)$, and if we follow the inclusion of $K \to R^m$ with the projection $R^m \to R^{m-l}$ on the last $m - l$ coordinates, each of these generators must be in the image. It follows that g is a submersion.

By the previous theorem, we can find new coordinates $(y_1, \ldots, y_k)$ near p in $\mathcal{U}$ and new coordinates $(y_{l+1}, \ldots, y_m)$ replacing $x_{l+1}, \ldots, x_m$ such that the map of the previous paragraph takes the form

$$(y_1, \ldots, y_k) \to (y_{l+1}, \ldots, y_m)$$

Here $m - l \leq k$. By a translation, we can further assume that the point $(0, \ldots 0)$ in the old coordinates maps to $(0, \ldots, 0)$ in the new coordinates.

A point in $\mathcal{U}$ belongs to $K \cap L$ just in case the original last $m-l$ coordinates were zero, and thus just in case the new $m-l$ coordinates are zero. So a point $(y_1, \ldots, y_k)$ in $\mathcal{U} \subset K$ is in $K \cap L$ just in case it's first $(m-l)$ coordinates are zero. This leaves $k-(m-l) = k+l-m$ coordinates, providing a coordinate neighborhood for $p \in K \cap L$. QED.

6.7 The Thom Transversality Theorem

We now come to the main theorems of this chapter. We'll begin with a special case which contains the germ of all later variants:

Theorem 28 *Let K and L be (embedded) submanifolds of R^n. For any $\epsilon > 0$, there is a vector $a \in R^n$ with $||a|| < \epsilon$ such that $K + a$ and L are transverse.*

Remark: Here $K+a$ means K translated by a. So the theorem says that after an arbitrarily small translation of K, this translation and L will intersect transversally.

We first sketch the proof without details. Imagine that L is fixed, but K is injected into R^n by a map f. Let A be the open ball in R^n of radius ϵ and let $g : K \times A \to R^n$ be the map $g(p, a) = f(p) + a$. Here the ball A represents *deformations* of f, and g is a sort of "bundle of deformed versions of f". To emphasize this point of view, define $f_a : K \to R^n$ for fixed $a \in A$ to be the map $K \to K \times a \to R^n$.

Clearly $g^\star : T_p(K \times A) \to T_p(M)$ is onto for each p, since fixing the K component and restricting the map to the second component is locally a diffeomorphism. It follows that $g^{-1}(L)$ is a submanifold $W \subset K \times A$. The dimension of this submanifold is *dim* K + *dim* L.

Consider the map $\rho : W \to A$ defined by projecting $W \subset K \times A$ to A. The space $\mathcal{W}$ consists of all intersection points of $K + a$ and L for all possible small translations a; we will prove that such an intersection point $k + a$ is not transversal if and only if $\rho^\star$ is not onto at that point. So the non-transversal intersection points correspond to singular points of ρ and the image of such singular points corresponds to translations for which $K + a$ and L intersect non-transversally. By Sard's theorem, these points have measure zero in A and therefore any open subset of A contains translations a for which $K + a$ and L only have transversal intersection points.

Example 1: Let K by the parabola $y = x^2$ and L be the line $y = 0$ in R^2. Notice that these meet non-transversely at the origin.

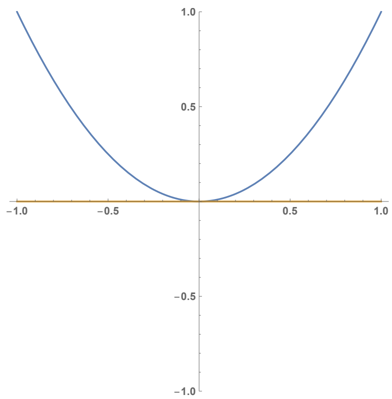

Figure 6.6: K and L

The map $g : K \times A \to R^2$ is given by $(t, t^2, a_1, a_2) \to (t + a_1, t^2 + a_2)$ and the inverse image of L consists of points where the last coordinate $t^2 + a_2 = 0$, so $a_2 = -t^2$. Thus $\{(t, t^2, a_1, -t^2)$ is exactly the set of intersection points of K and L for various translations (a_1, a_2). Note that no such points occur if $a_2 > 0$, so raising the parabola will make K and L transverse. On the other hand, if we lower the parabola by t^2, then $(\pm t, t^2)$ map to the two corresponding intersection points, both transverse. We can also translate left and right by a_1, but this leads to the same intersection points $(\pm t, t^2)$ now translated left or right and then down.

Eample 2: Let K be the cubic $y = x^3$, and L be the line $y = 0$ in R^2.

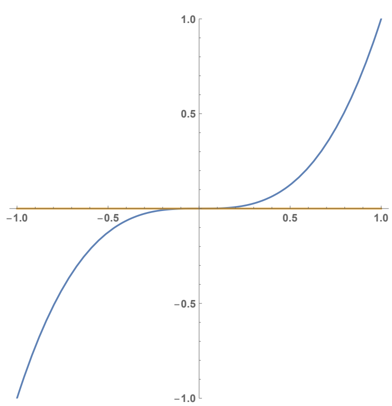

Figure 6.7: K and L

Then $g = (t, t^3, a_1, a_2) \to (t + a_1.t^3 + a_2)$. so $\mathcal{W} = \{(t.t^3, a_1, -t^3)\}$. This time there are intersection points when we raise or lower K. If we lower by t^3, there is one intersection point (t, t^3). We can also move left or right, but this remains the only intersection point after left or right translation and moving up or down.

Example 3: Let K be the cubic $y = x^3 - x$, and L be the line $y = 0$ in R^2.

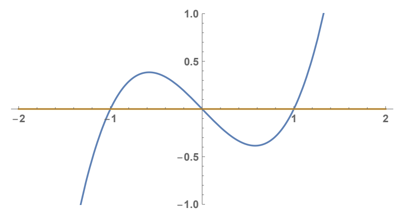

Figure 6.8: K and L

Now $g : (t, t^3 - t, a_1, a_2) \to (t + a_1, t^3 - t + a_2)$ and $\mathcal{W} = \{(t, t^3 - t, a_1, -t^3 + t)\}$. This time the intersection points are given by $a_2 = t^3 - t$. So if we raise or lower the curve K by a_2, we must solve for t to find intersection points; there may be one, or two, or three. These intersection points are non-transversal exactly at the local maximum and minimum of K, and thus when $3t^2 - 1 = 0$ or $t = \pm\frac{1}{\sqrt{3}}$.

The key step of the argument consists in showing that whenever $a \in A$ is a regular value of ρ, F_a is transversal to L. By Sard's theorem, there is a set of measure zero in A such that every $a \in A$ not in this set is a regular value. In particular, each open subset of A contains regular values. QED.

There are many results here that need checking and we'll do that shortly. But first a picture which explains what is going on. Suppose that K is the cubic $y = x^3$ in the plane, and L is the horizontal x-axis. These submanifolds meet non-transversally at the origin. The let $g : K \times A \to R^2$ sends $(t, a, b) \to (t, t^3) + (a, b) = (t + a, t^3 + b)$. Then $\mathcal{W} = g^{-1}(L)$ is the set of all (t, a, b) with $t^3 + b = 0$, and thus $b = -t^3$. Notice that we must calculate the regular values of ρ to determine whether or not the intersections are transverse.

Remark: We are now ready to fill in the details of the argument. However, we are going to generalize the theorem before doing so, since the more general theorem has exactly the same proof and since it leads to important results in the last section of this chapter.

Previously, we had submanifolds K and L of M with L fixed. We varied K by varying the one-to-one immersion $f : K \to M$. In the following generalization, we concentrate on the map and drop the condition that it be an immersion.

Definition 14 *Suppose K is a k dimensional manifold and $f : K \to M$ is a smooth map. Suppose $L \subset M$ is an embedded l dimensional submanifold. We say the map f is transverse to L is for each $p \in K$, $f^\star(T_p(K)) + T_p(L) = T_p(M)$.*

Remark: Previously when K was immersed in M, the first term of this last requirement was replaced by $T_p(K)$, where f was implicit and an immersion so $f^\star$ was one-to-one. In this definition, $f^\star$ might not be one-to-one, so $f^\star(T_p(K))$ does not necessarily have dimension $dim(K)$.

Theorem 29 *Let L be an (embedded) submanifold of a manifold M. Let K and A be other smooth manifolds, and suppose $g : K \times A \to M$ is a smooth map. Suppose $g^\star$ is always onto $T(M)$. For each fixed $a \in A$, define $g_a : K \to M$ to be $g(k, a)$. Then g_a is transversal to L except on a set of measure zero of A.*

Proof: Let $\mathcal{W} = g^{-1}(L)$. We claim $\mathcal{W}$ is a submanifold of $K \times A$. Indeed since g is a submersion, we can find local coordinates on $K \times A$ and M such that the map g is

$$(x_1, \ldots, x_s) \to (x_1, \ldots, x_m)$$

where $s \geq m$. Then whether or not a point in $K \times A$ maps to a point of L depends only on the first m coordinates of that point. By the immersion theorem, we can find new coordinates $(y_1, \ldots, y_m) \to (x_1, \ldots, x_m)$ so that L intersects this coordinate system exactly in $(y_1, \ldots, y_l, 0, \ldots, 0)$. It follows that $\mathcal{W}$ will be all $(y_1, \ldots, y_l, 0, \ldots, 0, x_{m+1}, \ldots, x_s)$ and thus have dimension $dim(K) + dim(A) - (m - l) = k + l + dim(A) - m$.

Let $\rho : \mathcal{W} \to A$ be the map $\mathcal{W} \subset K \times A \to A$ where the last piece is just projection onto A. For each fixed $a \in A$, let $g_a : K \to M$ be the map $K \to K \times A \to M$ where the first map sends k to $k \times a$. We claim that if a is a regular value of ρ, then g_a is transversal to L. If so, we are done, because by Sard's theorem the set of non-regular values of ρ has measure zero in A.

Therefore, suppose $a \in A$ is a regular value and $p = k \times a \in \mathcal{W}$ maps to this value. Let $v \in T_p(K \times A)$. We can write $v = v_K + v_A$ where v_K is tangent to K and v_A is tangent to A. Since a is a regular value of ρ, there is a vector w tangent to $\mathcal{W}$ at a and mapping to v_A. Write $w = w_K + v_A$. Then $v = w + (v_K - w_K)$ and so $T_p(K \times A) = T_p(\mathcal{W}) + T_p(K \times a)$.

But the first paragraph of the proof shows (using coordinates, no less!) that $T_p(L) + g^\star(T_p(K \times A)) = T_p(M)$. By the result in the previous paragraph,

$$T_p(L) + g^\star(T_p(\mathcal{W})) + g^\star(T_p(K \times a)) = T_(M)$$

The second term on the left is inside the first term, and the third term is $g_a^\star(T_p(K))$. So we get the equation of transversality for g_a at k:

$$T_p(L) + g_a^\star(T_p(K)) = T_p(M)$$

QED.

6.8 Transversal Homotopy Theorem

Theorem 30 *Let K and L be embedded submanifolds of a compact manifold M. Let $f : K \to M$ be the inclusion map. Then f is homotopic to a map which is transversal to L and still an embedding of K into M. The transversal K can be selected arbitrarily close to the original K.*

Proof: Since M is compact, an easy argument shows that M can be embedded in R^n. Consider the normal vector bundle $E \to M$. Each normal vector is a vector in R^n and E consists of pairs (p, n) where $p \in M$ and n is a vector in R^n orthogonal to $T_p(M)$. Define a map $f : E \to R^n$ by $f(p, n) = p + n$.

Clearly $f^\star : T_p(E) \to T_p(R^n)$ is an isomorphism on the zero section of E. By the inverse function theorem, locally near each $p \in M$ it is a diffeomorphism to an open neighborhood of R^n. In particular, it is a diffeomorphism on an open subset of M when restricted to normal vectors of length less than ϵ. By compactness of M, we can cover M by a finite number of such open sets and find a uniform ϵ which works for each of these open sets. In this way, a "tube" around the zero vectors consisting of all (p, n) with $p \in M$, n normal to M at p, and $||n|| < \epsilon$ can be constructed, and this tube maps to a similar tube N_ϵ that is an open set in R^n containing M. The map is a local diffeomorphism which is onto but might conceivably not be one-to-one. It is possible that some point can be reached from both (q, v) and (r, w) where q and r are in different neighborhoods in M.

We next claim that by shrinking ϵ, we can avoid this possibility. If not, then we can find a sequence of ϵ_n converging to zero, and (q_n, v_n) and (r_n, w_n) with $q_n \neq r_n$ and $||v_n|| < \epsilon_n$ and $||w_n|| < \epsilon_n$ and $q_n + v_n = r_n + w_n$. By selecting subsequences, we can assume that $q_n \to q, v_n \to 0, r_n \to r, w_n \to 0$. Recall that M is covered by a *finite number* of open sets where our map is a diffeomorphism. The point q is in one of these sets, and so eventually all q_n are in the set. Then r cannot belong to the set, since otherwise r_n is eventually in the set, but our map is a diffeomorphism on points with base in the same open set. So $q \neq r$. But $q_n + v_n \to q$ and $r_n + w_n \to r$ and $q_n + v_n = r_n + w_n$, a contradiction.

Consequently, we have a diffeomorphism from the tube consisting of points in the normal bundle with normal vector of length less than ϵ and the image tube in R^n.

One warning is in order, though. The normal bundle need not be trivial, so this tube is not necessarily diffeomorphic to $M \times B$ where B is the standard open ball. In the remaining parts of the proof, notice how we carefully avoid making that assumption!

Note that there is a standard submersion π : tube in $E \to M$ and corresponding submersion we also call $\pi : N_\epsilon \to M$.

Let B be the open ball of radius ϵ in R^n and define a map $g : K \times B \to M$ by $g(k, b) = \pi(k + b)$. Notice that b is an arbitrary vector in R^n, not necessarily normal to M. We claim that the point $k+b$ will be inside N_ϵ, and thus the map π will be defined. Indeed, since b has length less than ϵ, the point $k+b$ has distance in R^n less than ϵ from K. But N_ϵ contains all such points, because if p is a point with distance less than ϵ from K, then using compactness of K we can find a point on k closest to p, and then by calculus the line from k to p is normal to K, so p is in N_ϵ.

Clearly π is a submersion, since the restriction to B maps to the full R^n and thus onto $T_p(M)$ by the submersion π. So we can apply the previous transversality theorem, and deduce that $g(k,a)$ is a submanifold transverse to L for $a \in B$ not in a set of measure zero. This map is homotopic to the original identity map via the homotopy $\pi(k+ta)$ for $0 \le t \le 1$. We have to prove one other item, namely that $g(k,a)$ is an immersion for fixed a if a is sufficiently small.

Note that $g(k,0)$ is an immersion by assumption. Choose local coordinates $(x_1,\ldots,x_k)$ on K and consider $g^\star\left(\frac{\partial}{\partial x_1}\right),\ldots,g^\star\left(\frac{\partial}{\partial x_k}\right)$ at (k,a) as a varies. By continuity, these remain linearly independent near $(k,0)$, so we can choose ϵ for the tubular neighborhood near k so they are linearly independent within the tube. By compactness of K, we can choose a common ϵ so these are linearly independent within the entire tube, so g_a is always an immersion. Since K is compact, K will remain an embedded submanifold provided only that g_a is one-to-one for sufficiently small a. If this is not true, then we can find a sequence a_n converging to zero, and sequences e_n, f_n in K with $e_n \ne f_n$ and $g_{a_n}(e_n) = g_{a_n}(f_n)$. By compactness of K, we can assume e_n converges to $e \in K$ and f_n converges to f in K. Then continuity gives $g_0(e) = g_0(f)$. By assumption, g_0 is one-to-one, so $e = f$.

Note that e_n and f_n are in K near $e = f$, and thus lie in a common coordinate system $(x_1,\ldots,x_k)$. Since $e_n \ne f_n$, there is an i such that $(e_n)_i \ne (f_n)_i$ and

$$|(e_n)_j - (f_n)_j| \le |(e_n)_i - (f_n)_i|$$

Because i varies from 1 to k, there is at least one i which satisfies this inequality infinitely often. Fix this i, throw away other terms in the sequences, and renumber the rest; then we can assume that the inequality is always true.

By the mean-value theorem,

$$0 = \frac{g^i_{a_n}(e_n) - g^i_{a_n}(f_n)}{(e_n)_i - (f_n)_i} = \sum_j \left(\frac{\partial g^i_{a_n}}{\partial x_j}\right) \frac{(e_n)_j - (f_n)_j}{(e_n)_i - (f_n)_I}$$

Here the partial derivatives are evaluated at interior points very close to $e = f$. But at $e = f$ we have

$$\frac{\partial f^i}{\partial x_j} = \delta_{ij}$$

Therefore in the limit as n goes to infinity, the expression on the right goes to 1, a contradiction. QED.

Remark: We are going to use this result in our proof of the Lefshetz Fixed Point theorem. In that case we have a compact oriented manifold M and two graphs in $M \times M$: the graph of a map $f : M \to M$ and the diagonal graph of the map $id : M \to M$. Both are submanifolds of dimension m in $M \times M$, which has dimension $2m$.

The graph of f is given by an embedding $g : M \to M \times M$ by $g(x) = x \times f(x)$. By the previous theorem, we can find an arbitrarily small homotopy $h(u)$ from g to g_1 such that g_1 is transversal to the diagonal. But this homotopy might change the first component of g. If so, we have deformed *the submanifold of* $M \times M$ given by our original graph, but this isn't the same thing as deforming *our original map itself* and then considering the corresponding deformed submanifold.

We now claim that with a little care we can do both things at once. In the previous proof we studied an inclusion map $f : K \to M$ and introduced $g : K \times B \to M$. The special case of interest for the Lefshetz theorem starts with replacing K by M and replacing M by $M \times M$, and replacing f by $i(p) = p \times f(p)$ and replacing g by $g : M \times B \to M \times M$. This time we will embed $M \times M$ in R^n and consider a tube in the normal bundle E to $M \times M$ inside $R^n \times R^n$. Then $g(k, b) = \pi(k + b)$ will be appropriately replaced. Our previous proof then used the fact the every image is a regular point except for a set of b's of measure zero. We restricted attention to sufficiently small b's making other desired things true; but we can still find a b not in the bad set of measure zero.

In our present case, we do one more thing. We consider the map $M \to M \times M \to M$ by projecting on the first component. We claim that if we restrict to sufficiently small b's, then all such maps are diffeomorphisms. If so, we can compose our maps with the inverse of this map to obtain new maps $M \to M \times M$ in which the first component of the new map is the identity, so we are truly deforming functions from M to M and the graphs are just following along.

However, the argument that for all sufficiently small b these maps are diffeomorphisms is exactly the same as our earlier argument that $g(k, a)$ is an immersion for fixed a if a is small enough, so we will not repeat that argument again.

Remark: What condition on $f : M \to M$ makes the graph of f transversal to the diagonal in $M \times M$? Suppose p is a fixed point of f and thus an intersection point of the graph of f with the diagonal. A typical tangent vector to the graph has the form $X \times f_\star(X)$ where X is tangent to M. A typical tangent vector to the diagonal has the form $X \times X$. We must show that the sum of the two resulting subspaces of $T_{p \times p}(M \times M)$ give the entire subspace. Since both subspaces have dimension m and the full subspace has dimension $2m$, a necessary and sufficient condition for transversality in $M \times M$ is that the two subspaces intersect only at the zero vector. In other words, whenever $X \neq 0$, $X \neq f^\star(X)$. But this condition just says that 1 is not an eigenvalue of $f^\star$ and thus that $\det(I - f^\star) \neq 0$.

Chapter 7

Cohomology and Intersections

7.1 Introduction

To finish the proof of the Lefshetz Fixed Point Theorem, it suffices to prove that whenever M is compact and oriented, and K and L are embedded submanifolds which meet transversally, then the cohomology classes dual to K, L, and $K \cap L$, are related by

$$d_{K \cap L} = d_K \wedge d_L$$

To prove this, we need to understand d_K at a deeper level. In this section, we'll try some simple examples. In the examples which follow, suppose that M is a 2-dimensional manifold.

Consider first the case when K is a single point p. We inject this point into the manifold by $p \to M$ and the pullback of this map on 0-forms sends f to $f(p)$. This represents the element in $Hom(H^0(M), R)$ induced by the point. The dual element in $H^2(M)$ is an element $d_K \in H^2(M)$ represented by a 2-form. Moreover, $H^0(M)$ to R is given by $\xi \to \int \int_M \xi \wedge d_K$. Translating

$$f(0) = \int \int_M f \wedge d_K = \int \int_M f(m) d_K(m)\ dm$$

This statement should hold provided f represents an element of $H^0(M)$ and thus is a locally constant function. But we might hope that it holds more generally, for all f. A little thought shows that it holds this generally if and only if $d_K(m) = \delta(m) dx \wedge dy$ and δ is the Dirac delta function with singularity at p. This is illuminating since the delta function is geometrically related to K, the point p.

If M is connected and f is locally constant, then actually we can choose for d_K any two-form with integral equal to one. All such choices are equivalent, since the integral gives an isomorphism between $H^2(M)$ and R. So on the cohomological level, the connection between K and d_K is very tenuous.

Let's look at these same ideas one dimension higher. Suppose K is locally a line in R^2. The global picture here might be a closed curve on a torus or surface of higher genus. For concreteness, suppose M is a torus and work on the universal cover, i.e., the plane with standard $Z \times Z$ lattice. We'll imagine that K is the straight line along the x-axis from $(0,0)$ to $(1,0)$. This submanifold induces a map $H^1(M) \to R$ which is locally given by $\int_R \omega$. Here ω is a 1-form on R^2. so $\omega = \omega_x dx + \omega_y dy$. Thus our map is $\int_0^1 \omega_x(t,0)dt + 0$.

The corresponding dual element should be represented by a second 1-form on R^2 with the property that

$$\int_R \omega = \int \int_{R^2} \omega \wedge d_L$$

Written more concretely, this formula says

$$\int_0^1 \omega_x(t,0)\ dt = \int_0^1 \int_0^1 (\omega_x(t,u) d_{K,y}(t,u) - \omega_y(t,u) d_{K,x}(t,u))\ dtdu$$

This will be true if

$$d_{K,y}(t,u) = \delta(u) \text{ and } d_{K,x} = 0$$

So on the form level, d_K should be a delta function (or form) associated with the line K. More generally, d_L could be a one form which is zero except in a small tubular neighborhood of the line L.

I used to own a book by deRham in which he extended the deRham cohomology to *currents*, which he defined as distribution-valued forms. Someone borrowed that book and didn't return it. Conceivably, these currents could be used to give very explicit formulas for d_K, d_L, and $d_{K \cap L}$, but I don't know if deRham had that application in mind.

We do not have distributions, but we will prove an analogous result: the element d_K can be represented by an appropriate form which is non-zero in an arbitrarily small tubular neighborhood of K. This will be done in two stages. We will prove an *infinitesimal* form of the theorem, replacing the tubular neighborhood by a neighborhood of the zero section in the normal bundle of K. The tubular result will quickly follow.

7.2 Tubular Neighborhoods

Suppose $K \subset M$ is a submanifold of a compact manifold M. Assign a Riemannian metric to M. Let $N \to K$ be the normal bundle, so if $p \in K$ then N_p is the set of tangent vectors X to M at p such that X is orthogonal to $T_p(K) \subset T_p(M)$.

Define a map $T : N \to M$, called *the tubular map*, as follows: Suppose $(p, X) \in N$, so $p \in K$ and X is a normal vector to K at p. Let $\gamma(t)$ be the unique geodesic on M such that $\gamma(0) = p$ and $\gamma'(0) = X$. Define $T(p, X) = \gamma(1)$. This definition makes sense because geodesics on a compact M exist for all t.

There is another way to define this map that is slightly more intuitive. Recall that all geodesics move at constant speed. If $\gamma(t)$ is a geodesic and s is a fixed real number, then $\gamma(st)$ is also a geodesic which moves s times as fast as $\gamma(t)$. It follows that we can define $T(p, X)$ by finding the unique geodesic $\gamma(t)$ through p such that $\gamma(0) = p$ and $\gamma'(0) = \frac{X}{||X||}$. This geodesic moves in the direction of X at constant speed 1. Define $T(p, X) = \gamma(||X||)$. In other words, *follow a geodesic with constant speed 1 in the direction of X for a distance equal the length of X.*

Theorem 31 *There is a positive constant c such that if $\epsilon < c$ and we restrict the tubular map to vectors of length less than ϵ, it is a diffeomorphism from the open set of normal vectors of length less than ϵ to an open neighborhood of $K \subset M$.*

Remark: Such an open neighborhood of K is called a *tubular neighborhood of K*. Notice that it is locally diffeomorphic to $K \times B(\epsilon)$ where $B(\epsilon)$ is the open ball of radius ϵ in $R^{dimM - dimK}$.

Proof: We first prove this theorem locally. If $p \in K$, we can find local coordinates $(x_1, \ldots, x_m)$ near p on M such that K is the set of such points where $x_{k+1} = 0, \ldots, x_m = 0$. Choose an orthonormal basis of normal vectors $N^1(x_1, \ldots, x_k), \ldots, N^{m-k}(x_1, \ldots, x_k)$ on K near p. If (p, X) is a point in the normal bundle, we can describe p by coordinates $x_1, \ldots, x_k$ and we can describe X by numbers $t_1, \ldots, t_{m-k}$ such that $X = \sum t_i N^i$. Notice that each N^i is a vector tangent to M and thus in coordinates equals

$$N^i = \sum_{j=1}^{m} N^i_j \frac{\partial}{\partial x_j}$$

Since the normal vectors are only defined on K, the N^i_j are functions of $x_1, \ldots, x_k$.

We want to find a geodesic $\gamma(t)$ such that $\gamma(0) = p$ and $\gamma'(0) = X$. In coordinates, this geodesic is given by $\gamma_1(t), \ldots, \gamma_m(t)$ such that

$$\frac{d^2\gamma_i(t)}{dt^2} + \sum \Gamma^i_{uv}(\gamma(t)) \frac{d\gamma_u}{dt} \frac{d\gamma_v}{dt} = 0$$

The condition $\gamma(0) = p$ says that $\gamma_i(0) = x_i$ for $1 \le i \le k$ and $\gamma_i(0) = 0$ for $(k+1) \le i \le m$. The condition $\frac{d\gamma}{dt}(0) = X$ says that in coordinates $X = (t_1, \ldots, t_{m-k})$ and

$$\frac{d\gamma_i}{dt}(0) = \sum_{j=1}^{m-k} t_j N^j_i(x_1, \ldots, x_k)$$

The existence theorem for systems of ordinary differential equations says that these equations have a solution defined for $|t| < \delta$. These solutions depend on the boundary conditions, so γ_i is a function of $t, x_1, \ldots, x_k, t_1, \ldots, t_{m-k}$. A sharper form of the existence theorem says that we can find $\delta > 0$, an open neighborhood $\mathcal{V}$ of p, and $\tau > 0$ such that such solutions are defined for $|t| < \delta$, $(x_1, \ldots, x_k) \in \mathcal{V}$, and $\sum t_j^2 < \tau$. Finally it says that these functions are C^∞ in t, x_i, t_j.

Recall that if $\gamma(t)$ is a geodesic with $\gamma'(0) = X$, and if s is a constant, then $\gamma(st)$ is also a geodesic with $\gamma'(0) = sX$. In other words, geodesics are traced with constant speed, but if we change this speed, we still have a geodesic.

It follows that $\gamma(\frac{\delta t}{2})$ is also a geodesic which starts at $(x_1, \ldots, x_k)$ when $t = 0$, but is defined for $|\frac{\delta t}{2}| < \delta$, and thus for all $-2 < t < 2$. The tangent vector of this solution at $t = 0$ is $\frac{\delta}{2}$ times the original tangent vector, and so $\frac{\delta}{2}$ times a vector of norm at most τ. So its norm is at most $\frac{\delta \tau}{2}$. So if we shrink the lengths of the initial normal vectors, we can define the geodesic at least until time t. It follows that $\gamma(1)(x_1, \ldots, x_k, t_1, \ldots, t_{m-k})$ is defined and C^∞ for $(x_1, \ldots, x_k) \in \mathcal{V}$ and for $(t_1, \ldots, t_{m-k})$ of sufficiently small norm. This is exactly the tubular map we originally defined.

We now claim this map is a diffeomorphism if we shrink $\mathcal{V}$ and the norm of the normal vector determined by $(t_1, \ldots, t_{m-k})$ sufficiently. This will follow from the inverse function theorem if we can prove that the matrix of partial derivatives with respect to x_i and t_j has non-zero determinate at $(p, 0)$.

Since what we are about to do can be confusing, let's summarize it very carefully. We start with an initial point $(p, 0)$. If $\hat{x_i}$ denotes some fixed value of x_i, this initial point is

$$(\hat{x_1}, \ldots, \hat{x_k}, 0, \ldots, 0)$$

We fix all of these initial coordinates except one, which we allow to vary. For instance, perhaps in the initial condition we vary x_i around $\hat{x_i}$. Or we might let t_j vary around 0. We then compute the geodesic with this initial condition and find its value at $t = 1$. This value will be new coordinates $(\tilde{x}_1, \ldots, \tilde{x}_k, \tilde{y}_1, \ldots, \tilde{y}_{m-k})$ which all depend only on the varying initial coordinate. We then take the partials of these functions with respect to that initial coordinate.

If the varying coordinate is x_i, then the calculation is very easy. Since N is fixed at 0, the geodesic is just a constant, and so it ends where it started at $(\hat{x_1}, \ldots, x_i, \ldots \hat{x_k}, 0, \ldots, 0,)$. As predicted, all of these are functions of x_i, but actually all these functions except one do not depend on x_i. So all the partials with respect to x_i are zero except one.

Long ago we called our tubular map T. We have just shown that $\frac{\partial T^j}{\partial x_i} = \delta_{ij}$ where j ranges over all m possibilities, but i only ranges over the first k possibilities.

The remaining calculation is slightly trickier. This time we let y_j vary around 0, and form the geodesic starting at $(\hat{x}_1, \dots, \hat{x}_k, 0, \dots, t_j, \dots, 0)$. At time $t = 1$, this geodesic ends at $(\tilde{x}_1, \dots, \tilde{x}_k, \tilde{x}_{k+1}, \dots, \tilde{x}_{m-k})$ and all of these coordinates could depend on t_j. We again want to compute partials.

First, let's sketch the idea. The variable y_j will be very close to zero, and therefore the geodesic will move very slowly. So we should be able to approximate it by its linear approximation. We have

$$\frac{d\gamma_i}{dt}(0) = \sum_{s=1}^{m-k} t_s N_i^s(x_1, \dots, x_k) = t_j N_i^j(x_1, \dots, x_k)$$

and so

$$\gamma_i(t) \sim \gamma_i(0) + t\ t_j N_i^j(x_1, \dots, x_k)$$

and

$$T^i = \gamma_i(1) \sim \gamma_i(0) + t_j N_i^j(x_1, \dots, x_k)$$

The partial derivative of this expression with respect to t_j is $N_i^j(\hat{x}_1, \dots, \hat{x}_k)$. Here i goes from 1 to m, but j only goes from 1 to $m-k$. The vectors N_i are normal to K and linearly independent; the k vectors $(1, 0, \dots, 0), (0, 1, \dots,), \dots$ pointing in the first k directions form a basis for the tangent space to K, so these vectors together with $N^1, \dots, N^{m-k}$ form a basis for the full tangent space to M. Hence the determinant of the coordinates of these vectors is non-zero, and the inverse function theorem applies.

We need only make the above calculation rigorous. This involves an easy trick. Fix a normal vector N and let $\gamma_N(t)$ be the unique geodesic such that $\gamma_N(0) = p$ and $\gamma_N'(0) = N$. Then $T(p, N) = \gamma_N(1)$. Next consider the curve $\tau(t) = \gamma_N(t_j t)$, where for the moment t_j is just a fixed number. Then $\tau(0) = p$ and $\tau'(0) = t_j \gamma_N'(0) = t_j N$. So $T(p, t_j N) = \tau(1) = \gamma_N(t_j)$. It follows that $T(p, t_j N^j) = \gamma_{N_j}(t_j)$ and the partial derivative of these equal expressions with respect to t_j at $(p, 0)$ is N^j. This is exactly the result our previous calculation gave.

7.3 Completion of the Proof of the Tubular Neighborhood Theorem

In the previous section, we proved that for any $y \in K$ we can find a neighborhood $\mathcal{V}$ of p in K and an $\epsilon > 0$ such that the tubular map restricted to vectors in N of length less than ϵ is a diffeomorphism onto an open set in M. The resulting $\mathcal{V}$ cover K, which is compact since M is compact. So we can find a finite subcover, and select the minimum of the various ϵ attached to each set of the subcover. In the end we get a map T from all normal vectors to K of length less than ϵ, which maps this set in a C^∞ manner to an open neighborhood of K in M. To finish the tubular neighborhood theorem, we need only show that this map is one-to-one if ϵ is small enough.

If the result we want to prove is false, then we can find sequences (p_n, A_n) and (q_n, B_n) with $p_n \in K, q_n \in K, p_n \neq q_n$, and A_n and B_n vectors normal to K at p_n and q_n, $||A_n|| \to 0$ and $||B_n|| \to 0$, and $T(p_n, A_n) = T(q_n, B_n)$. Recall that M has a Riemannian metric, so we can compute the length of any tangent vector, and so of vectors normal to K.

Using the compactness of K, we can find a convergent subsequence of a_n. This yields a subsequence of the b_n and we can find a convergent subsequence of this sequence. So without loss of generality, we can assume that $a_n \to a_0$ and $b_n \to b_0$. (There are many ways to get these sequences. For example, cover M by a finite number of coordinates corresponding to balls in R^k of radius 2, such that the subsets corresponding to balls of radius 1 also cover. Then infinitely many p_n must be in an open set corresponding to one of the balls of radius 1, and so a subsequence converges to a point in the closure of this ball, and thus strictly inside the ball of radius 2.)

By continuity of T, we have $T(p_n, A_n) \to T(p_0, 0) = p_0$ and $T(q_n, B_n) \to T(q_0, 0) = q_0$. Since $T(p_n, A_n) = T(q_n, B_n)$, $p_0 = q_0$. But T is one-to-one on an open neighborhood of p_0, provided we restrict to normal vectors of norm below a fixed ϵ. This contradicts the assumption that $p_n \neq q_n$ and $T(p_n, A_n) = T(q_n, B_n)$. QED.

7.4 Vector Bundles

Recall that a vector bundle of dimension n over a manifold M is an assignment to each $p \in M$ of an n-dimensional vector space E_p, together with a topology and C^∞ structure on the union E of these E_p. This structure is required to be locally-trivial in the following sense: each $p \in M$ has an open neighborhood $\mathcal{U}$ with local coordinates $x_1, \ldots, x_m$ such that over this open set we can find vector fields $E_1, \ldots, E_n$, each a C^∞ map $M \to E$ with $E_i(q) \in E_q$ for each $q \in \mathcal{U}$, so that the $E_i(q)$ form a basis of E_q for each $q \in \mathcal{U}$.

Let $\pi : E \to M$ be the obvious map. Once we have a locally trivial structure over $\mathcal{U}$, we can assign coordinates to $\pi^{-1}\mathcal{U}$ of the form $x_1, \ldots, x_m, t_1, \ldots, t_n$. The first m coordinates determine a point $p \in M$ and the last n coordinates determine a vector $\sum t_i E_i(p)$ in E_p.

In that case, the change of coordinate map has a special form. Suppose $y_1, \ldots, y_m, s_1, \ldots, s_n$ are new coordinates over $\mathcal{V}$ obtained by choosing new basis fields $F_1, \ldots, F_n$. On $\mathcal{U} \cap \mathcal{V}$ we can write y_j as functions of the x_i; by abuse of notation, we write $y_j(x_1, \ldots, x_m)$. But then $\sum t_i E_i$ and $\sum s_j F_j$ are related by a matrix ρ_{ji}; when the column vector formed by the t_i is multiplied by this matrix, we obtain the column vector formed by the s_j. Note that the matrix elements depend on the base point, so each ρ_{ji} is a function of $x_1, \ldots, x_m$.

In a previous section we used the normal bundle to define a tubular neighborhood. The definition of the tubular map did not require selecting a local basis for this bundle, although such a basis was implicit in our proof that the map is a local diffeomorphism near the zero section.

But the change of coordinate formula for the normal bundle becomes more important in the following sections, so it is useful to sketch how it could be computed. Suppose $K \subset M$ is a submanifold of a manifold M. The normal bundle is a bundle over K, where each vector space N_p consists of tangent vectors to p in M which are perpendicular to the tangent space of K at p.

To compute this in local coordinates, we can choose coordinates $x_1, \ldots, x_m$ on M near p so that K is the set where $x_{k+1} = 0, \ldots, x_m = 0$. Then the $\frac{\partial}{\partial x_i}$ form a basis of $T_p(M)$ and the first k of these form a basis for $T_p(K)$. Apply the Gram-Schmidt process to these m vectors. The first k vectors from the process are irrelevant, but the remaining vectors $N_1, \ldots, N_{m-k}$ form an orthonormal basis for the set N_p of normal vectors near p. So we can introduce coordinates on the normal bundle as $x_1, \ldots, x_k, t_1, \ldots, t_{m-k}$ where the first k coordinates determine $p \in K$ and the remaining coordinates determine a normal vector $\sum t_i N_i$ at p.

If we had two coordinate systems of this type, we could compute the matrix $\rho_{ji}(x_1, \ldots, x_k)$ which maps the vector with coordinates $t_1, \ldots, t_{t-k}$ in the first coordinate system to the vector $s_1, \ldots, s_{m-k}$ in the new coordinate system. Details for the computation of this ρ are not important for us, and indeed ρ itself will be hidden away in our proofs. But if you completely forget about it, then our arguments may seem unreasonably abstract.

7.5 Vector Bundles and the Thom Space

Suppose $E \to M$ is a vector bundle with vector spaces of dimension n. In his thesis, Thom introduced an associated space now called the *Thom space*, $T(E)$. First replace each fibre E_p by its one-point compactification. Another way to think of this is that we replace each fibre $E_p \cong R^n$ with the corresponding sphere S^n. Each of the new fibers has a special point, the point at infinity. The Thom space is formed by identifying all of these special points. Thus it is a space with base point.

If E is oriented, Thom proved that there is an isomorphism

$$H^\star(M) \cong \tilde{H}^{\star^+ n}(T(E))$$

Notice that we have reduced cohomology on the right, so the zeroth cohomology group there is cancelled out. The isomorphism is given by forming a nth cohomology class Φ on $T(E)$ defined by the property that it induces the fundamental class on each fibre S^n, and then sending $\sigma \to \sigma \wedge \Phi$.

Consider the easiest case of this construction when M be the circle and E is the trivial line bundle over M. If we replace each fibre R in E by the sphere S^1, we get a torus $S^1 \times S^1$.

Below is a picture of this torus on the left. Imagine that the points at infinity form the inner circle on this torus; identifying these points gives the Thom space $T(E)$ on the right. The first cohomology group of a torus has two generators, illustrated on the left below. Only one of them survives and generates the first cohomology group of $T(E)$, as illustrated on the right. Clearly, then, the cohomology of this Thom space is $H^0 = R, H^1 = R, H^2 = R$ and its reduced cohomology is $H^0 = 0, H^1 = R, H^2 = R$. These are the cohomology groups of a circle, shifted up by one.

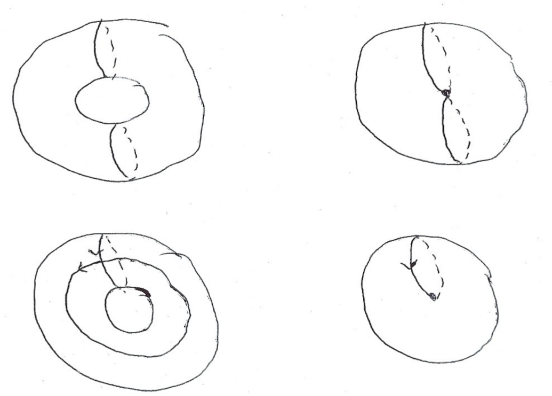

Figure 7.1: Thom Space

Remark: Thom applied his construction to the study of cobordism groups in the 1950's, obtaining spectacular results you can read about elsewhere.

We are going to apply the Thom isomorphism to the normal bundle of a submanifold $K \subset M$. This bundle has dimension $m - k$ and the element Φ belongs to H^{m-k} of the Thom space. Rather than constructing this Thom space, we will construct an analog of its cohomology groups formed by differential forms on N which are non-zero only near the zero section and vanish at infinity. In this cohomology group we will construct Φ.

However, the tubular neighborhood theorem allows us to map a neighborhood of the zero section in the normal bundle to a tubular neighborhood of K, and this map carries Φ to an $m - k$ form on M with support in this tubular neighborhood. We will prove that this element represents the element d_K dual to K. Thus we have found a representative which lives in a small tube about K.

Notice that $d_K \wedge d_L$ then lives is a small tube about $K \cap L$, so we will be close to proving our main result.

In some sense, the description of Φ in N is an infinitesimal form of the dual class, which is then realized by the tubular map. In the final step of the argument, we will prove that $d_K \wedge d_L = d_{K \cap L}$ infinitesimally, and show that this implies our desired result.

7.6 Compact Vertical Cohomology

Assume that $E \to M$ is a C^∞ vector bundle of dimension N. We will soon require that it be oriented, but for now any bundle will do. We do not initially assume that M is compact because we will prove the Thom isomorphism theorem using the Mayer-Vietoris sequence.

We form deRham cohomology groups by restricting to forms on E which have compact support on each fibre. To be specific, if ω is such a form, we require that each point p has an open neighborhood $\mathcal{U}$ on which E is isomorphic to $\mathcal{U} \times R^k$ and ω vanishes on (q, v) for $||v|| > B$. The bound depends on $\mathcal{U}$; we do not assume that there is a uniform B. Denote the resulting differential forms by $\lambda^k_{cv}(E)$ and the corresponding cohomology groups by $H^k_{cv}(E)$.

The Thom construction is approximated by these new cohomology groups. Since the forms defining the groups have compact support, they vanish near the additional point at infinity which Thom added to the fibers. So in some sense, these forms define *reduced* cohomology groups on $T(E)$.

We are going to define a map from $\Lambda^k_{cv}(E) \to \Lambda^{k-N}(M)$ by integration along the fibre. Start with a trivial bundle $E = M \times R^N$, and denote the coordinates on M by $x_1, \ldots, x_n$ and the coordinates on R^N by $t_1, \ldots, t_N$. Each form on E contains terms dx_i and dt_j and we will write the wedge products so the dx_i come before the dt_j. Send all terms to zero except terms that involve all dt_j. Map $w_{i_1 \ldots i_k} dx_{i_1} \wedge \ldots \wedge dx_{i_k} \wedge dt_1 \wedge \ldots \wedge dt_n$ to

$$\left(\int \ldots \int w_{i_1 \ldots i_k}(x_1, \ldots, x_n, t_1, \ldots, t_N) dt_1 \ldots dt_N \right) dx_{i_1} \wedge \ldots \wedge dx_{i_k}$$

This map is independent of the coordinate path on M, for if $y_1, \ldots, y_n$ is a second coordinate system, $dy_k = \sum \frac{\partial y_k}{\partial x_i} dx_i$ and

$$w_{i_1 \ldots i_k}(y_1, \ldots, y_n, t_1, \ldots, t_N) dy_{i_1} \wedge \ldots \wedge dy_{i_k} =$$

$$\sum w_{i_1 \ldots i_k}(y_1(x_1, \ldots, x_n), \ldots, y_n(x_1, \ldots, x_n), t_1, \ldots, t_N) \frac{\partial y_{i_k}}{\partial x_{j_1}} \ldots \frac{\partial y_{i_k}}{\partial x_{j_k}}$$

holds for each $t_1, \ldots, t_N$, so the integrals over $t_1, \ldots, t_n$ are equal.

Next suppose the vector bundle is not trivial. Then we define our map by choosing a partition of unity subordinate to a covering by open sets $\mathcal{U}$ where E is trivial over $\mathcal{U}$. This definition is easily shown to be independent of the choice of partition of unity, once we show that the integral is independent of coordinate changes in the fibers. So suppose $t_1, \ldots, t_N$ and $u_1, \ldots, u_N$ are coordinates for the fibers. Then $t_i = \sum a_{ij}(x_1, \ldots, x_n) u_j$. Note that this coordinate change map does not depend on the t_i. We have

$$w_{i_1 \ldots i_k}(x_1, \ldots, x_n, t_1, \ldots, t_N) dx_{i_1} \wedge \ldots \wedge dx_{i_k} \wedge dt_1 \wedge \ldots \wedge dt_N =$$

$$\sum w_{i_1 \ldots i_k} \left(x_1, \ldots, x_n, \sum a_{1 j_1} u_{j_1}, \ldots, \sum a_{N j_N} u_{j_N} \right) dx_{i_1} \wedge \ldots \wedge dx_{i_k} \det(a_{ij}) du_1 \wedge \ldots \wedge du_n$$

and the integrals with respect to the final variables will be equal provided $\det(a_{ij})$ is positive. So we need to assume the vector bundle is *oriented* in the general case.

Lemma 12 *Integration over the fibres commutes with the d operator, and thus induces a map in cohomology.*

Proof: We must prove that

$$\int dw = d \int w$$

First suppose that w does not contain all of $dt_1 \wedge \ldots \wedge dt_N$. Then the right side is zero; the left side is also zero unless each term is missing at most one dt_i. In that case dw is a sum of terms obtained by differentiation, but the only terms giving a non-zero integral have the form

$$\pm \int \ldots \int \frac{\partial w}{\partial t_i} dt_1 \wedge \ldots \wedge dt_N$$

These terms also give zero because we can integrate first with respect to t_i, and $\int_{-\infty}^{\infty} \frac{\partial w}{\partial t_i}\, dt_i = 0$ because w vanishes for very negative and very positive values of t_i.

In all remaining cases, w contains $dt_1 \wedge \ldots \wedge dt_N$. Then

$$d \int w = \sum \int \ldots \int \left(\frac{\partial w}{\partial x_i} dx_i \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_k} \right) dt_1 \ldots dt_N$$

and $\int dw$ is the same expression.

Remark: It immediately follows that integration over fibers induces a map we will call the *Thom map*

$$T : H^k_{cp}(E) \to H^{k-n}(M)$$

Here n is the dimension of the oriented vector bundle E. If $k < n$, this map sends everything to zero.

Remark: At the end of the next section, we need one more fact, which we prove now:

Lemma 13 *Let τ be an arbitrary form on M. By slight abuse of notation, let τ also indicate the form on E induced by the projection $\pi : E \to M$. If ω is a form on E with compact support along the fibres,*

$$T(\tau \wedge \omega) = \tau \wedge T(\omega)$$

Proof: In coordinates, the form τ depends only on $x_1, \ldots, x_m$ and only involves the basis vectors $dx_1, \ldots, dx_n$. So $\tau \wedge \omega$ is a sum of terms of the form

$$\left(\tau(x_1, \ldots, x_m) dx_{j_1} \wedge \ldots \wedge dx_{j_s} \right) \wedge \left(\omega(x_1, \ldots, x_m, t_1, \ldots, t_n) dx_{j_1} \wedge \ldots \wedge dx_{j_t} \wedge dt_1 \wedge \ldots \wedge dt_n \right)$$

When we apply T, we integrate with respect to $t_1, \ldots, t_n$ and omit $dt_1 \wedge \ldots \wedge dt_n$, and clearly τ does not affect this calculation, so we can wedge before or after integrating. QED.

7.7 The Thom Isomorphism Theorem

Theorem 32 *If E is an oriented vector bundle of dimension n over a compact manifold M, then the Thom map is an isomorphism for all k:*

$$E^k_{cv}(E) \to H^{k-n}(M)$$

Proof: Suppose E is an oriented bundle over a compact M. We earlier proved that M has a good cover. Since the open sets of this cover are diffeomorphic to R^m, they are contractible and so E is a trivial bundle over these open sets. But we can avoid using this theorem from vector bundle theory because our construction of an open cover could clearly be modified to give a good cover by open sets over which E is trivial.

We will prove the theorem by induction over the number of open sets in the good cover. But then the intermediate M in the argument will not be compact. So we prove the theorem more generally for M which possess a finite good cover by open sets over which E is trivial.

To start the induction, we prove the theorem in the special case when $M = R^m$ and the vector bundle is R^n.

Lemma 14 *The map $T : H^k_{cv}(R^n) \to H^{k-n}(M)$ is an isomorphism for $n \geq 1$.*

Proof of lemma: Notice that our map T is a composition of similar maps

$$H^k_{cv}(R^{n-1} \times R) \to H^{k-1}_{cv}(R^{n-1}) \to \ldots \to H^{k-n+1}_{cv}(R) \to H^{k-n}(M)$$

For instance, the map on the extreme left sends most forms on E to zero, but it sends forms ω involving some dx_i and all $dt_1, \ldots, dt_{n-1}$ and dt (i.e., dt_n) to

$$\int_{-\infty}^{\infty} \omega(x_1, \ldots, x_m, t_1, \ldots, t_{n-1}, u)\ du$$

wedged with the same dx_i and with all the dt_j except dt. Each successive map integrates one more t_i and drops one more dt_i until all the t's are gone. In the end, we get T previously defined. It suffices to prove that all of these maps are isomorphisms, and since they all have the same form, it suffices to prove this for the first map. In the proof of this lemma, let T stand for just the first map.

Define a map $S : H^{k-1}_{cv}(R^{n-1}) \to H^k_{cv}(R^{n-1} \times R)$ going backward as follows. We define it on the form level and then it will induce a map on the cohomology level. Define

$$S : \omega(x_1, \ldots, x_m, t_1, \ldots, t_{n-1}) \to b(t_n)\ \omega(x_1, \ldots, x_m, t_1, \ldots, t_{n-1}) dt_n$$

Here $b(t)$ is a C^∞ function with support in $(-\epsilon, \epsilon)$ and total integral 1. The form ω includes a wedge product of various dx_i and dt_j on both sides of the formula; we have omitted these to save space. The final dt_n is added as a wedge product from the extreme right.

We claim that $dS = Sd$ up to a sign, and thus that S induces a map in cohomology. This is easily checked. Both $Sd\omega$ and $dS\omega$ contain terms obtained by differentiating ω by some x_j or by some t_k for $k < n$, but it doesn't matter whether we multiply by $b(t_n)$ before or after the differentiation. In the expression dS we could also differentiate by t_n but this term vanishes since it would introduce a second dt_n.

We also claim that TS is the identity. Indeed, TS removes the final dt_n and replaces ω by

$$\int_{-\infty}^{\infty} b(u)\ \omega(x_1, \ldots, x_m, t_1, \ldots, t_{n-1}) du$$

and this is just ω because the integral of $b(u)$ is 1.

Finally we must prove that ST is the identity. This is harder because it is false on the form level and only holds on the cohomological level.

Define $K : \Lambda^k_{cv}(R^{n-1} \times R) \to \Lambda^{k-1}_{cv}(R^{n-1} \times R)$ as follows. If a term ω in a form has no dt, let $K(\omega) = 0$. If a term has a dt, then recall the function $b(t)$ introduced in the definition of S above, and let $K(\omega)$ be

$$\left(\int_{-\infty}^{t} \omega(x_1, \ldots, t_1, \ldots, u)\ du - \int_{-\infty}^{\infty} \omega(x_1, \ldots, t_1, \ldots, u)\ du \int_{-\infty}^{t} b(u)\ du \right) dx_i \wedge \ldots \wedge dt_j \wedge \ldots$$

where we omit the final dt at the end.

We claim that $dK - Kd = \mathrm{id} - S \circ T$ up to sign. It immediately follows that $S \circ T$ is the identity in cohomology, and thus that S and T are isomorphisms.

First, suppose that a term ω has no dt. Then

$$(dK - Kd)\omega = -(-1)^k K \frac{\partial \omega}{\partial t} dx_{i_1} \wedge \ldots \wedge dt_{i_1} \wedge \ldots \wedge dt =$$

$$(-1)^{k-1} \left(\int_{-\infty}^{t} \frac{\partial \omega}{\partial u} du - \int_{-\infty}^{\infty} \frac{\partial \omega}{\partial u}\ du \int_{-\infty}^{t} b(u) du \right) dx_{i_1} \wedge \ldots \wedge dt_{i_1} \wedge \ldots$$

The term $\int_{-\infty}^{t} \frac{\partial \omega}{\partial u} du$ equals ω because ω vanishes for very negative values. The term $\int_{-\infty}^{\infty} \frac{\partial \omega}{\partial u}$ equals zero because ω has compact support. So the above expression equals

$$(-1)^{k-1} \omega\ dx_{i_1} \wedge \ldots \wedge dt_{i_1} \wedge \ldots = (-1)^{k-1} (\mathrm{id} - S \circ T)(\omega)$$

Finally, suppose a term ω has a dt term. Then $(dK - Kd)\omega$ equals

$$d\left[\left(\int_{-\infty}^{t} \omega(x_1, \ldots, t_1, \ldots, u)\ du - \int_{-\infty}^{\infty} \omega(x_1, \ldots, t_1, \ldots, u)\ du \int_{-\infty}^{t} b(u)\ du\right) dx_i \wedge \ldots \wedge dt_j \wedge \ldots\right]$$

$$-K\left(\sum_j \frac{\partial \omega}{\partial x_j} dx_j \wedge dx_i \wedge \ldots \wedge dt_j \wedge \ldots \wedge dt\right) - K\left(\sum_j \frac{\partial \omega}{\partial t_j} dt_j \wedge dx_i \wedge \ldots \wedge dt_j \wedge \ldots \wedge dt\right)$$

In the very last term above, there is no partial with respect to t_n, i,e., t, because there is already one dt and a second one will give zero.

Notice that when we take d of the first line, differentiating with respect to x_j, we will get K of the second line applied to the initial $\frac{\partial \omega}{\partial x_j}$ terms. So these terms will cancel. Similarly when we compute d of the first line, differentiating with respect to t_j for $j < n$, we will get K of the second line applied to the second $\frac{\partial \omega}{\partial t_j}$ terms.

The remaining terms come from t_n, which we have been calling t. Only the first line will be involved because of the remark immediately following the formula. When we differentiate the first line with respect to t we will get

$$\left(\omega - b(t) \int_{-\infty}^{\infty} \omega(x_1, \ldots, t_{n-1}, u) du\right) dt \wedge dx_{i_1} \wedge \ldots \wedge dt_{j_1} \wedge \ldots$$

Since we are dealing with Λ^k, there are $k-1$ wedge terms to pass over if we want dt to come last, so the result is

$$(-1)^{k-1}\left(\omega - b(t) \int_{-\infty}^{\infty} \omega(x_1, \ldots, t_{n-1}, u) du\right) dx_{i_1} \wedge \ldots \wedge dt_{j_1} \wedge \ldots \wedge dt = (-1)^{k-1}(\mathrm{id} - S \circ T)\omega$$

Proof of Main Theorem, continued: Suppose $\mathcal{U}$ and $\mathcal{V}$ are open sets in M and suppose E is a vector bundle over M. We then have an exact sequence

$$0 \leftarrow \Lambda^k_{cv}(E|_{\mathcal{U}\cap\mathcal{V}}) \xleftarrow{j_1 - j_2} \Lambda^k_{cv}(E|_{\mathcal{U}}) \oplus \Lambda^k_{cv}(E|_{\mathcal{V}}) \xleftarrow{i_1 + i_2} \Lambda^k_{cv}(E|_{\mathcal{U}\cup\mathcal{V}}) \leftarrow 0$$

Each map is obtained by restriction and exactness is trivial except at the left side. But we can find a partition of unity $\varphi_{\mathcal{U}}, \varphi_{\mathcal{V}}$ with sum equal 1 such that $\varphi_{\mathcal{U}}$ is non-zero only on $\mathcal{U}$ and $\varphi_{\mathcal{V}}$ is nonzero only on $\mathcal{V}$ and both are C^∞ on the union. If ω is a form on $\mathcal{U} \cap \mathcal{V}$, then $\omega\ \varphi_{\mathcal{V}}$ is a form on all of $\mathcal{U}$, since it is certainly C^∞ on $\mathcal{U} \cap \mathcal{V}$ and it is identically zero on the rest of $\mathcal{U}$. Similarly $-\omega\ \varphi_{\mathcal{U}}$ is a form on all of $\mathcal{V}$ and the difference of these forms is $\omega(\varphi_{\mathcal{V}} + \varphi_{\mathcal{U}}) = \omega$. (See the intermission on page 30 for details of the argument that both extensions are C^∞ across the boundary.)

As usual, a purely algebraic argument then gives an exact sequence

$$H^{k+1}_{cv}(E|_{\mathcal{U}\cap\mathcal{V}}) \leftarrow H^{k+1}_{cv}(E|_{\mathcal{U}}) \oplus H^{k+1}_{cv}(E|_{\mathcal{V}}) \leftarrow H^{k+1}_{cv}(E|_{\mathcal{U}\cup\mathcal{V}}) \leftarrow H^{k}_{cv}(E|_{\mathcal{U}\cap\mathcal{V}}) \leftarrow H^{k}_{cv}(E|_{\mathcal{U}}) \oplus H^{k}_{cv}(E|_{\mathcal{V}})$$

We can combine this sequence with the corresponding sequence on M, and the integration along the fibre maps T to obtain the diagram below:

$$\begin{tikzcd}
H^{k+1}_{cv}(E|_{\mathcal{U}\cap\mathcal{V}}) \arrow[d, "T"] & \begin{array}{c} H^{k+1}_{cv}(E|_{\mathcal{U}}) \\ \oplus \\ H^{k+1}_{cv}(E|_{\mathcal{V}}) \end{array} \arrow[l] \arrow[d, "T"] & H^{k+1}_{cv}(E|_{\mathcal{U}\cup\mathcal{V}}) \arrow[l] \arrow[d, "T"] & H^{k}_{cv}(E|_{\mathcal{U}\cap\mathcal{V}}) \arrow[l, "D"'] \arrow[d, "T"] & \begin{array}{c} H^{k}_{cv}(E|_{\mathcal{U}}) \\ \oplus \\ H^{k}_{cv}(E|_{\mathcal{V}}) \end{array} \arrow[l] \arrow[d, "T"] \\
H^{k+1-n}(\mathcal{U}\cap\mathcal{V}) & \begin{array}{c} H^{k+1-n}(\mathcal{U}) \\ \oplus \\ H^{k+1-n}(\mathcal{V}) \end{array} \arrow[l] & H^{k+1-n}(\mathcal{U}\cup\mathcal{V}) \arrow[l] & H^{k-n}(\mathcal{U}\cap\mathcal{V}) \arrow[l, "D"'] & \begin{array}{c} H^{k-n}(\mathcal{U}) \\ \oplus \\ H^{k-n}(\mathcal{V}) \end{array} \arrow[l]
\end{tikzcd}$$

This diagram commutes; we supply the details at the end of the proof.

Proof of the main theorem, concluded: In the theorem, we require that M be compact. But we will prove the theorem more generally, for any C^∞ manifold with a finite good cover such that E restricted to each set of the cover is trivial. We prove the theorem by induction on the number of open sets in the cover.

If M is a single good open set, then the first step of the proof shows that T is an isomorphism. Suppose next that M has a good cover with two open sets $\mathcal{U}$ and $\mathcal{V}$. Since this is a good cover, $\mathcal{U}\cap\mathcal{V}$ is diffeomorphic to R^m or empty, and E is trivial over this set because it is trivial over $\mathcal{U}$. So T is an isomorphism on every vertical arrow except the middle one, and by the five lemma it is also an isomorphism in the middle.

Finally, we prove the induction step. Suppose M has a finite good cover with $N+1$ sets. Let $\mathcal{U}$ be the union of the first N sets and let $\mathcal{V}$ be the remaining set, so $M=\mathcal{U}\cup\mathcal{V}$. Notice that $\mathcal{U}$ has a good cover with N sets. To prove the theorem for M, we will apply the five lemma, and we can do that as soon as we prove the theorem for $\mathcal{U}\cap\mathcal{V}$. But if $1\le i\le N$, then $\mathcal{U}_i\cap\mathcal{V}$ is either empty or else diffeomorphic to R^m, so these sets form a good cover of $\mathcal{U}\cap\mathcal{V}$ with at most N sets and the induction hypothesis can be applied to it.

Proof of the theorem; commutativity of the diagram: Integration along the fibre is certainly commutative with restriction maps. So we need only consider the square containing the D maps.

Recall the definition of the D map. We start with an ω defined over $\mathcal{U}\cap\mathcal{V}$ with $d\omega=0$. We form $\varphi_{\mathcal{V}}\,\omega$ an extension to $\mathcal{U}$ and $-\varphi_{\mathcal{U}}\,\omega$ an extension to $\mathcal{V}$. We then compute $d\,(\varphi_{\mathcal{V}}\,\omega)$ and $d\,(-\varphi_{\mathcal{U}}\,\omega)$ on $\mathcal{U}$ and $\mathcal{V}$. These forms agree on $\mathcal{U}\cap\mathcal{V}$ because $d\,(\varphi_{\mathcal{V}}\,\omega)-d\,(-\varphi_{\mathcal{U}}\,\omega)=d(\varphi_{\mathcal{V}}+\varphi_{\mathcal{U}})\,\omega=d\omega=0$ there. Therefore the forms can be glued to form $D\omega$ on $\mathcal{U}\cup\mathcal{V}$.

We want to prove $T \circ D = D \circ T$. Since the arguments for $\varphi_{\mathcal{V}}\, \omega$ and $\varphi_{\mathcal{U}}\, \omega$ are similar, we study the first extension to $\mathcal{U}$. So we must show that $T\left(d(\varphi_{\mathcal{V}}\omega)\right) = d\left(T(\varphi_{\mathcal{V}}\, \omega)\right)$. Since $d\omega = 0$, the left side of this equation is $T\left(d(\varphi_{\mathcal{V}}) \wedge \omega\right)$. By the lemma at the end of the previous section, this equals $d(\varphi_{\mathcal{V}}) \wedge T(\omega)$. Note that $dT(\omega) = Td(\omega) = 0$, so we can write our expression as $d\left(\varphi_{\mathcal{V}} \wedge T(\omega)\right)$. Using the lemma at the end of the previous section again, we obtain $d(T(\varphi_{\mathcal{V}} \wedge \omega))$. Finally, $\varphi_{\mathcal{V}}$ is a function, so $\varphi_{\mathcal{V}} \wedge \omega = \varphi_{\mathcal{V}}\, \omega$.

7.8 The Thom Class

By the Thom isomorphism theorem, $T : H^n_{cv}(E) \to H^0(M)$ is an isomorphism. The group $H^0(M)$ is the group of locally constant functions on M, and contains the function which is identically 1. This element is the image of a unique element $\Phi \in H^n_{cv}(E)$, called the *Thom class of* E.

Theorem 33 *The inverse of the Thom isomorphism* $H^k_{cv}(E) \to H^{k-n}(M)$ *is the map* $\omega \to \omega \wedge \Phi$. *Here by abuse of notation, the second* ω *is the pullback of* ω *using the map* $\pi : E \to M$.

Proof: Since T is an isomorphism, it suffices to prove that $T(\omega \wedge \Phi) = \omega$. But by the lemma at the end of section 7.6, $T(\omega \wedge \Phi) = \omega \wedge T(\Phi)$ and this second element is $\omega \wedge 1 = \omega$.

Theorem 34 *Let* E_p *be a fibre of* E *over* $p \in M$. *The map* $E_p \to E$ *induces an obvious map* $H^n_c(E_p) \leftarrow H^n_{cv}(E)$. *The Thom class* Φ *maps under this map to the canonical generator of* $H^n_c(E_p)$. *Conversely, if an element* Φ *has this property for each* $p \in M$, *then that element is the Thom class.*

Remark: This theorem uses the orientation on E_p. The canonical element is determined by integration of n-forms with compact support, and represented by a "bump form" with integral one.

The proof in one direction follows because $T(\Phi)$ is given by integration over $t_1, \ldots, t_n$, and Φ maps to the function constantly equal to 1 under this map. Conversely, if $T(\Phi)$ equals 1 at each p, then T maps Φ to the generator of $H^0(M)$ and thus Φ is the Thom class.

7.9 The Thom Class and the Poincare Dual of a Submanifold

Theorem 35 *Let K be a compact oriented submanifold of a compact oriented Riemannian manifold M. Suppose K has dimension k and M has dimension m. Let $\pi : N \to K$ be the normal bundle of K. Thus N_p consists of all tangent vectors to M at p which are perpendicular to the tangent space of K at p.*

Give N an orientation as follows: a basis $e_1, \ldots, e_{m-k}$ of E_p is oriented if $v_1, \ldots, v_k$ is an oriented basis of $T_p(K)$ and $v_1, \ldots, v_k, e_1, \ldots, e_{m-k}$ is an oriented basis of $T_p(M)$.

Let $\mathcal{T}$ be a diffeomorphism from an ϵ neighborhood of the zero section of N to an open tubbular neighborhood of K in M and let Φ denote an $m-k$ form which represents the Thom class of N. By abuse of notation, use the same symbol for $\Phi \in H^{m-k}_{cv}(N)$ and $H^{m-k}(M)$. Then Φ is a representative of the cohomology class dual to K. In particular, this class has a representative which is non-zero only within ϵ of K.

Proof: Since $\mathcal{T}$ is a diffeomorphism, we can ignore it in the proof and work entirely in the normal bundle N. If ω is a k form on M, we must prove that

$$\int_K \omega = \int_M \omega \wedge \Phi$$

But the map $\omega \to \omega \wedge \Phi$ is the inverse of the Thom isomorphism theorem, so $T(\omega \wedge \Phi) = \omega$. Notice that T is defined by integration of the t_j variables on the normal bundle. The complete integral of $\omega \wedge \Phi$ on M is this integral followed by the integral of ω on K. QED.

More Precise Proof: The previous proof gives the central idea, but not enough attention was paid to the domains of the various objects. So it is not a rigorous proof. Here is the correct proof.

We start with a closed k-form on M. Restrict this form to the tubular neighborhood and use the diffeomorphism from N to this neighborhood to restrict the form to an ϵ neighborhood of the zero section in the normal bundle N. The resulting form need not vanish near the boundary of this neighborhood, so it does not define an element of $H^k_{cv}(N)$.

Let $i : K \to N$ be injection as the zero section and form $\tau = i^\star(\omega)$, a closed k-form on K. Let $\pi : N \to K$ and consider $\pi^\star(\tau)$. This is a closed form defined on all of N. Therefore, both ω and $\pi^\star(\tau)$ are closed k-forms on the set of vectors in N of length less than ϵ. Call this set N_ϵ.

However, the identity map on this set of vectors is homotopic to the map $i \circ \pi : N_\epsilon \to N_\epsilon$, by pushing vectors along themselves back to the origin. So ω and $(i \circ \pi)^\star(\omega) = \pi^\star \circ i^\star(\omega) =$

$\pi^\star(\tau)$ induce the same element of the cohomology group $H^k(N_\epsilon)$, and we can find a $k-1$ form σ such that $\omega = \pi^\star(\tau) + d\sigma$ on N_ϵ.

Notice that $\omega \wedge \Phi = (\pi^\star(\tau) + d\sigma) \wedge \Phi = \pi^\star(\tau) \wedge \Phi + d(\sigma \wedge \Phi)$ on N_ϵ, and all of these forms have compact supports on the fibres. So

$$\int_M \omega \wedge \Phi = \int_M \pi^\star(\tau) \wedge \Phi$$

The coefficients of ω may depend on both the x_i coordinates for L and the t_j coordinates for N, but the coefficients of $\pi^\star(\tau)$ depend only on the x_i. So computing $\int_M \pi^\star(\tau) \wedge \Phi$ is a matter of integrating over the t_j and then integrating over the x_i. Integration over the t_j is the same thing as $T(\pi^\star(\tau) \wedge \Phi)$, and since these operations are inverse, the result is just τ. Integrating over the x_i completes the integral, so

$$\int_M \omega \wedge \Phi = \int_M \pi^\star(\tau) \wedge \Phi = \int_K \tau = \int_K i^\star \omega$$

and this is the formula for the Poincare dual of L.

7.10 The Normal Bundle of a Transverse Intersection

Theorem 36 *Suppose K and L are oriented compact submanifolds of an oriented compact submanifold M, and suppose K and L intersect transversally. Then the intersection is a submanifold $K \cap L$. Let N_K and N_L be the normal bundles of K and L, and restrict these bundles to $K \cap L$. Then the normal bundle of $K \cap L$ is $N_K \oplus N_L$.*

Proof: Since $K \cap L \subset K$, $T_p(K \cap L) \subset T_p(K)$. An element of $N_p(K)$ is perpendicular to $T_p(K)$ and thus to $T_p(K \cap L)$. It follows that $N_p(K) + N_p(L) \subset N_p(K \cap L)$.

Next, $N_p(K) \cap N_p(L) = (0)$, for otherwise there is a non-zero vector perpendicular to both $T_p(K)$ and $T_p(L)$ and so perpendicular to $T_p(K) + T_p(L)$. But the intersection at p is transverse, so the sum of these spaces is $T_p(M)$ and no non-zero vector can be perpendicular to everything. So $N_p(K) + N_p(L)$ is a direct sum.

In particular, the dimension of $N_p(K) + N_p(L)$ is the sum of the dimensions of these spaces, and thus $(m-k) + (m-l) = 2m - (k+l) = m - (k+l-m)$. Since the dimension of $K \cap L$ is $(k+l) - m$, $N_p(K) + N_p(L)$ is the full space of normal vectors to $K \cap L$.

7.11 The Thom Class of a Direct Sum

Theorem 37 *Let E and F be oriented vector bundles over a compact M. Let Φ_E and Φ_F be the Thom classes of E and F in $H^k_{cv}(E)$ and $H^l_{cv}(F)$. Give $E \oplus F$ the natural orientation, so if $e_1, \ldots, e_k$ is an oriented basis of E and $f_1, \ldots, f_l$ is an oriented basis of F, then $e_1, \ldots, e_k, f_1, \ldots, f_l$ is an oriented basis of $E \oplus F$. Let $\pi_E : E \oplus F \to E$ and $\pi_F : E \oplus F \to F$ be natural projections. Then the Thom class of $E \oplus F$ is $\pi^\star_E(\Phi_E) \wedge \pi^\star_F(\Phi_F)$.*

Proof: By theorem 34, it suffices to prove that $\pi^\star_E(\Phi_E) \wedge \pi^\star_F(\Phi_F)$ restricted to a fiber generates $H^n_{cp}(E \oplus F)$. We already know that separately for each piece, so the theorem boils down to the assertion that $H^k_{cp}(R^k) \times H^l_{cp}(R^l) \xrightarrow{\wedge} H^{k+l}_{cp}(R^{k+l})$ maps generators to a generator. But the first group on the left is generated by any $b(x_1, \ldots, x_k) dx_1 \wedge \ldots \wedge dx_k$ where b has compact support and $\int \ldots \int b = 1$ and $dx_1, \ldots, dx_k$ is an oriented basis, and the second group has a similar generator, and the wedge of these is an expression in $k + l$ variables with the same properties.

7.12 The Fundamental Theorem of Intersection Theory

Theorem 38 *Let K and L be compact oriented submanifolds of a compact oriented manifold M. Suppose K and L intersect transversally. Let d_K and d_L be the dual forms to K and L. Then $K \cap L$ is a compact oriented submanifold of M and its dual form is $d_K \wedge d_L$.*

Proof: Let N_K and N_L be the normal bundles of K and L. Each has a natural orientation. For instance, if $x_1, \ldots, x_k$ is an oriented coordinate system on K near p, we call a basis of N_K oriented is $x_1, \ldots, x_k, n_1, \ldots, n_{m-k}$ gives the orientation on M.

By theorem 36, the normal bundle to $K \cap L$ is $N_K \oplus N_L$. We give this direct sum the orientation described in theorem 37, and then we orient $K \cap L$ so the pair $K \cap L$ and $N_K \oplus N_L$ satisfies the rule of the previous paragraph.

By section 7.8, the Thom Class Φ_K represents the Poincare Dual d_K. More precisely, we can find $\epsilon > 0$ such that the exponential map defined on normal vectors of length less than ϵ is a diffeomorphism to a tubular neighborhood of K, and we can find a representative of Φ_K with support on vectors of length less than or equal to δ for some $\delta < \epsilon$, and the diffeomorphism then carries Φ_K to a representative of d_K.

A similar statement holds for L and by shrinking ϵ if necessary, we can pick an ϵ which works for both K and L.

By shrinking this ϵ if necessary, we can make the exponential map defined on normal vectors to $K \cap L$ of length less than 2ϵ be a diffeomorphism to a tubular neighborhood of $K \cap L$.

Now consider $\pi^\star_K(\Phi_K)$, a form defined on $N_K \oplus N_L$ over $K \cap L$. This form only depends on the coordinates of K and N_K and is independent of the other coordinates. Notice that it need not have compact support on fibres because it is independent of the N_L component.

However, $\phi_K^\star(\Phi_K) \wedge \phi_L^\star(\Phi L)$, which we abbreviate $\Phi_K \wedge \Phi_L$, does have compact support on fibres. It is nonzero on a normal vector $X + Y \in N_K \oplus N_L$ only if $||X|| < \epsilon$ and $||Y|| < \epsilon$. So $||X + Y|| < 2\epsilon$ and consequently we can move the form over to the tubular neighborhood of $K \cap L$. By theorem 37, $\Phi_K \wedge \Phi_L$ represents the Thom class of $N_K \oplus N_L$ and thus its representative in the tubular neighborhood represents $d_{K\cap L}$. Since Φ_K on the tubular neighborhood of K represents d_K and Φ_L on the tubular neighborhood of L represents d_L, it follows that $d_K \wedge d_L = d_{K\cap L}$. QED.

7.13 Cautionary Notes

The theorem in the previous section almost completes the proof of the hard Lefshetz Fixed Point Formula, and supplies the central result for cohomological intersection theory. We add a series of cautionary notes about the previous argument.

Cautionary Note One: The fundamental idea of the proof is that a neighborhood of the zero section in the normal bundle for K, N_K, is diffeomorphic to a tubular neighborhood of K in M and thus to an open subset of M. Therefore, forms on N_K with support within ϵ of the zero section are equivalent to forms on M with support in the tubular neighborhood. One such form is Φ_K, the Thom Class of K. This form appears to be just the constant function on K multiplied by a bump function times $dt_1 \wedge \ldots \wedge dt_{m-k}$ on fibres. When it is moved to the tubular neighborhood, it becomes the dual class d_K.

Why not simply define the form this way to begin with, rather than using our complicated argument using the Meyer-Vietorius sequence and the Thom isomorphism? Because the K component of our form is *not* just a constant function; it also has terms coming from the change of coordinates of $dt_1 \wedge \ldots \wedge dt_{m-k}$ as we move from coordinate system to coordinate system, and these terms depend on $x_1, \ldots, x_m$. These terms are missing when we begin the proof of the Meyer-Vietoris sequence by dealing with a single open set $\mathcal{U} \subset M$, but they appear as soon as we add in other sets and thus need to change coordinates.

The first spot where this matters came when we defined T mapping forms on M to forms on K by integrating over the t_j. Go back to that spot and notice the matrix ρ_{ji}. Luckily, integration is invariant under this coordinate change.

Remark: In algebraic topology, cohomology classes of closed forms are determined by the integrals of these forms over certain elements of homology. One way to make this precise is via the universal coefficient theorem of singular theory. According to this theorem, if R is the field of real numbers, the singular cohomology group $H^k(M, R)$ is canonically isomorphic to $Hom(H_k(M,Z),R)$. In some sense, this isomorphism is induced by integration.

One way to make this precise is to define a C^∞ singular simplex to be a continuous singular simplex which is C^∞ in an open neighborhood of the simplex (in parameter space). There is an obvious inclusion map $C_k^\infty(M) \to C_k(M)$ from the set of all C^∞ singular chains on a C^∞ manifold to the set of all continuous singular chains. These inclusion maps induce isomorphisms in homology. So for a C^∞ manifold, we can assume that singular homology classes are defined by C^∞ simplices.

We can integrate k-forms over C^∞ k-simplicies. Integrating $d\omega$ over σ is the same as integrating ω over the boundary of σ, by Stokes' formula. It follows that this integration induces a map $H^k_{de-Rham}(M) \to Hom(H_k(M,Z),R)$. This map turns out to be an isomorphism. Another way of stating this isomorphism is as follows:

Theorem 39 *Let $L_1,\ldots,L_s$ be a basis for the free component of $H_k(M)$, that is, the portion of this homology isomorphic to $Z \oplus Z \oplus \ldots \oplus Z$. Then two closed k forms ω_1 and ω_2 represent the same cohomology element if and only if $\int_{L_i}\omega_1 = \int_{L_i}\omega_2$ for all i. Moreover, if arbitrary real numbers $r_1,\ldots,r_s$ are given, there is a unique element of the deRham cohomology group $H^k(M)$ represented by a k-form ω whose integral over L_i is r_i for each i.*

A consequence of this result is that deRham cohomology groups are isomorphic to singular cohomology groups.

Remark: Since we do not need this isomorphism to prove the fixed point theorem, we omitted the proof from these notes. Nevertheless, the many integrals of forms over embedded submanifolds $K \subset M$ in the notes show that the theorem is closely related to the isomorphism.

It is natural to ask two related questions about homology classes in $H_k(M,Z)$ where M is a C^∞ manifold:

- Which classes are represented by a compact oriented manifold K and arbitrary C^∞ map $i : K \to M$? Here we do not require that i be one-to-one or an embedding; to avoid confusion, call such classes "Steenrod representable."
- Which classes of $H_k(M,Z)$ are represented by embedded compact oriented manifolds?

Thom applied his construction of the Thom class to these questions. Among the known results are:

- If an element of $H_k(M,Z)$ is Steenrod representable and $2k < dim(M)$, then this element can be represented by an embedded submanifold.

- If M is orientable, then every class in $H_{m-1}(M, Z)$ and every class in $H_{m-2}(M, Z)$ is Steenrod representable.
- Every class in $H_k(M)$ for $k \leq 6$ is Steenrod representable.
- If $k \geq 7$, there are M and classes in $H_k(M)$ which are not Steenrod representable.
- Every class in $H_k(M, Z_2)$ is Steenrod representable.
- Every class in $H_k(M, Z)$ has a positive multiple which is Steenrod repressentable.

Cautionary Note 2: If a homology class in M is represented by an embedded submanifold K, then it induces a dual cohomology class $d_K \in H^{m-k}(M)$ which has support in an ϵ neighborhood of K. Cohomology classes are determined by their integrals over homology classes, which are often submanifolds. It is easy to confuse these ideas and begin thinking of K as directly representing a form in $H^k(M)$, perhaps a form whose integral over K is one.

Consider the following example:

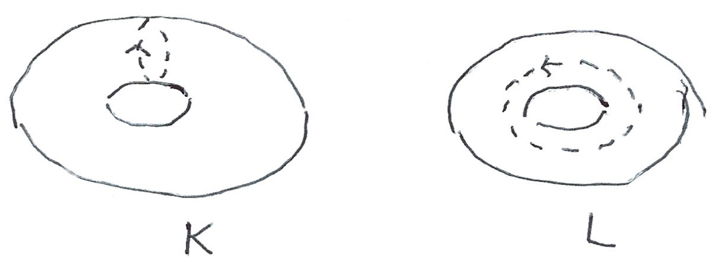

Figure 7.2: K and L in a Torus

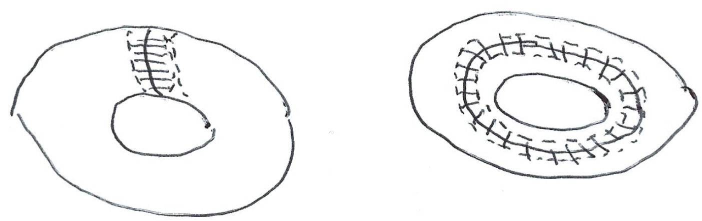

Figure 7.3: d_K and d_L

The first homology group of a torus has two generators, illustrated in the first picture above. The dual classes of these generators are illustrated in the second picture. Suppose we integrate d_K over K, perhaps under the false illusion that d_K is the unique form whose integral over K is 1 and over L is 0. In most cases this would not even make sense because d_K has degree $m-k$ and K has dimension k. But in this special case $m-k = 2-1 = 1 = k$.

Recall, however, that integration is a homotopy invariant. We could just slide K along the torus until the support of d_K no longer intersects K, and then clearly $\int_K d_K = 0$.

The truth is that $\int_L d_K \neq 0$ because any homotopy of L will still intersect the support of d_K. Indeed,

$$\int_L d_K = \int_M d_K \wedge d_L = \int_M d_{K \cap L} = 1$$

because $K \cap L$ is just a single point.

Remark: This special case can be greatly generalized. Suppose the dimension of M is even. Then there is a middle cohomology class $H^k(M)$ for $k = \frac{m}{2}$, and Poincare duality asserts that $\int_M \omega \wedge \tau$ is a non-degenerate form on this vector space.

If $k \equiv 2 \pmod 4$, then k^2 is odd and this form is skew-symmetric. By easy algebra, if a vector space has a non-degenerate skew form F, it is even dimensional and has a basis $b_1, b_2, \ldots, b_{2s}$ such that the matrix of F is

$$\begin{pmatrix} 0 & -1 & & & & & \\ 1 & 0 & & & & & \\ & & 0 & -1 & & & \\ & & 1 & 0 & & & \\ & & & & \ddots & & \\ & & & & & 0 & -1 \\ & & & & & 1 & 0 \end{pmatrix}$$

For instance, this theorem applies to tori with an arbitrary number of holes, showing that $H_1(T)$ can be generated by curves $C_1, C_2, \ldots, C_{2s-1}, C_{2s}$ such that the first pair $\{C_1, C_2\}$ intersect each other once and do not intersect the remaining C_i, and so forth for the remaining pairs. See the picture below.

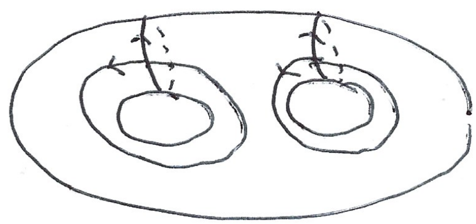

Figure 7.4: H_1(Two Torus)

If $k \equiv 0 \pmod 4$, then k^2 is even and the above form is symmetric. In that case, there is a basis $b_1, \ldots, b_k, b_{k+1}, \ldots, b_m$ such that the matrix for our symmetric non-degenerate form is as follows:

$$\begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & -1 & & \\ & & & & \ddots & \\ & & & & & -1 \end{pmatrix}$$

By definition, the *signature of M* is the number of 1's on the diagonal minus the number of -1's on the diagonal. This signature plays an important role on modern topology of manifolds.

Notice that the existence of 1's on the diagonal implies that $\int_K d_K \neq 0$ for some K in this case.

Chapter 8

Orientations

Several results in this report depend on orientations on L, M, and M. We collect here all spots where orientations matter.

Consider first the Lefshetz number itself: $L(f) = \sum(-1)^k tr\left[H^k(M) \leftarrow H^k(M)\right]$. Nothing in this formula depends on an orientation. The Lefshetz Fixed Point Theorem will claim that $L(f) = \sum_{\text{fixed points}} sign(f_p)$. These signs may depend on orientations, but the choices of these orientations must be canonical.

Orientations first appear in section 3.7, when we define an orientation on a manifold, and use it to integrate m-forms on an m-dimensional M. In particular, $H^m(M) \cong R$, and the isomorphism depends on the choice of orientation.

The Poincare Duality theorem depends on the pairing

$$H^k(M) \times H^{m-k}(M) \xrightarrow{\wedge} H^m(M) \xrightarrow{\int_M \omega} R$$

and thus on the choice of orientation on M. Therefore the map

$$Hom(H^k(M), R) \to H^{m-k}(M)$$

depends on the orientation of M.

Next we considered an oriented submanifold K of an oriented manifold M. Both orientations can be selected arbitrarily. But then $i : K \to M$ induces $i^\star : H^k(K) \leftarrow H^k(M)$ and integration over K defines a further map to R. This map certainly depends on the orientation of K. The map induces an element of $Hom(H^k(M), R)$, and by duality this gives an element $d_K \in H^{m-k}(M)$. Notice that this element depends on both the orientation of K and the orientation of M.

Next we studied tubular neighborhoods, the normal bundle, and the relation between Φ_K and d_K. This theory depends on an orientation on K and an orientation on the normal bundle. (Recall that Thom's theorem is about oriented vector bundles). We start with an arbitrary orientation on K. Find an oriented basis of $T_p(K)$, $\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_k}$. Define an orientation on N_K by calling a basis $N_1, \dots, N_{m-k}$ of N_p *oriented* if

$$\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_k}, N_1, \dots, N_{m-k}$$

is an oriented basis of $T_p(M)$.

Suppose next that K and L intersect transversally, and that both have arbitrary orientations. We proved that $K \cap L$ is a submanifold with normal bundle $N_K \oplus N_L$. Each of these is oriented by the previous paragraph, and their sum is thus oriented by calling an oriented basis of $N_p(K)$ followed by an oriented basis of $N_p(L)$ *oriented for* $N_K \oplus N_L$. Notice that this orientation depends on the order of K and L. Use this to define an orientation on $K \cap L$ by running the argument of the previous paragraph backward. Thus an oriented basis of $T_p(M)$ is obtained by starting with an oriented basis of $T_p(K \cap L)$ and extending it using an oriented basis of $N_p(K) \oplus N_p(L)$. Notice that this is exactly the orientation used in the proof that $d_{K \cap L} = d_K \wedge d_L$.

In the special case of the Lefshetz Fixed Point formula, we begin with an oriented manifold M. We then form $M \times M$, which inherits a natural orientation from M. Indeed if $\mathcal{U}$ and $\mathcal{V}$ are oriented coordinates on M, then $\mathcal{U} \times \mathcal{V}$ is an oriented coordinate on $M \times M$, obtained by first listing the coordinates of $\mathcal{U}$ and then listing the coordinates of $\mathcal{V}$.

We then form $K = G$ and $L = \Delta$, submanifolds of $M \times M$. These inherit natural orientations from M via $i : M \to G$ by $p \to (p, f(p))$ and $p \to \Delta$ by $p \to (p, p)$. We write G first and Δ second in the formula $d_G \wedge d_\Delta$ because that choice gives the correct value for the Lefshetz number of f.

In the final section which follows, we will compute the sign of f at fixed points based on these orientation choices.

Chapter 9

The Lefshetz Fixed Point Theorem (2)

9.1 Two Methods to Compute Signs of Fixed Points

Suppose K and L are compact, oriented submanifolds of a compact, oriented manifold M, and assume that $dim(K) + dim(L) = dim(M)$. If K and L meet transversally, $K \cap L$ is a finite collection of points, each assigned an orientation 1 or -1.

In this case, there are two equivalent methods to calculate the $sign(p)$ assigned to an intersection point. The first is straightforward and will be used in the following section. Select an oriented basis $X_1, \ldots, X_k$ for $T_p(K)$. Select a similar oriented basis $X_{k+1}, \ldots, X_m$ for $T_p(Y)$. Since the manifolds meet transversally at p, $T_p(K) \oplus T_p(L) = T_p(M)$, so $X_1, \ldots, X_m$ is a basis for $T_p(M)$. If this basis gives the orientation of M, we assign $sign(p) = 1$. Otherwise $sign(p) = -1$.

The second method of computing $sign(p)$ generalizes to the case when K, L, and M have arbitrary dimensions, and was used in the proof that $d_{K \cap L} = d_K \wedge d_L$. Determine spaces of normal vectors N_K and N_L to K and L. Then $T_p(K) \oplus N_K = T_p(M)$. Give N_K the orientation determined by the requirement that an oriented basis for $T_p(K)$, followed by an oriented basis of N_K, is an oriented basis of $T_p(M)$. Orient N_L using the same rule. Since the intersection is transversal, $N_K \oplus N_L = T_p(M)$. If an oriented basis for N_K, followed by an oriented basis of N_L, gives an oriented basis of $T_p(M)$, assign $sign(p) = 1$. Otherwise assign $sign(p) = -1$.

Since we want to compute $sign(p)$ using the first method, but apply it to our theory developed using the second method, we need to prove

Theorem 40 *Both methods give the same result.*

Proof: We first claim that we can find coordinates $x_1, \ldots, x_m$ near p on M such that K consists of points where $x_{k+1} = \ldots = x_m = 0$ and L consists of points where $x_1 = \ldots = x_k = 0$. If so, then we can use $x_1, \ldots, x_k$ as coordinates on K and use $x_{k+1}, \ldots, x_m$ as coordinates on L. By changing the signs of some x_i if necessary, we can assume that $x_1, \ldots, x_k$ is an oriented basis of K and $x_{k+1}, \ldots, x_m$ is an oriented basis of L. Notice that we have lost the freedom to make a final adjustment for the coordinates of M, so the full set of coordinates may or may not be oriented for M.

We claim that we can find a Riemannian metric near p such that the $\frac{\partial}{\partial x_i}$ are orthonormal. We respect to that metric, N_K and N_L come with natural ordered bases. For N_K we have $\frac{\partial}{\partial x_{k+1}}, \ldots, \frac{\partial}{\partial x_m}$ and for N_L we have $\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_k}$

Thus our coordinates give natural orientations to $T_p(K), T_p(L), T_p(M), N_K, N_L$. The trouble is that these natural orientations may be wrong for $T_p(M), N_K, N_L$. To distinguish between orientations, we will use the terminal “correct orientation” for the orientation determined by the theory of intersections, in contrast to the “natural orientation” determined by these coordinates.

There is one additional problem. We want an oriented basis of N_K, followed by an oriented basis of N_L to give an oriented basis for $T(M)$, but that is not necessarily true even for our natural orientation. To make it true, we must switch these two bases, changing the sign of the orientation of M by $(-1)^{k(m-k)}$.

Three results guide our juggling of all these orientations. First, we want an oriented basis of $T(K)$, followed by an oriented basis of N_K, to give the correct orientation of $T(M)$.

Lemma 15 *The natural orientation and the correct orientation for N_K agree if and only if the natural orientation for $T(M)$ is the correct orientation.*

Next consider the analogous situation for L. This time there is a complication, because the natural basis vectors for $T(L)$ are $\frac{\partial}{\partial x_i}$ for $i > k$ and the natural basis for N_L are these partials for $i \leq k$, so we must switch, introducing a sign change of $(-1)^{k(m-k)}$ in the natural basis for $T(M)$. So

Lemma 16 *The natural orientation of N_L is the correct orientation if and only if either $k(m-k)$ is even and the natural orientation of $T(M)$ is the correct orientation, or else $k(m-k)$ is odd and the natural orientation of $T(M)$ is not the correct orientation.*

Finally there is that switch because the natural basis vectors of N_K follow the natural basis vectors of N_L rather than the desired order, so

Lemma 17 *The natural orientation of $T(M)$ is the orientation obtained by writing a naturally oriented basis of N_K followed by a naturally oriented basis of N_L if and only if $k(m-k)$ is even.*

Let us put all this together. Suppose first that $k(m-k)$ is even. If the natural orientation of $T_p(M)$ is the correct orientation, then the natural orientations of N_K and N_L are both the correct orientations and an oriented basis of N_K followed by an oriented basis of N_L gives the natural orientation of M and thus the correct orientation of M, so $sign(p) = 1$.

If $k(m-k)$ is even, but the natural orientation of $T_p(M)$ is not the correct orientation, then both N_K and N_L have incorrect orientations, so if we write the naturally oriented basis of N_K followed by the naturally oriented basis of N_L, we actually get the same correct orientation as if we had used a correctly oriented basis of N_K followed by a correctly oriented basis of N_L, because there are two hidden sign changes there. On the other hand, we also get the naturally oriented basis of N_K, followed by the naturally oriented basis of N_L, so we get the naturally oriented basis of $T_p(M)$, which is not the correct orientation of $T_p(M)$. So $sign(p) = -1$.

Now suppose $k(m-k)$ is odd, but the natural orientation of $T_p(M)$ is the correct orientation. Then the natural orientation of N_K is the correct orientation, but the natural orientation of N_L is not the correct orientation. If we write the naturally oriented basis of N_K followed by the naturally oriented basis of N_L, we do not get the naturally oriented basis for $T_p(M)$ because $k(m-k)$ is odd. So we do not get the correct orientation on $T_p(M)$. However, if we change the sign of one of the basis vectors of N_L, then N_L will have the correct orientation and we will get the naturally oriented basis of $T_p(M)$, which is the correct basis of $T_p(M)$. So $sign(p) = 1$.

Finally, suppose $k(m-k)$ is odd and the natural orientation of $T_p(M)$ is not the correct orientation. Then the natural orientation of N_K is not the correct orientation, but the natural orientation of N_L is the correct orientation. Also if we write the naturally oriented basis of N_K followed by the naturally oriented basis of N_L, we do not get the naturally oriented basis of $T_p(M)$ because $k(m-k)$ is odd. Change the sign of one of the oriented basis vectors of N_K. Then both N_K and N_L are correctly oriented, and the new basis of N_K, followed by the natural oriented basis of N_L is the natural oriented basis of $T_p(M)$, and thus not the correct oriented basis of $T_p(M)$. So $sign(p) = -1$.

At the start of this proof, we made two assertions without proper proof and discussion. To complete the proof, we fill this gap.

Lemma 18 *Suppose K and L are submanifolds of M which meet transversally at p. Suppose their dimensions are k and l and $k + l = m$, the dimension of M. Then near p we can find local coordinates $x_1, \ldots, x_m$ such that K is given by $x_{k+1} = \ldots = x_m = 0$ and L is given by $x_1, \ldots, x_k = 0$.*

Proof: Let $i : K \to M$ be the map immersing K in M, and let $j : L \to M$ be the map immersing L. At $p \in K \cap L$, we have $i_\star(T_p(K)) \subset T_p(M)$ and $j_\star(T_p(L)) \subset T_p(M)$. Since this is a transverse intersection and $dim(K) + dim(L) = dim(M)$, we have

$$T_p(K) \oplus T_p(L) = T_p(M)$$

Select coordinates near $p \in M$, $x_1, \ldots, x_m$, such that p is the origin and $i_\star(T_p(K))$ has basis $\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_k}$ and $j_\star(T_p(L))$ has basis $\frac{\partial}{\partial x_{k+1}}, \ldots, \frac{\partial}{\partial x_m}$.

If $y_1, \ldots, y_k$ are local coordinates for K near p, consider the map

$$(y_1, \ldots, y_k) \to (i_1(y_1, \ldots, y_k), \ldots, i_k(y_1, \ldots, y_k)) : K \to R^k$$

By assumption, the derivative of this map at p is onto, and thus an isomorphism at p. Apply the inverse function theorem to obtain new local coordinates $(x_1, \ldots, x_k)$ on K near p., so the map i has the form $(x_1, \ldots, x_k) \to (x_1, \ldots, x_k, i_{k+1}(x_1, \ldots, x_k), \ldots, i_m(x_1, \ldots, x_k))$.

Define a map $\psi : R^m \to R^m$ by

$$(x_1, \ldots, x_k, x_{k+1}, \ldots x_n) \to (x_1, \ldots, x_k, x_{k+1} - i_{k+1}(x_1, \ldots, x_k), \ldots, x_m - i_m(x_1, \ldots, x_k))$$

$$= (y_1, \ldots, y_k, y_{k+1}, \ldots, y_m)$$

The derivative of this map is an isomorphism at p, and thus by the inverse function theorem we obtain new local coordinates on M near p. Notice that in the new coordinates, K maps to points whose final $m - k$ coordinates are zero. By abuse of notation we still call these coordinates $x_1, \ldots, x_m$ from now on, but now K is defined near p by coordinates whose last $m - k$ values are zero.

Notice that ψ maps $\frac{\partial}{\partial x_i}$ to itself if $i > k$, so the map j still maps $T_p(L)$ to the space spanned by these vectors.

The map $j : L \to M$ has coordinate form $j_1(z_1, \ldots, z_{m-k}), \ldots, j_m(z_1, \ldots, z_{m-k})$ and if we restrict to the last $m - k$ coordinates, this map is a local diffeomorphism by the inverse function theorem because of the images of $T_p(L)$. We conclude that the $x_{k+1}, \ldots, x_m$ provide new coordinates for L. With these new coordinates, the map j has the form

$$(x_{k+1}, \ldots, x_m) \to (j_1(x_{k+1}, \ldots, x_m), \ldots, j_k(x_{k+1}, \ldots, x_m), x_{k+1}, \ldots, x_m)$$

Consider the map $M \to M$ defined by

$$(x_1, \ldots, x_m) \to (x_1 - j_1(x_{k+1}, \ldots, x_m), \ldots x_k - j_k(x_{k+1}, \ldots, x_m), x_{k+1}, \ldots, x_m)$$

The derivative of this map at p is an isomorphism, so it defines new coordinates

$$(y_1, \ldots, y_k, y_{k+1}, \ldots, y_m)$$

on M. From now on we call these $(x_1, \ldots, x_m)$ by abuse of notation. Note that the $y's$ equal the $x"s$ except for the first k, so points in K still correspond to points with the last $m-k$ values zero. But now points in L also correspond to points with the first k values zero. QED.

Remark: The only other shaky point in the argument is the introduction of new inner products on $T_(M)$ near p, which could have the consequence of distorting the appearance of the normal vectors and tubular neighborhood in unexpected ways. In the end we are only interested in the orientation of our normal bundles, and we will justify this step by redefining the normal bundle in a way that is independent of the inner product, and showing that we can still orient this redefined bundle. Note that N_K is a bundle on K. At any point $q \in K$ we have an exact sequence

$$0 \to T_q(K) \to T_q(M) \to T_q(M)/T_q(K) \to 0$$

Our new normal bundle assigns to each $q \in K$ the quotient vector space $T_q(M)/T_q(K)$. If we have an inner product on $T_q(M)$, then we can find orthogonal complements and write $T_q(M) = T_q(K) \oplus N_q(K)$ and then there is a canonical isomorphism $T_q(M)/T_q(K) \to N_q(K)$. But the bundle can be defined using these quotients independent of this inner product. Moreover, suppose we have an oriented basis of $T_q(K)$, $v_1, \ldots, v_k$. We can extend this to an oriented basis of $T_q(M)$, $v_1, \ldots, v_k, v_{k+1}, \ldots, v_m$. The last $m-k$ vectors induce a basis of $T_q(M)/T_q(K)$ and define an orientation on this normal bundle.

Thus we conclude that there is no need to even introduce an inner product on $T(M)$ to discuss the normal bundle and its orientation.

9.2 The Lefshetz Fixed Point Theorem

Suppose M is a compact, oriented, C^∞ manifold and let $f : M \to M$ be a C^∞ map. Define $\Delta \subset M \times M$ to be the diagonal, that is, the set of all (p,p) with p in M. Define $G \subset M \times M$ to be the graph of f, that is, the set of all $(p, f(p))$ with p in M. In chapter 5, we proved that the Lefshetz number of f satisfies

$$L(f) = \sum_k (-1)^k \mathrm{Tr}\ f^\star : H^k(M) \leftarrow H^k(M) = \int_{M \times M} d_G \wedge d_\Delta$$

In chapter six, we proved that the map f can be replaced by a homotopic map such that G and Δ intersect transversally.

In chapter seven, we proved that in that case, the intersection is a zero-dimensional compact, oriented manifold and

$$\int_{M \times M} d_G \wedge d_\Delta = \int_{M \times M} d_{G \cap \Delta} = \sum_{\text{fixed points } p} \pm \int_{M \times M} d_p = \sum_{\text{fixed points } p} \pm 1$$

To finish the proof, it suffices to compute these signs. We will use the first method of the previous section.

Pick an oriented coordinate system $(x_1, \ldots, x_m)$ on M, where these coordinates are defined on an open $\mathcal{U}$ in M. Pick a second oriented coordinate system $(y_1, \ldots, y_m)$ on M, where these coordinates are defined on an open $\mathcal{V}$ in M. The combination of these coordinates gives a coordinate system on $\mathcal{U} \times \mathcal{V}$ in $M \times M : (x_1, \ldots, x_m, y_1, \ldots, y_m)$.

In particular, if we want coordinates near the diagonal Δ, we can pick the same coordinate system twice, so $(p, q) \in \mathcal{U} \times \mathcal{U}$ is given by $(x_1, \ldots, x_m)$ for p and $(\hat{x}_1, \ldots, \hat{x}_m)$ for q.

We can also form coordinates $\mathcal{U} \times \mathcal{V}$ on $M \times M$ where $\mathcal{V} \subset \mathcal{U}$ consists of points $p \in \mathcal{U}$ for which $f(p) \in \mathcal{U}$. These coordinates along the diagonal certainly contain all fixed points of f.

We have two maps, one from $\mathcal{U}$ to G and one from $\mathcal{U}$ to Δ:

$$(x_1, \ldots, x_m) \to (x_1, \ldots, x_m, f_1(x_1, \ldots, x_m), \ldots, f_m(x_1, \ldots, x_m)$$

$$(x_1, \ldots, x_m) \to (x_1, \ldots, x_m, x_1, \ldots, x_m)$$

The induced maps on tangent spaces map $\frac{\partial}{\partial x_i}$ to

$$\left(0, \ldots, \frac{\partial}{\partial x_i}, \ldots, 0, \sum \frac{\partial f_1}{\partial x_i} \frac{\partial}{\partial y_i}, \ldots, \sum \frac{\partial f_m}{\partial x_i} \frac{\partial}{\partial y_i}\right)$$

$$\left(0, \ldots, \frac{\partial}{\partial x_i}, \ldots, 0, 0, \ldots, \frac{\partial}{\partial y_i}, \ldots, 0\right)$$

The first m vectors in order, followed by the second m vectors in order, gives a basis for $T_p(G) \oplus T_p(\Delta)$ at a fixed point, and the sign assigned to this fixed point is 1 if this is equal to the canonical orientation of $T_p(M \times M)$ at that point, and -1 if it is the opposite, where the canonical orientation of $T_p(M \times M)$ at a diagonal point is given by the m vectors $\frac{\partial}{\partial x_i}$ followed by the m vectors $\frac{\partial}{\partial y_j}$. The matrix converting one basis to the other is

$$\begin{pmatrix}
1 & & 0 & & & \\
& \ddots & & & \frac{\partial f_i}{\partial x_j} & \\
0 & & 1 & & & \\
1 & & 0 & 1 & & 0 \\
& \ddots & & & \ddots & \\
0 & & 1 & 0 & & 1
\end{pmatrix}$$

We must take the determinant of this matrix and find its sign. Before doing that, we can subtract the top m rows from the bottom ones, giving the matrix

$$\begin{pmatrix} 1 & & 0 & 1 & & 0 \\ & \ddots & & & \ddots & \\ 0 & & 1 & 0 & & 1 \\ 0 & & 0 & & & \\ & \ddots & & & I - \left(\frac{\partial f_i}{\partial x_j}\right) & \\ 0 & & 0 & & & \end{pmatrix}$$

We can also subtract the m left columns from the m right columns, giving

$$\begin{pmatrix} 1 & & 0 & 0 & & 0 \\ & \ddots & & & \ddots & \\ 0 & & 1 & 0 & & 0 \\ 0 & & 0 & & & \\ & \ddots & & & I - \left(\frac{\partial f_i}{\partial x_j}\right) & \\ 0 & & 0 & & & \end{pmatrix}$$

The determinant of this expression, and thus the sign of p, is

$$\det\left[I - \left(\frac{\partial f_i}{\partial x_j}\right)\right]$$

Remark: Putting this altogether, we can triumphantly conclude that

Theorem 41 *Let $f : M \to M$ be a C^∞ map from a compact oriented manifold M to itself. We can form the Lefshetz number of f,*

$$L(f) = \sum_k (-1)^k \, Tr \; f^\star : H^k(M) \leftarrow H^k(M)$$

This number is invariant under homotopies of f. It is possible to find an arbitrarily small homotopy of f making the graph of f in $M \times M$ and the diagonal in $M \times M$ meet transversally. After this homotopy, the Lefshetz number is equal to the number of fixed points of f counted with appropriate signs

$$\sum_{\textit{fixed points } p} \textit{sign}(p)$$

where the sign of a fixed point p is the sign of

$$\det\left[I - \left.\left(\frac{\partial f_i}{\partial x_j}\right)\right|_p\right]$$